ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΧΗΜΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

ΤΟΜΕΑΣ ΙΙ: ΑΝΑΛΥΣΗΣ, ΣΧΕΔΙΑΣΜΟΥ & ΑΝΑΠΤΥΞΗΣ ΔΙΕΡΓΑΣΙΩΝ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ

ΜΟΝΑΔΑ ΑΥΤΟΜΑΤΗΣ ΡΥΘΜΙΣΗΣ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΑΝΑΠΤΥΞΗ ΑΛΓΟΡΙΘΜΩΝ ΥΠΟΛΟΓΙΣΤΙΚΗΣ ΝΟΗΜΟΣΥΝΗΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΣΤΗΝ ΕΞΕΛΙΞΗ ΤΟΥ ΠΑΓΚΟΣΜΙΟΥ ΙΣΤΟΥ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΕΥΑΓΓΕΛΙΑ ΧΟΝΔΡΟΔΗΜΑ

ΤΡΙΜΕΛΗΣ ΣΥΜΒΟΥΛΕΥΤΙΚΗ ΕΠΙΤΡΟΠΗ:

Χ. ΣΑΡΙΜΒΕΗΣ, ΚΑΘΗΓΗΤΗΣ ΕΜΠ (ΕΠΙΒΛΕΠΩΝ)
Χ. ΚΥΡΑΝΟΥΔΗΣ, ΚΑΘΗΓΗΤΗΣ ΕΜΠ
Γ. ΜΑΥΡΩΤΑΣ, ΑΝΑΠΛΗΡΩΤΗΣ ΚΑΘΗΓΗΤΗΣ ΕΜΠ

Α Θ Η Ν Α , Ι Ο Υ Λ Ι Ο Σ 2 0 1 7

Η διδακτορική διατριβή εκπονήθηκε με χρηματοδότηση από τον Ειδικό Λογαριασμό Έρευνας του Εθνικού Μετσόβιου Πολυτεχνείου.

.....

Η έγκριση της διδακτορικής διατριβής από την Ανώτατη Σχολή Χημικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου δεν υποδηλώνει αποδοχή των γνωμών του συγγραφέα.
(Ν. 5343/1932, Άρθρο 202)

Στη μητέρα μου,
έναν πολύ δυνατό άνθρωπο,
μία αξιοθαύμαστη γυναίκα,
μία αξιαγάπητη μητέρα,
για όλα όσα μου έχει προσφέρει,
στην οποία οφείλω όλα όσα είμαι, ή ελπίζω να γίνω
και που δεν θα πάψει ποτέ
να με εμπνέει σε ό,τι κάνω στη ζωή μου.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΕΧΟΜΕΝΑ i
ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ iv
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ vii
ΚΑΤΑΛΟΓΟΣ ΑΛΓΟΡΙΘΜΩΝ ix
ΚΑΤΑΛΟΓΟΣ ΣΥΝΤΟΜΟΓΡΑΦΙΩΝ x
ΕΥΧΑΡΙΣΤΙΕΣ xii
ΠΕΡΙΛΗΨΗ xiii
ABSTRACT xv
ΕΙΣΑΓΩΓΗ 1
ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΥΠΟΛΟΓΙΣΤΙΚΗ ΝΟΗΜΟΣΥΝΗ 13
1.1 Υπολογιστική νοημοσύνη 14
1.2 Νευρωνικά δίκτυα 16
1.2.1 Ιστορική αναδρομή 19
1.2.2 Η δομή του ενός νευρώνα 20
1.2.3 Νευρωνικά δίκτυα πολλών-στοιβάδων 24
1.2.4 Παλινδρομικά νευρωνικά δίκτυα 27
1.2.5 Νευρωνικά δίκτυα Kohonen 29
1.2.6 Διαδικασία Εκμάθησης 30
1.2.7 Γενίκευση τεχνητών νευρωνικών δικτύων 33
1.2.8 Προβλήματα μάθησης - Δείκτες απόδοσης μοντέλων τεχνητών νευρωνικών δικτύων 35
1.3 Ασαφής λογική 38
1.4 Εξελικτικός υπολογισμός 39
1.4.1 Ο αλγόριθμος εξελικτικής προσομοιωμένης ανόπτησης 41
1.5 Νοημοσύνη σμήνους 43
1.5.1 Ο αλγόριθμος βελτιστοποίησης σμήνους σωματιδίων 45
1.6 Τεχνητά Ανοσοποιητικά Συστήματα 48
ΚΕΦΑΛΑΙΟ 2. ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ RBF 50
2.1. Η δομή των δικτύων RBF 51

2.2 Εκπαίδευση δικτύων RBF και διαδικασία επιλογής μοντέλου 55
2.2.1 Ο αλγόριθμος *k*-means 59
2.2.2 Η έννοια της ασαφούς διαμέρισης - Ο συμμετρικός αλγόριθμος των ασαφών μέσων..... 61
2.2.3 Μη συμμμετρική διαμέριση – Ο μη συμμετρικός αλγόριθμος των ασαφών μέσων 64
2.3 Σύγκριση δικτύων RBF με άλλες αρχιτεκτονικές νευρωνικών δικτύων..... 68
ΚΕΦΑΛΑΙΟ 3. ΑΛΓΟΡΙΘΜΟΙ ΕΚΠΑΙΔΕΥΣΗΣ ΔΙΚΤΥΩΝ RBF 70
3.1 Εκπαίδευση νευρωνικών δικτύων RBF με τον αλγόριθμο βελτιστοποίησης σμήνους σωματιδίων..... 71
3.1.1 Βελτιστοποίηση της ασαφούς διαμέρισης με τον αλγόριθμο PSO 71
3.1.2 Μέθοδος εκπαίδευσης δικτύων RBF βασισμένη στη βελτιστοποίηση σμήνους.... 73
3.1.3 Αξιολόγηση του αλγορίθμου PSO-NSFM σε σύνολα δεδομένων τύπου benchmark – Αποτελέσματα και συζήτηση 75
3.1.4 Εφαρμογή του αλγορίθμου PSO-NSFM στην αναγνώριση μουσικού είδους – Αποτελέσματα και συζήτηση 83
3.2 Εκπαίδευση νευρωνικών δικτύων RBF με συνεργατική βελτιστοποίηση σμήνους σωματιδίων..... 89
3.2.1 Ενσωμάτωση μεταβλητού εύρους..... 90
3.2.2 Συνεργατικός αλγόριθμος PSO για πλήρη εκπαίδευση δικτύων RBF 92
3.2.3 Αξιολόγηση του αλγορίθμου COOP-VW 100
3.2.4 Αποτελέσματα και συζήτηση 103
3.3 Εκπαίδευση ταξινομητών RBF με τον αλγόριθμο εξελικτικής προσομοιωμένης ανόπτησης..... 111
3.3.1 Βελτιστοποίηση της ασαφούς διαμέρισης με τον αλγόριθμο ESA 111
3.3.2 Εκπαίδευση ταξινομητών RBF με βάση τον αλγόριθμο εξελικτικής προσομοιωμένης ανόπτησης 112
3.3.3 Εφαρμογή του αλγορίθμου ESA-NSFM στην ιατρική διάγνωση 115
3.3.4 Αποτελέσματα και συζήτηση 120
ΚΕΦΑΛΑΙΟ 4. ΚΑΤΗΓΟΡΙΚΑ ΚΑΙ ΜΙΚΤΑ ΔΕΔΟΜΕΝΑ..... 126
4.1 Εισαγωγή στα κατηγορικά δεδομένα 127
4.2 Μέτρα απόστασης για κατηγορικά δεδομένα..... 128
4.3 Νευρωνικά δίκτυα για τον χειρισμό κατηγορικών δεδομένων..... 130
4.3.1 Αλγόριθμος εκπαίδευσης νευρωνικών δικτύων CRBF 132
4.3.2 Επιλογή μοντέλου CRBF 134
4.3.3 Εφαρμογή δικτύων CRBF σε κατηγορικά σύνολα δεδομένων 135
4.3.4 Αποτελέσματα και συζήτηση 138
4.4 Νευρωνικά δίκτυα για τον χειρισμό μικτών δεδομένων 142

4.4.1 Αλγόριθμος εκπαίδευσης νευρωνικών δικτύων MRBF 145
4.4.2 Επιλογή μοντέλου MRBF..... 147
4.4.3 Εφαρμογή δικτύων MRBF σε μικτά σύνολα δεδομένων 149
4.4.4 Αποτελέσματα και συζήτηση 152
ΚΕΦΑΛΑΙΟ 5. ΕΦΑΡΜΟΓΕΣ ΣΤΗΝ ΕΞΕΛΙΞΗ ΤΟΥ ΠΑΓΚΟΣΜΙΟΥ ΙΣΤΟΥ .. 160
5.1 Εισαγωγή στον παγκόσμιο ιστό 161
5.1.1 Ιστορική αναδρομή..... 161
5.1.2 Το όραμα του σημασιολογικού ιστού..... 162
5.1.3 Τα επίπεδα του σημασιολογικού ιστού 163
5.2 Εισαγωγή στην εξέλιξη του παγκόσμιου ιστού..... 165
5.2.1 Βασικές έννοιες 166
5.2.2 Ταξινόμηση ιστοσελίδων, προεπεξεργασία και αναπαράσταση κειμένου..... 168
5.3 Ανάπτυξη συστήματος για την ταξινόμηση ιστοσελίδων 169
5.3.1 Εφαρμογή του προτεινόμενου συστήματος στην ταξινόμηση ιστοσελίδων 170
5.3.2 Αποτελέσματα και συζήτηση 172
ΣΥΜΠΕΡΑΣΜΑΤΑ..... 175
ΒΙΒΛΙΟΓΡΑΦΙΑ..... 178

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Σχήμα 1.1. Σχηματικό διάγραμμα νευρωνικού συστήματος. 16
Σχήμα 1.2. Σχηματικό διάγραμμα τυπικού βιολογικού νευρώνα. 18
Σχήμα 1.3. Σχηματικό διάγραμμα μοντέλου μη γραμμικού νευρώνα. 21
Σχήμα 1.4. Συνάρτηση κατωφλίου. 23
Σχήμα 1.5. Τμηματικά γραμμική συνάρτηση. 23
Σχήμα 1.6. Σιγμοειδής συνάρτηση (μπλε: α=0.5, κόκκινο: α=1, πράσινο: α=2). 23
Σχήμα 1.7. Σχηματικό διάγραμμα MLP με δύο κρυφές στοιβάδες. 26
Σχήμα 1.8. Απεικόνιση των κατευθύνσεων των δύο βασικών σημάτων σε MLP: Διάδοση των σημάτων συνάρτησης προς τα εμπρός και των σημάτων σφάλματος προς τα πίσω. 26
Σχήμα 1.9. Παλινδρομικό νευρωνικό δίκτυο χωρίς αυτό-ανατροφοδοτούμενους βρόχους και κρυφούς νευρώνες. 28
Σχήμα 1.10. Παλινδρομικό νευρωνικό δίκτυο με κρυφούς νευρώνες. 28
Σχήμα 1.11. Δισδιάστατο πλέγμα νευρώνων σε δίκτυο Kohonen. 29
Σχήμα 1.12. Διάγραμμα βαθμίδων νευρωνικού δικτύου που εστιάζει στον μοναδικό νευρώνα της στοιβάδας εξόδου. 32
Σχήμα 1.13. Πίνακας Σύγχυσης. 37
Σχήμα 2.1. Δομή δικτύου RBF. 51
Σχήμα 2.2. Τυπική δομή νευρώνα l κρυφής στοιβάδας δικτύου RBF. 52
Σχήμα 2.3. Συνάρτηση Gauss. 53
Σχήμα 2.4. Συνάρτηση Thin Plate Spline. 54
Σχήμα 2.5. Τροποποιημένη συνάρτηση Thin Plate Spline. 54
Σχήμα 2.6. Δισδιάστατο παράδειγμα: Συμμετρική ασαφής διαμέριση σε 5 ασαφή σύνολα και για τις δύο μεταβλητές εισόδου. 64
Σχήμα 2.7. Δισδιάστατο παράδειγμα: Μη συμμετρική ασαφής διαμέριση σε 7 ασαφή σύνολα για τη μεταβλητή εισόδου u1 και σε 5 ασαφή σύνολα για τη u2. 65
Σχήμα 2.8. Δισδιάστατο παράδειγμα: Ελλειψοειδής συνάρτηση συμμετοχής για μη συμμετρική ασαφή διαμέριση. 68
Σχήμα 3.1. Κωδικοποίηση ενός μη συμμετρικού 3-διάστατου χώρου εισόδου σε ένα σωματίδιο: Ο αριθμός των ασαφών συνόλων σε κάθε διάσταση κωδικοποιείται στο αντίστοιχο στοιχείο του σωματιδίου. 72

Σχήμα 3.2. Εξέλιξη του κριτηρίου RMSE ανά επανάληψη στο σύνολο δεδομένων αξιολόγησης για τις τρεις μεθοδολογίες. Για τον αλγόριθμο PSO-NSFM, απεικονίζεται η καλύτερη λύση που βρέθηκε σύμφωνα με τον δείκτη RMSE στο σύνολο αξιολόγησης.79

Σχήμα 3.3. Πίνακες σύγχυσης για τον αλγόριθμο PSO-NSFM στο σύνολο δεδομένων ελέγχου με τα 950 τραγούδια: (α) Το μοντέλο της εκτέλεσης που έδωσε το καλύτερο MCC στο σύνολο αξιολόγησης, (β) Μέσος όρος 20 εκτελέσεων..... 88

Σχήμα 3.4. Πίνακες σύγχυσης για τον αλγόριθμο PSO-NSFM στο σύνολο δεδομένων ελέγχου με τα 1000 τραγούδια: (α) Το μοντέλο της εκτέλεσης που έδωσε το καλύτερο MCC στο σύνολο αξιολόγησης, (β) Μέσος όρος 20 εκτελέσεων. 89

Σχήμα 3.5. Υπολογισμός ευρών για p=2 σε ένα μονοδιάστατο χώρο εισόδου που περιέχει (α) 3 κέντρα RBF και (b) 6 κέντρα RBF. 92

Σχήμα 3.6. Σχηματική αναπαράσταση των δύο συνεργατικών σμηνών, τα οποία συνεργάζονται για τη βελτιστοποίηση μοντέλων νευρωνικών δικτύων RBF. 99

Σχήμα 3.7. Η καλύτερη λύση που παράγει ο αλγόριθμος COOP-VW, στο σύνολο δεδομένων Samad: Αλλαγές του δείκτη RMSE ανά επανάληψη στο υποσύνολο αξιολόγησης. Η μεγεθυμένη περιοχή επικεντρώνεται στο σημείο όπου το σμήνος κάλυψης των πλησιέστερων κέντρων εντοπίζει το ολικό βέλτιστο, μετά από μια επιτυχημένη επανεκκίνηση 103

Σχήμα 3.8. Εξέλιξη του κριτηρίου RMSE ανά επανάληψη στο σύνολο αξιολόγησης για τις τέσσερις μεθοδολογίες, όπου απεικονίζεται η καλύτερη λύση που βρέθηκε σύμφωνα με τον δείκτη RMSE στο σύνολο αξιολόγησης. 105

Σχήμα 3.9. Εξέλιξη του κριτηρίου MCC ανά επανάληψη στο σύνολο αξιολόγησης για τις τέσσερις μεθοδολογίες, όπου απεικονίζεται η καλύτερη λύση που βρέθηκε σύμφωνα με τον δείκτη MCC στο σύνολο αξιολόγησης. 106

Σχήμα 3.10. Σχηματική αναπαράσταση του αλγορίθμου ESA-NSFM. 113

Σχήμα 3.11. Εξέλιξη του κριτηρίου ακρίβεια% ανά επανάληψη στο σύνολο αξιολόγησης για τις τρεις μεθοδολογίες. Για τον αλγόριθμο ESA-NSFM, απεικονίζεται η καλύτερη λύση που βρέθηκε σύμφωνα με τον δείκτη ακρίβεια% στο σύνολο αξιολόγησης. 125

Σχήμα 4.1. Δομή ενός δικτύου CRBF με κατηγορικές εισόδους. 131

Σχήμα 4.2. Δομή ενός δικτύου MRBF για την ταυτόχρονη διαχείριση αριθμητικών και κατηγορικών μεταβλητών. 143

Σχήμα 4.3. Κύριες συνιστώσες (PCs) και διανύσματα score και loading για έναν πίνακα **Z** διάστασης (10×2): 10 παραδειγμάτων και 2 διαστάσεων. 147

Σχήμα 5.1. Αρχιτεκτονική του σημασιολογικού ιστού σύμφωνα με τον Tim Berners-Lee [258]. 164

Σχήμα 5.2. Σχηματική αναπαράσταση του συστήματος MRBF-WPC. 171

Σχήμα 5.3. Πίνακες σύγχυσης για το μοντέλο που έδωσε το καλύτερο MCC στο σύνολο αξιολόγησης για το σύστημα MRBF-WPC στα σύνολα δεδομένων (α) αξιολόγησης και (β) ελέγχου. 174

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 3.1. Επισκόπηση συνόλων δεδομένων τύπου benchmark 76

Πίνακας 3.2. Λειτουργικές παράμετροι για τον αλγόριθμο PSO-NSFM 78

Πίνακας 3.3. Αποτελέσματα για όλα τα σύνολα δεδομένων – RMSE, ασαφής διαμέριση και αριθμός κόμβων 81

Πίνακας 3.4. Αποτελέσματα για όλα τα σύνολα δεδομένων – υπολογιστικοί χρόνοι και αριθμός επαναλήψεων 83

Πίνακας 3.5. Λειτουργικές παράμετροι για τον αλγόριθμο PSO-NSFM 86

Πίνακας 3.6. Αποτελέσματα για τα σύνολα δεδομένων GTZAN – MCC, ακρίβεια%, ασαφής διαμέριση και αριθμός κόμβων 87

Πίνακας 3.7. Επισκόπηση συνόλων δεδομένων τύπου benchmark 101

Πίνακας 3.8. Λειτουργικές παράμετροι για τον αλγόριθμο COOP-VW 102

Πίνακας 3.9. Αποτελέσματα για τα σύνολα δεδομένων προσέγγισης συνάρτησης – RMSE, ασαφής διαμέριση, ποσοστό κάλυψης πλησιέστερων κέντρων, αριθμός κόμβων και απαιτούμενες επαναλήψεις εύρεσης της καλύτερης λύσης 107

Πίνακας 3.10. Αποτελέσματα για τα σύνολα δεδομένων ταξινόμησης – MCC, ασαφής διαμέριση, ποσοστό κάλυψης πλησιέστερων κέντρων, αριθμός κόμβων και απαιτούμενες επαναλήψεις εύρεσης της καλύτερης λύσης 109

Πίνακας 3.11. Επισκόπηση συνόλων δεδομένων τύπου benchmark 117

Πίνακας 3.12. Λειτουργικές παράμετροι για τον αλγόριθμο ESA-NSFM 119

Πίνακας 3.13. Αποτελέσματα για όλα τα σύνολα δεδομένων – Ακρίβεια% και MCC 123

Πίνακας 3.14. Αποτελέσματα για όλα τα σύνολα δεδομένων – Ασαφής διαμέριση, βελτιστοποιημένες παράμετροι και υπολογιστικοί χρόνοι 124

Πίνακας 4.1. Επισκόπηση συνόλων δεδομένων τύπου benchmark 136

Πίνακας 4.2. Αποτελέσματα των δικτύων CRBF με τα έξι μέτρα απόστασης και των δύο μεθόδων επιλογής κέντρων (στο υποσύνολο ελέγχου) 140

Πίνακας 4.3. Σύγκριση με διάφορες τεχνικές ταξινόμησης (με βάση το υποσύνολο ελέγχου) …..... 141

Πίνακας 4.4. Επισκόπηση συνόλων δεδομένων τύπου benchmark 150

Πίνακας 4.5. Αποτελέσματα για τα σύνολα δεδομένων προσέγγισης συνάρτησης – RMSE και R2 154

Πίνακας 4.6. Αποτελέσματα για τα σύνολα δεδομένων ταξινόμησης – 157
Πίνακας 5.1. Αποτελέσματα για το σύνολο δεδομένων WebKB – MCC και ακρίβεια% ... 172
Πίνακας 5.2. Αποτελέσματα για το σύνολο δεδομένων WebKB – βελτιστοποιημένες παράμετροι και υπολογιστικοί χρόνοι 173

ΚΑΤΑΛΟΓΟΣ ΑΛΓΟΡΙΘΜΩΝ

Αλγόριθμος 1 – Αλγόριθμος FM 63
Αλγόριθμος 2 – Αλγόριθμος NSFM 67
Αλγόριθμος 3 – Αλγόριθμος PSO-NSFM 74
Αλγόριθμος 4 – Αλγόριθμος COOP-VW 97
Αλγόριθμος 5 – Αλγόριθμος ESA-NSFM 113
Αλγόριθμος 6 – Αλγόριθμος CCS 133
Αλγόριθμος 7 – Διαδικασία εκπαίδευσης CRBF 134
Αλγόριθμος 8 – Διαδικασία εκπαίδευσης MRBF 148

ΚΑΤΑΛΟΓΟΣ ΣΥΝΤΟΜΟΓΡΑΦΙΩΝ

ABC	:	Artificial Bee Colony
ACC%	:	percent accuracy
ACO	:	Ant Colony Optimization
AI	:	Artificial Intelligence
AIS	:	Artificial Immune System
ANN	:	Artificial Neural Network
CART	:	Classification and Regression Trees
CCS	:	Categorical Center Selection
CI	:	Computational Intelligence
COOP-VW	:	Cooperative Particle Swarm Optimization with Variable Width basis functions
CPSO	:	Cooperative Particle Swarm Optimization
CRBF	:	Categorical Radial Basis Function
EC	:	Evolutionary Computation
ELMs	:	Extreme Learning Machines
ERBF	:	Euclidean Radial Basis Function
ESA	:	Evolutionary Simulated Annealing
ESA-NSFM	:	Evolutionary Simulated Annealing - Non Symmetric Ffuzzy Means
FM	:	Fuzzy Means
FS	:	Fuzzy Systems
GCS	:	Greedy Center Selection
HTML	:	HyperText Markup Language
HTTP	:	HyperText Transfer Protocol
IOF	:	Inverse Occurrence Frequency
LMS	:	Least Mean Square
LR	:	Linear Regression
MCC	:	Matthews Correlation Coefficient
MDL	:	Minimum Description Length
MGC	:	Music Genre Classification
MLPs	:	Multi-Layer Perceptrons

MRBF	:	Mixed Radial Basis Function
MRBF-WPC	:	Mixed Radial Basis Function - Web Page Classification
NB	:	Naïve Bayes
NSFM	:	Non Symmetric Fuzzy Means
OF	:	Occurrence Frequency
PCA	:	Principal Component Analysis
PCs	:	Principal Components
PSO	:	Particle Swarm Optimization
PSO-NSFM	:	Particle Swarm Optimization-Non Symmetric Fuzzy Means
RBF	:	Radial Basis Function
RCS	:	Random Center Selection
RFT	:	Random Forest Trees
RMSE	:	Root Mean Square Error
RNN	:	Recurrent Neural Network
SA	:	Simulated Annealing
SFM	:	Symmetric Fuzzy Means
SI	:	Swarm Intelligence
STS	:	Simulation Time Step
SVMr	:	Support Vector Machines-Regression
SVMs	:	Support Vector Machines
TPS	:	Thin Plate Spline
URI	:	Uniform Resource Identifier
W3C	:	World Wide Web Consortium
WebKB	:	World Wide Knowledge Base
WWW	:	Word Wide Web
XML	:	eXtensible Markup Language

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα διδακτορική διατριβή, η οποία εκπονήθηκε κατά την περίοδο Ιούλιος 2013 - Ιούλιος 2017, είναι αποτέλεσμα προσωπικής προσπάθειας και συμβολής πολλών ανθρώπων τους οποίους νοιώθω την ανάγκη να ευχαριστήσω και να τους εκφράσω την ευγνωμοσύνη μου.

Αρχικά, θα ήθελα να εκφράσω την απεριόριστη εκτίμησή μου και ευγνωμοσύνη μου προς τον επιβλέποντα Καθηγητή κ. Χαράλαμπο Σαρίμβεη για την ευκαιρία που μου έδωσε να ασχοληθώ με την επιστήμη που αγαπώ στα πλαίσια ενός τόσο ενδιαφέροντος θέματος. Η υλοποίηση αυτής της διδακτορικής διατριβής δεν θα ήγαν εφικτή δίχως την εμπιστοσύνη που μου έδειξε, την ελευθερία που μου έδωσε, την κατανόηση που είχε απέναντί μου και την ουσιαστική καθοδήγησή του.

Καθοριστικό ρόλο στην επιτυχή εκπόνηση αυτής της διδακτορικής διατριβής είχε ο Αναπληρωτής Καθηγητής κ. Αλέξανδρος Αλεξανδρίδης, τον οποίο ευχαριστώ θερμά για την ανιδιοτελή βοήθεια και συμπαράσταση που μου παρείχε όλα αυτά τα χρόνια. Πρέπει να αναφέρω ότι δίπλα του έμαθα ότι για να μπορέσει κάποιος να φτάσει στο σημείο που επιθυμεί χρειάζονται τρία πράγματα: δουλειά, δουλειά και πάλι δουλειά.

Ιδιαίτερα, θα ήθελα να ευχαριστήσω τον Καθηγητή κ. Χρήστο Κυρανούδη και τον Αναπληρωτή Καθηγητή κ. Γεώργιο Μαυρωτά για τη συμμετοχή τους στην τριμελή συμβουλευτική επιτροπή και τις χρήσιμες συμβουλές και παρατηρήσεις τους, κατά την εκπόνηση αυτής της διατριβής.

Επίσης, θα ήθελα να ευχαριστήσω τα υπόλοιπα μέλη της επταμελούς επιτροπής για την τιμή που μου έκαναν να συμμετέχουν στην εξεταστική επιτροπή της παρούσας διατριβής.

Ακόμη, θα ήθελα να ευχαριστήσω τους Διδάκτορες κ. Φίλιππο Δογάνη, Θανάση Νικολακόπουλο, Γεώργιο Δρακάκη, τη Διδάκτορα κα. Γεωργία Τσιλίκη και τον Μεταπτυχιακό φοιτητή κ. Άγγελο Βαλσάμη για τη φιλική τους συνεργασία στα πλαίσια της κοινής παρουσίας μας στη μονάδα αυτόματης ρύθμισης και πληροφορικής.

Ένα μεγάλο ευχαριστώ από τα βάθη της καρδιάς μου θα ήθελα να εκφράσω στην Ηλεκτρονικό Μηχανικό κα. Κατερίνα Κουτσουδάκη για την αμέριστη συμπαράστασή της, και την ανιδιοτελή φιλία της.

Επίσης, νοιώθω την ανάγκη να ευχαριστήσω δύο υπέροχους και ξεχωριστούς ανθρώπους, τον κ. Κώστα Ζούμπερη και τη γυναίκα του κα. Μαρία.

Κλείνοντας, δεν υπάρχουν λόγια να εκφράσω την ατελείωτη ευγνωμοσύνη που νοιώθω για τη μητέρα μου Μαρία, την αδερφή μου Δέσποινα, τον φίλο και πατέρα μου Νίκο, τη γιαγιά μου Ευαγγελία και τον παππού μου Δημήτρη. Ιδιαίτερα, θα ήθελα να ευχαριστήσω τη μητέρα μου που η δίψα της για μάθηση ήταν οδηγός μου όλα αυτά τα χρόνια των σπουδών μου.

Ευαγγελία Χονδροδήμα

Ιούλιος 2017

ΠΕΡΙΛΗΨΗ

Η υπολογιστική νοημοσύνη (computational intelligence, CI) αποτελεί έναν κλάδο της τεχνητής νοημοσύνης που βασίζεται στη μίμηση στοιχείων ευφυΐας που παρατηρούνται στα έμβια όντα και προσφέρει τεχνικές ικανές να επιλύουν πολύπλοκα προβλήματα που δεν θα μπορούσαν να επιλυθούν με συμβατικές μεθοδολογίες. Το αντικείμενο της παρούσας διδακτορικής διατριβής είναι η ανάπτυξη νέων αλγορίθμων που εντάσσονται στο πεδίο της υπολογιστικής νοημοσύνης, και παρουσιάζουν τη δυνατότητα χειρισμού αριθμητικών, κατηγορικών και μικτών μεταβλητών, παρέχοντας βελτιωμένες ικανότητες πρόβλεψης σε σχέση με συμβατικές μεθοδολογίες. Οι τεχνικές που ανήκουν στον κλάδο της υπολογιστικής νοημοσύνης και αποτελούν βασικά εργαλεία της παρούσας διατριβής είναι τα νευρωνικά δίκτυα ακτινικής συνάρτησης βάσης (radial basis function, RBF), η τεχνική εξελικτικής προσομοιωμένης ανόπτησης (evolutionary simulated annealing, ESA) και ο αλγόριθμος βελτιστοποίησης σμήνους σωματιδίων (particle swarm optimization, PSO).

Σε ότι αφορά στον χειρισμό αριθμητικών εισόδων, στα πλαίσια της παρούσας διατριβής αναπτύχθηκαν 3 νέοι αλγόριθμοι εκπαίδευσης νευρωνικών δικτύων RBF, οι οποίοι επιτυγχάνουν την πλήρη βελτιστοποίηση των τελευταίων, με ταυτόχρονο προσδιορισμό όλων των παραμέτρων τους. Ο πρώτος αλγόριθμος που αναπτύχθηκε, ονομάστηκε PSO-NSFM και βασίζεται στον μη συμμετρικό αλγόριθμο των ασαφών μέσων (non symmetric fuzzy means, NSFM), ο οποίος ενισχύεται με έναν μηχανισμό βελτιστοποίησης PSO, προκειμένου να βελτιστοποιηθεί η ασαφής διαμέριση του χώρου εισόδου. Ο αλγόριθμος PSO-NSFM παράγει μοντέλα RBF που παρουσιάζουν απλούστερες δομές και ταυτόχρονα μεγαλύτερη ακρίβεια πρόβλεψης σε μικρότερους υπολογιστικούς χρόνους, έναντι καθιερωμένων τεχνικών μηχανικής μάθησης.

Ο δεύτερος αλγόριθμος που αναπτύχθηκε για χειρισμό αριθμητικών εισόδων ονομάστηκε COOP-VW και βελτιστοποιεί την κατασκευή δικτύων RBF που εκπαιδεύονται με τον αλγόριθμο NSFM, με βάση ένα συνεργατικό πλαίσιο PSO (cooperative PSO, CPSO). Ο αλγόριθμος COOP-VW επιτρέπει τη χρήση συναρτήσεων βάσης μεταβλητού εύρους, που αυξάνουν την ευελιξία των παραγόμενων μοντέλων, με αποτέλεσμα την περαιτέρω αύξηση της ακρίβειας πρόβλεψης, αλλά με σχετικά μικρή αύξηση του υπολογιστικού κόστους σε σχέση με την τεχνική PSO-NSFM.

Ο τρίτος αλγόριθμος που αναπτύχθηκε ονομάστηκε ESA-NSFM και προσαρμόζει την τεχνική ESA με στόχο τη βελτιστοποίηση των μοντέλων RBF που παράγονται από τον αλγόριθμο NSFM. Ο αλγόριθμος ESA-NSFM παράγει μοντέλα δικτύων RBF, τα οποία παρουσιάζουν αυξημένη ακρίβεια πρόβλεψης σε σχέση με τις καθιερωμένες τεχνικές μηχανικής μάθησης, και ταυτόχρονα εύκολη επιλογή λειτουργικών παραμέτρων.

Χρησιμοποιώντας αλγόριθμους όπως αυτοί που αναφέρθηκαν προηγουμένως, η βασική δομή των δικτύων RBF είναι σε θέση να χειριστεί με μεγάλη επιτυχία αριθμητικά δεδομένα εισόδου, ωστόσο εμφανίζει προβλήματα στον χειρισμό κατηγορικών ή μικτών δεδομένων. Προκειμένου να επιλυθεί αυτό το πρόβλημα, στα πλαίσια αυτής της διδακτορικής διατριβής αναπτύχθηκαν δύο νέες αρχιτεκτονικές νευρωνικών δικτύων. Η πρώτη από αυτές, ονομάστηκε CRBF και αποτελεί επέκταση των δικτυών RBF, δίνοντας στα τελευταία τη

δυνατότητα χειρισμού κατηγορικών δεδομένων. Αυτό επιτυγχάνεται χρησιμοποιώντας κατηγορικές πλειάδες ως κέντρα RBF, ενώ παράλληλα προτείνεται και ο αλγόριθμος categorical center selection (CCS) για την εκπαίδευση αυτών των δικτύων. Ο αλγόριθμος αυτός είναι ιδιαίτερα απλός και παρουσιάζει μεγάλη ταχύτητα εκτέλεσης, καθώς χρησιμοποιεί μόνο ένα πέρασμα των δεδομένων.

Η δεύτερη αρχιτεκτονική νευρωνικών δικτύων που αναπτύχθηκε ονομάστηκε MRBF και έχει την ικανότητα να χειρίζεται μικτά δεδομένα. Τα δίκτυα MRBF ουσιαστικά αποτελούν συνδυασμό των δικτύων RBF και CRBF, και εκπαιδεύονται με μία νέα μεθοδολογία η οποία βασίζεται στον αλγόριθμο των ασαφών μέσων και στην τεχνική CCS. Τα δίκτυα MRBF, λόγω της ικανότητας τους να λαμβάνουν υπόψη ξεχωριστά τη φύση των αριθμητικών και των κατηγορικών δεδομένων, μπορούν να δώσουν λύσεις με αυξημένη ακρίβεια πρόβλεψης σε σχέση με μεθοδολογίες που είναι εγγενώς φτιαγμένες για έναν μόνο τύπο δεδομένων, ενώ, λόγω του απλού και γρήγορου αλγορίθμου εκπαίδευσης που χρησιμοποιούν, ενδείκνυνται για εφαρμογή σε σύνολα δεδομένων μεγάλου όγκου (big data).

Όλες οι μέθοδοι που προαναφέρθηκαν, αξιολογήθηκαν σε ένα μεγάλο πλήθος συνόλων δεδομένων τύπου benchmark, που προέρχονται από ποικίλα επιστημονικά πεδία, και περιλαμβάνουν προβλήματα προσέγγισης συνάρτησης και ταξινόμησης. Τα αποτελέσματα και οι συγκρίσεις με ένα ευρύ φάσμα μεθόδων μηχανικής μάθησης, που περιλαμβάνουν διαφορετικές μεθόδους νευρωνικών δικτύων, μηχανές διανυσμάτων υποστήριξης (support vector machines, SVMs), ταξινομητές Bayes, δέντρα αποφάσεων, ταξινομητές τυχαίων δασών (random forest trees), κ.α., καταδεικνύουν την αποτελεσματικότητα των προτεινόμενων αλγορίθμων.

Επιπλέον, σε ότι αφορά στο επίπεδο της εφαρμογής, στα πλαίσια της παρούσας διατριβής, αναπτύχθηκε ένα νέο ευφυές σύστημα με την ικανότητα αυτόματης ταξινόμησης ιστοσελίδων με βάση το περιεχόμενό τους, το οποίο αποτελεί ένα εργαλείο γενικής χρήσης για την ταξινόμηση ιστοσελίδων με ικανοποιητική ακρίβεια και μπορεί για παράδειγμα να ταξινομήσει μία ιστοσελίδα ανάλογα με τη λειτουργία της, το είδος της, τη θεματική περιοχή της, κ.α. Το σύστημα αυτό ονομάστηκε MRBF-WPC και υιοθετεί τα νευρωνικά δίκτυα MRBF και στοιχεία από τον κλάδο της εξόρυξης γνώσης στον ιστό. Η εφαρμογή του συστήματος σε ένα ευρέως χρησιμοποιούμενο σύνολο ιστοσελίδων και η σύγκριση με άλλες μεθόδους, έδειξε ότι το σύστημα παρέχει αυξημένη ακρίβεια πρόβλεψης ως προς την ταξινόμηση των ιστοσελίδων. Έτσι, το σύστημα MRBF-WPC είναι ένα ολοκληρωμένο εργαλείο που μπορεί να χρησιμοποιηθεί ως βάση για τον εμπλουτισμό του περιεχομένου του ιστού με μεταδεδομένα, και άρα για την εξέλιξη του παγκόσμιου ιστού σε σημασιολογικό ιστό.

ABSTRACT

Computational intelligence (CI) is a branch of artificial intelligence, which is based on imitating elements of intelligence exhibited by living organisms and offers techniques that are capable of solving complex problems that could not be solved by using conventional methodologies. The objective of this doctoral dissertation is the development of novel algorithms belonging to the field of computational intelligence, which are capable of handling numerical, categorical and mixed attributes, while providing improved predictive abilities when compared to conventional methodologies. The computational intelligence techniques which constitute the basic tools used in this dissertation are the radial basis function (RBF) neural networks, the evolutionary simulated annealing (ESA) technique and the particle swarm optimization (PSO) algorithm.

As far as numerical attributes are concerned, 3 novel RBF network training algorithms where developed in this dissertation, with the ability to perform full network optimization by concurrently determining all the RBF parameters. The first algorithm, namely PSO-NSFM, is based on the non symmetric fuzzy means (NSFM) algorithm, which is enhanced by a PSO optimization engine, in order to optimize the fuzzy partition of the input space. The PSO-NSFM algorithm produces RBF models with simpler structures, while at the same time exhibiting higher prediction accuracy in shorter computational times, when compared to standard machine learning techniques.

The second algorithm which was developed for handling numerical inputs, namely COOP-VW, optimizes the construction of RBF networks trained with the NSFM algorithm, based on a cooperative PSO (CPSO) framework. The COOP-VW algorithm allows for using variable-width basis functions, which increase the flexibility of the produced models. The result is an additional improvement in prediction accuracy, albeit at the cost of a relatively small increase of the computational cost, when compared to the PSO-NSFM technique.

The third algorithm which was developed, namely ESA-NSFM, is based on a modification of the ESA technique, in order to optimize the RBF models produced by the NSFM algorithm. The ESA-NSFM algorithm produces RBF network models of increased prediction accuracy compared to standard machine learning techniques, and at the same time offers fast operational parameter customization.

Using algorithms similar to those mentioned above, the basic RBF network structure can handle numeric inputs with great success; however, it is unable to deal with categorical, or mixed data. In order to solve this problem, two novel neural network architectures were developed within this doctoral dissertation. The first one, namely CRBF networks, is an expansion of RBF networks, providing them with the ability of handling categorical data. This is accomplished by using categorical tuples as RBF centers, while the categorical center selection (CCS) algorithm is also proposed for training these networks.

The second neural network architecture that was developed, namely MRBF networks, is capable of handling mixed data. In fact, MRBF networks are essentially a synthesis of RBF and CRBF networks, and are trained with a novel methodology which combines the fuzzy means algorithm and the CCS technique. Due to their ability to handle in a different manner

the nature of numerical and categorical data, MRBF networks can provide solutions with increased prediction accuracy, in contrast to the methodologies that are inherently designed for a single type of data. Moreover, because of the simple and fast training algorithm that they use, MRBF networks are suitable to handle big data.

All the aforementioned methods were evaluated on a large number of benchmark datasets, derived from a variety of scientific fields, including function approximation and classification problems. Results and comparisons with a wide range of machine learning methods, including different neural network methods, support vector machines (SVMs), naïve Bayes classifiers, decision trees, random forest trees, etc., highlight the effectiveness of the proposed approaches.

In addition, at the level of application, a novel intelligent system was developed with the ability to automatically classify web pages based on their content. The latter is a generic tool for web page classification which offers satisfactory precision and is able to classify a webpage depending on functional, genre, subject classification, etc. This system was called MRBF-WPC and adopts the MRBF neural network architecture and web mining tools. Applying this system to a widely used dataset of webpages and comparing it with other methods, has shown that the system provides increased accuracy in classifying web pages. Thus, the MRBF-WPC system is an integrated tool that can be used as the basis for enrichment of the web content with metadata, and thus, for the World Wide Web growth into the Semantic Web.

ΕΙΣΑΓΩΓΗ

Η υπολογιστική νοημοσύνη (computational intelligence, CI) είναι ένας σχετικά νέος και αναπτυσσόμενος επιστημονικός κλάδος, ο οποίος ασχολείται με τη μοντελοποίηση της βιολογικής και φυσικής νοημοσύνης. Ο κλάδος της υπολογιστικής νοημοσύνης περιλαμβάνει τεχνικές που βασίζονται στη μίμηση των μηχανισμών που παρατηρούνται στη φύση και ειδικότερα στα έμβια όντα, όπως για παράδειγμα η διαδικασία μάθησης που χρησιμοποιεί ο ανθρώπινος εγκέφαλος, η διαδικασία εξέλιξης των ειδών, κτλ. Αυτές οι τεχνικές είναι ικανές να επιλύουν σύνθετα προβλήματα, τα οποία αποτελούσαν πάντα μια μεγάλη πρόκληση για την επιστήμη των υπολογιστών και της μηχανικής. Ορισμένες από τις βασικές τεχνικές που ανήκουν στον κλάδο της υπολογιστικής νοημοσύνης, είναι τα τεχνητά νευρωνικά δίκτυα (artificial neural networks, ANN), τα ασαφή συστήματα (fuzzy systems, FS), ο εξελικτικός υπολογισμός (evolutionary computation, EC), η νοημοσύνη σμήνους (swarm intelligence, SI), και τα τεχνητά ανοσοποιητικά συστήματα (artificial immune system, AIS) [1].

Τις τελευταίες δεκαετίες, έχει σημειωθεί μεγάλη πρόοδος σχετικά με την ανάπτυξη αλγορίθμων βασισμένων στον τρόπο μάθησης του ανθρώπινου εγκεφάλου, στην ευφυία σμήνους που παρουσιάζεται σε διάφορα είδη κοινωνικοποιημένων μελών του ζωικού βασιλείου, αλλά και στις φυσικές διαδικασίες εξέλιξης. Αυτή η διδακτορική διατριβή βασίστηκε στους τομείς της υπολογιστικής νοημοσύνης που περιλαμβάνουν τα τεχνητά νευρωνικά δίκτυα, τον εξελικτικό υπολογισμό και τη νοημοσύνη σμήνους. Πιο συγκεκριμένα, τα δίκτυα ακτινικής συνάρτησης βάσης (radial basis function, RBF) [2], η τεχνική εξελικτικής προσομοιωμένης ανόπτησης (evolutionary simulated annealing, ESA) [3] και ο αλγόριθμος βελτιστοποίησης σμήνους σωματιδίων (particle swarm optimization, PSO) [1] αποτελούν βασικά εργαλεία της παρούσας διδακτορικής διατριβής.

Τα τεχνητά νευρωνικά δίκτυα έχουν σημειώσει ραγδαία ανάπτυξη σε βαθμό που θεωρούνται το σημαντικότερο και πιο διαδεδομένο επιστημονικό πεδίο της υπολογιστικής νοημοσύνης. Τα τεχνητά νευρωνικά δίκτυα, αποτελούν μια βασική τεχνική μηχανικής μάθησης, δηλαδή εξόρυξης γνώσης μέσα από δεδομένα, τα οποία μιμούνται στοιχεία της ανθρώπινης συμπεριφοράς [2]. Ουσιαστικά, πρόκειται για ένα σύνολο από ισχυρά μαθηματικά εργαλεία που μπορούν να μοντελοποιήσουν άγνωστα μη γραμμικά συστήματα, χρησιμοποιώντας μόνο δεδομένα εισόδου-εξόδου από αυτά. Τα νευρωνικά δίκτυα έχουν σαν στόχο την προσομοίωση των βιολογικών νευρωνικών δικτύων του ανθρώπινου εγκεφάλου. Το νευρικό σύστημα των ζώντων οργανισμών αποτελείται από πολλά νευρωνικά δίκτυα τα οποία είναι εξειδικευμένα σε διεργασίες όπως η μάθηση, η μνήμη, η όραση, κτλ. Κάθε νευρωνικό δίκτυο αποτελείται από ένα μεγάλο αριθμό μονάδων, που λέγονται νευρώνες. Αντίστοιχα, τα τεχνητά νευρωνικά δίκτυα, μιμούμενα τα βιολογικά δίκτυα, απαρτίζονται από μονάδες επεξεργασίας που ονομάζονται τεχνητοί νευρώνες.

Η εκπαίδευση ενός νευρωνικού δικτύου αναφέρεται στη διαδικασία προσδιορισμού των παραμέτρων του δικτύου έτσι ώστε η συμπεριφορά του δικτύου να προσαρμόζεται σε ένα σύνολο πρωτοτύπων, δηλαδή ένα νευρωνικό δίκτυο μαθαίνει τις σχέσεις ανάμεσα στις μεταβλητές εισόδου και εξόδου. Οι πρώτες προσπάθειες αντιμετώπισης του προβλήματος εκπαίδευσης των νευρωνικών δικτύων βασίστηκαν σε συμβατικές τεχνικές βελτιστοποίησης

(conventional optimization techniques). Η τεχνική τάχιστης κατάβασης (gradient descent) αποτελεί τη βάση του πρωτοποριακού αλγορίθμου οπισθοδιάδοσης (back-propagation) και των πολλών παραλλαγών του [2]. Αργότερα, η τεχνική αυτή συνδυάστηκε με τη μέθοδο Newton, με αποτέλεσμα τη δημιουργία του αλγορίθμου Levenberg-Marquardt, ο οποίος είναι μια ευρέως γνωστή τεχνική που χρησιμοποιείται στην εκπαίδευση νευρωνικών δικτύων [4].

Ανάλογα με το είδος των συνδέσεων των νευρώνων που απαρτίζουν ένα νευρωνικό δίκτυο, διακρίνονται διαφορετικές αρχιτεκτονικές από τις οποίες οι πιο γνωστές είναι τα νευρωνικά δίκτυα πολλών-στοιβάδων (multi-layer perceptrons, MLPs), τα παλινδρομικά νευρωνικά δίκτυα (recurrent neural networks, RNNs), τα δίκτυα Kohonen και τα νευρωνικά δίκτυα ακτινικής συνάρτησης βάσης (RBF). Τα τελευταία αποτελούν μια ιδιαίτερη αρχιτεκτονική νευρωνικών δικτύων, λόγω των διαφορών της στη δομή, τη λειτουργία, αλλά και τον τρόπο εκπαίδευσής της.

Τα δίκτυα RBF παρουσιάστηκαν για πρώτη φορά στην επίλυση του προβλήματος παρεμβολής σε χώρους πολλών διαστάσεων. Αυτό το θέμα ερευνήθηκε σε βάθος από τον Powell [5, 6] και επί του παρόντος είναι ένας από τους κύριους τομείς έρευνας στην αριθμητική ανάλυση. Το μεγάλο ενδιαφέρον της ερευνητικής κοινότητας στην αρχιτεκτονική των δικτύων RBF οφείλεται στην εγγενή απλότητα της δομής της, σε αντίθεση με τα δίκτυα MLP, που έχουν μία πιο περίπλοκη δομή. Τα δίκτυα RBF περιλαμβάνουν μία μόνο κρυφή στοιβάδα, η οποία αρχικά συνδέεται γραμμικά με τη στοιβάδα εισόδου και έπειτα με τη στοιβάδα εξόδου. Τα παραπάνω χαρακτηριστικά απηχούν στους αλγορίθμους εκπαίδευσης που υιοθετούνται για τα δίκτυα RBF, οι οποίοι είναι συνήθως γρήγοροι και αποτελεσματικοί.

Ουσιαστικά, η διαδικασία εκπαίδευσης των δικτύων RBF, αντιστοιχεί με τον υπολογισμό της πολυδιάστατης επιφάνειας που προσεγγίζει καλύτερα ένα σύνολο παραδειγμάτων εκπαίδευσης. Αυτή η επιφάνεια είναι ένα άθροισμα πολλών απλούστερων επιφανειών, όπου οι βάσεις τους είναι ακτινικά συμμετρικές γύρω από τα κέντρα που βρίσκονται σε συγκεκριμένα σημεία του χώρου εισόδου. Ως εκ τούτου, η εκπαίδευση ενός δικτύου RBF ισοδυναμεί με την εύρεση βέλτιστων τιμών για τις παραμέτρους του δικτύου, που περιλαμβάνουν: (α) τον αριθμό κέντρων RBF, (β) τις συντεταγμένες των κέντρων RBF, (γ) τα εύρη των κέντρων RBF (ισχύει μόνο στην περίπτωση που χρησιμοποιείται συνάρτηση ενεργοποίησης με μεταβλητό εύρος, π.χ. συνάρτηση Gauss) και (δ) τα συναπτικά βάρη που ενώνουν τους νευρώνες της κρυφής στοιβάδας με τους νευρώνες της στοιβάδας εξόδου.

Στη βιβλιογραφία έχει προταθεί μια ποικιλία αλγορίθμων για τον καθορισμό των θέσεων των κέντρων RBF. Οι αρχικές προσπάθειες που έγιναν [7, 8] αφορούσαν σε μία τεχνική που επέλεγε τον αριθμό των κέντρων RBF να είναι ίσος με τον συνολικό αριθμό των δεδομένων εκπαίδευσης, τοποθετώντας έτσι ένα κέντρο RBF πάνω σε κάθε δεδομένο. Ωστόσο, η τεχνική αυτή καταλήγει να επιλέγει έναν υπερβολικά μεγάλο αριθμό κόμβων της κρυφής στοιβάδας, ειδικά στην περίπτωση των συνόλων δεδομένων που περιέχουν μεγάλο αριθμό παραδειγμάτων. Προκειμένου να αποφευχθεί το παραπάνω φαινόμενο, μπορεί να χρησιμοποιηθεί μια τεχνική συσταδοποίησης (clustering), με στόχο την κατάτμηση του χώρου εισόδου σε έναν αριθμό περιοχών, ο οποίος είναι μικρότερος σε σύγκριση με τον αριθμό των δεδομένων εκπαίδευσης. Στη συνέχεια, οι συναρτήσεις βάσης ακτινικής συμμετρίας, τοποθετούνται στο κέντρο κάθε συστάδας των δεδομένων. Η πιο δημοφιλής τεχνική, η οποία βασίζεται στην παραπάνω ιδέα, είναι ο αλγόριθμος συσταδοποίησης *k*-means, ο οποίος εφαρμόστηκε σε κάποιες πρώιμες ερευνητικές εργασίες σχετικές με τα δίκτυα RBF [9-11]. Ο αλγόριθμος *k*-means, όμως, παρουσιάζει τρία βασικά εγγενή μειονεκτήματα. Το πρώτο μειονέκτημα σχετίζεται με την επαναληπτική φύση της μεθόδου, η οποία αυξάνει τον υπολογιστικό χρόνο. Το δεύτερο μειονέκτημα αφορά στην αδυναμία της μεθόδου *k*-means να επιλέξει αυτόματα τον κατάλληλο αριθμό κέντρων RBF, με αποτέλεσμα

η κατάλληλη δομή του δικτύου να καθορίζεται μέσω της διαδικασίας δοκιμής – σφάλματος, η οποία επίσης αυξάνει τον υπολογιστικό χρόνο. Το τρίτο μειονέκτημα είναι ότι ο αλγόριθμος δεν έχει επαναληψιμότητα, καθώς εξαρτάται από μια τυχαία αρχική επιλογή της θέσης του κέντρου.

Ένα πλήθος εναλλακτικών τεχνικών έχουν προταθεί για την αντιμετώπιση αυτών των μειονεκτημάτων. Ένας σημαντικός αριθμός αυτών των μεθόδων χρησιμοποιεί μια προσθετική (constructive) τεχνική, η οποία χτίζει την κρυφή στοιβάδα σταδιακά μέχρι να ικανοποιηθεί ένα συγκεκριμένο κριτήριο. Στο πλαίσιο αυτό, έχει διερευνηθεί η εφαρμογή του αλγορίθμου ορθογωνίων ελαχίστων τετραγώνων (the orthogonal least squares) [12-16], αλλά επιπλέον έχουν προταθεί και άλλες μέθοδοι συμπεριλαμβανομένων της αφαιρετικής (constructive decay) τεχνικής [17], των δικτύων κατανομής πόρων (resource allocating networks) [18] και της αρχής της ελάχιστης περιγραφής (minimum description length, MDL) [19].

Μια διαφορετική μεθοδολογία σε σύγκριση με τις παραπάνω τεχνικές, αποτελεί ο αλγόριθμος των ασαφών μέσων (fuzzy means, FM) [20], ο οποίος βασίζεται στην ασαφή διαμέριση [21-23] του χώρου των μεταβλητών εισόδου, και επιλέγει τη δομή και τα κέντρα των νευρώνων κάνοντας μόνο ένα πέρασμα από τα δεδομένα εκπαίδευσης, ενώ ολοκληρώνει τη διαδικασία της εκπαίδευσης σε μικρούς υπολογιστικούς χρόνους. Η μέθοδος αυτή, σε αντίθεση με τον συνήθη αλγόριθμο εκπαίδευσης δικτύων RBF, επιλέγει αυτόματα τη δομή του δικτύου για μια συγκεκριμένη ασαφή διαμέριση του χώρου εισόδου, έχει επαναληψιμότητα, είναι πολύ γρήγορη και έχει αποδειχθεί ότι είναι πολύ αποδοτική στη μοντελοποίηση μη γραμμικών συστημάτων, αφού έχει χρησιμοποιηθεί με επιτυχία σε μια σειρά εφαρμογών που συμπεριλαμβάνουν την εκτίμηση της εμφάνισης σημαντικών σεισμικών γεγονότων [24], τον ευφυή έλεγχο [25, 26], τον αυτόματο έλεγχο βιομηχανικών διεργασιών [27, 28], την εκτίμηση κρίσιμων παραμέτρων σε υλικά [29-31], την πρόβλεψη σε ημερολόγια κίνησης [32, 33], την επίλυση προβλημάτων επιλογής μεταβλητών [28], κ.α. Έχει επίσης προταθεί μία online έκδοση του αλγορίθμου FM [34], η οποία επιτρέπει στη δομή του δικτύου RBF να προσαρμόζεται σε πραγματικό χρόνο. Μία παραλλαγή της μεθόδου των ασαφών μέσων είναι ο μη συμμετρικός αλγόριθμος των ασαφών μέσων (non symmetric fuzzy means, NSFM), ο οποίος εφαρμόζει μία μη συμμετρική διαμέριση του χώρου εισόδου. Προκαταρκτικά αποτελέσματα [35] έδειξαν ότι η μη συμμετρική διαμέριση του χώρου εισόδου μπορεί να οδηγήσει στην ανάπτυξη μοντέλων RBF που αποτελούνται από μικρότερο αριθμό κρυφών κόμβων και ταυτόχρονα παρουσιάζουν μεγαλύτερη ακρίβεια πρόβλεψης. Συμπερασματικά, ο αλγόριθμος NSFM έχει τα ίδια πλεονεκτήματα με τον συμμετρικό αλγόριθμο, αλλά υπερτερεί στο γεγονός ότι έχει μεγαλύτερη ευελιξία γεγονός που οδηγεί σε πιο αξιόπιστα μοντέλα RBF.

Όσον αφορά στον τρόπο με τον οποίο επιλέγονται τα εύρη των κόμβων, ο πρότυπος αλγόριθμος *k*-means χρησιμοποιεί μια ευρετική τεχνική γνωστή ως *p*-nearest neighbors [36], όπου τα πλάτη κάθε κόμβου RBF επιλέγονται έτσι ώστε να υπάρχει επικάλυψη των *p* πλησιέστερων κέντρων RBF. Βέβαια, προκειμένου να βελτιωθεί η ακρίβεια της μεθόδου, έχουν προταθεί και πιο περίπλοκες μέθοδοι [37-39], με σκοπό τη βελτιστοποίηση των ευρών των κόμβων RBF. Αξιοποιώντας τη γραμμική σχέση μεταξύ της κρυφής στοιβάδας και της στοιβάδας εξόδου, οι περισσότεροι αλγόριθμοι εκπαίδευσης υπολογίζουν τα συναπτικά βάρη των δικτύων RBF εφαρμόζοντας τη μέθοδο της γραμμικής παλινδρόμησης στις αποκρίσεις των κρυφών νευρώνων και στις επιθυμητές εξόδους του δικτύου [5, 36]. Εναλλακτικές μέθοδοι για τον υπολογισμό των βαρών περιλαμβάνουν τεχνικές τάχιστης κατάβασης [40], ασαφή λογική [41] και τον αλγόριθμο μεγιστοποίησης προσδοκίας (expectation-maximization) [42].

Στη βιβλιογραφία, επίσης, έχουν προταθεί αλγόριθμοι που προσδιορίζουν όλες τις παραμέτρους εκπαίδευσης του δικτύου RBF ταυτόχρονα σε ένα βήμα. Μερικοί από αυτούς αναφέρονται στη συνέχεια: Στην εργασία [43], με σκοπό την εκπαίδευση των δικτύων RBF, εισάγεται ένα ιεραρχικό μοντέλο Bayesian. Το μοντέλο αυτό αντιμετωπίζει όλες τις παραμέτρους εκπαίδευσης ως άγνωστες τυχαίες μεταβλητές και ο υπολογισμός Bayesian εκτελείται μέσω μίας μεθόδου αναδρομικής μαρκοβιανής αλυσίδας Monte Carlo (reversible-jump Markov chain Monte Carlo), ενώ τα δίκτυα βελτιστοποιούνται χρησιμοποιώντας τον αλγόριθμο προσομοιωμένης ανόπτησης (simulated annealing, SA). Στην εργασία [44], παρουσιάζεται ένας αλγόριθμος ο οποίος σε ένα βήμα προσδιορίζει τις παραμέτρους ενός δικτύου RBF, σε προβλήματα παρεμβολής, όπου οι κόμβοι έχουν ίσες αποστάσεις μεταξύ τους. Κατόπιν, αντικαθίσταται η ευκλείδεια απόσταση της γκαουσιανής συνάρτησης RBF με μία απόσταση Mahalanobis. Επίσης, στην εργασία [45], όλες οι παράμετροι του δικτύου RBF, συμπεριλαμβανομένων των βαρών εισόδου που αντιστοιχούν στις συνδέσεις μεταξύ της στοιβάδας εισόδου και της κρυφής στοιβάδας, ρυθμίζονται από έναν κανόνα ανανέωσης δεύτερης τάξης (second order update rule).

Με βάση τα παραπάνω, είναι φανερό ότι η διαδικασία εκπαίδευσης των νευρωνικών δικτύων RBF διαμορφώνεται φυσικά ως πρόβλημα βελτιστοποίησης, όπου οι μεταβλητές σχεδιασμού είναι οι παράμετροι του δικτύου και η αντικειμενική συνάρτηση είναι μια συνάρτηση κόστους που εκφράζει τη διαφορά μεταξύ των προβλέψεων που παράγει το δίκτυο και των πραγματικών ή επιθυμητών προβλέψεων-στόχων. Θα πρέπει, ωστόσο, να σημειωθεί ότι ο υπολογισμός των βέλτιστων τιμών για όλες τις παραμέτρους ενός δικτύου RBF είναι ένα δύσκολο έργο. Όταν η διαδικασία εκπαίδευσης των δικτύων RBF, αλλά και γενικότερα των νευρωνικών δικτύων, αντιμετωπίζεται ως πρόβλημα βελτιστοποίησης, τότε η αντικειμενική συνάρτηση, συνήθως, παρουσιάζει κάποια ανεπιθύμητα χαρακτηριστικά, που περιλαμβάνουν: (α) πολυτροπικότητα (multimodality), γεγονός που μπορεί να προκαλέσει ενδεχομένη παγίδευση του αλγορίθμου σε ένα μη-ικανοποιητικό τοπικό βέλτιστο, (β) μη-διαφορισιμότητα, η οποία εμποδίζει τη χρήση μεθόδων βασισμένων σε παραγώγους και (γ) υψηλά επίπεδα θορύβου, που συνήθως παρουσιάζονται στα δεδομένα εκπαίδευσης [2]. Αυτοί οι παράγοντες καθιστούν την εύρεση της βέλτιστης λύσης, σχεδόν απίθανη και μάλιστα εμποδίζουν την εύρεση ακόμη και ενός ικανοποιητικού υπο-βέλτιστου συνόλου παραμέτρων των νευρωνικών δικτύων. Επιπλέον, η μη-διαφορισιμότητα σε σχέση με ορισμένες από τις παραμέτρους των νευρωνικών δικτύων, για παράδειγμα, το μέγεθος της κρυφής στοιβάδας, απαγορεύει τη χρήση μεθόδων που βασίζονται σε παραγώγους και συχνά οδηγεί σε μία χρονοβόρα και "κουραστική" διαδικασία δοκιμής-σφάλματος για την επίλυση του προβλήματος εκπαίδευσης. Δεδομένου ότι αυτά τα χαρακτηριστικά καθιστούν τη χρήση των κλασικών μεθόδων βελτιστοποίησης αναποτελεσματική, δεν αποτελεί έκπληξη το γεγονός ότι ένας σημαντικός αριθμός επιστημονικών εργασιών επικεντρώνονται στη βελτιστοποίηση της διαδικασίας εκπαίδευσης των νευρωνικών δικτύων μέσω της χρήσης εναλλακτικών μεθόδων, όπως οι τεχνικές που βασίζονται στον εξελικτικό υπολογισμό ή στη νοημοσύνη σμήνους.

Ο εξελικτικός υπολογισμός [1] ανήκει σε μία οικογένεια πιθανοτικών τεχνικών βελτιστοποίησης (probabilistic optimization), οι οποίες μιμούνται τις διαδικασίες της φυσικής εξέλιξης, με κύριο στόχο την ιδέα της επιβίωσης του ισχυρότερου. Αυτές οι τεχνικές παρουσιάζουν αρκετές διαφορές σε σχέση με τις συμβατικές μεθόδους βελτιστοποίησης και αποτελούν ισχυρές μεθόδους με την ικανότητα να ξεφεύγουν από τα τοπικά βέλτιστα. Οι τεχνικές εξελικτικού υπολογισμού έχουν χρησιμοποιηθεί σε διάφορους και ποικίλους επιστημονικούς τομείς, όπως για παράδειγμα στη βελτιστοποίηση των παραμέτρων των νευρωνικών δικτύων. Οι πρώτες εργασίες, με βάση αυτό το θέμα, είχαν επικεντρωθεί στα

δίκτυα MLP [46]. Στο πρώτο δείγμα αυτών των εργασιών, οι τεχνικές εξελικτικού υπολογισμού αντικαθιστούσαν τις μεθόδους τάχιστης κατάβασης χρησιμοποιώντας πληθυσμούς ατόμων που κωδικοποιούσαν τα βάρη και τις πολώσεις (biases) του δικτύου. Επίσης, με βάση τις μεθοδολογίες εξελικτικού υπολογισμού, προτάθηκαν και πιο περίπλοκες τεχνικές για την προσαρμογή των βαρών, συμπεριλαμβάνοντας τη μέθοδο της διαφορικής εξέλιξης (differential evolution), και μια υβριδική τεχνική που συνδυάζει τη μέθοδο της διαφορικής εξέλιξης και το φίλτρο Kalman [47].

Με βάση τα αποτελέσματα των παραπάνω τεχνικών, διαπιστώθηκε ότι αυτές οι μέθοδοι αντιμετώπιζαν πιο αποδοτικά την πολυτροπική φύση της αντικειμενικής συνάρτησης, αλλά το πρόβλημα βελτιστοποίησης της δομής του δικτύου, δηλαδή ο προσδιορισμός του πλήθους των κρυφών στοιβάδων και ο αριθμός των νευρώνων ανά κρυφή στοιβάδα, εξακολουθούσε να παραμένει άλυτο. Φυσικά, το επόμενο βήμα, στα πλαίσια της βελτιστοποίησης νευρωνικών δικτύων, ήταν η ενσωμάτωση μηχανισμών ικανών να επιλέγουν τη δομή των δικτύων, προσαρμόζοντας κατάλληλα τεχνικές βελτιστοποίησης. Νέες μέθοδοι που προσδίδουν στις αρχιτεκτονικές νευρωνικών δικτύων τη δυνατότητα εξέλιξης, με εφαρμογές σε προβλέψεις χρονοσειρών, έχουν προταθεί στις εργασίες [48, 49]. Επίσης, στην εργασία [50], παρουσιάζεται η μέθοδος ανάπτυξης εξελικτικών ασαφών νευρωνικών δικτύων ARTMAP για την αντιμετώπιση προβλημάτων ταξινόμησης συνόλων δεδομένων με μη ομοιόμορφη κατανομή κλάσεων, με εφαρμογή στην κατασκευή ημιαγωγών.

Στη βιβλιογραφία συναντάμε μεθοδολογίες εξελικτικού υπολογισμού που έχουν επίσης χρησιμοποιηθεί στη βελτιστοποίηση νευρωνικών δικτύων RBF. Στην εργασία [51], προτείνεται ένας γενετικός αλγόριθμος για την ανάπτυξη δυναμικών μοντέλων RBF, όπου βελτιστοποιούνται ο αριθμός και οι θέσεις των κέντρων RBF. Μια υβριδική τεχνική βασισμένη σε μεθόδους multilogistic, η οποία εφαρμόζει εξελικτικό προγραμματισμό με σκοπό την ανάπτυξη δικτύων RBF με απλούστερες δομές, παρουσιάζεται στην επιστημονική δημοσίευση [52]. Στην εργασία [53], περιγράφεται ένας πολυκριτηριακός εξελικτικός αλγόριθμος για τη βελτιστοποίηση των δικτύων RBF, συμπεριλαμβανομένων ορισμένων νέων τελεστών που εφαρμόζονται στην εξελικτική διαδικασία. Ένας εξελικτικός αλγόριθμος, ο οποίος επιλέγει, ταυτόχρονα, τα χαρακτηριστικά και τα μοντέλα, με σκοπό τη δημιουργία ταξινομητών RBF, σε μικρούς υπολογιστικούς χρόνους παρουσιάζεται στην εργασία [54]. Επίσης, μια μέθοδος που συνδυάζει τις τεχνικές του εξελικτικού υπολογισμού και της τάχιστης κατάβασης για την εκπαίδευση ταξινομητών RBF, προσδιορίζοντας την αρχιτεκτονική, τα βάρη και την τοπολογία των κόμβων παρουσιάζεται στην εργασία [55]. Ακόμη, στην εργασία [56], προτείνεται μια μεθοδολογία για την κατασκευή υβριδικών δικτύων RBF, όπου χρησιμοποιείται ένας γενετικός αλγόριθμος για τη βελτιστοποίηση των βασικών παραμέτρων σχεδιασμού του μοντέλου.

Μια άλλη κατηγορία τεχνικών που αντλείται από το οπλοστάσιο της επιστήμης της υπολογιστικής νοημοσύνης και παρουσιάζει ομοιότητες με τις μεθόδους εξελικτικού υπολογισμού, και ταυτόχρονα έχει αξιοποιηθεί στην εκπαίδευση νευρωνικών δικτύων είναι οι αλγόριθμοι που ανήκουν στην κατηγορία της νοημοσύνης σμήνους. Η νοημοσύνη σμήνους [1] αναφέρεται στην ιδιότητα φυσικών ή τεχνητών συστημάτων που επιδεικνύουν συλλογικά ευφυή συμπεριφορά. Η νοημοσύνη σμήνους έχει τη βάση της στις μελέτες των αποικιών, των κοπαδιών, ή των σμηνών, που αποτελούνται από κοινωνικούς οργανισμούς, οι οποίες ώθησαν στον σχεδιασμό ιδιαίτερα αποτελεσματικών μεθόδων βελτιστοποίησης. Από την απαρχή του χρόνου, οι βιολογικοί οργανισμοί ανακάλυψαν ότι, αν οργανωθούν σε ομάδες με κοινούς στόχους, έχουν πολύ περισσότερες πιθανότητες να επιβιώσουν και να εξελιχθούν. Η επιστημονική παρατήρηση του τρόπου δράσης των κοινωνικοποιημένων βιολογικών οργανισμών, αποκάλυψε τον τρόπο με τον οποίο η διασπορά πληροφορίας μέσα σε μια

ομάδα, καθώς και η εκμετάλλευση της συλλογικής εμπειρίας, οδηγούν στην επίλυση προβλημάτων με τον βέλτιστο τρόπο. Μεταξύ των πιο κοινών παραδειγμάτων του κλάδου της νοημοσύνης σμήνους είναι, ο τρόπος πετάγματος των πουλιών σε σχηματισμό σμήνους που αποτέλεσε την αρχική ιδέα για τον αλγόριθμο βελτιστοποίησης σμήνους σωματιδίων (PSO) [57]. Ο τελευταίος αποτελεί μια απλή, αλλά ιδιαίτερα ισχυρή, στοχαστική μέθοδο βελτιστοποίησης με βάση τον πληθυσμό, όπου τα άτομα, που αναφέρονται ως σωματίδια, ομαδοποιούνται σε ένα σμήνος. Πιο συγκεκριμένα, ο αλγόριθμος PSO βασίζεται στην προσομοίωση της κοινωνικής συμπεριφοράς των πουλιών και υιοθετεί μία επαναληπτική διαδικασία εφαρμόζοντας ένα πληθυσμό πιθανών λύσεων, που κωδικοποιούνται ως σωματίδια που «πετούν» στον πολυδιάστατο χώρο αναζήτησης, ψάχνοντας για τη βέλτιστη λύση. Ο αλγόριθμος PSO έχει χρησιμοποιηθεί με επιτυχία σε συνδυασμό με άλλα εργαλεία υπολογιστικής νοημοσύνης [58, 59]. Επίσης, μια πρωτότυπη τεχνική που βασίζεται στη βελτιστοποίηση PSO για αυτόματη σχεδίαση δικτύων MLP παρουσιάζεται στην εργασία [60]. Ακόμη, ένας πολυκριτηριακός αλγόριθμος PSO με την ικανότητα προσαρμογής της τοπολογίας του νευρωνικού δικτύου εφαρμόστηκε με επιτυχία στην ταξινόμηση δορυφορικών εικόνων [61].

Ιδιαίτερη προσοχή δίνεται στην εφαρμογή του αλγορίθμου PSO για την επίλυση του προβλήματος εκπαίδευσης των δικτύων RBF. Ένας υποβοηθούμενος αλγόριθμος PSO προς τα εμπρός παλινδρόμησης (PSO-aided orthogonal forward regression algorithm), ο οποίος βασίζεται στην τεχνική leave-one-out, αναπτύσσεται στην εργασία [62], προκειμένου να κατασκευαστούν μοντέλα RBF μειωμένης πολυπλοκότητας με κόμβους που έχουν την ικανότητα προσαρμογής. Επίσης, στην εργασία [63], ένας αναδρομικός αλγόριθμος ορθογωνίων ελαχίστων τετραγώνων (orthogonal least squares) συνδυάστηκε με τη μέθοδο PSO, με αποτέλεσμα μια νέα ευρετική μέθοδο βελτιστοποίησης της δομής πιθανοκρατικών δικτύων RBF (RBF probabilistic networks). Οι συγγραφείς στην εργασία [64] παρουσιάζουν μια προηγμένη αρχιτεκτονική πολυωνυμικών δικτύων RBF, η οποία βασίζεται στην τεχνική *k*-means, και για τον σχεδιασμό της χρησιμοποιούνται ο αλγόριθμος PSO και η μέθοδος διαφορικής εξέλιξης. Άλλη μία μελέτη που αφορά σε πολυωνυμικά δίκτυα RBF, τα οποία εφαρμόζουν τον αλγόριθμο PSO, παρουσιάζεται στην εργασία [65]. Επιπλέον, μια βελτιωμένη τεχνική εκπαίδευσης δικτύων RBF, βασισμένη στον αλγόριθμο PSO, για εφαρμογές σε πραγματικό χρόνο, προτείνεται στην εργασία [66].

Ο **πρώτος στόχος** αυτής της διδακτορικής διατριβής αφορά στην ανάπτυξη ενός νέου ολοκληρωμένου πλαισίου για την πλήρη εκπαίδευση νευρωνικών δικτύων RBF, συνδυάζοντας τη μη συμμετρική έκδοση του αλγορίθμου των ασαφών μέσων με ένα ειδικά σχεδιασμένο μηχανισμό βελτιστοποίησης βασισμένο σε βελτιστοποίηση σμήνους PSO. Ο αλγόριθμος NSFM εφαρμόζεται στη διαδικασία καθορισμού του αριθμού και των συντεταγμένων των κέντρων των νευρώνων της κρυφής στοιβάδας, ενώ τα συναπτικά βάρη υπολογίζονται με γραμμική παλινδρόμηση. Αξιοποιώντας τους μικρούς υπολογιστικούς χρόνους που απαιτούνται από τον αλγόριθμο NSFM, η μεθοδολογία «τυλίγει» γύρω του ένα μηχανισμό βελτιστοποίησης PSO, προκειμένου να βελτιστοποιηθεί η ασαφής διαμέριση του χώρου εισόδου. Το προτεινόμενο σχήμα [67] ονομάστηκε PSO-NSFM και έχει ως αποτέλεσμα την κατασκευή νευρωνικών δικτύων RBF με μεγαλύτερη ακρίβεια και μικρότερη πολυπλοκότητα, ενώ ένα ικανοποιητικό μοντέλο RBF μπορεί να επιτευχθεί σε σχετικά μικρούς υπολογιστικούς χρόνους. Ο αλγόριθμος PSO-NSFM αξιολογείται μέσα από την εφαρμογή του σε 12 πραγματικά και συνθετικά σύνολα δεδομένων benchmark, και συγκρίνεται με άλλες μεθοδολογίες εκπαίδευσης νευρωνικών δικτύων. Ο αλγόριθμος PSO-NSFM, επίσης, χρησιμοποιήθηκε στην αναγνώριση μουσικού είδους (music genre classification, MGC) και προέκυψε ένα ευφυές σύστημα, το οποίο μπορεί να κατατάσσει

αυτόματα ένα τραγούδι μουσικής στο είδος που ανήκει. Το σύστημα αυτό εφαρμόστηκε στο σύνολο μουσικών δεδομένων τύπου benchmark γνωστό ως GTZAN [68], το οποίο είναι ένα από τα πρώτα σύνολα δεδομένων που έχουν καταστεί διαθέσιμα στο κοινό, και θεωρείται ότι είναι το πλέον χρησιμοποιούμενο στην αναγνώριση μουσικού είδους [69].

Ο αλγόριθμος PSO-NSFM προσφέρει σημαντικά πλεονεκτήματα, όμως παράγει δίκτυα RBF που οι συναρτήσεις ενεργοποίησής τους έχουν σταθερό εύρος. Το γεγονός αυτό αποτελεί έναν περιορισμό, διότι η χρήση συναρτήσεων βάσης σταθερού εύρους περιορίζει την ευελιξία της πολυδιάστατης επιφάνειας που δημιουργεί ένα δίκτυο RBF, ενδεχομένως μειώνοντας την ακρίβεια του παραγόμενου μοντέλου ή/και απαιτώντας μεγαλύτερες δομές δικτύων για να επιτευχθούν ικανοποιητικά αποτελέσματα. Για την εξάλειψη του περιορισμού που προκαλεί η χρήση συναρτήσεων ενεργοποίησης σταθερού εύρους, μπορεί να χρησιμοποιηθεί ένα συνεργατικό πλαίσιο νοημοσύνης σμήνους. Οι συνεργατικές μέθοδοι που ανήκουν στις τεχνικές εξελικτικού υπολογισμού ή νοημοσύνης σμήνους έχουν χρησιμοποιηθεί πρόσφατα με μεγάλη επιτυχία στην επίλυση σύνθετων προβλημάτων βελτιστοποίησης [70-72]. Επιπλέον, αυτές οι μέθοδοι προσαρμόζονται ιδιαίτερα καλά στη βελτιστοποίηση νευρωνικών δικτύων [73-76], καθώς προσεγγίζουν εγγενώς την τμηματική φύση του προβλήματος εκπαίδευσης. Παρόλα αυτά, υπάρχουν εξαιρετικά λίγες επιστημονικές δημοσιεύσεις όσον αφορά στο θέμα της εκπαίδευσης των δικτύων RBF με συνεργατικές μεθόδους [77, 78].

Ο **δεύτερος στόχος** αυτής της διδακτορικής διατριβής περιλαμβάνει την ανάπτυξη ενός νέου σχήματος συνεργατικής μάθησης με την ικανότητα να επιλύει προβλήματα προσέγγισης συνάρτησης και κατηγοριοποίησης, συνδυάζοντας βελτιωμένη ακρίβεια και ικανότητες γενίκευσης. Συγκεκριμένα, η προτεινόμενη μέθοδος [79], υιοθετεί ένα συνεργατικό αλγόριθμο PSO με συναρτήσεις βάσης μεταβλητού εύρους (cooperative PSO with variable width basis functions, COOP-VW), ο οποίος χρησιμοποιεί πολλαπλά σμήνη ώστε να βελτιστοποιεί διαφορετικές συνιστώσες του διανύσματος λύσης [80]. Η μέθοδος COOP-VW βελτιστοποιεί πλήρως την κατασκευή δικτύων RBF καθορίζοντας τις θέσεις των κέντρων, τα συναπτικά βάρη και το μέγεθος του δικτύου. Παράλληλα, η μέθοδος COOP-VW επιτρέπει τη χρήση οποιασδήποτε συνάρτησης ενεργοποίησης που έχει ακτινική συμμετρία, γεγονός που οδηγεί σε αυξημένη ευελιξία των παραγόμενων μοντέλων. Επίσης, προκειμένου να αποφευχθεί η χρήση υπερβολικού αριθμού μεταβλητών σχεδιασμού, που δυσκολεύει το έργο της βελτιστοποίησης, εισάγεται ένα συμπαγές σχήμα αντιπροσώπευσης, με χρήση δύο διακριτών σμηνών, τα οποία ρυθμίζουν διαφορετικά είδη παραμέτρων του δικτύου RBF. Το πρώτο σμήνος εφαρμόζει τον μη συμμετρικό αλγόριθμο των ασαφών μέσων για τον υπολογισμό της δομής του νευρωνικού δικτύου και των συντεταγμένων των κέντρων RBF, ενώ το δεύτερο κωδικοποιεί τα εύρη των συναρτήσεων βάσεως, εισάγοντας μια τροποποιημένη ευρετική μέθοδο *p*-nearest neighbor [11]. Σε αντίθεση με τις περισσότερες από τις υπάρχουσες μεθόδους που βελτιστοποιούν πλήρως το νευρωνικό δίκτυο, ο αλγόριθμος COOP-VW εισάγει αυτή τη νέα τροποποιημένη μέθοδο *p*-nearest neighbor, με σκοπό μια πιο συνεπτυγμένη κωδικοποίηση του προβλήματος βελτιστοποίησης που οδηγεί σε λιγότερες μεταβλητές σχεδιασμού. Τελικά, τα δύο σμήνη συνεργάζονται ανταλλάσσοντας πληροφορίες με στόχο την ανακάλυψη βελτιωμένων δικτύων RBF, ενώ εισάγεται και μια λειτουργία επαναφοράς, προκειμένου να αποφευχθεί το πρόβλημα της τελμάτωσης (stagnation). Το προτεινόμενο σύστημα συγκρίνεται με εναλλακτικές μεθόδους και ελέγχεται η απόδοσή του μέσω της εφαρμογής του σε μια ευρεία γκάμα 9 προβλημάτων προσέγγισης συνάρτησης και ταξινόμησης τύπου benchmark, τα οποία προέρχονται από τον πραγματικό κόσμο και από προσομοιώσεις.

Μία άλλου είδους μέθοδος που προέρχεται από το οπλοστάσιο της υπολογιστικής νοημοσύνης και αποτελεί έναν από τους περισσότερο μελετημένους στοχαστικούς αλγόριθμους αναζήτησης είναι ο αλγόριθμος προσομοιωμένης ανόπτησης (simulated annealing, SA) [81, 82]. Η βασική ιδέα του αλγορίθμου προέρχεται από τη διαδικασία φυσικής ανόπτησης, της θέρμανσης και της ελεγχόμενης ψύξης ενός στερεού υλικού μέχρι την κρυσταλλοποίησή του, με στόχο τη βελτίωση των μηχανικών ιδιοτήτων του. Η μέθοδος SA παρουσιάζει δύο βασικές διαφορές σε σχέση με τους εξελικτικούς αλγορίθμους: α) χρησιμοποιεί μία μόνο λύση την οποία εξελίσσει, με αποτέλεσμα τη μείωση του υπολογιστικού φορτίου, και β) δίνει τη δυνατότητα αποδοχής υποδεέστερων λύσεων, στις πρώτες επαναλήψεις, με βάση μια πιθανότητα, για την αποφυγή του εγκλεισμού του αλγορίθμου σε τοπικά ελάχιστα. Βέβαια, η χρήση μίας μόνο λύσης, παρουσιάζει και μειονεκτήματα, όπως είναι οι μειωμένες πιθανότητες, εύρεσης του ολικού βέλτιστου κατά τα τελευταία στάδια του αλγορίθμου. Για την αποφυγή αυτών των μειονεκτημάτων αναπτύχθηκε ο αλγόριθμος εξελικτικής προσομοιωμένης ανόπτησης (evolutionary simulated annealing, ESA). Η μέθοδος ESA αποτελεί μια υβριδική τεχνική που εμπλουτίζει τον παραδοσιακό αλγόριθμο SA με εξελικτικές δυνατότητες, δηλαδή συνδυάζει τις βασικές αρχές της τεχνικής SA και των εξελικτικών αλγορίθμων, βοηθώντας έτσι τη διαδικασία βελτιστοποίησης να ξεφύγει από τοπικά ελάχιστα που θα προέκυπταν με τον απλό αλγόριθμο SA και να ανακαλύπτει βελτιωμένες λύσεις. Ως εκ τούτου, ο αλγόριθμος ESA χρησιμοποιεί έναν πληθυσμό από λύσεις σε κάθε επανάληψη, ενώ επιτρέπει τις επιζήμιες κινήσεις κατά τη διάρκεια των πρώτων επαναλήψεων.

Ο **τρίτος στόχος** αυτής της διδακτορικής διατριβής είναι η ανάπτυξη ενός ολοκληρωμένου πλαισίου για την παραγωγή κατηγοριοποιητών RBF, με κατάλληλη προσαρμογή της τεχνικής εξελικτικής προσομοιωμένης ανόπτησης (ESA), με στόχο τη βελτιστοποίηση των μοντέλων RBF που παράγονται από τον αλγόριθμο NSFM. Το σχήμα [83] που αναπτύχθηκε ονομάστηκε ESA-NSFM και παρουσιάζει εύκολη επιλογή των απαιτούμενων λειτουργικών παραμέτρων, δηλαδή δεν χρειάζεται ιδιαίτερη εργασία για την κατάλληλη προσαρμογή τους, καθώς ο αλγόριθμος ESA χρησιμοποιεί λίγες λειτουργικές παραμέτρους, στις οποίες δεν παρουσιάζει ιδιαίτερη ευαισθησία. Έτσι, οι παραγόμενοι κατηγοριοποιητές, μπορούν να χρησιμοποιηθούν ως συστήματα υποστήριξης αποφάσεων σε ιατρική διάγνωση, αφού τα σύνολα δεδομένων ιατρικού ενδιαφέροντος παρουσιάζουν μεγάλη ποικιλία, ειδικά ως προς τις μεταβλητές εισόδου και εξόδου. Η προτεινόμενη μέθοδος συγκρίνεται με άλλες τεχνικές μηχανικής μάθησης και αξιολογείται σε 9 διαφορετικά σύνολα δεδομένων τύπου benchmark, όπου ο κοινός στόχος είναι η εκπαίδευση ενός κατηγοριοποιητή ικανού να παρέχει ιατρική διάγνωση.

Οι τρεις μέθοδοι εκπαίδευσης νευρωνικών δικτύων RBF που αναφέρθηκαν παραπάνω, χειρίζονται αριθμητικά δεδομένα, ωστόσο, πολύ συχνά, στα σύνολα δεδομένων συναντώνται κατηγορικά χαρακτηριστικά. Αντιπροσωπευτικά παραδείγματα κατηγορικών μεταβλητών αποτελούν το φύλο, με τιμές αρσενικό και θηλυκό, και η ομάδα αίματος με τιμές A, B, AB και O. Τα κατηγορικά δεδομένα έχουν απασχολήσει αρκετό καιρό την ερευνητική κοινότητα. Συγκεκριμένα, στα τέλη του 18^{ου} αιώνα, ο Pearson πρότεινε τον στατιστικό κανόνα *x*-τετράγωνο (chi-squared) [84], ο οποίος χρησιμοποιείται συχνά για να μετρηθεί ο βαθμός ανεξαρτησίας μεταξύ των κατηγορικών μεταβλητών. Τα κατηγορικά δεδομένα, από τη φύση τους, δεν έχουν εγγενώς μαθηματική σημασία, ωστόσο περιέχουν πληροφορίες που κρύβονται μέσα στις τιμές τους. Αυτές οι πληροφορίες, θα πρέπει με κάποιον τρόπο να εξαχθούν, έτσι ώστε να μπορεί να γίνει επεξεργασία αυτών των δεδομένων με μαθηματικές μεθόδους. Ένας τρόπος για να γίνει αυτό είναι να εφαρμοστούν μέθοδοι από τη θεωρία της πληροφορίας [85]. Αυτές οι μέθοδοι αναφέρονται σε μέτρα ομοιότητας που, ουσιαστικά,

λαμβάνουν υπόψη τη συχνότητα εμφάνισης των κατηγορικών τιμών στο σύνολο δεδομένων. Σε πρόσφατες δημοσιεύσεις [86, 87], προτείνονται νέα μέτρα για τον υπολογισμό της ομοιότητας μεταξύ κατηγορικών τιμών λαμβάνοντας υπόψη, όχι μόνο την ομοιότητα των κατηγοριών που ανήκουν σε μία συγκεκριμένη κατηγορική μεταβλητή (intra-coupled similarity), αλλά και την ομοιότητα μεταξύ κατηγοριών που ανήκουν σε άλλες μεταβλητές (inter-coupled similarity). Σε αντίθεση με την απλούστερη χωρίς επίβλεψη μέθοδο, που χρησιμοποιείται από αυτές τις τεχνικές, οι περισσότερες από τις πληροφορίες που κρύβονται σε ένα σύνολο κατηγορικών δεδομένων θα μπορούσαν να αξιοποιηθούν με έναν εποπτευόμενο τρόπο, δηλαδή λαμβάνοντας υπόψη τη συσχέτιση των κατηγορικών τιμών με τη μεταβλητή εξόδου. Αυτή η μέθοδος υιοθετήθηκε για να παραχθεί η τεχνική value difference metric [88], ενώ ακολούθησαν πολλές παραλλαγές της [89, 90]. Μερικές από τις προαναφερθείσες μεθόδους, ή άλλες παρόμοιας φύσης, ενσωματώθηκαν σε σχήματα μάθησης που χρησιμοποιούν αριθμητικές αποστάσεις προκειμένου να κάνουν συσταδοποίηση [87, 91-97] και/ή ταξινόμηση [98-100], εξοπλίζοντάς τα με την ικανότητα να χειρίζονται κατηγορικά δεδομένα.

Τα δίκτυα RBF έχουν χρησιμοποιηθεί εκτενώς και με μεγάλη επιτυχία για αριθμητικά δεδομένα [29], αλλά δυστυχώς, δεν είναι σε θέση να διαχειριστούν άμεσα κατηγορικά δεδομένα. Ο βασικός λόγος είναι ότι η λειτουργία των δικτύων RBF απαιτεί τον υπολογισμό των αποστάσεων μεταξύ των δεδομένων εισόδου, αλλά λόγω της φύσης των κατηγορικών δεδομένων, ο υπολογισμός αυτός, δεν είναι εφικτός.

Ο **τέταρτος στόχος** αυτής της διατριβής αφορά στην ανάπτυξη ενός νέου καινοτόμου συστήματος μάθησης επεκτείνοντας την αρχιτεκτονική των δικτύων RBF, ώστε να είναι εφικτό να χειριστεί κατηγορικά δεδομένα, εισάγοντας την έννοια των κατηγορικών δικτύων RBF (categorical RBF, CRBF). Η προτεινόμενη αρχιτεκτονική νευρωνικών δικτύων [101] αντικαθιστά τα αριθμητικά διανύσματα που είναι γνωστά ως κέντρα RBF, με αντίστοιχα κατηγορικής φύσης κέντρα και χρησιμοποιεί ειδικά σχεδιασμένα μέτρα για τον υπολογισμό της απόστασης ανάμεσα στα κέντρα και τα κατηγορικά δεδομένα εισόδου. Επιπλέον, προτείνεται ένας γρήγορος μη επαναληπτικός αλγόριθμος κατηγορικής συσταδοποίησης (categorical center selection, CCS), για την πραγματοποίηση του πρώτου σταδίου της εκπαίδευσης του δικτύου RBF, που περιλαμβάνει την επιλογή των κέντρων των νευρώνων της κρυφής στοιβάδας. Στο δεύτερο στάδιο της εκπαίδευσης του δικτύου RBF, τα συναπτικά βάρη υπολογίζονται με γραμμική παλινδρόμηση. Το νέο σύστημα εφαρμόζεται σε 22 κατηγορικά σύνολα δεδομένων και συγκρίνεται με αρκετά διαφορετικά σχήματα μάθησης.

Στον πραγματικό κόσμο, τα σύνολα δεδομένων συνήθως περιέχουν ταυτόχρονα ποσοτικές και ποιοτικές τιμές, δηλαδή μικτά δεδομένα. Ωστόσο, οι περισσότερες υπάρχουσες μέθοδοι διαχειρίζονται όλες τις μεταβλητές είτε ως αριθμητικές, είτε ως κατηγορικές [102]. Στην περίπτωση των μικτών δεδομένων, οι περισσότερες μέθοδοι μηχανικής μάθησης συνήθως χρησιμοποιούν μεθόδους μετασχηματισμού για να μετατρέψουν τις κατηγορικές μεταβλητές σε αριθμητικές, ή το αντίθετο, και στη συνέχεια εφαρμόζουν γνωστούς αλγορίθμους που μπορούν όμως να χειριστούν μόνο ομοιογενή δεδομένα [103]. Στις περισσότερες περιπτώσεις, όμως, οι μηχανισμοί μετασχηματισμού μπορεί να οδηγήσουν σε απώλεια πληροφοριών και στη συνέχεια σε λάθος αποτελέσματα [104]. Με σκοπό τη διαχείριση μικτού τύπου δεδομένων, αρκετοί ερευνητές έχουν διευρύνει τη λειτουργία του αλγορίθμου *k*-prototypes για τον χειρισμό μικτών δεδομένων. Οι Bezdek et. al. [105] πρότειναν τη μέθοδο fuzzy *k*-prototypes, η οποία παίρνει υπόψη την ασαφή φύση των δεδομένων. Στην εργασία [102], παρουσιάστηκε μία βελτιωμένη έκδοση του αλγορίθμου *k*-prototypes για τη συσταδοποίηση μικτών δεδομένων, όπου εισήχθη η έννοια του “distribution centroid”, ενώ παρουσιάστηκε και ένα είδος εξελικτικού αλγορίθμου *k*-prototypes [106].

Επίσης, προτάθηκαν ο αλγόριθμος KL-FCM-GM [107], ο οποίος βασίζεται στην υπόθεση ότι τα δεδομένα που προέρχονται από συστάδες ακολουθούν κατανομή Gauss, και το σύστημα SpectralCAT [108], για τον χειρισμό μικτών δεδομένων.

Όπως προκύπτει από μία ανασκόπηση στη βιβλιογραφία πολύ μικρή έρευνα έχει διεξαχθεί σχετικά με τα μικτά σύνολα δεδομένων [109], αν και η διαχείριση μικτού τύπου δεδομένων αποτελεί πρόκληση σε πολλές εφαρμογές. Όσον αφορά στα δίκτυα RBF, ένας μικρός αριθμός αλγορίθμων εκπαίδευσης μπορούν να διαχειρίζονται κατηγορικά δεδομένα, και ένας πολύ μικρότερος αριθμός μπορούν να διαχειρίζονται μικτά δεδομένα.

Ο **πέμπτος στόχος** αυτής της διατριβής περιλαμβάνει την ανάπτυξη μιας νέας αρχιτεκτονικής νευρωνικών δικτύων, με την ικανότητα να λαμβάνει υπόψη ξεχωριστά την ιδιαίτερη φύση των αριθμητικών και κατηγορικών μεταβλητών, και ταυτόχρονα να μπορεί να χειριστεί δεδομένα μεγάλου όγκου (big data). Το προτεινόμενο σχήμα [110] βασίζεται στα κλασσικά δίκτυα RBF και τα δίκτυα CRBF και ονομάστηκε μικτά δίκτυα RBF (mixed RBF, MRBF). Η εκπαίδευση των δικτύων MRBF χωρίζεται σε δύο στάδια: (α) στο πρώτο στάδιο, καθορίζονται τα αριθμητικά και τα κατηγορικά κέντρα με τη χρήση ενός νέου αλγορίθμου, ο οποίος δημιουργήθηκε στα πλαίσια αυτής της διατριβής και αποτελεί συνδυασμό των αλγορίθμων FM και CCS, (β) στο δεύτερο στάδιο, μειώνονται τα ενεργά κέντρα με τη μέθοδο ανάλυσης κύριων συνιστωσών (principal component analysis, PCA) [111] και υπολογίζονται τα συναπτικά βάρη με γραμμική παλινδρόμηση. Η προτεινόμενη μέθοδος συγκρίνεται με αρκετά διαφορετικά σχήματα μάθησης και εφαρμόζεται σε 12 σύνολα μικτών δεδομένων που περιλαμβάνουν προβλήματα προσέγγισης συνάρτησης και ταξινόμησης και κάποια από τα οποία ανήκουν στην κατηγορία δεδομένων μεγάλου όγκου.

Τα δίκτυα MRBF, λόγω της ικανότητάς τους να λαμβάνουν υπόψη αριθμητικές και κατηγορικές τιμές και να χειρίζονται δεδομένα μεγάλου όγκου, μπορούν να δώσουν λύση σε προβλήματα που αφορούν στον χώρο του παγκόσμιου ιστού. Κατά την τελευταία δεκαετία, ο παγκόσμιος ιστός (Word Wide Web, WWW) αποτελεί ένα δημοφιλές και διαδραστικό μέσο για τη διάδοση ενός τεράστιου όγκου πληροφοριών, το οποίο έχει διαπεράσει τις περισσότερες πλευρές της καθημερινής ζωής των ανθρώπων. Ο παγκόσμιος ιστός έχει τύχει ευρείας αποδοχής λόγω της απλότητας που παρουσιάζει στη χρήση του. Αυτή η απλότητα, όμως, ενδέχεται να αποτελέσει εμπόδιο στην εξέλιξη του παγκόσμιου ιστού. Καθημερινά στον παγκόσμιο ιστό προστίθεται ένα μεγάλος αριθμός νέων ιστοσελίδων, με αποτέλεσμα ο πρώτος, να αντιπροσωπεύει μια συνεχώς αυξανόμενη πηγή τεράστιου όγκου πληροφοριών, η οποία είναι διαθέσιμη σε όλους τους χρήστες του. Ωστόσο, το μεγαλύτερο μέρος αυτής της πληροφορίας δεν είναι προσπελάσιμο από τις μηχανές, καθώς η εισαγωγή των ιστοσελίδων δεν γίνεται με οργανωμένο τρόπο και οι πληροφορίες που εισάγονται δεν είναι δομημένες. Η περιγραφή και οργάνωση του τεράστιου περιεχομένου, που είναι διαθέσιμο μέσω του ιστού, είναι απαραίτητη για την αξιοποίηση του πλήρους δυναμικού του ως πηγή πληροφοριών, με απώτερο στόχο την εξέλιξη του παγκόσμιου ιστού σε μία ισχυρότερη έκδοση.

Ο σημασιολογικός ιστός (Semantic Web) αποτελεί την επέκταση του παγκόσμιου ιστού. Ουσιαστικά πρόκειται για μία ερευνητική προσπάθεια σε εξέλιξη, όπου η πληροφορία είναι εμπλουτισμένη με μεταδεδομένα (metadata), με αποτέλεσμα οι υπολογιστές να μπορούν να αντιλαμβάνονται τη σημασία της πληροφορίας που διαχειρίζονται [112]. Ουσιαστικά, ο σημασιολογικός ιστός είναι ο παγκόσμιος ιστός με την προσθήκη μεταδεδομένων, δηλαδή σχολιασμών. Ωστόσο, η διαδικασία της προσθήκης μεταδεδομένων στην υπάρχουσα πληροφορία, δηλαδή το έργο της επισημείωσης (annotation), με μη αυτόματο τρόπο αποτελεί εμπόδιο στην εξέλιξη του παγκόσμιου ιστού. Μία λύση στο πρόβλημα αυτό, αποτελούν οι αυτόματες ή ημιαυτόματες τεχνικές επισημείωσης. Παραδείγματα τέτοιων τεχνικών αποτελούν τα συστήματα σχολιασμού GATE [113], NCBO [114] και cTAKES [115].

Επίσης, στην εργασία [116], χρησιμοποιείται μία τεχνική μηχανικής μάθησης για τη δημιουργία μεταδεδομένων.

Η εξέλιξη του παγκόσμιου ιστού μπορεί να επέλθει με την προσθήκη μεταδεδομένων στον ιστό, το έργο, όμως, της επισήμανσης ολόκληρου του ιστού ταυτόχρονα είναι δύσκολο, δεδομένου ότι η διαθέσιμη τεχνολογία για την αυτόματη επισημείωση δεν είναι ακόμα ώριμη. Με στόχο την ευκολότερη αντιμετώπιση αυτού του έργου, προτάθηκε το τελευταίο να χωριστεί σε κομμάτια που αντιπροσωπεύουν διαφορετικά πεδία, όπου το πρώτο βήμα είναι η ταξινόμηση αντικειμένων ιστού, συνήθως ιστοσελίδων [117].

Στον παγκόσμιο ιστό, η ταξινόμηση παίζει ζωτικό ρόλο. Η ταξινόμηση του περιεχομένου των ιστοσελίδων είναι απαραίτητη για αρκετές λειτουργίες στο ιστό, όπως είναι για παράδειγμα οι εστιασμένοι ανιχνευτές (focused crawling) [118]. Σύμφωνα με την έρευνα που διεξήγαγαν οι Qi και Davison [118], το πρόβλημα ταξινόμησης ιστοσελίδων έχει διαιρεθεί σε υπο-προβλήματα, όπως είναι η ταξινόμηση ιστοσελίδας ανάλογα με τη λειτουργία της (functional classification) (π.χ. αν είναι μία ιστοσελίδα ατόμου, ή ιστοσελίδα πανεπιστημίου, κτλ.), το είδος της (genre classification) (π.χ. αν είναι μία ιστοσελίδα διαφήμισης, ή ερωτήσεων, ή online αγορών, κτλ.), τη θεματική της περιοχή (subject classification) (π.χ. αν είναι μία ιστοσελίδα που αναφέρεται στη βιολογία, ή στη χημεία, κτλ.), κ.α. Η ταξινόμηση ιστοσελίδων διαδραματίζει σημαντικό ρόλο τα τελευταία χρόνια με την ταχεία ανάπτυξη των πληροφοριών στον ιστό, ιδίως στα κοινωνικά δίκτυα, στα ιστολόγια και στα φόρουμ [119]. Είναι εύκολα κατανοητό ότι επισημειώνοντας το περιεχόμενο των ιστοσελίδων με ετικέτες και οργανώνοντάς το σε κατηγορίες, γίνεται ευκολότερη η αναζήτηση και η εύρεση της επιθυμητής πληροφορίας. Το πρόβλημα της ταξινόμησης είναι ένα υποσύνολο του γενικότερου προβλήματος της αυτόματης ανάλυσης του περιεχομένου των ιστοσελίδων, που αφορά στην απόδοση σημασιολογικών ετικετών σε αυτά. Με την ταχεία ανάπτυξη του όγκου των ιστοσελίδων, η αυτόματη ταξινόμηση του περιεχομένου τους είναι μια αποτελεσματική λύση στην οργάνωση της πληροφορίας και τη διαχείριση της γνώσης.

Ο **έκτος στόχος** αυτής της διατριβής αφορά στην ανάπτυξη ενός ευφυούς συστήματος, το οποίο μπορεί να ταξινομεί αυτόματα μία ιστοσελίδα με βάση το περιεχόμενό της. Ο απώτερος σκοπός είναι η υιοθέτηση του συστήματος ως βάση για τον εμπλουτισμό του περιεχομένου του ιστού με μεταδεδομένα, και άρα για την εξέλιξη του παγκόσμιου ιστού. Το προτεινόμενο σύστημα [120] ονομάστηκε MRBF-WPC (web page classification) και βασίζεται στα δίκτυα MRBF. Επίσης, χρησιμοποιεί βασικά στοιχεία του τομέα της εξόρυξης γνώσης από τον ιστό (web mining) και συγκεκριμένα του πιο ευρέως διερευνώμενου άξονά της, που είναι η εξόρυξη γνώσης από το περιεχόμενο του ιστού (web content mining). Το σύστημα αυτό, αποτελεί ένα ολοκληρωμένο σχήμα για την επίλυση του προβλήματος ταξινόμησης ιστοσελίδων, και χρησιμοποιεί ένα σχετικά μικρό αριθμό χαρακτηριστικών, από αριθμητικά και κατηγορικά δεδομένα. Επίσης, το σύστημα αυτό αποτελεί ένα γενικής χρήσης εργαλείο ταξινόμησης ιστοσελίδων, το οποίο μπορεί για παράδειγμα να ταξινομήσει μία ιστοσελίδα ανάλογα με τη λειτουργία της (functional classification), ή το είδος της (genre classification), ή τη θεματική περιοχή στην οποία ανήκει (subject classification), κ.α. Το σύστημα MRBF-WPC αξιολογείται με βάση το σύνολο ιστοσελίδων WebKB [121, 122] και συγκρίνεται με μία ποικιλία μεθόδων μηχανικής μάθησης.

Στη συνέχεια, παρουσιάζεται η δομή της παρούσας διδακτορικής διατριβής, η οποία είναι η ακόλουθη:

Το πρώτο κεφάλαιο αποτελεί μια εισαγωγή στους βασικούς κλάδους της επιστήμης της υπολογιστικής νοημοσύνης. Το ενδιαφέρον εστιάζεται στα τεχνητά νευρωνικά δίκτυα, στον εξελικτικό υπολογισμό και στη νοημοσύνη σμήνους. Πιο συγκεκριμένα παρουσιάζονται

οι πιο γνωστές αρχιτεκτονικές νευρωνικών δικτύων, η διαδικασία εκμάθησής τους, ο υβριδικός αλγόριθμος εξελικτικής προσομοιωμένης ανόπτησης και η τεχνική βελτιστοποίησης σμήνους σωματιδίων.

Στο δεύτερο κεφάλαιο επιχειρείται μία ανασκόπηση της αρχιτεκτονικής νευρωνικών δικτύων RBF, όπου αναλύεται η δομή τους και ο τυπικός αλγόριθμος εκπαίδευσής τους. Επίσης, αναλύεται η ασαφής διαμέριση και η μη συμμετρική ασαφής διαμέριση, ενώ παράλληλα παρουσιάζονται ο συμμετρικός και ο μη συμμετρικός αλγόριθμος των ασαφών μέσων. Επιπλέον, γίνεται μία σύγκριση μεταξύ των δικτύων RBF και της πιο βασικής αρχιτεκτονικής νευρωνικών δικτύων.

Στο τρίτο κεφάλαιο περιγράφονται οι τεχνικές εκπαίδευσης νευρωνικών δικτύων RBF που αναπτύχθηκαν στα πλαίσια αυτής της διδακτορικής διατριβής για αριθμητικές μεταβλητές εισόδου. Συγκεκριμένα, αναλύονται οι αλγόριθμοι PSO-NSFM, COOP-VW, και ESA-NSFM. Επίσης, παρουσιάζονται εφαρμογές αυτών των αλγορίθμων που περιλαμβάνουν ποικίλα επιστημονικά πεδία, και συγκρίσεις των προτεινόμενων αλγορίθμων με άλλες εναλλακτικές μεθόδους μηχανικής μάθησης.

Στο τέταρτο κεφάλαιο γίνεται μία σύντομη εισαγωγή στα κατηγορικά δεδομένα και στα μέτρα απόστασης που είναι κατάλληλα για κατηγορικές τιμές. Στη συνέχεια, αναλύονται οι δύο νέες αρχιτεκτονικές νευρωνικών δικτύων CRBF και MRBF και οι αλγόριθμοι εκπαίδευσής τους, που αναπτύχθηκαν στα πλαίσια αυτής της διατριβής για τον χειρισμό κατηγορικών και μικτών μεταβλητών, αντίστοιχα. Επίσης, παρουσιάζεται μία ποικιλία εφαρμογών αυτών των αρχιτεκτονικών, και συγκρίσεις με άλλα σχήματα μάθησης.

Στο πέμπτο κεφάλαιο, γίνεται μία σύντομη εισαγωγή στον παγκόσμιο ιστό και στην εξέλιξή του σε σημασιολογικό ιστό. Επίσης, γίνεται μία σύντομη αναφορά στα εμπόδια επέκτασης του παγκόσμιου ιστού. Στη συνέχεια, περιγράφεται το ολοκληρωμένο ευφυές σύστημα αυτόματης κατηγοριοποίησης ιστοσελίδων με βάση το περιεχόμενο τους, που μπορεί να αποτελέσει μία βάση στην εξέλιξη του παγκόσμιου ιστού, και αναπτύχθηκε στα πλαίσια αυτής της διδακτορικής διατριβής.

Τέλος, συνοψίζονται τα βασικότερα συμπεράσματα που προέκυψαν κατά τη διάρκεια εκπόνησης της παρούσας διδακτορικής διατριβής.

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΥΠΟΛΟΓΙΣΤΙΚΗ ΝΟΗΜΟΣΥΝΗ

Στο παρόν κεφάλαιο ορίζονται οι κύριες έννοιες και μεθοδολογίες που άπτονται της θεματολογίας της παρούσας διατριβής. Συγκεκριμένα, περιγράφονται οι κλάδοι της υπολογιστικής νοημοσύνης, τα τεχνητά νευρωνικά δίκτυα, η ασαφής λογική, ο εξελικτικός υπολογισμός, η νοημοσύνη σμήνους και τα τεχνητά ανοσοποιητικά συστήματα.

Το ενδιαφέρον εστιάζεται κυρίως στα τεχνητά νευρωνικά δίκτυα, όπου αναλύεται η σχέση τους με τα βιολογικά νευρωνικά δίκτυα. Παρουσιάζεται επίσης μία ιστορική αναδρομή κατά την οποία περιγράφεται σε γενικές γραμμές πως ξεκίνησε η ιδέα αυτή και ποια είναι τα σημαντικότερα βήματα στην ανάπτυξή της. Στη συνέχεια, αναλύεται το πρώτο απλό νευρωνικό δίκτυο δίνοντας έμφαση στη δομή και στη λειτουργία του. Παρατίθενται, επίσης, οι τρεις βασικοί τύποι συναρτήσεων ενεργοποίησης των νευρώνων και παρουσιάζονται συνοπτικά οι πιο γνωστές αρχιτεκτονικές νευρωνικών δικτύων. Τέλος, αναλύεται η διαδικασία εκμάθησης ενός νευρωνικού δικτύου, ο αλγόριθμος εκπαίδευσης με διόρθωση σφάλματος, τα προβλήματα μάθησης που μπορούν να επιλυθούν με χρήση των νευρωνικών δικτύων και τέσσερα κριτήρια απόδοσης των μοντέλων.

1.1 Υπολογιστική νοημοσύνη

Ένα από τα βασικά χαρακτηριστικά που ξεχωρίζει τον άνθρωπο από τα υπόλοιπα έμβια όντα είναι σαφέστατα η ικανότητα λογικής και δημιουργικής σκέψης. Ενώ υπάρχουν πολλά ζώα που επιδεικνύουν κάποιες ικανότητες ευρηματικότητας, το εύρος των ικανοτήτων τους είναι περιορισμένο συγκριτικά με αυτών του ανθρώπου. Κάτι ανάλογο ισχύει και για τους υπολογιστές. Οι υπολογιστές μπορούν να εκτελούν πολλές και διαφορετικές εργασίες ταυτόχρονα πολύ πιο γρήγορα από έναν άνθρωπο. Ωστόσο, ακόμα και ο ισχυρότερος υπολογιστής υστερεί σε σχέση με τον ανθρώπινο εγκέφαλο στον τομέα της αντίληψης και της δημιουργικής σκέψης.

Μπορούν οι μηχανές να αποκτήσουν νοημοσύνη;

Το 1950 ο Άγγλος μαθηματικός Alan Turing έθεσε το παραπάνω ερώτημα. Βεβαίως, το ερώτημα αυτό δεν μπορεί να απαντηθεί ακόμα με βεβαιότητα. Η υπόθεση του Alan Turing ότι μπορούν να αναπτυχθούν υπολογιστικά συστήματα που θα μιμούνται τη λειτουργία του ανθρώπινου εγκεφάλου [123], έχει εμπνεύσει τεράστια ερευνητική προσπάθεια προκειμένου οι υπολογιστές να αποκτήσουν νοημοσύνη.

Η νοημοσύνη είναι μια έννοια που συνδέεται με την ικανότητα επίλυσης προβλημάτων, τη μάθηση από την εμπειρία, την ικανότητα συλλογισμού, την κατανόηση, την ορθολογιστική και αναλυτική σκέψη, την εξαγωγή συμπερασμάτων, κτλ. Στο λεξικό Merriam-Webster η νοημοσύνη ορίζεται ως «η ικανότητα για μάθηση ή κατανόηση ή η αντιμετώπιση νέων ή δύσκολων καταστάσεων». Ένας πιο σύντομος, αλλά ίσως και πιο περιεκτικός ορισμός είναι ο εξής:

Νοημοσύνη είναι το υπολογιστικό κομμάτι της ικανότητας μιας οντότητας να επιτυγχάνει στόχους.

Η επιστήμη η οποία έχει σκοπό να δώσει, στα μέτρα που αυτό είναι δυνατό, ανθρώπινες δυνατότητες δηλαδή νοημοσύνη στους υπολογιστές είναι η τεχνητή νοημοσύνη (artificial intelligence, AI). Το επιστημονικό πεδίο της τεχνητής νοημοσύνης θεμελιώθηκε επίσημα το 1956 ύστερα από πρόταση του John McCarthy σε μια συνάντηση εργασίας με αντικείμενο τη μελέτη της ευφυΐας, στην οποία συμμετείχαν επιφανείς επιστήμονες όπως ο John McCarthy, ο Claude Shannon, ο Marvin Minsky και ο Nathaniel Rochester, στο Dartmouth της Μασαχουσέτης. Ένας από του πρώτους ορισμούς που δόθηκε για την τεχνητή νοημοσύνη είναι από τους Barr και Feigenbaum [124]:

Τεχνητή νοημοσύνη είναι ο τομέας που ασχολείται με τη σχεδίαση ευφυών υπολογιστικών συστημάτων, δηλαδή συστημάτων που επιδεικνύουν χαρακτηριστικά που σχετίζονται με τη νοημοσύνη στην ανθρώπινη συμπεριφορά.

Η τεχνητή νοημοσύνη διακρίνεται στην κλασική ή συμβολική τεχνητή νοημοσύνη (symbolic AI) και στην υπολογιστική νοημοσύνη (computational intelligence, CI). Η κλασσική τεχνητή νοημοσύνη βασίζεται στην κατανόηση των νοητικών διεργασιών και προσπαθεί να προσομοιώσει την ανθρώπινη νοημοσύνη, προσεγγίζοντάς τη με συστήματα βασισμένα στη γνώση που χρησιμοποιούν ως δομικές μονάδες τα σύμβολα (π.χ. συστήματα κανόνων). Η υπολογιστική νοημοσύνη βασίζεται στη μίμηση μηχανισμών που παρατηρούνται στη φύση και ειδικότερα στα έμβια όντα, όπως παραδείγματος χάρη η διαδικασία μάθησης του ανθρώπινου εγκεφάλου, η διαδικασία εξέλιξης των ειδών, κτλ. Στη

συνέχεια, δίνεται μια γενική εικόνα του κλάδου της υπολογιστικής νοημοσύνης, όπου επικεντρώνεται η παρούσα διδακτορική διατριβή.

Ο όρος της υπολογιστικής νοημοσύνης εισήχθη για πρώτη φορά το 1994 στο παγκόσμιο συνέδριο υπολογιστικής νοημοσύνης (World Congress on Computational Intelligence). Το συνέδριο εμπεριείχε τρία επιμέρους συνέδρια: το διεθνές συνέδριο νευρωνικών δικτύων της IEEE (IEEE International Conference on Neural Networks), το διεθνές συνέδριο ασαφούς λογικής της IEEE (IEEE International Conference on Fuzzy Systems), και το διεθνές συνέδριο εξελικτικού υπολογισμού της IEEE (IEEE Conference on Evolutionary Computation). Τότε, οι τεχνολογίες των νευρωνικών δικτύων, της ασαφούς λογικής και του εξελικτικού υπολογισμού συγκεντρώθηκαν κάτω από τον τίτλο της υπολογιστικής νοημοσύνης και ο Bezdek [125], ένας από τους πρωτοπόρους του χώρου, αναφέρει ότι:

> *... ένα σύστημα είναι υπολογιστικά νοήμον όταν χειρίζεται μόνο αριθμητικά δεδομένα (χαμηλού επιπέδου), έχει ένα τμήμα αναγνώρισης προτύπων, και δεν χρησιμοποιεί τη γνώση υπό την έννοια της τεχνητής νοημοσύνης, και επιπλέον, όταν (αρχίζει να) διαθέτει: (i) υπολογιστική προσαρμοστικότητα, (ii) ανοχή στα υπολογιστικά σφάλματα, (iii) ταχύτητα η οποία προσεγγίζει την ανθρώπινη ευστροφία, και iv) ποσοστά αποτυχίας που πλησιάζουν στα αντίστοιχα των ανθρώπων.*

Παρότι από τη στιγμή που εμφανίστηκε η έννοια της υπολογιστικής νοημοσύνης έχουν δημοσιευθεί πλήθος επιστημονικών εργασιών, ωστόσο δεν υπάρχει κάποιος κοινά αποδεκτός ορισμός. Σήμερα, η υπολογιστική νοημοσύνη αποτελεί ένα αυτοτελές επιστημονικό πεδίο και η γενική προσέγγιση που ακολουθείται από τους ερευνητές του χώρου είναι να αντιμετωπίζεται ως μία ομπρέλα υπό την οποία προστίθενται συνέχεια περισσότερες «ευφυείς» τεχνικές. Ένα γενικός ορισμός για την υπολογιστική νοημοσύνη θα μπορούσε να είναι ο παρακάτω:

> *Η υπολογιστική νοημοσύνη είναι ένας σχετικά νέος και αναπτυσσόμενος επιστημονικός κλάδος, ο οποίος ασχολείται με τη μοντελοποίηση της βιολογικής και φυσικής νοημοσύνης, έχοντας ως αποτέλεσμα τα λεγόμενα ”ευφυή συστήματα”.*

Ο Engelbrecht [1] υποστηρίζει ότι η υπολογιστική νοημοσύνη είναι ένας κλάδος της τεχνητής νοημοσύνης ο οποίος αφορά στη μελέτη προσαρμοστικών μηχανισμών οι οποίοι κάνουν δυνατή την ευφυή συμπεριφορά σε περίπλοκα και δυναμικά περιβάλλοντα. Αυτοί οι μηχανισμοί, περιλαμβάνουν τα παραδείγματα της τεχνητής νοημοσύνης που παρουσιάζουν την ικανότητα μάθησης ή προσαρμογής σε νέες καταστάσεις, καθώς και τις ικανότητες γενίκευσης, αφαίρεσης, ανακάλυψης και συσχέτισης. Οι τεχνικές που ανήκουν στον κλάδο της υπολογιστικής νοημοσύνης είναι ικανές να επιλύουν σύνθετα προβλήματα μηχανικής, τα οποία αποτελούσαν πάντα μεγάλες προκλήσεις. Σύμφωνα με τον Engelbrecht [1], οι τεχνικές της υπολογιστικής νοημοσύνης συνήθως χωρίζονται στις παρακάτω κατηγορίες:

- Τεχνητά νευρωνικά δίκτυα
- Ασαφή συστήματα
- Εξελικτικός υπολογισμός
- Νοημοσύνη σμήνους
- Τεχνητά ανοσοποιητικά συστήματα

Στη συνέχεια του κεφαλαίου, περιγράφονται οι βασικές τεχνικές υπολογιστικής νοημοσύνης, ενώ δίνεται βαρύτητα στις μεθόδους που χρησιμοποιήθηκαν σε αυτή τη διδακτορική διατριβή.

1.2 Νευρωνικά δίκτυα

Τα τεχνητά νευρωνικά δίκτυα (artificial neural networks, ANN) [2], ή πιο απλά νευρωνικά δίκτυα αποτελούν ένα σύνολο από ισχυρά μαθηματικά εργαλεία, η θεμελιώδης αρχή των οποίων έχει τις βάσεις της στη λειτουργία του ανθρωπίνου εγκεφάλου να μαθαίνει από τα δεδομένα που υπάρχουν στο περιβάλλον του, να αποθηκεύει αυτή τη γνώση και να μπορεί να τη χρησιμοποιεί στη συνέχεια σε νέες καταστάσεις. Τα νευρωνικά δίκτυα μπορούν να θεωρηθούν ως ένα είδος μαύρου-κουτιού (black-box), αφού μπορούν να μοντελοποιήσουν οποιοδήποτε άγνωστο μη γραμμικό σύστημα χρησιμοποιώντας μόνο δεδομένα εισόδου-εξόδου από αυτό. Στη βιβλιογραφία, κατά καιρούς, έχουν προταθεί αρκετοί ορισμοί για τα νευρωνικά δίκτυα. Ένα μεγάλο μέρος αυτών κάλυψε ο Haykin [2] με τον παρακάτω ορισμό:

> *Νευρωνικό δίκτυο ονομάζουμε έναν μαζικά κατανεμημένο παράλληλα επεξεργαστή, αποτελούμενο από απλές υπολογιστικές μονάδες που έχει την ικανότητα να αποθηκεύει εμπειρική γνώση και να την κάνει διαθέσιμη για χρήση. Μοιάζει με τον ανθρώπινο εγκέφαλο σε δυο σημεία:*
>
> 1. *Η γνώση αποκτάται από το περιβάλλον του δικτύου μέσα από μια διαδικασία μάθησης.*
> 2. *Η «ισχύς» των συνάψεων ανάμεσα στους νευρώνες, γνωστή και ως συναπτικά βάρη, χρησιμοποιείται για να αποθηκευθεί η αποκτούμενη γνώση.*

Ένα τεχνητό νευρωνικό δίκτυο κατασκευάζεται μελετώντας τα μεμονωμένα κύτταρα που αποτελούν τον ανθρώπινο εγκέφαλο. Ο ανθρώπινος εγκέφαλος είναι ένας εξαιρετικά δυνατός “υπολογιστής” που αποτελείται από ένα δίκτυο με πάνω από εκατό δισεκατομμύρια διασυνδεμένων νευρώνων [126] και μπορεί να αντιμετωπιστεί ως ένα σύστημα τριών επιπέδων, όπως φαίνεται στο σχήμα 1.1 [127]. Το κεντρικό σύστημα, που είναι ο εγκέφαλος, αναπαρίσταται ως ένα νευρωνικό δίκτυο, το οποίο συνέχεια λαμβάνει πληροφορίες, τις επεξεργάζεται και ύστερα αποφασίζει. Στο σχήμα 1.1 διακρίνονται δύο ζευγάρια από βέλη. Τα βέλη με κατεύθυνση από αριστερά προς τα δεξιά υποδεικνύουν την προώθηση της πληροφορίας στο σύστημα, ενώ τα βέλη με αντίθετη φορά δηλώνουν την ανατροφοδότηση του συστήματος. Οι δέκτες (receptors) μετατρέπουν τα ερεθίσματα από το ανθρώπινο σώμα, ή το εξωτερικό περιβάλλον, σε ηλεκτρικούς παλμούς οι οποίοι μεταφέρουν τις πληροφορίες στο νευρωνικό δίκτυο. Οι ενεργοποιητές (effectors) μετατρέπουν τους ηλεκτρικούς παλμούς που παράγονται από το νευρωνικό δίκτυο σε αντιδράσεις, που είναι η έξοδος του

Σχήμα 1.1. Σχηματικό διάγραμμα νευρωνικού συστήματος.

συστήματος.

Η προσπάθεια κατανόησης της λειτουργίας του εγκεφάλου ξεκίνησε με τον Ramón y Cajál ο οποίος παρουσίασε τους νευρώνες ως δομικά συστατικά στοιχεία του εγκεφάλου [128]. Τυπικά, οι νευρώνες του ανθρώπινου εγκεφάλου είναι πέντε με έξη τάξεις μεγέθους πιο αργοί από τις λογικές πύλες που χρησιμοποιούνται στα τσιπ σιλικόνης των ηλεκτρονικών υπολογιστών. Ωστόσο, ο ανθρώπινος εγκέφαλος αντισταθμίζει τον σχετικά αργό ρυθμό λειτουργίας ενός νευρώνα με την ύπαρξη ενός τεράστιου πλήθους νευρώνων, με εξίσου τεράστιο πλήθος διασυνδέσεων μεταξύ τους.

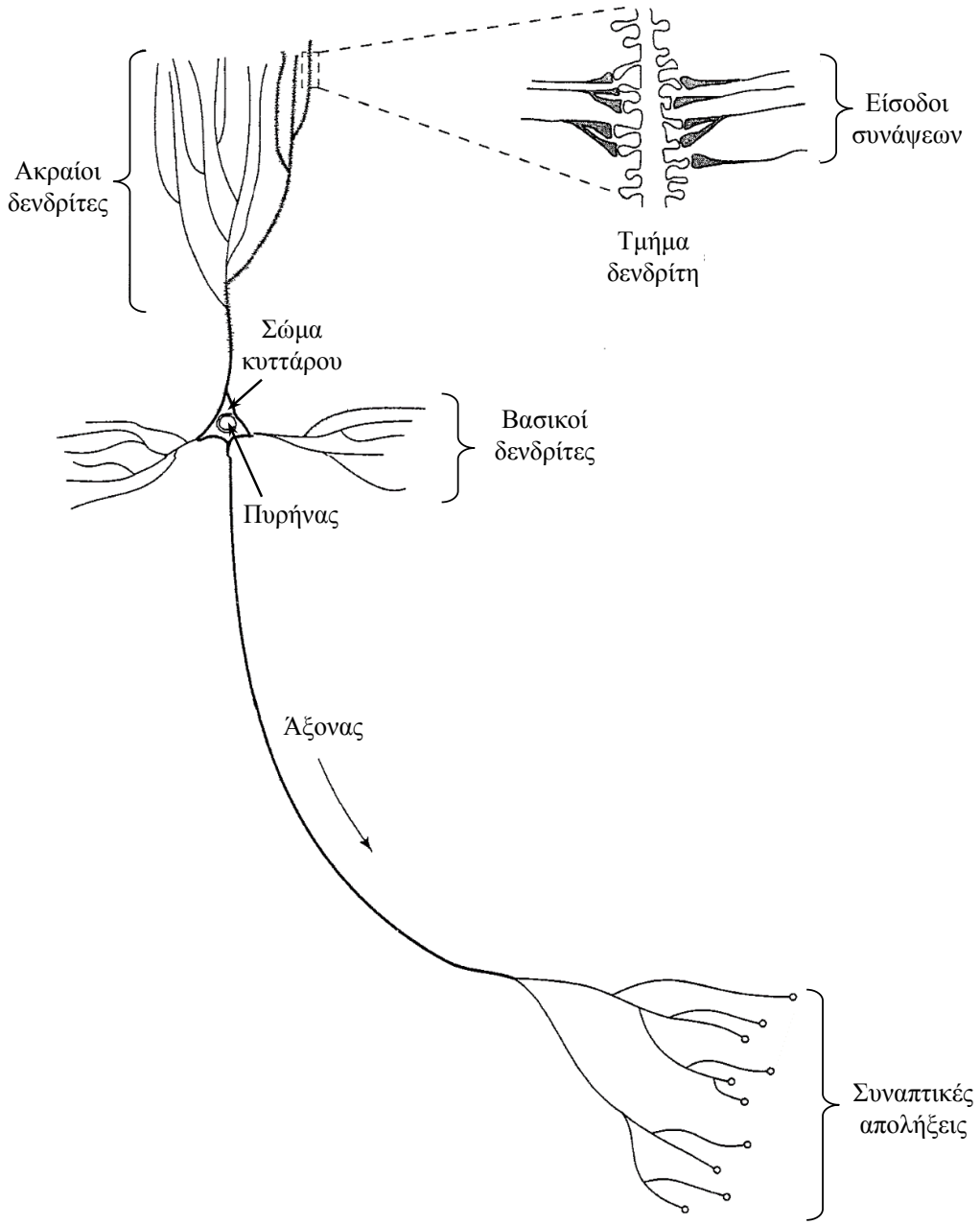
Το νευρικό σύστημα των ζώντων οργανισμών αποτελείται από πολλά νευρωνικά δίκτυα τα οποία είναι εξειδικευμένα σε διεργασίες όπως η μάθηση, η μνήμη, η όραση, κτλ. Η κεντρική μονάδα του νευρικού συστήματος όπως αναφέρθηκε προηγουμένως είναι ο εγκέφαλος, ο οποίος επίσης αποτελείται από νευρωνικά δίκτυα. Κάθε νευρωνικό δίκτυο αποτελείται από ένα μεγάλο αριθμό μονάδων, που λέγονται νευρώνες ή νευρώνια (neurons) οι οποίοι αποτελούνται από το κυρίως σώμα, τον άξονα και τους δενδρίτες, όπως φαίνεται στο σχήμα 1.2 [2].

Οι νευρώνες, που είναι η βασική δομική μονάδα του εγκεφάλου, είναι μεμονωμένα κύτταρα τα οποία επεξεργάζονται μικρές ποσότητες πληροφορίας και έπειτα ενεργοποιούν άλλους νευρώνες για να συνεχίσουν αυτή τη διαδικασία. Η διαβίβαση της πληροφορίας πραγματοποιείται μέσω των δενδριτών και στα σημεία που εφάπτονται οι δενδρίτες δημιουργείται μία σύναψη. Υπολογίζεται ότι υπάρχουν περίπου 10 δισεκατομμύρια νευρώνες στον ανθρώπινο φλοιό και 60 τρισεκατομμύρια συνάψεις ή συνδέσεις [129]. Είναι προφανές ότι η δομή του εγκεφάλου είναι απίστευτα αποδοτική. Συγκεκριμένα, η απόδοση ενέργειας του εγκεφάλου κυμαίνεται στα 10^{-16} joules (J) ανά λειτουργία ανά δευτερόλεπτο, ενώ η αντίστοιχη τιμή για τους σημερινούς υπολογιστές κυμαίνεται στα 10^{-6} joules (J) ανά λειτουργία ανά δευτερόλεπτο.

Οι συνάψεις είναι στοιχειώδεις δομικές και λειτουργικές μονάδες που επιτρέπουν τις αλληλεπιδράσεις μεταξύ των νευρώνων. Ο πιο συνηθισμένος τύπος συνάψεων είναι οι χημικές συνάψεις. Η σύναψη έχει πολύ περίπλοκη δομή και επιτελεί επίσης περίπλοκες διεργασίες κατά τη μετάδοση του σήματος. Η μεμβράνη του πρώτου νευρώνα που στέλνει το σήμα ονομάζεται προ-συναπτική μεμβράνη, ενώ αυτή του δεύτερου νευρώνα (δέκτη) ονομάζεται μετα-συναπτική μεμβράνη. Η προ-συναπτική μεμβράνη ενός νευρώνα ελευθερώνει μία ουσία μέσω των συναπτικών συνδέσεων η οποία επηρεάζει τη μετα-συναπτική μεμβράνη ενός άλλου νευρώνα. Έτσι, μία σύναψη μετατρέπει ένα προ-συναπτικό ηλεκτρικό σήμα σε χημικό και ξανά σε ηλεκτρικό. Οι νευρώνες δεν έχουν σταθερό αριθμό συνάψεων, αλλά συνήθως ο καθένας τους έχει περίπου 10^4 συνάψεις [2].

Κάθε νευρώνας έχει δύο δυνατές καταστάσεις στις οποίες μπορεί να βρίσκεται και τις ονομάζουμε ενεργό και μη ενεργό κατάσταση. Όταν ο νευρώνας είναι ενεργός λέμε ότι πυροδοτεί, ενώ όταν είναι μη ενεργός λέμε ότι είναι αδρανής. Ενδιάμεσες καταστάσεις δεν υπάρχουν. Κατά κάποιο τρόπο φαίνεται ότι ο νευρώνας είναι δυαδικό στοιχείο και μοιάζει σε αυτό το σημείο με τους ηλεκτρονικούς υπολογιστές. Όταν ο νευρώνας πυροδοτεί παράγει ένα ηλεκτρικό σήμα (παλμό). Όλα τα σήματα που καταφθάνουν σε ένα νευρώνα σε μια δεδομένη στιγμή αθροίζονται, οπότε αθροίζονται και τα ηλεκτρικά δυναμικά τους. Αν το άθροισμα των σημάτων φτάσει ή ξεπεράσει μία δεδομένη τιμή (κατώφλι), τότε ο νευρώνας στέλνει έναν παλμό (ενεργοποιείται), δηλαδή λαμβάνει μία απόφαση, αλλιώς παραμένει αδρανής. Οι νευρώνες που ενεργοποιούνται για την εκτέλεση μιας διεργασίας συγκροτούν ένα «βιολογικό νευρωνικό δίκτυο» και μέσα από τις τροποποιήσεις στις συνάψεις που ενώνουν τους νευρώνες επιτυγχάνεται η διαδικασία μάθησης του τελευταίου.

Πάνω σε αυτήν ακριβώς την ανάλυση βασίζεται η κεντρική ιδέα των νευρωνικών δικτύων. Μιμούμενα, τον ανθρώπινο εγκέφαλο, του οποίου το δομικό στοιχείο είναι ο βιολογικός νευρώνας, τα τεχνητά νευρωνικά δίκτυα έχουν σαν δομικό στοιχείο τον τεχνητό νευρώνα. Επίσης, σε αντιστοιχία με τον ανθρώπινο εγκέφαλο, τα νευρωνικά δίκτυα αποτελούνται από ένα μεγάλο αριθμό αλληλοσυνδεόμενων τεχνικών νευρώνων. Τέλος, όσον αφορά στον τρόπο με τον οποίο μαθαίνει ένα βιολογικό νευρωνικό δίκτυο, η ίδια διαδικασία ακολουθείται και στα τεχνητά νευρωνικά δίκτυα. Η μάθηση, στα τελευταία, προκύπτει μέσα από έναν αλγόριθμο που ονομάζεται "αλγόριθμος εκπαίδευσης" του νευρωνικού δικτύου, και κατά τον οποίο τροποποιούνται με επαναληπτικό τρόπο οι συνάψεις των νευρώνων με οδηγό ένα σύνολο από δεδομένα, που προέρχονται από το περιβάλλον του δικτύου.



Σχήμα 1.2. Σχηματικό διάγραμμα τυπικού βιολογικού νευρώνα.

1.2.1 Ιστορική αναδρομή

Στη διεθνή βιβλιογραφία η περιοχή των νευρωνικών δικτύων ξεκίνησε το 1943 με τους Warren McCulloch και Walter Pitts [130] οι οποίοι μελέτησαν ένα μοντέλο του βασικού κυττάρου του ανθρώπινου εγκεφάλου το οποίο ονόμασαν νευρόνιο και σήμερα είναι γνωστό ως νευρόνιο McCulloch και Pitts. Αυτό το πρώτο πρότυπο νευρωνικού δικτύου που μοντελοποίησαν οι McCulloch και Pitts περιελάμβανε όλα τα απαραίτητα στοιχεία για να εκτελεί λογικές πράξεις, όπως μεταβλητές αντιστάσεις και αθροιστικούς ενισχυτές, αλλά η ηλεκτρονική εφαρμογή του ήταν ανέφικτη εκείνη την εποχή.

Το επόμενο μεγάλο βήμα στον χώρο των νευρωνικών δικτύων έγινε το 1949 από τον Donald Hebb [131], ο οποίος διατύπωσε για πρώτη φορά ρητά την έννοια της μάθησης με την οποία ενημερώνονται οι διασυνδέσεις μεταξύ των νευρώνων. Πιο συγκεκριμένα, ο Hebb διατύπωσε την ιδέα ότι οι πληροφορίες μπορούν να αποθηκευτούν στις συνδέσεις. Αυτή η διατύπωση ήταν καθοριστική για το μέλλον του τομέα των νευρωνικών δικτύων, και γι' αυτό ως σήμερα αναφερόμαστε στον κανόνα μάθησης Hebbian.

Καθώς οι υπολογιστές συνέχισαν να αναπτύσσονται κατά τη δεκαετία του 1950, προσομοιώθηκε το πρώτο νευρωνικό δίκτυο [132] και το 1958 ο Frank Rosenblatt [133] εφηύρε το πρώτο μοντέλο perceptron. Ήταν ένας μηχανισμός εκπαίδευσης ο οποίος έθεσε τις βάσεις στους αλγόριθμους εκμάθησης και χρησιμοποιείται ακόμα και σήμερα.

Στις αρχές της δεκαετίας του 1960, οι Bernard Widrow και Marcian Hoff [134] ανέπτυξαν ένα μοντέλο αποκαλούμενο ADALINE (adaptive linear combinator) που χρησιμοποιούσε έναν ισχυρό κανόνα εκμάθησης. Ο κανόνας αυτός που ονομάστηκε κανόνας Widrow-Hoff ελαχιστοποιούσε το άθροισμα του τετραγωνικού σφάλματος κατά τη διάρκεια εκπαίδευσης συμπεριλαμβάνοντας ταξινόμηση προτύπων. Οι πρόωρες εφαρμογές ADALINE και οι επεκτάσεις της σε MADALINE (many ADALINES) περιλαμβάνουν αναγνώριση προτύπων, πρόβλεψη καιρικών φαινομένων και προσαρμοστικό έλεγχο. Ο Nilson στο βιβλίο του διατύπωσε τις μέχρι τότε εξελίξεις αλλά και τα προβλήματα πάνω σε αυτό το τομέα [135].

Τα νευρωνικά δίκτυα γνώρισαν μεγάλη άνθηση και προκάλεσαν υπερβολικό ενθουσιασμό για τις δυνατότητές τους. Δυστυχώς όμως, εκείνη την εποχή δεν υπήρχαν κατάλληλοι διαθέσιμοι υπολογιστικοί πόροι και τα μηχανήματα ήταν πολύ αδύναμα για να υποστηρίξουν σύνθετα υπολογιστικά προβλήματα. Αυτό είχε ως επακόλουθο να επικρατήσει απογοήτευση και η έρευνα στον τομέα των νευρωνικών δικτύων να γνωρίσει μία φάση στασιμότητας.

Κατά τη διάρκεια της περιόδου 1965-1984 αρκετοί ερευνητές ασχολήθηκαν με τα νευρωνικά δίκτυα. Οι Marvin Minsky και Seymour Papert δημοσίευσαν ένα βιβλίο στο οποίο εξέφραζαν τις αμφιβολίες τους για τη δυνατότητα εκμάθησης σε στοιβάδες (layers) των νευρωνικών δικτύων [136]. Στην Ιαπωνία, ο Kunihito Fukushima ανέπτυξε μία κατηγορία αρχιτεκτονικής νευρωνικών δικτύων γνωστή ως neocognitron [137]. Το neocognitron είναι ένα πρότυπο οπτικής αναγνώρισης το οποίο μιμείται αμφιβληστροειδείς εικόνες και τις επεξεργάζεται χρησιμοποιώντας δύο στοιβάδες νευρώνων. Η έρευνα μνήμης συνεχίστηκε, μεταξύ άλλων, από τον Tuevo Kohonen στη Φιλανδία [138-141] και τους Anderson et. al. [142]. Ο Stephen Grossberg και ο Gail Carpenter παρουσίασαν διάφορες αρχιτεκτονικές και θεωρίες και ανέπτυξαν τη θεωρία των προσαρμοστικών δικτύων συντονισμού (adaptive resonance networks) [143, 144].

Κατά τη διάρκεια της περιόδου 1982-1986 υπήρξαν διάφορες ενδιαφέρουσες δημοσιεύσεις που προήγαγαν σημαντικά τις δυνατότητες των νευρωνικών δικτύων. Η εποχή της αναγέννησης ξεκίνησε με τον John Hopfield ο οποίος παρουσίασε ένα παλινδρομικό

νευρωνικό δίκτυο (recurrent neural network, RNN) [145, 146]. Επίσης, νέοι δρόμοι ανοίχτηκαν από τους James McClelland και David Rumelhart με το βιβλίο τους «Parallel Distributed Processing» [147]. Οι νέοι κανόνες εκμάθησης και άλλες έννοιες που εισήχθησαν σε αυτή την εργασία, εξάλειψαν αρκετά από τα εμπόδια στην εκπαίδευση νευρωνικών δικτύων. Επίσης, το τελευταίο βιβλίο έχει επηρεάσει σημαντικά τη χρήση του αλγορίθμου οπισθοδιάδοσης, ο οποίος είναι ο πιο γνωστός αλγόριθμος εκπαίδευσης για τα εμπροσθοτροφοδοτούμενα συστήματα πολλών-στοιβάδων (multilayer feedforward systems) και η ανάπτυξή του αναφέρθηκε από τους Rumelhart, Hinton και Williams το 1986 [148, 149].

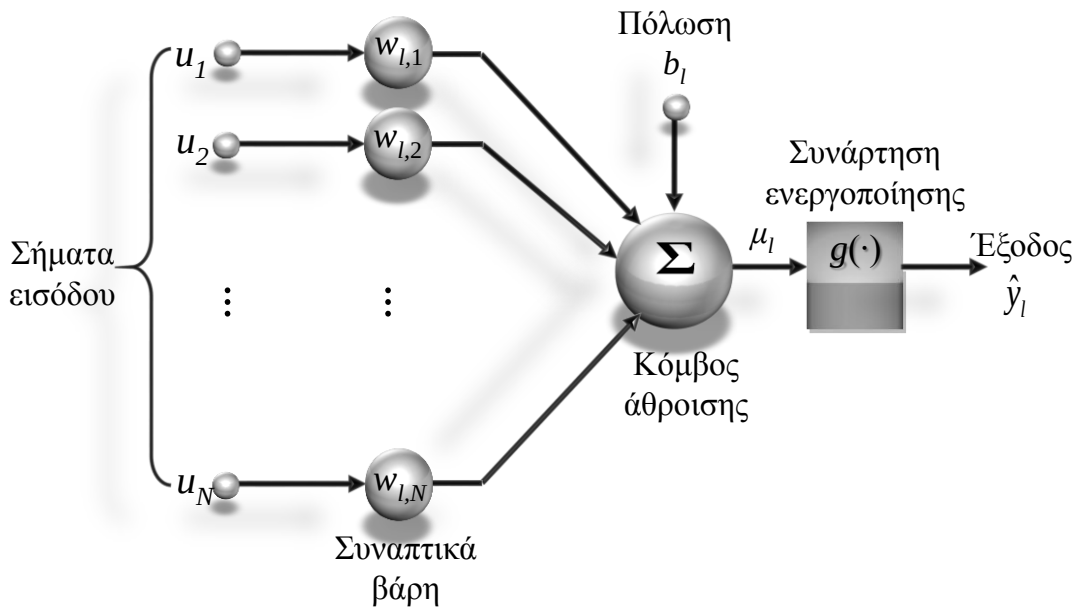
Στην πραγματικότητα, ο αλγόριθμος οπισθοδιάδοσης ανακαλύφθηκε ανεξάρτητα σε δύο άλλα μέρη, περίπου την ίδια εποχή (Parker, 1985, LeCun, 1985). Μετά την ανακάλυψη του αλγορίθμου οπισθοδιάδοσης στα μέσα της δεκαετίας του '80, αποδείχθηκε ότι είχε περιγραφεί νωρίτερα, το 1974, από τον Werbos στη διδακτορική διατριβή του [150], η οποία αποτελούσε την πρώτη τεκμηριωμένη περιγραφή του αλγορίθμου οπισθοδιάδοσης που εφαρμόστηκε σε μοντέλα νευρωνικών δικτύων. Η βασική ιδέα του αλγορίθμου οπισθοδιάδοσης μπορεί να εντοπιστεί ακόμα πιο παλιά. Σύμφωνα με τον Dreyfus [151], οι πρωτοπόροι της μεθόδου βελτιστοποίησης για τα εμπροσθοτροφοδοτούμενα συστήματα πολλών-στοιβάδων ήταν οι Bryson, Ho [152, 153] και Kelley [154], οι οποίοι βρήκαν μία λύση που βασίζεται στην κλίση (gradient) συναρτήσεων για την εκπαίδευση δικτύων πολλών-στοιβάδων.

Το 1986-87 έκαναν την εμφάνισή τους πολλά νέα ερευνητικά προγράμματα με θέμα τα νευρωνικά δίκτυα. Το μεγάλο ενδιαφέρον στη νευρο-υπολογιστική (neurocomputing) μπορεί να επιβεβαιωθεί εύκολα από τον σημαντικά αυξανόμενο αριθμό βιβλίων, συνεδρίων και δημοσιεύσεων σε περιοδικά που αναφέρονταν στον συγκεκριμένο τομέα. Τα προβλήματα που μπορούν να λυθούν με χρήση νευρωνικών δικτύων είναι πάρα πολλά. Αν και η νευρο-υπολογιστική έχει ήδη ένα πολύ ενδιαφέρον ιστορικό, ο τομέας αυτός βρίσκεται ακόμα στα πρώτα στάδια ανάπτυξής του.

1.2.2 Η δομή του ενός νευρώνα

Σε αναλογία με το βιολογικό νευρώνα του εγκεφάλου, ο τεχνητός νευρώνας (artificial neuron) είναι η θεμελιώδης μονάδα για τη λειτουργία ενός τεχνητού νευρωνικού δικτύου. Σε αυτόν συντελείται όλη η επεξεργασία της πληροφορίας. Κάθε νευρώνας δέχεται πληροφορία και οι είσοδοί του είναι είτε οι έξοδοι άλλων νευρώνων, είτε το πρωταρχικό σήμα εισόδου του δικτύου. Στο σχήμα 1.3 παρουσιάζεται το βασικό μοντέλο ενός νευρώνα. Στο νευρώνα αυτό, η πληροφορία ρέει πάντα προς μία κατεύθυνση, από αριστερά προς τα δεξιά και μπορούμε να διακρίνουμε τρία βασικά στοιχεία του νευρωνικού μοντέλου:

1. Ένα σύνολο συνάψεων ή συνδέσεων, κάθε μία από τις οποίες χαρακτηρίζεται από ένα βάρος (weight) ή μία δύναμη (strength). Συγκεκριμένα, ένα σήμα u_n στην είσοδο της σύναψης n που συνδέεται με τον νευρώνα l πολλαπλασιάζεται με το συναπτικό βάρος $w_{l,n}$. Είναι σημαντικό να σημειωθεί ο τρόπος με τον οποίο συμβολίζονται τα συναπτικά βάρη $w_{l,n}$. Το πρώτο σύμβολο αναφέρεται στον εν λόγω νευρώνα και το δεύτερο σύμβολο αναφέρεται στην τελευταία είσοδο της σύναψης στην οποία αντιστοιχεί το βάρος. Αντίθετα από μία σύναψη του εγκεφάλου, το συναπτικό βάρος ενός τεχνητού νευρώνα περιλαμβάνει τόσο αρνητικές, όσο και θετικές τιμές.



Σχήμα 1.3. Σχηματικό διάγραμμα μοντέλου μη γραμμικού νευρώνα.

2. Έναν αθροιστή που αθροίζει τα σήματα των εισόδων που σταθμίζονται από τις αντίστοιχες συνάψεις του νευρώνα. Στη διαδικασία που περιγράφεται εδώ, ο αθροιστής αντιστοιχεί σε ένα γραμμικό συνδυαστή (linear combinator).
3. Μια συνάρτηση ενεργοποίησης (activation function) για να περιορίζει το εύρος της εξόδου ενός νευρώνα. Η συνάρτηση ενεργοποίησης αναφέρεται επίσης και ως συνάρτηση συμπίεσης (activation squashing) δεδομένου ότι περιορίζει το επιτρεπτό εύρος του σήματος εξόδου σε κάποια πεπερασμένη τιμή. Τυπικά το κανονικοποιημένο εύρος της εξόδου ενός νευρώνα βρίσκεται στο κλειστό μοναδιαίο διάστημα [0,1] ή εναλλακτικά στο [- 1,1].

Στο μοντέλο του νευρώνα στο σχήμα 1.3 διακρίνεται η ύπαρξη ενός σήματος πόλωσης (bias), b_l. Το σήμα πόλωσης b_l επιδρά προσθετικά ή αφαιρετικά στην είσοδο του δικτυώματος της συνάρτησης ενεργοποίησης, ανάλογα με το εάν είναι θετικό ή αρνητικό, αντίστοιχα.

Με μαθηματικούς όρους, μπορούμε να περιγράψουμε έναν νευρώνα l με τις ακόλουθες εξισώσεις:

$$\hat{\mu}_l = \sum_{n=1}^{N} w_{l,n} u_n \tag{1.1}$$

και

$$\hat{y}_l = g\left(\hat{\mu}_l + b_l\right) \tag{1.2}$$

όπου $u_1, u_2, \ldots, u_N$ είναι τα σήματα εισόδου, $w_{l,1}, w_{l,2}, \ldots, w_{l,N}$ τα συναπτικά βάρη του νευρώνα l, $\hat{\mu}_l$ η έξοδος του αθροιστή που αφορά στα σήματα εισόδου, b_l η πόλωση (bias), $g(\cdot)$ η

συνάρτηση ενεργοποίησης και $\hat{y}_l$ η έξοδος του νευρώνα. Η χρήση της πόλωσης b_l έχει ως αποτέλεσμα την εφαρμογή ενός αφινικού μετασχηματισμού (affine transformation) στην έξοδο $\hat{\mu}_l$ του αθροιστή στο μοντέλο του σχήματος 1.3, που περιγράφεται μαθηματικά ως εξής:

$$\mu_l = \hat{\mu}_l + b_l \tag{1.3}$$

Η συνάρτηση ενεργοποίησης $g(\mu_l)$ καθορίζει την έξοδο ενός νευρώνα l σε σχέση με το μ. Στη συνέχεια παρατίθενται τρεις βασικοί τύποι συναρτήσεων ενεργοποίησης:

1. *Συνάρτηση κατωφλίου (Threshold function ή Hard-limit function)*: Ο νευρώνας που υπακούει σε αυτή τη συνάρτηση ονομάζεται McCulloch-Pitts. Η συνάρτηση περιγράφεται γραφικά στο σχήμα 1.4 και μαθηματικά με την παρακάτω έκφραση:

$$g(\mu_l) = \begin{cases} 1 & \text{αν} \quad \mu_l \geq 0 \\ 0 & \text{αν} \quad \mu_l < 0 \end{cases} \tag{1.4}$$

2. *Τμηματικά γραμμική συνάρτηση (Piecewise-linear function)*: Περιγράφεται γραφικά στο σχήμα 1.5 και μαθηματικά με την παρακάτω έκφραση:

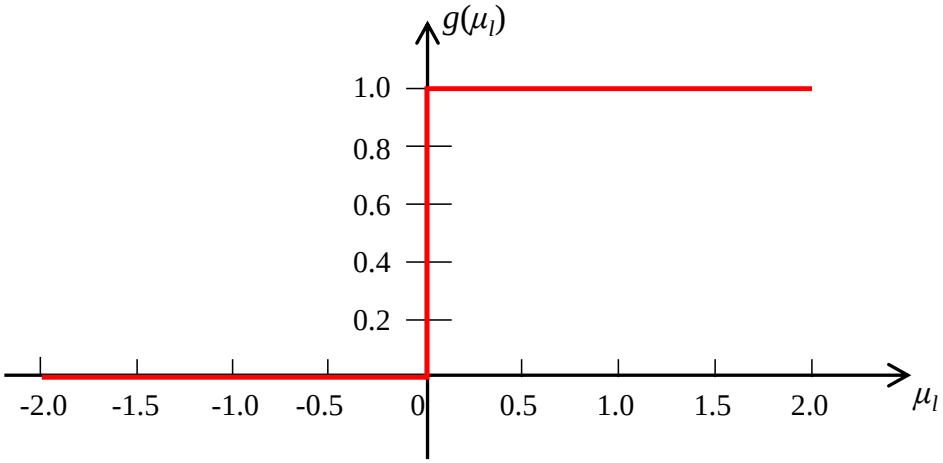
$$g(\mu_l) = \begin{cases} 1 & \text{αν} \quad \mu_l \geq +\frac{1}{2} \\ \mu_l & \text{αν} \quad +\frac{1}{2} > \mu_l > -\frac{1}{2} \\ 0 & \text{αν} \quad \mu_l \leq -\frac{1}{2} \end{cases} \tag{1.5}$$

3. *Σιγμοειδής συνάρτηση (Sigmoidal function)*: Είναι η πλέον χρησιμοποιούμενη στην κατασκευή των τεχνητών νευρωνικών δικτύων. Η γραφική της παράσταση έχει σχήμα “S”. Πρόκειται για μία αυστηρά αύξουσα συνάρτηση που διατηρεί μια ισορροπία ανάμεσα στη γραμμική και στη μη γραμμική συμπεριφορά. Η πιο συνήθης μορφή είναι η λογιστική συνάρτηση (logistic function), η οποία φαίνεται στο σχήμα 1.6 για κάποια τιμή της παραμέτρου a, όπου a είναι η παράμετρος κλίσης της σιγμοειδούς συνάρτησης.

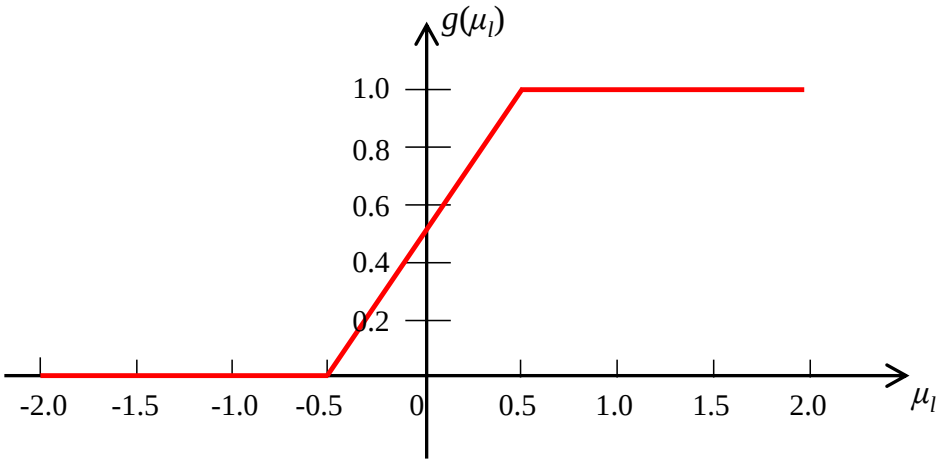
$$g(\mu_l) = \frac{1}{1 + \exp(-a \cdot \mu_l)} \tag{1.6}$$

Παρακάτω, παρουσιάζονται συνοπτικά οι πιο γνωστές αρχιτεκτονικές νευρωνικών δικτύων και συγκεκριμένα τα νευρωνικά δίκτυα πολλών-στοιβάδων, τα παλινδρομικά νευρωνικά

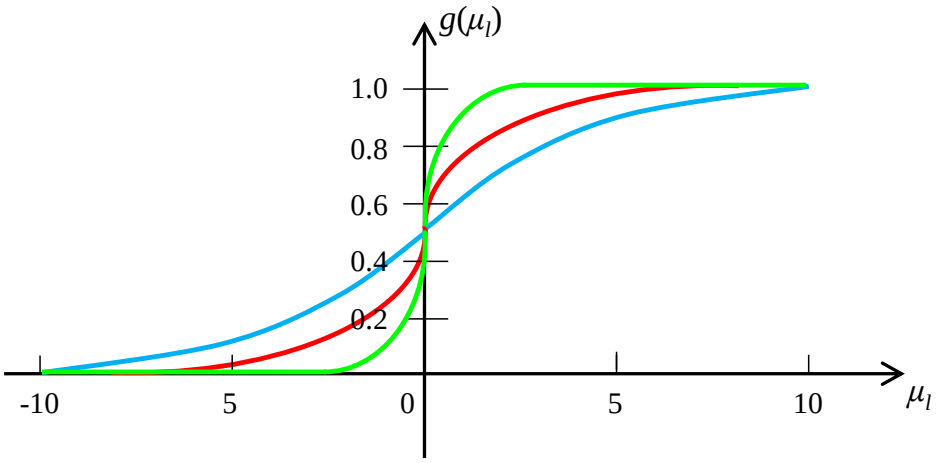
δίκτυα και τα δίκτυα Kohonen. Στο κεφάλαιο 2 αναλύονται διεξοδικά τα νευρωνικά δίκτυα ακτινικής συνάρτησης βάσης (RBF), τα οποία αποτελούν την αρχιτεκτονική που χρησιμοποιήθηκε σε αυτή τη διδακτορική διατριβή.



Σχήμα 1.4. Συνάρτηση κατωφλίου.



Σχήμα 1.5. Τμηματικά γραμμική συνάρτηση.



Σχήμα 1.6. Σιγμοειδής συνάρτηση (μπλε: α=0.5, κόκκινο: α=1, πράσινο: α=2).

1.2.3 Νευρωνικά δίκτυα πολλών-στοιβάδων

Τα νευρωνικά δίκτυα πολλών-στοιβάδων (multi-layer perceptrons, MLPs) είναι εμπροσθοτροφοδοτούμενα νευρωνικά δίκτυα (feed-forward neural networks), και αντιπροσωπεύουν μια γενίκευση των μονοστρωματικών νευρωνικών δικτύων (single-layer perceptrons). Χαρακτηριστικά, ένα τέτοιο δίκτυο αποτελείται από:

- Ένα σύνολο αισθητήριων μονάδων (κόμβων εισόδου) που απαρτίζουν τη στοιβάδα εισόδου.
- Μία ή περισσότερες κρυφές στοιβάδες υπολογιστικών κόμβων.
- Μία στοιβάδα εξόδου υπολογιστικών κόμβων.

Όλες οι στοιβάδες αποτελούνται από έναν αριθμό νευρώνων. Κάθε νευρώνας που ανήκει στη στοιβάδα εξόδου ή στην/στις κρυφή/κρυφές στοιβάδα/στοιβάδες ενός δικτύου MLP έχει σχεδιαστεί για να εκτελεί δύο υπολογισμούς :

- Υπολογίζει το σήμα που εμφανίζεται στην έξοδο του νευρώνα και εκφράζεται σαν μια συνεχής μη γραμμική συνάρτηση των σημάτων που εισέρχονται στον νευρώνα και των αντίστοιχων βαρών των συνάψεων που σχετίζονται με τον νευρώνα.
- Υπολογίζει μια στιγμιαία προσέγγιση του διανύσματος κλίσης (δηλαδή τις κλίσεις της επιφάνειας σφάλματος ως προς τα βάρη που σχετίζονται με τις συνάψεις), κατά την προς τα πίσω διάδοση του σήματος στο δίκτυο.

Το σήμα εισόδου διαδίδεται μέσω του δικτύου με μία προς τα εμπρός κατεύθυνση από στοιβάδα σε στοιβάδα.

Τα MLPs, τα οποία έχουν εφαρμοστεί επιτυχώς στη λύση πολλών και διαφορετικών πολύπλοκων προβλημάτων, εκπαιδεύονται με επιβλεπόμενο τρόπο, χρησιμοποιώντας έναν ιδιαίτερα δημοφιλή αλγόριθμο γνωστό ως αλγόριθμο σφάλματος οπισθοδιάδοσης (error back-propagation). Αυτός ο αλγόριθμος είναι βασισμένος στον κανόνα εκμάθησης διόρθωσης σφάλματος (error-correction learning rule). Υπό αυτήν την έννοια, μπορεί να αντιμετωπισθεί ως γενίκευση ενός εξίσου δημοφιλούς προσαρμοστικού αλγόριθμου φίλτρου: Του αλγόριθμου ελαχίστων μέσων τετραγώνων (least mean square, LMS) για την ειδική περίπτωση ενός μοναδικού γραμμικού νευρώνα (single linear neuron).

Ο αλγόριθμος error back-propagation αποτελείται από δύο περάσματα μέσω των διαφορετικών στοιβάδων του δικτύου: Ένα προς τα εμπρός πέρασμα και ένα προς τα πίσω πέρασμα. Στο προς τα εμπρός πέρασμα, ένα διάνυσμα εισόδου εφαρμόζεται στους αισθητήριους κόμβους του δικτύου και η επίδρασή του διαδίδεται από στοιβάδα σε στοιβάδα. Τέλος, ένα σύνολο τιμών παράγεται ως έξοδος του δικτύου. Κατά τη διάρκεια του προς τα εμπρός περάσματος τα συναπτικά βάρη των δικτύων είναι όλα αμετάβλητα.

Αντίθετα, κατά τη διάρκεια του προς τα πίσω περάσματος όλα τα συναπτικά βάρη προσαρμόζονται σύμφωνα με τον κανόνα διόρθωσης σφάλματος. Συγκεκριμένα, η πραγματική τιμή της εξόδου του δικτύου αφαιρείται από την επιθυμητή τιμή της εξόδου για να παραγάγει ένα σήμα σφάλματος. Αυτό το σήμα σφάλματος διαδίδεται έπειτα προς τα πίσω μέσω του δικτύου, αντίθετα στην κατεύθυνση των συναπτικών συνδέσεων, εξ' ου και η ονομασία «error back-propagation». Τα συναπτικά βάρη προσαρμόζονται έτσι ώστε η πραγματική τιμή της εξόδου του δικτύου να γίνει όσο το δυνατόν πιο όμοια με την επιθυμητή τιμή της εξόδου, υπό μια στατιστική έννοια. Η διαδικασία εκμάθησης που εκτελείται με τον αλγόριθμο back-propagation καλείται εκμάθηση back-propagation.

Ο αλγόριθμος back-propagation έγινε ιδιαίτερα γνωστός μετά το 1985, όταν διαδόθηκε η χρήση του μέσω της δημοσίευσης του βιβλίου με τίτλο Parallel Distributed Processing [147]. Η ανάπτυξη του αλγόριθμου back-propagation είναι ορόσημο στα νευρωνικά δίκτυα, δεδομένου ότι παρέχει μία υπολογιστικά αποδοτική μέθοδο για την εκπαίδευση των δικτύων MLP.

Τα νευρωνικά δίκτυα MLP έχουν τρία σημαντικά χαρακτηριστικά:

- Το μοντέλο κάθε νευρώνα στο δίκτυο περιλαμβάνει μία μη γραμμική συνάρτηση ενεργοποίησης. Μια σημαντική διαφορά από το perceptron του Rosenblatt είναι η ομαλότητα αυτής της συνάρτησης σε σχέση με τη συνάρτηση κατωφλίου και το γεγονός ότι είναι παραγωγίσιμη σε κάθε σημείο. Συναρτήσεις που ικανοποιούν αυτές τις απαιτήσεις και χρησιμοποιούνται συνήθως είναι η λογιστική συνάρτηση (1.7) ή η υπερβολική εφαπτομένη (εξίσωση (1.8)):

$$\hat{y}_l = \frac{1}{1+\exp\left(-\mu_l\right)} \tag{1.7}$$

$$\hat{y}_l = \frac{1-\exp\left(-\mu_l\right)}{1+\exp\left(-\mu_l\right)} \tag{1.8}$$

  όπου μ_l είναι η σταθμισμένη τιμή όλων των συναπτικών εισόδων του νευρώνα l και $\hat{y}$ είναι η έξοδος του νευρώνα.
- Το δίκτυο περιέχει μία ή περισσότερες κρυφές στοιβάδες νευρώνων που δεν αποτελούν μέρος της εισόδου ή της εξόδου του δικτύου. Αυτοί οι νευρώνες καθιστούν το δίκτυο ικανό να μαθαίνει σύνθετες διαδικασίες με σταδιακή εξαγωγή των σημαντικότερων χαρακτηριστικών γνωρισμάτων από τα πρότυπα/διανύσματα εισόδου.
- Το δίκτυο παρουσιάζει υψηλό βαθμό συνδεσιμότητας που καθορίζεται από τις συνάψεις μεταξύ των νευρώνων του δικτύου. Μια αλλαγή στη συνδεσιμότητα του δικτύου απαιτεί μια αλλαγή στο πλήθος ή τα βάρη των συναπτικών συνδέσεων.

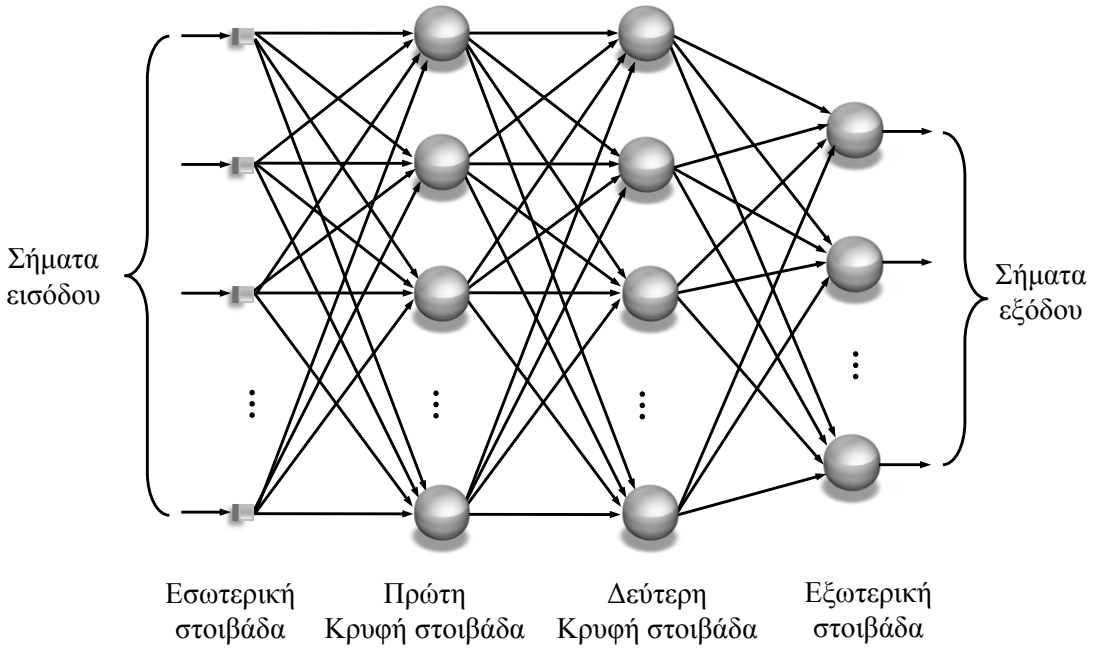
Η υπολογιστική ισχύς (computing power) των MLPs προέρχεται από το συνδυασμό των παραπάνω χαρακτηριστικών και από τη διαδικασία εκπαίδευσης. Τα ίδια αυτά χαρακτηριστικά είναι επίσης υπεύθυνα και για τα μειονεκτήματα σχετικά με τη συμπεριφορά αυτών των δικτύων. Κατ' αρχάς, η μη γραμμική κατανεμημένη μορφή και η υψηλή περιπλοκότητα συνδέσεων του δικτύου κάνουν δύσκολα κατανοητή τη θεωρητική ανάλυση των MLPs. Δεύτερον, η χρήση των κρυφών στοιβάδων των νευρώνων δυσκολεύει πάρα πολύ την οπτικοποίηση της διαδικασίας εκμάθησης. Είναι αυτονόητο ότι τα χαρακτηριστικά γνωρίσματα της εισόδου που αντιπροσωπεύονται από τους κρυφούς νευρώνες καθορίζονται από τη διαδικασία εκμάθησης. Η διαδικασία εκμάθησης γίνεται όλο και περισσότερο δύσκολη, εφόσον η αναζήτηση διεξάγεται ανάμεσα σε ένα πολύ μεγάλο αριθμό πιθανών συναρτήσεων και η επιλογή γίνεται μεταξύ εναλλακτικών αναπαραστάσεων της εισόδου [155].

Το σχήμα 1.7 παρουσιάζει την αρχιτεκτονική ενός MLP που αποτελείται από δύο κρυφές στοιβάδες και μία στοιβάδα εξόδου. Για να παρουσιαστεί η γενική μορφή των MLPs, το δίκτυο που φαίνεται στο σχήμα 1.7 είναι πλήρως συνδεδεμένο. Αυτό σημαίνει ότι κάθε νευρώνας σε οποιαδήποτε στοιβάδα του δικτύου, συνδέεται με όλους τους νευρώνες της

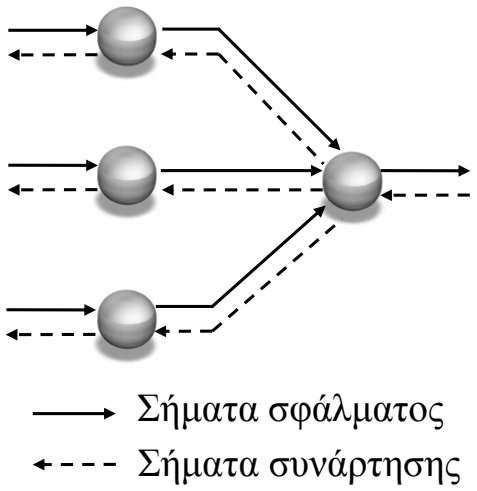
προηγούμενης στοιβάδας. Η ροή της πληροφορίας στο δίκτυο γίνεται με μία προς τα εμπρός κατεύθυνση, από το αριστερό στο δεξιό μέρος και από στοιβάδα σε στοιβάδα.

Το σχήμα 1.8 απεικονίζει ένα μέρος του MLP. Δύο είδη σημάτων προσδιορίζονται σε αυτό το δίκτυο [156]:

1. *Σήματα Λειτουργίας (Function Signals).* Ένα σήμα λειτουργίας είναι ένα σήμα εισόδου (ερέθισμα) που μπαίνει στην είσοδο του δικτύου, διαδίδεται προς τα εμπρός (από νευρώνα σε νευρώνα) μέσω του δικτύου και προκύπτει στην έξοδο του δικτύου ως σήμα εξόδου. Το σήμα λειτουργίας αναφέρεται επίσης και ως σήμα εισόδου (input signal).
2. *Σήματα σφάλματος (Error Signals).* Ένα σήμα σφάλματος δημιουργείται σε έναν



Σχήμα 1.7. Σχηματικό διάγραμμα MLP με δύο κρυφές στοιβάδες.



Σχήμα 1.8. Απεικόνιση των κατευθύνσεων των δύο βασικών σημάτων σε MLP: Διάδοση των σημάτων συνάρτησης προς τα εμπρός και των σημάτων σφάλματος προς τα πίσω.

νευρώνα εξόδου του δικτύου και διαδίδεται προς τα πίσω (από στοιβάδα σε στοιβάδα) μέσω του δικτύου, κατά τη διάρκεια της εκπαίδευσης μόνο.

Οι νευρώνες εξόδων συγκροτούν τις εξωτερικές στοιβάδες του δικτύου. Οι υπόλοιποι νευρώνες συγκροτούν τις κρυφές στοιβάδες του δικτύου. Οι κρυφές στοιβάδες δεν είναι μέρος της εξόδου ή της εισόδου του δικτύου και γι' αυτό καλούνται κρυφές. Η πρώτη κρυφή στοιβάδα τροφοδοτείται από τη στοιβάδα της εισόδου που αποτελείται από αισθητήριες μονάδες. Στη συνέχεια, οι έξοδοι της πρώτης κρυφής στοιβάδας εφαρμόζονται σαν είσοδο στην επόμενη κρυφή στοιβάδα και ούτω καθεξής.

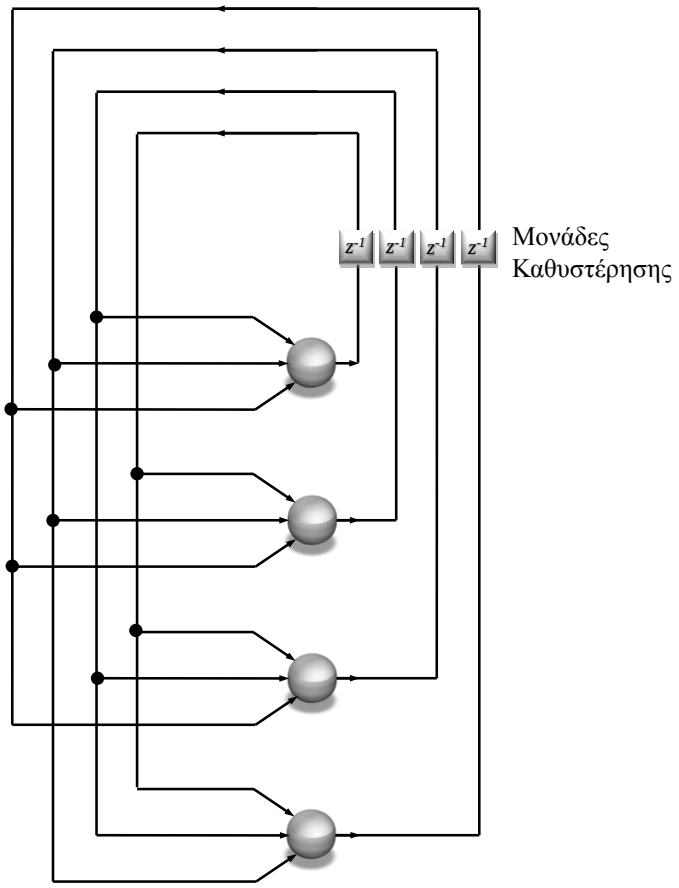
1.2.4 Παλινδρομικά νευρωνικά δίκτυα

Τα παλινδρομικά νευρωνικά δίκτυα (recurrent neural networks, RNNs) δεν ανήκουν στην κατηγορία των δικτύων MLP αν και μοιάζουν πολύ με αυτά. Συγκεκριμένα, αυτό που κάνει τα RNNs να διαφέρουν από τα νευρωνικά δίκτυα τύπου MLP είναι ότι περιέχουν τουλάχιστον έναν βρόχο ανατροφοδότησης. Επίσης, αντίθετα από τα MLPs, τα RNNs μπορούν να χρησιμοποιήσουν την εσωτερική μνήμη τους για να επεξεργαστούν αυθαίρετες ακολουθίες των εισόδων. Για τον λόγο αυτό χρησιμοποιούνται με μεγάλη επιτυχία σε εφαρμογές όπως η αναγνώριση γραφής.

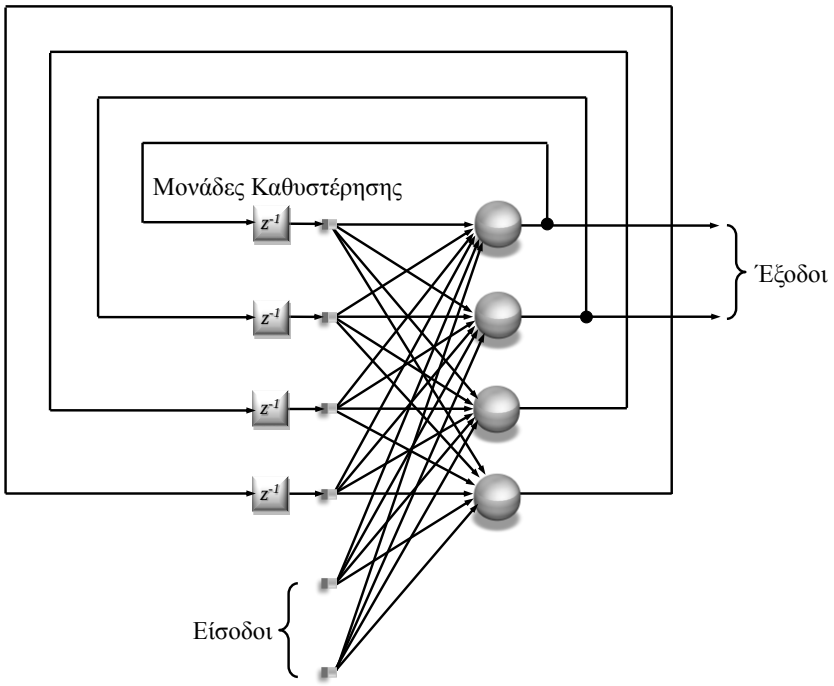
Ένα RNN μπορεί να συγκροτείται από μία μόνο στοιβάδα νευρώνων όπου το σήμα εξόδου του κάθε νευρώνα ανατροφοδοτείται στις εισόδους όλων των νευρώνων της στοιβάδας, όπως απεικονίζεται στη γραφική παράσταση του σχήματος 1.9. Πρέπει να σημειωθεί ότι στη δομή που απεικονίζεται σε αυτό το σχήμα δεν υπάρχουν αυτό-ανατροφοδοτούμενοι βρόχοι στο δίκτυο. Με τον όρο αυτό-ανατροφοδοτούμενος βρόχος νοείται μία κατάσταση όπου η έξοδος ενός νευρώνα ανατροφοδοτεί την είσοδο του ίδιου νευρώνα. Επίσης, το δίκτυο RNN του σχήματος 1.9 δεν έχει κανέναν κρυμμένο νευρώνα.

Μια άλλη κατηγορία δικτύων RNN παρουσιάζεται στο σχήμα 1.10. Αυτή η μορφή περιέχει κρυφές στοιβάδες. Οι ανατροφοδοτήσεις γίνονται από την έξοδο των νευρώνων της κρυφής στοιβάδας προς τη στοιβάδα εισόδου. Οι βρόχοι ανατροφοδότησης, είτε στη δομή του σχήματος 1.9 είτε στη δομή του σχήματος 1.10, επιδρούν καταλυτικά στην ικανότητα εκμάθησης του δικτύου και στην απόδοσή του. Επιπλέον, οι βρόχοι ανατροφοδότησης περιλαμβάνουν ιδιαίτερους κλάδους που αποτελούνται από τους καθυστερητές και συμβολίζονται με z^{-1}. Οι καθυστερητές οδηγούν το νευρωνικό δίκτυο σε μία μη γραμμική δυναμική συμπεριφορά όταν αυτό περιέχει μη γραμμικές μονάδες.

Τα δίκτυα RNN χρησιμοποιούνται κυρίως στη μοντελοποίηση δυναμικών συστημάτων. Ο συνήθης αλγόριθμος που χρησιμοποιείται για την εκπαίδευσή τους είναι μία τροποποιημένη μορφή του αλγόριθμου back-propagation ο οποίος χρησιμοποιείται στην εκπαίδευση δικτύων τύπου MLP [157].



Σχήμα 1.9. Παλινδρομικό νευρωνικό δίκτυο χωρίς αυτό-ανατροφοδοτούμενους βρόχους και κρυφούς νευρώνες.



Σχήμα 1.10. Παλινδρομικό νευρωνικό δίκτυο με κρυφούς νευρώνες.

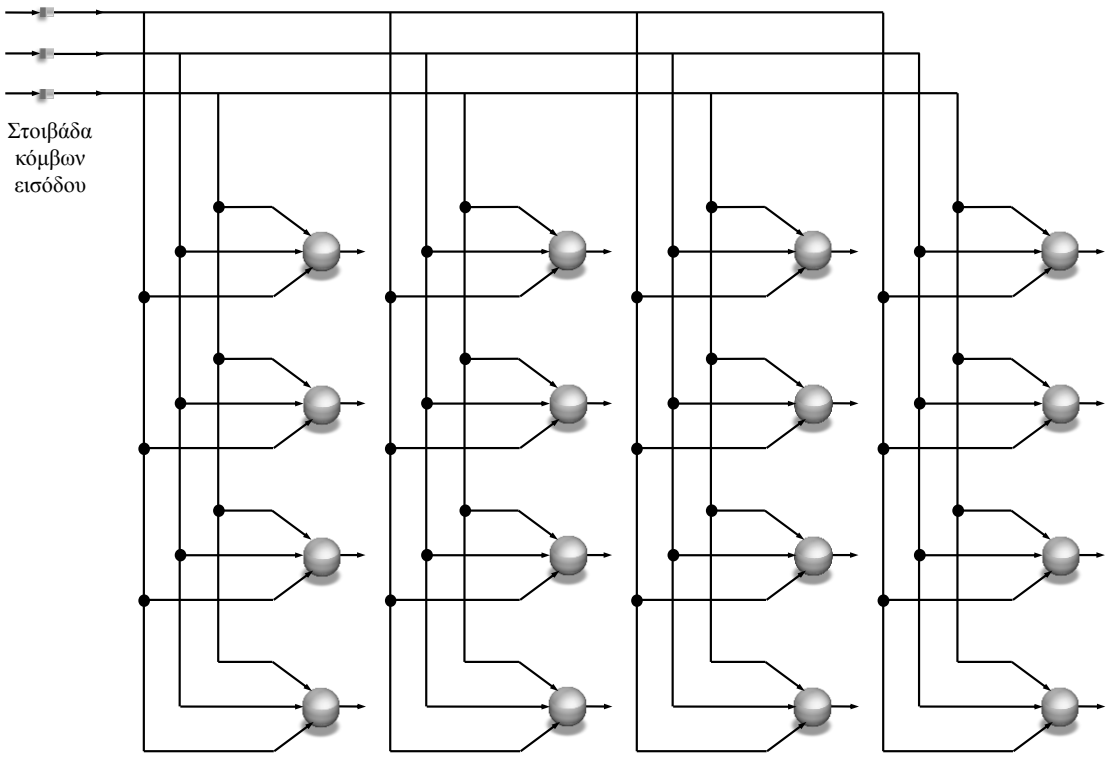
1.2.5 Νευρωνικά δίκτυα Kohonen

Μία ειδική κατηγορία αυτό-οργανούμενων συστημάτων, που είναι γνωστά και ως νευρωνικά δίκτυα τύπου Kohonen ή αλλιώς Self-Organizing Maps ή εν συντομία SOMs, μελετάται λόγω των ιδιαίτερων χαρακτηριστικών της. Συγκεκριμένα, αυτά τα δίκτυα είναι βασισμένα στην ανταγωνιστική μάθηση (competitive learning) [2].

Οι νευρώνες εξόδου του δικτύου ανταγωνίζονται μεταξύ τους για το ποιος θα ενεργοποιηθεί, γιατί μόνο ένας νευρώνας εξόδου, ή νευρώνας ανά συστάδα, ενεργοποιείται κάθε φορά. Ο νευρώνας που ενεργοποιείται αποκαλείται winner-takes-all neuron ή απλά winning neuron. Ένας τρόπος διεξαγωγής διαγωνισμού τύπου winner-takes-all ανάμεσα στους νευρώνες είναι η χρήση πλευρικών ανασταλτικών συνδέσεων, για παράδειγμα αρνητικών ανατροφοδοτήσεων [133].

Σε ένα self-organizing map, οι νευρώνες του δικτύου τοποθετούνται πάνω σ' ένα πλέγμα, συνήθως δισδιάστατο όπως φαίνεται στο σχήμα 1.11. Κατά τη διάρκεια της ανταγωνιστικής μάθησης οι νευρώνες συντονίζονται επιλεκτικά σε σχέση με τις διάφορες εισόδους. Ένα SOM χαρακτηρίζεται από το σχηματισμό του τοπολογικού χάρτη των προτύπων εισόδου, όπου οι χωρικές θέσεις, δηλαδή οι συντεταγμένες των νευρώνων στο δικτυωτό πλέγμα, είναι ενδεικτικές των εγγενών στατιστικών χαρακτηριστικών γνωρισμάτων που περιλαμβάνονται στα πρότυπα εισόδου, εξ' ου και το όνομα "self-organizing map".

Επειδή ένα SOM είναι εγγενώς μη γραμμικό, μπορεί να αντιμετωπισθεί ως μη γραμμική γενίκευση της ανάλυσης κύριων συνιστωσών (principal component analysis, PCA) [158]. Η ανάπτυξη των SOMs ως νευρωνικά μοντέλα προέρχεται από ένα ευδιάκριτο χαρακτηριστικό γνώρισμα του ανθρώπινου εγκεφάλου: Ο εγκέφαλος οργανώνεται κατά τέτοιο τρόπο ώστε τα διάφορα ερεθίσματα να αντιπροσωπεύονται από τοπολογικά διαταγμένους υπολογιστικούς χάρτες.



Σχήμα 1.11. Δισδιάστατο πλέγμα νευρώνων σε δίκτυο Kohonen.

Στην πραγματικότητα, τα μοντέλα Kohonen ανήκουν στην κατηγορία των αλγορίθμων vector-coding όπου τοποθετούν βέλτιστα έναν σταθερό αριθμό διανυσμάτων, για παράδειγμα λέξεις κώδικα, σε έναν χώρο εισόδου πολλών-διαστάσεων. Με αυτόν τον τρόπο διευκολύνεται η συμπίεση στοιχείων.Τα μοντέλα Kohonen μπορούν να παραχθούν με δύο τρόπους. Μπορούμε να χρησιμοποιήσουμε τις βασικές ιδέες αυτό-οργάνωσης που είναι η συνήθης προσέγγιση [159-161]. Εναλλακτικά, μπορούμε να χρησιμοποιήσουμε μια διανυσματική προσέγγιση κβάντισης που περιλαμβάνει έναν κωδικοποιητή και έναν αποκωδικοποιητή [162, 163].

Τα νευρωνικά δίκτυα Kohonen βρίσκουν πολλές εφαρμογές στις περιοχές της αναγνώρισης προτύπων, της κατηγοριοποίησης αλλά και της συμπίεσης δεδομένων.

1.2.6 Διαδικασία Εκμάθησης

Η ιδιότητα που αποτελεί πρωταρχική σημασία για ένα νευρωνικό δίκτυο είναι η δυνατότητα του δικτύου να μαθαίνει από το περιβάλλον του και να βελτιώνει την απόδοσή του μέσω της διαδικασίας εκμάθησης ή εκπαίδευσης. Η βελτίωση της απόδοσής του πραγματοποιείται με την πάροδο του χρόνου σύμφωνα με κάποιο ορισμένο μέτρο. Ένα νευρωνικό δίκτυο μαθαίνει το περιβάλλον του μέσω μιας διαδραστικής διαδικασίας προσαρμογής που εφαρμόζεται στα συναπτικά βάρη και στα επίπεδα πόλωσής του. Ιδανικά, το δίκτυο καθίσταται “πιο πεπειραμένο” όσον αφορά στο περιβάλλον του, μετά από κάθε επαναληπτική διαδικασία εκμάθησης.

Υπάρχουν πάρα πολλές δραστηριότητες που συνδέονται με την έννοια της “εκπαίδευσης” και ανάλογα με το ποια δραστηριότητα συνδέεται, δίνεται και ο ανάλογος ακριβής ορισμός. Ο ορισμός της διαδικασίας εκμάθησης είναι υποκειμενικό θέμα. Για παράδειγμα, η έννοια της εκμάθησης για έναν ψυχολόγο είναι αρκετά διαφορετική σε σχέση με την αίσθηση μάθησης στην τάξη. Αναγνωρίζοντας το ιδιαίτερο ενδιαφέρον για τα νευρωνικά δίκτυα σε αυτή τη διατριβή, δίνεται ο ορισμός της εκμάθησης όπως διατυπώθηκε από τους Mendel και McClaren [164].

> *Η εκμάθηση ή εκπαίδευση είναι μια διαδικασία με την οποία οι ελεύθερες παράμετροι ενός νευρωνικού δικτύου προσαρμόζονται μέσω μιας διαδικασίας διεγέρσεων από το περιβάλλον στο οποίο το δίκτυο είναι ενταγμένο. Ο τύπος εκμάθησης καθορίζεται από τον τρόπο με τον οποίο οι μεταβολές των παραμέτρων λαμβάνουν χώρα.*

Αυτός ο ορισμός υποδηλώνει την εξής ακολουθία γεγονότων:

- Το νευρωνικό δίκτυο διεγείρεται από το περιβάλλον.
- Το νευρωνικό δίκτυο υποβάλλεται σε μεταβολές των ελεύθερων παραμέτρων ως αποτέλεσμα αυτής της διέγερσης.
- Το νευρωνικό δίκτυο ανταποκρίνεται με ένα νέο τρόπο στο περιβάλλον λόγω των αλλαγών που έχουν συμβεί στην εσωτερική δομή του.

Σύμφωνα με τα παραπάνω, δίνεται ο ορισμός του αλγορίθμου εκμάθησης ή εκπαίδευσης (learning algorithm) από τον Haykin [2]:

> *Ένα καθορισμένο σύνολο καλώς ορισμένων κανόνων για την επίλυση ενός προβλήματος εκμάθησης ονομάζεται αλγόριθμος εκμάθησης.*

Υπάρχουν διάφοροι αλγόριθμοι εκμάθησης για τη σχεδίαση νευρωνικών δικτύων και ο καθένας τους έχει τα δικά του πλεονεκτήματα. Βασικά, οι αλγόριθμοι εκμάθησης διαφέρουν μεταξύ τους, στον τρόπο με τον οποίο συγκροτούνται οι νευρώνες μέσα στο δίκτυο όσον αφορά στο περιβάλλον του, και στον τρόπο με τον οποίο τροποποιούνται τα συναπτικά βάρη των νευρώνων.

Οι αλγόριθμοι εκμάθησης από τη σκοπιά της αλληλεπίδρασης του δικτύου με το περιβάλλον του χωρίζονται σε 3 διαφορετικά είδη:

- **Επιβλεπόμενη μάθηση (supervised learning):** Η επιβλεπόμενη μάθηση αναφέρεται και ως μάθηση με εκπαιδευτή και βασίζεται σε ένα διαθέσιμο σύνολο δεδομένων εισόδου-εξόδου, το οποίο εκφράζει τη γνώση του περιβάλλοντος που παρέχεται στο νευρωνικό δίκτυο. Το διαθέσιμο αυτό σύνολο είναι ένα δείγμα παραδειγμάτων, όπου κάθε παράδειγμα απαρτίζεται από ένα σήμα εισόδου (ερέθισμα) και το αντίστοιχο σήμα εξόδου που είναι η αντίστοιχη επιθυμητή απόκριση (στόχος). Το σύνολο αυτών των παραδειγμάτων αποκαλείται σύνολο δεδομένων εκπαίδευσης (training dataset). Ο στόχος της επιβλεπόμενης μάθησης είναι η κατάλληλη προσαρμογή των παραμέτρων του νευρωνικού δικτύου (των συναπτικών βαρών), δηλαδή ώστε το σφάλμα που εκφράζει τη διαφορά μεταξύ της πραγματικής και της επιθυμητής εξόδου του δικτύου να είναι ελάχιστο. Πιο απλά, θα πρέπει να γίνει μια συγκεκριμένη αντιστοίχιση εισόδου-εξόδου, ελαχιστοποιώντας μια συνάρτηση κόστους. Η διαδικασία επαναλαμβάνεται έως ότου το δίκτυο να μάθει να "μιμείται" τον εκπαιδευτή επαρκώς, σύμφωνα με κάποιο προκαθορισμένο στατιστικό κριτήριο.
- **Μη επιβλεπόμενη μάθηση (unsupervised learning):** Στη μη επιβλεπόμενη μάθηση δεν υπάρχει κάποιος εκπαιδευτής που να δίνει την επιθυμητή απόκριση και να επιβλέπει τη διαδικασία της μάθησης. Η μη επιβλεπόμενη μάθηση στηρίζεται σε ένα σύνολο παραδειγμάτων, όπου το κάθε παράδειγμα απαρτίζεται μόνο από ένα σήμα εισόδου. Εδώ, η επιθυμητή τιμή εξόδου δεν εμφανίζεται στο δίκτυο, αλλά παρέχεται ένα μέτρο της ποιότητας της αναπαράστασης που καλείται να μάθει το δίκτυο με αυτό-οργανώμενο τρόπο. Η βελτιστοποίηση των ελεύθερων παραμέτρων του δικτύου γίνεται με βάση αυτό το μέτρο. Πιο απλά, το νευρωνικό δίκτυο προσπαθεί να ανακαλύψει πρότυπα στα δεδομένα εισόδου.
- **Ενισχυτική μάθηση (reinforcement learning):** Η ενισχυτική μάθηση βρίσκεται κάπου ανάμεσα στην επιβλεπόμενη και στη μη επιβλεπόμενη μάθηση. Εδώ, δεν εμφανίζεται στο δίκτυο η επιθυμητή τιμή εξόδου, αλλά μόνο η πληροφορία αν η έξοδος του δικτύου είναι αποδεκτή ή όχι. Το νευρωνικό δίκτυο μαθαίνει μέσω διαρκούς αλληλεπίδρασης με το περιβάλλον του, με σκοπό την ελαχιστοποίηση ενός βαθμωτού δείκτη απόδοσης. Το νευρωνικό δίκτυο εκτελεί μια ενέργεια και μαθαίνει από την απόκριση που του παρέχει το περιβάλλον ως προς αυτή την ενέργεια. Ο στόχος είναι είτε να επιβραβεύεται είτε να τιμωρείται ο νευρώνας, όταν αποδίδει καλά ή όταν η απόδοση του δεν είναι αποδεκτή, αντίστοιχα. Ουσιαστικά, ο ρόλος που έχει ο εκπαιδευτής αντικαθίσταται από τον ρόλο ενός κριτή.

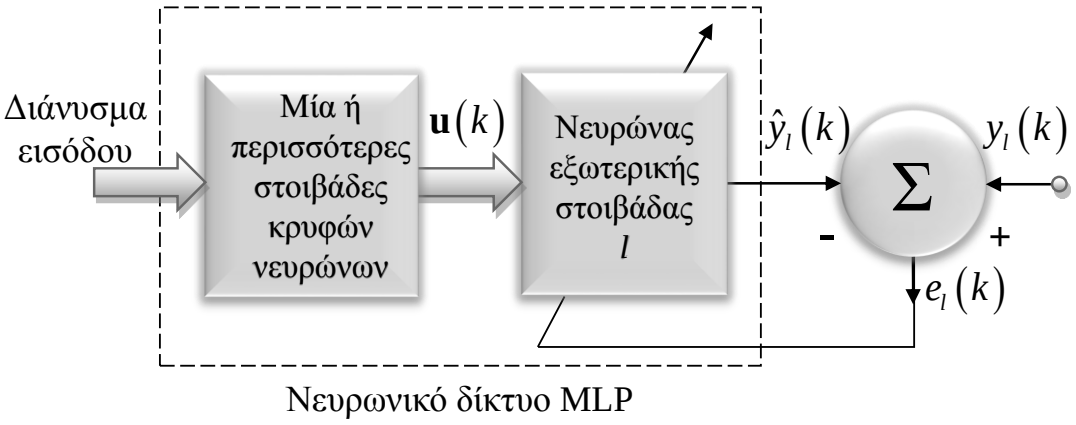
Από τη σκοπιά του τρόπου εκπαίδευσης του δικτύου (δηλαδή του τρόπου με τον οποίο ρυθμίζονται τα συναπτικά βάρη ενός νευρώνα), μπορούμε να διακρίνουμε διάφορους κανόνες εκμάθησης (learning rules), οι σημαντικότεροι εκ των οποίων είναι:

- **Εκμάθηση διόρθωσης σφάλματος (Error-correction learning)**: στοχεύει στην προσαρμογή των παραμέτρων του νευρωνικού δικτύου ώστε η έξοδός του να πλησιάζει όσο γίνεται περισσότερο την επιθυμητή τιμή.
- **Εκμάθηση με βάση τη μνήμη (Memory-based learning)**: βασίζεται στη ρητή απομνημόνευση των δεδομένων της εκπαίδευσης
- **Εκμάθηση Hebbian (Hebbian learning)**: προέρχεται από το χώρο της νευροβιολογίας και πριμοδοτεί τις συνάψεις τα άκρα των οποίων ενεργοποιούνται με ταυτόχρονο τρόπο.
- **Ανταγωνιστική εκμάθηση (Competitive learning):** προέρχεται από το χώρο της νευροβιολογίας, όπου βασίζεται στον ανταγωνισμό μεταξύ των νευρώνων της στοιβάδας εξόδου για το ποιος νευρώνας θα ενεργοποιηθεί.
- **Εκμάθηση Boltzmann (Boltzmann learning)**: προέρχεται από τον χώρο της στατιστικής μηχανικής

Ο πιο ευρέως διαδεδομένος κανόνας μάθησης είναι ο αλγόριθμος διόρθωσης σφάλματος ο οποίος αναλύεται παρακάτω. Για να γίνει εύκολα κατανοητός αυτός ο κανόνας εκμάθησής ας θεωρήσουμε την απλή περίπτωση ενός νευρώνα *l*, αποτελούμενο από μόνο έναν υπολογιστικό κόμβο στην εξωτερική στοιβάδα ενός δικτύου MLP, όπως απεικονίζεται στο σχήμα 1.12. Ο νευρώνας *l* οδηγείται από ένα διάνυσμα σήματος *u*(*k*) που παράγεται από μία ή περισσότερες κρυφές στοιβάδες νευρώνων, οι οποίοι με τη σειρά τους οδηγούνται από ένα διάνυσμα εισόδου που εφαρμόζεται στους κόμβους της εισόδου (δηλαδή, στη στοιβάδα εισόδου) του νευρωνικού δικτύου. Το *k* δείχνει το διακριτό χρόνο, ή ακριβέστερα το χρονικό βήμα μιας επαναληπτικής διαδικασίας που περιλαμβάνεται στον καθορισμό των συναπτικών βαρών του νευρώνα *l*. Το σήμα εξόδου ενός νευρώνα *l* συμβολίζεται με $\hat{y}_l(k)$. Αυτό το σήμα εξόδου που αντιπροσωπεύει τη μοναδική έξοδο του νευρωνικού δικτύου συγκρίνεται με την επιθυμητή έξοδο που συμβολίζεται με $y_l(k)$. Συνεπώς, παράγεται ένα σφάλμα που συμβολίζεται με $e_l(k)$. Έτσι, εξ' ορισμού έχουμε:

$$e_l(k) = y_l(k) - \hat{y}_l(k) \tag{1.9}$$

Το σφάλμα $e_l(k)$ ενεργοποιεί ένα μηχανισμό ελέγχου, ο σκοπός του οποίου είναι να



Σχήμα 1.12. Διάγραμμα βαθμίδων νευρωνικού δικτύου που εστιάζει στον μοναδικό νευρώνα της στοιβάδας εξόδου.

εφαρμόσει μια ακολουθία διορθώσεων στα συναπτικά βάρη του νευρώνα *l*. Οι διορθώσεις σχεδιάζονται για να κάνουν την τιμή του σήματος εξόδου $\hat{y}_l(k)$ ίση με την τιμή της επιθυμητής εξόδου $y_l(k)$ βαθμιαία. Αυτός ο στόχος επιτυγχάνεται με την ελαχιστοποίηση μιας συνάρτησης κόστους (cost function) ή ενός δείκτη απόδοσης, $\mathscr{E}(k)$, που καθορίζεται συναρτήσει του σφάλματος από τον παρακάτω τύπο:

$$\mathscr{E}(k) = \frac{1}{2} e_l^2(k) \tag{1.10}$$

Οι βαθμιαίες προσαρμογές στα συναπτικά βάρη του νευρώνα *k* συνεχίζονται έως ότου το σύστημα φτάσει σε μία σταθερή κατάσταση (δηλαδή, έως ότου σταθεροποιηθούν τα συναπτικά βάρη). Σε αυτό το σημείο, η διαδικασία εκμάθησης ολοκληρώνεται.

1.2.7 Γενίκευση τεχνητών νευρωνικών δικτύων

Βασικός στόχος της διαδικασίας της εκπαίδευσης ενός τεχνητού νευρωνικού δικτύου είναι να μπορεί το δίκτυο να αποκρίνεται καλά σε άγνωστα δεδομένα, τα οποία δεν χρησιμοποιήθηκαν ποτέ κατά τη δημιουργία ή εκπαίδευσή του, δηλαδή να πραγματοποιεί καλή γενίκευση. Ένα σημαντικό πρόβλημα που δυσχεραίνει την υλοποίηση του συγκεκριμένου στόχου είναι το φαινόμενο της υπερπροσαρμογής (overfitting) ή υπερεκπαίδευσης (overtraining) του δικτύου. Κατά τη διάρκεια εκπαίδευσής του, ένα νευρωνικό δίκτυο, μπορεί να οδηγηθεί στο φαινόμενο της υπερπροσαρμογής, δηλαδή σε μια κατάσταση όπου απομνημονεύει τα δεδομένα εκπαίδευσης και λειτουργεί μόνο βάσει αυτών. Αυτό μπορεί να συμβεί εάν βρει κάποιο χαρακτηριστικό (π.χ. οφειλόμενο σε θόρυβο) το οποίο υπάρχει στα δεδομένα εκπαίδευσης, αλλά όχι στην υποκείμενη συνάρτηση που πρέπει να μοντελοποιεί. Όταν το δίκτυο είναι υπερεκπαιδευμένο, δηλαδή εστιασμένο αποκλειστικά στα δεδομένα εκπαίδευσης, χάνει τη δυνατότητα να γενικεύει βασιζόμενο σε παρόμοια πρότυπα εισόδου-εξόδου και παρουσιάζει κακή απόδοση όταν εφαρμόζεται σε νέα δεδομένα που δεν έχουν χρησιμοποιηθεί κατά τη διαδικασία της εκπαίδευσης.

Για την αποφυγή του παραπάνω προβλήματος χρησιμοποιείται η μέθοδος της διασταυρωμένης επικύρωσης (cross-validation). Τα δύο πιο γνωστά είδη αυτής της μεθόδου είναι οι τεχνικές hold-out και *k*-επαναλήψεων (*k*-fold) [2]. Κατά τη διαδικασία της μεθόδου διασταυρωμένης επικύρωσης hold-out, το διαθέσιμο σύνολο δεδομένων χωρίζεται με τυχαίο τρόπο, συνήθως σε τρία ξεχωριστά υποσύνολα δεδομένων:

- **Σύνολο δεδομένων εκπαίδευσης (training subset)**: Τα μοντέλα νευρωνικών δικτύων, εκπαιδεύονται με βάση τα συγκεκριμένα δεδομένα, δηλαδή οι παράμετροι των δικτύων προσδιορίζονται χρησιμοποιώντας αυτό το σύνολο δεδομένων.
- **Σύνολο δεδομένων αξιολόγησης (validation subset)**: Τα εκπαιδευμένα υποψήφια μοντέλα νευρωνικών δικτύων αξιολογούνται με βάση αυτό το σύνολο δεδομένων και έπειτα επιλέγεται το μοντέλο που αποδίδει καλύτερα σε αυτό.
- **Σύνολο δεδομένων ελέγχου (testing dataset)**: Χρησιμοποιώντας αυτό το σύνολο δεδομένων ελέγχεται η αξιοπιστία του μοντέλου που επιλέχτηκε με βάση το σύνολο δεδομένων αξιολόγησης.

Κατ' αυτόν τον τρόπο, το σύνολο αξιολόγησης χρησιμοποιείται για να εκτιμηθεί η απόδοση των διάφορων υποψήφιων μοντέλων που εκπαιδεύτηκαν στο σύνολο εκπαίδευσης. Έτσι, επιλέγεται το «βέλτιστο» μοντέλο, δηλαδή επιλέγονται οι τιμές των παραμέτρων που δίνουν την καλύτερη απόδοση στο σύνολο αξιολόγησης. Υπάρχει, ωστόσο, σαφής πιθανότητα το μοντέλο που επιλέγεται με αυτό τον τρόπο να παρουσιάσει το φαινόμενο της υπερπροσαρμογής δηλαδή, να είναι υπερεκπαιδευμένο στα δεδομένα του υποσυνόλου αξιολόγησης. Για την εξάλειψη αυτού του κινδύνου, η απόδοση γενίκευσης του επιλεγμένου μοντέλου μετριέται βάσει ενός συνόλου ελέγχου το οποίο είναι εντελώς ξεχωριστό από τα σύνολα δεδομένων εκπαίδευσης και αξιολόγησης.

Σε γενικές γραμμές, στη βιβλιογραφία δεν υπάρχει κάποια συγκεκριμένη αναφορά σχετικά με το ποσοστό των δεδομένων που πρέπει να κατανέμονται σε κάθε υποσύνολο δεδομένων [165]. Σε αυτή τη διδακτορική διατριβή, συνήθως, ακολουθείται η κατανομή ποσοστών 50% - 25% - 25% που μοιράζονται τα διαθέσιμα δεδομένα με τυχαίο τρόπο στα υποσύνολα εκπαίδευσης - αξιολόγησης - ελέγχου, αντίστοιχα.

Κατά τη διαδικασία της μεθόδου διασταυρωμένης επικύρωσης k-επαναλήψεων, το διαθέσιμο σύνολο δεδομένων χωρίζεται σε k υποσύνολα, όπου $k>1$. Το δίκτυο εκπαιδεύεται πλήρως, χρησιμοποιώντας όλα τα υποσύνολα δεδομένων εκτός από ένα. Ύστερα, το μοντέλο που προκύπτει ελέγχεται στο υποσύνολο που έχει μείνει εκτός της διαδικασίας της εκπαίδευσης. Πιο απλά, η μέθοδος holdout επαναλαμβάνεται k φορές, όπου κάθε φορά ένα από τα k υποσύνολα χρησιμοποιείται ως σύνολο ελέγχου, ενώ τα υπόλοιπα k-1 υποσύνολα ενώνονται και σχηματίζουν ένα ενιαίο σύνολο που χρησιμοποιείται ως σύνολο εκπαίδευσης και ως σύνολο αξιολόγησης. Στο τέλος, υπολογίζεται ο μέσος όρος της απόδοσης των μοντέλων. Το πλεονέκτημα αυτής της μεθόδου είναι ότι η απόδοση του δικτύου δεν εξαρτάται από τον τρόπο διαχωρισμού του συνόλου δεδομένων. Από την άλλη, το μειονέκτημα αυτής της μεθόδου είναι ότι έχει αυξημένο υπολογιστικό φορτίο, αφού επαναλαμβάνεται όλη η διαδικασία k φορές.

Η πιο ακραία μορφή της μεθόδου διασταυρωμένης επικύρωσης k-επαναλήψεων, είναι η εξαντλητική τεχνική γνωστή ως μέθοδος leave-one-out, που προκύπτει όταν ο αριθμός των υποσυνόλων k επιλέγεται να είναι ίσος με τον αριθμό k των δεδομένων εκπαίδευσης. Αυτό σημαίνει ότι τα σύνολα ελέγχου αποτελούνται από ένα μόνο παράδειγμα. Κατά τη διαδικασία αυτή, ένα παράδειγμα εξαιρείται από το διαθέσιμο σύνολο δεδομένων και τα υπόλοιπα παραδείγματα χρησιμοποιούνται για την πλήρη εκπαίδευση ενός δικτύου. Το εκπαιδευμένο δίκτυο, στη συνέχεια, χρησιμοποιείται για να προβλέψει αυτό το ένα παράδειγμα που είχε εξαιρεθεί από τη διαδικασία της εκπαίδευσης. Αυτή η διαδικασία επαναλαμβάνεται k φορές, χρησιμοποιώντας κάθε φορά ένα διαφορετικό δεδομένο για τον έλεγχο του μοντέλου. Στο τέλος, όταν έχει ολοκληρωθεί αυτή η διαδικασία, υπολογίζεται ο μέσος όρος της απόδοσης των μοντέλων. Η διαδικασία αυτή συνήθως χρησιμοποιείται στα πολύ μικρά σύνολα δεδομένων.

Στην περίπτωση που υπάρχουν παράμετροι του αλγορίθμου που πρέπει να επιλεχθούν έτσι ώστε να επιτευχθεί η υψηλότερη αποτελεσματικότητα, το σύνολο εκπαίδευσης πρέπει να χωριστεί επιπλέον σε ένα "πραγματικό" σύνολο εκπαίδευσης και ένα σύνολο αξιολόγησης, το οποίο χρησιμοποιείται "εσωτερικά" από τον αλγόριθμο για να προσδιοριστούν οι βέλτιστες τιμές των παραμέτρων. Μόνο όταν αυτές σταθεροποιηθούν, εξετάζεται ο αλγόριθμος πάνω στο σύνολο ελέγχου.

1.2.8 Προβλήματα μάθησης - Δείκτες απόδοσης μοντέλων τεχνητών νευρωνικών δικτύων

Σκοπός του ελέγχου της απόδοσης ενός μοντέλου είναι η εκτίμηση για το πόσο χρήσιμο είναι το τελευταίο για ένα συγκεκριμένο σκοπό. Η εκτίμηση της αξιοπιστίας είναι επίσης ένα σημαντικό βήμα για την ανάπτυξη μοντέλων τεχνητών νευρωνικών δικτύων. Τα τεχνικά νευρωνικά δίκτυα μπορούν να αντιμετωπίσουν δύο βασικά προβλήματα μάθησης:

- **Προβλήματα προσέγγισης συνάρτησης (function approximation) ή Παλινδρόμησης (regression)**: Αυτού του είδους τα προβλήματα αφορούν στη πρόβλεψη συνεχών αριθμητικών τιμών. Με βάση ένα σύνολο δεδομένων εισόδου-εξόδου που δίνονται στο νευρωνικό δίκτυο, το τελευταίο πρέπει να προσεγγίσει μια άγνωστη συνάρτηση-στόχο, δηλαδή πρέπει να βρει τη σχέση που συνδέει τις εισόδους με τις εξόδους.
- **Προβλήματα αναγνώρισης προτύπων (pattern recognition) ή κατηγοριοποίησης / ταξινόμησης (classification)**: Αυτού του είδους τα προβλήματα αφορούν στη πρόβλεψη διακριτών τάξεων ή κλάσεων ή κατηγοριών. Στο νευρωνικό δίκτυο παρέχεται ένα σύνολο δεδομένων εισόδου-εξόδου που αφορούν σε στοιχεία εισόδου και κατηγορίες εξόδου. Με βάση αυτό το σύνολο δεδομένων, το νευρωνικό δίκτυο πρέπει να βρει τη συσχέτιση μεταξύ των δεδομένων εισόδου-εξόδου, ώστε να ταξινομήσει την έξοδο σε μια εκ των κλάσεων, δηλαδή το δίκτυο πρέπει να αναγνωρίσει σε ποια κατηγορία πρέπει να κατατάξει το κάθε πρότυπο.

Για την αξιολόγηση της απόδοσης των προβλέψεων ενός μοντέλου έχουν αναπτυχθεί διάφορες μέθοδοι. Σε αυτή τη διατριβή τα κριτήρια αξιολόγησης που χρησιμοποιούνται, για τα προβλήματα προσέγγισης συνάρτησης είναι οι μέθοδοι της τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (root mean square error, RMSE) και του συντελεστή προσδιορισμού R^2. Για τα προβλήματα ταξινόμησης αντίστοιχα χρησιμοποιούνται ο στατιστικός δείκτης ακρίβεια επί τοις εκατό (percent accuracy, ACC%) και ο συντελεστής συσχέτισης Matthews (Matthews correlation coefficient, MCC) [166], τα οποία περιγράφονται παρακάτω.

<u>Δείκτης τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος, RMSE</u>

Ο δείκτης RMSE είναι ένα απλό στατιστικό μέτρο και συνδέεται άμεσα με το τυπικό σφάλμα της παλινδρόμησης. Όσο πιο μικρή τιμή RMSE παρουσιάζει ένα μοντέλο, τόσο μεγαλύτερη είναι η ικανότητα πρόβλεψης που διαθέτει. Η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος υπολογίζεται από τη σχέση:

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{K}\left(\hat{y}_i - y_i\right)^2}{K}} \tag{1.11}$$

όπου

y : οι πραγματικές τιμές

$\hat{y}$: οι υπολογιζόμενες/προβλεπόμενες τιμές

K : ο συνολικός αριθμός των δεδομένων

Συντελεστής προσδιορισμού R^2

Ο συντελεστής προσδιορισμού R^2 μετρά το βαθμό "αξιοπιστίας" της γραμμικής παλινδρόμησης, δηλαδή δείχνει το πόσο καλά ένα σύνολο δεδομένων ταιριάζει σε ένα μοντέλο. Επίσης, ο συντελεστής προσδιορισμού R^2 δεν έχει μονάδα μέτρησης και το εύρος τιμών του είναι μεταξύ -∞ έως 1, όπου η τιμή 1 αντιστοιχεί στην τέλεια εφαρμογή. Τέλος, ο συντελεστής προσδιορισμού R^2 υπολογίζεται σύμφωνα με τον παρακάτω τύπο:

$$
\begin{aligned}
R^2 &= 1 - \frac{SS_{err}}{SS_{tot}} \\
SS_{err} &= \sum_{i=1}^{K} \left(y_i - \hat{y}_i \right)^2 \\
SS_{tot} &= \sum_{i=1}^{K} \left(y_i - \overline{y} \right)^2
\end{aligned}
\tag{1.12}
$$

όπου

y : οι πραγματικές τιμές

$\hat{y}$: οι προβλεπόμενες τιμές

$\overline{y}$: ο μέσος όρος των πραγματικών τιμών y

K : ο συνολικός αριθμός των δεδομένων

Στατιστικός δείκτης ακρίβεια %

Ο στατιστικός δείκτης ακρίβεια% είναι μια τυπική μέθοδος για την αξιολόγηση της απόδοσης ενός ταξινομητή και αποτελεί μια φυσική επιλογή, καθώς αξιολογεί τη συνολική επιτυχία κάθε ταξινομητή. Ο δείκτης ακρίβεια% υπολογίζεται σύμφωνα με την παρακάτω σχέση:

$$
\text{Ακρίβεια\%} = \frac{\text{Αριθμός συνολικών επιτυχών προβλέψεων}}{\text{Αριθμός συνολικών δεδομένων}} \times 100 \tag{1.13}
$$

Ωστόσο, υπάρχουν περιπτώσεις όπου είναι απαραίτητες πρόσθετες πληροφορίες από τον πίνακα σύγχυσης για να χαρακτηριστεί πλήρως η απόδοση ενός ταξινομητή. Ένα κοινό παράδειγμα είναι η ταξινόμηση των στοιχείων με μη ισορροπημένη κατανομή των κλάσεων, όπου ακόμη και μια συνολική αποτυχία στην πρόβλεψη μιας σπάνιας κατηγορίας (δηλαδή μιας κλάσης που περιέχει λίγα δεδομένα), θα έχει μόνο μια μικρή επίπτωση στο συνολικό ποσοστό ακρίβειας %.

Συντελεστής συσχέτισης Matthews, MCC

Για να ξεπεραστεί η αδυναμία που παρουσιάζει ο στατιστικός δείκτης ακρίβεια %, προτάθηκε ο συντελεστής MCC, ο οποίος για την αξιολόγηση της απόδοσης ενός προβλήματος ταξινόμησης δύο κλάσεων συνοψίζει τα αποτελέσματα όλου του πίνακα σύγχυσης (confusion matrix) σε μια μόνο τιμή. Δηλαδή, ο συντελεστής Matthews αντιπροσωπεύει τα ποσοστά επιτυχίας και αποτυχίας του ταξινομητή ανά επιμέρους κατηγορία.

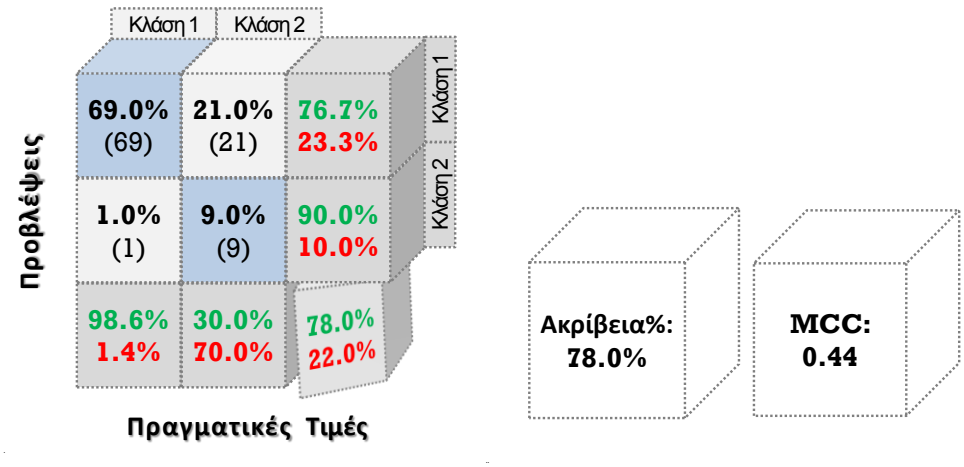
Έτσι, ο δείκτης MCC είναι κατάλληλος για την αξιολόγηση των επιδόσεων των ταξινομητών, ακόμη και για σύνολα δεδομένων που η κατανομή των δεδομένων τους σε κλάσεις δεν είναι ισορροπημένη. Στην περίπτωση προβλημάτων ταξινόμησης με περισσότερες από δύο κλάσεις, έχει υιοθετηθεί ένας γενικευμένος συντελεστής συσχέτισης Matthews [167], ο οποίος έχει αποδειχθεί ότι αποτελεί μια καλή βάση γενικής σύγκρισης, σε σχέση με διάφορα άλλα κριτήρια απόδοσης ταξινομητών [168]. Ο δείκτης MCC υπολογίζεται ως εξής:

$$\mathrm{MCC} = \frac{\sum_{k,l,m=1}^{M} C_{kk}C_{ml} - C_{lk}C_{km}}{\sqrt{\sum_{k=1}^{M}\left(\sum_{l=1}^{M} C_{lk}\right)\left(\sum_{\substack{f,g=1\\ f\neq k}}^{M} C_{gf}\right)}\sqrt{\sum_{k=1}^{M}\left(\sum_{l=1}^{M} C_{kl}\right)\left(\sum_{\substack{f,g=1\\ f\neq k}}^{M} C_{fg}\right)}} \tag{1.14}$$

όπου C_{ij} είναι ο αριθμός των στοιχείων της πραγματικής κλάσης i που ο ταξινομητής έχει ταξινομήσει στην κλάση j. Οι τιμές του δείκτη MCC ανήκουν στο πεδίο [-1 1], όπου η τιμή 1 αντιστοιχεί στην τέλεια ταξινόμηση.

Τέλος, θα πρέπει να γίνει μία μικρή αναφορά στους πίνακες σύγχυσης (confusion matrices). Η αρχή του πίνακα σύγχυσης είναι ότι αναγνωρίζει τη φύση των σφαλμάτων, καθώς και την ποσότητά τους. Ένας πίνακας σύγχυσης είναι ένας τετραγωνικός πίνακας, με αριθμό γραμμών και στηλών ίσο με τον αριθμό των κλάσεων (σχήμα 1.13).

Οι στήλες του πίνακα σύγχυσης αντιστοιχούν στις πραγματικές κλάσεις εξόδου, ενώ οι γραμμές στις προβλεπόμενες. Έτσι, το εκάστοτε στοιχείο του πίνακα αναπαριστά τον αριθμό των προτύπων που ανήκουν στην κατηγορία που δείχνει η γραμμή του πίνακα και ανατέθηκαν στην κατηγορία που δείχνει η στήλη του πίνακα. Η κύρια διαγώνιος του πίνακα σύγχυσης περιέχει τα στοιχεία που κατηγοριοποιήθηκαν σωστά, ενώ τα υπόλοιπα στοιχεία του πίνακα αφορούν σε εσφαλμένες προβλέψεις. Επίσης, η στήλη που βρίσκεται δεξιότερα στον πίνακα, παρουσιάζει τις πιθανότητες που έχει το μοντέλο να προβλέπει σωστά μία συγκεκριμένη κλάση με πράσινο χρώμα και τις πιθανότητες που έχει να προβλέπει λάθος μία κλάση με κόκκινο χρώμα. Αντίθετα, η γραμμή που βρίσκεται στο κάτω μέρος του πίνακα, παρουσιάζει τις πιθανότητες που έχει η κλάση να προβλεφθεί σωστά από το μοντέλο με πράσινο χρώμα και τις πιθανότητες που έχει η κλάση να προβλεφθεί εσφαλμένα από το μοντέλο με κόκκινο χρώμα.



Σχήμα 1.13. Πίνακας Σύγχυσης.

1.3 Ασαφής λογική

Η έννοια της ασαφούς λογικής (fuzzy logic) εισήχθηκε το 1965 από τον Lotfi Aliasker Zadeh, έναν καθηγητή του Πανεπιστημίου Μπέρκλεϋ στην Καλιφόρνια [169]. Η μαθηματική θεμελίωση της ασαφούς λογικής βασίζεται στη θεωρία των ασαφών συνόλων (fuzzy set theory). Επιπλέον, μαθηματικά, ο όρος ασάφεια (fuzziness) σημαίνει πολλαπλότητα τιμών και προέρχεται από την αρχή της απροσδιοριστίας ή αρχή της αβεβαιότητας του Heisenberg.

Η ασαφής λογική μπορεί να θεωρηθεί ως μια επέκταση της κλασσικής αριστοτέλειας λογικής. Μια πρόταση μπορεί να είναι αληθής με κάποιο βαθμό αληθείας και όχι απλά αληθής ή ψευδής. Δηλαδή, η ασαφής λογική λέει ότι τα πράγματα συχνά δεν είναι «άσπρο - μαύρο» αλλά «αποχρώσεις του γκρι». Η ιδέα αυτή ήταν επανάσταση στη θεωρία της λογικής, γιατί ξέφυγε από το μοντέλο που κυριαρχούσε εδώ και τόσα χρόνια, δηλαδή το μοντέλο του «0-1» ή αντίστοιχα «αληθές - ψευδές» (Boolean λογική).

Η ασαφής λογική παρέχει τη δυνατότητα στα συστήματα να μπορούν να λειτουργούν και να εξάγουν συμπεράσματα μέσα σε συνθήκες αβεβαιότητας, όπως κάνουν οι εμπειρογνώμονες. Δηλαδή, μπορούν να δώσουν λύσεις σε δύσκολα προβλήματα τα οποία δεν περιγράφονται με μαθηματικές εξισώσεις ή μοντέλα. Ο όρος αυτής της αβεβαιότητας διατυπώθηκε από τον Zadeh ως μια αρχή, την οποία ονόμασε "αρχή του ασυμβίβαστου". Η αρχή του ασυμβίβαστου λέει ότι:

> «*Καθώς η πολυπλοκότητα ενός συστήματος αυξάνει, η ικανότητα μας να προβαίνουμε σε ακριβείς και σημαντικές δηλώσεις για τη συμπεριφορά του μειώνεται μέχρι που να φθάσουμε σε ένα όριο πέρα από το οποίο ακρίβεια και σημαντικότητα (ή σχετικότητα) καθίστανται σχεδόν αμοιβαίως αποκλειόμενα χαρακτηριστικά*».

Στην ουσία, ένα ασαφές σύστημα μπορεί να δημιουργηθεί κατά τέτοιον τρόπο έτσι ώστε να εμπεριέχει τις γνώσεις ενός ειδικού πάνω στο θέμα που επιχειρεί να αντιμετωπίσει. Ουσιαστικά, η γνώση κωδικοποιείται μέσα στο σύστημα ασαφούς λογικής. Για αυτό το λόγο τα ασαφή συστήματα (fuzzy systems) ανήκουν στην κατηγορία των έμπειρων συστημάτων (expert systems).

Η αρχή του ασυμβίβαστου χρησιμοποιώντας τη θεωρία των ασαφών συνόλων έχει ως σκοπό της την αναπαράσταση της συγκεχυμένης και μη ακριβούς γνώσης του ανθρώπου άμεσα, δηλαδή, χωρίς τη μεσολάβηση κάποιου ακριβούς μαθηματικού τύπου ή μοντέλου.

Η κλασική θεωρία των συνόλων ουσιαστικά λέει ότι ένα στοιχείο ή θα ανήκει ή δεν θα ανήκει σε ένα σύνολο. Αυτή η λογική μπορεί να περιγραφεί και με τους αριθμούς «1» (ανήκει) και «0» (δεν ανήκει). Η θεωρία όμως των ασαφών συνόλων λέει ότι ένα στοιχείο μπορεί να ανήκει σε ένα σύνολο σύμφωνα με ένα βαθμό συμμετοχής ο οποίος περιγράφεται με έναν αριθμό μεταξύ του μηδέν «0» και της μονάδας «1», δηλαδή έναν αριθμό που ανήκει στο διάστημα [0,1]. Ο βαθμός συμμετοχής προσδιορίζεται από τις συναρτήσεις συμμετοχής (membership function) που μπορούν να αναπαραστήσουν μία υποκειμενική ιδέα μίας αόριστης έκφρασης. Τέλος, θα πρέπει να αναφερθεί ότι οι συναρτήσεις συμμετοχής μπορούν να υπολογιστούν και με βάση στατιστικά δεδομένα.

Με τη βοήθεια της ασαφούς λογικής μπορεί να δοθεί λύση σε δύσκολα προβλήματα για τα οποία δεν υπάρχουν διαθέσιμες μαθηματικές εξισώσεις ή μοντέλα, παρά μόνο ανθρώπινη εμπειρία. Τα συστήματα ασαφούς λογικής βρίσκουν εφαρμογές σε περιοχές όπως:

- Αυτόματος έλεγχος
- Λήψη αποφάσεων

- Αναγνώριση και ταξινόμηση προτύπων
- Μοντελοποίηση συστημάτων

1.4 Εξελικτικός υπολογισμός

Με τον όρο εξελικτικός υπολογισμός (evolutionary computation, EC) προσδιορίζεται μια κλάση στοχαστικών μεθόδων βελτιστοποίησης. Οι εξελικτικοί αλγόριθμοι συνιστούν μια κατηγορία μεθόδων αναζήτησης γενικού σκοπού, οι οποίες επιτυγχάνουν μια αξιοσημείωτη ισορροπία μεταξύ ολικής (exploration) και τοπικής (exploitation) εξερεύνησης - δύο φαινομενικά αντικρουόμενων στόχων σε οποιαδήποτε τεχνική βελτιστοποίησης. Υπάρχουν ορισμένες διαφορές μεταξύ της τεχνικής του εξελικτικού υπολογισμού και των συμβατικών μεθόδων βελτιστοποίησης, συμπεριλαμβανομένων :

- Ο εξελικτικός υπολογισμός χειρίζεται τις παραμέτρους του προβλήματος σε κωδικοποιημένη μορφή, αντί των ίδιων των παραμέτρων.
- Ο εξελικτικός υπολογισμός δεν χρησιμοποιεί βοηθητικές πληροφορίες σχετικά με την αντικειμενική συνάρτηση, όπως οι παράγωγοι.
- Ο εξελικτικός υπολογισμός χρησιμοποιεί έναν πληθυσμό από λύσεις που λειτουργούν παράλληλα, σε αντίθεση με τις συμβατικές μεθόδους βελτιστοποίησης που χειρίζονται μία μόνο λύση.

Ως εκ τούτου, οι τεχνικές εξελικτικού υπολογισμού θεωρούνται αποτελεσματικές και ισχυρές μέθοδοι βελτιστοποίησης που είναι σε θέση να ξεφύγουν από τα τοπικά βέλτιστα. Αυτές οι ιδιότητες έχουν κάνει τις τεχνικές εξελικτικού υπολογισμού ευρέως γνωστές σε διάφορους και ποικίλους επιστημονικούς τομείς.

Ο εξελικτικός υπολογισμός έχει τη βάση του στις εξελικτικές διαδικασίες που παρατηρούνται στη φύση, όπου ο βασικός κανόνας είναι ότι τα άτομα που προσαρμόζονται καλύτερα στο περιβάλλον τους, έχουν περισσότερες πιθανότητες να επιβιώσουν. Μιμούμενοι τις φυσικές διαδικασίες εξέλιξης, οι τεχνικές του εξελικτικού υπολογισμού χρησιμοποιούν πληθυσμούς ατόμων, όπου το κάθε άτομο αντιπροσωπεύει μια πιθανή λύση ενός προβλήματος. Καθώς ο πληθυσμός εξελίσσεται, οι λύσεις βελτιώνονται με τελικό σκοπό την εύρεση της βέλτιστης λύσης.

Σύμφωνα με τον Engelbrecht η έννοια του εξελικτικού υπολογισμού εισήχθη τη δεκαετία του 1950 από τους Fraser [170, 171], Bremermann [172] και Reed [173]. Παρόλα αυτά, πατέρας του εξελικτικού υπολογισμού, και πιο συγκεκριμένα των γενετικών αλγορίθμων, θεωρείται ο John Holland. Τη δεκαετία του 1960, ο Ingo Rechenberg [174, 175] εισήγαγε την ιδέα των “στρατηγικών εξέλιξης” ως μια μέθοδο βελτιστοποίησης, η οποία αναπτύχθηκε περισσότερο από τον Hans-Paul Schwefel [176, 177]. Το 1966 οι Fogel et.al. [178] ανέπτυξαν τον εξελικτικό προγραμματισμό, μια τεχνική στην οποία οι υποψήφιες λύσεις σε δεδομένα προβλήματα παραστάθηκαν ως μηχανές πεπερασμένων καταστάσεων, οι οποίες εξελίχθηκαν μεταλλάσσοντας τυχαία τα διαγράμματα μετάπτωσης των καταστάσεών τους και επιλέγοντας το πλέον κατάλληλο. Οι γενετικοί αλγόριθμοι προτάθηκαν από τον John Holland [179], όπου ο αρχικός του στόχος ήταν να μελετήσει το φαινόμενο της προσαρμογής, όπως λαμβάνει χώρα στη φύση, και να αναπτύξει τρόπους με τους οποίους οι μηχανισμοί της φυσικής προσαρμογής θα μπορούσαν να εμφυτευτούν στα υπολογιστικά

συστήματα. Επιπλέον, ο Holland ήταν αυτός ο οποίος προσπάθησε πρώτος να θέσει την έννοια της υπολογιστικής εξέλιξης σε μια σταθερή θεωρητική βάση.

Η εξέλιξη είναι μια μέθοδος βελτιστοποίησης, με στόχο να βελτιωθεί η ικανότητα της επιβίωσης ενός συστήματος σε ένα ανταγωνιστικό περιβάλλον που αλλάζει δυναμικά. Στη φυσική εξέλιξη, όπου η κύρια ιδέα είναι η επιβίωση του ισχυρότερου (ο αδύναμος πρέπει να πεθάνει), η επιβίωση επιτυγχάνεται μέσω της αναπαραγωγής. Οι απόγονοι, αναπαράγονται από δύο γονείς (μερικές φορές περισσότερους από δύο) και περιέχουν γενετικό υλικό και των δύο (ή του συνόλου) των γονέων. Εκείνα τα άτομα που κληρονομούν τα «λάθος» χαρακτηριστικά είναι αδύναμα και χάνουν τη μάχη της επιβίωσης.

Ο Jean-Baptiste Lamarck (1744 – 1829) ήταν ο πρώτος που διατύπωσε μια ολοκληρωμένη θεωρία εξέλιξης, η οποία βασίστηκε στην κληρονομικότητα των ειδών. Η κεντρική ιδέα, είναι ότι τα είδη προσαρμόζονται κατά τη διάρκεια της ζωής τους, και μεταφέρουν τα χαρακτηριστικά τους, στους απογόνους τους, όπου οι τελευταίοι εξακολουθούν να προσαρμόζονται. Επίσης, σύμφωνα με τον Lamarck, η μέθοδος της προσαρμοστικότητας στηρίζεται στην ιδέα της «χρήσης» και της «αχρηστίας»: Με την πάροδο του χρόνου τα είδη χάνουν τα χαρακτηριστικά που δεν χρησιμοποιούν και αναπτύσσουν αυτά τα οποία τους είναι χρήσιμα.

Το 1858 ο Charles Robert Darwin και ο Alfred Russel Wallace διατύπωσαν τη θεωρία της εξέλιξης μέσω της φυσικής επιλογής. Η θεωρία της εξέλιξης του Δαρβίνου, αποτέλεσε θεμέλιο λίθο στη βιολογική εξέλιξη, η οποία συνοψίζεται ως εξής: Σε έναν κόσμο με περιορισμένες πηγές και σταθερούς πληθυσμούς, υπάρχει ανταγωνισμός μεταξύ των ειδών με σκοπό την επιβίωση. Τα είδη με τα βέλτιστα χαρακτηριστικά είναι πιθανότερο να επιβιώσουν και να αναπαραχθούν, με αποτέλεσμα αυτά τα χαρακτηριστικά να τα κληρονομήσουν οι απόγονοι τους. Αυτά τα επιθυμητά χαρακτηριστικά που περνούν στις επόμενες γενιές γίνονται κυρίαρχα με την πάροδο του χρόνου στους πληθυσμούς.

Σε ένα δεύτερο τμήμα της θεωρίας του Δαρβίνου αναφέρεται ότι κατά την ανάπτυξη ενός παιδικού οργανισμού, τυχαία γεγονότα προκαλούν τυχαίες μεταβολές στα χαρακτηριστικά του. Αν οι μεταβολές που υφίσταντο τα χαρακτηριστικά είναι προς όφελος του οργανισμού, τότε οι πιθανότητες επιβίωσής του, αυξάνονται.

Οι εξελικτικοί αλγόριθμοι αναφέρονται σε συστήματα που λύνουν υπολογιστικά προβλήματα, τα οποία χρησιμοποιούν υπολογιστικά μοντέλα εξελικτικών διαδικασιών, όπως είναι η φυσική επιλογή (natural selection) και η επιβίωση του καλύτερου (survival of the fittest), ως τις θεμελιώδεις συνιστώσες τους. Οι εξελικτικοί αλγόριθμοι χρησιμοποιούν ένα πληθυσμό ατόμων (individuals), δηλαδή υποψήφιων λύσεων, όπου κάθε άτομο αποτελείται από χρωμοσώματα (chromosomes) και κάθε χρωμόσωμα αναπαριστά το σύνολο των μεταβλητών του προβλήματος. Κάθε χρωμόσωμα (που ουσιαστικά αποτελεί μια πιθανή λύση ενός δεδομένου προβλήματος) καθορίζει τα χαρακτηριστικά των ατόμων στον πληθυσμό και αποτελείται από γονίδια (genes). Κάθε γονίδιο επηρεάζει την κληρονομικότητα ενός ή περισσοτέρων χαρακτηριστικών. Κάθε χαρακτηριστικό μπορεί να εμφανιστεί με διάφορες μορφές, ανάλογα με την τιμή του αντίστοιχου γονιδίου που το επηρεάζει. Οι διάφορες τιμές των γονιδίων αποκαλούνται αλληλόμορφες (alleles).

Σε κάθε γενιά, τα άτομα ανταγωνίζονται για να αναπαράγουν απογόνους, όπου τα άτομα με τις καλύτερες δυνατότητες επιβίωσης έχουν τη μεγαλύτερη ευκαιρία για να αναπαραχθούν. Τρεις βασικοί μηχανισμοί που αποκαλούνται γενετικοί τελεστές (bio-inspired operators), συμβάλλουν στην αναπαραγωγή των ατόμων: η επιλογή (selection), η διασταύρωση (crossover) και η μετάλλαξη (mutation). Η επιλογή προσδιορίζει τα άτομα του πληθυσμού που επιβιώνουν με σκοπό τη μεταφορά των γονιδίων τους στην επόμενη γενιά και την αναπαραγωγή επιτυχημένων απογόνων. Η διασταύρωση εξασφαλίζει την ανάμειξη των

επιτυχημένων γονιδίων με σκοπό τη διατήρηση μιας ποικιλομορφίας επιτυχημένων απογόνων που οδηγεί στη βέλτιστη προσαρμογή στο περιβάλλον τους. Η μετάλλαξη πραγματοποιεί μια τυχαία αλλαγή των γονιδίων με κάποια πιθανότητα, δηλαδή μεταβάλλει μερικά από το αλληλόμορφα του χρωμοσώματος. Η πιθανότητα επιβίωσης ενός ατόμου μετράται χρησιμοποιώντας μια συνάρτηση καταλληλότητας (fitness function) που αντικατοπτρίζει τους στόχους και τους περιορισμούς του προβλήματος που πρέπει να επιλυθεί. Ύστερα από κάθε γενιά, τα άτομα μπορούν να επιβιώσουν ή να μην επιβιώσουν στην επόμενη γενιά. Επιπλέον, τα χαρακτηριστικά της συμπεριφοράς που αναφέρονται και ως φαινότυποι (phenotypes) μπορεί να χρησιμοποιηθούν για να επηρεάσουν την εξελικτική διαδικασία με δύο τρόπους: μπορεί να υπάρξουν αλλαγές στη γενετική τους δομή, και/ή τα χαρακτηριστικά της συμπεριφοράς να εξελίσσονται χωριστά.

Ανάλογα με τους τρόπους που εφαρμόζονται τα παραπάνω, οι εξελικτικοί αλγόριθμοι χωρίζονται στις παρακάτω βασικές κατηγορίες:

- **Γενετικοί αλγόριθμοι (Genetic algorithms):** είναι οι πιο ευρέως γνωστοί εξελικτικοί αλγόριθμοι και μοντελοποιούνται σύμφωνα με τη γενετική εξέλιξη των ειδών.
- **Γενετικός προγραμματισμός (Genetic programming):** αποτελεί ειδική εφαρμογή των γενετικών αλγορίθμων όπου τα άτομα είναι προγράμματα που αναπαρίστανται ως δέντρα.
- **Εξελικτικές Στρατηγικές (Evolution strategies):** προσανατολίζονται προς τη μοντελοποίηση των στρατηγικών παραμέτρων που ελέγχουν τη μεταβολή της εξέλιξης.
- **Εξελικτικός προγραμματισμός (Evolutionary programming):** έχει τον ίδιο στόχο με τους γενετικούς αλγορίθμους και τον γενετικό προγραμματισμό, αλλά διαφέρει ουσιαστικά στο ότι δίνει έμφαση στην ανάπτυξη μοντέλων συμπεριφοράς και όχι γενετικών μοντέλων.
- **Διαφορική εξέλιξη (Differential evolution):** είναι παρόμοια με τους γενετικούς αλγορίθμους, με διαφορετικό μηχανισμό στην αναπαραγωγή.

1.4.1 Ο αλγόριθμος εξελικτικής προσομοιωμένης ανόπτησης

Ο αλγόριθμος εξελικτικής προσομοιωμένης ανόπτησης (evolutionary simulated annealing, ESA) [3] αποτελεί μία υβριδική μέθοδο, που συνδυάζει τις βασικές αρχές δύο ευρέως γνωστών ευρετικών μεθόδων αναζήτησης, της προσομοιωμένης ανόπτησης και των εξελικτικών αλγορίθμων.

Η προσομοιωμένη ανόπτηση (simulated annealing, SA) είναι μία τεχνική στοχαστικής βελτιστοποίησης, η οποία επινοήθηκε από τον Nicholas Metropolis [81]. Ανόπτηση αποκαλείται η διεργασία ανακατανομής των ατόμων κατά την ψύξη ενός θερμοδυναμικού συστήματος. Η ανόπτηση των μετάλλων είναι η διαδικασία κατά την οποία το κράμα θερμαίνεται σε πολύ υψηλή θερμοκρασία και στη συνέχεια ψύχεται αργά με σκοπό να βελτιωθεί η ευκαμψία του. Σε υψηλές θερμοκρασίες, παρατηρείται ελεύθερη κινητικότητα των μορίων προς όλες τις κατευθύνσεις. Κατά τη διάρκεια της ψύξης, η κινητικότητα των μορίων περιορίζεται. Έτσι, τα τελευταία ανακατανέμονται και όταν η θερμοκρασία μειωθεί αρκετά και ο ρυθμός ψύξης είναι αργός, μπορούν να δημιουργήσουν κρυσταλλικές δομές, που αποτελούν τις καταστάσεις ελάχιστης ενέργειας του συστήματος. Στην περίπτωση, που ο ρυθμός ψύξης δεν είναι αρκετά αργός, το σύστημα δεν φτάνει στην κατάσταση ελάχιστης ενέργειας, αλλά καταλήγει να έχει μεγαλύτερη ενέργεια.

Οι πρώτοι που εφάρμοσαν την ιδέα της ανόπτησης των μετάλλων στη βελτιστοποίηση ήταν οι Kirkpatrick et al. [82]. Η μέθοδος της προσομοιωμένης ανόπτησης προσομοιάζει την αναζήτηση της ελάχιστης ενεργειακής κατάστασης, με την αναζήτηση του ελαχίστου μιας μαθηματικής συνάρτησης, μέσα από μια επαναληπτική διαδικασία. Ο πρότυπος αλγόριθμος προσομοιωμένης ανόπτησης χρησιμοποιεί μία μόνο λύση. Η τελευταία, συνήθως, εκφράζεται ως ένα διάνυσμα που αποτελείται από ακέραιες ή πραγματικές τιμές.

Ο αλγόριθμος ξεκινά επιλέγοντας με τυχαίο τρόπο μία λύση του προβλήματος. Σε κάθε επανάληψη, η λύση εξελίσσεται και μια νέα λύση παράγεται εφαρμόζοντας μια τυχαία αλλαγή στην προηγούμενη. Η νέα λύση αξιολογείται με βάση την αντικειμενική συνάρτηση και μπορεί να γίνει αποδεκτή ή να διαγραφεί, αναλόγως με το τι διαφορά έχει σε σχέση με την προηγούμενη αξιολόγηση. Κατά τη διάρκεια των πρώτων επαναλήψεων είναι πιθανό να γίνει δεκτή ακόμη και μια χειρότερη λύση σε σύγκριση με την προηγούμενη, διενεργώντας έτσι τυχαία έρευνα. Αυτή η διαδικασία προσομοιάζει στην τυχαία κίνηση των μορίων σε υψηλές θερμοκρασίες. Η υποβάθμιση της ποιότητας των λύσεων μπορεί να αποδειχθεί απαραίτητη, όσον αφορά στη δυνατότητα του αλγορίθμου να αποφύγει τον εγκλεισμό σε τοπικά ελάχιστα. Καθώς ο αλγόριθμος προχωρά, η πιθανότητα να γίνει δεκτή μια χειρότερη λύση μικραίνει σταδιακά, μιμούμενη τη μειωμένη κινητικότητα των μορίων καθώς η θερμοκρασία ελαττώνεται. Στην πραγματικότητα, αυτή η πιθανότητα δίνεται από τη συνάρτηση κατανομής Boltzmann-Gibbs:

$$p_B = \begin{cases} 1 & \text{, αν} \quad f\left(\mathbf{x}_{B_i}\right) < f\left(\mathbf{x}_{B_{i-1}}\right) \\ \exp\left(-\dfrac{f\left(\mathbf{x}_{B_i}\right) - f\left(\mathbf{x}_{B_{i-1}}\right)}{k_B T_B}\right) & \text{, διαφορετικά} \end{cases} \tag{1.15}$$

όπου T_B είναι η θερμοκρασία σε κάθε επανάληψη, k_B είναι η σταθερά Boltzmann, f είναι η αντικειμενική συνάρτηση και τα $\mathbf{x}_{B_i}$, $\mathbf{x}_{B_{i-1}}$ αντιπροσωπεύουν δύο διαδοχικές λύσεις κατά τη διάρκεια της διαδικασίας της ανόπτησης.

Η απόδοση του αλγορίθμου προσομοιωμένης ανόπτησης εξαρτάται από το πρόγραμμα ανόπτησης που εφαρμόζεται, δηλαδή αναδιάταξης των μορίων, ξεκινώντας από μια αρχική-υψηλή θερμοκρασία $T_{B_{max}}$, η οποία μειώνεται σταδιακά μέχρι να φτάσει στην τελική-χαμηλότερη θερμοκρασία $T_{B_{min}}$. Η επιλογή του προγράμματος ανόπτησης εξαρτάται κυρίως από τη φύση του προβλήματος. Το πιο κλασικό πρόγραμμα ανόπτησης είναι η εκθετική ψύξη, όπου η θερμοκρασία μειώνεται με διαδοχικά διακριτά βήματα t σύμφωνα με την ακόλουθη εξίσωση:

$$T_B\left(t+1\right) = a_B T_B\left(t\right), \quad \text{όπου } 0 < a_B < 1 \tag{1.16}$$

όπου a_B είναι η εκθετική σταθερά ψύξης

Σε αντίθεση με τις κλασικές εξελικτικές τεχνικές, όπως για παράδειγμα οι γενετικοί αλγόριθμοι, ο πρότυπος αλγόριθμος προσομοιωμένης ανόπτησης χρησιμοποιεί μόνο μία λύση, αντί ενός πληθυσμού λύσεων. Η χρήση μίας μόνο λύσης αποτελεί πλεονέκτημα όσον αφορά στο υπολογιστικό φορτίο του αλγορίθμου, αλλά από την άλλη μειώνει την πιθανότητα εύρεσης του ολικού βέλτιστου. Μια επιπλέον διαφορά του αλγορίθμου προσομοιωμένης

ανόπτησης σε αντίθεση με τις προαναφερθείσες τεχνικές, είναι ότι αν και στον πρώτο επιτρέπονται οι επιζήμιες κινήσεις κατά τη διάρκεια των πρώτων επαναλήψεων, στη συνέχεια, η μέθοδος συμπεριφέρεται σχεδόν όπως ένας αλγόριθμος αναρρίχησης λόφου (hill-climbing algorithm), που αποδέχεται μόνο μία βελτιωμένη λύση. Το γεγονός αυτό καθιστά την αποφυγή των τοπικών ελαχίστων κατά τα τελευταία στάδια του αλγορίθμου προσομοιωμένης ανόπτησης ακόμα πιο δύσκολη.

Για να αποφευχθούν τα μειονεκτήματα αυτά, ο αλγόριθμος προσομοιωμένης ανόπτησης συνδυάστηκε με εξελικτικές τεχνικές, δημιουργώντας έτσι τον αλγόριθμο εξελικτικής προσομοιωμένης ανόπτησης. Ο αλγόριθμος εξελικτικής προσομοιωμένης ανόπτησης ξεκινά με έναν πληθυσμό από τυχαία αρχικοποιημένες λύσεις, οι οποίες κωδικοποιούνται ως χρωμοσώματα. Σε αντίθεση, όμως, με τους γενετικούς αλγορίθμους, δεν υπάρχουν γενετικοί τελεστές που να εμπλέκονται, όπως η διασταύρωση, ή η μετάλλαξη. Στην πραγματικότητα, ο μόνος τελεστής που δρα στα χρωμοσώματα, είναι ο τελεστής προσομοιωμένης ανόπτησης. Αυτό συνεπάγεται ότι κάθε χρωμόσωμα της τρέχουσας γενιάς χρησιμοποιείται ως αρχική λύση στην εκτέλεση του κλασικού αλγόριθμου προσομοιωμένης ανόπτησης. Η τελική λύση που παράγεται από τον αλγόριθμο προσομοιωμένης ανόπτησης, στη συνέχεια, συγκρίνεται με την αρχική λύση όπου η καλύτερη εκ των δύο επιβιώνει, δηλαδή μεταφέρεται στην επόμενη γενιά. Ο αλγόριθμος επαναλαμβάνεται μέχρι να επιτευχθεί τουλάχιστον ένα από τα κριτήρια τερματισμού. Κριτήρια που θα μπορούσαν να χρησιμοποιηθούν ως προϋποθέσεις τερματισμού του αλγορίθμου εξελικτικής προσομοιωμένης ανόπτησης περιλαμβάνουν:

- Την ολοκλήρωση ενός μέγιστου αριθμού επαναλήψεων
- Την αδυναμία βελτίωσης της αντικειμενικής συνάρτησης επί σειρά διαδοχικών γενεών
- Την εύρεση μιας αποδεκτής λύσης

Θα πρέπει να σημειωθεί ότι τα δύο πρώτα κριτήρια αναφέρονται στις γενιές που αντιστοιχούν στην εξελικτική διαδικασία. Ο αριθμός των επαναλήψεων που εκτελούνται από τον αλγόριθμο προσομοιωμένης ανόπτησης έχει μια σταθερή τιμή την οποία επιλέγει ο χρήστης.

1.5 Νοημοσύνη σμήνους

Η κοινωνική συμπεριφορά που επιδεικνύουν οι βιολογικοί οργανισμοί που οργανώνονται σε ομάδες ήταν πάντα αντικείμενο θαυμασμού από τον άνθρωπο. Μελέτες της κοινωνικής συμπεριφοράς των εντόμων που ζουν σε αποικίες όπως είναι τα μυρμήγκια, οι μέλισσες, οι τερμίτες αλλά και των διαφόρων ψαριών που κινούνται σε κοπάδια και των πουλιών που κινούνται σε σμήνη, ώθησαν στον σχεδιασμό ιδιαίτερα αποτελεσματικών μεθόδων βελτιστοποίησης. Η επιστημονική παρατήρηση του τρόπου δράσης των κοινωνικοποιημένων βιολογικών οργανισμών, αποκάλυψε τον τρόπο με τον οποίο η διασπορά πληροφορίας μέσα σε μια ομάδα, καθώς και η εκμετάλλευση της συλλογικής εμπειρίας, οδηγούν στον ευφυή τρόπο επίλυσης προβλημάτων. Ένας όρος που θα μπορούσε να περιγράψει την έννοια της ομάδας κοινωνικοποιημένων βιολογικών οργανισμών είναι το σμήνος. Τυπικά, ένα σμήνος μπορεί να οριστεί ως μια ομάδα συστημάτων ή οντοτήτων που επικοινωνούν μεταξύ τους (είτε άμεσα είτε έμμεσα), δρώντας στο περιβάλλον τους.

Η νοημοσύνη σμήνους (swarm intelligence, SI) [1] αναφέρεται στην ιδιότητα φυσικών ή τεχνητών συστημάτων που επιδεικνύουν συλλογικά ευφυή συμπεριφορά. Ενώ, η

υπολογιστική νοημοσύνη σμήνους (computational swarm intelligence) αναφέρεται στα αλγοριθμικά μοντέλα των συμπεριφορών των οντοτήτων που αλληλεπιδρούν. Η νοημοσύνη σμήνους μπορεί επίσης να αναφέρεται ως συλλογική νοημοσύνη.

Τα συστήματα νοημοσύνης σμήνους αποτελούνται από έναν πληθυσμό μελών ατόμων ή σωματιδίων, τα οποία αλληλεπιδρούν μεταξύ τους και με το περιβάλλον τους σε τοπικό επίπεδο. Το κάθε άτομο έχει μια σχετικά απλή δομή, αλλά η συλλογική συμπεριφορά που επιδεικνύουν πολλά άτομα ως σμήνος είναι συνήθως πολύ περίπλοκη. Η σύνθετη συμπεριφορά ενός σμήνους δεν αποτελεί ιδιότητα ενός μεμονωμένου ατόμου, αντιθέτως αυτή η συμπεριφορά είναι άγνωστη στα μέλη του σμήνους και δεν μπορούν να την εμφανίσουν μόνα τους. Η σύνθετη συμπεριφορά ενός σμήνους είναι απόρροια των σχετικά απλών ενεργειών των μελών του ή των αλληλεπιδράσεων των ατόμων μεταξύ τους και με το περιβάλλον τους, η οποία αναφέρεται ως ανάδυση (emergence). Συγκεκριμένα, ανάδυση είναι η διαδικασία με την οποία προκύπτουν νέες και συνεκτικές δομές, πρότυπα και ιδιότητες (ή συμπεριφορές) σε ένα σύνθετο σύστημα. Αυτές οι δομές, τα πρότυπα και οι συμπεριφορές προκύπτουν από τις αλληλεπιδράσεις των ατόμων με το περιβάλλον τους δίχως να υπάρχει κάποιος κεντρικός μηχανισμός ελέγχου, που να υπαγορεύει πως θα πρέπει να συμπεριφέρεται το κάθε μέλος ξεχωριστά.

Ως εκ τούτου, η συλλογική συμπεριφορά ενός σμήνους κοινωνικοποιημένων βιολογικών οργανισμών αναδύεται από τις συμπεριφορές των ατόμων του που αλληλεπιδρούν με έναν μη γραμμικό τρόπο. Η ατομική και η συλλογική συμπεριφορά των μελών του σμήνους είναι στενά συνδεδεμένες: η συλλογική συμπεριφορά των ατόμων διαμορφώνει και υπαγορεύει τη συμπεριφορά του σμήνους. Από την άλλη πλευρά, η συμπεριφορά του σμήνους επηρεάζει τις συνθήκες κάτω από τις οποίες κάθε μέλος του σμήνους ενεργεί. Οι ενέργειες αυτές μπορεί να αλλάξουν το περιβάλλον, και έτσι οι συμπεριφορές του εν λόγω ατόμου και των γειτόνων του μπορεί επίσης να αλλάξουν - το οποίο και πάλι μπορεί να αλλάξει τη συλλογική συμπεριφορά του σμήνους.

Με βάση τα παραπάνω το πιο σημαντικό συστατικό της νοημοσύνης σμήνους και της διευκόλυνσης των αναδυόμενων συμπεριφορών, είναι η αλληλεπίδραση (interaction) ή συνεργασία (cooperation). Η αλληλεπίδραση μεταξύ των ατόμων βοηθά στη βελτίωση της γνώσης του κάθε ατόμου για το περιβάλλον του. Η αλληλεπίδραση στα βιολογικά συστήματα πραγματοποιείται με πολλούς τρόπους, εκ των οποίων η κοινωνική αλληλεπίδραση είναι η πιο σημαντική. Η κοινωνική αλληλεπίδραση μπορεί να είναι άμεση (μέσω της σωματικής επαφής, της οπτικής επαφής, του ήχου, ή της χημικής αντίληψης) ή έμμεση (μέσω των τοπικών αλλαγών του περιβάλλοντος). Ο όρος στιγμετρία (stigmergy) αναφέρεται στην έμμεση μορφή επικοινωνίας μεταξύ των ατόμων.

Μερικά από τα χαρακτηριστικά παραδείγματα νοήμονος συμπεριφοράς στη φύση αποτελούν οι τερμίτες που δημιουργούν φωλιές οι οποίες είναι τεράστιες, εντυπωσιακές και πολύπλοκες κατασκευές που μπορεί να περιέχουν τούνελ, σπηλιές και να επεκτείνονται δεκάδες μέτρα σε βάθος ή σε μήκος. Άλλο παράδειγμα είναι τα κοπάδια ψαριών που καθορίζουν τη συμπεριφορά τους, όπως την ταχύτητα και την κατεύθυνση προς την οποία κολυμπούν, με βάση ένα μικρό αριθμό γειτόνων. Οι αγέλες αρπακτικών ζώων όπως είναι οι λύκοι αποτελούν μια επιθετική περίπτωση νοήμονος συμπεριφοράς, η οποία αποσκοπεί στη στρατηγική παγίδευση της λείας.

Μια από τις πρώιμες μελέτες της νοημοσύνης σμήνους ασχολήθηκε με τα μυρμήγκια κατά τη διαδικασία αναζήτησης της τροφής τους. Η μοντελοποίηση της φερομόνης που εναποθέτουν τα μυρμήγκια κατά τη διαδικασία αναζήτησης των συντομότερων διαδρομών που οδηγούν από την αποικία τους στην τροφή, έχει ως αποτέλεσμα την ανάπτυξη αλγορίθμων βελτιστοποίησης αποικίας μυρμηγκιών (ant colony optimization, ACO) [180].

Μία άλλη μέθοδος που εντάσσεται στις τεχνικές νοημοσύνης σμήνους είναι ο αλγόριθμος τεχνητής αποικίας μελισσών (artificial bee colony, ABC) [181], ο οποίος προσομοιώνει τη διαδικασία του 'χορού των μελισσών' (waggled dance) που ένα σμήνος από μέλισσες πραγματοποιεί κατά τη διαδικασία αναζήτησης τροφής. Όταν οι μέλισσες βρουν τροφή, συλλέγουν μία ποσότητα από αυτήν και εφόσον επιστρέψουν στην κυψέλη, την ξεφορτώνουν και ενημερώνουν τις υπόλοιπες μέλισσες για την πηγή της συγκεκριμένης τροφής πραγματοποιώντας ένα είδος χορού. Τέλος, μία από τις κυριότερες μεθόδους νοημοσύνης σμήνους αποτελεί ο αλγόριθμος βελτιστοποίησης σμήνους σωματιδίων (Particle Swarm Optimization, PSO) [1, 57], ο οποίος είναι μια ιδιαίτερα αποτελεσματική μέθοδος βελτιστοποίησης. Μια σύντομη περιγραφή του βασικού αλγορίθμου PSO δίνεται στην επόμενη ενότητα.

1.5.1 Ο αλγόριθμος βελτιστοποίησης σμήνους σωματιδίων

Η βελτιστοποίηση σμήνους σωματιδίων, ή εν συντομία PSO, εισήχθη από τους Kennedy και Eberhart το 1995 στην εργασία [57], η οποία προέκυψε αρχικά ως εργαλείο μελέτης της κοινωνικής συμπεριφοράς που παρουσιάζουν τα μέλη ενός πληθυσμού ειδών στη φύση, όπως για παράδειγμα τα ψάρια που κολυμπούν σε σχηματισμό κοπαδιού. Υστέρα, μετεξελίχθηκε σε μέθοδο βελτιστοποίησης, η οποία προσομοιώνει την κοινωνική συμπεριφορά που επιδεικνύουν τα πουλιά που πετούν σε σχηματισμό σμήνους. Το σμήνος αποτελείται από ένα συνολικό πληθυσμό σωματιδίων, όπου κάθε σωματίδιο αποτελεί και μια διαφορετική πιθανή λύση του προβλήματος. Η μέθοδος PSO χρησιμοποιεί ένα πληθυσμό πιθανών λύσεων με παρόμοιο τρόπο με τους εξελικτικούς αλγορίθμους (σμήνος=πληθυσμός, σωματίδιο=άτομο). Έτσι, στην περίπτωση του αλγορίθμου PSO οι λύσεις κωδικοποιούνται ως σωματίδια, τα οποία «πετούν» στον πολυδιάστατο χώρο αναζήτησης, ψάχνοντας για τη βέλτιστη λύση. Κάθε σωματίδιο αλλάζει τη θέση του με στοχαστικό τρόπο, προσπαθώντας να επαναλάβει τις ατομικές του προηγούμενες επιτυχημένες θέσεις, και ταυτόχρονα να μιμηθεί τις επιτυχίες των γειτονικών του σωματιδίων. Αυτή η διαδικασία, οδηγεί σε μια απλή συλλογική συμπεριφορά με την οποία ανακαλύπτεται η βέλτιστη λύση, ακόμα και σε χώρους πολλών διαστάσεων.

Ο αλγόριθμος PSO λειτουργεί θεωρώντας έναν χώρο N-διαστάσεων, ίσο σε διαστάσεις με τον αριθμό των μεταβλητών του προβλήματος που πρέπει να βελτιστοποιηθούν. Σε αυτό τον χώρο τοποθετείται, με τυχαίο τρόπο, ένα σύνολο σωματιδίων P, το οποίο είναι γνωστό ως σμήνος. Κάθε σωματίδιο i, την επανάληψη t βρίσκεται μέσα στο χώρο αναζήτησης στη θέση $\mathbf{x}_i(t)$, όπου i=1,2,…,P. Η θέση του κάθε σωματιδίου του σμήνους αντιστοιχεί σε μία κωδικοποιημένη πιθανή λύση του προβλήματος βελτιστοποίησης, και κάθε επανάληψη t ανανεώνεται με βάση την ακόλουθη αναδρομική σχέση:

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}_i(t+1) \tag{1.17}$$

Από την παραπάνω εξίσωση, προκύπτει ότι η αλλαγή των θέσεων των σωματιδίων εξαρτάται από το διάνυσμα της ταχύτητας **v**. Η ταχύτητα του σωματιδίου i υπολογίζεται ως εξής:

$$v_{ij}(t+1) = w_{inertia} v_{ij}(t) + c_1 r_{1j}(t)\left[y_{ij}(t) - x_{ij}(t)\right] + c_2 r_{2j}(t)\left[\hat{y}_j(t) - x_{ij}(t)\right] \tag{1.18}$$

όπου:

$w_{inertia}$: είναι η σταθερά αδράνειας (inertia),

$v_{ij}(t)$ είναι η ταχύτητα του σωματιδίου i=1,2,…,P, στη διάσταση j=1,2,…,N, την επανάληψη t,

$x_{ij}(t)$ είναι η θέση του σωματιδίου i=1,2,…,P, στη διάσταση j=1,2,…,N, την επανάληψη t,

$y_{ij}(t)$ είναι η καλύτερη προσωπική θέση που έχει πετύχει το σωματίδιο i=1,2,…,P, στη διάσταση j=1,2,…,N, την επανάληψη t,

$\hat{y}_j(t)$ είναι η καλύτερη θέση που έχει πετύχει συνολικά το σμήνος στη διάσταση j=1,2,…,N, την επανάληψη t,

c_1 και c_2 είναι σταθερές επιτάχυνσης,

$r_{1j}(t)$ και $r_{2j}(t)$ παίρνουν τυχαίες τιμές στο διάστημα [0, 1], που δειγματοληπτούνται κάθε επανάληψη t από μία ομοιόμορφη κατανομή.

Η εξίσωση (1.18) περιλαμβάνει τρεις όρους:

1. Ο πρώτος όρος είναι γνωστός ως ο όρος μνήμης (memory term) και αντιπροσωπεύει την επίδραση που έχει η προηγούμενη ταχύτητα του σωματιδίου στην τρέχουσα ταχύτητα του. Το $w_{inertia}$ αντιπροσωπεύει την αδράνεια (inertia), δηλαδή την τάση που έχουν τα σωματίδια να διατηρούν σταθερή την προηγούμενη τους ταχύτητα. Μπορούμε να πούμε πως συνολικά ο όρος αυτός δείχνει την ορμή του σωματιδίου.
2. Ο δεύτερος όρος είναι γνωστός ως *γνωστικός όρος* (cognitive term) και αντιπροσωπεύει την προσωπική εμπειρία του κάθε σωματιδίου. Ο όρος αυτός είναι ανάλογος της απόστασης του ενός σωματιδίου από τη δική του καλύτερη θέση. Η σταθερά c_1 δρα σαν ενίσχυση του γνωστικού όρου και είναι επίσης γνωστή ως *νοσταλγία* (nostalgia) [57], διότι εκφράζει την επιθυμία του κάθε σωματιδίου να ανακτήσει την καλύτερή του θέση.
3. Ο τρίτος όρος είναι γνωστός ως *κοινωνικός όρος* (social term) και αντιπροσωπεύει τη συνολική εμπειρία του σμήνους. Ο όρος αυτός είναι ανάλογος της απόστασης του ενός σωματιδίου από την καλύτερη θέση του σμήνους που έχει επιτευχθεί συνολικά. Η σταθερά c_2 δρα σαν ενίσχυση του κοινωνικού όρου και είναι επίσης γνωστή ως *ζήλεια* (envy), διότι εκφράζει την επιθυμία του κάθε σωματιδίου να κινείται τόσο καλά όσο και τα γειτονικά του σωματίδια.

Μια σημαντική παράμετρος που ουσιαστικά καθορίζει την ακρίβεια και την απόδοση του αλγορίθμου PSO είναι η εξισορρόπηση μεταξύ ολικής και τοπικής εξερεύνησης. Ολική εξερεύνηση (exploration) είναι η ικανότητα του σμήνους να εξερευνεί διαφορετικές περιοχές του χώρου για να βρει καλύτερη λύση. Προκειμένου να συμβεί αυτό, θα πρέπει το σμήνος να διασκορπίζεται στο χώρο αναζήτησης, που σημαίνει ότι οι ταχύτητες των σωματιδίων θα πρέπει να έχουν μεγάλες τιμές. Τοπική εξερεύνηση (exploitation) είναι η ικανότητα του σμήνους να επικεντρώνει την αναζήτησή του σε μια περιορισμένη περιοχή του χώρου με σκοπό να ανακαλύψει καλύτερες λύσεις. Σε αυτήν την περίπτωση όμως θα πρέπει το σμήνος να βρίσκεται συγκεντρωμένο σε μια συγκεκριμένη περιοχή, που σημαίνει ότι οι ταχύτητες των σωματιδίων θα πρέπει να έχουν μικρές τιμές.

Η εξισορρόπηση μεταξύ ολικής και τοπικής εξερεύνησης έχει ζωτική σημασία στην επιτυχή επίλυση ενός προβλήματος βελτιστοποίησης, και μπορεί να ρυθμιστεί μέσω του μηχανισμού περιορισμού ταχύτητας (velocity clamping), μέσω του όρου αδράνειας, κτλ. [1]. Ο μηχανισμός περιορισμού ταχύτητας περιορίζει τις ταχύτητες των σωματιδίων για να παραμείνουν εντός κάποιων προκαθορισμένων ορίων, προκειμένου να επιτευχθεί καλύτερη τοπική εξερεύνηση:

$$v_{ij}(t+1) = \begin{cases} v_{ij}(t+1), & \text{if } \left\| v_{ij}(t+1) \right\| < V_{max} \\ \pm V_{\max}, & \text{διαφορετικά} \end{cases} \tag{1.19}$$

Μεγάλες τιμές της σταθεράς περιορισμού ταχύτητας (velocity clamping constant) V_{max} οδηγούν σε αναζήτηση σε μεγαλύτερο κομμάτι του χώρου, αυξάνοντας τις πιθανότητες να ξεφύγει η μέθοδος από κάποιο τοπικό ελάχιστο και να βρεθεί κάποια απομακρυσμένη καλύτερη λύση. Αντίθετα, μικρές τιμές οδηγούν σε καλύτερη τοπική εξερεύνηση σε συγκεκριμένα κομμάτια του χώρου, αυξάνοντας τις πιθανότητες να ανακαλυφθεί μια καλύτερη τοπική λύση.

Όπως αναφέρθηκε προηγουμένως, η εξισορρόπηση μεταξύ ολικής και τοπικής εξερεύνησης μπορεί να ρυθμιστεί και μέσω του όρου αδράνειας. Ο όρος της αδράνειας εισήχθει από τους Shi και Eberhart [182] ως ένας μηχανισμός για τον έλεγχο της τοπικής και της ολικής εξερεύνησης του σμήνους [183]. Η αδράνεια $w_{inertia}$, ελέγχει την ορμή του σωματιδίου σταθμίζοντας τη συμβολή της προηγούμενης ταχύτητας - ουσιαστικά ελέγχει κατά πόσο η προηγούμενη κατεύθυνση της πτήσης θα επηρεάσει τη νέα ταχύτητα. Η τιμή της αδράνειας είναι εξαιρετικά σημαντική, διότι επηρεάζει τη συμπεριφορά σύγκλισης και την ισορροπία ανάμεσα στην τοπική και στην ολική εξερεύνηση.

- Για $w_{inertia} \geq 1$, οι ταχύτητες αυξάνονται με το χρόνο, επιταχύνοντας προς τη μέγιστη ταχύτητα, και το σμήνος αποκλίνει (ολική εξερεύνηση), με αποτέλεσμα τα σωματίδια να αδυνατούν να αλλάξουν κατεύθυνση για να μετακινηθούν προς υποσχόμενες περιοχές.
- Για $w_{inertia} < 1$, τα σωματίδια επιβραδύνουν μέχρις ότου η ταχύτητά τους να φτάσει στο μηδέν (ανάλογα με τις τιμές των σταθερών επιτάχυνσης c_1 και c_2), με αποτέλεσμα τα σωματίδια να κάνουν τοπική αναζήτηση και έτσι μπορούν εύκολα να εντοπίσουν με μεγάλη ακρίβεια τη θέση ενός τοπικού ελαχίστου, αλλά δεν μπορούν να εξερευνήσουν καλά όλο τον χώρο.

Οι μεγάλες τιμές της αδράνειας διευκολύνουν την ολική εξερεύνηση, ενώ οι μικρές τιμές διευκολύνουν την τοπική εξερεύνηση. Ωστόσο, πολύ μικρές τιμές της αδράνειας σχεδόν μηδενίζουν την ικανότητα εξερεύνησης του σμήνους [1]. Η επιλογή της τιμής για το $w_{inertia}$ πρέπει να γίνεται σε συνδυασμό με την επιλογή των τιμών για τις σταθερές επιτάχυνσης c_1 και c_2.

Η καλύτερη προσωπική θέση y_i είναι η καλύτερη θέση που έχει βρει το σωματίδιο i κατά τη διάρκεια υλοποίησης της μεθόδου. Για προβλήματα ελαχιστοποίησης, η καλύτερη προσωπική θέση την επόμενη επανάληψη υπολογίζεται ως εξής:

$$\mathbf{y}_i(t+1)=\begin{cases}\mathbf{y}_i(t) & , \text{ αν } f(\mathbf{x}_i(t+1)) \geq f(\mathbf{y}_i(t)) \\ \mathbf{x}_i(t+1) , & \text{διαφορετικά}\end{cases} \tag{1.20}$$

όπου f είναι η συνάρτηση καταλληλότητας. Για προβλήματα βελτιστοποίησης η καλύτερη προσωπική θέση υπολογίζεται ως εξής:

$$\mathbf{y}_i(t+1)=\begin{cases}\mathbf{y}_i(t) & , \text{ αν } f(\mathbf{x}_i(t+1)) \leq f(\mathbf{y}_i(t)) \\ \mathbf{x}_i(t+1) , & \text{διαφορετικά}\end{cases} \tag{1.21}$$

Η καλύτερη ολική θέση $\hat{\mathbf{y}}(t)$ κάθε επανάληψη t υπολογίζεται ως η καλύτερη θέση μεταξύ των καλύτερων προσωπικών θέσεων που έχουν βρει τα σωματίδια τις προηγούμενες επαναλήψεις. Για προβλήματα ελαχιστοποίησης, η καλύτερη ολική θέση υπολογίζεται ως εξής:

$$f(\hat{\mathbf{y}}(t)) = \min\left(f(\mathbf{y}_1(t)),\ldots, f(\mathbf{y}_P(t))\right) \tag{1.22}$$

Ενώ, για προβλήματα μεγιστοποίησης η καλύτερη ολική θέση υπολογίζεται ως εξής:

$$f(\hat{\mathbf{y}}(t)) = \max\left(f(\mathbf{y}_1(t)),\ldots, f(\mathbf{y}_P(t))\right) \tag{1.23}$$

Τέλος, για τον αλγόριθμο PSO έχουν προταθεί διάφορα κριτήρια ως συνθήκες διακοπής (stopping conditions), συμπεριλαμβανομένων των παρακάτω:

- Όταν επιτυγχάνεται ένας μέγιστος αριθμός επαναλήψεων
- Όταν βρίσκεται μια αποδεκτή λύση
- Όταν η κλίση της συνάρτησης καταλληλότητας γίνεται πολύ μικρή
- Όταν η κανονικοποιημένη ακτίνα του σμήνους πλησιάζει κοντά στο μηδέν

Τα ανωτέρω κριτήρια χρησιμοποιούνται συνήθως σε συνδυασμό μεταξύ τους για καλύτερα (και συνήθως ταχύτερα) αποτελέσματα.

1.6 Τεχνητά Ανοσοποιητικά Συστήματα

Το βιολογικό ανοσοποιητικό σύστημα είναι ένα πολύπλοκο δίκτυο που αποτελείται από εξειδικευμένους ιστούς, όργανα, κύτταρα και χημικά μόρια. Το φυσικό ανοσοποιητικό σύστημα (natural immune system) είναι υπεύθυνο για την άμυνα του οργανισμού και έχει μια αξιοθαύμαστη ικανότητα να διακρίνει τα ξένα κύτταρα (αντιγόνα) που εισέρχονται στον τελευταίο από τα φυσιολογικά κύτταρα που ανήκουν στον οργανισμό. Η ικανότητα αυτή ονομάζεται ανοσία, καθώς όταν το φυσικό ανοσοποιητικό σύστημα συναντά ένα αντιγόνο, η

προσαρμοστική φύση του πρώτου ενεργοποιείται και απομνημονεύει τη δομή αυτού του αντιγόνου για ταχύτερη απόκριση στο μέλλον.

Σε ερευνητικό επίπεδο, το φυσικό ανοσοποιητικό σύστημα εμπεριέχει τέσσερα μοντέλα:

- **Κλασικό μοντέλο:** Ένα από τα κύρια χαρακτηριστικά γνωρίσματα σύμφωνα με το κλασσικό μοντέλο του φυσικού ανοσοποιητικού συστήματος είναι η ικανότητά του να διακρίνει τα φυσιολογικά κύτταρα από τα ξένα κύτταρα, χρησιμοποιώντας λεμφοκύτταρα που παράγονται στα λεμφοειδή όργανα.
- **Θεωρία Επιλογής Κλώνων:** Ένα ενεργό Β-κύτταρο παράγει αντισώματα και εξαπλώνεται μέσω μιας διαδικασίας κλωνοποίησης. Οι παραγόμενοι κλώνοι επίσης περνούν από μία διαδικασία μετάλλαξης.
- **Θεωρία Κινδύνου:** Το ανοσοποιητικό σύστημα έχει την ικανότητα να διακρίνει τα επικίνδυνα από τα ακίνδυνα αντιγόνα.
- **Θεωρία Δικτύου:** Τα Β-κύτταρα σχηματίζουν ένα δίκτυο κυττάρων. Όταν ένα Β-κύτταρο του δικτύου αντιδρά σε ένα αντιγόνο, ενεργοποιείται και διεγείρει όλα τα άλλα Β-κύτταρα με τα οποία είναι συνδεδεμένο.

Το τεχνητό ανοσοποιητικό σύστημα (artificial immune system, AIS) είναι εμπνευσμένο από το φυσικό ανοσοποιητικό σύστημα και μιμείται τις βασικές λειτουργίες του τελευταίου που αφορούν στους παθογόνους εισβολείς όπως:

- η αναγνώριση
- η καταστροφή
- η εκμάθηση
- η μνήμη

Τα AIS εφαρμόζονται κυρίως στην επίλυση των προβλημάτων αναγνώρισης προτύπων, στους τομείς της ασφάλειας πληροφοριακών συστημάτων, της ανίχνευσης σφαλμάτων και της διάγνωσης βλαβών.

ΚΕΦΑΛΑΙΟ 2

ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ RBF

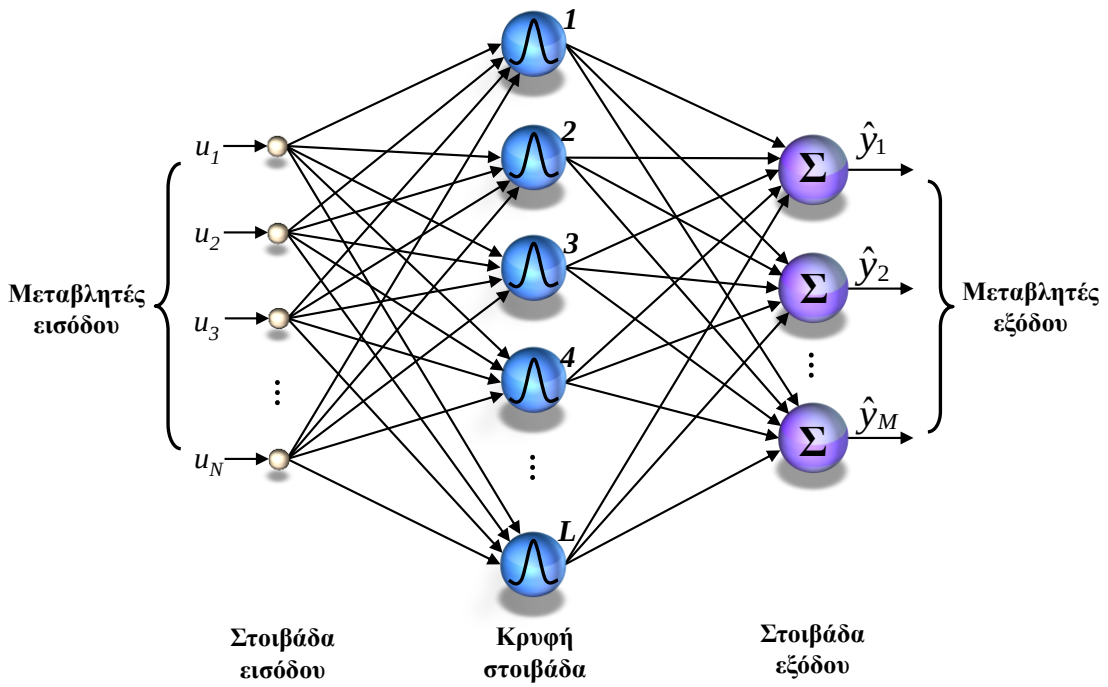
Αρχικά, στο παρόν κεφάλαιο παρουσιάζονται αναλυτικά τα νευρωνικά δίκτυα RBF, η δομή τους, και ο τυπικός αλγόριθμος εκπαίδευσης που χρησιμοποιούν. Στη συνέχεια, περιγράφεται η ασαφής διαμέριση και η μη συμμετρική ασαφής διαμέριση και αντίστοιχα, παρουσιάζεται ο συμμετρικός αλλά και ο μη συμμετρικός αλγόριθμος των ασαφών μέσων, οι οποίοι καθορίζουν από μόνοι τους τη διάσταση της κρυφής στοιβάδας. Τέλος, τα δίκτυα RBF συγκρίνονται με άλλες αρχιτεκτονικές και συγκεκριμένα με τα δίκτυα MLP.

2.1. Η δομή των δικτύων RBF

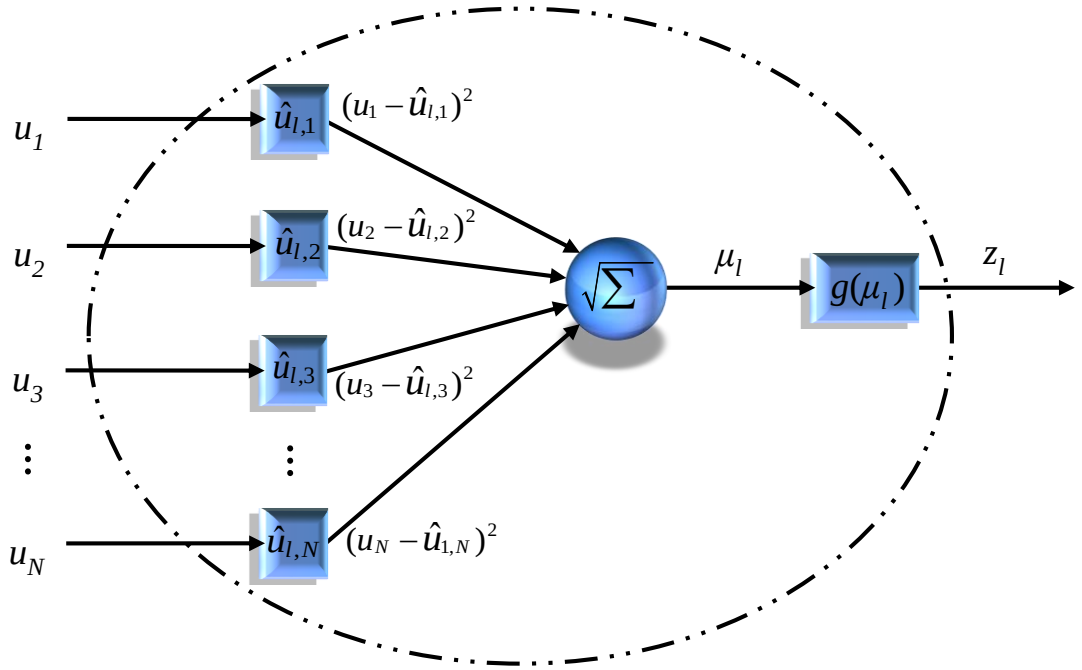
Τα νευρωνικά δίκτυα ακτινικής συνάρτησης βάσης (radial basis function, RBF) είναι ένας σχετικά νέος τύπος αρχιτεκτονικής νευρωνικών δικτύων που εισήχθη στη δεκαετία του 1980 ως πρόταση βελτίωσης των νευρωνικών δικτύων πρόσθιας τροφοδότησης [5, 6]. Ένα δίκτυο RBF μπορεί να θεωρηθεί ως ένα ειδικού τύπου νευρωνικό δίκτυο, το οποίο αποτελείται μόνο από τρεις στοιβάδες με τελείως διαφορετικούς ρόλους και η τυπική δομή του φαίνεται στο σχήμα 2.1. Τα δίκτυα RBF αποτελούνται από τη στοιβάδα εισόδου, την κρυφή στοιβάδα και τη στοιβάδα εξόδου, και χρησιμοποιούνται, είτε σε προβλήματα προσέγγισης συνάρτησης, είτε σε προβλήματα ταξινόμησης. Δεδομένου ότι οι συναρτήσεις ενεργοποίησης είναι μη γραμμικές, δεν είναι απαραίτητο να υπάρχουν περισσότερες από μία κρυφές στοιβάδες για να προσεγγίσουν οποιαδήποτε συνάρτηση. Κάθε στοιβάδα αποτελείται από κάποιους νευρώνες. Ο ρόλος των νευρώνων της κάθε στοιβάδας περιγράφεται στη συνέχεια:

Η πρώτη στοιβάδα είναι η στοιβάδα εισόδου, η οποία αποτελείται από N κόμβους εισόδου που συνδέουν το νευρωνικό δίκτυο με το περιβάλλον του. Οι νευρώνες της στοιβάδας εισόδου είναι ίσοι σε αριθμό, με τη διάσταση του χώρου εισόδου, δηλαδή είναι τόσοι, όσες είναι και οι μεταβλητές εισόδου. Η στοιβάδα εισόδου δεν εκτελεί υπολογισμούς, παρά μόνο μεταφέρει το σήμα των μεταβλητών εισόδου προς την κρυφή στοιβάδα.

Η δεύτερη στοιβάδα, που είναι και η μόνη κρυφή στοιβάδα, αποτελείται από L υπολογιστικές μονάδες, επίσης γνωστές ως κόμβοι ή νευρώνες ή πυρήνες RBF. Ένας τυπικός νευρώνας μίας κρυφής στοιβάδας απεικονίζεται στο σχήμα 2.2. Κάθε κρυφός νευρώνας l=1,2,…,L χαρακτηρίζεται από ένα κέντρο $\hat{\mathbf{u}}_l$, όπου $\hat{\mathbf{u}}_l^T = \left[\hat{u}_{1,l}, \hat{u}_{2,l}, \ldots, \hat{u}_{N,l}\right]$ είναι ένα διάνυσμα με διάσταση ίση με τον αριθμό των μεταβλητών εισόδου N. Οι κρυφοί νευρώνες εκτελούν έναν μη γραμμικό μετασχηματισμό δημιουργώντας μια χαρτογράφηση του χώρου εισόδου σε έναν νέο χώρο υψηλότερων διαστάσεων. Πιο συγκεκριμένα, για κάθε ένα διάνυσμα εισόδου $\mathbf{u}(k)$, όπου k=1,2,…,K είναι τα διαθέσιμα δεδομένα του δικτύου, κάθε



Σχήμα 2.1. Δομή δικτύου RBF.



Σχήμα 2.2. Τυπική δομή νευρώνα *l* κρυφής στοιβάδας δικτύου RBF.

νευρώνας *l* προκαλεί μια διαφορετική τιμή ενεργότητας $\mu_l\left(\mathbf{u}(k)\right)$. Η κάθε ενεργότητα υπολογίζεται ως η ευκλείδεια απόσταση (norm) της διαφοράς μεταξύ του διανύσματος εισόδου και του αντίστοιχου κέντρου ως εξής:

$$\mu_l\left(\mathbf{u}(k)\right)=\left\|\mathbf{u}(k)-\hat{\mathbf{u}}_l\right\|=\sqrt{\sum_{i=1}^{N}\left(u_i(k)-\hat{u}_{i,l}\right)^2},\ k=1,\ldots,K \tag{2.1}$$

Στη συνέχεια, η ενεργότητα του νευρώνα δίνεται ως είσοδος σε μια συνάρτηση που ονομάζεται συνάρτηση ενεργοποίησης (activation function). Η συνάρτηση αυτή έχει ακτινική συμμετρία, από την οποία προκύπτει και η ονομασία των δικτύων RBF. Τυπικές επιλογές συναρτήσεων ενεργοποίησης ενός νευρώνα *l* είναι οι παρακάτω:

- **Συνάρτηση Gauss**:

$$g(\mu_l)=\exp\left(-\frac{\mu_l^2}{\sigma_l^2}\right) \tag{2.2}$$

όπου σ_l είναι το εύρος της συνάρτησης.

- **Συνάρτηση Thin Plate Spline**:

$$g(\mu_l)=\mu_l^2\log(\mu_l) \tag{2.3}$$

- **Τροποποιημένη συνάρτηση Thin Plate Spline**:

$$g(\mu_l)=\mu_l^2\log(\mu_l+1) \tag{2.4}$$

Η συνάρτηση Gauss (σχήμα 2.3) διαφέρει από τις άλλες δύο συναρτήσεις λόγω του μεταβλητού της εύρους. Η διαδικασία υπολογισμού του εύρους σ_l για κάθε νευρώνα l της κρυφής στοιβάδας, επιτυγχάνεται επιλέγοντας την τιμή σ_l έτσι ώστε να είναι μεγαλύτερη από την απόσταση ως τον αμέσως κοντινότερο νευρώνα. Η τιμή σ_l υπολογίζεται με βάση τον κανόνα p-nearest neighbor [36]:

$$\sigma_l = \sqrt{\frac{1}{P}\sum_{j=1}^{p} \| \hat{\mathbf{u}}_l - \mathbf{u}(j) \|^2} \tag{2.5}$$

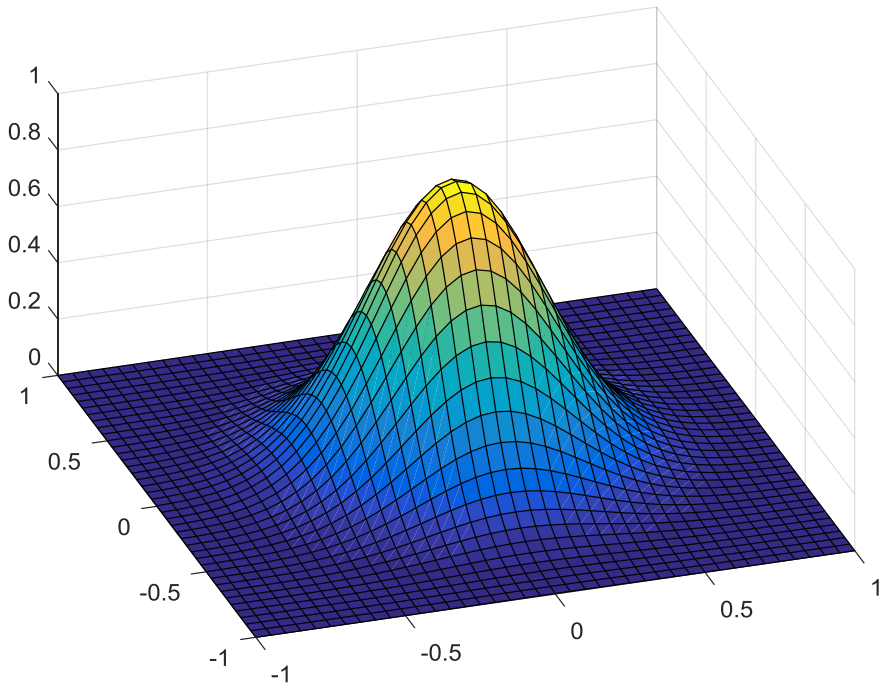
όπου $\mathbf{u}(j)$ είναι τα p κοντινότερα κέντρα στο κέντρο $\hat{\mathbf{u}}_l$.

Η συνάρτηση Thin Plate Spline (TPS) όπως φαίνεται στο σχήμα 2.4 δεν ορίζεται στο σημείο μηδέν. Αντιθέτως, η τροποποιημένη συνάρτηση TPS (σχήμα 2.5) ορίζεται στο σημείο μηδέν.

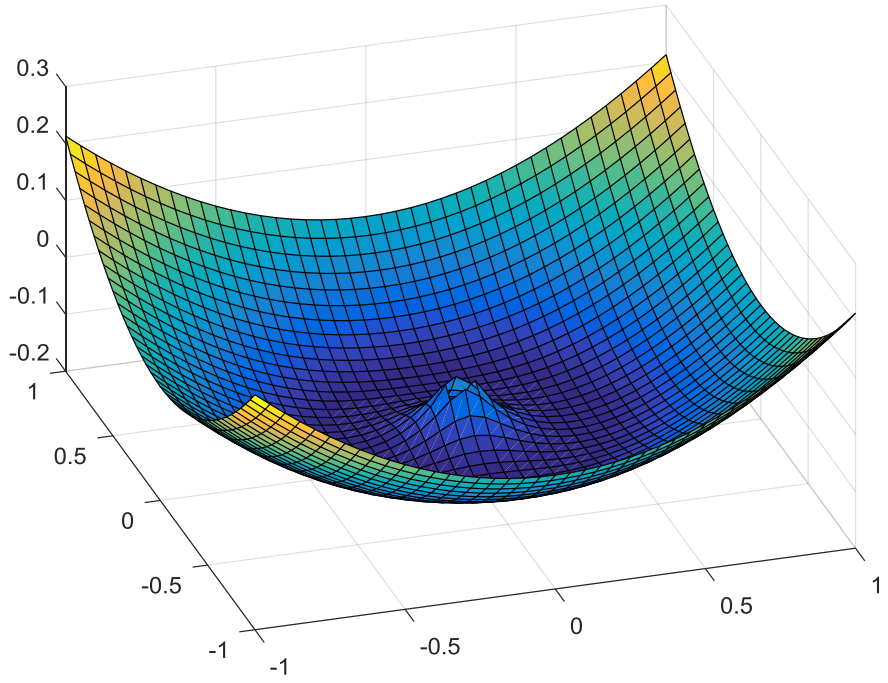
Εφόσον, η ενεργότητα του κάθε νευρώνα l=1,2,…,L περνά από τη συνάρτηση ενεργοποίησης, οι έξοδοι των νευρώνων της κρυφής στοιβάδας καταλήγουν να είναι οι παρακάτω:

$$\mathbf{z}(k) = \left[z_1(k), z_2(k), ..., z_L(k) \right] = \left[g\left(\mu_1(\mathbf{u}(k))\right), g\left(\mu_2(\mathbf{u}(k))\right), ..., g\left(\mu_L(\mathbf{u}(k))\right) \right]$$

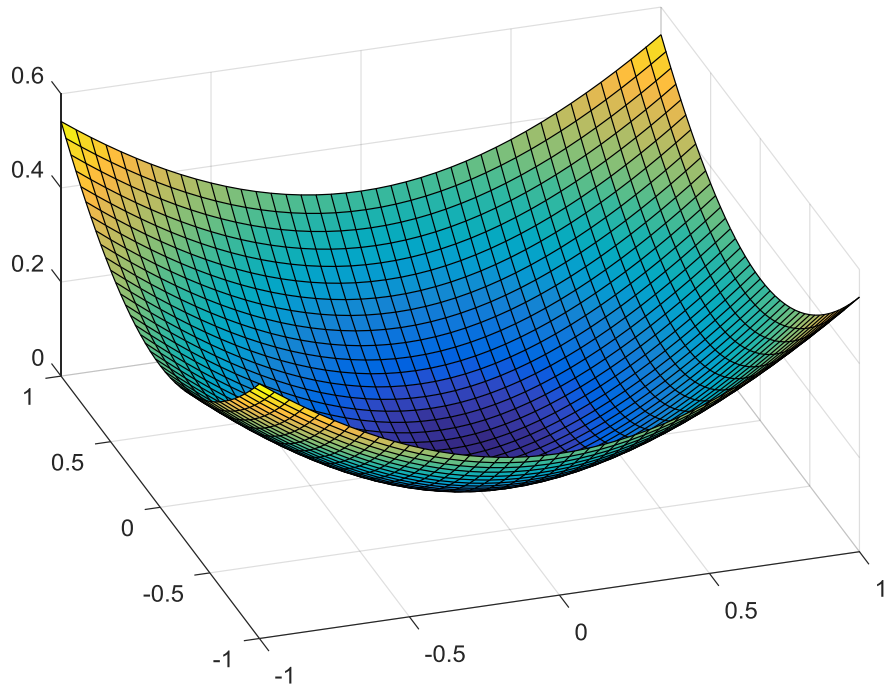
(2.6)



Σχήμα 2.3. Συνάρτηση Gauss.



Σχήμα 2.4. Συνάρτηση Thin Plate Spline.



Σχήμα 2.5. Τροποποιημένη συνάρτηση Thin Plate Spline.

όπου το σήμα αυτό πολλαπλασιάζεται με το αντίστοιχο συναπτικό βάρος w_l και καταλήγει στη στοιβάδα εξόδου.

Η τρίτη και η τελευταία στοιβάδα του δικτύου RBF είναι η στοιβάδα εξόδου. Αυτή αποτελείται από έναν ή περισσότερους νευρώνες, που σημαίνει ότι ένα δίκτυο RBF μπορεί να έχει μία ή περισσότερες εξόδους. Η στοιβάδα εξόδου περιέχει *M* νευρώνες, όπου επίσης, *M* είναι ο αριθμός των μεταβλητών εξόδου. Αναλόγως το πρόβλημα που πρέπει να αντιμετωπίσει το δίκτυο, δηλαδή προσέγγισης συνάρτησης ή ταξινόμησης, το *M* είναι ίσο με τη μονάδα ή ίσο με τον αριθμό των κλάσεων, αντίστοιχα. Σε κάθε νευρώνα εξόδου *m*, όπου $m = 1,2,\ldots,M$, αντιστοιχεί ένα διάνυσμα $\mathbf{w}_m$ που περιέχει τα σταθμισμένα σήματα της κρυφής στοιβάδας:

$$\mathbf{w}_m = \left[w_{1,m}, w_{2,m}, \ldots, w_{L,m} \right] \tag{2.7}$$

Οι νευρώνες της στοιβάδας εξόδου λειτουργούν μόνο ως μονάδες άθροισης, οι οποίες αθροίζουν τα σταθμισμένα σήματα της κρυφής στοιβάδας και παράγουν την τελική έξοδο του δικτύου. Για κάθε νευρώνα εξόδου η αριθμητική έξοδος $\hat{y}_m(k)$, παράγεται συνδυάζοντας γραμμικά τις αποκρίσεις των κρυφών νευρώνων όπως φαίνεται στην ακόλουθη εξίσωση:

$$\hat{y}_m(k) = \mathbf{z}(k) \cdot \mathbf{w}_m = \sum_{l=1}^{L} w_{l,m} g\left(\mu_l\left(\mathbf{u}(k)\right)\right) = \sum_{l=1}^{L} w_{l,m} g\left(\sqrt{\sum_{i=1}^{N}\left(u_i(k) - \hat{u}_{i,l}\right)^2}\right) \tag{2.8}$$

Όταν, όμως, το δίκτυο RBF λειτουργεί ως ταξινομητής, η πρόβλεψη για την κατηγορία εξόδου $\mathscr{C}(k)$ αντιστοιχεί στη μονάδα άθροισης που ενεργοποιεί τη μέγιστη αριθμητική έξοδο [2]:

$$\mathscr{C}(k) = \arg\max_m \hat{y}_m(k) \tag{2.9}$$

Τέλος, μια μαθηματική αιτιολόγηση για τη λογική του μη γραμμικού μετασχηματισμού που ακολουθείται από έναν γραμμικό μετασχηματισμό επισημαίνεται από τον Cover [184]. Σύμφωνα με τον Cover, ένα πρόβλημα αναγνώρισης προτύπων σε ένα χώρο πολλών διαστάσεων είναι πιθανότερο να είναι γραμμικά διαχωριζόμενο απ’ ότι σε ένα χώρο λίγων διαστάσεων και γι' αυτό το λόγο η κρυφή στοιβάδα σε ένα δίκτυο RBF είναι συνήθως πολλών διαστάσεων. Ένα άλλο σημαντικό σημείο είναι το γεγονός ότι η διάσταση του χώρου της κρυφής στοιβάδας συσχετίζεται άμεσα με την ικανότητα του δικτύου να προσεγγίσει μια ομαλή χαρτογράφηση εισόδου - εξόδου (όσο πιο πολλών διαστάσεων είναι ο χώρος της κρυφής στοιβάδας, τόσο ακριβέστερη είναι η προσέγγιση) [185, 186].

2.2 Εκπαίδευση δικτύων RBF και διαδικασία επιλογής μοντέλου

Τα δίκτυα RBF διαμορφώνουν μία ιδιαίτερη αρχιτεκτονική νευρωνικών δικτύων, η οποία παρουσιάζει δύο κύρια πλεονεκτήματα: Την απλότητα της δομής και την ταχύτητα των αλγορίθμων εκπαίδευσης που υιοθετεί. Το σημαντικότερο πρόβλημα στο σχεδιασμό ενός δικτύου RBF είναι η επιλογή των κέντρων, το οποίο θεωρείται ως ένα μη πολυωνυμικό πρόβλημα βελτιστοποίησης (non-polynomial) [187]. Προκειμένου να λυθεί το πρόβλημα

επιλογής των κατάλληλων νευρώνων της κρυφής στοιβάδας, προτάθηκε μία νέα κατηγορία αλγορίθμων, όπου η δομή του δικτύου καθορίζεται αυτόματα.
Οι περισσότερες από τις υπάρχουσες τεχνικές εκπαίδευσης νευρωνικών δικτύων βασίζονται σε αλγόριθμους μη γραμμικής βελτιστοποίησης και απαιτούν μεγάλους υπολογιστικούς χρόνους. Στις άλλες αρχιτεκτονικές νευρωνικών δικτύων η επίλυση του προβλήματος βελτιστοποίησης αντιστοιχεί σε μία προσπάθεια να υπολογιστούν όλες οι παράμετροι του δικτύου σε ένα βήμα. Η διαφορά στην ταχύτητα των διαδικασιών μάθησης μεταξύ των δικτύων RBF και των άλλων αρχιτεκτονικών οφείλεται στο ότι η εκπαίδευση στα δίκτυα RBF μπορεί να εκτελεστεί σε διακριτές φάσεις και συγκεκριμένα στα εξής δύο στάδια:

- Πρώτο στάδιο: Σε αυτό το στάδιο, καθορίζεται η δομή του δικτύου, δηλαδή υπολογίζεται ο αριθμός και οι θέσεις (συντεταγμένες) των κέντρων των νευρώνων της κρυφής στοιβάδας. Η πιο συνηθισμένη τεχνική μάθησης που προσεγγίζει το πρόβλημα σε δύο φάσεις χρησιμοποιεί τον αλγόριθμο *k*-means για να επιλέξει τα κέντρα της κρυφής στοιβάδας [9-11].
- Δεύτερο στάδιο: Αυτό το στάδιο είναι πιο εύκολο. Εδώ, υπολογίζονται οι υπόλοιπες παράμετροι του δικτύου. Συγκεκριμένα, χρησιμοποιείται η μέθοδος γραμμικής παλινδρόμησης (linear regression, LR) για να υπολογιστούν οι σταθμισμένες συνδέσεις μεταξύ της κρυφής στοιβάδας και της στοιβάδας εξόδου [5, 36].

Ένα ιδιαίτερα σημαντικό στάδιο που προηγείται της διαδικασίας εκπαίδευσης είναι αυτό της στάθμισης ή κανονικοποίησης (scaling) των δεδομένων εισόδου [188]. Όταν οι τάξεις μεγεθών των δεδομένων διαφέρουν πολύ μεταξύ τους μπορεί να δημιουργήσουν δυσλειτουργίες στη διαδικασία της εκπαίδευσης του δικτύου. Γι' αυτό τον λόγο, πρώτα σταθμίζονται τα δεδομένα, ύστερα ακολουθεί η εκπαίδευση χρησιμοποιώντας τα σταθμισμένα δεδομένα, και μετά οι προβλέψεις επανέρχονται στην αρχική τους μορφή (unscaling) ακολουθώντας την αντίστροφη πορεία της στάθμισης. Η στάθμιση των δεδομένων μπορεί να πραγματοποιηθεί με πολλούς τρόπους. Οι πιο συχνά χρησιμοποιούμενοι τρόποι είναι δύο και είναι οι εξής:

- **Μέθοδος της Γραμμικής Παρεμβολής (linear interpolation)**: Ορίζεται ένα διάστημα [α,β] όπου α είναι η ελάχιστη τιμή των μεταβλητών εισόδου και β είναι η μέγιστη τιμή. Ύστερα, όλα τα δεδομένα εισόδου – εξόδου ανάγονται στο παραπάνω διάστημα μέσω ενός γραμμικού μετασχηματισμού. Συνήθως, το διάστημα που χρησιμοποιείται στα νευρωνικά δίκτυα είναι το [-1,1].
- **Μέθοδος της στάθμισης με μηδενικό μέσο και μοναδιαία διασπορά (zero mean – unit variance)**: Για μία μεταβλητή βρίσκουμε τον μέσο όρο όλων των δεδομένων της και τον αφαιρούμε από την τιμή αυτής της μεταβλητής. Στη συνέχεια, για την ίδια μεταβλητή βρίσκουμε τη διασπορά όλων των δεδομένων της και διαιρούμε την τιμή αυτής της μεταβλητής με την τιμή της διασποράς της. Αυτή η διαδικασία εκτελείται για κάθε μεταβλητή με αποτέλεσμα, οι νέες μεταβλητές να έχουν μέσο όρο μηδέν και διασπορά ίση με τη μονάδα.

Μετά το στάδιο της στάθμισης των δεδομένων, το σύνολο των δεδομένων εισόδου-εξόδου που αποτελείται από N μεταβλητές εισόδου, M μεταβλητές εξόδου και K διαθέσιμα δεδομένα είναι έτοιμο για να περάσει στο νευρωνικό δίκτυο RBF και να ξεκινήσει η διαδικασία εκπαίδευσης. Η εκπαίδευση ενός δικτύου RBF αντιστοιχεί με την εύρεση βέλτιστων τιμών για τις ακόλουθες παραμέτρους του δικτύου:

1. Μέγεθος του δικτύου ή διάσταση της κρυφή στοιβάδας (αριθμός κέντρων L).
2. Τοποθεσίες των κέντρων των νευρώνων της κρυφής στοιβάδας $\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, ..., \hat{\mathbf{u}}_L$.
3. Εύρη των νευρώνων της κρυφής στοιβάδας $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, ..., \sigma_L]$. Αν η συνάρτηση ενεργοποίησης έχει μεταβλητό πλάτος, όπως η συνάρτηση Gauss, τότε θα πρέπει να βελτιστοποιείται και το εύρος των συναρτήσεων που σε άλλες περιπτώσεις (π.χ. TPS) θα είχε σταθερή τιμή.
4. Συναπτικά βάρη που ενώνουν τους νευρώνες της κρυφής στοιβάδας με τους νευρώνες της στοιβάδας εξόδου $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_M]$.

Σύμφωνα με τα παραπάνω, για ένα σταθερό αριθμό κέντρων L των νευρώνων της κρυφής στοιβάδας, η διαδικασία εκπαίδευσης του δικτύου RBF αντιστοιχεί στην επίλυση του παρακάτω προβλήματος βελτιστοποίησης, το οποίο για προβλήματα ελαχιστοποίησης εκφράζεται με τη σχέση (2.10), ενώ για προβλήματα μεγιστοποίησης εκφράζεται με τη σχέση (2.11):

$$\min_{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, ..., \hat{\mathbf{u}}_L, \boldsymbol{\sigma}, \mathbf{W}} f\left(\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, ..., \hat{\mathbf{u}}_L, \boldsymbol{\sigma}, \mathbf{W}, \mathbf{U}_{train}, \mathbf{Y}_{train}\right) \tag{2.10}$$

$$\max_{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, ..., \hat{\mathbf{u}}_L, \boldsymbol{\sigma}, \mathbf{W}} f\left(\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, ..., \hat{\mathbf{u}}_L, \boldsymbol{\sigma}, \mathbf{W}, \mathbf{U}_{train}, \mathbf{Y}_{train}\right) \tag{2.11}$$

όπου f είναι η αντικειμενική συνάρτηση, η οποία παρέχει ένα μέτρο για την απόκλιση μεταξύ των προβλέψεων του δικτύου και των πραγματικών τιμών, $\mathbf{U}_{train}$ και $\mathbf{Y}_{train}$ είναι πίνακες που περιέχουν τα K σταθμισμένα διαθέσιμα δεδομένα εκπαίδευσης εισόδου και εξόδου, αντίστοιχα, όπου το διάνυσμα των δεδομένων εισόδου είναι το παρακάτω:

$$\mathbf{U}_{train} = \begin{bmatrix} \mathbf{u}(1) \\ \mathbf{u}(2) \\ \vdots \\ \mathbf{u}(K) \end{bmatrix} = \begin{bmatrix} u_1(1) & u_1(2) & \cdots & u_1(K) \\ u_2(1) & u_2(2) & \cdots & u_2(K) \\ \vdots & \vdots & \ddots & \vdots \\ u_N(1) & u_N(2) & \cdots & u_N(K) \end{bmatrix} \tag{2.12}$$

Η διαδικασία ελαχιστοποίησης ή μεγιστοποίησης θα μπορούσε να πραγματοποιηθεί χρησιμοποιώντας κάποιον αλγόριθμο μη γραμμικής βελτιστοποίησης. Αυτό, όμως, θα καθυστερούσε αρκετά τη διαδικασία της εκπαίδευσης και θα μεγάλωνε το υπολογιστικό φορτίο, λόγω του ότι τέτοιου είδους αλγόριθμοι συνήθως παγιδεύονται σε τοπικά ελάχιστα.

Με βάση τα παραπάνω, ο συνολικός αριθμός των βαθμωτών (scalar) μεταβλητών σχεδιασμού ισούται με $L(N+M+1)$. Ο αριθμός αυτός μπορεί εύκολα να γίνει πολύ μεγάλος. Για παράδειγμα, για ένα μικρού μεγέθους πρόβλημα εκπαίδευσης, με πέντε μεταβλητές εισόδου, μία μεταβλητή εξόδου και 15 κρυφούς νευρώνες, οι συνολικές βαθμωτές μεταβλητές σχεδιασμού υπερβαίνουν ήδη τις 100, ενώ ακόμα και για τα μεσαίου μεγέθους προβλήματα, ο αριθμός των βαθμωτών μεταβλητών σχεδιασμού μπορεί εύκολα να αυξηθεί σε δεκάδες χιλιάδες. Έχοντας κατά νου ότι η αντικειμενική συνάρτηση παρουσιάζει μεγάλο αριθμό τοπικών ελαχίστων, είναι εύκολα κατανοητό ότι στις περισσότερες περιπτώσεις ο εντοπισμός του ολικού βέλτιστου είναι αρκετά απίθανος. Στην πραγματικότητα, ακόμη και η εύρεση μιας ικανοποιητικής υπο-βέλτιστης λύσης μπορεί να είναι δύσκολη.

Η μεγάλη πολυπλοκότητα του έργου της βελτιστοποίησης μπορεί να αντιμετωπιστεί με την ανάλυση του προβλήματος εκπαίδευσης του νευρωνικού δικτύου RBF σε δύο ξεχωριστά στάδια (όπως αναφέρθηκε στην αρχή του κεφαλαίου). Ως προς το δεύτερο στάδιο, που αποτελεί και το πιο εύκολο από τα δύο, ο στόχος είναι η εύρεση ενός συνόλου συναπτικών βαρών που να ελαχιστοποιεί το τετράγωνο της νόρμας των υπολοίπων:

$$E(\mathbf{W}) = \frac{1}{2}\sum_{k=1}^{K} \| \mathbf{y}(k) - \hat{\mathbf{y}}(k) \|^2 = \frac{1}{2}\sum_{k=1}^{K} \| \mathbf{y}(k) - \mathbf{W}\cdot\mathbf{z}(k) \|^2 \tag{2.13}$$

Εκμεταλλευόμενοι τη γραμμική σχέση που υπάρχει μεταξύ της κρυφής στοιβάδας και της στοιβάδας εξόδου, το E ελαχιστοποιείται επιλύοντας τη μέθοδο της γραμμικής παλινδρόμησης (LR) σε μορφή πινάκων, δηλαδή τα συναπτικά βάρη υπολογίζονται με βάση την ακόλουθη εξίσωση ελαχίστων τετραγώνων:

$$\mathbf{W}^T = \mathbf{Y}_{train}{}^T \cdot \mathbf{Z} \cdot \left(\mathbf{Z}^T \cdot \mathbf{Z}\right)^{-1} \tag{2.14}$$

όπου **Z**, **W**, $\mathbf{Y}_{train}$, είναι πίνακες που περιέχουν τις εξόδους της κρυφής στοιβάδας, τα συναπτικά βάρη, τις επιθυμητές εξόδους του δικτύου, και περιγράφονται με τις εξισώσεις (2.16), (2.15), (2.17), αντίστοιχα:

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_M \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,M} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ w_{L,1} & w_{L,2} & \cdots & w_{L,M} \end{bmatrix} \tag{2.15}$$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}(1) \\ \mathbf{z}(2) \\ \vdots \\ \mathbf{z}(K) \end{bmatrix} = \begin{bmatrix} z_1(1) & z_2(1) & \cdots & z_L(1) \\ z_1(2) & z_2(2) & \cdots & z_L(2) \\ \vdots & \vdots & \ddots & \vdots \\ z_1(K) & z_2(K) & \cdots & z_L(K) \end{bmatrix} \tag{2.16}$$

$$\mathbf{Y}_{train} = \left[\mathbf{y}(1),\ \mathbf{y}(2),\ \ldots,\ \mathbf{y}(K)\right] \tag{2.17}$$

Σε αυτό το σημείο πρέπει να αναφερθεί ότι αν το δίκτυο RBF λειτουργεί ως ταξινομητής, κάθε διάνυσμα πραγματικών τιμών $\mathbf{y}(k)$ παρέχει την πληροφορία της κλάσης στην οποία ανήκει πραγματικά το k-οστό διάνυσμα εισόδου. Όταν ένα διάνυσμα παρατήρησης $\mathbf{y}(k)$ ανήκει στην κλάση εξόδου m, τότε όλα τα στοιχεία του ισούνται με μηδέν, εκτός από το m-οστό στοιχείο, το οποίο τίθεται ίσο με τη μονάδα, αυτού του είδους η κωδικοποίηση ονομάζεται “1 of C” και περιγράφεται με την παρακάτω σχέση:

$$y_i(k) = \begin{cases} 1, & \text{αν } i = m \\ 0, & \text{αν } i \neq m \end{cases}, \quad i = 1, 2, .., M \tag{2.18}$$

Ο υπολογισμός των συναπτικών βαρών εφαρμόζοντας τη μέθοδο της γραμμικής παλινδρόμησης αποτελεί ένα σημαντικό πλεονέκτημα, διότι έχοντας δεδομένες τιμές για τα κέντρα και τα εύρη, το ολικό βέλτιστο των βαρών είναι εγγυημένο. Παρά την ύπαρξη αυτού του πλεονεκτήματος, ο υπολογισμός των υπολοίπων παραμέτρων του δικτύου παραμένει ένα δύσκολο έργο, χωρίς εγγυημένη βέλτιστη λύση.

Τυπικά η διαδικασία της εκπαίδευσης ακολουθείται από την αξιολόγηση των μοντέλων που εκπαιδευτήκαν και τελικά από την επιλογή του καταλληλότερου μοντέλου, δηλαδή την επιλογή του αριθμού των κέντρων RBF. Σύμφωνα με τις συναρτήσεις (2.10) ή (2.11), ένας μεγάλος αριθμός κέντρων RBF μπορεί να παράγει βελτιωμένες λύσεις, ωστόσο αυτό θα μπορούσε να οδηγήσει στο φαινόμενο της υπερπροσαρμογής του δικτύου το οποίο αναλύθηκε στο κεφάλαιο 1.2.7. Πιο απλά, ένας μεγάλος αριθμός κέντρων θα μπορούσε να οδηγήσει το μοντέλο να μην αποδίδει ικανοποιητικά όταν του δίνονται νέα δεδομένα που δεν έχουν χρησιμοποιηθεί κατά τη διαδικασία της εκπαίδευσης.

Επίσης, σύμφωνα με το κεφάλαιο 1.2.7, η αξιολόγηση του δικτύου RBF αφορά στην ικανότητα γενίκευσης του μοντέλου, δηλαδή στην ακρίβεια πρόβλεψης του μοντέλου όταν αυτό εφαρμόζεται σε ένα νέο σύνολο δεδομένων εισόδου-εξόδου $\{\mathbf{U}_{val}, \mathbf{Y}_{val}\}$. Τελικά, το μοντέλο RBF που επιλέγεται είναι αυτό που αποδίδει καλύτερα κατά τη διαδικασία της αξιολόγησης. Βέβαια, η επιλογή του αριθμού των κέντρων του δικτύου διαμέσου αυτού του επιπλέον σταδίου αυξάνει περαιτέρω το μέγεθος του χώρου αναζήτησης.

2.2.1 Ο αλγόριθμος *k*-means

Η τεχνική *k*-means είναι η πιο ευρέως γνωστή τεχνική μάθησης που χρησιμοποιείται για να βρεθούν οι συντεταγμένες των κέντρων ενός δικτύου RBF. Ο αλγόριθμος *k*-means δεν παρέχει κάποιο τρόπο για την επιλογή του κατάλληλου αριθμού κέντρων των νευρώνων της κρυφής στοιβάδας και γι' αυτό η κατάλληλη δομή δικτύου επιλέγεται με τη μέθοδο δοκιμή – σφάλμα.

Ο αλγόριθμος *k*-means έχει ως σκοπό του την εύρεση κατάλληλων κέντρων, ώστε να ελαχιστοποιηθεί το *E* που είναι το άθροισμα των ευκλείδειων αποστάσεων μεταξύ του κάθε κέντρου (cluster center) και των δεδομένων εκπαίδευσης που ανήκουν στην ίδια συστάδα:

$$E = \sum_{l=1}^{L} \sum_{k=1}^{K} B_{lk} \, \| \hat{u}_l - u(k) \|^2 \tag{2.19}$$

όπου B_{lk} είναι ένας *L*x*K* πίνακας του οποίου κάθε γραμμή αντιπροσωπεύει ένα κέντρο και κάθε στήλη ένα δεδομένο εκπαίδευσης. Το κάθε δεδομένο αντιστοιχεί μόνο σε ένα κέντρο και αυτό είναι το κοντινότερό του. Το κάθε δεδομένο έχει ένα «1» στο τετράγωνο που διασταυρώνονται η στήλη του συγκεκριμένου δεδομένου και η γραμμή του κοντινότερου κέντρου σε αυτό. Δηλαδή, η στήλη που αντιστοιχεί στο συγκεκριμένο δεδομένο έχει ένα «0» σε όλα τα υπόλοιπα τετράγωνά της [9, 36].

Ο αλγόριθμος *k*-means ξεκινά επιλέγοντας *L* τυχαία δεδομένα και τοποθετεί επάνω τους κέντρα. Στη συνέχεια, κάθε δεδομένο εκμάθησης εκπροσωπείται από μόνο ένα κέντρο. Το σε πιο κέντρο θα ανήκει το κάθε κέντρο επιλέγεται με βάση την ευκλείδεια απόσταση. Δηλαδή, για κάθε ένα δεδομένο υπολογίζεται το κοντινότερο κέντρο $\hat{\mathbf{u}}_{near}$ που του

αντιστοιχεί. Ύστερα, η θέση του κάθε κέντρου $\hat{\mathbf{u}}_{near}$ τροποποιείται λαμβάνοντας υπόψη μόνο τα δεδομένα που εκπροσωπεί (τα άλλα δεδομένα δεν επηρεάζουν τη θέση του) με βάση την παρακάτω σχέση:

$$\Delta\hat{\mathbf{u}}_{near} = \eta \cdot \left(\mathbf{u}(k) - \hat{\mathbf{u}}_{near}\right) \tag{2.20}$$

όπου η είναι ο ρυθμός μάθησης, ο οποίος καθορίζει το πόσο γρήγορα συγκλίνει ο αλγόριθμος [9]. Οι κυριότεροι ρυθμοί μάθησης είναι οι εξής:

- **Σταθερός ρυθμός μάθησης** :

$$\eta = \text{σταθερά} \tag{2.21}$$

Στην περίπτωση αυτή, η σταθερή τιμή του ρυθμού μάθησης υπολογίζεται συνήθως με τη μέθοδο δοκιμή – σφάλμα.

- **Ρυθμός μάθησης MacQueen** :

$$\eta = \frac{1}{N_{near}} \tag{2.22}$$

όπου N_{near} είναι ο συνολικός αριθμός δεδομένων εκπαίδευσης που αντιστοιχούν στο $\hat{\mathbf{u}}_{near}$ κέντρο.

- **Ρυθμός μάθησης "τετραγωνικής ρίζας"** :

$$\eta = \frac{1}{\sqrt{N_{near}}} \tag{2.23}$$

Ο αλγόριθμος επαναλαμβάνεται αρκετές φόρες μέχρι οι τοποθεσίες των κέντρων να συγκλίνουν. Εφόσον έχουν καθοριστεί οι τοποθεσίες των κέντρων, και με την προϋπόθεση ότι η συνάρτηση ενεργοποίησης είναι η Gauss, ακολουθεί η διαδικασία υπολογισμού του εύρους για κάθε νευρώνα l της κρυφής στοιβάδας έτσι ώστε να ενεργοποιούνται παραπάνω από ένα κέντρα για κάθε παράδειγμα εκπαίδευσης.

Με την ολοκλήρωση της πρώτης φάσης του προβλήματος στη συνέχεια, σε δεύτερη φάση, υπολογίζονται τα συναπτικά βάρη με τη μέθοδο των ελαχίστων τετραγώνων (εξίσωση (2.14)). Έτσι, ολοκληρώνεται η εκπαίδευση των δικτύων RBF. Από τα παραπάνω συμπεραίνουμε ότι ο αλγόριθμος *k*-means, παρά τα πλεονεκτήματα του, παρουσιάζει τρία σημαντικά μειονεκτήματα:

- Απαιτούνται πολλές δοκιμές όλων των δεδομένων εκπαίδευσης. Αυτή η επαναληπτική διαδικασία αυξάνει τον υπολογιστικό χρόνο, ειδικότερα μάλιστα όταν ο αριθμός των δεδομένων εκμάθησης είναι πολύ μεγάλος.
- Ο αριθμός των νευρώνων της κρυφής στοιβάδας πρέπει να προκαθοριστεί από τον χρήστη. Αυτό έχει ως επακόλουθο την αναγκαιότητα χρήσης της διαδικασία δοκιμής –

σφάλματος προκειμένου να ληφθεί ο βέλτιστος αριθμός των νευρώνων της κρυφής στοιβάδας.

- Εξαρτάται από μια τυχαία αρχική επιλογή της θέσης του κέντρου, δηλαδή ο αλγόριθμος δεν έχει επαναληψιμότητα. Έτσι, η ίδια δομή δικτύου για διαφορετικά τρεξίματα δίνει διαφορετικά σύνολα κέντρων.

Προκειμένου να ξεπεραστούν τα παραπάνω προβλήματα, οι Sarimveis et.al. πρότειναν τον αλγόριθμο εκμάθησης των ασαφών μέσων [20], ο οποίος παρουσιάζεται στη συνέχεια.

2.2.2 Η έννοια της ασαφούς διαμέρισης - Ο συμμετρικός αλγόριθμος των ασαφών μέσων

Ο αλγόριθμος των ασαφών μέσων (fuzzy means, FM) ή συμμετρικός αλγόριθμος των ασαφών μέσων (symmetric FM, SFM) είχε προταθεί πριν από μια δεκαετία περίπου, για την εξάλειψη των μειονεκτημάτων του αλγορίθμου *k*-means, όσον αφορά στην επιλογή των κόμβων της κρυφής στοιβάδας του δικτύου RBF [20, 34]. Ο αλγόριθμος των ασαφών μέσων χρησιμοποιεί μία διαφορετική τεχνική επιλογής κέντρων, η οποία βασίζεται στην ασαφή λογική. Πιο συγκεκριμένα, η μέθοδος αυτή κάνοντας μόνο ένα πέρασμα από τα δεδομένα εκπαίδευσης, επιλέγει τη δομή (αριθμό των νευρώνων) και τις παραμέτρους (κέντρα των νευρώνων) του δικτύου με βάση μία ασαφή διαμέριση του χώρου των μεταβλητών εισόδου. Επιπλέον, για μία συγκεκριμένη ασαφή διαμέριση του χώρου των μεταβλητών εισόδου, ο αλγόριθμος παράγει πάντα τον ίδιο αριθμό κέντρων, αφού τα κέντρα δεν επιλέγονται τυχαία. Ο αλγόριθμος των ασαφών μέσων έχει χρησιμοποιηθεί με επιτυχία σε μια σειρά εφαρμογών [24-33].

Η βάση όλων των ασαφών συστημάτων είναι η διαμέριση του χώρου των μεταβλητών εισόδου σε έναν αριθμό ασαφών συνόλων. Έστω ότι ένα σύστημα έχει N κανονικοποιημένες μεταβλητές εισόδου u_i, όπου i=1,2,…,N. Στην περίπτωση της συμμετρικής διαμέρισης, το πεδίο ορισμού της κάθε μεταβλητής εισόδου διαμερίζεται σε έναν αριθμό ίσων μονοδιάστατων τριγωνικών ασαφών συνόλων, s. Κάθε ασαφές σύνολο μπορεί να περιγραφεί πλήρως από το αντίστοιχο κέντρο και εύρος του:

$$A_{i,j} = \left\{ a_{i,j}, \delta\alpha \right\},\ i = 1,\ldots,N,\ j = 1,\ldots,s \tag{2.24}$$

Όπου $\alpha_{i,j}$ είναι το κέντρο του ασαφούς συνόλου $A_{i,j}$ και $\delta\alpha$ είναι το μισό του αντίστοιχου εύρους (από τη στιγμή που έχουμε θεωρήσει ισομερή ασαφή διαμέριση σε κάθε διάσταση, τα εύρη όλων των αντίστοιχων ασαφών συνόλων είναι ίσα). Αυτή η τεχνική ασαφούς διαμέρισης του συνολικού χώρου δημιουργεί ένα σύνολο s^N πολυδιάστατων ασαφών υπόχωρων $\mathbf{A}^l$, όπου l=1,2,…, s^N. Κάθε πολυδιάστατος ασαφής υπόχωρος ορίζεται ως συνδυασμός N μονοδιάστατων ασαφών συνόλων, ένα για κάθε διάνυσμα εισόδου. Κάθε ασαφής υπόχωρος $\mathbf{A}^l$ μπορεί να περιγραφεί πλήρως από το αντίστοιχο κέντρο $\boldsymbol{\alpha}^l$ και εύρος του $\delta\boldsymbol{\alpha}$:

$$\mathbf{A}^l = \left\{ \boldsymbol{\alpha}^l, \delta\boldsymbol{\alpha} \right\} = \left\{ \left[a^l_{1,j_1}, a^l_{2,j_2}, \ldots, a^l_{N,j_N} \right], \underbrace{\left[\delta\alpha\ , \delta\alpha, \ldots, \delta\alpha \right]}_{N} \right\},\ l = 1,\ldots,s^N \tag{2.25}$$

όπου a_{i,j_i}^l είναι το κέντρο του μονοδιάστατου ασαφούς συνόλου A_{i,j_i} που ανήκει στην είσοδο i. Καθένας από τους παραχθέντες ασαφείς υπόχωρους είναι υποψήφιος για να γίνει κέντρο του δικτύου RBF. Ωστόσο, το τελικό σύνολο των κέντρων θα το συγκροτούν μόνο οι επιλεγμένοι ασαφείς υπόχωροι. Η επιλογή είναι βασισμένη στην ιδέα της πολυδιάστατης συνάρτησης συμμετοχής (multidimensional membership function) $\mu_{\mathbf{A}^l}\left(\mathbf{u}(k)\right)$ ενός διανύσματος εισόδου $\mathbf{u}(k)$ σε ένα ασαφή υπόχωρο $\mathbf{A}^l$, που δίνεται από τον Nie [189]:

$$\mu_{\mathbf{A}^l}\left(\mathbf{u}(k)\right)=\begin{cases}1-d_r^l\left(\mathbf{u}(k)\right)\ , & \text{αν } d_r^l\left(\mathbf{u}(k)\right)\le 1\\ 0 & ,\ \text{διαφορετικά}\end{cases} \tag{2.26}$$

όπου με $d_r^l\left(\mathbf{u}(k)\right)$ συμβολίζεται η σχετική ευκλείδεια απόσταση μεταξύ του $\mathbf{A}^l$ και του διανύσματος εισόδου $\mathbf{u}(k)$:

$$d_r^l\left(\mathbf{u}(k)\right)=\sqrt{\sum_{i=1}^{N}\left(a_{i,j_i}^l-u_i(k)\right)^2}\Big/\delta a\sqrt{N} \tag{2.27}$$

Η παραπάνω εξίσωση (2.27) καθορίζει μια υπερ-σφαίρα στον χώρο των μεταβλητών εισόδου με ακτίνα ίση με $\sqrt{N}\delta a$. Ο αντικειμενικός στόχος του αλγόριθμου εκπαίδευσης είναι να επιλέξει ένα υποσύνολο των ασαφών υπόχωρων ως κέντρα RBF έτσι ώστε όλα τα στοιχεία εκπαίδευσης να καλύπτονται από μία τουλάχιστον υπερ-σφαίρα. Η παραπάνω ιδέα λαμβάνοντας υπόψη την εξίσωση (2.26) προδιαγράφει ότι το υποσύνολο των ασαφών υπόχωρων επιλέγεται έτσι ώστε να υπάρχει τουλάχιστον ένας ασαφής υπόχωρος που να ορίζει ένα μη μηδενικό πολυδιάστατο βαθμό σε κάθε διάνυσμα εισόδου εκμάθησης.

Ένα παράδειγμα αυτού του είδους ασαφούς διαμέρισης φαίνεται στο σχήμα 2.6, όπου για λόγους απεικόνισης σχεδιάζεται ένας δισδιάστατος χώρος των μεταβλητών εισόδου. Τα πεδία ορισμού των δύο μεταβλητών εισόδου u_1, u_2 διαμερίζονται σε 5 ασαφή σύνολα, δημιουργώντας με αυτόν τον τρόπο 16 ορθογώνια. Προφανώς αυτή η συμμετρική διαδικασία διαμέρισης οδηγεί σε ένα τετραγωνικό πλέγμα, όπου η σχετική ευκλείδεια απόσταση καθορίζει έναν κύκλο με ακτίνα ίση με $\sqrt{2}\delta a$. Η χρήση περισσότερων ασαφών συνόλων για κάθε μεταβλητή εισόδου καθιστά το πλέγμα πυκνότερο και οδηγεί στην επιλογή περισσότερων κέντρων RBF.

Η κατανομή των υποψηφίων κέντρων κατά μήκος της κατεύθυνσης της κάθε μεταβλητής εισόδου είναι η ίδια, εφόσον χρησιμοποιείται ο ίδιος αριθμός ασαφών συνόλων για όλες τις μεταβλητές. Εντούτοις, αυτό περιορίζει την ευελιξία του αλγόριθμου, δεδομένου ότι μια διαφορετική διαμέριση για κάθε μεταβλητή εισόδου μπορεί να οδηγήσει σε ένα καλύτερο δίκτυο υπό την έννοια της ακρίβειας ή/και της πολυπλοκότητας του μοντέλου.

Ο αλγόριθμος των ασαφών μέσων περιγράφεται αναλυτικά από τους Sarimveis et al. [20]. Παρακάτω παρουσιάζεται μια επισκόπηση αυτού του αλγορίθμου. Τέλος, στο επόμενο υποκεφάλαιο, παρουσιάζονται οι απαραίτητες τροποποιήσεις προκειμένου να επεκταθεί η

δυνατότητα εφαρμογής του αλγόριθμου των ασαφών μέσων σε μία μη συμμετρική διαμέριση του χώρου των μεταβλητών εισόδου.

Αλγόριθμος 1 – Αλγόριθμος FM

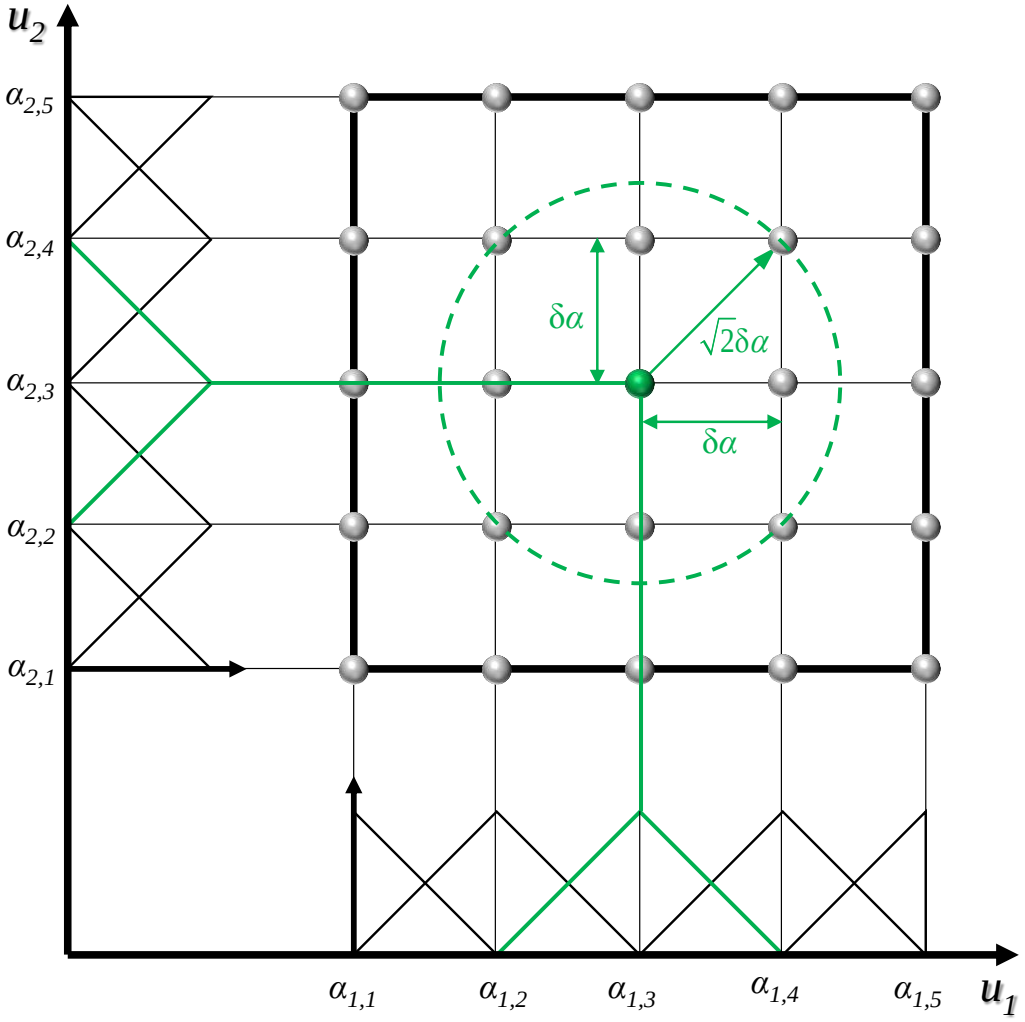
Δεδομένα Εισόδου: $\{\mathbf{U}_{train}, \mathbf{Y}_{train}\}$: Σύνολο δεδομένων εκπαίδευσης,

s: Αριθμός των ασαφών συνόλων που διαμερίζεται το πεδίο ορισμού κάθε μεταβλητής εισόδου

Δεδομένα Εξόδου: L : Αριθμός επιλεγμένων κέντρων RBF,

$\hat{\mathbf{U}} = [\hat{\mathbf{u}}_1,\ \hat{\mathbf{u}}_2,\ \ldots,\ \hat{\mathbf{u}}_L]^T$: Τοποθεσίες επιλεγμένων κέντρων RBF

1. Πάρε το πρώτο σημείο δεδομένων: $k \leftarrow 1$
2. Ξεκίνα τους υπολογισμούς για το πρώτο κέντρο RBF: $L \leftarrow 1$
3. **Για** i=1:N **Εκτέλεσε**:
4. Υπολόγισε το ασαφές σύνολο με τη μέγιστη δυνατή συμμετοχή σε κάθε διάσταση i:
   $A_i^1 = \{a_i^1, \delta a\} \leftarrow \max_{1 \le j \le s_i} \left[\mu_{A_{i,j}} \left(u_i(1) \right) \right]$
5. **Τέλος Για**
6. Δημιούργησε το πρώτο κέντρο RBF $\hat{\mathbf{u}}_1$: $\hat{\mathbf{u}}_1 = [a_1^1, a_2^1, \ldots, a_N^1]$
7. **Για** k=2:K **Εκτέλεσε**:
8. **Αν** το σημείο k βρίσκεται εκτός των υπέρ-ελλείψεων που ορίζονται από τα ήδη επιλεγμένα κέντρα: $\min_{1 \le l \le L} \left[d_r^l \left(\mathbf{u}(k) \right) \right] > 1$, όπου d_r^l υπολογίζεται από τον τύπο (2.27)
9. Πρόσθεσε ένα νέο κέντρο RBF: $L \leftarrow L+1$
10. **Για** i=1:N **Εκτέλεσε**:
11. Υπολόγισε το ασαφές σύνολο με τη μέγιστη δυνατή συμμετοχή σε κάθε διάσταση i: $A_i^L = \{a_i^L, \delta a\} \leftarrow \max_{1 \le j \le s_i} \left[\mu_{A_{i,j}} \left(u_i(k) \right) \right]$
12. **Τέλος Για**
13. Δημιούργησε το L-οστό κέντρο RBF $\hat{\mathbf{u}}_L$: $\hat{\mathbf{u}}_L = [a_1^L, a_2^L, \ldots, a_N^L]$
14. **Τέλος Αν**
15. **Τέλος Για**



Σχήμα 2.6. Δισδιάστατο παράδειγμα: Συμμετρική ασαφής διαμέριση σε 5 ασαφή σύνολα και για τις δύο μεταβλητές εισόδου.

2.2.3 Μη συμμμετρική διαμέριση – Ο μη συμμετρικός αλγόριθμος των ασαφών μέσων

Για τη περίπτωση της μη συμμετρικής ασαφούς διαμέρισης, προτάθηκε ο μη συμμετρικός αλγόριθμος των ασαφών μέσων (non symmetric fuzzy means, NSFM). Στη μη συμμετρική ασαφή διαμέριση, το πεδίο ορισμού κάθε μεταβλητής εισόδου διαμερίζεται σε s_i ασαφή σύνολα όπου i=1,2,…,N, γεγονός που οδηγεί σε S ασαφείς υπόχωρους:

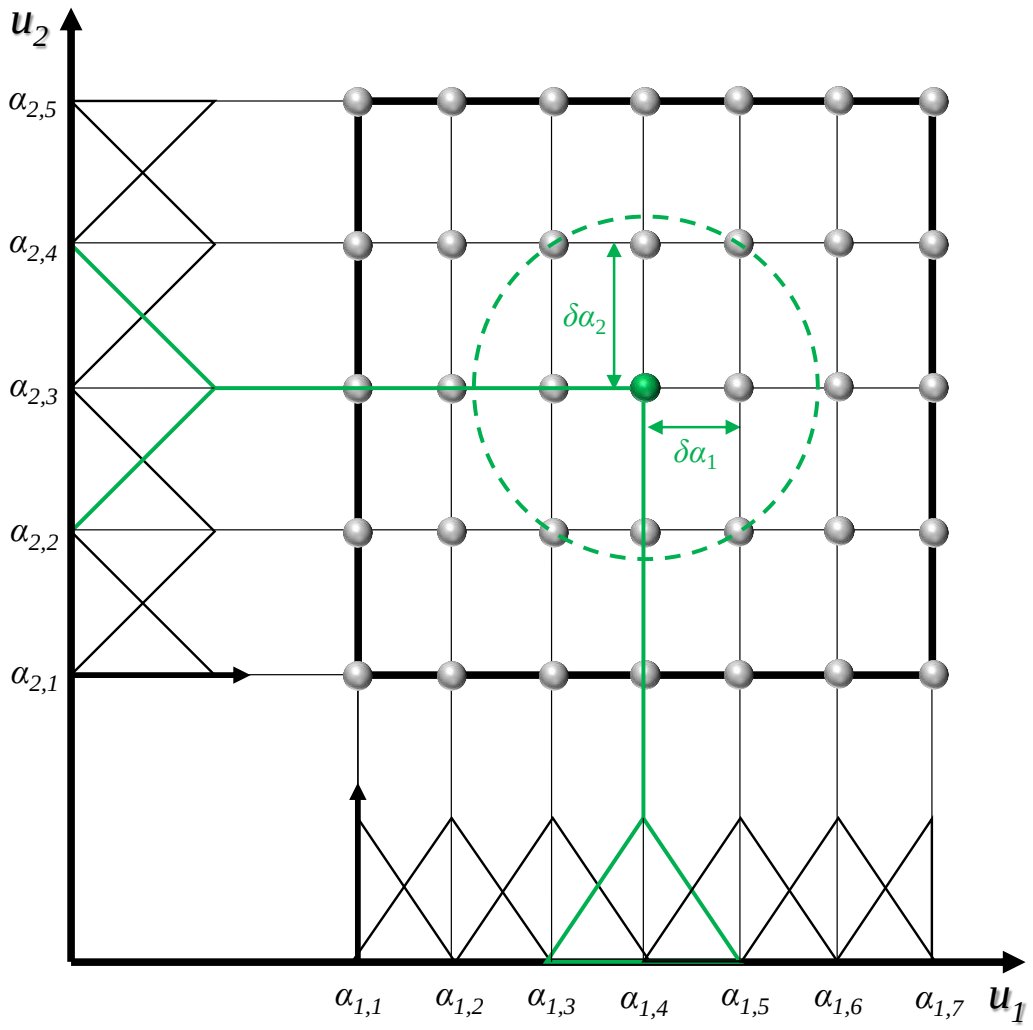
$$S = \prod_{i=1}^{N} s_i \ , \ i = 1, \ldots, N \tag{2.28}$$

Τα εύρη των ασαφών συνόλων είναι διαφορετικά για κάθε μεταβλητή εισόδου και οι ασαφείς υπόχωροι περιγράφονται ως εξής:

$$\mathbf{A}^l = \left\{\boldsymbol{\alpha}^l, \delta\boldsymbol{\alpha}\right\} = \left\{[\,a^l_{1,j_1},\ a^l_{2,j_2},\ \ldots,\ a^l_{N,j_N}\,],[\,\delta\alpha_1,\ \delta\alpha_2,\ \ldots,\ \delta\alpha_N\,]\right\}\ ,\ l = 1,\ldots,S \qquad (2.29)$$

Ένα δισδιάστατο παράδειγμα μη συμμετρικής διαμέρισης φαίνεται στο σχήμα 2.7. Εδώ, το πεδίο ορισμού της μεταβλητής εισόδου u_1 διαμερίζεται σε 7 ασαφή σύνολα ενώ το πεδίο ορισμού της u_2 διαμερίζεται σε 5 ασαφή σύνολα. Το πλέγμα ορίζει 24 ίσα ορθογώνια στο χώρο των μεταβλητών εισόδου και οι δύο πλευρές κάθε ορθογωνίου είναι δύο φορές τα εύρη των αντίστοιχων ασαφών συνόλων $\delta\alpha_1$,$\delta\alpha_2$.

Οι σχέσεις της πολυδιάστατης συνάρτησης συμμετοχής και της σχετικής ευκλείδειας απόστασης που περιγράφονται από τις εξισώσεις (2.26) και (2.27) αντίστοιχα, χρησιμοποιούνται μόνο στην περίπτωση της συμμετρικής διαμέρισης. Κατά συνέπεια, παρακάτω εξετάζεται πως αυτές οι εξισώσεις μπορούν να τροποποιηθούν έτσι ώστε να περιγράφουν και τη μη συμμετρική διαμέριση. Πιο συγκεκριμένα, μία N –διαστάσεων υπερ-έλλειψη συνδέεται με τον κάθε ασαφή υπόχωρο:



Σχήμα 2.7. Δισδιάστατο παράδειγμα: Μη συμμετρική ασαφής διαμέριση σε 7 ασαφή σύνολα για τη μεταβλητή εισόδου u_1 και σε 5 ασαφή σύνολα για τη u_2.

$$\sum_{i=1}^{N}\frac{\left(a_{i,j_i}^{l}-u_i\left(k\right)\right)^2}{\left(b_i\right)^2}=1\ ,\ l=1,\ldots,S \tag{2.30}$$

Όπου b_i είναι το μισό του αντίστοιχου άξονα της έλλειψης για κάθε μεταβλητή i=1,2,…,N. Σε αντίθεση με την υπερ-σφαίρα για την οποία όλες οι παράμετροι b_i είναι ίσες με τη μονάδα, εδώ οι παράμετροι b_i πρέπει να καθοριστούν. Μια βασική προϋπόθεση είναι ότι η επιφάνεια της υπερ-έλλειψης πρέπει να περάσει από όλες τις γωνίες του υπερ-ορθογωνίου που καθορίζεται από τον ασαφή υπόχωρο, π.χ. στο δισδιάστατο παράδειγμα πρέπει να διασχίσει τις τέσσερις γωνίες του υπερ-ορθογωνίου. Εντούτοις, αυτή η προϋπόθεση δεν είναι επαρκής από μόνη της για να καθορίσει τη μοναδική υπερ-έλλειψη, δεδομένου ότι υπάρχουν άπειρες υπερ-ελλείψεις που περνούν από όλες τις γωνίες του υπερ-ορθογωνίου. Μια ειδική περίπτωση υπερ-έλλειψης είναι η υπερ-σφαίρα που περιγράφεται από την εξίσωση (2.27). Μεταξύ τους, επιλέγουμε την υπερ-έλλειψη της οποίας οι άξονες είναι ανάλογοι προς τις βάσεις του υπερ-ορθογωνίου, δηλαδή:

$$\frac{\delta\alpha_1}{b_1}=\frac{\delta\alpha_2}{b_2}=\ldots=\frac{\delta\alpha_N}{b_N} \tag{2.31}$$

Ο συνδυασμός των δύο παραπάνω προϋποθέσεων καθορίζει με μοναδικό τρόπο την υπερ-έλλειψη, η οποία περιγράφεται από την ακόλουθη εξίσωση:

$$\sum_{i=1}^{N}\left(\frac{\left(a_{i,j_i}^{l}-u_i\left(k\right)\right)^2}{N\left(\delta a_i\right)^2}\right)=1 \tag{2.32}$$

Το σχήμα 2.8 απεικονίζει την έλλειψη που προκύπτει τελικά στη δισδιάστατη περίπτωση. Με βάση την εξίσωση (2.32) προκύπτει μια νέα πολυδιάστατη συνάρτηση συμμετοχής, όπου η σχετική απόσταση μεταξύ του ασαφούς υπόχωρου $\mathbf{A}^l$ και του διανύσματος των δεδομένων εισόδου $\mathbf{u}(k)$ δίνεται από τη σχέση:

$$d_r^l\left(\mathbf{u}\left(k\right)\right)=\sqrt{\sum_{i=1}^{N}\left(\frac{\left(a_{i,j_i}^{l}-u_i\left(k\right)\right)^2}{N\left(\delta a_i\right)^2}\right)} \tag{2.33}$$

Η αντικατάσταση της αρχικής σφαιρικής εξίσωσης της σχετικής απόστασης με μια εξίσωση ελλειψοειδούς μορφής καθιστά το μη συμμετρικό αλγόριθμο πιο “ευέλικτο”, δεδομένου ότι δίνει στο χρήστη την ευκαιρία να διαμερίσει το πεδίο ορισμού της κάθε μεταβλητής εισόδου με διαφορετικό τρόπο. Κατά συνέπεια, οι προκύπτουσες συναρτήσεις βάσης με ακτινική συμμετρία καλύπτουν αποτελεσματικότερα τα πεδία ορισμού του χώρου των μεταβλητών εισόδου όπου τα δεδομένα εισόδου είναι διαθέσιμα και παράγουν δίκτυα με μεγαλύτερη ακρίβεια (μικρότερο σφάλμα μοντελοποίησης) και μικρότερη πολυπλοκότητα. Μετά από την αντικατάσταση της εξίσωσης της απόστασης, ο αλγόριθμος συνεχίζει με τον ίδιο τρόπο όπως η αρχική μέθοδος των ασαφών μέσων για να βρεθεί το υποσύνολο των ασαφών υπόχωρων

που ορίζουν ένα μη μηδενικό πολυδιάστατο βαθμό σε όλα τα διανύσματα εισόδου που χρησιμοποιούνται στην εκπαίδευση του δικτύου.

Παρακάτω παρουσιάζεται μια επισκόπηση του αλγορίθμου NSFM.

Αλγόριθμος 2 – Αλγόριθμος NSFM

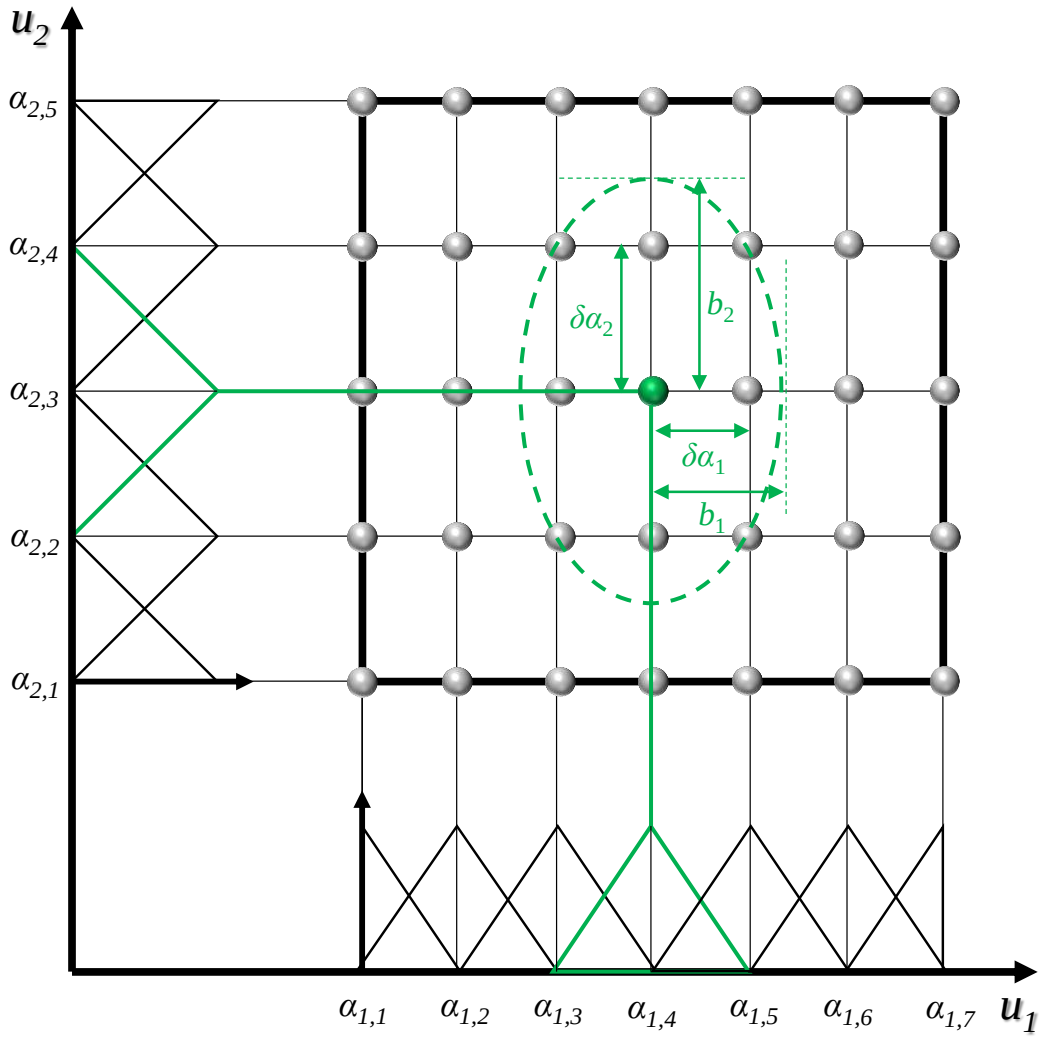
Δεδομένα Εισόδου: $\{\mathbf{U}_{train}, \mathbf{Y}_{train}\}$: Σύνολο δεδομένων εκπαίδευσης,

$\mathbf{s} = [s_1, s_2, ..., s_N]$: Αριθμός των ασαφών συνόλων που διαμερίζεται το πεδίο ορισμού κάθε μεταβλητής εισόδου

Δεδομένα Εξόδου: L : Αριθμός επιλεγμένων κέντρων RBF,

$\hat{\mathbf{U}} = [\hat{\mathbf{u}}_1,\ \hat{\mathbf{u}}_2,\ \ldots,\ \hat{\mathbf{u}}_L]^T$: Τοποθεσίες επιλεγμένων κέντρων RBF

1. Πάρε το πρώτο σημείο δεδομένων: $k \leftarrow 1$
2. Ξεκίνα τους υπολογισμούς για το πρώτο κέντρο RBF: $L \leftarrow 1$
3. **Για** *i*=1:*N* **Εκτέλεσε**:
4. Υπολόγισε το ασαφές σύνολο με τη μέγιστη δυνατή συμμετοχή σε κάθε διάσταση *i*:
   $A_i^1 = \{a_i^1, \delta a\} \leftarrow \max_{1 \le j \le s_i} \left[\mu_{A_{i,j}} \left(u_i(1) \right) \right]$
5. **Τέλος Για**
6. Δημιούργησε το πρώτο κέντρο RBF $\hat{\mathbf{u}}_1$: $\hat{\mathbf{u}}_1 = [a_1^1, a_2^1, \ldots, a_N^1]$
7. **Για** *k*=2:*K* **Εκτέλεσε**:
8. **Αν** το σημείο *k* βρίσκεται εκτός των υπέρ-ελλείψεων που ορίζονται από τα ήδη επιλεγμένα κέντρα: $\min_{1 \le l \le L} \left[d_r^l \left(\mathbf{u}(k) \right) \right] > 1$, όπου d_r^l υπολογίζεται από τον τύπο (2.33)
9. Πρόσθεσε ένα νέο κέντρο RBF: $L \leftarrow L + 1$
10. **Για** *i*=1:*N* **Εκτέλεσε**:
11. Υπολόγισε το ασαφές σύνολο με τη μέγιστη δυνατή συμμετοχή σε κάθε διάσταση *i*: $A_i^L = \{a_i^L, \delta a_i\} \leftarrow \max_{1 \le j \le s_i} \left[\mu_{A_{i,j}} \left(u_i(k) \right) \right]$
12. **Τέλος Για**
13. Δημιούργησε το *L*-οστό κέντρο RBF $\hat{\mathbf{u}}_L$: $\hat{\mathbf{u}}_L = [a_1^L, a_2^L, \ldots, a_N^L]$
14. **Τέλος Αν**
15. **Τέλος Για**



Σχήμα 2.8. Δισδιάστατο παράδειγμα: Ελλειψοειδής συνάρτηση συμμετοχής για μη συμμετρική ασαφή διαμέριση.

2.3 Σύγκριση δικτύων RBF με άλλες αρχιτεκτονικές νευρωνικών δικτύων

Μετά τη συνοπτική παρουσίαση των διαφόρων αρχιτεκτονικών νευρωνικών δικτύων του κεφαλαίου 1 και ύστερα από την αναλυτική περιγραφή των δικτύων RBF σε αυτό το κεφάλαιο, οι διαφορές που υπάρχουν είναι σαφείς.

Τα δίκτυα RBF και τα δίκτυα MLP έχουν παρόμοιους ρόλους δεδομένου ότι και τα δύο παρέχουν τεχνικές για την ανάπτυξη μη γραμμικών συναρτήσεων χαρτογράφησης μεταξύ πολυδιάστατων χώρων. Επίσης, πάντα υπάρχει ένα δίκτυο RBF ικανό να μιμείται ακριβώς ένα MLP ή αντίστροφα [2]. Παρ' όλα αυτά, οι δομές αυτών των δύο νευρωνικών δικτύων διαφέρουν κατά πολύ και γι' αυτό είναι τόσο σημαντικό να τα συγκρίνουμε λεπτομερώς. Μερικές από τις πιο σημαντικές διαφορές που κάνουν τα δίκτυα RBF να

πλεονεκτούν σε σχέση με τις άλλες αρχιτεκτονικές νευρωνικών δικτύων και πιο συγκεκριμένα σε σχέση με τα MLPs, είναι οι παρακάτω:

1. Τα νευρωνικά δίκτυα RBF (στην πιο βασική τους μορφή) έχουν απλούστερη δομή. Η πιο βασική μορφή ενός δικτύου RBF περιέχει μόνο μία ενιαία κρυφή στοιβάδα, ενώ ένα δίκτυο που ανήκει στην κατηγορία των MLPs μπορεί να έχει μία ή περισσότερες κρυφές στοιβάδες. Τα παλινδρομικά δίκτυα είναι ακόμα πιο περίπλοκα λόγω του ότι περιέχουν πολλές κρυφές στοιβάδες αλλά και βρόχους ανατροφοδότησης. Έτσι, τα δίκτυα RBF μπορούν να ερμηνευτούν ευκολότερα ως προς τη φυσική τους σημασία.
2. Οι αλγόριθμοι εκμάθησης που υιοθετούν τα δίκτυα RBF είναι αρκετά πιο γρήγοροι λόγω της διφασικής προσέγγισης του προβλήματος. Αντιθέτως, τα δίκτυα που ανήκουν στην κατηγορία των MLPs αλλά και τα παλινδρομικά δίκτυα επιλύουν όλο το πρόβλημα βελτιστοποίησης σε ένα μόνο βήμα.
3. Οι νευρώνες στην κρυφή στοιβάδα ενός δικτύου RBF είναι αρκετά διαφορετικοί και εξυπηρετούν διαφορετικούς σκοπούς σε σχέση με εκείνους που βρίσκονται στη στοιβάδα εξόδου. Επίσης, η κρυφή στοιβάδα ενός δικτύου RBF είναι μη γραμμική, ενώ η στοιβάδα εξόδου είναι γραμμική. Οι νευρώνες ενός δικτύου MLP που βρίσκονται στη κρυφή στοιβάδα ή στη στοιβάδα εξόδου, μοιράζονται ένα κοινό νευρωνικό μοντέλο, λειτουργούν και οι δύο ως ταξινομητές προτύπων και είναι μη γραμμικοί.
4. Η συνάρτηση ενεργοποίησης κάθε νευρώνα της κρυφής στοιβάδας σε ένα δίκτυο RBF δέχεται σαν είσοδο την ευκλείδεια απόσταση μεταξύ του διανύσματος εισόδου και του κέντρου του συγκεκριμένου νευρώνα. Αντίστοιχα, η συνάρτηση ενεργοποίησης κάθε νευρώνα της κρυφής στοιβάδας σε ένα δίκτυο MLP δέχεται σαν είσοδο το εσωτερικό γινόμενο του διανύσματος εισόδου και του διανύσματος συναπτικού βάρους του συγκεκριμένου νευρώνα.
5. Τα δίκτυα RBF έχουν τοπολογική συμπεριφορά και ο κάθε νευρώνας επηρεάζει μόνο μία μικρή περιοχή του υπερ-χώρου των διανυσμάτων εισόδου, δηλαδή προσεγγίζει τα διανύσματα εισόδου που εμπίπτουν στην περιοχή του. Αντίθετα, ο κάθε νευρώνας των δικτύων MLP ενεργοποιείται λαμβάνοντας υπόψη όλα τα διανύσματα εισόδου, δηλαδή τα δίκτυα MLP έχουν ολική συμπεριφορά. Έτσι, τα νευρωνικά δίκτυα RBF απαιτούν μικρότερους χρόνους εκπαίδευσης και προσομοίωσης.

ΚΕΦΑΛΑΙΟ 3

ΑΛΓΟΡΙΘΜΟΙ ΕΚΠΑΙΔΕΥΣΗΣ ΔΙΚΤΥΩΝ RBF

Σκοπός αυτού του κεφαλαίου είναι η παρουσίαση τριών νέων καινοτόμων μεθόδων εκπαίδευσης νευρωνικών δικτύων τύπου RBF που χειρίζονται αριθμητικά δεδομένα και αναπτύχθηκαν στα πλαίσια αυτής της διδακτορικής διατριβής. Συγκεκριμένα, παρουσιάζονται οι εξής τεχνικές:

- PSO-NSFM: παρουσιάζει αυξημένη ακρίβεια πρόβλεψης σε σχέση με τις μεθόδους της βιβλιογραφίας
- COOP-VW: επιτρέπει τη χρήση συναρτήσεων βάσης μεταβλητού πλάτους με αποτέλεσμα μοντέλα που παρέχουν ακόμα μεγαλύτερη ακρίβεια πρόβλεψης με σχετικά μικρή αύξηση του υπολογιστικού κόστους
- ESA-NSFM: παρουσιάζει εύκολη επιλογή λειτουργικών παραμέτρων

3.1 Εκπαίδευση νευρωνικών δικτύων RBF με τον αλγόριθμο βελτιστοποίησης σμήνους σωματιδίων

Στο κεφάλαιο 2 αναλύθηκαν δύο αλγόριθμοι εκπαίδευσης νευρωνικών δικτύων, ο απλός αλγόριθμος των ασαφών μέσων (FM ή SFM) και μία παραλλαγή του, που είναι ο μη συμμετρικός αλγόριθμος των ασαφών μέσων (NSFM). Προφανώς, ο αρχικός αλγόριθμος FM είναι λιγότερο ευέλικτος και παρέχει λιγότερες παραμέτρους σχεδιασμού σε σύγκριση με τη μη-συμμετρική έκδοσή του. Σε μία σχετικά πρόσφατη δημοσίευση [35], αποδεικνύεται ότι ο αλγόριθμος NSFM υπερτερεί έναντι του αρχικού αλγορίθμου FM στο ότι έχει μεγαλύτερη ευελιξία, γεγονός που οδηγεί σε καλύτερα δίκτυα υπό την έννοια της ακρίβειας, ή και της πολυπλοκότητας του μοντέλου. Από την άλλη πλευρά, η μη συμμετρική διαμέριση ορίζει ένα πιο περίπλοκο πρόβλημα όσον αφορά στην κατασκευή του δικτύου. Αυτό συμβαίνει, λόγω της ανάγκης που υπάρχει να βελτιστοποιηθεί ένας αυξημένος αριθμός λειτουργικών παραμέτρων. Για προβλήματα που περιλαμβάνουν μικρό αριθμό μεταβλητών εισόδου, η βέλτιστη διαμέριση μπορεί να υπολογιστεί εκτελώντας μια διεξοδική διαδικασία αναζήτησης (exhaustive search), ίδια με αυτή που χρησιμοποιείται στον αρχικό αλγόριθμο FM, όπου μπορούν να δοκιμαστούν όλες οι πιθανές διαμερίσεις [35]. Αυτή η μέθοδος είναι εφικτή για προβλήματα μικρής κλίμακας, εξαιτίας της διαδικασίας γρήγορης εκπαίδευσης που χρησιμοποιεί ο αλγόριθμος των ασαφών μέσων. Ωστόσο, το επιπλέον υπολογιστικό φορτίο που απαιτείται από τα προβλήματα μεσαίας ή μεγάλης κλίμακας, κάνει απαγορευτική τη χρήση της διαδικασίας της διεξοδικής αναζήτησης. Οι παραπάνω λόγοι καθιστούν απαραίτητη τη χρήση ενός μηχανισμού βελτιστοποίησης για να βελτιστοποιηθεί η ασαφής διαμέριση του χώρου εισόδου.

Ο αλγόριθμος βελτιστοποίησης σμήνους σωματιδίων PSO αποτελεί έναν από τους πιο επιτυχημένους αλγορίθμους βελτιστοποίησης. Είναι μία στοχαστική μέθοδος, η οποία χρησιμοποιεί έναν πληθυσμό, που αποκαλείται σμήνος, για την εύρεση λύσεων εντός ενός χώρου αναζήτησης. Τα μέλη του σμήνους, που ονομάζονται σωματίδια, «πετούν» στον χώρο αναζήτησης και ανταλλάσσουν μεταξύ τους πληροφορίες σχετικά με την προσωπική καλύτερη λύση (θέση) που έχει βρει το κάθε ένα. Σύμφωνα με αυτές τις πληροφορίες καθορίζεται η κίνηση του σμήνους (θέση και ταχύτητα).

Σε αυτή τη διδακτορική διατριβή, παρουσιάζεται ένα νέο σύστημα για τον υπολογισμό της βέλτιστης μη συμμετρικής διαμέρισης του χώρου εισόδου, το οποίο βασίζεται στον αλγόριθμο PSO. Το σύστημα αυτό οδηγεί σε μια νέα ολοκληρωμένη διαδικασία εκπαίδευσης νευρωνικών δικτύων RBF, τον αλγόριθμο PSO-NSFM, ο οποίος αποτελεί ένα ολοκληρωμένο πλαίσιο για τον πλήρη καθορισμό των παραμέτρων του νευρωνικού δικτύου βασισμένο σε βελτιστοποίηση σμήνους σωματιδίων.

3.1.1 Βελτιστοποίηση της ασαφούς διαμέρισης με τον αλγόριθμο PSO

Ο αλγόριθμος βελτιστοποίησης σμήνους σωματιδίων μπορεί να χρησιμοποιηθεί ως μηχανισμός αναζήτησης για την εύρεση της βέλτιστης μη-συμμετρικής διαμέρισης. Σε αυτή την περίπτωση, τα σωματίδια του σμήνους αποτελούν διαφορετικές πιθανές διαμερίσεις του χώρου εισόδου. Υιοθετώντας αυτού του είδους την κωδικοποίηση, τα στοιχεία $\mathbf{s}_i(t)$ του κάθε σωματιδίου την επανάληψη (simulation time step, STS) t, περιέχουν τον αριθμό των ασαφών συνόλων σε κάθε διάσταση:

$$\mathbf{s}_i(t) = \left[s_1(t), s_2(t), ..., s_N(t) \right]^T \tag{3.1}$$

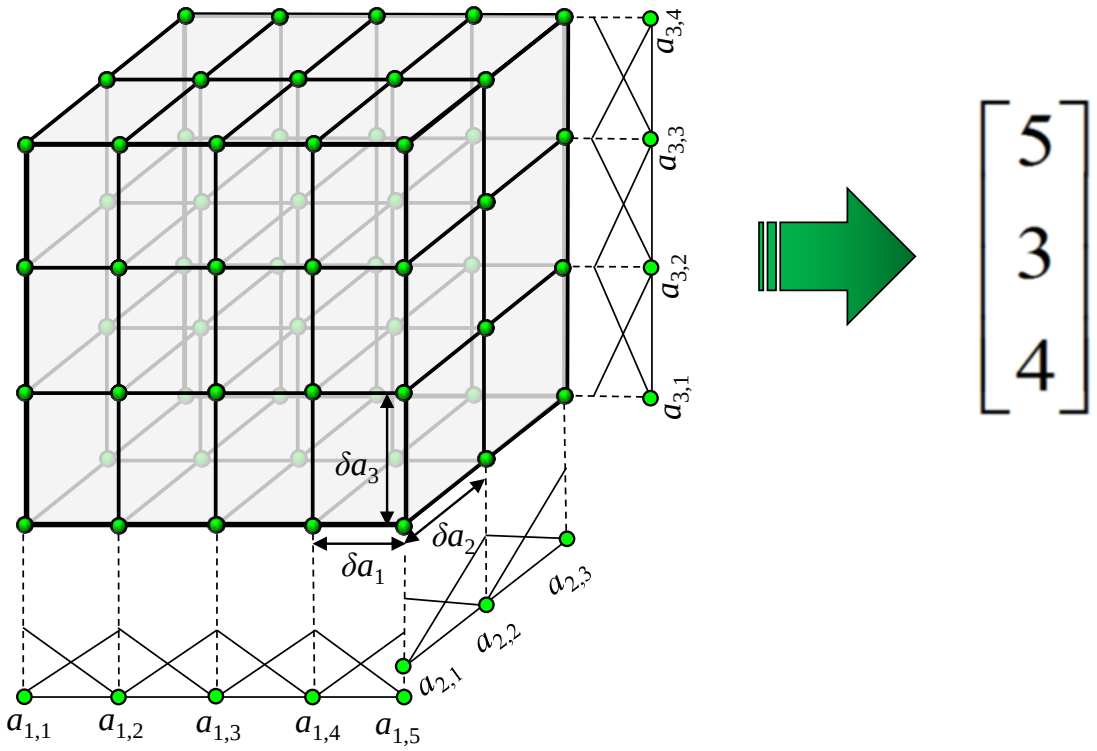
Στο σχήμα 3.1 παρουσιάζεται μία σχηματική αναπαράσταση της κωδικοποίησης της μη-συμμετρικής διαμέρισης σε σωματίδια, χρησιμοποιώντας ένα 3-διάστατο χώρο εισόδου. Σύμφωνα με τη συγκεκριμένη κωδικοποίηση, οι εξισώσεις (1.17) και (1.18) που διέπουν ένα βασικό μοντέλο PSO πλήρους συνδεδεμένης τοπολογίας, μετατρέπονται σε:

$$\mathbf{s}_i(t+1) = \mathbf{s}_i(t) + \mathbf{v}_i(t+1) \tag{3.2}$$

$$v_{ij}(t+1) = \text{round}\left\{ \begin{array}{r} v_{ij}(t) + c_1 r_{1j}(t)\left[y_{ij}(t) - s_{ij}(t) \right] + ... \\ ... + c_2 r_{2j}(t)\left[\hat{y}_j(t) - s_{ij}(t) \right] \end{array} \right\} \tag{3.3}$$

Στην παραπάνω εξίσωση, ο όρος “round” χρησιμοποιείται για να στρογγυλοποιηθούν οι ταχύτητες στον πλησιέστερο ακέραιο, δεδομένου ότι ο αριθμός των ασαφών συνόλων σε κάθε διάσταση μπορεί να λάβει μόνο ακέραιες τιμές. Επίσης, ο όρος της αδράνειας $w_{inertia}$. είναι ίσος με τη μονάδα, γι’ αυτό και δεν καταγράφεται στην παραπάνω εξίσωση.

Στην περίπτωση ενός προβλήματος ελαχιστοποίησης, η προσωπική καλύτερη θέση του κάθε σωματιδίου υπολογίζεται με βάση τη σχέση (1.20), αφού πρώτα αντικαθίστανται τα σωματίδια $\mathbf{x}_i(t)$ με τα σωματίδια $\mathbf{s}_i(t)$ που περιέχουν τη διαμέριση εισόδου. Επίσης, η καλύτερη ολική θέση και ο παράγοντας περιορισμού ταχύτητας δίνονται από τις σχέσεις



Σχήμα 3.1. Κωδικοποίηση ενός μη συμμετρικού 3-διάστατου χώρου εισόδου σε ένα σωματίδιο: Ο αριθμός των ασαφών συνόλων σε κάθε διάσταση κωδικοποιείται στο αντίστοιχο στοιχείο του σωματιδίου.

(1.22) και (1.19), αντίστοιχα. Τέλος, στην περίπτωση ενός προβλήματος μεγιστοποίησης, ισχύει ότι και στην περίπτωση της ελαχιστοποίησης με τη διαφορά ότι αντί για τις σχέσεις (1.20) και (1.22), χρησιμοποιούνται οι σχέσεις (1.21) και (1.23), αντίστοιχα.

3.1.2 Μέθοδος εκπαίδευσης δικτύων RBF βασισμένη στη βελτιστοποίηση σμήνους

Η βελτιστοποίηση της μη συμμετρικής ασαφούς διαμέρισης του χώρου εισόδου χρησιμοποιώντας τη μέθοδο PSO, μπορεί να αποτελέσει τη βάση για μια ολοκληρωμένη μεθοδολογία εκπαίδευσης των δικτύων RBF. Ο αλγόριθμος ξεκινά κατατάσσοντας με τυχαίο τρόπο τα διαθέσιμα δεδομένα σε τρία υποσύνολα δεδομένων και συγκεκριμένα στα σύνολα δεδομένων εκπαίδευσης, αξιολόγησης και ελέγχου, όπως αναφέρεται στο κεφάλαιο 1.2.7. Η ύπαρξη ενός τρίτου ανεξάρτητου συνόλου δεδομένων για τον έλεγχο του παραγόμενου μοντέλου RBF είναι κρίσιμη, δεδομένου ότι η διαδικασία βελτιστοποίησης μπορεί να οδηγήσει το μοντέλο στο φαινόμενο της υπερπροσαρμογής στο σύνολο αξιολόγησης (το φαινόμενο της υπερπροσαρμογής αναλύθηκε στο κεφάλαιο 1.2.7).

Έπειτα, ο αλγόριθμος προχωρά στο στάδιο της αρχικοποίησης των σωματιδίων. Εδώ, επιλέγονται με τυχαίο τρόπο P διαφορετικές διαμερίσεις του χώρου εισόδου, οι οποίες κωδικοποιούνται ως σωματίδια. Στη συνέχεια, κάθε σωματίδιο δημιουργεί ένα δίκτυο RBF, εφαρμόζοντας τον αλγόριθμο NSFM. Ο τελευταίος, υπολογίζει τον αριθμό και τις συντεταγμένες των κέντρων RBF με βάση την εξίσωση (2.33). Το στάδιο του καθορισμού των κέντρων RBF, ακολουθεί το στάδιο του υπολογισμού των συναπτικών βαρών χρησιμοποιώντας τη μέθοδο των ελαχίστων τετραγώνων, σύμφωνα με την εξίσωση (2.14).

Με βάση τα παραγόμενα δίκτυα RBF, για κάθε σωματίδιο υπολογίζεται μια τιμή για τη συνάρτηση καταλληλότητας (fitness function), εφαρμόζοντας ένα κριτήριο σφάλματος. Ο αλγόριθμος PSO-NSFM χρησιμοποιεί ως κριτήριο σφάλματος το κριτήριο της τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (root mean square error, RMSE), αλλά μπορούν επίσης να χρησιμοποιηθούν και άλλα εναλλακτικά κριτήρια σφάλματος (περισσότερες πληροφορίες αναφέρονται στο κεφάλαιο 1.2.8). Οι τιμές των συναρτήσεων καταλληλότητας που προκύπτουν, συγκρίνονται με τις προσωπικές και τις ολικά καλύτερες θέσεις των σωματιδίων. Υπό κάποιους όρους, οι θέσεις αυτές ανανεώνονται σύμφωνα με τις σχέσεις (1.20) και (1.22) αφού πρόκειται για πρόβλημα ελαχιστοποίησης. Στην περίπτωση ενός προβλήματος μεγιστοποίησης οι θέσεις αυτές θα ανανεώνονταν σύμφωνα με τις σχέσεις (1.21) και (1.23).

Έπειτα, ο αλγόριθμος προχωρά στο επόμενο στάδιο, το οποίο περιλαμβάνει τον υπολογισμό των καινούριων ταχυτήτων των σωματιδίων για την επανάληψη t+1 και εν συνεχεία την ενημέρωση των θέσεων των σωματιδίων, χρησιμοποιώντας τις σχέσεις (3.3) και (3.2), αντίστοιχα. Κατόπιν, ο αλγόριθμος επιστρέφει ξανά στο στάδιο της κατασκευής του δικτύου RBF και σταματά αυτή την επαναληπτική διαδικασία όταν ισχύσει τουλάχιστον μία από τις δύο συνθήκες διακοπής. Η πρώτη συνθήκη διακοπής ικανοποιείται όταν επιτυγχάνεται ο μέγιστος αριθμός βημάτων προσομοίωσης. Η δεύτερη συνθήκη διακοπής ικανοποιείται όταν η κανονικοποιημένη ακτίνα του σμήνους γίνει μικρότερη από μία συγκεκριμένη τιμή κατωφλίου. Η κανονικοποιημένη ακτίνα του σμήνους υπολογίζεται ως εξής:

$$R_{norm}(t) = \frac{R_{max}(t)}{R_{max}(1)} \tag{3.4}$$

Όπου $R_{max}(t)$ είναι η μέγιστη ακτίνα του σμήνους την επανάληψη *t*, όπου υπολογίζεται παρακάτω:

$$R_{max}(t) = \max_{1 \le i \le P}\left[\left\|\mathbf{s}_i(t) - \hat{\mathbf{y}}(t)\right\|\right] \tag{3.5}$$

Όταν η κανονικοποιημένη ακτίνα του σμήνους R_{norm} μικραίνει, η δυναμική του σμήνους για περαιτέρω βελτίωση μειώνεται [1], καθώς όλα τα σωματίδια βρίσκονται σε μικρή απόσταση από την ολική καλύτερή τους θέση.

Παρακάτω παρουσιάζεται μια επισκόπηση του αλγορίθμου PSO-NSFM για ένα πρόβλημα ελαχιστοποίησης.

Αλγόριθμος 3 – Αλγόριθμος PSO-NSFM

Δεδομένα Εισόδου: $\{\mathbf{U}_{train}, \mathbf{Y}_{train}\}$: Σύνολο δεδομένων εκπαίδευσης,

$\{\mathbf{U}_{val}, \mathbf{Y}_{val}\}$: Σύνολο δεδομένων αξιολόγησης

s_{min}, s_{max}: Ελάχιστος και μέγιστος αριθμός των ασαφών συνόλων που διαμερίζεται το πεδίο ορισμού κάθε μεταβλητής εισόδου,

P: Πληθυσμός σμήνους,

c_1, c_2, V_{max}: Λειτουργικές παράμετροι του αλγορίθμου PSO που επηρεάζουν τον γνωστικό όρο, τον κοινωνικό όρο και τη σταθερά περιορισμού ταχύτητας, αντίστοιχα,

ξ_1, ξ_2: παράμετροι για τις συνθήκες διακοπής

Δεδομένα Εξόδου: L_f: Αριθμός επιλεγμένων κέντρων RBF,

$\hat{\mathbf{U}}_f$: Τοποθεσίες επιλεγμένων κέντρων RBF,

$\mathbf{w}_f$: Συναπτικά βάρη επιλεγμένων κέντρων RBF

1. Αρχικοποίηση των συντεταγμένων των σωματιδίων με τυχαίους ακέραιους αριθμούς μεταξύ των s_{min} και s_{max}: $\mathbf{s}_i(0) \leftarrow \left[\underbrace{\text{rand, rand, …, rand}}_{N}\right]$
2. Ξεκινά με την πρώτη επανάληψη: $t \leftarrow 1$
3. **Όσο** καμία από τις δύο συνθήκες διακοπής (ξ_1, ξ_2) έχουν επιτευχθεί **Εκτέλεσε**:
4. **Για** i=1:P **Εκτέλεσε**:
5. Δώσε τους αριθμούς των ασαφών συνόλων $\mathbf{s}_i(t)$ και το σύνολο δεδομένων εκπαίδευσης $\{\mathbf{U}_{train}, \mathbf{Y}_{train}\}$ στον Αλγόριθμο 2, προκειμένου να υπολογιστεί ο συνολικός αριθμός των κέντρων RBF $L_i(t)$ και οι θέσεις τους $\hat{\mathbf{u}}_i(t)$
6. Υπολόγισε τα συναπτικά βάρη $\mathbf{w}_i(t)$ σύμφωνα με τον τύπο (2.14)

7. Υπολόγισε τη συνάρτηση καταλληλότητας $f\left(\mathbf{s}_i\left(t\right)\right)$ για το σύνολο δεδομένων αξιολόγησης $\left\{\mathbf{U}_{val}, \mathbf{Y}_{val}\right\}$, και αν χρειαστεί, ενημέρωσε τις προσωπικές $\mathbf{y}_i\left(t\right)$ και τις ολικά $\hat{\mathbf{y}}\left(t\right)$ καλύτερες θέσεις, με τους τύπους (1.20) και (1.22)
8. **Τέλος Για**
9. **Για** *i*=1:*P* **Εκτέλεσε**:
10. **Για** *j*=1:*N* **Εκτέλεσε**:
11. Ενημέρωση τα στοιχεία του διανύσματος της ταχύτητας $v_{ij}\left(t+1\right)$ με τον τύπο (3.3)
12. Εκτέλεσε τον μηχανισμό περιορισμού ταχύτητας, αν χρειάζεται, με τον τύπο (1.19)
13. **Τέλος Για**
14. Ενημέρωσε τις θέσεις των σωματιδίων $\mathbf{s}_i\left(t+1\right)$, με τον τύπο (3.2)
15. **Τέλος Για**
16. Προχώρα στην επόμενη επανάληψη: $t \leftarrow t+1$
17. **Τέλος Όσο**

3.1.3 Αξιολόγηση του αλγορίθμου PSO-NSFM σε σύνολα δεδομένων τύπου benchmark – Αποτελέσματα και συζήτηση

Η προτεινόμενη μεθοδολογία αξιολογήθηκε σε διάφορα σύνολα δεδομένων τύπου benchmark, τα οποία περιλαμβάνουν είτε συνθετικά, είτε πραγματικά δεδομένα. Αυτά τα σύνολα δεδομένων ανακτήθηκαν από τη βάση UCI Machine Learning Repository [190], εκτός από το σύνολο Gas-Furnace, το οποίο μπορεί να βρεθεί στο [191]. Στον πίνακα 3.1 παρουσιάζονται όλα τα σύνολα δεδομένων που χρησιμοποιήθηκαν σε αυτή την εφαρμογή, μαζί με τον αριθμό των μεταβλητών εισόδου και τον αριθμό των δεδομένων τους, ενώ μια σύντομη περιγραφή για κάθε ένα από αυτά δίνεται παρακάτω.

- **Ailerons–Elevators**: Αυτά τα δύο σύνολα δεδομένων περιέχουν πληροφορίες για ένα αεροσκάφος F16. Οι μεταβλητές περιέχουν πληροφορίες για την κατάσταση του αεροπλάνου, ενώ ο στόχος είναι να προβλεφθεί η λειτουργία ρύθμισης των πτερυγίων κλίσης (ailerons) και του πηδαλίου ανόδου-καθόδου της ουράς (elevators) του αεροσκάφους, αντίστοιχα.
- **Auto MPG**: Σε αυτή την εφαρμογή χρησιμοποιήθηκε μια τροποποιημένη έκδοση του αρχικού συνόλου δεδομένων, όπως αναφέρεται στην εργασία [192]. Ο στόχος είναι να προβλεφθεί η κατανάλωση καυσίμου ενός αυτοκινήτου, χρησιμοποιώντας ως εισόδους μερικά από τα χαρακτηριστικά του.
- **Auto Price**: Αυτό το σύνολο δεδομένων είναι επίσης γνωστό ως “Auto Mobile” και αφορά στην πρόβλεψη του κόστους ενός αυτοκινήτου, ως συνάρτηση των προδιαγραφών του, π.χ. το μέγεθος του κινητήρα, η ιπποδύναμη, κτλ. Μόνο οι συνεχείς αριθμητικές μεταβλητές χρησιμοποιήθηκαν ως είσοδοι.

- **Boston Housing**: Αυτό το σύνολο δεδομένων αφορά στην πρόβλεψη τιμών πώλησης σπιτιών στη Βοστώνη, με βάση τα χαρακτηριστικά τους. Τα δεδομένα ελήφθησαν από την απογραφή του 1970.
- **CPU Small**: Το CPU Small είναι μια μικρή παραλλαγή του συνόλου δεδομένων "Computer Activity". Περιλαμβάνει μετρήσεις δραστηριοτήτων ενός υπολογιστικού συστήματος Sun SPARCstation. Ο σκοπός αυτού του συνόλου δεδομένων, είναι να προβλεφθεί ο χρόνος που οι επεξεργαστές τρέχουν σε κατάσταση χρήστη, λαμβάνοντας υπόψη έναν περιορισμένο αριθμό χαρακτηριστικών.
- **Friedman**: Αυτό το σύνολο δεδομένων παράγεται με βάση τη συνάρτηση [193]:

$$y = 5\left[2\sin\left(\pi x_1 x_2\right) + 4\left(x_3 - 0.5\right)^2 + 2x_4 + x_5\right] + \varepsilon_{noise} \tag{3.6}$$

όπου ε_{noise} είναι γκαουσιανός θόρυβος ~$N(0,\ 0.8)$. Οι τιμές για τις μεταβλητές εισόδου προέρχονται από μια ομοιόμορφη κατανομή στο διάστημα [0, 1].

- **Gas Furnace**: Αυτό το σύνολο περιλαμβάνει την πρόβλεψη $y(t)$ της συγκέντρωσης του CO_2 των καυσαερίων ενός συστήματος κλίβανου με αέριο, ενώ αλλάζει ο ρυθμός τροφοδοσίας του μεθανίου $u(t)$. Σε μια προηγούμενη μελέτη [27] φάνηκε ότι ο βέλτιστος συνδυασμός των μεταβλητών εισόδου, οδηγεί σε ένα μοντέλο της παρακάτω μορφής:

Πίνακας 3.1. Επισκόπηση συνόλων δεδομένων τύπου benchmark

Σύνολο δεδομένων	**Αριθμός μεταβλητών εισόδου**	**Αριθμός δεδομένων**
ΠΡΑΓΜΑΤΙΚΑ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ		
Ailerons	40	13750
Auto MPG	6	392
Auto Price	15	159
Boston Housing	13	506
CPU small	12	8192
Elevators	18	16599
Gas Furnace	3	290
House Price-16H	16	22784
Machine CPU	6	209
ΣΥΝΘΕΤΙΚΑ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ		
Friedman	5	1000
Mackey-Glass	6	1995
Samad	3	1500

$$y(t) = RBF\left(u(t-2), y(t-2), y(t-1)\right) \tag{3.7}$$

- **House Price-16H**: Αυτό το σύνολο δεδομένων αφορά στην πρόβλεψη της διάμεσης τιμής των σπιτιών, με βάση τη δημογραφική σύσταση και την κατάσταση της αγοράς κατοικίας.
- **Machine CPU**: Αυτό το σύνολο δεδομένων είναι επίσης γνωστό ως "Computer Hardware". Ο στόχος είναι να προβλεφθεί η σχετική απόδοση ενός επεξεργαστή, με βάση κάποια χαρακτηριστικά του.
- **Mackey-Glass**: Το σύνολο Mackey-Glass προέρχεται από μια χαοτική χρονοσειρά που έχει χρησιμοποιηθεί εκτενώς στην αξιολόγηση μοντέλων νευρωνικών δικτύων [194-196]. Για τη διακριτή περίπτωση, το σύνολο παράγεται από την ακόλουθη εξίσωση διαφορών:

$$y(t+1) = (1-b)\,y(t) + a\frac{y(t-\tau)}{\left(1+y^{10}(t-\tau)\right)} \tag{3.8}$$

Ο στόχος είναι να δημιουργηθεί ένα μοντέλο που να προβλέπει την τρέχουσα τιμή της μεταβλητής $y(t)$, με βάση τις προηγούμενες έξι τιμές της:

$$y(t) = RBF\left(y(t-1), y(t-2), \ldots y(t-6)\right) \tag{3.9}$$

Για τη δημιουργία του συνόλου δεδομένων, εφαρμόστηκαν οι τιμές των παραμέτρων a=0.2, b=0.1 and τ=30 στην εξίσωση (3.8). Θα πρέπει να σημειωθεί ότι οι πρώτες 1000 τιμές διαγράφηκαν, προκειμένου το σύστημα να μπορέσει να φθάσει σε σταθερή κατάσταση. Επίσης, προστέθηκε γκαουσιανός θόρυβος ~N(0, 0.05) στην έξοδο $y(t)$.

- **Samad**: Αυτό το σύνολο δεδομένων παράγεται με βάση τη συνάρτηση [197]:

$$y = \frac{1}{1+\exp\left[-\exp(x_1)+(x_2-0.5)^2+\sin(\pi x_3)\right]} + \varepsilon_{noise} \tag{3.10}$$

όπου ε_{noise} είναι γκαουσιανός θόρυβος ~N(0, 0.025). Οι τιμές για τις μεταβλητές εισόδου προέρχονται από μια ομοιόμορφη κατανομή στο διάστημα [0, 1].

Κάθε ένα από τα παραπάνω σύνολα δεδομένων χωρίστηκε στα υποσύνολα εκπαίδευσης, αξιολόγησης και ελέγχου. Τα δεδομένα του κάθε συνόλου δεδομένων μοιράστηκαν με τυχαίο τρόπο και σε ποσοστά 50% - 25% - 25% στα παραπάνω 3 υποσύνολα, αντίστοιχα. Για τα σύνολα Gas-Furnace και Mackey-Glass δεν ισχύει ο παραπάνω διαχωρισμός των δεδομένων. Ο λόγος είναι ότι συνήθως, στις εφαρμογές που αφορούν στα δύο αυτά σύνολα, τα δεδομένα τους χρησιμοποιούνται με την αρχική τους σειρά.

Στον πίνακα 3.2 αναφέρονται οι παράμετροι λειτουργίας του αλγορίθμου PSO-NSFM (συμπεριλαμβανομένων των παραμέτρων που χρησιμοποιούνται για τους αλγορίθμους NSFM και PSO), που εφαρμόζονται σε όλα τα πειράματα. Συγκεκριμένα, ο ελάχιστος και ο μέγιστος αριθμός των ασαφών συνόλων καθορίζουν τα κατώτερα και τα ανώτερα όρια του χώρου αναζήτησης, αντίστοιχα. Προηγούμενα πειράματα που είχαν εκτελεστεί χρησιμοποιώντας τη συμμετρική και τη μη-συμμετρική έκδοση του αλγορίθμου FM έδειξαν ότι διαμερίσεις εκτός αυτών των ορίων δεν είναι πιθανό να οδηγήσουν σε πετυχημένα μοντέλα. Επίσης, οι σταθερές c_1 και c_2 που επηρεάζουν τον γνωστικό και τον

Πίνακας 3.2. Λειτουργικές παράμετροι για τον αλγόριθμο PSO-NSFM

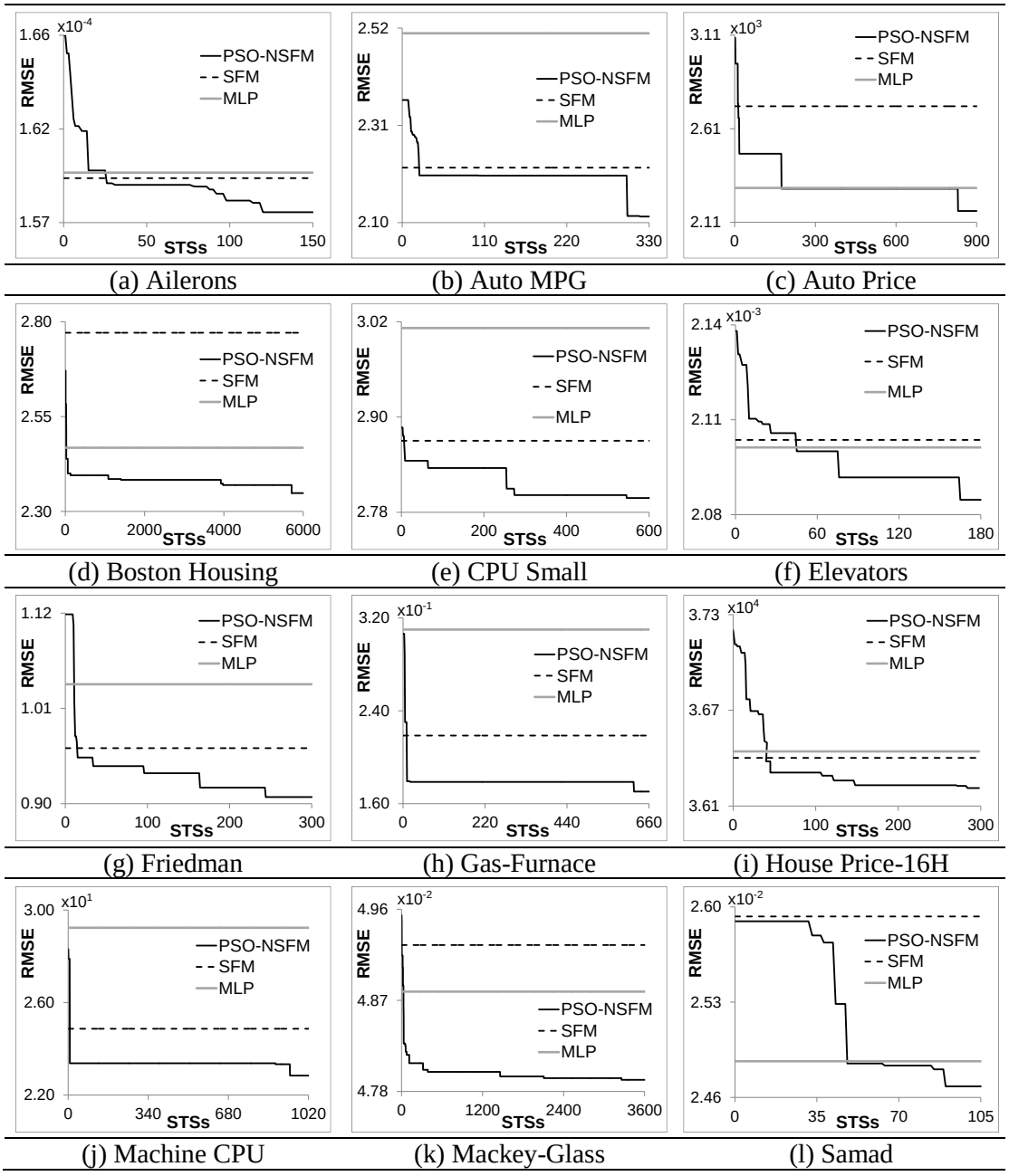
Παράμετρος	**Σύμβολο**	**Τιμή**
Ελάχιστος αριθμός ασαφών συνόλων	$s_{\min}$	4
Μέγιστος αριθμός ασαφών συνόλων	$s_{\max}$	50
Πληθυσμός	P	20
Νοσταλγία	c_1	0.05
Ζήλεια	c_2	0.05
Σταθερά περιορισμού ταχύτητας[α]	V_{max}	S:5, M:15, L:25
Μέγιστος αριθμός βημάτων προσομοίωσης	ξ_1	8000
Ελάχιστη κανονικοποιημένη ακτίνα σμήνους	ξ_2	0.1

[α]Τιμή που εξαρτάται από τη διαστατικότητα του χώρου εισόδου: Μικρά προβλήματα (Small - S): 1-4 μεταβλητές εισόδου, Μέτρια προβλήματα (Medium - M): 5-8 μεταβλητές εισόδου, Μεγάλα προβλήματα (Large - L): περισσότερες από 8 μεταβλητές εισόδου

κοινωνικό όρο, αντίστοιχα, διαπιστώθηκε ότι είχαν μικρή επίδραση στη διαδικασία βελτιστοποίησης, όταν διατηρούσαν μικρές τιμές. Επιπλέον, επελέγησαν ίσες τιμές για τους όρους c_1 και c_2, καθώς προηγούμενες μελέτες έχουν δείξει ότι ο αλγόριθμος PSO είναι συνήθως πιο αποτελεσματικός όταν υπάρχει καλή ισορροπία μεταξύ των δύο αυτών σταθερών [1]. Η εξισορρόπηση των δυνατοτήτων ολικής και τοπικής εξερεύνησης του σμήνους ρυθμίζεται με τη σταθερά περιορισμού ταχύτητας V_{max}. Καθώς το μέγεθος του χώρου αναζήτησης αυξάνεται ανάλογα με τη διάσταση του χώρου εισόδου, με ανάλογο τρόπο θα πρέπει να αυξάνονται και οι ταχύτητες των σωματιδίων, προκειμένου ο χώρος αναζήτησης να μπορεί να διερευνηθεί επαρκώς εντός εύλογου αριθμού επαναλήψεων. Ως εκ τούτου, χρησιμοποιήθηκαν τρεις διαφορετικές τιμές για τη σταθερά περιορισμού ταχύτητας, οι οποίες εξαρτώνται από τη διάσταση του χώρου εισόδου του προβλήματος (Μικρά προβλήματα: 1-4 μεταβλητές εισόδου, Μέτρια προβλήματα: 5-8 μεταβλητές εισόδου, Μεγάλα προβλήματα: πάνω από 8 μεταβλητές εισόδου).

Για λόγους σύγκρισης, χρησιμοποιήθηκαν δύο επιπλέον μεθοδολογίες: τα δίκτυα RBF που εκπαιδεύτηκαν με τον συμμετρικό αλγόριθμο των ασαφών μέσων (SFM), και τα δίκτυα MLPs που εκπαιδεύτηκαν με τον αλγόριθμο Levenberg-Marquardt [4]. Οι παράμετροι και των δύο δικτύων προσδιορίστηκαν χρησιμοποιώντας το σύνολο εκπαίδευσης και στη συνέχεια η επιλογή του μοντέλου έγινε με βάση το σύνολο αξιολόγησης. Στην περίπτωση των δικτύων RBF που εκπαιδεύτηκαν με τον αλγόριθμο SFM, η επιλογή του μοντέλου ελέγχεται από μία μόνο παράμετρο, τον αριθμό των ασαφών συνόλων s, σύμφωνα με τον οποίο διαμερίζεται κάθε μεταβλητή εισόδου. Η παράμετρος s επιλέχτηκε αφού δοκιμάστηκαν όλες οι διαμερίσεις από 4 ως 50 ασαφή σύνολα. Όσον αφορά στα δίκτυα MLP, χρησιμοποιήθηκε μία αρχιτεκτονική δύο κρυφών στοιβάδων όπου εκτελέστηκε διεξοδική αναζήτηση, δοκιμάζοντας όλους τους δυνατούς συνδυασμούς, όταν ο αριθμός των κόμβων σε κάθε κρυφή στοιβάδα κυμαίνεται από 5 έως 40.

Τα αποτελέσματα όλων των συνόλων δεδομένων εμφανίζονται στους πίνακες 3.3 και 3.4. Στον πίνακα 3.3 απεικονίζονται ο δείκτης RMSE στα σύνολα αξιολόγησης και ελέγχου, ο αριθμός των νευρώνων για τις τρεις μεθοδολογίες, και η επιλεγμένη ασαφής διαμέριση για τα δύο μοντέλα RBF. Ο υπολογιστικός χρόνος για τις τρεις μεθοδολογίες και οι επαναλήψεις που απαιτούνται για την ολοκλήρωση του αλγορίθμου PSO-NSFM φαίνονται στον πίνακα



(a) Ailerons (b) Auto MPG (c) Auto Price

(d) Boston Housing (e) CPU Small (f) Elevators

(g) Friedman (h) Gas-Furnace (i) House Price-16H

(j) Machine CPU (k) Mackey-Glass (l) Samad

Σχήμα 3.2. Εξέλιξη του κριτηρίου RMSE ανά επανάληψη στο σύνολο δεδομένων αξιολόγησης για τις τρεις μεθοδολογίες. Για τον αλγόριθμο PSO-NSFM, απεικονίζεται η καλύτερη λύση που βρέθηκε σύμφωνα με τον δείκτη RMSE στο σύνολο αξιολόγησης.

3.4. Επίσης, στον πίνακα 3.4 συμπεριλαμβάνονται ο υπολογιστικός χρόνος και οι επαναλήψεις που απαιτούνται ώστε ο αλγόριθμος PSO-NSFM να κατασκευάσει ένα μοντέλο που υπερτερεί των μοντέλων του αλγορίθμου SFM, όσον αφορά στο δείκτη RMSE. Όλοι οι υπολογιστικοί χρόνοι μετρήθηκαν σε έναν υπολογιστή με επεξεργαστή Intel Core 2 Quad (2.83 GHz) και μνήμη RAM 4GB. Λόγω της στοχαστικής φύσης της μεθόδου PSO, ο αλγόριθμος PSO-NSFM επιστρέφει διαφορετικά αποτελέσματα κάθε φορά που εκτελείται. Ωστόσο, για να προκύψουν στατιστικά αξιόπιστα αποτελέσματα, ο αλγόριθμος εκτελέστηκε 30 φορές για κάθε σύνολο δεδομένων. Έτσι, στους πίνακες απεικονίζεται το καλύτερο αποτέλεσμα και σε παρένθεση ο μέσος όρος και η τυπική απόκλιση των αποτελεσμάτων που έδωσε ο αλγόριθμος PSO-NSFM τις 30 φορές που εκτελέστηκε. Στο σχήμα 3.2

παρουσιάζεται η εξέλιξη του δείκτη RMSE στο σύνολο δεδομένων αξιολόγησης, ανά αριθμό επαναλήψεων (STSs), για το μοντέλο που έδωσε το καλύτερο αποτέλεσμα σε κάθε περίπτωση.

Ο αλγόριθμος PSO-NSFM φαίνεται ότι σε όλες τις περιπτώσεις, ξεπερνά τον αλγόριθμο SFM και τα δίκτυα MLP, όσον αφορά στην καλύτερη λύση που βρέθηκε στα σύνολα αξιολόγησης και ελέγχου. Σε ότι αφορά, στη μέση τιμή του δείκτη RMSE που υπολογίστηκε από τον αλγόριθμο PSO-NSFM, είναι χαμηλότερη σε σύγκριση με αυτή που βρέθηκε από τις άλλες δύο μεθοδολογίες σε 10 από τις 12 περιπτώσεις, λαμβάνοντας υπόψη είτε το σύνολο αξιολόγησης είτε το σύνολο ελέγχου. Επίσης, οι τιμές τυπικής απόκλισης του δείκτη RMSE είναι σχετικά μικρές. Αυτό δείχνει ότι ο αλγόριθμος παρέχει αποτελέσματα, τα οποία δεν έχουν μεγάλες αποκλίσεις μεταξύ τους.

Μία άλλη σημαντική βελτίωση του αλγορίθμου SFM αφορά στο μέγεθος των παραγόμενων δικτύων. Σε 10 από τα 12 σύνολα, ο αλγόριθμος PSO-NSFM κατορθώνει να μειώσει τον αριθμό των κόμβων, ενώ ταυτόχρονα εξακολουθεί να παρέχει μικρότερα σφάλματα προσέγγισης. Στην πραγματικότητα, σε 5 περιπτώσεις η μείωση των κέντρων RBF, που επιλέγονται από τον αλγόριθμο PSO-NSFM, υπερέβη το 20%, σε σύγκριση με τον αριθμό των κέντρων που επιλέγονται από τον αλγόριθμο SFM.

Όσον αφορά στις επαναλήψεις, ο αλγόριθμος PSO-NSFM χρειάζεται μεγαλύτερη υπολογιστική προσπάθεια μέχρι να ολοκληρωθεί, σε σύγκριση με τη μέθοδο SFM. Αυτό είναι αναμενόμενο λόγω της σημαντικής αύξησης του μεγέθους του χώρου αναζήτησης που πρέπει να εξερευνήσει ο προτεινόμενος αλγόριθμος. Ωστόσο, είναι σημαντικό να σημειωθεί ότι το μοντέλο RBF που παράγεται από τον αλγόριθμο PSO-NSFM, μέσα σε μόλις λίγα βήματα, ξεπερνά το τελικό αποτέλεσμα της μεθόδου SFM, όσον αφορά στην ακρίβεια πρόβλεψης. Αυτό μπορεί να μεταφραστεί σε κέρδος στον υπολογιστικό χρόνο, ο οποίος είναι μικρότερος σε όλες τις περιπτώσεις (λαμβάνοντας υπόψη είτε το μοντέλο που έδωσε το καλύτερο αποτέλεσμα μεταξύ των 30 εκτελέσεων, είτε τον μέσο όρο των αποτελεσμάτων που έδωσαν τα μοντέλα των 30 εκτελέσεων), σε σύγκριση με το χρόνο που απαιτείται από τον αλγόριθμο SFM για να ολοκληρωθεί. Στην πραγματικότητα, ο αλγόριθμος PSO-NSFM κατάφερε να ξεπεράσει τον αλγόριθμο SFM ακριβώς από την πρώτη επανάληψη σε 5 περιπτώσεις, σε ότι αφορά στο καλύτερο μοντέλο από τις 30 εκτελέσεις. Με λίγα λόγια, ο προτεινόμενος αλγόριθμος έχει την ικανότητα να κατασκευάζει μοντέλα δικτύων RBF που δίνουν εξαιρετικά ακριβή αποτελέσματα μέσα στις πρώτες επαναλήψεις. Παρόλα αυτά, μέχρι την ολοκλήρωση του αλγορίθμου PSO-NSFM, τα αποτελέσματα συνεχίζουν να βελτιώνονται.

Θα πρέπει να σημειωθεί ότι μία αύξηση του αριθμού των παραδειγμάτων και των διαστάσεων, οδηγεί σε μεγαλύτερο υπολογιστικό φορτίο. Η αύξηση της πολυπλοκότητας θα μπορούσε τελικά να οδηγήσει σε τόσο μεγάλους υπολογιστικούς χρόνους, που πρακτικά θα ήταν αδύνατο να ολοκληρωθεί ο αλγόριθμος. Ωστόσο, η απόδοση της προτεινόμενης μεθόδου σε περιπτώσεις υψηλών διαστάσεων και μεγάλου αριθμού παραδειγμάτων δείχνει ότι ο αλγόριθμος μπορεί να χειριστεί σύνολα δεδομένων υψηλής πολυπλοκότητας.

Πίνακας 3.3. Αποτελέσματα για όλα τα σύνολα δεδομένων – RMSE, ασαφής διαμέριση και αριθμός κόμβων

Σύνολο δεδομένων	**Αλγόριθμος**	**RMSE συνόλου αξιολόγησης**	**RMSE συνόλου ελέγχου**
Ailerons	PSO-NSFM	1.57×10^{-4} ($1.58\times10^{-4}\pm1.08\times10^{-6}$)	1.56×10^{-4} ($1.56\times10^{-4}\pm1.70\times10^{-6}$)
	SFM	1.59×10^{-4}	1.57×10^{-4}
	MLP	1.59×10^{-4}	1.58×10^{-4}
Auto MPG	PSO-NSFM	2.11 ($2.14\pm3.19\times10^{-2}$)	1.95 ($2.09\pm8.96\times10^{-2}$)
	SFM	2.22	2.12
	MLP	2.51	2.34
Auto Price	PSO-NSFM	2.17×10^{3} ($2.38\times10^{3}\pm1.28\times10^{2}$)	2.11×10^{3} ($2.30\times10^{3}\pm9.06\times10^{1}$)
	SFM	2.73×10^{3}	2.86×10^{3}
	MLP	2.29×10^{3}	2.38×10^{3}
Boston Housing	PSO-NSFM	2.35 ($2.38\pm3.13\times10^{-2}$)	2.72 ($3.12\pm2.26\times10^{-1}$)
	SFM	2.77	2.93
	MLP	2.47	3.38
CPU Small	PSO-NSFM	2.80 ($2.83\pm1.34\times10^{-2}$)	2.81 ($2.83\pm2.39\times10^{-2}$)
	SFM	2.87	2.84
	MLP	3.01	3.02
Elevators	PSO-NSFM	2.08×10^{-3} ($2.08\times10^{-3}\pm7.27\times10^{-6}$)	2.11×10^{-3} ($2.11\times10^{-3}\pm1.08\times10^{-5}$)
	SFM	2.10×10^{-3}	2.12×10^{-3}
	MLP	2.10×10^{-3}	2.14×10^{-3}
Friedman	PSO-NSFM	9.08×10^{-1} ($9.33\times10^{-1}\pm1.61\times10^{-2}$)	1.06 ($1.07\pm3.03\times10^{-2}$)
	SFM	9.64×10^{-1}	1.09
	MLP	1.04	1.08
Gas Furnace	PSO-NSFM	1.70×10^{-1} ($1.80\times10^{-1}\pm3.70\times10^{-3}$)	4.27×10^{-1} ($4.04\times10^{-1}\pm1.65\times10^{-2}$)
	SFM	2.19×10^{-1}	4.44×10^{-1}
	MLP	3.10×10^{-1}	4.60×10^{-1}
House Price-16H	PSO-NSFM	3.62×10^{4} ($3.63\times10^{4}\pm1.24\times10^{2}$)	3.61×10^{4} ($3.59\times10^{4}\pm4.98\times10^{2}$)
	SFM	3.64×10^{4}	3.62×10^{4}
	MLP	3.64×10^{4}	3.63×10^{4}
Machine CPU	PSO-NSFM	2.28×10^{1} ($2.30\times10^{1}\pm4.03\times10^{-1}$)	2.16×10^{1} ($2.63\times10^{1}\pm4.50$)
	SFM	2.49×10^{1}	2.77×10^{1}
	MLP	2.92×10^{1}	2.99×10^{1}
Mackey-Glass	PSO-NSFM	4.79×10^{-2} ($4.81\times10^{-2}\pm 1.48\times10^{-4}$)	5.30×10^{-2} ($5.31\times10^{-2}\pm2.29\times10^{-4}$)
	SFM	4.92×10^{-2}	5.34×10^{-2}
	MLP	4.88×10^{-2}	5.35×10^{-2}
Samad	PSO-NSFM	2.47×10^{-2} ($2.48\times10^{-2}\pm2.11\times10^{-4}$)	2.80×10^{-2} ($2.79\times10^{-2}\pm4.40\times10^{-4}$)
	SFM	2.59×10^{-2}	2.83×10^{-2}
	MLP	2.49×10^{-2}	2.81×10^{-2}

Ο αλγόριθμος PSO-NSFM εκτελέστηκε 30 φορές για κάθε σύνολο δεδομένων. Απεικονίζεται η εκτέλεση για την οποία βρέθηκε το καλύτερο αποτέλεσμα σύμφωνα με τον δείκτη RMSE στο σύνολο αξιολόγησης και ακολουθείται από τον μέσο όρο και την τυπική απόκλιση των 30 εκτελέσεων μέσα σε παρένθεση.

[α]Ο αριθμός των κόμβων των στοιβάδων MLP δίνεται στη μορφή:[1η στοιβάδα 2η στοιβάδα]

Πίνακας 3.3. Αποτελέσματα για όλα τα δύνολα δεδομένων – RMSE, ασαφής διαμέριση και αριθμός κόμβων *(συνέχεια)*

Σύνολο δεδομένων	**Αλγόριθμος**	**Ασαφής διαμέριση**	**Αριθμός κόμβων[α]**
Ailerons	PSO-NSFM	[4 6 4 4 13 7 5 4 14 11 22 32 5 4 4 10 5 37 12 9 4 50 4 24 50 37 48 44 5 50 4 27 6 50 42 6 4 50 9 9]	237 (314±127)
	SFM	13	311
	MLP	-	[35 25]
Auto MPG	PSO-NSFM	[27 4 4 32 4 8]	36 (39±24)
	SFM	10	41
	MLP	-	[23 19]
Auto Price	PSO-NSFM	[4 4 29 50 50 23 7 34 5 10 24 4 35 4 37]	57 (57±5)
	SFM	11	44
	MLP	-	[17 11]
Boston Housing	PSO-NSFM	[48 50 39 4 50 24 4 6 28 45 18 5 11]	138 (141±17)
	SFM	16	146
	MLP	-	[18 19]
CPU Small	PSO-NSFM	[4 7 18 15 16 32 24 29 26 16 35 20]	467 (463±261)
	SFM	23	480
	MLP	-	[24 30]
Elevators	PSO-NSFM	[16 17 5 18 49 17 13 7 9 16 15 11 4 11 5 15 16 12]	981 (805±274)
	SFM	18	1776
	MLP	-	[40 24]
Friedman	PSO-NSFM	[8 4 4 32 4]	114 (116±66)
	SFM	8	179
	MLP	-	[15 15]
Gas Furnace	PSO-NSFM	[8 5 23]	22 (17±5)
	SFM	15	29
	MLP	-	[8 19]
House Price-16H	PSO-NSFM	[5 9 4 26 39 10 6 4 11 10 24 16 13 12 14 30]	984 (902±299)
	SFM	17	1065
	MLP	-	[37 14]
Machine CPU	PSO-NSFM	[12 4 13 6 16 5]	19 (21±6)
	SFM	11	22
	MLP	-	[21 7]
Mackey-Glass	PSO-NSFM	[4 4 4 15 5 43]	24 (25±4)
	SFM	17	34
	MLP	-	[9 13]
Samad	PSO-NSFM	[4 6 38]	89 (87±19)
	SFM	11	112
	MLP	-	[13 7]

Ο αλγόριθμος PSO-NSFM εκτελέστηκε 30 φορές για κάθε σύνολο δεδομένων. Απεικονίζεται η εκτέλεση για την οποία βρέθηκε το καλύτερο αποτέλεσμα σύμφωνα με τον δείκτη RMSE στο σύνολο αξιολόγησης και ακολουθείται από τον μέσο όρο και την τυπική απόκλιση των 30 εκτελέσεων μέσα σε παρένθεση.

[α]Ο αριθμός των κόμβων των στοιβάδων MLP δίνεται στη μορφή:[1η στοιβάδα 2η στοιβάδα]

Πίνακας 3.4. Αποτελέσματα για όλα τα σύνολα δεδομένων – υπολογιστικοί χρόνοι και αριθμός επαναλήψεων

Σύνολο δεδομένων	**Υπολογιστικοί χρόνοι**				**Αριθμός επαναλήψεων**	
	PSO-NSFM (Complete)	**PSO-NSFM (Outperform)**	**SFM**	**MLP**	**PSO-NSFM (Complete)**	**PSO-NSFM (Outperform)**
Ailerons	4220 (7138±3176)	941 (1525±1003)	2613	8141	118 (227±167)	21 (32±21)
Auto MPG	92 (388±113)	4 (6±4)	8	1174	319 (1725±584)	7 (13±9)
Auto Price	339 (357±112)	0.7 (5±2)	6	624	832 (825±295)	1 (9±4)
Boston Housing	2588 (2824±877)	0.8 (7±3)	13	2402	5709 (6677±2118)	1 (7±2)
CPU Small	4505 (2377±1053)	17 (705±289)	790	2814	547 (261±121)	1 (58±25)
Elevators	7197 (8257±2877)	1745 (2587±1201)	3873	10715	166 (194±77)	34 (47±20)
Friedman	133 (436±123)	23 (41±23)	47	1300	245 (949±285)	21 (37±19)
Gas Furnace	77 (35±22)	0.4 (1.5±1)	1.6	1040	620 (250±112)	1 (7±5)
House Price-16H	13201 (11708±3651)	869 (2109±1108)	3508	15745	284 (249±86)	18 (44±21)
Machine CPU	83 (69±32)	1 (3±2)	4	954	943 (698±274)	5 (18±13)
Mackey-Glass	539 (303±218)	1 (7±4)	27	1140	3259 (2124±1807)	3 (11±5)
Samad	60 (104±30)	2 (14±11)	34	1298	89 (178±72)	1 (10±8)

Ο αλγόριθμος PSO-NSFM εκτελέστηκε 30 φορές για κάθε σύνολο δεδομένων. Απεικονίζεται η εκτέλεση για την οποία βρέθηκε το καλύτερο αποτέλεσμα σύμφωνα με τον δείκτη RMSE στο σύνολο αξιολόγησης και ακολουθείται από τον μέσο όρο και την τυπική απόκλιση των 30 εκτελέσεων μέσα σε παρένθεση.
Στην περίπτωση του PSO-NSFM (Complete), το καλύτερο αποτέλεσμα αντιστοιχεί στην εκτέλεση που έδωσε τη μικρότερη τιμή RMSE στο σύνολο δεδομένων αξιολόγησης μέχρι και την ολοκλήρωση του αλγορίθμου. Στην περίπτωση του PSO-NSFM (Outperform), το καλύτερο αποτέλεσμα αντιστοιχεί στην εκτέλεση που χρειάστηκε τον λιγότερο υπολογιστικό χρόνο για να ξεπεράσει τον αλγόριθμο SFM.

3.1.4 Εφαρμογή του αλγορίθμου PSO-NSFM στην αναγνώριση μουσικού είδους – Αποτελέσματα και συζήτηση

Εξαιτίας του μειωμένου κόστους και της αυξανόμενης χωρητικότητας των ηλεκτρονικών μέσων αποθήκευσης, είναι διαθέσιμες τεράστιες ποσότητες ψηφιακών ηχογραφημένων κομματιών μουσικής. Αυτές οι πληροφορίες έχουν συγκεντρωθεί από τα αρχεία ραδιοφωνικών σταθμών και τηλεοπτικών μουσικών καναλιών. Επίσης, αυτές οι πληροφορίες είναι εύκολα προσβάσιμες στο διαδίκτυο μέσω online streaming. Η ανάπτυξη έμπειρων συστημάτων ικανών να ανακτούν αυτόματα σημαντικές πληροφορίες από τα αποθηκευμένα terabytes ακατέργαστων ψηφιακών δεδομένων προσφέρει αρκετά πλεονεκτήματα, όπως ότι

μπορεί να διευκολύνει την οργάνωση και την εκτέλεση αναζητήσεων σε τεράστιες βάσεις μουσικών δεδομένων. Επίσης, άλλο ένα πλεονέκτημα είναι ότι παρέχει ένα χρήσιμο εργαλείο για την ανάλυση των μουσικών τάσεων. Ένα παράδειγμα ουσιαστικής ανάκτησης πληροφοριών είναι η αυτόματη αναγνώριση του μουσικού είδους ενός τραγουδιού, που βασίζεται αποκλειστικά στα αποσπασματικά χαρακτηριστικά που εξάγονται από την ηχογραφημένη κυματομορφή. Αυτή η διαδικασία, είναι γνωστή ως αναγνώριση μουσικού είδους (music genre classification, MGC). Λαμβάνοντας υπόψη παράγοντες, όπως η αφηρημένη φύση της μουσικής, η δυσκολία στην εξαγωγή ουσιαστικών χαρακτηριστικών από τα αρχικά ακατέργαστα δεδομένα, καθώς και το γεγονός ότι ακόμη και οι άνθρωποι ως ακροατές δεν συμφωνούν πάντα σχετικά με το είδος που ανήκει ένα τραγούδι, είναι εύκολα κατανοητό ότι η αναγνώριση μουσικού είδους δεν είναι εύκολο έργο για μία μηχανή.

Στην εφαρμογή αυτή με βάση τον αλγόριθμο PSO-NSFM αναπτύχθηκε ένα ευφυές σύστημα, το οποίο μπορεί να κατατάσσει αυτόματα ένα τραγούδι μουσικής στο είδος που ανήκει. Το σύστημα αυτόματης ταξινόμησης μουσικών ειδών εφαρμόστηκε στο σύνολο μουσικών δεδομένων τύπου benchmark γνωστό ως GTZAN, με σκοπό την ταξινόμηση τραγουδιών σε δέκα διαφορετικά είδη μουσικής.

Το GTZAN είναι ένα σύνολο δεδομένων τύπου benchmark [68], το οποίο είναι ένα από τα πρώτα σύνολα δεδομένων που έχουν καταστεί διαθέσιμα στο κοινό, και θεωρείται ότι είναι το πλέον χρησιμοποιούμενο στην αναγνώριση μουσικού είδους [69]. Αυτό το σύνολο δεδομένων αποτελείται από 1000 αποσπάσματα τραγουδιών των 30 δευτερολέπτων που ταξινομούνται σε 10 διαφορετικά είδη μουσικής και συγκεκριμένα: blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, και rock. Κάθε ένα από αυτά τα είδη μουσικής περιλαμβάνει 100 αποσπάσματα τραγουδιών. Τα αποσπάσματα αυτά λήφθηκαν από διάφορες πηγές, όπως ραδιοφωνικοί σταθμοί, ψηφιακοί δίσκοι και αρχεία ήχου MP3, έτσι ώστε να εξασφαλιστεί η ποικιλία από διαφορετικές ποιότητες εγγραφής.

Κατά την τελευταία δεκαετία, μελέτες που χρησιμοποιούν το σύνολο δεδομένων GTZAN έχουν δείξει [198] ότι υπάρχουν 50 ακριβή αντίγραφα των τραγουδιών. Προφανώς, αν χρησιμοποιηθούν αυτά τα αντίγραφα σε έναν αλγόριθμο μηχανικής μάθησης, τότε αυτός θα οδηγηθεί σε λανθασμένα αποτελέσματα. Σε αυτή την εφαρμογή, τα 50 αντίγραφα διαγράφηκαν και χρησιμοποιήθηκαν τα υπόλοιπα 950 τραγούδια (σύνολο δεδομένων με 950 τραγούδια). Ωστόσο, καθώς αυτά τα ακριβή αντίγραφα χρησιμοποιούνται στη συντριπτική πλειοψηφία των δημοσιευμένων εργασιών, σε αυτή την εφαρμογή περιλαμβάνονται και τα αποτελέσματά σε ολόκληρο το σύνολο δεδομένων (σύνολο δεδομένων με 1000 τραγούδια).

Δεδομένου ότι αυτή η εφαρμογή δεν έχει ως αντικείμενο την εξαγωγή χαρακτηριστικών, χρησιμοποιείται το πρότυπο σύνολο χαρακτηριστικών ηχοχρωματικής υφής (timbral texture) που επίσης έχει χρησιμοποιηθεί στην εργασία [68]. Τα χαρακτηριστικά βασίζονται στον μετασχηματισμό Fourier βραχέως χρόνου (short time Fourier transform, STFT) και υπολογίζονται για κάθε πλαίσιο ήχου μικρής διάρκειας. Περισσότερες λεπτομέρειες σχετικά με τον μετασχηματισμό Fourier βραχέως χρόνου και των συντελεστών cepstral συχνότητας Mel (Mel – frequency cepstral coefficients) μπορούν να βρεθούν στην εργασία [199]. Παρακάτω δίνεται μια σύντομη περιγραφή των χρησιμοποιούμενων χαρακτηριστικών για την αναράσταση ηχοχρωματικής υφής:

- **Spectral Centroid**: Ορίζεται ως το κέντρο βάρους του φασματικού μεγέθους του μετασχηματισμού Fourier βραχέως χρόνου.
- **Spectral Rolloff**: Ορίζεται ως η συχνότητα κάτω από την οποία συγκεντρώνεται το 85% της κατανομής μεγέθους.

- **Spectral Flux**: Ορίζεται ως η τετραγωνική διαφορά ανάμεσα σε κανονικοποιημένα μεγέθη διαδοχικών φασματικών διασπορών.
- **Time Domain Zero Crossings**: Παρέχουν το μέτρο του θορύβου του σήματος.
- **Mel – Frequency Cepstral Coefficients**: Αναφέρονται σε μια σειρά από χαρακτηριστικά τα οποία βασίζονται στον μετασχηματισμό Fourier βραχέως χρόνου, που συχνά χρησιμοποιούνται στην αναγνώριση ομιλίας. Συνήθως στην αναγνώριση ομιλίας χρησιμοποιούνται 13 συντελεστές, αλλά στην παρούσα εφαρμογή χρησιμοποιήθηκαν οι πρώτοι πέντε από αυτούς, καθώς, σύμφωνα με τους Tzanetakis και Cook [68], οι πρώτοι 5 συντελεστές παρέχουν την καλύτερη απόδοση στην αναγνώριση μουσικού είδους.

Τα πραγματικά χαρακτηριστικά που υπολογίζονται για να δοθούν στο σύστημα, είναι οι τρέχοντες μέσοι όροι και τυπικές αποκλίσεις των παραπάνω εξαγόμενων χαρακτηριστικών πάνω σε έναν αριθμό μικρών χρονικών παραθύρων που είναι γνωστά ως παράθυρα ανάλυσης. Τελικά, στο σύστημα δίνεται ένα διάνυσμα ηχοχρωματικής υφής που αποτελείται από 18 μεταβλητές εισόδου.

Ο αλγόριθμος PSO-NSFM χρησιμοποιεί ως κριτήριο σφάλματος το RMSE, αλλά σε αυτή την εφαρμογή γίνεται χρήση του συντελεστή συσχέτισης Matthews πολλών κλάσεων, ο οποίος λειτουργεί ως μέτρο για την αξιολόγηση της απόδοσης του ταξινομητή και περιγράφεται στο κεφάλαιο 1.2.8. Άρα, με βάση το συγκεκριμένο κριτήριο σφάλματος, ο αλγόριθμος PSO-NSFM αντιμετωπίζει ένα πρόβλημα μεγιστοποίησης.

Σε αυτή την εφαρμογή, το κάθε ένα από τα δύο διαθέσιμα σύνολα δεδομένων GTZAN (950 και 1000 τραγούδια) χωρίστηκε με τυχαίο τρόπο στα υποσύνολα εκπαίδευσης, αξιολόγησης και ελέγχου, σε ποσοστά 50% - 25% - 25%, αντίστοιχα. Τα αναφερθέντα ποσοστά χρησιμοποιήθηκαν όχι μόνο σε όλο το σύνολο δεδομένων, αλλά και για κάθε κατηγορία ξεχωριστά. Έτσι, μπορεί να διασφαλιστεί το γεγονός ότι κάθε μουσικό είδος αναπαρίσταται επαρκώς στο σύνολο δεδομένων ελέγχου. Για παράδειγμα, στο σύνολο δεδομένων GTZAN των 950 τραγουδιών, τα δεδομένα χωρίστηκαν σε: 477 για εκπαίδευση, 239 για αξιολόγηση και 234 για έλεγχο.

Ο προτεινόμενος ταξινομητής εκπαιδεύεται στο σύνολο δεδομένων εκπαίδευσης, ενώ το σύνολο δεδομένων αξιολόγησης χρησιμοποιείται για την επιλογή του μοντέλου. Εναλλακτικά, η επιλογή του μοντέλου μπορεί να επιτευχθεί χρησιμοποιώντας τη μέθοδο διασταυρωμένης επικύρωσης *k*-επαναλήψεων (*k*-fold), αν και ο αλγόριθμος θα επιβαρυνθεί με επιπλέον υπολογιστικό φορτίο. Καθώς η μέθοδος διασταυρωμένης επικύρωσης *k*-επαναλήψεων είναι από μόνη της μια υπολογιστικά δαπανηρή τεχνική, η προσθήκη της στον αλγόριθμο PSO μπορεί να σημαίνει την επανάληψη της διαδικασίας εκπαίδευσης του δικτύου RBF πολλές φορές για κάθε σωματίδιο και για κάθε επανάληψη. Ανεξαρτήτως της διαδικασίας επιλογής του μοντέλου, ένα τρίτο ανεξάρτητο σύνολο δεδομένων, που χρησιμοποιείται μόνο στον έλεγχο του παραγόμενου δικτύου RBF, είναι απαραίτητο. Η διαδικασία της βελτιστοποίησης μπορεί να οδηγήσει στο φαινόμενο της υπερπροσαρμογής του μοντέλου (το οποίο αναλύθηκε στο κεφάλαιο 1.2.7), ακόμα και στην περίπτωση της μεθόδου διασταυρωμένης επικύρωσης *k* επαναλήψεων.

Για λόγους σύγκρισης, δοκιμάστηκαν δύο επιπλέον μέθοδοι, οι ταξινόμητες RBF που εκπαιδεύτηκαν με τη συμμετρική εκδοχή του αλγορίθμου των ασαφών μέσων (SFM) [34] και οι ταξινομητές MLP που εκπαιδεύτηκαν με τον αλγόριθμο Levenberg – Marquardt [2]. Οι παράμετροι των παραπάνω δύο ταξινομητών, προσδιορίστηκαν χρησιμοποιώντας το σύνολο δεδομένων εκπαίδευσης και στη συνέχεια, με βάση το σύνολο δεδομένων αξιολόγησης, επιλέχτηκε το μοντέλο. Στην περίπτωση των ταξινομητών RBF που εκπαιδεύονται με τον αλγόριθμο SFM, το μοντέλο επιλέγεται με βάση μία μόνο παράμετρο, τον αριθμό των

ασαφών συνόλων *s* σύμφωνα με τον οποίο διαμερίζεται το πεδίο ορισμού κάθε μεταβλητής εισόδου. Η επιλογή του αριθμού των ασαφών συνόλων *s* έγινε αφού δοκιμάστηκαν όλες οι διαμερίσεις από 4 έως 50 ασαφή σύνολα. Όσον αφορά στα δίκτυα MLP, χρησιμοποιήθηκε μία αρχιτεκτονική που αποτελείται από δύο κρυφές στοιβάδες και στη συνέχεια εκτελέστηκε μία διεξοδική αναζήτηση, δοκιμάζοντας όλους τους πιθανούς συνδυασμούς, όπου ο αριθμός των κόμβων της πρώτης κρυφής στοιβάδας κυμαίνεται από 2 έως 30 και της δεύτερης από 2 έως 20.

Ο πίνακας 3.5 παρουσιάζει τις παραμέτρους σχεδιασμού του αλγορίθμου PSO-NSFM. Οι παράμετροι αυτές, ρυθμίστηκαν ειδικά για το σύνολο δεδομένων GTZAN σύμφωνα με τις κατευθυντήριες γραμμές της εργασίας [67]. Επίσης, στον πίνακα 3.6 συνοψίζονται τα αποτελέσματα για τα σύνολα δεδομένων με τα 950 και 1000 τραγούδια, όπου απεικονίζονται οι δείκτες MCC και ακρίβεια% στα σύνολα δεδομένων αξιολόγησης και ελέγχου, καθώς και ο αριθμός των κόμβων για τις τρεις μεθοδολογίες, μαζί με τον αριθμό των ασαφών συνόλων για τα δύο μοντέλα RBF. Ο αριθμός των κόμβων των νευρωνικών δικτύων δίνεται στη μορφή: [1η κρυφή στοιβάδα, 2η κρυφή στοιβάδα]. Επίσης, με έντονους χαρακτήρες δηλώνεται η καλύτερη απόδοση ανά σύνολο δεδομένων, για καθένα από τα κριτήρια αξιολόγησης. Καθώς ο αλγόριθμος PSO-NSFM και η μέθοδος εκπαίδευσης των δικτύων MLP περιλαμβάνουν στοχαστικά στοιχεία, οι μέθοδοι αυτές εκτελέστηκαν 20 φορές για κάθε μία από τις δύο περιπτώσεις. Στον πίνακα 3.6 απεικονίζεται το καλύτερο αποτέλεσμα που βρέθηκε σύμφωνα με τον δείκτη MCC, ενώ, στις παρενθέσεις φαίνονται οι τιμές του μέσου όρου και της τυπικής απόκλισης από τις 20 εκτελέσεις. Ο αλγόριθμος SFM, είναι μία ντετερμινιστική διαδικασία, συνεπώς δεν υπάρχει λόγος να εκτελεστεί παρά μόνο μια φορά.

Η σύγκριση μεταξύ των τριών ταξινομητών που δοκιμάστηκαν σε αυτή την εφαρμογή, δείχνει ότι και στα δύο σύνολα δεδομένων που αποτελούνται από 950 και 1000 τραγούδια, η καλύτερη λύση βρέθηκε από τον αλγόριθμο PSO-NSFM. Επίσης, ο τελευταίος υπερτερεί έναντι του αλγορίθμου SFM αναφορικά με το καλύτερο MCC και την καλύτερη ακρίβεια% που υπολογίστηκαν βάσει του συνόλου ελέγχου. Η υπεροχή του αλγορίθμου PSO-NSFM είναι ακόμα περισσότερο εμφανής όσον αφορά στα αποτελέσματα του συνόλου δεδομένων αξιολόγησης. Παρόλα αυτά, το σύνολο δεδομένων αξιολόγησης, δεν είναι κατάλληλο για να εξαχθούν ασφαλή συμπεράσματα, καθώς χρησιμοποιείται στη διαδικασία επιλογής του μοντέλου. Επίσης, ο αλγόριθμος PSO-NSFM υπερτερεί έναντι της μεθόδου SFM, όσον αφορά στη σύγκριση μεταξύ της μέσης λύσης των 20 εκτελέσεων του πρώτου, με την καλύτερη πιθανή λύση που μπορεί να πετύχει ο ταξινομητής SFM.

Πίνακας 3.5. Λειτουργικές παράμετροι για τον αλγόριθμο PSO-NSFM

Παράμετρος	**Σύμβολο**	**Τιμή**
Ελάχιστος αριθμός ασαφών συνόλων	s_{min}	4
Μέγιστος αριθμός ασαφών συνόλων	s_{max}	50
Πληθυσμός	P	20
Νοσταλγία	c_1	0.5
Ζήλεια	c_2	0.5
Σταθερά Velocity clamping	V_{max}	15
Μέγιστος αριθμός βημάτων προσομοίωσης	ξ_1	2000

Πίνακας 3.6. Αποτελέσματα για τα σύνολα δεδομένων GTZAN – MCC, ακρίβεια%, ασαφής διαμέριση και αριθμός κόμβων

Σύνολο δεδομένων	**Αλγόριθμος**	**MCC συνόλου αξιολόγησης**	**MCC συνόλου ελέγχου**	**Ακρίβεια% συνόλου αξιολόγησης**	**Ακρίβεια% συνόλου ελέγχου**
950 τραγούδια	PSO-NSFM	**0.76** (0.74 ± 0.01)	**0.72** (0.69 ± 0.01)	**77.82** (76.61 ± 0.61)	**74.79** (72.32 ± 0.95)
	SFM	0.67	0.69	70.29	71.79
	MLP	0.67 (0.61 ± 0.06)	0.60 (0.59 ± 0.06)	70.29 (64.73 ± 5.74)	64.10 (62.57 ± 5.92)
1000 τραγούδια	PSO-NSFM	**0.78** (0.77 ± 0.02)	**0.72** (0.68 ± 0.02)	**80.40** (79.43 ± 1.60)	**74.40** (70.77 ± 1.65)
	SFM	0.72	0.67	74.40	70.40
	MLP	0.71 (0.64 ± 0.08)	0.65 (0.60 ± 0.07)	74.00 (67.19 ± 7.61)	68.00 (63.27 ± 7.05)

Ο αλγόριθμος PSO-NSFM εκτελέστηκε 20 φορές για κάθε περίπτωση. Με έντονους αριθμούς δηλώνεται η καλύτερη απόδοση ανά σύνολο δεδομένων, για καθένα από τα κριτήρια αξιολόγησης.
[α]Ο αριθμός των κόμβων των στοιβάδων MLP δίνεται στη μορφή:[1η στοιβάδα 2η στοιβάδα]

Πρέπει σημειωθεί ότι ο αλγόριθμος PSO-NSFM προβλέπει καλύτερα στο σύνολο δεδομένων αξιολόγησης από ότι στο σύνολο ελέγχου. Αυτό μπορεί να οφείλεται στο φαινόμενο της υπερπροσαρμογής στο σύνολο δεδομένων αξιολόγησης. Επομένως, μπορεί να υποτεθεί ότι η τεχνική διασταυρωμένης επικύρωσης k επαναλήψεων είναι δυνατό να οδηγήσει σε καλύτερη επιλογή μοντέλου και ακολούθως σε υψηλότερη ακρίβεια στο σύνολο δεδομένων ελέγχου, παρόλο τον αυξημένο υπολογιστικό χρόνο που απαιτείται. Τέλος, οι τιμές της τυπικής απόκλισης των δεικτών MCC και ακρίβεια% είναι σχετικά μικρές, επιβεβαιώνοντας το γεγονός ότι ο αλγόριθμος PSO-NSFM καταλήγει σε συνεπή αποτελέσματα. Όσον αφορά στα δίκτυα MLP, ολοφάνερα κατατάσσονται τρίτα καθώς τα αποτελέσματα τους είναι χειρότερα σχεδόν σε όλες τις κατηγορίες.

Μια σύγκριση μεταξύ των συνόλων δεδομένων με τα 950 και τα 1000 τραγούδια δείχνει ότι και οι τρεις αλγόριθμοι τείνουν να παράγουν καλύτερα αποτελέσματα στο γνήσιο σύνολο δεδομένων, αναφορικά με τα σύνολα αξιολόγησης. Ωστόσο, όσον αφορά στα δεδομένα ελέγχου, όλοι οι αλγόριθμοι παράγουν καλύτερα αποτελέσματα στο σύνολο δεδομένων με τα λιγότερα τραγούδια. Αυτό μπορεί να αιτιολογηθεί αν ληφθεί υπόψη το γεγονός ότι το σύνολο δεδομένων με τα 1000 τραγούδια περιλαμβάνει αντίγραφα. Αυτά τα αντίγραφα βοηθούν τους αλγόριθμους να αποδώσουν καλύτερα στα δεδομένα αξιολόγησης, αλλά αυτό είναι αποτέλεσμα του φαινομένου της υπερπροσαρμογής. Έτσι, όταν οι ταξινομητές προσπαθούν να κάνουν προβλέψεις για το σύνολο ελέγχου, το οποίο είναι εντελώς ανεξάρτητο από τη διαδικασία της εκπαίδευσης, καταλήγουν να αποδίδουν χειρότερα σε σύγκριση με το σύνολο δεδομένων των 950 τραγουδιών.

Μια καλύτερη εικόνα των αποτελεσμάτων του αλγορίθμου PSO-NSFM για τα σύνολα δεδομένων με τα 950 και τα 1000 τραγούδια, δίνουν οι πίνακες σύγχυσης των σχημάτων 3.3 και 3.4, αντίστοιχα. Αυτοί οι πίνακες σύγχυσης παρουσιάζουν την καλύτερη λύση και τους μέσους όρους των 20 εκτελέσεων στα σύνολα δεδομένων ελέγχου. Παρατηρώντας τους πίνακες σύγχυσης, είναι εμφανές ότι ο αλγόριθμος PSO-NSFM, όχι μόνο κατορθώνει ένα ικανοποιητικό συνολικό ποσοστό επιτυχίας, αλλά επίσης προβλέπει με

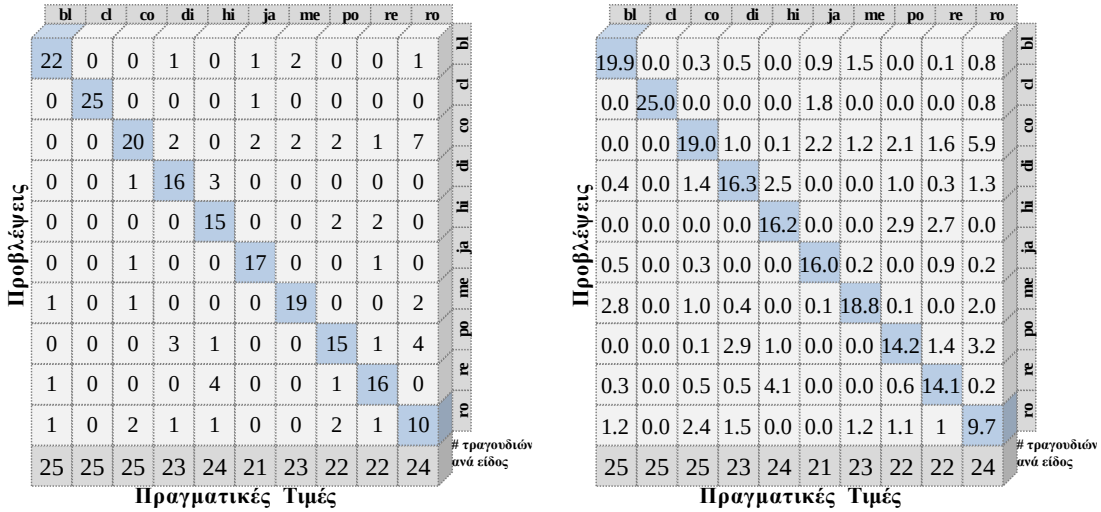
Πίνακας 3.6. Αποτελέσματα για τα σύνολα δεδομένων GTZAN – MCC, ακρίβεια%, ασαφής διαμέριση και αριθμός κόμβων *(συνέχεια)*

Σύνολο δεδομένων	**Αλγόριθμος**	**Ασαφής διαμέριση**	**Αριθμός κόμβων[α]**
950 τραγούδια	PSO-NSFM	[4 4 50 9 4 33 8 4 4 50 4 4 4 4 4 22 4 4]	108 (104 ± 17)
	SFM	13	200
	MLP	-	[25 17]
1000 τραγούδια	PSO-NSFM	[50 37 50 4 4 18 4 4 10 4 50 5 4 4 7 9 7 7]	131 (129 ± 100)
	SFM	10	92
	MLP	-	[21 20]

Ο αλγόριθμος PSO-NSFM εκτελέστηκε 20 φορές για κάθε περίπτωση. Με έντονους αριθμούς δηλώνεται η καλύτερη απόδοση ανά σύνολο δεδομένων, για καθένα από τα κριτήρια αξιολόγησης.
[α]Ο αριθμός των κόμβων των στοιβάδων MLP δίνεται στη μορφή:[1η στοιβάδα 2η στοιβάδα]

καλή απόδοση κάθε μουσικό είδος ξεχωριστά. Τη μόνη εξαίρεση, αποτελεί το είδος μουσικής rock, το οποίο μπερδεύεται με τα είδη μουσικής pop, metal και country. Ωστόσο, πρέπει να σημειωθεί ότι για την rock μουσική επιτυγχάνεται η χειρότερη ακρίβεια ταξινόμησης στις περισσότερες δημοσιευμένες εργασίες που χρησιμοποιείται το σύνολο δεδομένων GTZAN. Ακόμα πιο σημαντικό, είναι το γεγονός ότι ακόμα και οι άνθρωποι ως ακροατές συχνά διαφωνούν μεταξύ τους σχετικά με τη διάκριση ενός τραγουδιού rock με άλλα είδη, όπως pop, metal και country.

Η καλή απόδοση πρόβλεψης που εμφανίζει ο αλγόριθμος PSO-NSFM μπορεί να αποδοθεί στην αξιοποίηση του συντελεστή MCC ως συνάρτηση κόστους του προβλήματος βελτιστοποίησης που λύνεται με τη μέθοδο PSO. Τέλος, θα πρέπει να επισημανθεί ότι αυτά τα αποτελέσματα επιτεύχθηκαν χρησιμοποιώντας μόνο ένα τυπικό σετ των χαρακτηριστικών timbral texture. Αν συμπεριληφθούν πρόσθετες μεταβλητές εισόδου, π.χ. με βάση τα ρυθμικά



Σχήμα 3.3. Πίνακες σύγχυσης για τον αλγόριθμο PSO-NSFM στο σύνολο δεδομένων ελέγχου με τα 950 τραγούδια: (α) Το μοντέλο της εκτέλεσης που έδωσε το καλύτερο MCC στο σύνολο αξιολόγησης, (β) Μέσος όρος 20 εκτελέσεων.

Προβλέψεις \ Πραγματικές Τιμές	bl	cl	co	di	hi	ja	me	po	re	ro	# τραγουδιών ανά είδος
bl	21	0	0	0	0	1	2	0	0	4	
cl	0	24	0	0	0	2	0	0	0	1	
co	1	0	14	0	0	1	0	2	1	3	
di	1	0	3	19	2	0	0	1	1	1	
hi	0	0	1	0	16	0	0	0	1	1	
ja	0	1	2	1	0	21	1	0	0	0	
me	1	0	1	0	0	0	22	0	0	0	
po	0	0	1	4	3	0	0	20	2	6	
re	0	0	1	1	4	0	0	0	20	0	
ro	1	0	2	0	0	0	0	2	0	9	
# τραγουδιών ανά είδος	25	25	25	25	25	25	25	25	25	25	

(α)

Προβλέψεις \ Πραγματικές Τιμές	bl	cl	co	di	hi	ja	me	po	re	ro	# τραγουδιών ανά είδος
bl	18.9	0.0	1.6	1.9	0.2	0.5	1.2	0.0	0.6	0.1	
cl	0.0	24.4	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.1	
co	0.8	0.5	13.7	4.1	0.8	1.7	1.7	0.5	0.7	0.5	
di	0.2	0.0	0.1	17.5	0.8	0.2	0.0	4.1	1.2	0.9	
hi	0.1	0.0	0.0	1.9	16.7	0.0	0.2	1.9	4.1	0.1	
ja	1.5	2.4	0.5	0.0	0.0	19.9	0.3	0.0	0.2	0.3	
me	1.8	0.0	0.0	0.2	0.1	0.3	22.1	0.0	0.0	0.5	
po	0.1	0.0	1.6	1.6	0.7	0.0	0.0	19.7	0.1	1.3	
re	0.1	0.0	0.4	0.8	3.9	0.6	0.2	2.2	16.4	0.3	
ro	3.4	0.1	4.3	2.0	1.0	1.5	0.3	4.4	0.5	7.6	
# τραγουδιών ανά είδος	25	25	25	25	25	25	25	25	25	25	

(β)

Σχήμα 3.4. Πίνακες σύγχυσης για τον αλγόριθμο PSO-NSFM στο σύνολο δεδομένων ελέγχου με τα 1000 τραγούδια: (α) Το μοντέλο της εκτέλεσης που έδωσε το καλύτερο MCC στο σύνολο αξιολόγησης, (β) Μέσος όρος 20 εκτελέσεων.

και/ή χαρακτηριστικά τονικού ύψους [1], αποδεικνύεται ότι ενισχύεται η απόδοση της ταξινόμησης. Επομένως, είναι αναμενόμενο ότι τα αποτελέσματα μπορούν να βελτιωθούν περαιτέρω χρησιμοποιώντας επιπλέον χαρακτηριστικά, σε συνδυασμό με τον αλγόριθμο PSO-NSFM.

3.2 Εκπαίδευση νευρωνικών δικτύων RBF με συνεργατική βελτιστοποίηση σμήνους σωματιδίων

Ο αλγόριθμος PSO-NSFM που αναλύθηκε στο προηγούμενο κεφάλαιο παρουσιάζει σημαντικά πλεονεκτήματα, όπως μεγάλη ακρίβεια πρόβλεψης, γρήγορους υπολογιστικούς χρόνους και απλούστερες δομές δικτύων. Από την άλλη πλευρά, όμως, ο αλγόριθμος PSO-NSFM παράγει δίκτυα RBF που οι συναρτήσεις ενεργοποίησής τους έχουν σταθερό εύρος. Το γεγονός αυτό αποτελεί έναν σημαντικό περιορισμό, διότι ο αλγόριθμος PSO-NSFM δεν μπορεί να εφαρμόσει κάποιες ευρέως γνωστές συναρτήσεις RBF, όπως η συνάρτηση Gauss. Η χρήση των συναρτήσεων βάσης σταθερού εύρους περιορίζει την ευελιξία της πολυδιάστατης επιφάνειας που δημιουργεί το δίκτυο RBF, ενδεχομένως μειώνοντας την ακρίβεια του παραγόμενου μοντέλου ή/και απαιτώντας περισσότερα κέντρα για να επιτευχθούν ικανοποιητικά αποτελέσματα.

Για την εξάλειψη των μειονεκτημάτων του αλγορίθμου PSO-NSFM, προτάθηκε μία νέα μέθοδος, η οποία χρησιμοποιεί συναρτήσεις μεταβλητού πλάτους και βασίζεται σε ένα συνεργατικό πλαίσιο εξελικτικού υπολογισμού. Συγκεκριμένα, η προτεινόμενη μέθοδος, βασίζεται σε μια συνεργατική τεχνική PSO με μεταβλητού εύρους συναρτήσεις βάσης (cooperative PSO with variable width basis functions, COOP-VW). Η μέθοδος COOP-VW βελτιστοποιεί πλήρως τα δίκτυα RBF και μπορεί να εφαρμόσει οποιαδήποτε συνάρτηση ενεργοποίησης που έχει ακτινική συμμετρία. Επίσης, η συγκεκριμένη μέθοδος υιοθετεί ένα συνεργατικό αλγόριθμο βελτιστοποίησης σμήνους σωματιδίων (cooperative PSO, CPSO), ο

οποίος χρησιμοποιεί πολλαπλά σμήνη ώστε να βελτιστοποιεί διάφορες συνιστώσες του διανύσματος λύσης [80].

Η προτεινόμενη μέθοδος χρησιμοποιεί δύο συνεργατικά σμήνη, τα οποία ρυθμίζουν διαφορετικά είδη παραμέτρων του δικτύου RBF, αλλά τελικά εργάζονται από κοινού με στόχο την εκπαίδευση δικτύων με βελτιωμένες ιδιότητες. Σε αντίθεση με τις περισσότερες από τις υπάρχουσες μεθόδους εξελικτικού υπολογισμού που βελτιστοποιούν πλήρως το δίκτυο, ο αλγόριθμος COOP-VW εισάγει μια πιο συνεπτυγμένη κωδικοποίηση του προβλήματος βελτιστοποίησης με λιγότερες μεταβλητές σχεδιασμού, η οποία βοηθά την τεχνική υπολογιστικής νοημοσύνης που υιοθετήθηκε ώστε να βρει βελτιωμένες λύσεις, οδηγώντας τελικά σε μοντέλα με μεγαλύτερη ακρίβεια πρόβλεψης. Αυτό επιτυγχάνεται εισάγοντας μια τροποποιημένη τεχνική *p*-nearest neighbor [11], με σκοπό αυτή η τεχνική να μπορεί να χρησιμοποιηθεί για να κωδικοποιηθούν τα εύρη των συναρτήσεων των δικτύων RBF εντός του πλαισίου της μεθόδου νοημοσύνης σμήνους.

3.2.1 Ενσωμάτωση μεταβλητού εύρους

Καθώς ο αλγόριθμος PSO-NSFM δεν ενσωματώνει κάποιο μηχανισμό για να βελτιστοποιηθούν τα εύρη των συναρτήσεων βάσης, μπορεί να λειτουργήσει μόνο με συναρτήσεις βάσης σταθερού εύρους, όπως είναι η συνάρτηση TPS. Αυτός ο περιορισμός στερεί από τη μεθοδολογία την πρόσθετη ευελιξία που θα μπορούσε ενδεχομένως να βοηθήσει στο να προσαρμοστεί καλύτερα η πολυδιάστατη επιφάνεια που δημιουργεί το δίκτυο RBF, έτσι ώστε να προσεγγιστούν καλύτερα τα παραδείγματα εκπαίδευσης.

Ο αλγόριθμος PSO-NSFM σε έναν χώρο N διαστάσεων (μεταβλητών εισόδου) χρησιμοποιεί ένα σμήνος P σωματιδίων. Την επανάληψη t, ο αλγόριθμος χρησιμοποιεί ένα είδος κωδικοποίησης όπου το κάθε στοιχείο j του κάθε σωματιδίου i, περιέχει τον αριθμό των ασαφών συνόλων s_j σε κάθε διάσταση j=1,…,N. Το σωματίδιο $\mathbf{s}_i(t)$ περιγράφεται από την εξίσωση (3.1).

Μια απλοϊκή τεχνική, προκειμένου να βελτιστοποιηθούν τα εύρη των συναρτήσεων βάσης σ, θα μπορούσε απλά να προσθέτει τα εύρη ως επιπλέον στοιχεία σε ένα σωματίδιο $\mathbf{\Lambda}_i$ της μορφής:

$$\mathbf{\Lambda}_i(t) = \left[\underbrace{s_1(t),\ \ldots,\ s_N(t)}_{\text{Ασαφής Διαμέριση}}, \underbrace{\sigma_1(t),\ \ldots,\ \sigma_{L_i(t)}(t)}_{\text{Εύρη RBF}} \right]^T,\ i = 1,\ldots,P \tag{3.11}$$

όπου $L_i(t)$ είναι ο αριθμός των κέντρων που επιλέχθηκαν από τον αλγόριθμο NSFM (ο οποίος αναλύεται στο κεφάλαιο 2.2.3) για το δίκτυο RBF που αντιστοιχεί στο i-οστό σωματίδιο την επανάληψη t. Βέβαια, κάτι τέτοιο οδηγεί στο να αποτελούνται τα σωματίδια του ίδιου πληθυσμού από διαφορετικό αριθμό διαστάσεων, αλλά οδηγεί και σε σωματίδια τα οποία ανά επανάληψη αλλάζουν μέγεθος. Κλασσικές τεχνικές που βασίζονται στον αλγόριθμο PSO δεν μπορούν να διαχειριστούν τέτοιου είδους καταστάσεις, αλλά ακόμη και αν μπορούσε να γίνει κάποια κατάλληλη τροποποίηση, ο αριθμός των μεταβλητών σχεδιασμού, όσον αφορά στο προκύπτον πρόβλημα βελτιστοποίησης, θα αυξανόταν σημαντικά.

Προκειμένου να εξαληφθεί το παραπάνω εμπόδιο, χρησιμοποιήθηκε μια ευρετική τεχνική γνωστή ως *p*-nearest neighbor [11], η οποία τροποποιείται κατάλληλα έτσι ώστε να

εφαρμοστεί παράλληλα με μεθόδους εξελικτικού υπολογισμού. Ο πρότυπος αλγόριθμος *p*-nearest neighbor επιλέγει το εύρος κάθε συνάρτησης βάσης $\sigma_j(t)$ ως τη μέση τετραγωνική ρίζα της απόστασης των *p* πλησιέστερων γειτόνων, χρησιμοποιώντας την εξίσωση (2.5). Στην προκειμένη περίπτωση ο αλγόριθμος *p*-nearest neighbor τροποποιείται έτσι ώστε να υπολογίζει το εύρος κάθε συνάρτησης βάσης $\sigma_j(t)$ ως τη μέση τετραγωνική ρίζα της απόστασης των *p* πλησιέστερων κέντρων ως εξής:

$$\sigma_j(t) = \sqrt{\frac{1}{P}\sum_{k=1}^{p} \| \hat{\mathbf{u}}_k - \hat{\mathbf{u}}_j \|^2}, \; j = 1, 2, ..., L_i(t) \tag{3.12}$$

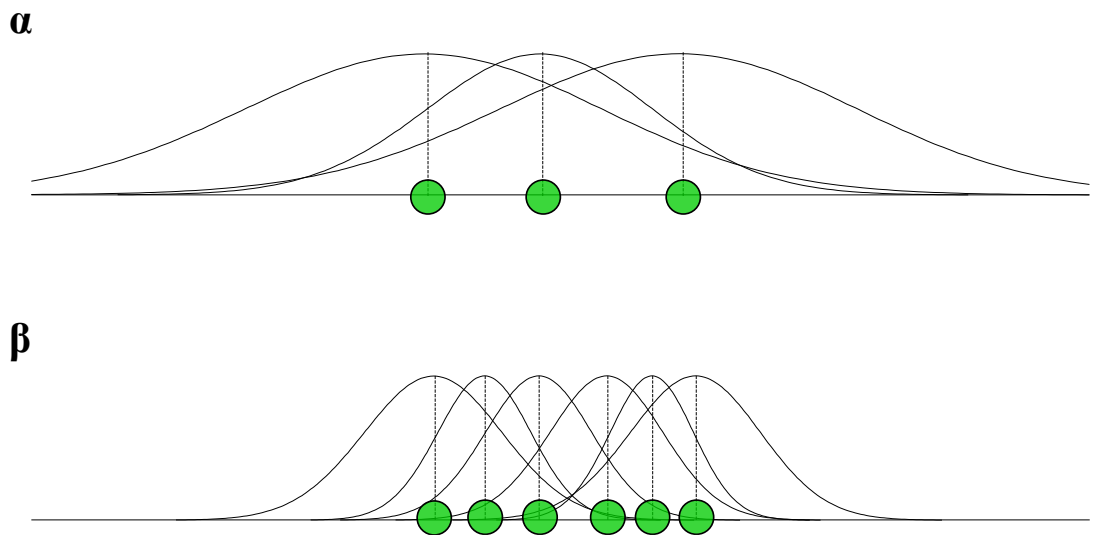
όπου $\hat{\mathbf{u}}_k$ είναι τα *p* πλησιέστερα κέντρα στο κέντρο $\hat{\mathbf{u}}_j$. Η υιοθέτηση αυτής της ευρετικής μεθόδου επιτρέπει την αντικατάσταση των εύρων όλων των επιμέρους κέντρων RBF στο σωματίδιο, με μία μόνο παράμετρο. Έτσι, όχι μόνο είναι δυνατό να διατηρηθεί σταθερή η διάσταση ανά σωματίδιο, αλλά και ανά επανάληψη. Επίσης, οι επιπλέον μεταβλητές σχεδιασμού που ρυθμίζουν τα εύρη, μειώνονται σε μία μόνο μεταβλητή.

Ωστόσο, απαιτεί ιδιαίτερη προσοχή η επιλογή της παραμέτρου με την οποία ελέγχονται τα εύρη των συναρτήσεων βάσης RBF. Η συγκεκριμένη παράμετρος εισάγεται ως πρόσθετη διάσταση σε κάθε σωματίδιο. Αν και μπορεί να φαίνεται ότι ο αριθμός των *p* πλησιέστερων κέντρων είναι ο πλέον κατάλληλος, μια τέτοια επιλογή δεν μπορεί να χρησιμοποιηθεί σε όλες τις περιπτώσεις. Ο λόγος είναι ότι το *p* χαρακτηρίζει τα εύρη RBF μόνο έμμεσα, μέσω της κάλυψης των πλησιέστερων κέντρων. Τα παραγόμενα εύρη όμως, εξαρτώνται και από τον αριθμό των κέντρων. Αυτό σημαίνει ότι η ίδια τιμή για το *p* θα μπορούσε να οδηγήσει σε διαφορετικά αποτελέσματα εάν επιλεγεί διαφορετικός αριθμός κέντρων, με βάση μία διαφορετική ασαφή διαμέριση. Αυτό είναι σημαντικό, διότι, καθώς εκτυλίσσεται η εξελικτική διαδικασία PSO, ο αριθμός των κέντρων που αντιστοιχεί στο δίκτυο που δημιουργεί κάθε σωματίδιο μπορεί να διαφέρει σημαντικά, μην επιτρέποντας έτσι στον αλγόριθμο να ανταλλάσσει κατάλληλες πληροφορίες μεταξύ των διαφόρων επαναλήψεων. Για παράδειγμα, ο υπολογισμός της ταχύτητας απαιτεί τη σύγκριση μεταξύ των τρεχουσών θέσεων των σωματιδίων, με τις καλύτερες προσωπικές θέσεις τους, οι οποίες ενδεχομένως υπολογίζονται σε μία προηγούμενη επανάληψη. Εάν ο αριθμός των κέντρων τυχαίνει να έχει σημαντική διαφορά μεταξύ αυτών των δύο επαναλήψεων, τότε η τιμή του *p* που αντιστοιχεί στις καλύτερες θέσεις των σωματιδίων, δεν σχετίζεται με τις τρέχουσες θέσεις των σωματιδίων. Ένα οπτικό παράδειγμα μιας μονοδιάστατης εισόδου δίνεται στο σχήμα 3.5, όπου φαίνεται ότι η τιμή *p*=2 κάνει τις συναρτήσεις Gauss ιδιαίτερα πλατιές στην περίπτωση των τριών κέντρων RBF, σε σύγκριση με την περίπτωση των έξι κέντρων RBF.

Προκειμένου τα εύρη των συναρτήσεων να μην εξαρτώνται από τον αριθμό των κέντρων, και επομένως από την ασαφή διαμέριση, ορίζουμε το ποσοστό κάλυψης των πλησιέστερων κέντρων κ_i για κάθε σωματίδιο:

$$\kappa_i(t) = \frac{p_i(t)}{L_i(t)} \cdot 100 \tag{3.13}$$

Κανονικοποιώντας τον αριθμό των πλησιέστερων κέντρων p_i, με το σύνολο των κέντρων RBF ανά σωματίδιο L_i, δημιουργείται μια παράμετρος η οποία μπορεί να ελέγχει τα εύρη RBF, ενώ παράλληλα επιτρέπει την ανταλλαγή πληροφοριών μεταξύ των σωματιδίων



Σχήμα 3.5. Υπολογισμός ευρών για p=2 σε ένα μονοδιάστατο χώρο εισόδου που περιέχει (α) 3 κέντρα RBF και (b) 6 κέντρα RBF.

ακόμα και με διαφορετικό αριθμό κέντρων. Έτσι, τα σωματίδια παίρνουν την ακόλουθη μορφή:

$$\mathbf{\Lambda}_i(t) = \left[\underbrace{s_1(t),\ \ldots,\ s_N(t),}_{\text{Ασαφής Διαμέριση}} \quad \underset{\text{Κάλυψη Πλησιέστερων Κέντρων}}{\kappa_i(t)} \right]^T \tag{3.14}$$

3.2.2 Συνεργατικός αλγόριθμος PSO για πλήρη εκπαίδευση δικτύων RBF

Η δομή των σωματιδίων που υιοθετεί ο αλγόριθμος PSO-NSFM θα μπορούσε εύκολα να τροποποιηθεί για να ενσωματώσει σωματίδια της μορφής που περιγράφονται από τη σχέση (3.14), επιτυγχάνοντας έτσι την ταυτόχρονη βελτιστοποίηση των κέντρων και των ευρών των συναρτήσεων, ενώ διατηρείται ένας σχετικά χαμηλός αριθμός μεταβλητών σχεδίασης, που είναι ίσος με N+1. Ωστόσο, μπορεί κανείς να διακρίνει δύο ξεχωριστά μέρη στο σωματίδιο, τα οποία ελέγχουν ένα διαφορετικό χαρακτηριστικό του παραγόμενου δικτύου RBF: Το πρώτο μέρος, το οποίο περιέχει την ασαφή διαμέριση, διαχειρίζεται τον υπολογισμό των κέντρων, ενώ το δεύτερο μέρος, που περιέχει την κάλυψη των πλησιέστερων κέντρων, ελέγχει τα εύρη των συναρτήσεων Gauss. Στη συνέχεια, τα δύο σμήνη μπορούν να εξελιχθούν ξεχωριστά, βελτιώνοντας τις επιδόσεις τους μέσω μιας συνεργατικής διαδικασίας.

Η βελτίωση της διαδικασίας της βελτιστοποίησης χρησιμοποιώντας συνεργατικούς πράκτορες (agents) είναι μια ισχυρή τεχνική και έχει εφαρμοστεί με επιτυχία σε μεθόδους γενετικών αλγορίθμων [200], διαχωρίζοντας το διάνυσμα λύσης σε πολλαπλά χρωμοσώματα που ανήκουν σε διαφορετικούς πληθυσμούς. Μια παρόμοια ιδέα έχει επίσης χρησιμοποιηθεί στα πλαίσια του αλγορίθμου PSO, όπου το διάνυσμα λύσης χωρίζεται σε έναν αριθμό σωματιδίων τα οποία ανήκουν σε διαφορετικά σμήνη και εξελίσσονται αυτόνομα, αλλά

ανταλλάσσουν πληροφορίες κατά τη διάρκεια της διαδικασίας. Η πρώτη διαφορά με τον συνεργατικό αλγόριθμο PSO [201] είναι ότι αποσυνθέτει ένα διάνυσμα λύσης που αποτελείται από N_d στοιχεία, σε ακριβώς N_d μονοδιάστατα σωματίδια. Ένας νεότερος αλγόριθμος, γνωστός ως CPSO-S_K [80], χωρίζει το διάνυσμα λύσης σε K τμήματα με $K \leq N_d$. Βέβαια, είναι πιθανό ότι κάθε ένα από τα παραχθέντα τμήματα έχει και μια διαφορετική διάσταση.

Σε αυτή τη διατριβή, ο αλγόριθμος CPSO-S_K προσαρμόζεται κατάλληλα για να αντιμετωπίζει το πρόβλημα της παράλληλης βελτιστοποίησης των κέντρων και των ευρών των συναρτήσεων σε ένα συνεργατικό πλαίσιο μάθησης, ώστε να εκπαιδεύονται δίκτυα RBF. Με βάση το διάνυσμα λύσης, το οποίο περιλαμβάνει δύο διακριτά τμήματα , όπως φαίνεται στη σχέση (3.14), η υιοθέτηση δύο σμηνών είναι απόλυτα λογική επιλογή. Τα σωματίδια που ανήκουν στο πρώτο σμήνος το οποίο κωδικοποιεί την ασαφή διαμέριση συμβολίζονται ως $P_1.\mathbf{\Lambda}_i$, ενώ τα σωματίδια που ανήκουν στο δεύτερο σμήνος, συμβολίζονται ως $P_2.\mathbf{\Lambda}_i$:

$$P_1.\mathbf{\Lambda}_i(t) = \left[s_1(t),\ s_2(t),\ \ldots,\ s_N(t) \right]^T \tag{3.15}$$

$$P_2.\mathbf{\Lambda}_i(t) = \kappa_i(t) \tag{3.16}$$

Θα πρέπει να σημειωθεί ότι το σμήνος $P_1.\mathbf{\Lambda}_i$ θα μπορούσε να διαχωριστεί περαιτέρω σε μονοδιάστατα σωματίδια, όπου κάθε ένα σωματίδιο θα περιείχε μόνο τον αριθμό των ασαφών συνόλων για έναν μονοδιάστατο χώρο εισόδου. Ωστόσο, είναι πιθανό ότι οι αριθμοί των ασαφών συνόλων s_i συσχετίζονται με κάποιες, ή ακόμη και με όλες τις μεταβλητές εισόδου. Σε αυτή την περίπτωση, οι ανεξάρτητες αλλαγές που γίνονται από τα διάφορα σμήνη θα έχουν αρνητικές επιπτώσεις για τις συσχετιζόμενες μεταβλητές.

Για κάθε σμήνος, τα σωματίδια ενημερώνονται σύμφωνα με σχέση:

$$P_k.\mathbf{\Lambda}_i(t+1) = P_k.\mathbf{\Lambda}_i(t) + P_k.\mathbf{v}_i(t+1), \quad k=1,2 \tag{3.17}$$

Επίσης, τα στοιχεία του διανύσματος της ταχύτητας του πρώτου σμήνους $P_1.\mathbf{v}_i(t)$ υπολογίζονται ως εξής:

$$P_1.v_{ij}(t+1) = \text{round}\left\{ \begin{array}{r} w_{inertia} P_1.v_{ij}(t) + c_1 r_{1j}(t)\left[P_1.y_{ij}(t) - P_1.s_{ij}(t) \right] + \ldots \\ \ldots c_2 r_{2j}(t)\left[P_1.\hat{y}_j(t) - P_1.s_{ij}(t) \right] \end{array} \right\} \tag{3.18}$$

όπου:

$P_1.v_{ij}(t)$ είναι η ταχύτητα του σωματιδίου i=1,2,…,P, στη διάσταση j=1,2,…,N, την επανάληψη t,

$P_1.s_{ij}(t)$ είναι η θέση του σωματιδίου i=1,2,…,P, στη διάσταση j=1,2,…,N, την επανάληψη t,

$P_1.y_{ij}(t)$ είναι η καλύτερη προσωπική θέση που έχει πετύχει το σωματίδιο i=1,2,…,P, στη διάσταση j=1,2,…,N, την επανάληψη t,

$P_1.\hat{y}_j(t)$ είναι η καλύτερη θέση που έχουν πετύχει συνολικά τα σωματίδια που ανήκουν στο πρώτο σμήνος στη διάσταση *j*=1,2,…,*N*, την επανάληψη *t*,

$w_{inertia}$: είναι η σταθερά αδράνειας (inertia),

c_1 και c_2 είναι σταθερές επιτάχυνσης,

$r_{1j}(t)$ και $r_{2j}(t)$ παίρνουν τυχαίες τιμές στο διάστημα [0, 1], που δειγματοληπτούνται κάθε επανάληψη t από μία ομοιόμορφη κατανομή.

Όσον αφορά στο δεύτερο σμήνος, οι αντίστοιχες ταχύτητες $P_2.\mathrm{v}_i(t)$ ενημερώνονται χρησιμοποιώντας τον τύπο:

$$P_2.v_i(t+1) = w_{inertia}P_2.v_i(t) + c_1r_1(t)\left[P_2.y_i(t) - P_2.\kappa_i(t)\right] + \ldots$$
$$+c_2r_2(t)\left[P_2.\hat{y}(t) - P_2.\kappa_i(t)\right]\frac{n!}{r!(n-r)!}$$

(3.19)

όπου:

$P_2.v_i(t)$ είναι η ταχύτητα του σωματιδίου *i*=1,2,…,*P*, στη διάσταση *j*=1,2,…,*N*, την επανάληψη *t*,

$P_2.\kappa_i(t)$ είναι η θέση του σωματιδίου *i*=1,2,…,*P*, την επανάληψη *t*,

$P_2.y_i(t)$ είναι η καλύτερη προσωπική θέση που έχει πετύχει το σωματίδιο *i*=1,2,…,*P*, την επανάληψη *t*,

$P_2.\hat{y}(t)$ είναι η καλύτερη θέση που έχουν πετύχει συνολικά τα σωματίδια που ανήκουν στο δεύτερο σμήνος, την επανάληψη *t*,

Η εξισορρόπηση μεταξύ της ολικής και της τοπικής εξερεύνησης σε κάθε σμήνος ελέγχεται με έναν μηχανισμό περιορισμού ταχύτητας, ο οποίος περιορίζει τις ταχύτητες των σωματιδίων για να παραμείνουν εντός κάποιων προκαθορισμένων ορίων:

$$P_k.v_{ij}(t+1) = \begin{cases} P_k.v_{ij}(t+1), & \text{αν } \left\|P_k.v_{ij}(t+1)\right\| < P_k.V_{max} \\ \pm P_k.V_{max} & , \text{ διαφορετικά} \end{cases} \quad (3.20)$$

όπου $P_k.V_{max}$ είναι η σταθερά περιορισμού ταχύτητας, η οποία μπορεί να είναι διαφορετική για κάθε σμήνος.

Κάθε σμήνος ελέγχει ένα συγκεκριμένο τμήμα της διαδικασίας βελτιστοποίησης του δικτύου RBF, όπου το πρώτο σμήνος βασίζεται στον αλγόριθμο NSFM και το δεύτερο σμήνος χρησιμοποιεί την ευρετική μέθοδο κάλυψης των πλησιέστερων κέντρων (nearest neighbor coverage, NNC). Βέβαια, απαιτούνται πληροφορίες και από τα δύο σμήνη προκειμένου να εκπαιδευτεί πλήρως ένα δίκτυο RBF και, εν τέλει, να υπολογιστεί η συνάρτηση καταλληλότητας. Έτσι, είναι αναγκαίο ένα κατάλληλο διάνυσμα που να παρέχει το πλαίσιο για την αξιολόγηση των στοιχείων του κάθε σμήνους. Αυτό το κατάλληλο

διάνυσμα δημιουργείται μέσω της συνένωσης κάθε σωματιδίου από το ένα σμήνος, με την καλύτερη ολική θέση των σωματιδίων του άλλου σμήνους, όπως προτάθηκε στη δημοσίευση [80]. Αυτό επιτυγχάνεται εκπαιδεύοντας το δίκτυο $\mathrm{RBF}\left(P_k.\mathbf{\Lambda}_i, P_{(2k)^{2-k}}.\hat{\mathbf{y}}\right),\ k=1,2$ χρησιμοποιώντας από το ένα σμήνος το σωματίδιο $\mathbf{\Lambda}_i$, i=1,2,…,P, και από το άλλο σμήνος την ολική καλύτερη θέση όλων των σωματιδίων $\hat{\mathbf{y}}$. Έτσι, προκειμένου να υπολογιστεί η συνάρτηση καταλληλότητας όλων των σωματιδίων του σμήνους P_k, P διαφορετικά δίκτυα RBF εκπαιδεύονται εναλλάσσοντας τα σωματίδια $P_k.\mathbf{\Lambda}_i$, ενώ πάντοτε εφαρμόζεται αμετάβλητα το ολικά καλύτερο σωματίδιο του συμπληρωματικού σμήνους $P_{(2k)^{2-k}}.\hat{\mathbf{y}}$.

Μετά την πλήρη εκπαίδευση των δικτύων RBF, η αντίστοιχη τιμή της συνάρτησης καταλληλότητας υπολογίζεται εφαρμόζοντας ένα κατάλληλα επιλεγμένο κριτήριο σφάλματος στο σύνολο αξιολόγησης. Εξαιτίας της διαφορετικής φύσης των προβλημάτων προσέγγισης συνάρτησης και των προβλημάτων ταξινόμησης, διακρίνουμε δύο κριτήρια, ανάλογα με την περίπτωση. Συγκεκριμένα, σε αυτόν τον αλγόριθμο, για τα προβλήματα προσέγγισης συνάρτησης υιοθετήθηκε ο δείκτης RMSE (περιγράφεται από την εξίσωση (1.11)) και για τα προβλήματα ταξινόμησης ο συντελεστής MCC [167] (περιγράφεται από την εξίσωση (1.14)). Η προτεινόμενη μέθοδος θα μπορούσε, επίσης να ενσωματώσει εναλλακτικά κριτήρια σφάλματος.

Το πλαίσιο που περιγράφηκε παραπάνω για την αξιολόγηση των σωματιδίων που ανήκουν σε δύο σμήνη παρουσιάζει δύο σημαντικά πλεονεκτήματα όσον αφορά στη διαδικασία βελτιστοποίησης των δικτύων RBF. Πρώτον, η συνάρτηση καταλληλότητας αξιολογείται μετά την ενημέρωση κάθε ξεχωριστού μέρους του διανύσματος λύσης, σε αντίθεση με τον πρότυπο αλγόριθμο PSO, όπου υπολογίζεται μόνο μετά την ενημέρωση όλων των συνιστωσών του. Αυτό οδηγεί σε μια πιο λεπτομερή ανάθεση σφάλματος, αποφεύγοντας καταστάσεις όπου μια ενημέρωση θα μπορούσε να βελτιώσει το ένα μέρος του διανύσματος λύσης, αλλά ενδεχομένως να “χειροτερέψει” το άλλο.

Το δεύτερο πλεονέκτημα σχετίζεται με τον αυξημένο αριθμό συνδυασμών των διαφόρων σωματιδίων από διαφορετικά σμήνη που εξετάζεται, ο οποίος ενισχύει την ποικιλία των λύσεων. Η μεγαλύτερη ποικιλία λύσεων, όμως, έχει ως μειονέκτημα το αυξημένο υπολογιστικό κόστος, καθώς ο αριθμός των αξιολογήσεων της συναρτήσης καταλληλότητας ανά επανάληψη ουσιαστικά διπλασιάζεται. Αυτό θα μπορούσε να είναι ένα σοβαρό μειονέκτημα όσον αφορά στην εκπαίδευση των δικτύων RBF, καθώς κάθε κύκλος εκπαίδευσης θα μπορούσε να έχει σημαντική υπολογιστική επιβάρυνση, ανάλογα με τα μεγέθη του δικτύου και του συνόλου των δεδομένων. Παρόλα αυτά, το προτεινόμενο πλαίσιο αξιολόγησης μπορεί να βοηθήσει στην αποσυμφόρηση ενός σημαντικού μέρους του υπολογιστικού φορτίου που επιβλήθηκε με την ανταλλαγή συμπληρωματικών πληροφοριών μεταξύ των σμηνών. Το κομμάτι της διαδικασίας της εκπαίδευσης που απαιτεί το πιο μεγάλο υπολογιστικό φορτίο είναι ο υπολογισμός των κέντρων RBF, λόγω του μεγάλου αριθμού των αποστάσεων που πρέπει να υπολογιστούν [34]. Θέτοντας το σμήνος που χειρίζεται την ασαφή διαμέριση να εκτελείται πάντα πρώτο και στη συνέχεια να περνά τα κέντρα RBF του ολικού καλύτερου σωματιδίου στο σμήνος που χειρίζεται την κάλυψη των πλησιέστερων κέντρων, η διαδικασία αυτή εγγυάται ότι το στάδιο που υπολογίζει τα κέντρα RBF και έχει το πιο σημαντικό υπολογιστικό φορτίο, θα εκτελεστεί μόνο για τα σωματίδια του πρώτου σμήνους. Έτσι, αν και η συνεργατική μέθοδος διπλασιάζει τον αριθμό των αξιολογήσεων της συνάρτησης καταλληλότητας, το αντίστοιχο υπολογιστικό φορτίο αυξάνεται με σημαντικά μικρότερο συντελεστή.

Η προσωπική καλύτερη θέση $P_k.\mathbf{y}_i(t)$ για κάθε σωματίδιο *i* του κάθε σμήνους *k* αντιπροσωπεύει το καλύτερο αποτέλεσμα που βρέθηκε για αυτό το συγκεκριμένο σωματίδιο μέχρι την επανάληψη *t*. Ο υπολογισμός βασίζεται στην αξιολόγηση της απόδοσης του σωματιδίου στο σύνολο δεδομένων αξιολόγησης, εφαρμόζοντας την επιλεγμένη συνάρτηση καταλληλότητας *f*:

$$P_k.\mathbf{y}_i(t+1)=\begin{cases} P_k.\mathbf{y}_i(t) & \text{, αν } f\left(\mathrm{RBF}\left(P_k.\mathbf{\Lambda}_i(t+1), P_{(2k)^{2-k}}.\hat{\mathbf{y}}(t)\right)\right) \\ & \geq f\left(\mathrm{RBF}\left(P_k.\mathbf{y}_i(t), P_{(2k)^{2-k}}.\hat{\mathbf{y}}(t-1)\right)\right) \\ P_k.\mathbf{\Lambda}_i(t+1) & \text{, διαφορετικά} \end{cases} \tag{3.21}$$

Η ολική καλύτερη θέση $P_k.\hat{\mathbf{y}}(t)$ για κάθε σμήνος *k* αντιπροσωπεύει το καλύτερο αποτέλεσμα που βρέθηκε από το αντίστοιχο σμήνος μέχρι την επανάληψη *t*:

$$P_k.\hat{\mathbf{y}}(t)=\arg\min_{P_k.\mathbf{y}_i(t),} f\left(\mathrm{RBF}\left(P_k.\mathbf{y}_i(t), P_{(2k)^{2-k}}.\hat{\mathbf{y}}(t-1)\right)\right),\quad i=1,2,...P \tag{3.22}$$

Πρέπει να σημειωθεί ότι, οι τύποι (3.21)-(3.22) είναι κατάλληλοι για προβλήματα ελαχιστοποίησης, αλλά μπορούν να τροποποιηθούν έτσι ώστε να είναι συμβατοί με κριτήρια σφάλματος τα οποία χρειάζεται να μεγιστοποιηθούν, π.χ. MCC.

Ένα κοινό πρόβλημα που παρουσιάζουν οι τεχνικές που βασίζονται στη μέθοδο PSO είναι το πρόβλημα της στασιμότητας (stagnation). Το τελευταίο αφορά μια κατάσταση όπου, καθώς ο αριθμός των επαναλήψεων αυξάνεται, τα σωματίδια κινούνται πολύ κοντά στην ολική καλύτερη λύση του σμήνους τους. Αυτό έχει ως αποτέλεσμα, τα σωματίδια να έχουν πολύ περιορισμένες δυνατότητες για να ερευνήσουν τον χώρο αναζήτησης και ενδεχομένως, να μη μπορούν να βελτιωθούν περαιτέρω. Στην περίπτωση των συνεργατικών σμηνών, οι αρνητικές επιπτώσεις του προβλήματος της στασιμότητας μπορεί να αυξηθούν λόγω του γεγονότος ότι ο υπολογισμός της καταλληλότητας του κάθε σμήνους, περιλαμβάνει επίσης και την ολική καλύτερη λύση του συμπληρωματικού σμήνους. Ας θεωρήσουμε την εξής υποθετική κατάσταση: για ένα συγκεκριμένο ολικά καλύτερο διάνυσμα που ανήκει στο σμήνος που διαχειρίζεται την ασαφή διαμέριση, η επιλογή στενών συναρτήσεων Gauss οδηγεί στη βελτίωση της συνάρτησης καταλληλότητας. Εάν το ολικά καλύτερο σωματίδιο του σμήνους που χειρίζεται την ασαφή διαμέριση παραμένει σταθερό για μεγάλο χρονικό διάστημα, αυτό ενδεχομένως να σημαίνει ότι τα σωματίδια του άλλου σμήνους (που χειρίζεται την κάλυψη πλησιέστερων κέντρων) έχουν προσαρμοστεί κατάλληλα, και ως εκ τούτου έχουν συγκλίνει μέσα σε μια μικρή περιοχή του χώρου αναζήτησης, που αντιστοιχεί σε χαμηλά ποσοστά κάλυψης, τα οποία μάλιστα ευνοούν τις στενές γκαουσιανές. Σε αυτό το σημείο, το ολικά καλύτερο σωματίδιο του σμήνους που χειρίζεται την ασαφή διαμέριση μετακινείται σε μια νέα θέση, όπου θα μπορούσε να επιτευχθεί μία ακόμη καλύτερη επίδοση με πιο φαρδιές γκαουσιανές συναρτήσεις. Δυστυχώς, τώρα, δεν είναι δυνατό για τα σωματίδια που ανήκουν στο σμήνος που χειρίζεται την κάλυψη πλησιέστερων κέντρων να μετατοπιστούν ανάλογα, καθώς όλα τα σωματίδια έχουν μείνει στάσιμα σε μια περιοχή, η οποία μπορεί να παράγει μόνο στενές γκαουσιανές.

Για να μπορέσει ο αλγόριθμος να χειριστεί τέτοιες καταστάσεις, τα σωματίδια των σμηνών επανεκκινούνται από νέες τυχαίες θέσεις (reset), όταν ισχύσουν ταυτόχρονα οι δύο ακόλουθες προϋποθέσεις:

1. ένα σμήνος έχει συγκλίνει σε μια μικρή περιοχή του χώρου εισόδου και
2. η ολική καλύτερη θέση του συμπληρωματικού σμήνους έχει αλλάξει.

Η σύγκλιση του σμήνους k εκτιμάται υπολογίζοντας το κανονικοποιημένο άθροισμα των αποστάσεων από το ολικά καλύτερο σωματίδιο του σμήνους $P_k.S_{dnorm}$, ως εξής:

$$P_k.S_{dnorm}(t) = \frac{P_k.S_d(t)}{P_k.S_d(0)} \tag{3.23}$$

όπου $P_k.S_d(t)$ είναι το άθροισμα των αποστάσεων όλων των σωματιδίων, από το ολικά καλύτερο σωματίδιο την επανάληψη t:

$$P_k.S_d(t) = \sum_{i=1}^{P}\left[\left\|P_k.\mathbf{\Lambda}_i(t) - P_k.\hat{\mathbf{y}}(t)\right\|\right] \tag{3.24}$$

Όταν το $P_k.S_{dnorm}(t)$ γίνεται μικρότερο από μία σταθερά σύγκλισης ε, τότε το σμήνος k θεωρείται ότι έχει συγκλίνει.

Τέλος, μία σχηματική αναπαράσταση των δύο συνεργατικών σμηνών, που συνεργάζονται για τη βελτιστοποίηση μοντέλων RBF παρουσιάζεται στο σχήμα 3.6, ενώ μια επισκόπηση του αλγορίθμου COOP-VW δίνεται στη συνέχεια.

Αλγόριθμος 4 – Αλγόριθμος COOP-VW

Δεδομένα Εισόδου: $\{\mathbf{U}_{train}, \mathbf{Y}_{train}\}$: Σύνολο δεδομένων εκπαίδευσης,

$\{\mathbf{U}_{val}, \mathbf{Y}_{val}\}$: Σύνολο δεδομένων αξιολόγησης

$s_{\min}$, $s_{\max}$, $\kappa_{\min}$, $\kappa_{\max}$: Παράμετροι αλγορίθμων NSFM και NNC,
P: Πληθυσμός σμήνους,

c_1, c_2, w, $P_k.V_{\max}$: Παράμετροι του αλγορίθμου PSO,

ε, ξ: Παράμετροι για τη σύγκλιση του σμήνους και τις συνθήκες διακοπής

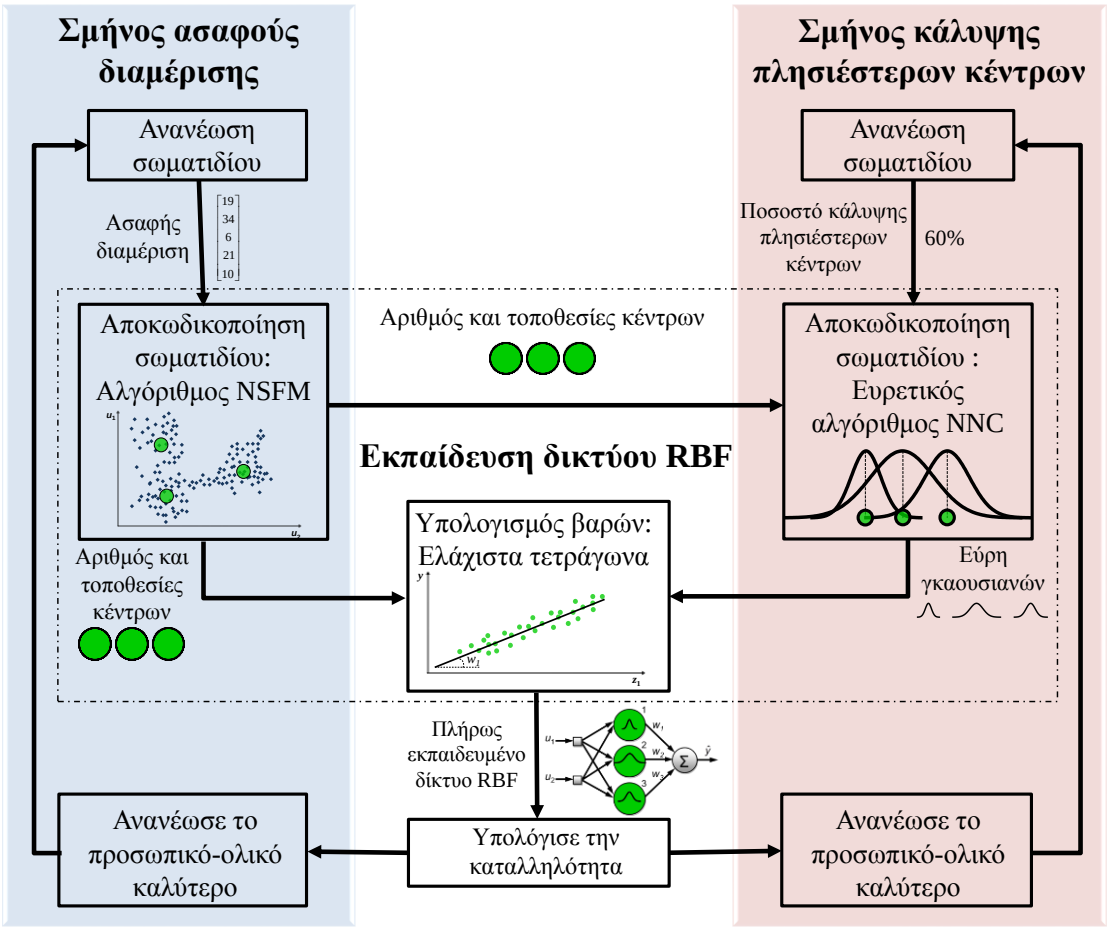
Δεδομένα Εξόδου: $L_f, \hat{\mathbf{U}}_f, \mathbf{w}_f$: Βελτιστοποιημένες παράμετροι του δικτύου RBF

1. **Για** i=1:P **Εκτέλεσε**:
2. Αρχικοποίησε τα σωματίδια $P_1.\mathbf{\Lambda}_i(0)$ και $P_2.\mathbf{\Lambda}_i(0)$ με τυχαίους αριθμούς
3. Χρησιμοποίησε τις τεχνικές NSFM, NNC (3.12)-(3.13) και LR (2.14) για να εκπαιδευτεί πλήρως το μοντέλο $\mathrm{RBF}\left(P_1.\mathbf{\Lambda}_i(0), P_2.\mathbf{\Lambda}_i(0)\right)$ στο σύνολο δεδομένων $\{\mathbf{U}_{train}, \mathbf{Y}_{train}\}$

4. Υπολόγισε τη συνάρτηση καταλληλότητας $\mathrm{RBF}\left(P_1.\mathbf{\Lambda}_i(0), P_2.\mathbf{\Lambda}_i(0)\right)$ στο σύνολο δεδομένων $\left\{\mathbf{U}_{val}, \mathbf{Y}_{val}\right\}$ και θέσε $P_1.\mathbf{y}_i(0)$ και $P_2.\mathbf{y}_i(0)$
5. **Τέλος Για**
6. Υπολόγισε τις ολικά καλύτερες θέσεις $P_1.\hat{\mathbf{y}}(0)$ και $P_2.\hat{\mathbf{y}}(0)$ σύμφωνα με τη σχέση (3.22)
7. Ξεκίνα με την πρώτη επανάληψη: $t \leftarrow 1$
8. **Όσο** η συνθήκη διακοπής ξ δεν έχει επιτευχθεί **Εκτέλεσε**:
9. Ξεκίνα με το σμήνος που χειρίζεται την ασαφή διαμέριση: $k \leftarrow 1$
10. **Αν** $t > 1$ και $P_2.\hat{\mathbf{y}}(t-1) \neq P_2.\hat{\mathbf{y}}(t-2)$ και $P_1.S_{dnorm}(t-1) < \varepsilon$
11. **Τότε** επανεκκίνησε τα σωματίδια $P_1.\mathbf{\Lambda}_i(t),\ i = 1, 2, ..., P$
12. **Για** *i*=1:*P* **Εκτέλεσε**:
13. Χρησιμοποίησε τις τεχνικές NSFM, NNC (3.12)-(3.13) και LR (2.14) για να εκπαιδευτεί πλήρως το μοντέλο $\mathrm{RBF}\left(P_1.\mathbf{\Lambda}_i(t), P_2.\hat{\mathbf{y}}(t-1)\right)$ στο σύνολο δεδομένων $\left\{\mathbf{U}_{train}, \mathbf{Y}_{train}\right\}$
14. Υπολόγισε τη συνάρτηση καταλληλότητας $\mathrm{RBF}\left(P_1.\mathbf{\Lambda}_i(t), P_2.\hat{\mathbf{y}}(t-1)\right)$ στο σύνολο δεδομένων $\left\{\mathbf{U}_{val}, \mathbf{Y}_{val}\right\}$ και υπολόγισε το $P_1.\mathbf{y}_i(t)$ σύμφωνα με τη σχέση (3.21)
15. **Τέλος Για**
16. Υπολόγισε την ολική καλύτερη θέση $P_1.\hat{\mathbf{y}}(t)$ σύμφωνα με τη σχέση (3.22)
17. **Για** *i*=1:*P* **Εκτέλεσε**:
18. **Για** *j*=1:*N* **Εκτέλεσε**:
19. Ενημέρωσε το διάνυσμα της ταχύτητας $P_1.v_{ij}(t+1)$ σύμφωνα με τις σχέσεις (3.18) και (3.20)
20. **Τέλος Για**
21. Ενημέρωσε τις θέσεις των σωματιδίων $P_1.\mathbf{\Lambda}_i(t+1)$ σύμφωνα με τη σχέση (3.17)
22. **Τέλος Για**
23. Συνέχισε με το σμήνος που χειρίζεται την κάλυψη των πλησιέστερων κέντρων: $k \leftarrow 2$
24. **Αν** $t > 1$ και $P_1.\hat{\mathbf{y}}(t) \neq P_1.\hat{\mathbf{y}}(t-1)$ και $P_2.S_d(t-1) < \varepsilon$
25. **Τότε** επανεκκίνησε τα σωματίδια $P_2.\mathbf{\Lambda}_i(t),\ i = 1, 2, ..., P$
26. **Για** *i*=1:*P* **Εκτέλεσε**:
27. Χρησιμοποίησε τα κέντρα RBF που αντιστοιχούν στα $P_1.\hat{\mathbf{y}}(t)$, NSFM, NNC (3.12)-(3.13) και LR (2.14) για να εκπαιδευτεί πλήρως το μοντέλο $\mathrm{RBF}\left(P_2.\mathbf{\Lambda}_i(t), P_1.\hat{\mathbf{y}}(t)\right)$ στο σύνολο δεδομένων $\left\{\mathbf{U}_{train}, \mathbf{Y}_{train}\right\}$
28. Υπολόγισε τη συνάρτηση καταλληλότητας $\mathrm{RBF}\left(P_2.\mathbf{\Lambda}_i(t), P_1.\hat{\mathbf{y}}(t)\right)$ στο

σύνολο δεδομένων $\{\mathbf{U}_{val}, \mathbf{Y}_{val}\}$ και υπολόγισε το $P_2.\mathbf{y}_i(t)$ σύμφωνα με τη σχέση (3.21)

29. **Τέλος Για**
30. Υπολόγισε την ολική καλύτερη θέση $P_2.\hat{\mathbf{y}}(t)$ σύμφωνα με τη σχέση (3.22)
31. **Για** i=1:P **Εκτέλεσε**:
32. Ενημέρωσε το διάνυσμα της ταχύτητας $P_2.v_i(t+1)$ σύμφωνα με τις σχέσεις (3.19) και (3.20)
33. Ενημέρωσε τις θέσεις των σωματιδίων $P_2.\mathbf{\Lambda}_i(t+1)$ σύμφωνα με τη σχέση (3.17)
34. **Τέλος Για**
35. Συνέχισε με την επόμενη επανάληψη: $t \leftarrow t+1$
36. **Τέλος Όσο**



Σχήμα 3.6. Σχηματική αναπαράσταση των δύο συνεργατικών σμηνών, τα οποία συνεργάζονται για τη βελτιστοποίηση μοντέλων νευρωνικών δικτύων RBF.

3.2.3 Αξιολόγηση του αλγορίθμου COOP-VW

Ο αλγόριθμος COOP-VW δοκιμάστηκε σε μια ποικιλία από σύνολα δεδομένων τύπου benchmark που προέρχονται από τον πραγματικό κόσμο αλλά και συνθετικά σύνολα δεδομένων, που αφορούν τόσο σε προβλήματα προσέγγισης συνάρτησης όσο και σε προβλήματα ταξινόμησης. Τα σύνολα δεδομένων που προέρχονται από τον πραγματικό κόσμο είναι διαθέσιμα στη βάση UCI Machine Learning Repository [190]. Τα σύνολα δεδομένων που χρησιμοποιήθηκαν σε αυτή την εφαρμογή δίνονται στον πίνακα 3.7, μαζί με τους αριθμούς των μεταβλητών εισόδου και εξόδου (για τα σύνολα δεδομένων ταξινόμησης, ο αριθμός μεταβλητών εξόδου αντιστοιχεί στον αριθμό των κλάσεων), και το συνολικό αριθμό των παραδειγμάτων. Επίσης, μια σύντομη περιγραφή για κάθε σύνολο δεδομένων δίνεται παρακάτω.

- **Auto Price**: Αυτό το σύνολο δεδομένων περιγράφεται στο κεφάλαιο 3.1.3.
- **Energy efficiency**: Ο στόχος σε αυτό το σύνολο δεδομένων είναι να εκτιμηθούν οι απαιτήσεις θέρμανσης των κτιρίων, δηλαδή η ενεργειακή απόδοση των κτιρίων ως συνάρτηση των παραμέτρων τους [202].
- **Friedman**: Αυτό το σύνολο δεδομένων περιγράφεται στο κεφάλαιο 3.1.3.
- **Parkinsons Telemonitoring**: Αυτό το σύνολο αποτελείται από μια σειρά μετρήσεων της φωνής 42 ατόμων που πάσχουν από τη νόσο Πάρκινσον σε πρώιμο στάδιο [203]. Ο στόχος είναι να εκτιμηθούν οι βαθμολογίες της κλίμακας αξιολόγησης της νόσου Πάρκινσον UPDRS (unified Parkinson's disease rating scale) με βάση 16 βιοϊατρικές μετρήσεις φωνής.
- **Samad**: Αυτό το σύνολο δεδομένων περιγράφεται στο κεφάλαιο 3.1.3.
- **Cardiotocography (FHR, NSP)**: Αυτά τα δύο σύνολα δεδομένων περιέχουν έναν αριθμό εμβρυϊκών καρδιοτοκογραφημάτων (CTGs), τα οποία υπέστησαν αυτόματη επεξεργασία μαζί με τα αντίστοιχα διαγνωστικά χαρακτηριστικά τους. Ο στόχος του συνόλου δεδομένων «Cardiotocography FHR» είναι να βρεθεί το μορφολογικό πρότυπο (1-10) του εμβρυϊκού καρδιακού ρυθμού (FHR), ενώ ο σκοπός του συνόλου δεδομένων «Cardiotocography NSP» είναι να προβλεφθεί η εμβρυϊκή κατάσταση (φυσιολογική, ύποπτη, παθολογική). Τα διαθέσιμα χαρακτηριστικά είναι κοινά για τα δύο σύνολα δεδομένων, όσον αφορά στην κατασκευή του ταξινομητή. Σε αυτή την εφαρμογή χρησιμοποιήθηκε μόνο το σύνολο δεδομένων Cardiotocography FHR.
- **Iris**: Αυτή είναι ίσως η πιο γνωστή βάση δεδομένων που υπάρχει στη βιβλιογραφία στον τομέα της αναγνώρισης προτύπων. Αυτό το σύνολο δεδομένων περιέχει 3 διαφορετικές κατηγορίες εξόδου που αποτελούνται από 50 περιπτώσεις η κάθε μία. Η κάθε κατηγορία, αναφέρεται σε ένα διαφορετικό είδος φυτού ίριδας.
- **Leaf**: Αυτή η βάση δεδομένων αποτελείται από 30 διαφορετικά είδη φυτών, που ταξινομούνται με βάση δεδομένα που αφορούν στο σχήμα και στην υφή, τα οποία εξάγονται από φωτογραφίες από τα δείγματα φύλλων [204].
- **Waveform Database Generator**: Αυτό είναι ένα συνθετικό σύνολο δεδομένων που παράγεται από έναν κώδικα σε γλώσσα C. Αποτελείται από 21 χαρακτηριστικά και τα δεδομένα χωρίζονται σε 3 διαφορετικές κατηγορίες.

Όλα τα σύνολα δεδομένων χωρίζονται με τυχαίο τρόπο σε τρία υποσύνολα. Εκτός από τα σύνολα δεδομένων εκπαίδευσης και αξιολόγησης, τα οποία χρησιμοποιούνται για τον υπολογισμό των παραμέτρων του δικτύου και την επιλογή του μοντέλου, αντίστοιχα, είναι σημαντικό να δοκιμαστεί η απόδοση του δικτύου RBF σε ένα τρίτο ανεξάρτητο υποσύνολο

δεδομένων. Το τελευταίο είναι σημαντικό λόγω του φαινομένου της υπερπροσαρμογής που θα μπορούσε να προκύψει στο σύνολο αξιολόγησης, καθώς η διαδικασία βελτιστοποίησης χρησιμοποιεί το τελευταίο για να αξιολογήσει τη συνάρτηση καταλληλότητας. Τα διαθέσιμα δεδομένα κατανέμονται στα τρία υποσύνολα με τυχαίο τρόπο, χρησιμοποιώντας το 50% των δεδομένων για την εκπαίδευση, το 25% για την αξιολόγηση και το υπόλοιπο 25% για τον έλεγχο του δικτύου. Στην περίπτωση των προβλημάτων κατηγοριοποίησης, τα δεδομένα που ανήκουν στις ίδιες κατηγορίες κατανεμήθηκαν ομοιόμορφα μεταξύ των τριών υποσυνόλων, σύμφωνα με τα προαναφερθέντα ποσοστά.

Για λόγους σύγκρισης, δοκιμάστηκαν επίσης δύο διαφορετικές τεχνικές εξελικτικού υπολογισμού και βελτιστοποίησης δικτύων RBF. Η πρώτη τεχνική είναι ο αρχικός αλγόριθμος PSO-NSFM [67], ο οποίος χρησιμοποιεί ένα μόνο σμήνος και μια συνάρτηση βάσης σταθερού εύρους, τη συνάρτηση TPS. Η σύγκριση με αυτή τη μέθοδο, βοηθάει στο να αξιολογηθεί η αποτελεσματικότητα των δικτύων RBF με χρήση συναρτήσεων μεταβλητού εύρους έναντι της επιτυχημένης τεχνικής βελτιστοποίησης δικτύων RBF με συναρτήσεις βάσης σταθερού εύρους.

Η δεύτερη τεχνική σύγκρισης εξελικτικού υπολογισμού, περιλαμβάνει μία μη συνεργατική παραλλαγή χρησιμοποιώντας ένα μόνο σμήνος και τη συνάρτηση βάσης μεταβλητού εύρους Gauss και αποκαλείται SINGLE-VW. Σε αυτή την περίπτωση, χρησιμοποιείται ένα μόνο σμήνος, το οποίο αποτελείται από σωματίδια των οποίων τα στοιχεία περιλαμβάνουν τόσο την ασαφή διαμέριση, όσο και την κάλυψη των πλησιέστερων κέντρων, ακολουθώντας τη δομή που περιγράφεται από την εξίσωση (3.14). Η μέθοδος αυτή χρησιμοποιείται για να αξιολογηθούν τα οφέλη της συνεργατικής τεχνικής έναντι μιας παρόμοιας, αλλά μη συνεργατικής, μεθοδολογίας.

Τέλος, δοκιμάστηκε επιπλέον, μια τεχνική που δεν ανήκει στον χώρο του εξελικτικού

Πίνακας 3.7. Επισκόπηση συνόλων δεδομένων τύπου benchmark

Σύνολο δεδομένων	**Προέλευση Συνόλου Δεδομένων[α]**	**Αριθμός μεταβλητών εισόδου**	**Αριθμός μεταβλητών εξόδου[β]**	**Αριθμός δεδομένων**
ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ ΠΡΟΣΕΓΓΙΣΗΣ ΣΥΝΑΡΤΗΣΗΣ				
Auto Price	RW	15	1	159
Energy	RW	8	1	768
Friedman	S	5	1	1000
Parkinsons	RW	16	1	5875
Samad	S	3	1	1500
ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ ΤΑΞΙΝΟΜΗΣΗΣ				
Cardiotocography FHR	RW	21	10	2126
Iris	RW	4	3	150
Leaf	RW	14	30	340
Waveform	S	21	3	5000

[α]RW: Πραγματικά Σύνολα Δεδομένων (Real-World), S: Συνθετικά Σύνολα Δεδομένων (Synthetic)
[β]Για τα σύνολα δεδομένων ταξινόμησης, ο αριθμός μεταβλητών εξόδου αντιστοιχεί στον αριθμό των κλάσεων

Πίνακας 3.8. Λειτουργικές παράμετροι για τον αλγόριθμο COOP-VW

Παράμετρος	**Σύμβολο**	**Τιμή**
Πεδίο ορισμού ασαφών συνόλων	$[s_{\min}\, s_{\max}]$	[4 50]
Πεδίο ορισμού κάλυψης πλησιέστερων κέντρων	$[\kappa_{\min}\, \kappa_{\max}]$	[2% 99%]
Πληθυσμός	P	20
Αδράνεια	w	0.8
Νοσταλγία	c_1	1.5
Ζήλεια	c_2	1.5
Σταθερά περιορισμού ταχύτητας ασαφών συνόλων[α]	$P_1.V_{max}$	S:5, M:15, L:25
Σταθερά περιορισμού ταχύτητας κάλυψης πλησιέστερων κέντρων	$P_2.V_{max}$	30
Μέγιστος αριθμός βημάτων προσομοίωσης	ξ	2000
Σταθερά σύγκλησης	ε	0.1

[α]Τιμή που εξαρτάται από τη διαστατικότητα του χώρου εισόδου: Μικρά προβλήματα (Small - S): 1-4 μεταβλητές εισόδου, Μέτρια προβλήματα (Medium - M): 5-8 μεταβλητές εισόδου, Μεγάλα προβλήματα (Large - L): περισσότερες από 8 μεταβλητές εισόδου

υπολογισμού και βασίζεται σε μια διαφορετική αρχιτεκτονική από τα δίκτυα RBF και σε ένα διαφορετικό αλγόριθμο εκπαίδευσης. Η τεχνική αυτή αφορά στα δίκτυα MLP τα οποία εκπαιδεύονται με τον αλγόριθμο Levenberg-Marquardt [4].

Για την καλύτερη αξιολόγηση των αποτελεσμάτων, εφαρμόστηκε μια διεξοδική διαδικασία αναζήτησης στο σύνολο δεδομένων Samad, το οποίο περιέχει τον μικρότερο αριθμό μεταβλητών εισόδου, έτσι ώστε να καθοριστεί η ολική βέλτιστη λύση των παραμέτρων του δικτύου RBF. Δεδομένου ότι αυτή είναι μια υπολογιστικά δαπανηρή μέθοδος, θα ήταν πρακτικά ανέφικτο να εφαρμοστεί σε σύνολα δεδομένων περισσότερων διαστάσεων.

Οι παράμετροι λειτουργίας του αλγορίθμου COOP-VW, οι οποίες εφαρμόζονται σε όλα τα πειράματα, δίνονται στον Πίνακα 3.8. Οι ίδιες παράμετροι, ανάλογα με την περίπτωση, χρησιμοποιήθηκαν επίσης για τον αλγόριθμο SINGLE-VW, ενώ οι παράμετροι που αναφέρονται στο κεφάλαιο 3.1.3 χρησιμοποιήθηκαν για την τεχνική PSO-NSFM.

Η επιλογή των παραμέτρων έγινε με βάση τη βιβλιογραφία σε συνδυασμό με τη μέθοδο δοκιμής-σφάλματος [1]. Για κάθε σμήνος χρησιμοποιήθηκε μια διαφορετική τιμή για τη σταθερά περιορισμού ταχύτητας. Μία μέθοδος που συνδέεται με τη διάσταση του χώρου εισόδου [67] του συγκεκριμένου προβλήματος, εφαρμόστηκε για την επιλογή της σταθεράς περιορισμού ταχύτητας του σμήνους που διαχειρίζεται την ασαφή διαμέριση,. Η χρήση μεγαλύτερων τιμών για τις ταχύτητες σε προβλήματα με εισόδους πολλών διαστάσεων βελτιώνει τα αποτελέσματα. Από την άλλη πλευρά, και στα δύο σμήνη χρησιμοποιήθηκαν οι ίδιοι συντελεστές αδράνειας $w_{inertia}$ και επιτάχυνσης c_1 και c_2.

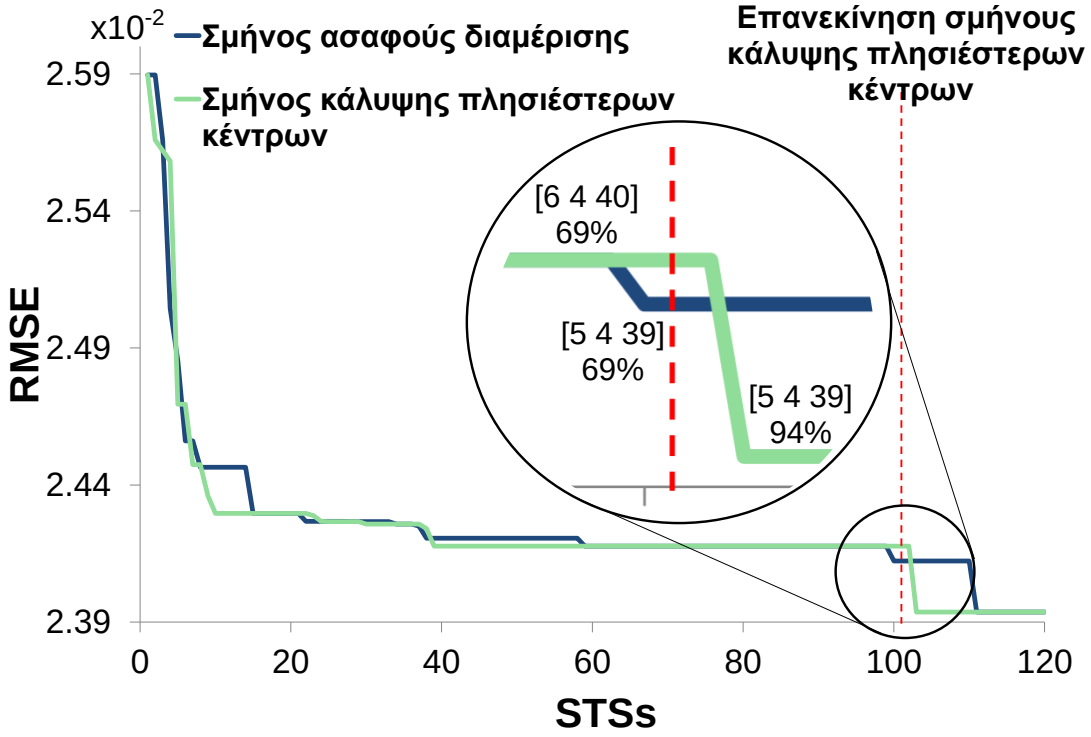
Όσον αφορά στις παραμέτρους που σχετίζονται με την εκπαίδευση των δικτύων RBF και τον αλγόριθμο NSFM, και για τις τρεις μεθοδολογίες που βασίζονται στα δίκτυα RBF, η επιλογή των παραμέτρων ήταν κοινή σύμφωνα με τις κατευθύνσεις που δίνονται στο κεφάλαιο 3.1.3. Τέλος, πρέπει να σημειωθεί ότι μεγαλύτερες τιμές για το μέγεθος του

πληθυσμού, βελτιώνουν περισσότερο τα αποτελέσματα, αφού αυτό σημαίνει ότι το σμήνος θα περιέχει περισσότερα σωματίδια και άρα περισσότερες πιθανές λύσεις. Ωστόσο, οι μεγάλες τιμές πληθυσμού δεν συνιστώνται στα μεγαλύτερα σύνολα δεδομένων, λόγω του αυξημένου υπολογιστικού φορτίου.

Λόγω της στοχαστικής τους φύσης οι μέθοδοι που χρησιμοποιήθηκαν σε αυτή την εφαρμογή εκτελέστηκαν 30 φορές η κάθε μία. Στην περίπτωση των δικτύων MLP, χρησιμοποιήθηκαν δύο κρυφές στοιβάδες όπου για την επιλογή του αντίστοιχου αριθμού νευρώνων εφαρμόστηκε μία διαδικασία διεξοδικής αναζήτησης. Ο αριθμός των νευρώνων σε κάθε κρυφή στοιβάδα κυμαίνεται μεταξύ 5 και 40, όπου κάθε πιθανός συνδυασμός των νευρώνων ανά κρυφή στοιβάδα δοκιμάστηκε 30 φορές, ξεκινώντας από διαφορετικά αρχικά βάρη σε κάθε εκτέλεση.

3.2.4 Αποτελέσματα και συζήτηση

Στο σχήμα 3.7 απεικονίζονται οι μεταβολές του δείκτη RMSE στο υποσύνολο αξιολόγησης ανά αριθμό επανάληψης, όσον αφορά στο καλύτερο αποτέλεσμα που παράχθηκε από τον αλγόριθμο COOP-VW στο σύνολο δεδομένων Samad. Το συγκεκριμένο σύνολο δεδομένων χρησιμοποιήθηκε, διότι είναι το μόνο που επιτρέπει, σε εύλογο χρόνο, τον υπολογισμό του ολικού βέλτιστου μέσω μιας διαδικασίας διεξοδικής αναζήτησης. Η διαδικασία αυτή κατέληξε στη βέλτιστη λύση, η οποία περιλαμβάνει τα ασαφή σύνολα [5 4 39] και το ποσοστό κάλυψης κέντρων 93%. Είναι ενδιαφέρον το γεγονός ότι, με βάση τα αποτελέσματα της διαδικασίας της διεξοδικής αναζήτησης, ο χώρος αναζήτησης γύρω από το σημείο [5 4



Σχήμα 3.7. Η καλύτερη λύση που παράγει ο αλγόριθμος COOP-VW, στο σύνολο δεδομένων Samad: Αλλαγές του δείκτη RMSE ανά επανάληψη στο υποσύνολο αξιολόγησης. Η μεγεθυμένη περιοχή επικεντρώνεται στο σημείο όπου το σμήνος κάλυψης των πλησιέστερων κέντρων εντοπίζει το ολικό βέλτιστο, μετά από μια επιτυχημένη επανεκκίνηση .

39] δημιουργεί μια σειρά καλών υπο-βέλτιστων λύσεων οι οποίες, ωστόσο, συνδυάζονται με τιμές χαμηλότερης κάλυψης κέντρων που κυμαίνονται μεταξύ των ποσοστών 50-70 %.

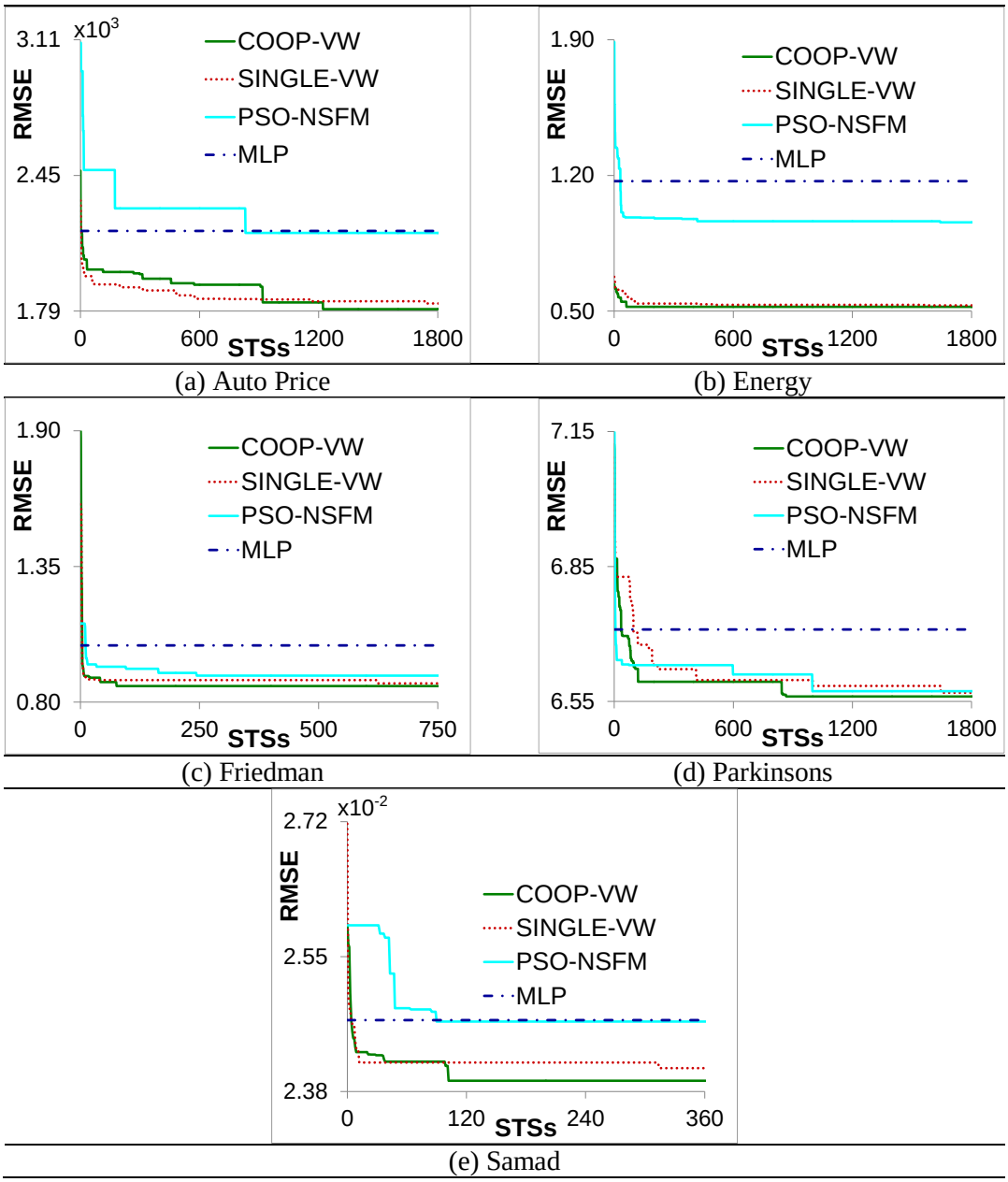
Ο αλγόριθμος COOP-VW πετυχαίνει να εντοπίσει την ολικά βέλτιστη λύση στις 2 από τις 30 εκτελέσεις, ενώ ο αλγόριθμος SINGLE-VW αποτυγχάνει. Ωστόσο, ο αλγόριθμος SINGLE-VW καταφέρνει να βρει τη δεύτερη καλύτερη λύση, η οποία προσδιορίστηκε μέσω της διαδικασίας της διεξοδικής αναζήτησης και περιλαμβάνει τα ασαφή σύνολα [17 4 14] με ποσοστό κάλυψης κέντρων 73%. Στο σχήμα 3.7 φαίνεται ο τρόπος με τον οποίο συνεργάζονται μεταξύ τους τα σμήνη και τα ωφέλιμα αποτελέσματα της διαδικασίας της επανεκκίνησης σμήνους. Επίσης, φαίνεται ότι το σμήνος που διαχειρίζεται την ασαφή διαμέριση πλησιάζει την περιοχή που βρίσκεται η ολικά βέλτιστη λύση και στην εκατοστή επανάληψη το σωματίδιο με την καλύτερη λύση σταματά στα ασαφή σύνολα [5 4 39]. Το συγκεκριμένο σωματίδιο συνιστά το ολικά καλύτερο ως προς την ασαφή διαμέριση, αλλά όταν συνδυάζεται με το 69% της κάλυψης των κέντρων, οδηγεί σε μια υπο-βέλτιστη λύση. Στη συγκεκριμένη επανάληψη, το άθροισμα των αποστάσεων των ολικών καλύτερων σωματιδίων του σμήνους που διαχειρίζεται την κάλυψη των πλησιέστερων κέντρων έχει μικρή τιμή. Αυτό έχει ως αποτέλεσμα όλα τα αντίστοιχα σωματίδια να συγκλίνουν κοντά στην τιμή 69%. Ως εκ τούτου, θα ήταν εξαιρετικά απίθανο για ένα σωματίδιο του σμήνους που διαχειρίζεται την κάλυψη των πλησιέστερων κέντρων να μεταβεί στην τιμή 93% και να εντοπίσει την ολικά βέλτιστη λύση. Έτσι, ο αλγόριθμος θα μπορούσε ενδεχομένως να εγκλωβιστεί σε μια κατάσταση στασιμότητας. Στην προτεινόμενη μεθοδολογία, η κατάσταση αυτή προκαλεί την επανεκκίνηση του σμήνους κάλυψης των πλησιέστερων κέντρων, και του επιτρέπει να ξεφύγει από το τοπικό σημείο στασιμότητας και να ανακαλύψει την ολικά βέλτιστη λύση.

Τα αποτελέσματα για τα σύνολα δεδομένων προσέγγισης συνάρτησης και ταξινόμησης, συνοψίζονται στους πίνακες 3.9 και 3.10, αντίστοιχα. Στους συγκεκριμένους πίνακες απεικονίζονται για όλες τις μεθόδους η τιμή της συνάρτησης κόστους στα σύνολα δεδομένων αξιολόγησης και ελέγχου, και ο αριθμός των νευρώνων. Επιπλέον, για τις τρεις μεθοδολογίες που βασίζονται στην αρχιτεκτονική RBF, απεικονίζονται στους προαναφερθέντες πίνακες, η ασαφής διαμέριση, το ποσοστό κάλυψης των πλησιέστερων κέντρων και ο αριθμός των επαναλήψεων που χρειάζεται για να βρεθεί η καλύτερη λύση. Ακόμη, στα σχήματα 3.8 και 3.9 παρουσιάζεται η εξέλιξη του δείκτη σφάλματος στα σύνολα δεδομένων αξιολόγησης, ανά αριθμό επαναλήψεων (STSs), για το μοντέλο που έδωσε το καλύτερο αποτέλεσμα σε κάθε περίπτωση. Συγκεκριμένα, στο σχήμα 3.8 φαίνεται η εξέλιξη του δείκτη RMSE για τα προβλήματα προσέγγισης συνάρτησης, και στο σχήμα 3.9 φαίνεται η εξέλιξη του δείκτη MCC για τα προβλήματα ταξινόμησης.

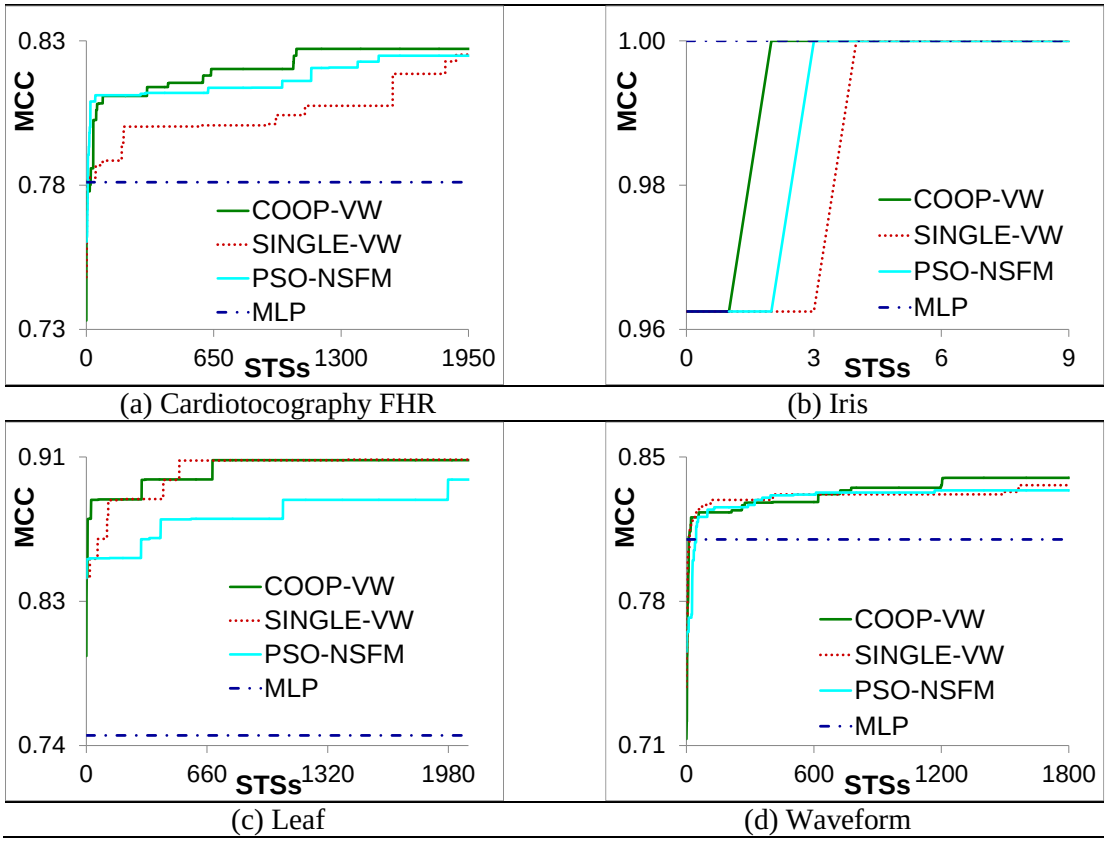
Φαίνεται ότι τα μοντέλα RBF της προτεινόμενης μεθοδολογίας παράγουν μεγαλύτερη ακρίβεια από την άποψη της βέλτιστης και της μέσης λύσης, στη συντριπτική πλειοψηφία των συνόλων δεδομένων προσέγγισης συνάρτησης, ή ταξινόμησης. Πιο συγκεκριμένα, ο αλγόριθμος COOP-VW επιτυγχάνει σημαντικά καλύτερα αποτελέσματα έναντι των δικτύων MLP, που ξεκάθαρα κατατάσσονται στην τέταρτη θέση, παρά το γεγονός ότι είναι η μόνη μέθοδος που χρησιμοποιεί μία διαδικασία διεξοδικής αναζήτησης για τη βελτιστοποίηση της δομής του δικτύου. Οι αλγόριθμοι COOP-VW και SINGLE-VW, επίσης, ξεπερνούν σημαντικά τη μέθοδο PSO-NSFM, αναδεικνύοντας τα πλεονεκτήματα της επέκτασης του αρχικού αλγόριθμου με χρήση συναρτήσεων βάσης με μεταβλητό εύρος. Εδώ, θα πρέπει να σημειωθεί ότι ο αλγόριθμος PSO-NSFM έχει ήδη αποδειχτεί ότι παράγει καλύτερα αποτελέσματα σε σύγκριση με την πρότυπη μέθοδο FM [67].

Παρά το γεγονός ότι οι διαφορές στην ακρίβεια στα σύνολα ελέγχου είναι μικρότερες κατά τη σύγκριση μεταξύ των δύο αλγορίθμων που χρησιμοποιούν συναρτήσεις βάσης

μεταβλητού εύρους, η προτεινόμενη μέθοδος εξακολουθεί να διατηρεί ένα σαφές προβάδισμα. Επιπλέον, ο συνεργατικός αλγόριθμος παρουσιάζει σημαντικά μικρότερες τιμές τυπικών αποκλίσεων γύρω από τον μέσο όρο. Αυτό αποδεικνύει την ευρωστία του αλγορίθμου COOP-VW, ο οποίος παρέχει σχεδόν αμετάβλητα αποτελέσματα. Είναι σημαντικό να αναφερθεί ότι σε όλα τα σύνολα δεδομένων, η τεχνική COOP-VW απαιτεί λιγότερες επαναλήψεις για να παράγει μεγαλύτερη ακρίβεια από αυτή της καλύτερης λύσης που παρουσιάζει η μη συνεργάτική μέθοδος SINGLE-VW. Επίσης, πρέπει να σημειωθεί ότι, όπως αναφέρθηκε και στο κεφάλαιο 3.2, αν και στη συνεργατική μέθοδο μία επανάληψη περιλαμβάνει περισσότερες εκτιμήσεις της συνάρτησης καταλληλότητας σε σύγκριση με τις μη συνεργατικές μεθόδους, το υπολογιστικό φορτίο δεν είναι αναλογικά αυξημένο. Αυτό οφείλεται στο γεγονός ότι ο αλγόριθμος έχει σχεδιαστεί έτσι ώστε να εκτελεί τον ίδιο αριθμό



Σχήμα 3.8. Εξέλιξη του κριτηρίου RMSE ανά επανάληψη στο σύνολο αξιολόγησης για τις τέσσερις μεθοδολογίες, όπου απεικονίζεται η καλύτερη λύση που βρέθηκε σύμφωνα με τον δείκτη RMSE στο σύνολο αξιολόγησης.



(a) Cardiotocography FHR (b) Iris

(c) Leaf (d) Waveform

Σχήμα 3.9. Εξέλιξη του κριτηρίου MCC ανά επανάληψη στο σύνολο αξιολόγησης για τις τέσσερις μεθοδολογίες, όπου απεικονίζεται η καλύτερη λύση που βρέθηκε σύμφωνα με τον δείκτη MCC στο σύνολο αξιολόγησης.

υπολογισμών κέντρων RBF ανά επανάληψη με τις μη συνεργατικές τεχνικές. Τέλος, θα πρέπει να σημειωθεί ότι σε ένα μεγάλο αριθμό των συνόλων δεδομένων, η τεχνική COOP-VW ευνοεί τα δίκτυα με τις μικρότερες δομές.

Πίνακας 3.9. Αποτελέσματα για τα σύνολα δεδομένων προσέγγισης συνάρτησης – RMSE, ασαφής διαμέριση, ποσοστό κάλυψης πλησιέστερων κέντρων, αριθμός κόμβων και απαιτούμενες επαναλήψεις εύρεσης της καλύτερης λύσης

Σύνολο δεδομένων	**Αλγόριθμος**	**RMSE συνόλου αξιολόγησης**	**RMSE συνόλου ελέγχου**
Auto Price	COOP-VW	**1.80×10^{3}** (**1.98×10^{3}**$\pm5.93\times10^{1}$)	**1.70×10^{3}** (**1.81×10^{3}**$\pm5.70\times10^{1}$)
	SINGLE-VW	1.83×10^{3} ($1.99\times10^{3}\pm6.33\times10^{1}$)	1.84×10^{3} ($1.83\times10^{3}\pm5.79\times10^{1}$)
	PSO-NSFM	2.17×10^{3} ($2.38\times10^{3}\pm1.28\times10^{2}$)	2.11×10^{3} ($2.30\times10^{3}\pm9.06\times10^{1}$)
	MLP	2.18×10^{3} ($2.61\times10^{3}\pm2.87\times10^{2}$)	2.14×10^{3} ($2.53\times10^{3}\pm3.24\times10^{2}$)
Energy	COOP-VW	**5.22×10^{-1}** (**5.38×10^{-1}**$\pm7.68\times10^{-3}$)	**5.10×10^{-1}** (**5.16×10^{-1}**$\pm1.99\times10^{-2}$)
	SINGLE-VW	5.29×10^{-1} ($5.56\times10^{-1}\pm2.32\times10^{-2}$)	5.31×10^{-1} ($5.40\times10^{-1}\pm3.19\times10^{-2}$)
	PSO-NSFM	9.57×10^{-1} ($1.00\times10^{0}\pm5.08\times10^{-2}$)	8.95×10^{-1} ($9.84\times10^{-1}\pm9.81\times10^{-2}$)
	MLP	1.17×10^{0} ($1.68\times10^{0}\pm3.89\times10^{-1}$)	1.08×10^{0} ($1.58\times10^{0}\pm3.13\times10^{-1}$)
Friedman	COOP-VW	**8.65×10^{-1}** (**8.84×10^{-1}**$\pm9.54\times10^{-3}$)	**1.01×10^{0}** (**1.01×10^{0}**$\pm2.16\times10^{-2}$)
	SINGLE-VW	8.76×10^{-1} ($9.40\times10^{-1}\pm1.30\times10^{-1}$)	1.04×10^{0} ($1.05\times10^{0}\pm1.01\times10^{-1}$)
	PSO-NSFM	9.08×10^{-1} ($9.33\times10^{-1}\pm1.61\times10^{-2}$)	1.06×10^{0} ($1.07\pm3.03\times10^{-2}$)
	MLP	1.03×10^{0} ($1.19\times10^{0}\pm1.22\times10^{-1}$)	1.11×10^{0} ($1.20\times10^{0}\pm1.30\times10^{-1}$)
Parkinsons	COOP-VW	**6.56×10^{0}** (**6.62×10^{0}**$\pm3.10\times10^{-2}$)	**6.55×10^{0}** (**6.60×10^{0}**$\pm3.71\times10^{-2}$)
	SINGLE-VW	6.57×10^{0} (**6.62×10^{0}**$\pm4.82\times10^{-2}$)	6.56×10^{0} ($6.63\times10^{0}\pm8.61\times10^{-2}$)
	PSO-NSFM	6.57×10^{0} ($6.64\times10^{0}\pm3.02\times10^{-2}$)	6.58×10^{0} ($6.63\times10^{0}\pm2.58\times10^{-1}$)
	MLP	6.71×10^{0} ($6.91\times10^{0}\pm1.22\times10^{-1}$)	7.04×10^{0} ($6.82\times10^{0}\pm1.49\times10^{-1}$)
Samad	COOP-VW	**2.39×10^{-2}** (**2.42×10^{-2}**$\pm9.14\times10^{-5}$)	**2.71×10^{-2}** (**2.73×10^{-2}**$\pm3.05\times10^{-4}$)
	SINGLE-VW	2.41×10^{-2} ($2.43\times10^{-2}\pm3.01\times10^{-4}$)	2.73×10^{-2} ($2.74\times10^{-2}\pm6.27\times10^{-4}$)
	PSO-NSFM	2.47×10^{-2} ($2.48\times10^{-2}\pm2.11\times10^{-4}$)	2.80×10^{-2} ($2.79\times10^{-2}\pm4.40\times10^{-4}$)
	MLP	2.47×10^{-2} ($3.10\times10^{-2}\pm3.70\times10^{-3}$)	2.82×10^{-2} ($3.47\times10^{-2}\pm4.16\times10^{-3}$)

Οι αλγόριθμοι εκτελέστηκαν 30 φορές για κάθε σύνολο δεδομένων. Απεικονίζεται η εκτέλεση για την οποία βρέθηκε το καλύτερο αποτέλεσμα σύμφωνα με τον δείκτη RMSE στο σύνολο αξιολόγησης και ακολουθείται από τον μέσο όρο και την τυπική απόκλιση των 30 εκτελέσεων μέσα σε παρένθεση. Το καλύτερο αποτέλεσμα παρουσιάζεται με έντονους χαρακτήρες.

[α]Ο αριθμός των κόμβων των στοιβάδων MLP δίνεται στη μορφή:[1η στοιβάδα 2η στοιβάδα]

Πίνακας 3.9. Αποτελέσματα για τα σύνολα δεδομένων προσέγγισης συνάρτησης – RMSE, ασαφής διαμέριση, ποσοστό κάλυψης πλησιέστερων κέντρων, αριθμός κόμβων και απαιτούμενες επαναλήψεις εύρεσης της καλύτερης λύσης *(συνέχεια)*

Σύνολο δεδομένων	**Αλγόριθμος**	**Ασαφής διαμέριση**	**Ποσοστό κάλυψης πλησιέστερων κέντρων**	**Αριθμός κόμβων**[α]	**Επαναλήψεις εύρεσης της καλύτερης λύσης**
Auto Price	COOP-VW	[4 5 17 4 13 9 50 4 13 21 4 12 50 7 29]	2.00%	**50** (60±6)	1224 (**801**±421)
	SINGLE-VW	[4 5 17 4 4 12 50 4 4 4 50 50 6 47 50]	2.00%	56 (59±8)	1738 (1293±560)
	PSO-NSFM	[4 4 29 50 50 23 7 34 5 10 24 4 35 4 37]	-	57 (**57**±5)	**832** (825±295)
	MLP	-	-	[5 5]	-
Energy	COOP-VW	[50 15 4 4 17 6 7 14]	99.00%	**174** (**198**±20)	**64** (**732**±365)
	SINGLE-VW	[6 6 50 9 50 4 50 49]	99.00%	212 (208±23)	1702 (1239±687)
	PSO-NSFM	[9 44 4 5 50 20 8 22]	-	313 (345±31)	1643 (1367±417)
	MLP	-	-	[39 27]	-
Friedman	COOP-VW	[4 7 10 4 11]	99.00%	99 (**104**±11)	**78** (**51**±33)
	SINGLE-VW	[4 28 4 6 4]	93.64%	**84** (131±48)	426 (97±59)
	PSO-NSFM	[8 4 4 32 4]	-	114 (116±66)	245 (949±285)
	MLP	-	-	[12 6]	-
Parkinsons	COOP-VW	[8 12 15 4 50 33 4 4 9 18 50 12 48 27 50 4]	5.34%	426 (**403**±97)	**868** (**657**±442)
	SINGLE-VW	[4 17 9 50 4 8 21 4 6 10 50 17 50 50 32 4]	2.00%	434 (**403**±82)	1646 (1078±594)
	PSO-NSFM	[32 26 14 4 4 9 19 4 12 15 12 4 48 50 6 50]	-	**341** (423±73)	999 (949±122)
	MLP	-	-	[12 10]	-
Samad	COOP-VW	[5 4 39]	93.41%	**83** (88±7)	104 (**85**±50)
	SINGLE-VW	[17 4 14]	73.15%	91 (91±10)	315 (146±103)
	PSO-NSFM	[4 6 38]	-	89 (**87**±19)	**89** (178±72)
	MLP	-	-	[11 11]	-

Οι αλγόριθμοι εκτελέστηκαν 30 φορές για κάθε σύνολο δεδομένων. Απεικονίζεται η εκτέλεση για την οποία βρέθηκε το καλύτερο αποτέλεσμα σύμφωνα με τον δείκτη RMSE στο σύνολο αξιολόγησης και ακολουθείται από τον μέσο όρο και την τυπική απόκλιση των 30 εκτελέσεων μέσα σε παρένθεση. Το καλύτερο αποτέλεσμα παρουσιάζεται με έντονους χαρακτήρες.

[α]Ο αριθμός των κόμβων των στοιβάδων MLP δίνεται στη μορφή:[1η στοιβάδα 2η στοιβάδα]

Πίνακας 3.10. Αποτελέσματα για τα σύνολα δεδομένων ταξινόμησης – MCC, ασαφής διαμέριση, ποσοστό κάλυψης πλησιέστερων κέντρων, αριθμός κόμβων και απαιτούμενες επαναλήψεις εύρεσης της καλύτερης λύσης

Σύνολο δεδομένων	**Αλγόριθμος**	**MCC συνόλου αξιολόγησης**	**MCC συνόλου ελέγχου**
Cardio-tocography FHR	COOP-VW	**8.27×10^{-1}** (**8.08×10^{-1}**$\pm1.31\times10^{-2}$)	**7.88×10^{-1}** ($7.84\times10^{-1}\pm1.44\times10^{-2}$)
	SINGLE-VW	8.25×10^{-1} ($8.05\times10^{-1}\pm1.34\times10^{-2}$)	7.87×10^{-1} ($7.84\times10^{-1}\pm1.61\times10^{-2}$)
	PSO-NSFM	8.25×10^{-1} ($8.07\times10^{-1}\pm1.35\times10^{-2}$)	**7.88×10^{-1}** (**7.85×10^{-1}**$\pm1.63\times10^{-2}$)
	MLP	7.81×10^{-1} ($7.54\times10^{-1}\pm2.61\times10^{-2}$)	7.86×10^{-1} ($7.77\times10^{-1}\pm2.53\times10^{-2}$)
Iris	COOP-VW	**1.00×10^{0}** (**1.00×10^{0}**$\pm0.00\times10^{0}$)	**1.00×10^{0}** (**9.73×10^{-1}**$\pm3.66\times10^{-2}$)
	SINGLE-VW	**1.00×10^{0}** (**1.00×10^{0}**$\pm0.00\times10^{0}$)	**1.00×10^{0}** ($9.64\times10^{-1}\pm3.89\times10^{-2}$)
	PSO-NSFM	**1.00×10^{0}** (**1.00×10^{0}**$\pm0.00\times10^{0}$)	**1.00×10^{0}** ($9.64\times10^{-1}\pm4.60\times10^{-2}$)
	MLP	**1.00×10^{0}** ($9.70\times10^{-1}\pm1.58\times10^{-2}$)	9.59×10^{-1} ($9.47\times10^{-1}\pm2.83\times10^{-2}$)
Leaf	COOP-VW	**9.08×10^{-1}** (**8.96×10^{-1}**$\pm9.48\times10^{-3}$)	**8.50×10^{-1}** (**8.00×10^{-1}**$\pm2.82\times10^{-2}$)
	SINGLE-VW	**9.08×10^{-1}** ($8.93\times10^{-1}\pm1.37\times10^{-2}$)	7.81×10^{-1} ($7.94\times10^{-1}\pm4.09\times10^{-2}$)
	PSO-NSFM	8.97×10^{-1} ($8.83\times10^{-1}\pm9.86\times10^{-3}$)	7.12×10^{-1} ($7.62\times10^{-1}\pm3.22\times10^{-2}$)
	MLP	7.46×10^{-1} ($6.52\times10^{-1}\pm4.74\times10^{-2}$)	7.01×10^{-1} ($6.31\times10^{-1}\pm5.47\times10^{-2}$)
Waveform	COOP-VW	**8.40×10^{-1}** (**8.33×10^{-1}**$\pm3.49\times10^{-3}$)	**8.02×10^{-1}** (**8.05×10^{-1}**$\pm7.79\times10^{-3}$)
	SINGLE-VW	8.36×10^{-1} ($8.30\times10^{-1}\pm3.59\times10^{-3}$)	7.88×10^{-1} ($8.01\times10^{-1}\pm8.01\times10^{-3}$)
	PSO-NSFM	8.34×10^{-1} ($8.29\times10^{-1}\pm3.51\times10^{-3}$)	8.01×10^{-1} ($8.00\times10^{-1}\pm7.98\times10^{-3}$)
	MLP	8.10×10^{-1} ($7.94\times10^{-1}\pm7.62\times10^{-3}$)	7.95×10^{-1} ($7.98\times10^{-1}\pm5.52\times10^{-3}$)

Οι αλγόριθμοι εκτελέστηκαν 30 φορές για κάθε σύνολο δεδομένων. Απεικονίζεται η εκτέλεση για την οποία βρέθηκε το καλύτερο αποτέλεσμα σύμφωνα με τον δείκτη MCC στο σύνολο αξιολόγησης και ακολουθείται από τον μέσο όρο και την τυπική απόκλιση των 30 εκτελέσεων μέσα σε παρένθεση. Το καλύτερο αποτέλεσμα παρουσιάζεται με έντονους χαρακτήρες.

[α]Ο αριθμός των κόμβων των στοιβάδων MLP δίνεται στη μορφή:[1η στοιβάδα 2η στοιβάδα]

Πίνακας 3.10. Αποτελέσματα για τα σύνολα δεδομένων ταξινόμησης – MCC, ασαφής διαμέριση, ποσοστό κάλυψης πλησιέστερων κέντρων, αριθμός κόμβων και απαιτούμενες επαναλήψεις εύρεσης της καλύτερης λύσης *(συνέχεια)*

Σύνολο δεδομένων	**Αλγόριθμος**	**Ασαφής διαμέριση**	**Ποσοστό κάλυψης πλησιέστερων κέντρων**	**Αριθμός κόμβων[α]**	**Επαναλήψεις εύρεσης της καλύτερης λύσης**
Cardio-tocography FHR	COOP-VW	[17 4 4 4 4 50 4 4 4 49 4 4 4 31 4 15 50 4 7 50 4]	96.25%	**248** (**495**±228)	**1073** (**995**±325)
	SINGLE-VW	[4 7 45 7 4 4 4 4 50 50 4 50 6 4 4 4 48 50 4 50 4]	4.53%	480 (526±229)	1892 (1502±255)
	PSO-NSFM	[4 50 50 4 4 17 50 4 16 4 4 4 4 4 48 4 4 4 5 50 50]	-	342 (635±252)	1493 (1097±445)
	MLP	-	-	[36 33]	-
Iris	COOP-VW	[24 4 14 6]	66.56%	**21** (**41**±13)	**2** (**2.13**±0.43)
	SINGLE-VW	[16 20 31 9]	84.56%	44 (46±13)	4 (3.09±1.09)
	PSO-NSFM	[31 27 16 50]	-	64 (52±12)	3 (2.20±0.41)
	MLP	-	-	[39 30]	-
Leaf	COOP-VW	[4 4 7 4 17 50 11 4 4 50 50 4 8 4]	58.17%	**94** (105±13)	**692** (**728**±305)
	SINGLE-VW	[4 23 4 4 4 43 7 43 4 49 50 4 19 4]	99.00%	98 (**104**±16)	1418 (1398±589)
	PSO-NSFM	[10 46 48 28 5 45 14 13 38 11 4 44 48 12]	-	147 (129±15)	1979 (1223±528)
	MLP	-	-	[35 33]	-
Waveform	COOP-VW	[4 4 4 4 30 4 4 4 39 6 4 5 4 4 4 5 4 4 4 4 4]	47.00%	**43** (61±23)	1219 (**660**±441)
	SINGLE-VW	[4 4 4 4 16 4 49 4 4 5 4 4 6 4 4 4 9 4 10 4 4]	2.00%	54 (70±25)	1565 (1022±586)
	PSO-NSFM	[4 4 4 4 47 26 4 4 4 4 5 4 9 4 4 4 4 4 4 4 4]	-	57 (**49**±18)	**1170** (1044±524)
	MLP	-	-	[39 5]	-

Οι αλγόριθμοι εκτελέστηκαν 30 φορές για κάθε σύνολο δεδομένων. Απεικονίζεται η εκτέλεση για την οποία βρέθηκε το καλύτερο αποτέλεσμα σύμφωνα με τον δείκτη MCC στο σύνολο αξιολόγησης και ακολουθείται από τον μέσο όρο και την τυπική απόκλιση των 30 εκτελέσεων μέσα σε παρένθεση. Το καλύτερο αποτέλεσμα παρουσιάζεται με έντονους χαρακτήρες.

[α]Ο αριθμός των κόμβων των στοιβάδων MLP δίνεται στη μορφή:[1η στοιβάδα 2η στοιβάδα]

3.3 Εκπαίδευση ταξινομητών RBF με τον αλγόριθμο εξελικτικής προσομοιωμένης ανόπτησης

Ο αλγόριθμος εξελικτικής προσομοιωμένης ανόπτησης (ESA), ο οποίος αναλύεται στο κεφάλαιο 1.4.1, συνδυάζει βασικά πλεονεκτήματα της μεθόδου προσομοιωμένης ανόπτησης και των εξελικτικών αλγορίθμων. Με βάση αυτά τα πλεονεκτήματα, ο αλγόριθμος εξελικτικής προσομοιωμένης ανόπτησης θα μπορούσε να αποτελέσει ένα κατάλληλο εργαλείο για τη βελτιστοποίηση της μη συμμετρικής διαμέρισης, με σκοπό την κατασκευή μοντέλων RBF.

Σε αυτή τη διδακτορική διατριβή, παρουσιάζεται μια νέα μεθοδολογία για τη δημιουργία και τη βελτιστοποίηση μοντέλων RBF που παρουσιάζουν αυξημένες δυνατότητες ταξινόμησης. Η προτεινόμενη μέθοδος χρησιμοποιεί τον μη συμμετρικό αλγόριθμο των ασαφών μέσων, ο οποίος έχει τροποποιηθεί κατάλληλα για να εκπαιδεύει ταξινομητές RBF. Επιπλέον, η τεχνική εξελικτικής προσομοιωμένης ανόπτησης έχει προσαρμοστεί και ενσωματωθεί κατάλληλα στο προτεινόμενο σχήμα για να βελτιστοποιηθεί η ασαφής διαμέριση των παραγόμενων ταξινομητών, ως προς την ακρίβεια ταξινόμησης. Η προκύπτουσα μέθοδος ταξινόμησης εφαρμόζεται σε ιατρικά σύνολα δεδομένων που σχετίζονται με τη διάγνωση.

3.3.1 Βελτιστοποίηση της ασαφούς διαμέρισης με τον αλγόριθμο ESA

Ο αλγόριθμος ESA μπορεί να χρησιμοποιηθεί ως μέθοδος αναζήτησης για τον υπολογισμό της βέλτιστης μη συμμετρικής διαμέρισης, που δίνεται ως είσοδος στον αλγόριθμο NSFM. Στην περίπτωση αυτή, τα χρωμοσώματα που μετέχουν στην εξελικτική διαδικασία αντιπροσωπεύουν διαφορετικούς τρόπους διαμέρισης του χώρου εισόδου. Εφαρμόζοντας αυτού του είδους τον σχηματισμό, κάθε χρωμόσωμα $\mathbf{s}_p(q)$, αντιστοιχεί στον αριθμό των ασαφών συνόλων σε κάθε διάσταση:

$$\mathbf{s}_p(q) = \left[s_1(q),\ s_2(q),\ \ldots,\ s_N(q) \right]^T, \quad p = 1, 2, \ldots, P, \quad q = 1, 2, \ldots, Q \tag{3.25}$$

όπου P και Q συμβολίζουν το μέγεθος του πληθυσμού και τον συνολικό αριθμό των γενεών, αντίστοιχα. Σε κάθε γενιά, όλα τα χρωμοσώματα του πληθυσμού μεταβάλλονται σύμφωνα με τον τελεστή προσομοιωμένης ανόπτησης. Η τυχαία αλλαγή που εφαρμόζεται στα χρωμοσώματα σε κάθε επανάληψη του αλγορίθμου SA είναι της μορφής:

$$\mathbf{s}_p(q+1) = \mathrm{round}\left\{ \mathbf{s}_p(q) + \mathbf{r}(q) \right\} \tag{3.26}$$

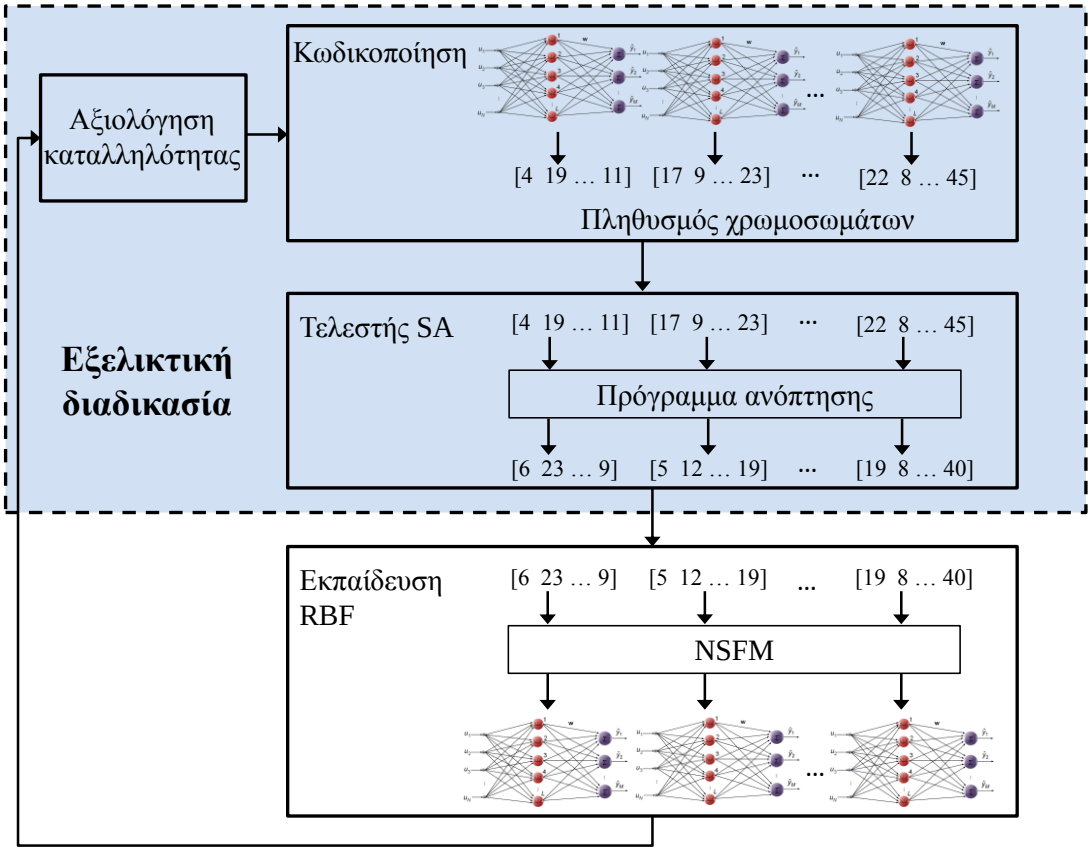
όπου $\mathbf{r}(q)$ είναι ένα διάνυσμα που περιλαμβάνει N τυχαίους αριθμούς οι οποίοι προέρχονται από μια κατανομή Gauss $\sim N(0, \sigma)$. Η παράμετρος σ προσδιορίζει την τυπική απόκλιση των τυχαία εφαρμοσμένων αλλαγών στα χρωμοσώματα. Τέλος, ο τελεστής «round» στρογγυλοποιεί τις τιμές των χρωμοσωμάτων στον πλησιέστερο ακέραιο. Αυτό συμβαίνει διότι τα χρωμοσώματα αποτελούνται από ασαφή σύνολα, των οποίων ο αριθμός σε κάθε διάσταση μπορεί να λάβει μόνο ακέραιες τιμές.

3.3.2 Εκπαίδευση ταξινομητών RBF με βάση τον αλγόριθμο εξελικτικής προσομοιωμένης ανόπτησης

Η βελτιστοποίηση της μη συμμετρικής ασαφούς διαμέρισης του χώρου εισόδου γίνεται χρησιμοποιώντας τον αλγόριθμο εξελικτικής προσομοιωμένης ανόπτησης, ο οποίος αποτελεί τη βάση για μια ολοκληρωμένη μεθοδολογία εκπαίδευσης ταξινομητών RBF. Ο αλγόριθμος ξεκινά χωρίζοντας τα διαθέσιμα δεδομένα με τυχαίο τρόπο σε τρία σύνολα δεδομένων: σύνολο εκπαίδευσης, σύνολο αξιολόγησης και σύνολο ελέγχου, όπως αναφέρεται στο κεφάλαιο 1.2.7.

Έπειτα, ο αλγόριθμος προχωρά στο στάδιο της αρχικοποίησης των σωματιδίων. Κατά το στάδιο της αρχικοποίησης των σωματιδίων, P διαφορετικές διαμερίσεις του χώρου εισόδου επιλέγονται με τυχαίο τρόπο, και κωδικοποιούνται ως χρωμοσώματα. Στη συνέχεια, κάθε χρωμόσωμα δημιουργεί έναν ταξινομητή RBF εφαρμόζοντας τον μη συμμετρικό αλγόριθμο των ασαφών μέσων (NSFM), ο οποίος υπολογίζει τον αριθμό και τις συντεταγμένες των κέντρων RBF με βάση την εξίσωση (2.26). Μετά τον καθορισμό των κέντρων RBF, τα συναπτικά βάρη υπολογίζονται χρησιμοποιώντας την εξίσωση (2.14). Με βάση τους παραγόμενους ταξινομητές RBF, για κάθε χρωμόσωμα υπολογίζεται μια τιμή για τη συνάρτηση καταλληλότητας, εφαρμόζοντας ένα κριτήριο σφάλματος. Ο προτεινόμενος αλγόριθμος χρησιμοποιεί ως κριτήριο σφάλματος την ακρίβεια% (σχέση (1.13)), αλλά μπορούν επίσης να χρησιμοποιηθούν και άλλα κριτήρια σφάλματος.

Το επόμενο βήμα περιλαμβάνει τον υπολογισμό μιας νέας γενιάς χρωμοσωμάτων εφαρμόζοντας τον τελεστή προσομοιωμένης ανόπτησης. Ο αλγόριθμος προσομοιωμένης ανόπτησης ξεκινά με το αρχικό χρωμόσωμα και εκτελεί τυχαίες αλλαγές χρησιμοποιώντας την εξίσωση (3.26). Η λύση που προκύπτει μπορεί να γίνει αποδεκτή, ή να απορριφθεί σύμφωνα με μία πιθανότητα η οποία υπολογίζεται από την κατανομή Boltzmann-Gibbs (εξίσωση (1.15)). Μόλις η θερμοκρασία φτάσει στην ελάχιστη τιμή της, ο αλγόριθμος προσομοιωμένης ανόπτησης επιστρέφει την παραγόμενη λύση, η οποία συγκρίνεται με την αρχική λύση. Η καλύτερη από τις δύο λύσεις περνά στην επόμενη γενιά. Στη συνέχεια, ο αλγόριθμος επιστρέφει στο στάδιο της δημιουργίας του ταξινομητή RBF, έως ότου ισχύσει μία προϋπόθεση διακοπής. Σε αυτή την περίπτωση, η προϋπόθεση διακοπής αφορά στον μέγιστο αριθμό γενεών που έχει επιτευχθεί. Μια αναπαράσταση της μεθόδου ESA-NSFM δίνεται στο σχήμα 3.10, ενώ μια επισκόπηση του αλγορίθμου ESA-NSFM παρουσιάζεται παρακάτω.



Σχήμα 3.10. Σχηματική αναπαράσταση του αλγορίθμου ESA-NSFM.

Αλγόριθμος 5 – Αλγόριθμος ESA-NSFM

Δεδομένα Εισόδου:

$\{\mathbf{U}_{train}, \mathbf{Y}_{train}\}$: Σύνολο δεδομένων εκπαίδευσης,

$\{\mathbf{U}_{val}, \mathbf{Y}_{val}\}$: Σύνολο δεδομένων αξιολόγησης,

s_{min}, s_{max}, σ: Ελάχιστος και μέγιστος αριθμός των ασαφών συνόλων που διαμερίζεται το πεδίο ορισμού κάθε μεταβλητής εισόδου, και τυπική απόκλιση των τυχαίων εφαρμοσμένων στον αριθμό των ασαφών συνόλων,

P: Πληθυσμός,

T_{Bmin}, T_{Bmax}, k_B, a_B: Λειτουργικές παράμετροι του αλγορίθμου ESA (Μέγιστες και ελάχιστες θερμοκρασίες, σταθερά Boltzmann, και σταθερά ψύξης, αντίστοιχα)

Q: Μέγιστος αριθμός γενεών

Δεδομένα Εξόδου:

L_f: Αριθμός επιλεγμένων κέντρων RBF,

$\hat{\mathbf{U}}_f$: Τοποθεσίες επιλεγμένων κέντρων RBF,

$\mathbf{W}_f$: Συναπτικά βάρη επιλεγμένων κέντρων RBF

1. **Για** $p = 1{:}P$ **Εκτέλεσε:**
2. Αρχικοποίησε τις συντεταγμένες των χρωμοσωμάτων με τυχαίους ακέραιους

αριθμούς μεταξύ των s_{min} και s_{max}: $\mathbf{s}_p(0) \leftarrow \left[\underbrace{\text{rand, rand, …, rand}}_{N}\right]$

3. **Τέλος Για**
4. **Για** q = 1:Q **Εκτέλεσε:**
5. **Για** p = 1:P **Εκτέλεσε:**
6. Δώσε το χρωμόσωμα $\mathbf{s}_p(q)$ και το σύνολο δεδομένων εκπαίδευσης $\{\mathbf{U}_{train}, \mathbf{Y}_{train}\}$ στον αλγόριθμο 2, προκειμένου να υπολογιστεί ο συνολικός αριθμός των κέντρων RBF $L_i(t)$ και οι θέσεις τους $\hat{\mathbf{u}}_p(t)$
7. Χρησιμοποίησε τον τύπο (2.14) για τον υπολογισμό των συναπτικών βαρών $\mathbf{W}_p(q)$
8. Υπολόγισε τη συνάρτηση καταλληλότητας $f(\mathbf{s}_p(q))$ για το σύνολο δεδομένων αξιολόγησης $\{\mathbf{U}_{val}, \mathbf{Y}_{val}\}$
9. Αρχικοποίησε τη λύση $\mathbf{s}'_p(0)$ για τη διαδικασία ανόπτησης $\mathbf{s}'_p(0) \leftarrow \mathbf{s}_p(q)$ και τις επαναλήψεις ανόπτησης $t \leftarrow 0$
10. **Όσο** η τιμή της θερμοκρασίας T είναι μεγαλύτερη από την τιμή T_{Bmin} **Εκτέλεσε:**
11. Προχώρα στην επόμενη επανάληψη ανόπτησης $t \leftarrow t+1$
12. Υπολόγισε την καινούρια λύση $\mathbf{s}'_p(t)$ χρησιμοποιώντας τον τύπο (3.26)
13. **Αν** η παλιά λύση $\mathbf{s}'_p(t-1)$ είναι καλύτερη από την καινούρια λύση $\mathbf{s}'_p(t)$
14. Υπολόγισε την πιθανότητα Boltzmann p_B χρησιμοποιώντας τον τύπο (1.15)
15. **Αν** ένας τυχαίος αριθμός είναι μεγαλύτερος από την p_B
16. Αντικατέστησε την παλιά λύση με την καινούρια $\mathbf{s}'_p(t) \leftarrow \mathbf{s}'_p(t-1)$
17. **Τέλος Αν**
18. **Τέλος Αν**
19. Ενημέρωσε τη θερμοκρασία χρησιμοποιώντας τον τύπο (1.16)
20. **Τέλος Όσο**
21. **Αν** η τελική λύση της προσομοιωμένης ανόπτησης $\mathbf{s}'_p(t)$ είναι καλύτερη από τη λύση της τρέχουσας γενιάς $\mathbf{s}_p(q)$
22. Αντικατέστησε τη λύση της τρέχουσας γενιάς με την τελική λύση της προσομοιωμένης ανόπτησης $\mathbf{s}_p(q) \leftarrow \mathbf{s}'_p(t)$
23. **Τέλος Αν**
24. **Τέλος Για**
25. **Τέλος Για**

3.3.3 Εφαρμογή του αλγορίθμου ESA-NSFM στην ιατρική διάγνωση

Η ιατρική διάγνωση αναφέρεται στην αναγνώριση μιας ασθένειας με βάση τις ενδείξεις και τα συμπτώματα, ή, με άλλα λόγια, την ταξινόμηση των συμπτωμάτων σε κάποια γνωστή κατηγορία ασθένειας. Καθώς η ιατρική διάγνωση, συχνά, εμπεριέχει έναν βαθμό αβεβαιότητας, οι γιατροί θα μπορούσαν να επωφεληθούν από τα οφέλη κατάλληλων συστημάτων υποστήριξης αποφάσεων, τα οποία είναι σε θέση να αξιολογούν ιατρικές καταστάσεις. Τα άφθονα δεδομένα που συσσωρεύονται σε μορφή ιατρικών αρχείων θα μπορούσαν να αποτελέσουν μεγάλη βοήθεια στον σχεδιασμό τέτοιων συστημάτων. Ωστόσο, η απελευθέρωση της γνώσης που εμπεριέχεται σε αυτά τα ιστορικά δεδομένα συμπεριλαμβάνει μια σειρά από δυσκολίες. Η σχέση μεταξύ της διαδικασίας της διάγνωσης και των συναφών συμπτωμάτων ή των κλινικών ευρημάτων είναι συνήθως πολύπλοκη. Επιπλέον, η φύση της προαναφερθείσας σχέσης είναι μη-γραμμική και οι μηχανισμοί που διέπουν τη λειτουργία του ανθρώπινου σώματος είναι αρκετά συχνά ελάχιστα κατανοητοί. Αυτά τα γεγονότα εμποδίζουν τη διατύπωση ρητών νόμων που θα μπορούσαν να προσαρμοστούν στα διαθέσιμα στοιχεία και έτσι η διαδικασία της εξόρυξης γνώσης από τα ιατρικά αρχεία γίνεται ένα αρκετά δύσκολο έργο.

Αν και οι προαναφερθείσες δυσκολίες δυσχεραίνουν το συγκεκριμένο έργο, πρόσφατες εξελίξεις στις μεθοδολογίες της μηχανικής μάθησης και της εξόρυξης δεδομένων καταλύουν την ανάπτυξη συστημάτων υποστήριξης αποφάσεων σε ιατρική διάγνωση. Υπό αυτό το πρίσμα, για την κατασκευή ιατρικών διαγνωστικών εργαλείων έχει χρησιμοποιηθεί μια ποικιλία από τεχνικές διάγνωσης [205]. Ένα σημαντικό μέρος των τεχνικών αυτών περιλαμβάνει τη γνωστή οικογένεια των νευρωνικών δικτύων.

Ένας σημαντικός αριθμός επιστημονικών δημοσιεύσεων, που αξιοποιούν τα νευρωνικά δίκτυα για τη δημιουργία ιατρικών διαγνωστικών εργαλείων, έχει εμφανιστεί κατά τη διάρκεια των τελευταίων δεκαετιών. Μια μέθοδος που κάνει χρήση των διακυμάνσεων της ηλεκτρικής εμπέδησης σε σχέση με τη συχνότητα, με σκοπό τη διάκριση του βασικοκυτταρικού καρκινώματος από τις καλοήθεις βλάβες του δέρματος, η οποία βασίζεται σε ταξινομητές νευρωνικών δικτύων, παρουσιάζεται στην εργασία [206]. Οι συγγραφείς στην εργασία [207] παρουσιάζουν μία μέθοδο εξόρυξης πληροφοριών από βιοδείκτες, η οποία βασίζεται σε ένα νευρωνικό δίκτυο και χρησιμοποιείται στην ταξινόμηση του καρκίνου. Ένα σύστημα αποκαλούμενο ως GEMS (gene expression model selector) έχει αναπτυχθεί με σκοπό την αυτοματοποιημένη δημιουργία και αξιολόγηση υψηλής ποιότητας διαγνωστικών μοντέλων καρκίνου, συμπεριλαμβανομένων των μεθοδολογιών των νευρωνικών δικτύων [208]. Αρκετές τεχνικές ταξινομητών που βασίζονται σε νευρωνικά δίκτυα έχουν προταθεί για τη διάγνωση καρδιακών παθήσεων [209]. Στη δημοσίευση [210] μελετάται η εκπαίδευση ταξινομητών νευρωνικών δικτύων που χρησιμοποιούνται στη λήψη ιατρικών αποφάσεων. Η εν λόγω εργασία επικεντρώνεται στις επιπτώσεις που έχουν τα μη ισορροπημένα σύνολα δεδομένων ως προς την κατανομή των κατηγοριών, στην απόδοση ταξινόμησης. Στην εργασία [211] διερευνάται η επιλογή των χαρακτηριστικών στην κατασκευή ταξινομητών που βασίζονται σε μηχανές διανυσμάτων υποστήριξης αποφάσεων και εφαρμόζονται σε βιοϊατρικά διαγνωστικά σύνολα δεδομένων. Η χρήση των νευρωνικών δικτύων στην ταξινόμηση ιατρικών δεδομένων εξετάζεται στη δημοσίευση [212]. Άλλες παρόμοιες τεχνικές μηχανικής μάθησης που χρησιμοποιούνται στην κατασκευή ιατρικών διαγνωστικών εργαλείων περιλαμβάνουν μεθόδους Bayesian [213-215], ασαφή λογική [216], δέντρα απόφασης [217] και πολυμεταβλητών στατιστικών τεχνικών [218].

Λόγω των πλεονεκτημάτων που παρουσιάζουν τα νευρωνικά δίκτυα RBF, τα οποία έχουν αναλυθεί στο κεφάλαιο 2, τα τελευταία έχουν χρησιμοποιηθεί επιτυχώς σε πολλές

εφαρμογές ανάπτυξης ιατρικών διαγνωστικών εργαλείων [219-221]. Ειδικά στον τομέα αυτό, η βελτίωση της ακρίβειας των προβλέψεων του δικτύου είναι ζωτικής σημασίας για την επίτευξη υψηλών επιδόσεων διάγνωσης. Δυστυχώς, η εύρεση βέλτιστων τιμών για τις παραμέτρους του δικτύου RBF, όπως και με άλλες αρχιτεκτονικές νευρωνικών δικτύων, είναι αρκετά δύσκολη. Στη βιβλιογραφία φαίνεται ότι το ενδιαφέρον μετατοπίζεται στη βελτιστοποίηση της διαδικασίας εκπαίδευσης των δικτύων RBF μέσω της χρήσης εναλλακτικών μεθόδων, όπως οι τεχνικές που βασίζονται στον εξελικτικό υπολογισμό.

Λόγω της ικανότητάς τους να ξεφεύγουν από τα τοπικά βέλτιστα, οι τεχνικές εξελικτικού υπολογισμού έχουν χρησιμοποιηθεί ευρέως σε διάφορους και ποικίλους επιστημονικούς τομείς, συμπεριλαμβανομένων των εφαρμογών ιατρικού ενδιαφέροντος [222-225].

Η προτεινόμενη μέθοδος συνδυάζει τα πλεονεκτήματα των νευρωνικών δικτύων RBF και της τεχνικής εξελικτικής προσομοιωμένης ανόπτησης. Για να αποδειχθεί η αποτελεσματικότητα της προτεινόμενης μεθόδου ως διαγνωστικό εργαλείο στην ιατρική, ο προτεινόμενος αλγόριθμος εφαρμόστηκε σε διάφορα σύνολα δεδομένων ιατρικού ενδιαφέροντος τύπου benchmark, τα οποία είναι διαθέσιμα στη βάση UCI Machine Learning Repository [190]. Κοινό στόχο σε όλα τα σύνολα δεδομένων αποτελεί η πρόβλεψη της κατηγορίας εξόδου, η οποία βασίζεται σε μια σειρά από μεταβλητές εισόδου, οι οποίες εμπεριέχουν συνεχείς ή διατάξιμες (ordinal) τιμές. Στον πίνακα 3.11 παρουσιάζονται όλα τα σύνολα δεδομένων που χρησιμοποιήθηκαν σε αυτή την εφαρμογή, μαζί με τον αριθμό των μεταβλητών εισόδου, το σύνολο των δεδομένων, τον αριθμό των κλάσεων εξόδου και τις θετικές περιπτώσεις των τελευταίων (μόνο για τα σύνολα δεδομένων με δύο κλάσεις εξόδου). Μερικά από τα σύνολα δεδομένων εμπεριείχαν παραδείγματα με ελλιπείς τιμές, οι οποίες για τους σκοπούς της παρούσας μελέτης διαγράφηκαν. Μια σύντομη περιγραφή για κάθε σύνολο δεδομένων δίνεται παρακάτω.

- **Cardiotocography (FHR, NSP)**: Αυτά τα δύο σύνολα δεδομένων δεδομένων περιγράφονται στο κεφάλαιο 3.2.3.
- **EEG Eye State**: Αυτό το σύνολο δεδομένων έχει διαμορφωθεί από συνεχείς μετρήσεις EEG που προέρχονται από το Emotiv EEG Neuroheadset. Ο στόχος είναι ο προσδιορισμός της κατάστασης του οφθαλμού, ο οποίος βασίζεται σε 14 διαφορετικά χαρακτηριστικά EEG.
- **Pima Indians Diabetes**: Ο σκοπός αυτού του συνόλου δεδομένων είναι η δημιουργία ενός ταξινομητή που θα προβλέπει κατά πόσο ένας ασθενής θα αναπτύξει διαβήτη, με βάση τα κλινικά ευρήματα. Η διαθέσιμη βάση δεδομένων περιέχει δεδομένα από γυναίκες ασθενείς, τουλάχιστον 21 ετών, οι οποίες είναι απόγονοι των Pima Indians. Το σύνολο δεδομένων περιέχει 39 περιπτώσεις με ελλιπείς τιμές
- **Thyroid ANN**: Ο στόχος σε αυτό το σύνολο δεδομένων είναι η διάκριση μεταξύ των τριών πιθανών καταστάσεων του θυρεοειδούς αδένα (ευθυρεοειδισμό, υπερθυρεοειδισμό ή υποθυρεοειδισμό). Τα δεδομένα εισόδου αποτελούνται από 5 διαφορετικές ορμόνες που προκύπτουν από εργαστηριακούς ελέγχους σε δείγματα αίματος των ασθενών.
- **Vertebral Column (3 Classes–2 Classes)**: Τα σύνολα δεδομένων «Vertebral Column» αφορούν στη διάγνωση των ατόμων που παρουσιάζουν διαταραχές στη σπονδυλική στήλη. Κάθε ασθενής περιγράφεται με 6 χαρακτηριστικά, τα οποία αφορούν στο σχήμα και στον προσανατολισμό της λεκάνης και της οσφυϊκής μοίρας της σπονδυλικής του στήλης. Οι συνολικά 310 ασθενείς κατατάσσονται στην περίπτωση του συνόλου δεδομένων «Vertebral Column Dataset – 3 Classes» σε μία από τις τρεις κατηγορίες: φυσιολογική (υγιείς άνθρωποι), δισκοκήλη (άνθρωποι που πάσχουν από δισκοκήλη) και

σπονδυλολίσθηση (άνθρωποι που πάσχουν από σπονδυλολίσθηση). Ενώ, στην περίπτωση του συνόλου δεδομένων «Vertebral Column Dataset – 2 Classes» οι ίδιοι ασθενείς διακρίνονται στους υγιείς ανθρώπους (κατηγορία: φυσιολογική), και σε εκείνους που πάσχουν είτε από δίσκοκήλη είτε από σπονδυλολίσθηση (κατηγορία: αφύσικη/ανώμαλη).

- **Wisconsin Original Breast Cancer (WOBC)–Wisconsin Diagnostic Breast Cancer (WDBC)**: Αυτά τα δύο σύνολα δεδομένων [46,47] περιλαμβάνουν μετρήσεις που προκύπτουν από ψηφιοποιημένες εικόνες μαζών του μαστού. Αμφότερα τα σύνολα δεδομένων σκοπεύουν στη διάκριση των καλοηθών όγκων από τους κακοήθεις. Όμως, κάθε σύνολο δεδομένων βασίζεται σε διαφορετικές μεταβλητές εισόδου. Τέλος, το σύνολο δεδομένων WOBC περιέχει 16 περιπτώσεις με ελλιπείς τιμές, οι οποίες έχουν διαγραφεί.

Για κάθε σύνολο δεδομένων, τα δεδομένα χωρίστηκαν με τυχαίο τρόπο στα υποσύνολα εκπαίδευσης, αξιολόγησης και ελέγχου και σε ποσοστά 50% - 25% - 25%, αντίστοιχα. Ιδιαίτερη προσοχή δόθηκε στον διαχωρισμό των κλάσεων, έτσι ώστε τα δεδομένα που ανήκουν σε κάθε κατηγορία εξόδου να ακολουθούν την παραπάνω αναλογία για να μοιραστούν στα 3 υποσύνολα δεδομένων.

Η επιλογή των λειτουργικών παραμέτρων, βασίστηκε στο θεωρητικό υπόβαθρο, καθώς και σε πειράματα που διεξήχθησαν. Όσον αφορά στις παραμέτρους που σχετίζονται με το κομμάτι του αλγορίθμου NSFM, ο ελάχιστος και ο μέγιστος αριθμός των ασαφών συνόλων, καθορίζουν τα κατώτερα και τα ανώτερα όρια του χώρου αναζήτησης, αντίστοιχα. Προηγούμενα πειράματα που είχαν εκτελεστεί χρησιμοποιώντας τη συμμετρική και τη μη-συμμετρική έκδοση του αλγορίθμου FM έδειξαν ότι οι διαμερίσεις εκτός αυτών των ορίων δεν είναι πιθανό να οδηγήσουν σε πετυχημένα μοντέλα. Η τυπική απόκλιση σ των αλλαγών που υιοθετούνται, στην πραγματικότητα, καθορίζει το προσδοκώμενο μέγεθος της

Πίνακας 3.11. Επισκόπηση συνόλων δεδομένων τύπου benchmark

Σύνολο δεδομένων	**Αριθμός μεταβλητών εισόδου**	**Αριθμός δεδομένων**	**Αριθμός κατηγοριών εξόδου**	**Αριθμός θετικών περιπτώσεων[α]**
Cardiotocography FHR	21	2126	10	-
Cardiotocography NSP	21	2126	3	-
EEG Eye State	14	14980	2	6723
Pima Indians Diabetes	8	768	2	268
Thyroid ANN	6	7200	3	-
Vertebral Column – 2 Classes	6	310	2	210
Vertebral Column – 3 Classes	6	310	3	-
Wisconsin Diagnostic Breast Cancer (WDBC)	30	569	2	212
Wisconsin Original Breast Cancer (WOBC)	10	699	2	241

[α]Η στήλη αυτή αναφέρεται μόνο στα σύνολα δεδομένων που περιέχουν δύο κατηγορίες εξόδου

διαφοροποίησης που θα εφαρμοστεί στα ασαφή σύνολα, σε κάθε επανάληψη της διαδικασίας προσομοιωμένης ανόπτησης. Αυτή η παράμετρος μπορεί να χρησιμοποιηθεί στον χειρισμό της εξισορρόπησης μεταξύ ολικής και τοπικής του αλγορίθμου ESA-NSFM.

Όταν η παράμετρος *σ* έχει μικρή τιμή, τότε επιτυγχάνεται διεξοδική τοπική αναζήτηση του χώρου, αλλά η ολική εξερεύνησή του δεν είναι ικανοποιητική. Αντιθέτως, οι μεγάλες τιμές της παραμέτρου *σ* οδηγούν σε διεξοδική ολική διερεύνησή του χώρου αναζήτησης, αλλά σε τοπικό επίπεδο η εξερεύνησή του είναι ελλιπής, ενώ οι μεγάλες τιμές έχουν το αντίθετο αποτέλεσμα. Προφανώς, η επιλογή αυτής της παραμέτρου είναι στενά συνδεδεμένη με το μέγεθος του χώρου αναζήτησης. Η τελευταία εξαρτάται, από τις τιμές s_{min} και s_{max} οι οποίες καθορίζουν τα όρια της αναζήτησης, και από τη διάσταση του χώρου εισόδου η οποία αντιστοιχεί στη διάσταση του χώρου αναζήτησης. Όταν η διάσταση του χώρου εισόδου αυξάνεται, η παράμετρος *σ* θα πρέπει επίσης να μπορεί να αυξηθεί, προκειμένου να διασφαλιστεί ότι ο χώρος αναζήτησης μπορεί να εξερευνηθεί επαρκώς εντός εύλογου αριθμού επαναλήψεων προσομοίωσης. Σε αυτή την εφαρμογή, οι δύο διαφορετικές τιμές για την παράμετρο *σ* χρησιμοποιούνται, ανάλογα με τον χώρο εισόδου της διάστασης του προβλήματος. Οι τιμές επιλέχθηκαν μετά από πειραματικές δοκιμές, αφού διαπιστώθηκε ότι παρέχουν καλή ισορροπία μεταξύ ολικής και τοπικής εξερεύνησης του χώρου.

Σε ό,τι αφορά στις παραμέτρους του αλγορίθμου ESA, το μέγεθος του πληθυσμού και ο μέγιστος αριθμός των γενεών είναι παράμετροι που επηρεάζουν σημαντικά το υπολογιστικό φορτίο της μεθόδου. Ομοίως, οι ελάχιστες και οι μέγιστες τιμές θερμοκρασίας, μαζί με τη σταθερά ψύξης, στην πραγματικότητα καθορίζουν τον συνολικό αριθμό των επαναλήψεων που θα εκτελεστούν από τον τελεστή προσομοιωμένης ανόπτησης σε κάθε γενιά. Σε αυτή την εφαρμογή, οι αντίστοιχες τιμές των προαναφερθέντων παραμέτρων επιλέχθηκαν έτσι ώστε να εκτελούνται 500 επαναλήψεις. Πειράματα έδειξαν ότι η αύξηση αυτών των τριών παραμέτρων (πληθυσμός, αριθμός γενεών και αριθμός επαναλήψεων προσομοιωμένης ανόπτησης), θα μπορούσε να οδηγήσει σε βελτιωμένη ακρίβεια ταξινόμησης, παρόλο το κόστος του πρόσθετου υπολογιστικού φορτίου. Για τους σκοπούς της παρούσας διατριβής χρησιμοποιήθηκαν μικρές τιμές για τις παραμέτρους αυτές, καθώς τα πειράματά που διεξήχθησαν έδειξαν ότι ικανοποιητικά αποτελέσματα μπορούν να επιτευχθούν σε σχετικά γρήγορους υπολογιστικούς χρόνους. Τέλος, η σταθερά Boltzmann ελέγχει κατά πόσο μια χειρότερη λύση σε σύγκριση με την τρέχουσα λύση θα γίνει δεκτή, ή θα απορριφθεί στο εσωτερικό του βρόχου της προσομοιωμένης ανόπτησης. Επίσης, η επιλογή αυτής της παραμέτρου επηρεάζει την εξισορρόπηση μεταξύ ολικής και τοπικής εξερεύνησης του χώρου. Η επίδραση όμως της σταθεράς Boltzmann, κατά κάποιον τρόπο εξασθενεί, λόγω του εξελικτικού κύκλου που επαναλαμβάνει τη διαδικασία προσομοιωμένης ανόπτησης χρησιμοποιώντας ως αρχική τιμή την τελική λύση που έχει παραχθεί από την προηγούμενη εκτέλεση του αλγορίθμου. Έτσι, μια πιο ομοιόμορφη κατανομή των τυχαίων κινήσεων παρέχεται κατά τη διαδικασία του αλγορίθμου ESA. Αυτό επαληθεύτηκε και πειραματικά, καθώς τα αποτελέσματα έδειξαν ότι υπήρχε μικρή ευαισθησία κατά την αλλαγή της σταθεράς Boltzmann γύρω από την τιμή 0.05, ενώ τα αποτελέσματα χειροτέρεψαν μόνο για σημαντικά χαμηλότερες ή υψηλότερες τιμές της.

Στον πίνακα 3.12 παρουσιάζονται οι παράμετροι λειτουργίας του αλγορίθμου ESA-NSFM. Όσον αφορά στα αποτελέσματα που παρουσιάζονται σε αυτή την εφαρμογή, σε όλα τα σύνολα δεδομένων χρησιμοποιήθηκαν οι ίδιες λειτουργικές παράμετροι, με εξαίρεση την παράμετρο *σ*. Το σκεπτικό πίσω από αυτή την επιλογή ήταν να αναδείξει την ικανότητα γενικής χρήσης της μεθόδου. Η προτεινόμενη μέθοδος μπορεί να εφαρμοστεί με επιτυχία σε ιατρικά αρχεία δεδομένων διαφορετικής φύσης, χωρίς να είναι αναγκαία μία διαδικασία βαθμονόμησης (tuning) των παραμέτρων. Βέβαια, θα μπορούσαν να επιτευχθούν περαιτέρω

Πίνακας 3.12. Λειτουργικές παράμετροι για τον αλγόριθμο ESA-NSFM

Αλγόριθμος	Παράμετρος	Σύμβολο	Τιμή
ESA	Πληθυσμός	P	12
	Μέγιστος αριθμός γενεών	G	20
	Μέγιστη θερμοκρασία	T_{Bmax}	100
	Ελάχιστη θερμοκρασία	T_{Bmin}	1
	Σταθερά ψύξης	a_B	0.991
	Σταθερά Boltzmann	k_B	0.05
NSFM	Ελάχιστος αριθμός ασαφών συνόλων	s_{min}	4
	Μέγιστος αριθμός ασαφών συνόλων	s_{max}	50
	Τυπική απόκλιση των τυχαία εφαρμοσμένων αλλαγών[α]	σ	S:8, L:12

[α]Τιμή που εξαρτάται από τη διαστατικότητα του χώρου εισόδου: Μικρά προβλήματα (Small - S): 1-10 μεταβλητές εισόδου, Μεγάλα προβλήματα (Large - L): περισσότερες από 10 μεταβλητές εισόδου

βελτιώσεις προσαρμόζοντας τις παραμέτρους ξεχωριστά για κάθε σύνολο δεδομένων, σύμφωνα με όσα ειπώθηκαν προηγουμένως.

Για λόγους σύγκρισης, δύο επιπλέον μεθοδολογίες κατηγοριοποίησης δοκιμάστηκαν, οι ταξινομητές RBF που εκπαιδεύτηκαν με τον αλγόριθμο των ασαφών μέσων [226] και οι μηχανές διανυσμάτων υποστήριξης αποφάσεων (Support Vector Machines, SVMs) [227] με γκαουσιανή συνάρτηση πυρήνα [228]. Σε όλες τις περιπτώσεις, οι παράμετροι των ταξινομητών προσδιορίστηκαν χρησιμοποιώντας το σύνολο δεδομένων εκπαίδευσης και στη συνέχεια η επιλογή μοντέλου επιτεύχθηκε με βάση το σύνολο δεδομένων αξιολόγησης, χρησιμοποιώντας το κριτήριο ακρίβεια%. Στην περίπτωση των ταξινομητών RBF που εκπαιδεύτηκαν με τον αλγόριθμο SFM, η επιλογή του μοντέλου ελέγχεται από μία μόνο παράμετρο, τον αριθμό των ασαφών συνόλων s στον οποίο διαμερίζεται κάθε μεταβλητή εισόδου. Η επιλογή της παραμέτρου s έγινε αφού δοκιμάστηκαν όλες οι διαμερίσεις από 4 ως 50 ασαφή σύνολα. Οι ταξινομητές SVM βελτιστοποιήθηκαν με βάση το εύρος των γκαουσιανών πυρήνων σ και τον συντελεστή ποινής (penalty factor) C. Μία διαδικασία διεξοδικής αναζήτησης εκτελέστηκε η οποία βασίστηκε σε μία μέθοδο αναζήτησης τύπου πλέγματος [229], δοκιμάζοντας όλους τους δυνατούς συνδυασμούς των σ και C. Στην περίπτωση των συνόλων δεδομένων με περισσότερες από δύο κατηγορίες εξόδου, εφαρμόστηκε μία τεχνική ένας-εναντίον-όλων (one-against-all), όπως προτάθηκε από τον Vapnik [228].

Μία τυπική μέθοδος σύγκρισης των επιδόσεων των ταξινομητών, η οποία χρησιμοποιείται επίσης στην παρούσα εφαρμογή, είναι το κριτήριο ακρίβεια %. Αυτού του είδους η μέτρηση αξιολογεί το συνολικό ποσοστό επιτυχίας του κάθε ταξινομητή. Ωστόσο, για να χαρακτηριστεί πλήρως η απόδοση του ταξινομητή χρησιμοποιήθηκε και ο συντελεστής MCC [167] (περιγράφεται στο κεφάλαιο 1.2.8).

3.3.4 Αποτελέσματα και συζήτηση

Τα αποτελέσματα όλων των συνόλων δεδομένων συνοψίζονται στους πίνακες 3.13 και 3.14. Στον πίνακα 3.13 απεικονίζονται οι δείκτες ακρίβεια% και MCC στα σύνολα δεδομένων αξιολόγησης και ελέγχου, και ακολουθούνται από τα διαστήματα εμπιστοσύνης ποσοστού 95%, μέσα σε παρένθεση. Οι βελτιστοποιημένες παράμετροι και οι υπολογιστικοί χρόνοι για τις τρεις μεθοδολογίες, μαζί με τις επιλεγμένες ασαφείς διαμερίσεις για τα δύο μοντέλα δικτύων RBF, παρουσιάζονται στον πίνακα 3.14. Όλοι οι υπολογιστικοί χρόνοι μετρήθηκαν σε ένα υπολογιστικό σύστημα με επεξεργαστή Intel Core 2 Quad (2.83 GHz) και μνήμη RAM 4GB. Όλες οι τιμές που παρουσιάζονται στους δύο πίνακες αντιστοιχούν στο μοντέλο που επιτυγχάνει το καλύτερο αποτέλεσμα σε σχέση με την ακρίβεια% στο σύνολο δεδομένων αξιολόγησης. Επίσης, όσον αφορά στον αλγόριθμο ESA-NSFM, ο μέσος όρος του πληθυσμού και τα διαστήματα εμπιστοσύνης ποσοστού 95% δίνονται σε παρενθέσεις. Στο σχήμα 3.11 παρουσιάζεται η εξέλιξη του κριτηρίου ακρίβεια% στο σύνολο δεδομένων αξιολόγησης για το καλύτερο χρωμόσωμα που παράγεται από τον αλγόριθμο ESA-NSFM σε κάθε περίπτωση. Ο οριζόντιος άξονας απεικονίζει τον αριθμό των επαναλήψεων (STSs) που διεξάγει η μέθοδος, συνδυάζοντας τις γενιές της εξελικτικής διαδικασίας με τις επαναλήψεις του αλγορίθμου προσομοιωμένης ανόπτησης. Θα πρέπει να σημειωθεί ότι, για λόγους καλύτερης απεικόνισης, κάθε γράφημα εμφανίζει την εξέλιξη του κριτηρίου ακρίβεια% μέχρι τον αριθμό των επαναλήψεων που για πρώτη φορά επιτυγχάνεται η αντίστοιχη καλύτερη λύση. Οι υπολογιστικοί χρόνοι που φαίνονται στον πίνακα 3.14 για τον αλγόριθμο ESA-NSFM, αντιστοιχούν στον πλήρη χρόνο που απαιτείται για τη διεξαγωγή των 10000 επαναλήψεων που πραγματοποιήθηκαν σε κάθε περίπτωση, και όχι στον χρόνο που χρειάστηκε πραγματικά η μέθοδος για να προσεγγίσει για πρώτη φορά την καλύτερη λύση.

Κάθε επανάληψη στις γραφικές παραστάσεις του σχήματος 3.11, ουσιαστικά αντιπροσωπεύει την απόδοση του αλγορίθμου NSFM σε μια συγκεκριμένη ασαφή διαμέριση που δοκιμάστηκε από τη μέθοδο ESA. Είναι προφανές ότι η απόδοση του αλγορίθμου NSFM ποικίλλει σημαντικά ανάλογα με την ασαφή διαμέριση, αλλά σε όλες τις περιπτώσεις που δοκιμάστηκαν, η μέθοδος ESA κατάφερε τελικά να βελτιώσει σημαντικά την απόδοση του πρώτου. Θα πρέπει, επίσης, να σημειωθεί ότι η ακρίβεια ταξινόμησης της μεθόδου ESA-NSFM συχνά παραμένει σταθερή για έναν αριθμό επαναλήψεων. Αυτό σημαίνει ότι ο αλγόριθμος παγιδεύεται γρήγορα σε μία λύση που αντιστοιχεί σε ένα τοπικό ελάχιστο. Η χρήση της μεθόδου ESA βοηθά τον αλγόριθμο να ξεφύγει από το τοπικό ελάχιστο και να ανακαλύψει ένα μονοπάτι που οδηγεί τελικά σε μια βελτιωμένη λύση. Θα πρέπει να σημειωθεί ότι το μονοπάτι προς τη βελτίωση της λύσης, συχνά περνά από χειρότερες λύσεις.

Μια σύγκριση μεταξύ των τριών ταξινομητών δείχνει ότι σχεδόν σε όλες τις περιπτώσεις η καλύτερη λύση που βρέθηκε από τον αλγόριθμο ESA-NSFM ξεπερνά τον αλγόριθμο SFM και τις μηχανές SVM όσον αφορά στην καλύτερη ακρίβεια% που υπολογίζεται τόσο στο σύνολο αξιολόγησης, όσο και στο σύνολο ελέγχου. Κάτι παρόμοιο παρατηρείται κατά τη σύγκριση των μεθόδων, όσον αφορά στον συντελεστή MCC, καθώς η καλύτερη λύση που βρέθηκε από τον αλγόριθμο ESA-NSFM υπερτερεί των αντιπάλων του στα 8 από τα 9 σύνολα δεδομένων, λαμβάνοντας υπόψη το σύνολο αξιολόγησης ή το σύνολο ελέγχου. Δεδομένου ότι ο προτεινόμενος αλγόριθμος είναι μια στοχαστική μέθοδος, σε κάθε εκτέλεση και για κάθε χρωμόσωμα παράγει ένα διαφορετικό αποτέλεσμα. Για τον λόγο αυτό, είναι επίσης ενδιαφέρον να γίνει μία σύγκριση με βάση τη μέση λύση του πληθυσμού του αλγορίθμου ESA-NSFM, αντί για την καλύτερη λύση. Αποδεικνύεται ότι η προτεινόμενη μέθοδος και πάλι υπερτερεί έναντι των άλλων τεχνικών στην πλειοψηφία των συνόλων δεδομένων, λαμβάνοντας υπόψη είτε την ακρίβεια%, είτε τον συντελεστή MCC, ακόμα και

όταν ως μέτρο σύγκρισης εφαρμόζεται το χαμηλότερο όριο του διαστήματος εμπιστοσύνης 95%. Η υπεροχή του αλγορίθμου ESA-NSFM είναι πιο προφανής, αν λάβουμε υπόψη ότι η λύση που βρέθηκε από τους άλλους ταξινομητές ήταν η ολικά βέλτιστη λύση που θα μπορούσαν να επιτύχουν, ως αποτέλεσμα της διαδικασίας διεξοδικής αναζήτησης που χρησιμοποιείται για την εκπαίδευσή τους. Ακόμα και αν συγκριθεί η μέση λύση ενός αριθμού πειραμάτων της προτεινόμενης μεθόδου, με την καλύτερη δυνατή λύση που βρέθηκε από τους άλλους ταξινομητές, όπου ο αλγόριθμος ESA-NSFM είναι σε μειονεκτική θέση, ο τελευταίος εξακολουθεί να παρέχει στατιστικά σημαντικά υψηλότερη ακρίβεια.

Όσον αφορά στην ταχύτητα, ο αλγόριθμος ESA-NSFM απαιτεί περισσότερη υπολογιστική ισχύ για ολοκληρωθεί σε σύγκριση με τη μέθοδο εκπαίδευσης SFM. Αυτό είναι αναμενόμενο λόγω της σημαντικής αύξησης του μεγέθους του χώρου αναζήτησης που πρέπει να διερευνήσει ο αλγόριθμος ESA-NSFM. Ωστόσο, είναι σημαντικό να σημειωθεί ότι στις περισσότερες περιπτώσεις, το μοντέλο RBF που παράγει ο αλγόριθμος ESA-NSFM ξεπερνά το τελικό αποτέλεσμα της μεθόδου SFM όσον αφορά στην ακρίβεια%, μέσα σε λίγες επαναλήψεις. Έτσι, η προτεινόμενη τεχνική παράγει ένα πολύ ακριβές μοντέλο μέσα στις πρώτες κιόλας επαναλήψεις, αλλά απαιτούνται περαιτέρω βελτιώσεις οι οποίες γίνονται μέχρι την ολοκλήρωση του αλγορίθμου. Από την άλλη πλευρά, η εκπαίδευση των μηχανών SVM αποδεικνύεται ότι είναι η πιο αργή μέθοδος, καθώς σε αυτή την περίπτωση η διεξοδική αναζήτηση εκτελείται παράλληλα για δύο παραμέτρους.

Ο συνολικός υπολογιστικός χρόνος που απαιτείται για όλους τους ταξινομητές επηρεάζεται έντονα από το μέγεθος του συνόλου δεδομένων, που προσδιορίζεται από τον αριθμό των δεδομένων εκπαίδευσης και των μεταβλητών εισόδου, αν και ο τελευταίος δεν φαίνεται να έχει σαφείς επιπτώσεις στις επιδόσεις της ακρίβειάς τους. Αντιθέτως, ο αριθμός των κατηγοριών εξόδου, επηρεάζει την απόδοση ταξινομητή. Πιο συγκεκριμένα, η ακρίβεια πρόβλεψης φαίνεται να μειώνεται, καθώς ο αριθμός των κατηγοριών εξόδου στο σύνολο δεδομένων αυξάνεται. Ένα καλό παράδειγμα, αποτελούν τα δύο σύνολα δεδομένων Cardiotocography, όπου υπάρχει σημαντική απόκλιση μεταξύ των επιδόσεων των ταξινομητών για την περίπτωση NSP που περιέχει 3 κλάσεις και για την περίπτωση FHR που αποτελείται από 10 κλάσεις. Αυτό μπορεί να αποδοθεί στο γεγονός ότι το έργο της κατάταξης γίνεται πιο δύσκολο, καθώς ο αριθμός των πιθανών κατηγοριών εξόδου αυξάνεται. Ωστόσο, είναι ενδιαφέρον να σημειωθεί, ότι ο αλγόριθμος ESA-NSFM φαίνεται να έχει ένα ελαφρώς μεγαλύτερο πλεονέκτημα σε σχέση με τον ταξινομητή SVM στα σύνολα δεδομένων με περισσότερες από δύο κατηγορίες εξόδου. Μια πιθανή εξήγηση είναι ότι οι μηχανές SVM είναι δυαδικοί ταξινομητές που πρέπει να τροποποιηθούν ανάλογα για να προβλέπουν περισσότερες από δύο κλάσεις εξόδου, σε αντίθεση με τον αλγόριθμο ESA-NSFM που μπορεί να χειριστεί εγγενώς τέτοιες περιπτώσεις.

Όπως συμβαίνει με όλες τις τεχνικές ταξινόμησης που βασίζονται σε δεδομένα, η παρουσία outliers στα δεδομένα εκπαίδευσης μπορεί να έχει αρνητικές συνέπειες στην απόδοση του προτεινόμενου ταξινομητή. Ως εκ τούτου, συνιστάται να εκτελείται προεπεξεργασία δεδομένων πριν από την εφαρμογή της μεθόδου στα αρχικά ιατρικά αρχεία (π.χ. βλέπε [230]). Όσον αφορά στις ελλιπείς τιμές, η τεχνική που υιοθετείται σε αυτή την εργασία, δηλαδή η διαγραφή των αντίστοιχων παραδειγμάτων, λειτουργεί με επιτυχία στα χρησιμοποιούμενα σύνολα δεδομένων. Ωστόσο, η διαγραφή παραδειγμάτων θα μπορούσε να δημιουργήσει προβλήματα σε μικρότερα σύνολα δεδομένων, καθώς τα δεδομένα που μένουν μπορεί να μην είναι αρκετά για την εκπαίδευση του δικτύου RBF. Σε αυτή την περίπτωση, ένας αριθμός από πιο εξειδικευμένες τεχνικές, οι οποίες έχουν αποδειχθεί ότι συνεργάζονται με τους ταξινομητές RBF [231], θα μπορούσαν να εφαρμοστούν σε συνδυασμό με τον αλγόριθμο ESA-NSFM για τον υπολογισμό των τιμών που λείπουν. Τέλος, θα πρέπει να

σημειωθεί ότι η αύξηση του αριθμού των παραδειγμάτων και το πρόβλημα της διαστατικότητας, οδηγούν σε μεγαλύτερους υπολογιστικούς χρόνους. Τελικά, η αύξηση της πολυπλοκότητας θα μπορούσε να οδηγήσει σε υπολογιστικούς χρόνους που δεν είναι πρακτικά εφικτοί. Ωστόσο, η απόδοση του προτεινόμενου αλγορίθμου σε σύνολα δεδομένων με σχετικά υψηλή πολυπλοκότητα και το πλεονέκτημά του να παράγει ικανοποιητικά μοντέλα μέσα στις πρώτες κιόλας επαναλήψεις του, δείχνουν ότι μπορεί να χειριστεί με επιτυχία τις τέτοιου είδους περιπτώσεις.

Πίνακας 3.13. Αποτελέσματα για όλα τα σύνολα δεδομένων – Ακρίβεια% και MCC

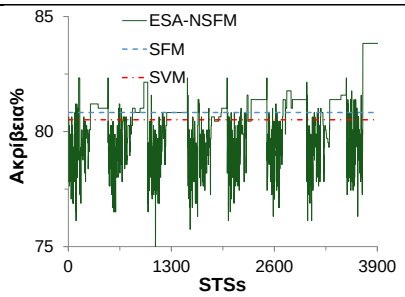
Σύνολο δεδομένων	Αλγόριθμος	Ακρίβεια% συνόλου αξιολόγησης	Ακρίβεια% συνόλου ελέγχου	MCC συνόλου αξιολόγησης	MCC συνόλου ελέγχου
Cardiotocography FHR	ESA-NSFM	**83.8** (83.2 [83.0-83.4])	**82.9** (82.5 [82.1-82.9])	**0.80** (0.79 [0.79-0.79])	**0.79** (0.79 [0.79-0.79])
	SFM	80.8	80.8	0.76	0.77
	SVM	80.5	80.7	0.74	0.75
Cardiotocography NSP	ESA-NSFM	**95.1** (94.9 [94.8-95.0])	93.2 (93.1 [92.8-93.4])	**0.86** (0.85 [0.85-0.85])	**0.82** (0.81 [0.81-0.82])
	SFM	94.0	**93.6**	0.83	0.80
	SVM	92.9	93.0	0.80	0.80
EEG Eye State	ESA-NSFM	**96.7** (96.4 [96.4-96.5])	**96.8** (96.7 [96.7-96.8])	**0.93** (0.93 [0.93-0.93])	**0.94** (0.93 [0.93-0.94])
	SFM	95.9	96.2	0.92	0.92
	SVM	96.3	96.6	**0.93**	0.93
Pima Indians Diabetes	ESA-NSFM	**82.8** (81.6 [81.0-82.1])	**81.8** (79.8 [78.8-80.8])	**0.75** (0.72 [0.71-0.73])	**0.76** (0.72 [0.70-0.74])
	SFM	80.2	78.7	0.68	0.67
	SVM	78.1	78.7	0.70	0.71
Thyroid ANN	ESA-NSFM	**95.5** (95.4 [95.4-95.5])	**95.7** (95.4 [95.4-95.5])	**0.81** (0.79 [0.79-0.79])	**0.79** (0.78 [0.78-0.79])
	SFM	94.8	95.1	0.75	0.74
	SVM	95.3	95.2	**0.81**	0.78
Vertebral Column – 2 Classes	ESA-NSFM	**93.6** (93.5 [93.5-93.6])	**88.3** (87.0 [86.5-87.5])	**0.83** (0.82 [0.81-0.83])	**0.80** (0.79 [0.78-0.80])
	SFM	91.0	84.4	0.76	0.71
	SVM	89.7	84.4	0.76	0.73
Vertebral Column – 3 Classes	ESA-NSFM	**94.9** (93.3 [92.8-93.8])	**88.3** (86.4 [85.9-86.9])	**0.92** (0.88 [0.87-0.89])	**0.89** (0.86 [0.85-0.87])
	SFM	91.0	80.5	0.83	0.77
	SVM	88.6	81.6	0.81	0.78
Wisconsin Diagnostic Breast Cancer (WDBC)	ESA-NSFM	99.3 (99.0 [98.8-99.2])	**96.5** (94.3 [93.1-95.5])	0.99M(0.98 [0.98-0.98])	**0.94** (0.91 [0.86-0.96])
	SFM	97.9	93.0	0.96	0.88
	SVM	**100.0**	96.5	**1**	0.93
Wisconsin Original Breast Cancer (WOBC)	ESA-NSFM	**99.4** (98.9 [98.8-99.0])	**99.4** (98.9 [98.8-99.0])	**0.99** (0.98 [0.98-0.98])	**0.99** (0.98 [0.98-0.99])
	SFM	97.7	98.2	0.95	0.96
	SVM	98.3	98.8	0.96	0.97

Απεικονίζεται η εκτέλεση που βρέθηκε το καλύτερο αποτέλεσμα σύμφωνα με τον δείκτη ακρίβεια% στο σύνολο αξιολόγησης για όλους τους ταξινομητές.
Στην περίπτωση του ταξινομητή ESA-NSFM, το καλύτερο αποτέλεσμα ακολουθείται από τον μέσο όρο και τα διαστήματα εμπιστοσύνης 95% μέσα σε παρένθεση. Το καλύτερο αποτέλεσμα παρουσιάζεται με έντονους χαρακτήρες.

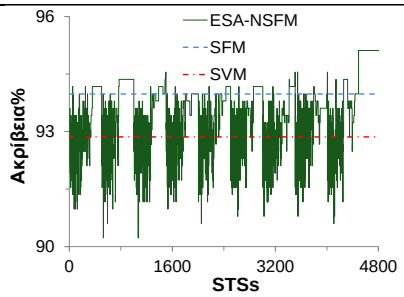
Πίνακας 3.14. Αποτελέσματα για όλα τα σύνολα δεδομένων – Ασαφής διαμέριση, βελτιστοποιημένες παράμετροι και υπολογιστικοί χρόνοι

Σύνολο δεδομένων	Αλγόριθμος	Ασαφής διαμέριση	Βελτιστοποιημένες παράμετροι[α]	Υπολογιστικοί χρόνοι
Cardiotocography FHR	ESA-NSFM	[23 6 19 4 4 44 44 4 4 45 4 5 22 15 25 4 23 9 36 10 4]	352 (453 [402-504])	603 (834 [739-929])
	SFM	10	219	69
	SVM	-	36.76 - 3.48	13535
Cardiotocography NSP	ESA-NSFM	[24 4 17 50 28 15 50 38 12 18 15 12 5 28 41 39 30 37 4 23 4]	825 (800 [755-845])	604 (811 [733-889])
	SFM	26	845	70
	SVM	-	21.11 - 2.55	5612
EEG Eye State	ESA-NSFM	[21 20 18 37 48 35 48 32 50 37 32 33 30 26]	2801 (2737 [2673-2801])	1968 (1832 [1800-1864])
	SFM	23	1867	181
	SVM	-	16.00 - 1.15	23412
Pima Indians Diabetes	ESA-NSFM	[4 4 9 4 6 4 36 4]	31 (45 [38-52])	102 (99 [93-105])
	SFM	7	35	8
	SVM	-	2.00 - 3.25	1383
Thyroid ANN	ESA-NSFM	[6 50 43 46 4 50]	183 (270 [252-288])	832 (819 [730-908])
	SFM	28	300	104
	SVM	-	48.50 - 2.30	20675
Vertebral Column – 2 Classes	ESA-NSFM	[9 31 4 28 4 4]	36 (29 [26-32])	14 (13 [11-15])
	SFM	9	25	1
	SVM	-	3.61 - 1.23	212
Vertebral Column – 3 Classes	ESA-NSFM	[4 14 4 4 5 50]	22 (23 [20-26])	11 (13 [12-14])
	SFM	13	45	2
	SVM	-	25.99 - 3.73	674
Wisconsin Diagnostic Breast Cancer (WDBC)	ESA-NSFM	[34 4 23 50 36 43 26 7 7 13 50 4 20 50 18 4 4 14 21 45 4 4 4 10 10 42 14 4 5 9]	149 (239 [215-263])	95 (88 [78-98])
	SFM	19	221	8
	SVM	-	0.38 - 4.93	912
Wisconsin Original Breast Cancer (WOBC)	ESA-NSFM	[7 4 35 4 50 28 8 8 50 4]	171 (198 [189-207])	42 (45 [44-46])
	SFM	17	203	7
	SVM	-	0.03 - 3.03	851

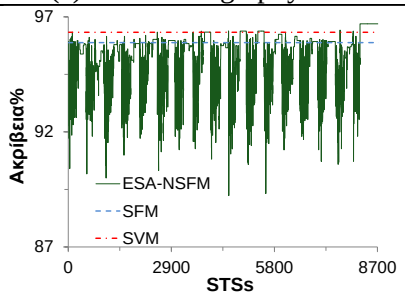
Απεικονίζεται η εκτέλεση που βρέθηκε το καλύτερο αποτέλεσμα σύμφωνα με τον δείκτη ακρίβεια% στο σύνολο αξιολόγησης για όλους τους ταξινομητές.
Στην περίπτωση του ταξινομητή ESA-NSFM, το καλύτερο αποτέλεσμα ακολουθείται από τον μέσο όρο και τα διαστήματα εμπιστοσύνης 95% μέσα σε παρένθεση.
[α]Για τους δύο ταξινομητές RBF, οι βελτιστοποιημένες παράμετροι που αφορούν στον αριθμό επιλεγμένων κόμβων. Για τους ταξινομητές SVM, οι βελτιστοποιημένες παράμετροι που αφορούν στον συντελεστή ποινής C και το εύρος των γκαουσιανών πυρήνων σ, δίνονται στη μορφή: C - σ



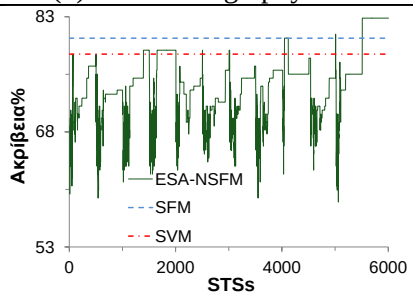
(a) Cardiotocography FHR



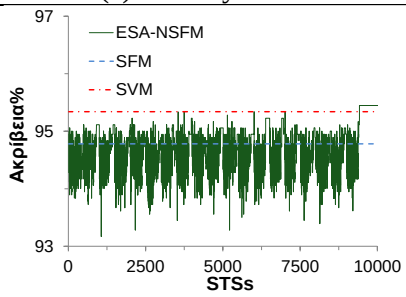
(b) Cardiotocography NSP



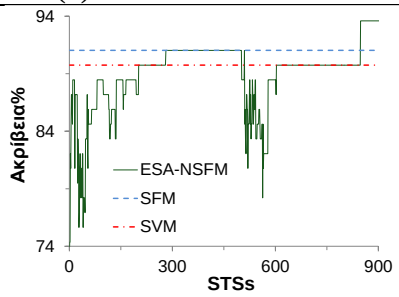
(c) EEG Eye State



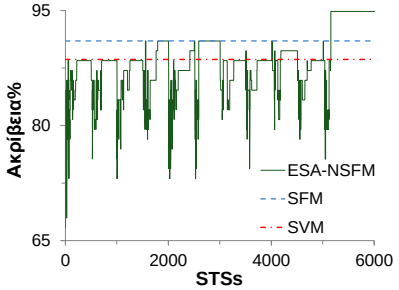
(d) Pima Indians Diabetes



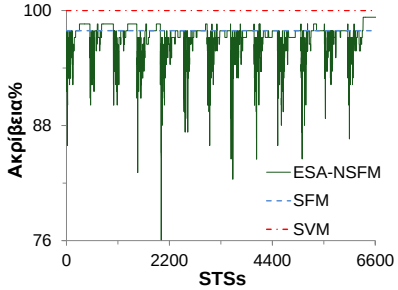
(e) Thyroid ANN



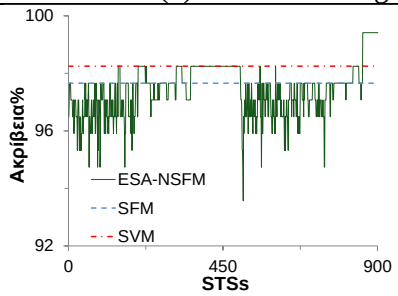
(f) Vertebral Column – 2 Classes



(g) Vertebral Column – 3 Classes



(h) Wisconsin Diagnostic Breast Cancer (WDBC)



(i) Wisconsin Original Breast Cancer (WOBC)

Σχήμα 3.11. Εξέλιξη του κριτηρίου ακρίβεια% ανά επανάληψη στο σύνολο αξιολόγησης για τις τρεις μεθοδολογίες. Για τον αλγόριθμο ESA-NSFM, απεικονίζεται η καλύτερη λύση που βρέθηκε σύμφωνα με τον δείκτη ακρίβεια% στο σύνολο αξιολόγησης.

ΚΕΦΑΛΑΙΟ 4

ΚΑΤΗΓΟΡΙΚΑ ΚΑΙ ΜΙΚΤΑ ΔΕΔΟΜΕΝΑ

Στο παρόν κεφάλαιο γίνεται μία σύντομη εισαγωγή στα κατηγορικά δεδομένα και στα μέτρα απόστασης που είναι κατάλληλα για κατηγορικές τιμές. Στόχος αυτού του κεφαλαίου είναι η παρουσίαση δύο νέων καινοτόμων αρχιτεκτονικών νευρωνικών δικτύων, που έχουν την ικανότητα να χειρίζονται κατηγορικά και μικτά δεδομένα ως εισόδους. Οι αρχιτεκτονικές αυτές, που ονομάστηκαν CRBF και MRBF, αντίστοιχα, βασίζονται στο γενικό πρότυπο των δικτύων RBF, αλλά έχουν τροποποιηθεί κατάλληλα ώστε να μπορούν να χειριστούν τις ιδιαιτερότητες που παρουσιάζουν αυτοί οι τύποι δεδομένων. Επίσης, παρουσιάζονται δύο νέοι αλγόριθμοι εκπαίδευσης με την ικανότητα να επιτυγχάνουν αποτελεσματική εκπαίδευση για τους αντίστοιχους τύπους αρχιτεκτονικών.

4.1 Εισαγωγή στα κατηγορικά δεδομένα

Τα κατηγορικά δεδομένα αναφέρονται σε τιμές ή παρατηρήσεις περιγραφικής φύσης, και σε αντίθεση με τα αριθμητικά δεδομένα, δεν είναι εγγενώς διατεταγμένα, αλλά ίσως να μπορούν να ταξινομηθούν σε ποιοτικές ή κατηγορικές ομάδες. Οι κατηγορικές μεταβλητές συναντώνται πολύ συχνά σε σύνολα δεδομένων, ανεξάρτητα από την προέλευσή τους, και έτσι έχουν μεγάλη σημασία στον τομέα της εξόρυξης δεδομένων και της ανακάλυψης γνώσης. Ωστόσο, η έλλειψη εγγενούς μαθηματικής σημασίας στα κατηγορικά δεδομένα καθιστά δύσκολη την εκτίμηση της απόστασης μεταξύ τους, γεγονός που εμποδίζει σημαντικές λειτουργίες που σχετίζονται με τη μηχανική μάθηση, όπως ταξινόμηση, προσέγγιση συνάρτησης, συσταδοποίηση, κ.λπ.

Ένας ευρέως γνωστός τρόπος για να δημιουργηθούν συστήματα μηχανικής μάθησης που να διαχειρίζονται κατηγορικά δεδομένα είναι να παρακαμφθεί η ανάγκη υπολογισμού των αποστάσεων, χρησιμοποιώντας ταξινομητές δέντρων απόφασης. Οι τελευταίοι βρίσκονται σε πλεονεκτική θέση, διότι μπορούν να χειριστούν κατηγορηματικά δεδομένα εκ φύσεως. Από την άλλη, οι τεχνικές που βασίζονται σε δέντρα αποφάσεων μπορεί να παρουσιάζουν μειονεκτήματα όσον αφορά σε κατηγορικά σύνολα δεδομένων που οι είσοδοί τους έχουν υψηλή συσχέτιση μεταξύ τους, καθώς επίσης και σε συνεχή ή μικτά σύνολα δεδομένων [98].

Μια απλοϊκή μέθοδος για να πραγματοποιηθεί η εκπαίδευση νευρωνικών δικτύων με βάση την απόσταση κατηγορικών δεδομένων, περιλαμβάνει την ανάθεση μιας αυθαίρετης αριθμητικής τιμής σε κάθε κατηγορία και στη συνέχεια τον υπολογισμό των αποστάσεων με βάση τις αριθμητικές αυτές τιμές, χρησιμοποιώντας την Ευκλείδεια απόσταση ή την απόσταση Manhattan. Αν και η παραπάνω τεχνική δίνει τη δυνατότητα εκμετάλλευσης ενός πλήθους μεθόδων από το οπλοστάσιο της μηχανικής μάθησης, η αυθαίρετη ανάθεση τιμών δεν έχει νόημα και επιπλέον, θα μπορούσαν να επιτευχθούν καλύτερα αποτελέσματα με τη χρήση πιο σύνθετων μεθοδολογιών. Εναλλακτικά, μια απόσταση, ή ισοδύναμα ένα μέτρο ομοιότητας, θα μπορούσε να οριστεί με βάση τις πληροφορίες που είναι κρυμμένες μέσα σε πολλές κατηγορικές τιμές, δηλαδή σε ένα κατηγορικό σύνολο δεδομένων. Ένας τρόπος εκτίμησης αυτού του είδους των πληροφοριών είναι να ληφθεί υπόψη η συχνότητα εμφάνισης των κατηγορικών τιμών στο σύνολο δεδομένων. Οι ανομοιότητες στις συχνά εμφανιζόμενες τιμές θα μπορούσαν να υποδεικνύουν μια μικρότερη απόσταση σε σύγκριση με τις ανομοιότητες στις σπάνιες τιμές, ή αντίστροφα, δημιουργώντας έτσι τα μέτρα απόστασης της συχνότητας εμφάνισης (occurrence frequency, OF) και της αντίστροφης συχνότητας εμφάνισης (inverse occurrence frequency, IOF) [232], αντίστοιχα. Άλλες παρόμοιες μέθοδοι που βασίζονται στη θεωρία της πληροφορίας περιλαμβάνουν τα μέτρα Lin, Burnaby και Anderberg [85].

Η αρχιτεκτονική των δικτύων ακτινικής συνάρτησης βάσης προσφέρει απλούς, αλλά ισχυρούς αλγορίθμους μάθησης, οι οποίοι βασίζονται στην απόσταση, και έχουν χρησιμοποιηθεί εκτενώς και με μεγάλη επιτυχία για αριθμητικά δεδομένα [29]. Δυστυχώς, τα δίκτυα RBF δεν είναι σε θέση να διαχειριστούν άμεσα κατηγορικά δεδομένα. Αυτό οφείλεται αφενός στην εγγενή ανάγκη των δικτύων αυτού του τύπου για τον υπολογισμό των αποστάσεων μεταξύ των δεδομένων εισόδου και των διανυσμάτων που είναι γνωστά ως κέντρα RBF και αφετέρου για την εκτέλεση συσταδοποίησης στο χώρο εισόδου, όπως ορίζει η διαδικασία εκπαίδευσης των δύο-σταδίων που συνήθως χρησιμοποιείται για την εκπαίδευση των δικτύων RBF.

Σε αυτή τη διδακτορική διατριβή, η αρχιτεκτονική των δικτύων RBF επεκτάθηκε ώστε να είναι εφικτό να χειριστεί κατηγορικά δεδομένα, εισάγοντας την έννοια των

κατηγορικών δικτύων RBF (categorical RBF, CRBF). Τα τελευταία χρησιμοποιούν ως κέντρα RBF κατηγορικές πλειάδες (tuples) αντί αριθμητικών διανυσμάτων και ενσωματώνουν τα εργαλεία για τον υπολογισμό των απαιτούμενων αποστάσεων μεταξύ αυτών και των πλειάδων εισόδου. Επιπλέον, παρουσιάζεται μία νέα μέθοδος για την εκπαίδευση των δικτύων CRBF. Η τεχνική αυτή βασίζεται σε ένα γρήγορο μη επαναληπτικό αλγόριθμο συσταδοποίησης που χρησιμοποιείται για την επιλογή των κέντρων πλειάδων RBF, η οποία ακολουθείται από τη διαδικασία υπολογισμού των συναπτικών βαρών χρησιμοποιώντας τη μέθοδο των ελαχίστων τετραγώνων.

4.2 Μέτρα απόστασης για κατηγορικά δεδομένα

Έστω ένα σύνολο N_c κατηγορικών μεταβλητών x_i, i=1,2,…, N_c, όπου κάθε μεταβλητή x_i μπορεί να λάβει n_i διαφορετικές κατηγορικές τιμές. Το πεδίο ορισμού της μεταβλητής x_i είναι το σύνολο $\mathbf{B}_i = \{x_i.c_1 \;\; x_i.c_2 \; \ldots \; x_i.c_{n_i}\}$, όπου με $x_i.c_j$, j=1,2,…, n_i δηλώνεται η j-οστή κατηγορική τιμή της μεταβλητής x_i. Η κατηγορική πλειάδα $\mathbf{x}$ ορίζεται ως ένα σύνολο μεγέθους N_c, που περιέχει ακριβώς μία κατηγορική τιμή από κάθε μεταβλητή. Σε αυτή την περίπτωση, το σύνολο $\mathbf{B}=\mathbf{B}_1\times\mathbf{B}_2\times\ldots\times\mathbf{B}_{N_c}$ περιέχει όλες τις πιθανές κατηγορικές πλειάδες.

Λόγω της έλλειψης ενός εγγενούς μέτρου απόστασης για κατηγορικά δεδομένα, θα πρέπει να σχεδιαστεί ένα κατάλληλο μέτρο $D(\mathbf{x}(1),\mathbf{x}(2))$ που να μπορεί να προσδιορίσει μια απόσταση μεταξύ δύο κατηγορικών πλειάδων $\mathbf{x}(1)$ και $\mathbf{x}(2)$. Μια απλή προσέγγιση είναι να χρησιμοποιηθεί η απόσταση Hamming, δηλαδή ο αριθμός των μεταβλητών που λαμβάνουν μια διαφορετική κατηγορική τιμή μεταξύ των δύο πλειάδων. Ο υπολογισμός της απόστασης Hamming είναι απλός και πολύ γρήγορος. Από την άλλη πλευρά, το μέτρο αυτό δεν μπορεί, πάντα να χαρακτηρίσει την πραγματική ανομοιότητα μεταξύ δύο πλειάδων, καθώς δίνει την ίδια σημασία σε όλες τις κατηγορικές τιμές που δεν είναι όμοιες, χωρίς να λαμβάνει υπόψη τη μεταβλητή ή την κατηγορία που ανήκουν οι συγκεκριμένες ανόμοιες τιμές. Περισσότερες πληροφορίες σχετικά με την ανομοιότητα δύο πλειάδων θα μπορούσαν να επιτευχθούν, μέσω της εκτίμησης συμπληρωματικών στοιχείων που προκύπτουν από έναν αριθμό επιλεγμένων πλειάδων, με τη βοήθεια τεχνικών που βασίζονται σε δεδομένα (data-driven). Για το σκοπό αυτό, έχουν προταθεί διάφορα μέτρα στη βιβλιογραφία [85], τα περισσότερα από τα οποία υπολογίζουν την ομοιότητα βάσει του περιεχομένου της πληροφορίας. Σε αυτή τη διδακτορική διατριβή χρησιμοποιήθηκαν έξι διαφορετικά μέτρα απόστασης, τα οποία περιγράφονται παρακάτω. Περισσότερες λεπτομέρειες σχετικά με τα μέτρα αυτά μπορούν να βρεθούν στην εργασία [85]. Επίσης, τα παρακάτω μέτρα χρησιμοποιούνται για τον υπολογισμό των αποστάσεων ανά μεταβλητή $d(x_i(1),x_i(2))$, ενώ η συνολική απόσταση μεταξύ δύο κατηγορικών πλειάδων $\mathbf{x}(1)$ και $\mathbf{x}(2)$ μπορεί να υπολογιστεί ως εξής:

$$D\left(\mathbf{x}(1),\mathbf{x}(2)\right)=\sum_{i=1}^{N_C} d\left(x_i(1),x_i(2)\right) \tag{4.1}$$

Πρέπει να σημειωθεί ότι τα μέτρα που παρουσιάζονται παρακάτω είναι κανονικοποιημένα σύμφωνα με τον συνολικό αριθμό των μεταβλητών N_c.

- **Hamming**: Η απόσταση Hamming υπολογίζει το πλήθος των χαρακτηριστικών που λαμβάνουν διαφορετική κατηγορική τιμή μεταξύ δύο πλειάδων.

$$d\left(x_i(1), x_i(2)\right) = \begin{cases} 0 & , \text{ αν } x_i(1) = x_i(2) \\ \dfrac{1}{N_c} & , \text{ αν } x_i(1) \neq x_i(2) \end{cases} \tag{4.2}$$

- **Inverse Occurrence Frequency (IOF)**: Το μέτρο IOF λαμβάνει υπόψιν τη συχνότητα εμφάνισης των τιμών των μεταβλητών [232]. Πιο συγκεκριμένα, η συχνότητα εμφάνισης $f(x_i.c_j)$ ορίζεται ως το πλήθος των περιπτώσεων που η μεταβλητή x_i εμφανίζει την τιμή $x_i.c_j$, σε ένα σύνολο δεδομένων K πλειάδων. Στις ανομοιότητες των πιο συχνά εμφανιζόμενων τιμών, ανατίθενται μεγαλύτερες αποστάσεις.

$$d\left(x_i(1), x_i(2)\right) = \begin{cases} 0 & , \text{ αν } x_i(1) = x_i(2) \\ \dfrac{\log f\left(x_i(1)\right) \cdot \log f\left(x_i(2)\right)}{N_c} & , \text{ αν } x_i(1) \neq x_i(2) \end{cases} \tag{4.3}$$

- **Lin**: Το μέτρο Lin [85] αρχικά προτάθηκε για την εκτίμηση της ομοιότητας κατηγορικών δεδομένων. Σε αυτή τη διατριβή το μέτρο Lin τροποποιήθηκε έτσι ώστε η απόσταση να υπολογίζεται άμεσα, αναθέτοντας μικρότερη απόσταση σε ομοιότητες συχνά εμφανιζόμενων τιμών και μεγαλύτερη απόσταση σε ανομοιότητες σπάνια εμφανιζόμενων τιμών. Επίσης, σε αυτό το μέτρο εμπεριέχεται ο υπολογισμός της πιθανότητας $p(x_i.c_j)$ που έχει μια μεταβλητή x_i να εμφανίσει την τιμή $x_i.c_j$.

$$d\left(x_i(1), x_i(2)\right) = \begin{cases} \dfrac{A_{Lin} - 2\log\left(p\left(x_i(1)\right)\right)}{2N_c \log\left(p\left(x_i(1)\right)\right)} & , \text{αν } x_i(1) = x_i(2) \\ \dfrac{A_{Lin} - 2\log\left(p\left(x_i(1)\right) + p\left(x_i(2)\right)\right)}{2N_c \log\left(p\left(x_i(1)\right) + p\left(x_i(2)\right)\right)} & , \text{αν } x_i(1) \neq x_i(2) \end{cases} \tag{4.4}$$

όπου

$$p\left(x_i.c_j\right) = \frac{f\left(x_i.c_j\right)}{K} \tag{4.5}$$

και

$$A_{Lin} = \sum_{h=1}^{N_c} \left[\log\left(p\left(x_h(1)\right)\right) + \log\left(p\left(x_h(2)\right)\right)\right] \tag{4.6}$$

- **Occurrence Frequency (OF)**: Αυτό το μέτρο παρουσιάζει την αντίθετη λειτουργία σε σύγκριση με το μέτρο IOF όσον αφορά στις ανομοιότητες, καθώς αναθέτει μικρότερη απόσταση στις ανομοιότητες των πιο συχνά εμφανιζόμενων τιμών [232].

$$d\left(x_i(1), x_i(2)\right) = \begin{cases} 0 & , \text{ αν } x_i(1) = x_i(2) \\ \dfrac{\log \dfrac{K}{f\left(x_i(1)\right)} \cdot \log \dfrac{K}{f\left(x_i(2)\right)}}{N_c} & , \text{ αν } x_i(1) \neq x_i(2) \end{cases} \tag{4.7}$$

- **Burnaby**: Η βασική ιδέα που κρύβεται πίσω από αυτό το μέτρο ομοιότητας με βάση τη θεωρία της πληροφορίας [85], είναι ότι οι τιμές που εμφανίζονται σπάνια περιέχουν περισσότερη πληροφορία, και ως εκ τούτου θα πρέπει να σταθμίζονται ανάλογα. Σε αυτή τη διατριβή το μέτρο Burnaby τροποποιήθηκε έτσι ώστε ο υπολογισμός της απόστασης να είναι άμεσος, αναθέτοντας μεγαλύτερη απόσταση σε ανομοιότητες σπάνια εμφανιζόμενων τιμών, και μικρότερη απόσταση σε ανομοιότητες συχνά εμφανιζόμενων τιμών.

$$d\left(x_i(1), x_i(2)\right) = \begin{cases} 1 & , \text{ αν } x_i(1) = x_i(2) \\ \dfrac{\log \dfrac{p\left(x_i(1)\right) p\left(x_i(2)\right)}{\left(1 - p\left(x_i(1)\right)\right)\left(1 - p\left(x_i(2)\right)\right)}}{N_c \sum_{j=1}^{n_i} 2 \log\left(1 - p\left(x_i.c_j\right)\right)} & , \text{ αν } x_i(1) \neq x_i(2) \end{cases} \tag{4.8}$$

- **Eskin**: Το μέτρο Eskin [85] χειρίζεται τις μεταβλητές που οι τιμές τους ταυτίζονται, όμοια με την απόσταση Hamming, αλλά σε περίπτωση που οι μεταβλητές δεν ταυτίζονται, λαμβάνει υπόψη τον αριθμό n_i των διαφορετικών κατηγορικών τιμών που δέχεται η μεταβλητή i, αναθέτοντας μικρότερη απόσταση στις ανομοιότητες των μεταβλητών με μεγάλο πλήθος n_i.

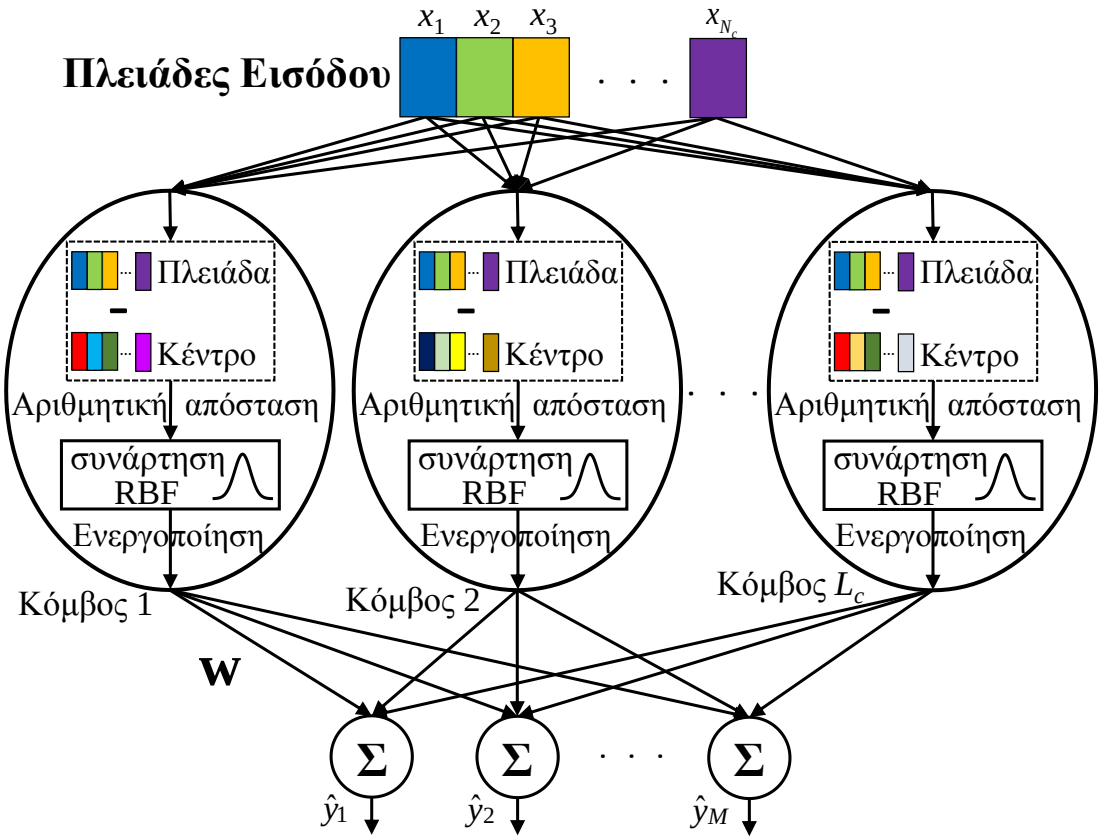
$$d\left(x_i(1), x_i(2)\right) = \begin{cases} 0 & , \text{ αν } x_i(1) = x_i(2) \\ \dfrac{2}{N_c n_i^2} & , \text{ αν } x_i(1) \neq x_i(2) \end{cases} \tag{4.9}$$

4.3 Νευρωνικά δίκτυα για τον χειρισμό κατηγορικών δεδομένων

Η βασική δομή των δικτύων RBF, είναι σε θέση να χειριστεί με μεγάλη επιτυχία αριθμητικά δεδομένα [29], ωστόσο εμφανίζει προβλήματα στον χειρισμό κατηγορικών δεδομένων. Τα δίκτυα RBF μπορούν να χειριστούν με επιτυχία αριθμητικές εισόδους προβάλλοντας τον χώρο εισόδου στον χώρο της κρυφής στοιβάδας, ο οποίος συνήθως είναι υψηλότερων διαστάσεων. Η προβολή αυτή βασίζεται σε ένα σύνολο διανυσμάτων που είναι γνωστά ως κέντρα RBF, όπου κάθε κέντρο αντιστοιχεί σε έναν διαφορετικό νευρώνα της κρυφής στοιβάδας. Τα διανύσματα εισόδου οδηγούνται προς τους κρυφούς νευρώνες, όπου υπολογίζεται η Ευκλείδεια απόσταση μεταξύ του διανύσματος εισόδου και του αντίστοιχου διανύσματος κέντρου. Αυτή η απόσταση δίνεται ως είσοδος σε μια συνάρτηση με ακτινική

συμμετρία, ώστε να παραχθεί η ενεργότητα του αντίστοιχου νευρώνα. Η τελική απόκριση του δικτύου RBF είναι το σταθμισμένο άθροισμα των εξόδων των νευρώνων της κρυφής στοιβάδας.

Η αρχιτεκτονική των νευρωνικών δικτύων CRBF είναι ικανή να χειρίζεται κατηγορικά δεδομένα. Η δομή ενός δικτύου CRBF που διαθέτει N_c κατηγορικές μεταβλητές εισόδου, L_c κρυφούς νευρώνες και M κατηγορίες εξόδου εμφανίζεται στο σχήμα 4.1. Τα δίκτυα CRBF λαμβάνουν ως εισόδους κατηγορικές πλειάδες. Η στοιβάδα εισόδου διανέμει τις πλειάδες εισόδου στην κρυφή στοιβάδα. Κάθε νευρώνας l της κρυφής στοιβάδας χαρακτηρίζεται από ένα κέντρο πλειάδας $\hat{\mathbf{x}}_l$. Τα κέντρα πλειάδων είναι πλειάδες μεγέθους N_c και περιέχουν κατηγορικές τιμές, οι οποίες προέρχονται από τις κατηγορίες που εμπεριέχονται στο κατηγορικό σύνολο δεδομένων **B**. Όπως και με τις πλειάδες εισόδου, κάθε κέντρο πλειάδας περιέχει ακριβώς μία κατηγορική τιμή από κάθε μία μεταβλητή. Λόγω αυτών των ιδιοτήτων, οποιαδήποτε από τα κατηγορικά μέτρα απόστασης που περιγράφηκαν παραπάνω μπορεί να χρησιμοποιηθεί προκειμένου να υπολογιστεί η απόσταση μεταξύ της εισόδου και του κέντρου πλειάδας. Σε κάθε κρυφό νευρώνα, λαμβάνει χώρα ένας υπολογισμός απόστασης, αποδίδοντας με κάποιο κατάλληλο μέτρο μια αριθμητική τιμή $D\left(\mathbf{x}(k),\hat{\mathbf{x}}_l\right)$ στην ανομοιότητα μεταξύ μιας κατηγορικής πλειάδας εισόδου $\mathbf{x}(k)$ και του αντίστοιχου κατηγορικού κέντρου $\hat{\mathbf{x}}_l$. Από αυτό το σημείο και μετά το δίκτυο CRBF δρα σαν ένα αριθμητικό δίκτυο RBF, υπολογίζοντας τις ενεργότητες των νευρώνων μέσω μιας συνάρτησης RBF. Σε αυτή την εργασία, χρησιμοποιήθηκε η συνάρτηση TPS, η οποία



Σχήμα 4.1. Δομή ενός δικτύου CRBF με κατηγορικές εισόδους.

περιγράφεται από την εξίσωση (2.3). Έτσι, για κάθε κατηγορική πλειάδα εισόδου $\mathbf{x}(k)$, κάθε κρυφός νευρώνας παράγει μια αριθμητική έξοδο $z_{c,l}$, η οποία δίνεται από την παρακάτω σχέση:

$$z_{c,l}(k) = g\left(D\left(\mathbf{x}(k), \hat{\mathbf{x}}_l\right)\right) = \left(D\left(\mathbf{x}(k), \hat{\mathbf{x}}_l\right)\right)^2 \cdot \log\left(D\left(\mathbf{x}(k), \hat{\mathbf{x}}_l\right)\right), \quad l = 1, ..., L_C \tag{4.10}$$

Η στοιβάδα εξόδου αποτελείται από M μονάδες. Μια αριθμητική έξοδος $\hat{y}_m(k)$ για κάθε μονάδα εξόδου παράγεται συνδυάζοντας γραμμικά τις αποκρίσεις των κρυφών νευρώνων:

$$\hat{y}_m(k) = \mathbf{z}_C(k) \cdot \mathbf{w}_m = \sum_{l=1}^{L_C} w_{l,m} z_{c,l}(k) \tag{4.11}$$

όπου $\mathbf{z}_C(k)$ είναι ένα διάνυσμα που περιέχει τις αριθμητικές εξόδους τη κρυφής στοιβάδας της κατηγορικής πλειάδας εισόδου $\mathbf{x}(k)$:

$$\begin{aligned} \mathbf{z}_C(k) &= \left[z_{c,1}(k), z_{c,2}(k), ..., z_{c,L_C}(k)\right] = \\ &= \left[g\left(D\left(\mathbf{x}(k), \hat{\mathbf{x}}_1\right)\right), g\left(D\left(\mathbf{x}(k), \hat{\mathbf{x}}_2\right)\right), ..., g\left(D\left(\mathbf{x}(k), \hat{\mathbf{x}}_{L_C}\right)\right)\right] \end{aligned} \tag{4.12}$$

και $\mathbf{w}_m = \left[w_{1,m},\ w_{2,m},\ \ldots,\ w_{L_C,m}\right]^T$ είναι ένα συναπτικό διάνυσμα βάρους που αντιστοιχεί στην έξοδο m. Η πρόβλεψη για την κατηγορία εξόδου $\mathscr{C}(k)$ αντιστοιχεί στη μονάδα άθροισης ενεργοποιώντας τη μέγιστη αριθμητική έξοδο:

$$\mathscr{C}(k) = \arg\max_m \hat{y}_m(k), \qquad m = 1, 2, ..., M \tag{4.13}$$

4.3.1 Αλγόριθμος εκπαίδευσης νευρωνικών δικτύων CRBF

Με βάση την προηγούμενη ανάλυση, η εκπαίδευση ενός δικτύου CRBF αφορά στην εύρεση αριθμητικών τιμών για τον αριθμό των κέντρων πλειάδων L_C και τα συναπτικά βάρη, και κατηγορικών τιμών για τα περιεχόμενα των κέντρων πλειάδων $\hat{\mathbf{x}}_l$, $l = 1, 2, \ldots L_C$. Σε αυτή την ενότητα παρουσιάζεται ένα απλό σύστημα κατάλληλο για την ολοκληρωμένη εκπαίδευση δικτύων CRBF, επιλέγοντας τις προαναφερθείσες παραμέτρους.

Έστω ότι είναι διαθέσιμο ένα σύνολο κατηγορικών δεδομένων εισόδου-εξόδου, το οποίο περιέχει K πλειάδες εισόδου μεγέθους N_C και μια κατηγορική μεταβλητή εξόδου C, η οποία περιέχει M κατηγορίες. Τα διαθέσιμα δεδομένα χωρίζονται με τυχαίο τρόπο σε τρία υποσύνολα: εκπαίδευσης, αξιολόγησης και ελέγχου, τα οποία περιέχουν K_{train}, K_{val} and K_{test} πλειάδες, αντίστοιχα.

Η προτεινόμενη μεθοδολογία χωρίζει τη διαδικασία εκπαίδευσης σε δύο στάδια: στο πρώτο στάδιο, ο αλγόριθμος επιλογής κατηγορικών κέντρων (categorical center selection,

CCS) χρησιμοποιείται για να επιλέξει τον αριθμό και τα περιεχόμενα των κέντρων των πλειάδων χρησιμοποιώντας μια μη-επαναληπτική διαδικασία, ενώ στο δεύτερο στάδιο προσδιορίζονται οι υπόλοιπες παράμετροι του δικτύου. Ο αλγόριθμος CCS ξεκινά ορίζοντας την πρώτη πλειάδα εισόδου εκπαίδευσης ως κέντρο πλειάδας. Για όλες τις υπόλοιπες πλειάδες εισόδου του υποσυνόλου εκπαίδευσης, ο αλγόριθμος υπολογίζει τις αποστάσεις από τις ήδη επιλεγμένες πλειάδες κέντρων και συγκρίνει την ελάχιστη απόσταση αυτών με ένα κατώφλι απόστασης δ. Αν η ελάχιστη απόσταση είναι μικρότερη από το κατώφλι δ, αυτό σημαίνει ότι η συγκεκριμένη πλειάδα εισόδου καλύπτεται επαρκώς από τουλάχιστον ένα υπάρχον κέντρο πλειάδας. Σε αντίθετη περίπτωση, οι νευρώνες της κρυφής στοιβάδας αυξάνονται κατά ένα, επιλέγοντας την πλειάδα εισόδου ως ένα νέο κέντρο πλειάδας. Οποιοδήποτε από τα μέτρα απόστασης που περιγράφηκαν στο κεφάλαιο 4.2 μπορεί να χρησιμοποιηθεί σε συνδυασμό με τον αλγόριθμο. Μετά την ολοκλήρωση ενός περάσματος των δεδομένων εκπαίδευσης, ο αλγόριθμος τερματίζει, έχοντας επιλέξει ένα υποσύνολο των πλειάδων εισόδου ως πλειάδες κέντρων. Αυτό το υποσύνολο εγγυάται: α) μια ελάχιστη απόσταση μεταξύ των επιλεγμένων κέντρων και β) ότι κάθε πλειάδα εισόδου καλύπτεται επαρκώς από τουλάχιστον ένα κέντρο πλειάδας, βάση του επιλεγμένου μέτρου απόστασης. Το τελευταίο βοηθά στο να αποφευχθούν ιδιάζοντα (ill-conditioned) προβλήματα, τα οποία θα μπορούσαν να προκληθούν από κέντρα που βρίσκονται πολύ κοντά το ένα στο άλλο, κατά τη διάρκεια του σταδίου υπολογισμού των συναπτικών βαρών. Παρακάτω παρουσιάζεται μια επισκόπηση του αλγορίθμου CCS.

Αλγόριθμος 6 – Αλγόριθμος CCS

Δεδομένα Εισόδου: $\mathbf{X}_{train} = \left[\mathbf{x}(1)\ \mathbf{x}(2)\ \ldots\ \mathbf{x}(K_{train})\right]^T$: Πλειάδες εισόδου εκπαίδευσης,
δ: Τιμή κατωφλίου

Δεδομένα Εξόδου: L_C: Αριθμός επιλεγμένων κέντρων πλειάδων,
$\hat{\mathbf{X}} = \left[\hat{\mathbf{x}}_1\ \hat{\mathbf{x}}_2\ \ldots\ \hat{\mathbf{x}}_{L_C}\right]^T$: Περιεχόμενα κέντρων πλειάδων

1. Ξεκίνα τους υπολογισμούς για το πρώτο κέντρο πλειάδας: $L_C \leftarrow 1$
2. Όρισε τη πρώτη πλειάδα εισόδου ως το πρώτο κέντρο πλειάδας: $\hat{\mathbf{x}}_L \leftarrow \mathbf{x}(1)$
3. **Για** $k = 2{:}K_{train}$ **Εκτέλεσε:**
4. **Για** $l = 1{:}\ L_C$ **Εκτέλεσε:**
5. Υπολόγισε την αριθμητική απόσταση $D_{k,l}\left(\mathbf{x}(k), \hat{\mathbf{x}}_l\right)$
6. **Τέλος Για**
7. **Αν** $\min\limits_{l=1,..,L_C}\left[D_{k,l}\left(\mathbf{x}(k), \hat{\mathbf{x}}_l\right)\right] > \delta$
8. Πρόσθεσε ένα νέο κέντρο πλειάδας: $L_C \leftarrow L_C + 1$
9. Τοποθέτησε το νέο κέντρο πλειάδας στη k-οστή πλειάδα εισόδου: $\hat{\mathbf{x}}_{L_C} \leftarrow \mathbf{x}(k)$
10. **Τέλος Αν**
11. **Τέλος Για**

Έχοντας δημιουργήσει την κρυφή στοιβάδα του δικτύου CRBF μέσω του αλγορίθμου CCS, στη συνέχεια υπολογίζονται τα συναπτικά βάρη $\mathbf{W}=\left[\mathbf{w}_1,\ \mathbf{w}_2,\ \ldots,\ \mathbf{w}_M\right]$ χρησιμοποιώντας τη μέθοδο γραμμικής παλινδρόμησης μεταξύ των αποκρίσεων των κρυφών νευρώνων $\mathbf{Z}=\left[\mathbf{z}_c\left(\mathbf{x}(1)\right),\ \mathbf{z}_c\left(\mathbf{x}(2)\right),\ \ldots,\mathbf{z}_c\left(\mathbf{x}\left(K_{train}\right)\right)\right]^T$ για όλες τις πλειάδες εκπαίδευσης, και των πραγματικών κατηγορικών εξόδων στο υποσύνολο της εκπαίδευσης, όπως περιγράφεται από την εξίσωση (2.14). Οι κατηγορίες εξόδου $\mathbf{Y}_{train}=\left[\mathbf{y}(1),\ \mathbf{y}(2),\ \ldots,\ \mathbf{y}\left(K_{train}\right)\right]^T$ του συνόλου εκπαίδευσης είναι σε αριθμητική μορφή υιοθετώντας την κωδικοποίηση “1 of C”.

4.3.2 Επιλογή μοντέλου CRBF

Το κατώφλι δ είναι μια σημαντική παράμετρος, δεδομένου ότι καθορίζει τον αριθμό των κατηγορικών κέντρων και άρα το μέγεθος του δικτύου. Η κατάλληλη επιλογή του κατωφλίου δ, και κατά συνέπεια του μοντέλου CRBF, μπορεί να επιτευχθεί χρησιμοποιώντας μια διαδικασία δοκιμής-σφάλματος στο υποσύνολο αξιολόγησης. Πιο συγκεκριμένα, εφαρμόζεται μια διαδικασία διεξοδικής αναζήτησης, δοκιμάζοντας τιμές του κατωφλίου δ μεταξύ ενός μέγιστου και ενός κατώτερου ορίου δ_{max} και δ_{min} με ένα σταθερό βήμα. Το μέγεθος του βήματος μπορεί στη συνέχεια να μειωθεί για να εκτελεστεί μία πιο λεπτομερής αναζήτηση μεταξύ των τιμών του κατωφλίου δ που παράγουν καλά αποτελέσματα. Τέλος, επιλέγεται το μοντέλο που αντιστοιχεί στην τιμή του κατωφλίου δ, η οποία δίνει το καλύτερο αποτέλεσμα στο υποσύνολο αξιολόγησης. Παρακάτω παρουσιάζεται μία επισκόπηση της προτεινόμενης διαδικασίας εκπαίδευσης του δικτύου CRBF.

Αλγόριθμος 7 – Διαδικασία εκπαίδευσης CRBF

Δεδομένα Εισόδου: $\left\{\mathbf{X}_{train},\mathbf{Y}_{train}\right\},\left\{\mathbf{X}_{val},\mathbf{Y}_{val}\right\}$: Υποσύνολα εκπαίδευσης-αξιολόγησης

Δεδομένα Εξόδου: $\hat{\mathbf{X}},\mathbf{W}$: Επιλεγμένες παράμετροι δικτύου CRBF

1. **Για** $\delta = \delta_{\min}$: $\delta_{\max}$ **Εκτέλεσε:**
2. Πέρνα τα δ και $\mathbf{X}_{train}$ στον αλγόριθμο 6, για να υπολογιστούν ο αριθμός των κέντρων των πλειάδων L_C και τα περιεχόμενα $\hat{\mathbf{X}}$
3. **Για** $k = 1{:}K_{train}$ **Εκτέλεσε:**
4. Υπολόγισε τις αριθμητικές αποκρίσεις $\mathbf{z}_c\left(\mathbf{x}(k)\right)$ των πλειάδων εισόδου εκπαίδευσης χρησιμοποιώντας τον τύπο (4.10)
5. **Τέλος Για**
6. Υπολόγισε τα συναπτικά βάρη χρησιμοποιώντας τον τύπο (2.14)
7. Υπολόγισε τις προβλέψεις του μοντέλο στο υποσύνολο αξιολόγησης χρησιμοποιώντας τους τύπους (4.11)-(4.13)
8. **Τέλος Για**
9. Επέλεξε το μοντέλο που παράγει τα καλύτερα αποτελέσματα στο υποσύνολο αξιολόγησης

4.3.3 Εφαρμογή δικτύων CRBF σε κατηγορικά σύνολα δεδομένων

Ο αλγόριθμος CCS εφαρμόστηκε σε 22 διαφορετικά σύνολα κατηγορικών δεδομένων τύπου benchmark. Τα σύνολα δεδομένων είναι διαθέσιμα στη βάση UCI Machine Learning Repository [190] και παρουσιάζονται στον πίνακα 4.1, όπου επίσης απεικονίζονται οι αριθμοί των μεταβλητών εισόδου, των κατηγοριών εξόδου και των παραδειγμάτων, ανά σύνολο δεδομένων. Τα σύνολα δεδομένων που χρησιμοποιούνται, περιλαμβάνουν εφαρμογές από διαφορετικούς επιστημονικούς τομείς, συμπεριλαμβανομένης της μοριακής βιολογίας, της ιατρικής, των κοινωνικών επιστημών, της θεωρίας παιγνίων, κ.λπ., ενώ προσφέρουν καλή ποικιλία όσον αφορά στο μέγεθος του συνόλου δεδομένων και τον αριθμό των κατηγοριών εξόδου. Μια σύντομη περιγραφή για κάθε σύνολο δεδομένων δίνεται παρακάτω, ενώ περισσότερες πληροφορίες μπορούν να βρεθούν στο [190].

- **Breast cancer, Lymphography, Primary tumor**: Αυτά τα τρία σύνολα δεδομένων προήλθαν από το Ινστιτούτο Ογκολογίας, που έχει επανειλημμένα εμφανιστεί στη βιβλιογραφία της μηχανικής μάθησης. Στα σύνολα δεδομένων Breast cancer, Lymphography και Primary tumor, ο στόχος είναι να προβλεφθεί η ύπαρξη υποτροπής ή όχι, να ταξινομηθεί το εύρημα σε 4 κλάσεις (φυσιολογικό, μετάσταση, κακοήθεις λεμφαδένες) και να βρεθεί σε ποιο από τα 18 πιθανά σημεία του σώματος ανήκει ο όγκος (πνευμόνια, στομάχι, πάγκρεας, ήπαρ, νεφρά, κτλ.), αντίστοιχα.
- **Car Evaluation**: Αυτό το σύνολο δεδομένων περιέχει χαρακτηριστικά αυτοκινήτων κατά την πώλησή τους, όπως, το μέγεθος χώρου αποσκευών, η τιμή, κτλ. Με βάση αυτά τα χαρακτηριστικά, θα πρέπει να κατηγοριοποιηθεί η ικανοποίηση των αγοραστών των αυτοκινήτων σε τέσσερις πιθανές κλάσεις (πολύ χαμηλή, χαμηλή, υψηλή, πολύ υψηλή).
- **Chess (**KR vs. KP**)**: Αυτό το σύνολό δεδομένων αφορά στο επιτραπέζιο παιχνίδι σκάκι. Ο στόχος εδώ είναι να προβλεφθεί ποιος από τους δύο μπορεί να νικήσει με βάση 36 μεταβλητές εισόδου που περιγράφουν τα τετράγωνα της σκακιέρας.
- **Clave-Direction**: Αυτό το σύνολό δεδομένων στη βάση UCI Machine Learning Repository αναφέρεται ως “Firm-Teacher Clave-Direction Classification Data Set” και περιέχει 16 δυαδικές εισόδους οι οποίες κατηγοριοποιούνται σε μία από τις τέσσερις μουσικές κλάσεις (neutral, reverse clave, forward clave, incoherent).
- **Congressional Voting**: Αυτό το σύνολο δεδομένων περιλαμβάνει στοιχεία για τις ψήφους των Αντιπροσώπων του Κογκρέσου της Αμερικανικής Βουλής το 1984. Σε αυτό το σύνολο δεδομένων, ο στόχος είναι η διάκριση των 16 βασικών ψήφων που προσδιορίζονται από τη CQA, στα κόμματα των Ρεμπουπλικάνων ή των Δημοκρατικών.
- **Connect-4**: Αυτή η βάση δεδομένων περιέχει όλες τις πιθανές θέσεις στο παιχνίδι connect-4. Ο στόχος εδώ είναι η πρόβλεψη μιας από τις τρεις κλάσεις που αναφέρονται στη νίκη, την ήττα και την ισοπαλία.
- **Dermatology**: Ο σκοπός αυτού του συνόλου δεδομένων είναι η πρόβλεψη του τύπου της ασθένειας που σχετίζεται με τις ερυθηματολεπιδώδεις πλάκες, με βάση κλινικά και ιστοπαθολογικά γνωρίσματα.
- **Hayes-Roth**: Αυτό το σύνολο δεδομένων περιγράφει ανθρώπινα κοινωνικά χαρακτηριστικά, τα οποία κατατάσσονται σε 3 κλάσεις.
- **HIV-1**: Αυτό το σύνολό δεδομένων στη βάση αποτελείται από αλληλουχίες αμινοξέων, οι οποίες διακρίνονται σε δύο κατηγορίες.
- **Mammographic Mass**: Ο στόχος σε αυτό το σύνολο δεδομένων είναι να προβλεφθεί αν ο όγκος είναι καλοήθης ή κακοήθης, με βάση τα χαρακτηριστικά από μία μαστογραφία και την ηλικία της ασθενούς.

- **MONK's Problems**: Αυτή η βάση περιέχει 3 τεχνητά σύνολα δεδομένων που προέρχονται από το χώρο της ρομποτικής. Σε αυτή την εφαρμογή χρησιμοποιήθηκε το σύνολο δεδομένων MONK's-1 και ο στόχος είναι η κατηγοριοποίηση των δεδομένων σε δύο κλάσεις.
- **Mushroom**: Αυτή η βάση δεδομένων περιέχει χαρακτηριστικά από 23 είδη μανιταριών, όπως η μυρωδιά, το χρώμα, κ.α. Με βάση αυτά τα χαρακτηριστικά, κάθε είδος μανιταριού πρέπει να ταξινομηθεί ως δηλητηριώδες ή φαγώσιμο.
- **Nursery**: Αυτό το σύνολο δεδομένων αναφέρεται σε ένα ιεραρχικό μοντέλο αποφάσεων που αναπτύχθηκε για την κατάταξη των αιτήσεων σε νηπιαγωγεία της Σλοβενίας το 1980, με βάση χαρακτηριστικά όπως οι συνθήκες στέγασης, η οικονομική και η κοινωνική κατάσταση της οικογένειας, κτλ.
- **Promoter**: Αυτό το σύνολό δεδομένων στη βάση UCI Machine Learning Repository αναφέρεται ως “Molecular Biology (Promoter Gene Sequences)” και έχει ως στόχο την κατηγοριοποίηση αλληλουχιών πρωτεϊνικών βάσεων γενετικού κώδικα DNA σε δύο κλάσεις.
- **SPECT Heart**: Αυτό το σύνολο δεδομένων περιέχει εικόνες από ένα σύστημα

Πίνακας 4.1. Επισκόπηση συνόλων δεδομένων τύπου benchmark

Σύνολο δεδομένων	**Συντομογραφία**	**Αριθμός μεταβλητών εισόδου**	**Αριθμός κατηγοριών εξόδου**	**Αριθμός δεδομένων**
Breast cancer	BRE	9	2	286
Car Evaluation	CAR	6	4	1728
Chess (KR vs. KP)	CHE	36	2	3196
Clave-Direction	CLA	16	4	10800
Congressional Voting	COG	16	2	232
Connect4	CON	42	3	67557
Dermatology	DER	34	6	358
Hayes-Roth	HAY	4	3	160
HIV-1	HIV	8	2	6590
Lymphography	LYM	18	4	148
Mammographic Mass	MAM	4	2	831
MONK's Problems	MON	6	2	556
Mushrooms	MUS	21	2	5644
Nursery	NUR	8	5	12960
Primary Tumor	PRI	17	18	132
Promoter	PRO	60	2	106
SPECT Heart	SPE	22	2	267
Splice	SPL	60	3	3190
Teaching Assistant	TEA	5	3	151
Tic-Tac-Toe	TIC	9	2	958
Turkiye Student	TUR	32	3	5820
Wisconsin (Original)	WIS	9	2	683

υπολογιστικής τομογραφίας εκπομπής μονού φωτονίου (Single Photon Emission Computed Tomography, SPECT). Με βάση αυτές τις εικόνες, κάθε ένας από τους ασθενείς ταξινομείται σε μία από τις δύο κατηγορίες: φυσιολογικό και μη φυσιολογικό.

- **Splice**: Αυτό το σύνολό δεδομένων στη βάση UCI Machine Learning Repository αναφέρεται ως "Molecular Biology (Splice-junction Gene Sequences)" και έχει ως στόχο την αναγνώριση προτύπων (δότης, δέκτης, κανένα από τα δύο) σε αλληλουχίες πρωτεϊνικών βάσεων γενετικού κώδικα DNA.
- **Teaching Assistant**: Αυτό το σύνολό δεδομένων περιέχει στοιχεία από αξιολογήσεις φοιτητών σχετικά με τη διδασκαλία στο Τμήμα Στατιστικής του Πανεπιστημίου Wisconsin-Madison. Ο στόχος είναι να προβλεφθεί η κλάση της βαθμολογίας (χαμηλή, μέση, υψηλή).
- **Tic-Tac-Toe**: Η βάση αυτή αφορά στο επιτραπέζιο παιχνίδι τρίλιζα, η οποία κωδικοποιεί το πλήρες σύνολο των πιθανών διατάξεων των πιονιών επάνω στον πίνακα της τρίλιζας. Δεδομένου ότι το "x" έχει παίξει πρώτο, ο στόχος εδώ είναι να προβλεφθεί αν το "x" μπορεί να δημιουργήσει τρίλιζα (3 πιόνια στη σειρά) με έναν από τους 8 πιθανούς τρόπους ή όχι.
- **Turkiye Student**: Αυτό το σύνολό δεδομένων περιέχει στοιχεία από αξιολογήσεις φοιτητών σχετικά με τη διδασκαλία μαθημάτων στο Πανεπιστήμιο Gazi στην Τουρκία. Ο στόχος είναι να προσδιοριστεί ποιος είναι ο καθηγητής, με βάση τις αξιολογήσεις.
- **Wisconsin (Original)**: Αυτό το σύνολο δεδομένων στη βάση UCI Machine Learning Repository αναφέρεται ως "Breast Cancer Wisconsin (Original)" και περιέχει στοιχεία από ασθενείς που είχαν υποβληθεί σε χειρουργική επέμβαση για καρκίνο του μαστού. Ο στόχος είναι να προβλεφθεί εάν ο ανιχνευόμενος όγκος είναι καλοήθης ή κακοήθης.

Για κάθε σύνολο δεδομένων, τα δεδομένα κατανεμήθηκαν με τυχαίο τρόπο σε τρία υποσύνολα εκπαίδευσης, αξιολόγησης και ελέγχου, χρησιμοποιώντας το 50% των δεδομένων για εκπαίδευση, το 25% για αξιολόγηση και το υπόλοιπο 25% για έλεγχο των νευρωνικών δικτύων. Ιδιαίτερη προσοχή δόθηκε στο να κατανεμηθούν τα δεδομένα που ανήκουν στις ίδιες κατηγορίες, ομοιόμορφα μεταξύ των τριών υποσυνόλων σύμφωνα με την παραπάνω αναλογία.

Η διαδικασία εκπαίδευσης των δικτύων CRBF δοκιμάστηκε χρησιμοποιώντας και τα 6 μέτρα απόστασης που παρουσιάστηκαν στο κεφάλαιο 4.2. Με στόχο να αξιολογηθεί ανεξάρτητα η απόδοση του προτεινόμενου αλγορίθμου εκπαίδευσης CCS, δοκιμάστηκαν επιπλέον δύο εναλλακτικές τεχνικές, η τυχαία επιλογή κέντρων (random center selection, RCS), και η "άπληστη" επιλογή (greedy center selection, GCS). Όσον αφορά στον αλγόριθμο RCS, τα κέντρα επιλέγονται τυχαία ανάμεσα στα δεδομένα εκπαίδευσης και ο αριθμός τους τίθεται ίσος με τον αριθμό των κέντρων που παράγουν το καλύτερο αποτέλεσμα με τον αλγόριθμο CCS. Ο αλγόριθμος GCS, είναι μια επαναληπτική άπληστη μέθοδος [233, 234], η οποία χτίζει σταδιακά την κρυφή στοιβάδα του δικτύου δοκιμάζοντας σε κάθε επανάληψη όλες τις υπόλοιπες πλειάδες εκπαίδευσης ως κέντρα. Στο τέλος κάθε επανάληψης προστίθεται ως επιπλέον κέντρο η πλειάδα, η οποία βελτιώνει περισσότερο την απόδοση του δικτύου. Σε περίπτωση που υπάρχουν περισσότερες από μία πλειάδες που αντιστοιχούν στο ίδιο καλύτερο μέτρο αξιολόγησης, η μέθοδος επιλέγει με τυχαίο τρόπο μία από αυτές τις πλειάδες και την προσθέτει ως κέντρο. Όταν η απόδοση του δικτύου σταματά να βελτιώνεται, τότε σταματά και η επαναληπτική διαδικασία του αλγορίθμου. Τα συναπτικά βάρη στους αλγορίθμους RCS και GCS υπολογίστηκαν χρησιμοποιώντας την ψευδο-αντίστροφη μέθοδο Moore-Penrose για να αποφευχθούν ιδιάζοντα προβλήματα που οφείλονται στην κακή

επιλογή των κέντρων, σε αντίθεση με τον αλγόριθμο CCS, ο οποίος είναι εγγενώς απρόσβλητος σε αυτού του είδους τα προβλήματα.

Η μέθοδος συγκρίθηκε επίσης με ένα ευρύ φάσμα διαφορετικών τεχνικών ταξινόμησης, οι οποίες περιλαμβάνουν:

- τα δίκτυα MLP, τα οποία εκπαιδεύονται με τον αλγόριθμο Levenberg-Marquardt [4], χρησιμοποιώντας την κωδικοποίηση "1 of C" για όλες τις κατηγορικές μεταβλητές
- οι μηχανές διανυσμάτων υποστήριξης αποφάσεων (SVMs) με γκαουσιανή συνάρτηση πυρήνα [228],
- οι ταξινομητές Naive Bayes (NB) [235],
- τα δέντρα ταξινόμησης και παλινδρόμησης (classification and regression trees, CART) [236]
- τους ταξινομητές τυχαίων δασών (random forest trees, RFTs) [237]

Επιπλέον, χρησιμοποιήθηκε ένα αριθμητικό δίκτυο RBF που εφαρμόζει την Ευκλείδεια απόσταση (euclidean RBF, ERBF). Το δίκτυο ERBF υιοθετεί μία παρόμοια τεχνική με αυτή των δικτύων CRBF, αλλά αντί να εφαρμόζει κάποια ειδική συνάρτηση για τον υπολογισμό των κατηγορικών αποστάσεων, απονέμει σε κάθε κατηγορία μια αυθαίρετη αριθμητική τιμή και στη συνέχεια χρησιμοποιεί την Ευκλείδεια απόσταση.

Όλοι οι ταξινομητές εκτελέστηκαν 30 φορές και βελτιστοποιήθηκαν ξεχωριστά για κάθε σύνολο δεδομένων. Όσον αφορά στα δίκτυα MLP, χρησιμοποιήθηκε μία αρχιτεκτονική δύο κρυφών στοιβάδων όπου εκτελέστηκε μια διαδικασία διεξοδικής αναζήτησης, δοκιμάζοντας όλους τους πιθανούς συνδυασμούς ανάμεσα στον αριθμό των νευρώνων της πρώτης κρυφής στοιβάδας που κυμαίνεται από 4 έως 40 και της δεύτερης κρυφής στοιβάδας που κυμαίνεται από 3 έως 30. Επίσης, οι παράμετροι των ταξινομητών SVM βελτιστοποιήθηκαν με βάση μια διαδικασία διεξοδικής αναζήτησης πλέγματος, παρόμοια με αυτή που χρησιμοποιήθηκε στα δίκτυα MLP, βελτιστοποιώντας έτσι το εύρος των γκαουσιανών σ και τον συντελεστή ποινής C. Τέλος, οι ταξινομητές που βασίζονται σε δέντρα βελτιστοποιήθηκαν με βάση τον αριθμό των δέντρων [238].

Τέλος, η απόδοση των ταξινομητών χαρακτηρίστηκε χρησιμοποιώντας δύο κριτήρια σφάλματος: τον στατιστικό δείκτη ακρίβεια% (ACC%), ο οποίος λαμβάνει υπόψη μόνο το συνολικό ποσοστό επιτυχίας του ταξινομητή, και τον συντελεστή MCC ο οποίος χρησιμοποιεί όλο τον πίνακα σύγχυσης συμπεριλαμβανομένων των επιτυχιών και των αποτυχιών ανά επιμέρους κατηγορία.

4.3.4 Αποτελέσματα και συζήτηση

Ο πίνακας 4.2 παρουσιάζει τα αποτελέσματα των δικτύων CRBF, συμπεριλαμβανομένου του δείκτη σφάλματος MCC στο σύνολο ελέγχου και του αριθμού των επιλεγμένων κέντρων πλειάδων RBF. Επίσης, αναφέρονται τα αποτελέσματα των τεχνικών RCS και GCS, χρησιμοποιώντας το πιο επιτυχημένο μέτρο απόστασης. Στην ανάλυση που ακολουθεί όσον αφορά στον πίνακα 4.2, για να συγκριθούν τα αποτελέσματα των μεθόδων χρησιμοποιείται ο δείκτης MCC αντί της ακρίβειας%, διότι σε κάποια σύνολα δεδομένων οι κλάσεις δεν είναι ισορροπημένες. Οι δύο δείκτες σφάλματος αξιολογούνται στο υποσύνολο ελέγχου, καθώς είναι εντελώς ανεξάρτητο και δεν χρησιμοποιείται ούτε στη διαδικασία της εκπαίδευσης του δικτύου, αλλά ούτε και στη διαδικασία της επιλογής του μοντέλου. Έτσι, μπορούν να αξιολογηθούν οι δυνατότητες γενίκευσης του κάθε μοντέλου.

Σύμφωνα με τα αποτελέσματα του πίνακα 4.2, φαίνεται ότι το μέτρο IOF αποδίδει καλύτερα σε σχέση με τα άλλα μέτρα, και ακολουθούν τα μέτρα OF και Eskin. Ωστόσο, όλα τα μέτρα συνεργάζονται καλά με τα δίκτυα CRBF και παράγουν ικανοποιητικά μοντέλα στα περισσότερα σύνολα δεδομένων. Επίσης, η σύγκριση του προτεινόμενου αλγορίθμου CCS με τους αλγορίθμους RCS και GCS αναδεικνύει την επιτυχία του πρώτου. Η μέθοδος με την τυχαία επιλογή κέντρων RCS, είναι λιγότερο ακριβής στα 19 από τα 22 σύνολα δεδομένων και στο ίδιο επίπεδο με την προτεινόμενη μέθοδο στα υπόλοιπα 3 σύνολα. Επιπλέον, είναι λιγότερο εύρωστη, δηλαδή παρουσιάζει υψηλότερες τιμές τυπικής απόκλισης. Η άπληστη μέθοδος GCS, σε ορισμένα σύνολα δεδομένων παράγει μοντέλα με μειωμένη πολυπλοκότητα. Ωστόσο, ο αλγόριθμος GCS παρουσιάζει δύο σημαντικά μειονεκτήματα: 1) τα αποτελέσματά του είναι σε κάθε περίπτωση χειρότερα από αυτά της τεχνικής CCS και 2) είναι πολύ πιο αργός, ειδικά στα μεγάλα σύνολα δεδομένων, λόγω της επαναληπτικής φύσης της μεθόδου GCS.

Στον πίνακα 4.3 παρουσιάζεται μια σύγκριση με διαφορετικές τεχνικές ταξινόμησης, περιλαμβάνοντας τους δείκτες σφάλματος MCC και ακρίβεια% (ACC%) στα υποσύνολα ελέγχου για όλες τις μεθόδους. Ο πίνακας 4.3 εμπεριέχει τα αποτελέσματα μόνο του πιο επιτυχημένου μέτρου απόστασης, δηλαδή του μέτρου IOF, σε συνδυασμό με την αρχιτεκτονική CRBF. Επίσης, στον πίνακα 4.3 παρουσιάζεται μια σύγκριση με βάση τον στατιστικό έλεγχο υποθέσεων. Συγκεκριμένα, αναφέρεται η τιμή της παραμέτρου *p* (*p*-value) της μεθόδου *t*-test μεταξύ των δικτύων CRBF και κάθε μιας από τις ανταγωνιστικές μεθοδολογίες. Η μηδενική υπόθεση (null hypothesis) είναι ότι τα αποτελέσματα των δύο ανταγωνιστικών μεθοδολογιών παράγονται από πληθυσμούς που έχουν την ίδια μέση τιμή. Μια άμεση σύγκριση μεταξύ της μεθόδου CRBF και την τεχνικής ERBF καταδεικνύει την αξία των κατηγορικών συναρτήσεων απόστασης που εφαρμόστηκαν στην πρώτη μέθοδο αντί των αριθμητικών τιμών που ορίστηκαν αυθαίρετα και των Ευκλείδειων αποστάσεων που χρησιμοποιήθηκαν, καθώς η μέθοδος CRBF παράγει σημαντικά καλύτερα αποτελέσματα σχεδόν σε όλα τα σύνολα δεδομένων. Η σύγκριση με τους ταξινομητές MLP και SVM, ευνοεί επίσης την τεχνική CRBF, η οποία ξεπερνά και τις δύο τεχνικές στα 19 από τα 22 σύνολα δεδομένων. Όσον αφορά στη σύγκριση με ταξινομητές NB, αλλά και με τις δύο τεχνικές που βασίζονται σε δέντρα απόφασης (CART και RFT), τα οποία έχουν σχεδιαστεί εγγενώς για να χειρίζονται κατηγορικά δεδομένα, φαίνεται ότι το προτεινόμενο σχήμα υπερτερεί όλων αυτών των μεθόδων στα 15 από τα 22 σύνολα δεδομένων, ενώ στα 6 από τα 7 υπόλοιπα σύνολα δεδομένων, το σχήμα CRBF αποδίδει εξίσου καλά με το καλύτερο μεταξύ αυτών. Είναι σημαντικό να σημειωθεί ότι στην πλειοψηφία των συνόλων δεδομένων, η ανωτερότητα της τεχνικής CRBF έναντι των άλλων μεθοδολογιών είναι στατιστικά σημαντική με επίπεδο εμπιστοσύνης 99%, ή μεγαλύτερο.

Όσον αφορά στους χρόνους εκπαίδευσης, τα πειραματικά αποτελέσματα έδειξαν ότι το προτεινόμενο σχήμα είναι:

- καλύτερο από τις τεχνικές MLP και SVM,
- στο ίδιο επίπεδο ή ελαφρώς χειρότερο από τις δύο μεθοδολογίες που βασίζονται στα δέντρα απόφασης,
- χειρότερο από τον ταξινομητή NB, ο οποίος ήταν ο ταχύτερος σε σχέση με τις μεθόδους που δοκιμάστηκαν,
- τάξεις μεγέθους πιο γρήγορο σε σύγκριση με τον αλγόριθμο GCS ή κλασικές επαναληπτικές μεθόδους όπως ο αλγόριθμος *k*-means, λόγω της μη επαναληπτικής φύσης του αλγορίθμου CCS που υιοθετεί το προτεινόμενο σχήμα.

Πίνακας 4.2. Αποτελέσματα των δικτύων CRBF με τα έξι μέτρα απόστασης και των δύο μεθόδων επιλογής κέντρων (στο υποσύνολο ελέγχου)

Μέθοδος		**Hamming - CCS**		**Eskin - CCS**		**OF - CCS**		**Lin - CCS**		**Burnaby - CCS**		**IOF - CCS**		**IOF - RCS**	**IOF - GCS**	
Παράμετρος		MCC ($x10^{-2}$)	Αριθμός κέντρων *L*	MCC ($x10^{-2}$)	Αριθμός κέντρων *L*	MCC ($x10^{-2}$)	Αριθμός κέντρων *L*	MCC ($x10^{-2}$)	Αριθμός κέντρων *L*	MCC ($x10^{-2}$)	Αριθμός κέντρων *L*	MCC ($x10^{-2}$)	Αριθμός κέντρων *L*	MCC ($x10^{-2}$)	MCC ($x10^{-2}$)	Αριθμός κέντρων *L*
Σύνολο Δεδομένων	BRE	34±9	12±1	31±5	4±0	30±11	21±2	33±3	21±5	32±8	9±2	**38**±4	14±1	31±9	32±9	22±11
	CAR	88±1	224±5	89±1	288±6	87±1	224±5	89±1	288±5	81±2	92±3	**92**±0	865±0	**92**±0	78±6	89±40
	CHE	95±1	901±7	94±1	833±7	94±1	734±8	90±2	1518±0	97±1	869±8	**98**±1	674±11	96±2	87±3	90±11
	CLA	69±1	98±3	70±1	1023±10	70±1	330±12	66±2	387±14	70±1	330±12	**72**±0	509±7	71±1	65±1	25±11
	COG	**100**±0	28±1	**100**±0	28±1	**100**±1	32±1	75±13	89±0	98±3	28±1	**100**±0	28±1	97±4	91±4	17±14
	CON	62±0	2714±16	62±0	2713±17	**65**±0	4397±27	56±7	2562±35	62±0	5297±11	64±0	6006±26	64±1	35±2*	17±0*
	DER	96±2	72±3	95±2	80±3	76±3	54±6	74±22	28±8	93±2	57±5	**97**±2	64±2	94±3	94±2	13±1
	HAY	77±6	22±1	74±7	22±1	71±5	28±1	73±7	29±2	68±10	17±3	**78**±6	27±2	64±7	57±14	21±13
	HIV	75±1	333±7	75±1	333±6	73±1	490±11	77±1	1551±16	78±1	1733±12	**79**±1	1112±14	76±2	69±1	330±13
	LYM	60±8	22±1	45±7	22±2	34±14	27±1	23±14	39±0	53±5	50±6	**65**±5	26±2	58±8	55±11	26±14
	MAM	57±1	25±1	**58**±2	25±1	51±3	33±2	52±1	10±3	53±4	11±2	57±1	14±2	56±2	56±2	18±18
	MON	**100**±0	248±0	**100**±0	248±0	**100**±0	156±3	87±5	267±0	**100**±1	62±3	**100**±1	71±3	98±2	91±13	39±10
	MUS	94±1	201±4	94±1	909±6	93±1	419±9	57±12	530±22	67±8	71±21	**100**±1	486±8	86±3	87±2	16±9
	NUR	93±0	244±4	**98**±0	2519±0	**98**±0	4856±0	96±0	1985±14	93±0	515±8	**98**±0	2519±0	97±1	91±1	148±36
	PRI	52±5	22±1	53±5	38±1	45±7	20±2	32±7	50±0	48±4	16±1	**57**±5	32±2	44±7	33±10	8±3
	PRO	42±19	5±1	61±15	5±1	57±12	8±1	69±10	19±2	78±11	16±2	**84**±7	23±1	47±12	62±18	14±7
	SPE	17±6	67±2	**35**±6	14±2	33±12	2±0	26±12	114±0	34±9	9±1	34±7	14±2	33±9	30±6	14±14
	SPL	94±0	1500±1	93±1	1093±7	91±1	853±7	**95**±0	1459±1	73±7	5±1	**95**±1	1083±8	94±2	81±4	64±10
	TEA	27±7	36±1	20±2	2±0	42±5	49±0	26±6	1±0	34±8	41±2	**46**±3	53±1	36±7	16±12	6±3
	TIC	53±5	93±3	65±3	479±0	94±2	273±3	97±1	461±1	97±1	110±3	**98**±1	452±1	97±2	35±7	61±21
	TUR	96±0	762±6	83±1	874±7	**97**±0	565±10	96±0	959±6	**97**±0	197±21	**97**±0	990±7	**97**±0	96±0	52±17
	WIS	94±1	19±2	94±1	27±2	88±5	17±21	96±2	20±2	92±4	48±34	**98**±1	15±2	95±2	97±1	18±12

Απεικονίζεται ο μέσος όρος των 30 εκτελέσεων και ακολουθείται από την τυπική απόκλισή τους. Το καλύτερο αποτέλεσμα παρουσιάζεται με έντονους χαρακτήρες.

* Για το σύνολο δεδομένων CON, οι 30 εκτελέσεις της μεθόδου IOF-GCS τερματίστηκαν έπειτα από 24 ώρες σε έναν υπολογιστή με επεξεργαστή Intel Core i7-5820K και μνήμη 16GBs.

Πίνακας 4.3. Σύγκριση με διάφορες τεχνικές ταξινόμησης (με βάση το υποσύνολο ελέγχου)

Μέθοδος		**CRBF (CCS-IOF)**		**ERBF**			**MLP**			**SVM**			**NB**			**CART**			**RFT**		
Παράμετρος		MCC ($x10^{-2}$)	ACC %	MCC ($x10^{-2}$)	ACC %	*p*-value	MCC ($x10^{-2}$)	ACC %	*p*-value	MCC ($x10^{-2}$)	ACC %	*p*-value	MCC ($x10^{-2}$)	ACC %	*p*-value	MCC ($x10^{-2}$)	ACC %	*p*-value	MCC ($x10^{-2}$)	ACC %	*p*-value
Σύνολο Δεδομένων	BRE	**38**±4	75±2	21±6	71±2	0.00	21±7	70±2	0.00	36±0	76±0	0.01	31±0	72±0	0.00	22±6	73±1	0.00	24±7	70±3	0.00
	CAR	**92**±0	97±0	76±4	91±24	0.00	91±2	96±1	0.23	90±0	95±0	0.00	70±0	87±0	0.00	**92**±1	97±1	0.39	**92**±1	96±1	0.22
	CHE	**98**±1	99±0	63±45	81±23	0.00	96±1	98±0	0.00	96±0	98±0	0.00	72±0	86±0	0.00	**98**±0	99±0	0.25	96±0	98±0	0.01
	CLA	72±0	82±0	70±1	81±1	0.00	**75**±1	84±1	0.00	**75**±0	83±0	0.00	66±0	78±0	0.00	61±0	75±0	0.00	70±0	81±0	0.00
	COG	**100**±0	100±0	**100**±2	100±1	0.33	98±2	99±1	0.00	97±0	98±0	0.00	93±0	97±0	0.00	**100**±0	100±0	0.28	99±2	99±1	0.00
	CON	**64**±0	83±0	39±11	69±16	0.00	57±1	80±1	0.00	52±0	76±0	0.00	37±0	73±0	0.00	53±1	78±0	0.00	55±0	80±0	0.00
	DER	**97**±2	97±1	95±1	96±1	0.01	94±2	95±2	0.00	94±0	95±0	0.00	94±0	95±0	0.00	91±2	92±2	0.00	95±1	96±1	0.00
	HAY	78±6	86±4	63±5	76±3	0.00	**80**±6	87±3	0.12	73±0	79±0	0.00	72±0	82±0	0.00	64±0	77±0	0.00	73±3	82±2	0.00
	HIV	**79**±1	93±0	42±2	83±1	0.00	78±1	93±0	0.01	47±0	82±0	0.00	71±0	91±0	0.00	70±1	90±0	0.00	67±1	90±0	0.00
	LYM	**65**±5	82±3	37±12	68±6	0.00	62±8	80±4	0.07	57±0	78±0	0.00	57±0	78±0	0.00	64±3	82±1	0.42	54±3	77±2	0.00
	MAM	**57**±1	78±1	56±2	78±1	0.02	52±2	76±1	0.00	45±0	72±0	0.00	56±0	78±0	0.01	56±1	78±1	0.04	56±0	78±0	0.00
	MON	**100**±1	100±1	74±6	87±3	0.00	99±2	100±1	0.48	88±0	93±0	0.00	55±0	73±0	0.00	85±7	92±4	0.00	96±2	98±1	0.00
	MUS	**100**±1	100±1	79±23	89±13	0.00	86±3	93±1	0.00	98±0	99±0	0.03	87±0	94±0	0.00	97±3	99±1	0.04	**100**±0	100±0	0.00
	NUR	**98**±0	99±0	57±42	65±36	0.00	**98**±1	99±0	0.31	94±0	96±0	0.00	86±0	90±0	0.00	**98**±0	99±0	0.52	95±0	96±0	0.00
	PRI	**57**±5	63±4	50±4	56±3	0.00	44±8	50±7	0.00	26±0	32±0	0.00	56±0	61±0	0.09	50±1	54±0	0.00	53±5	59±4	0.00
	PRO	84±7	92±4	58±16	79±8	0.00	60±12	79±6	0.00	79±0	88±0	0.00	**93**±0	96±0	0.00	36±8	68±4	0.00	75±7	87±3	0.00
	SPE	**34**±7	82±2	30±6	79±2	0.03	26±7	77±3	0.00	30±0	77±0	0.00	28±0	76±0	0.00	26±3	80±1	0.00	31±6	80±2	0.03
	SPL	**95**±1	97±0	70±2	82±1	0.00	92±1	95±1	0.00	75±0	85±0	0.00	93±0	96±0	0.00	91±0	94±0	0.00	92±1	95±0	0.00
	TEA	**46**±3	62±2	27±8	51±5	0.00	43±8	62±5	0.14	27±0	51±0	0.00	11±0	41±0	0.00	31±9	54±6	0.00	36±1	57±1	0.00
	TIC	**98**±1	99±0	87±26	92±17	0.02	86±4	94±2	0.00	95±0	98±0	0.00	23±0	67±0	0.00	80±1	91±0	0.00	80±1	91±1	0.00
	TUR	**97**±0	98±0	43±15	67±14	0.00	96±0	98±0	0.00	42±0	71±0	0.00	55±0	74±0	0.00	**97**±0	98±0	0.00	88±1	93±0	0.00
	WIS	**98**±1	99±0	92±2	97±1	0.00	95±2	98±1	0.00	94±0	97±0	0.00	96±0	98±0	0.00	93±1	97±0	0.00	97±1	99±0	0.01

Απεικονίζεται ο μέσος όρος των 30 εκτελέσεων και ακολουθείται από την τυπική απόκλισή τους. Το καλύτερο αποτέλεσμα, σύμφωνα με το MCC, παρουσιάζεται με έντονους χαρακτήρες.

4.4 Νευρωνικά δίκτυα για τον χειρισμό μικτών δεδομένων

Στον πραγματικό κόσμο, τα σύνολα δεδομένων συνήθως αποτελούνται από μικτά δεδομένα, τα οποία είναι ετερογενή δεδομένα που περιέχουν ταυτόχρονα αριθμητικές και κατηγορικές τιμές, δηλαδή ποσοτικές και ποιοτικές τιμές, αντίστοιχα. Η συντριπτική πλειοψηφία των μεθόδων που υπάρχουν στη βιβλιογραφία χειρίζονται ομοιογενή δεδομένα, δηλαδή δεδομένα του ίδιου τύπου [239]. Οι περισσότερες μέθοδοι που εφαρμόζονται σε μικτά δεδομένα, είναι τεχνικές που χειρίζονται ομοιογενή δεδομένα με αποτέλεσμα να αντιμετωπίζουν όλα τα δεδομένα ως κατηγορικά ή αριθμητικά, μέσω κάποιου είδους μετατροπής [240]. Ωστόσο, κατά τη διαδικασία μετατροπής των χαρακτηριστικών από ποσοτικά σε ποιοτικά δεδομένα, ή το αντίθετο, η σημασία τους έμμεσα μεταβάλλεται. Η μοντελοποίηση μικτών δεδομένων είναι ένα δύσκολο έργο και αποτελεί πρόκληση για τον κλάδο της μηχανικής μάθησης και ειδικότερα των νευρωνικών δικτύων.

Μια σημαντική καινοτομία αυτής της διδακτορικής διατριβής αφορά στη δημιουργία μιας νέας αρχιτεκτονικής νευρωνικών δικτύων, η οποία βασίζεται στα δίκτυα RBF και CRBF και έχει την ικανότητα να λαμβάνει υπόψη την ιδιαίτερη φύση των αριθμητικών και κατηγορικών μεταβλητών, ξεχωριστά. Επίσης, στα πλαίσια της διαδικασίας εκπαίδευσης αυτής της νέας αρχιτεκτονικής νευρωνικών δικτύων, αναπτύχθηκε ένας νέος αλγόριθμος, ο οποίος ουσιαστικά αποτελεί συνδυασμό των αλγορίθμων των ασαφών μέσων και CCS.

Τα νευρωνικά δίκτυα RBF και CRBF συνδυάστηκαν, με σκοπό τον χειρισμό μικτών δεδομένων. Η νέα αρχιτεκτονική νευρωνικών δικτύων που προέκυψε, περιλαμβάνει τρεις στοιβάδες νευρώνων, και ονομάστηκε μικτά δίκτυα RBF (mixed RBF, MRBF). Η δομή ενός δικτύου MRBF, που διαθέτει N αριθμητικές μεταβλητές εισόδου, N_c κατηγορικές μεταβλητές εισόδου, L αριθμητικούς κρυφούς νευρώνες, L_c κατηγορικούς κρυφούς νευρώνες και M εξόδους εμφανίζεται στο σχήμα 4.2.

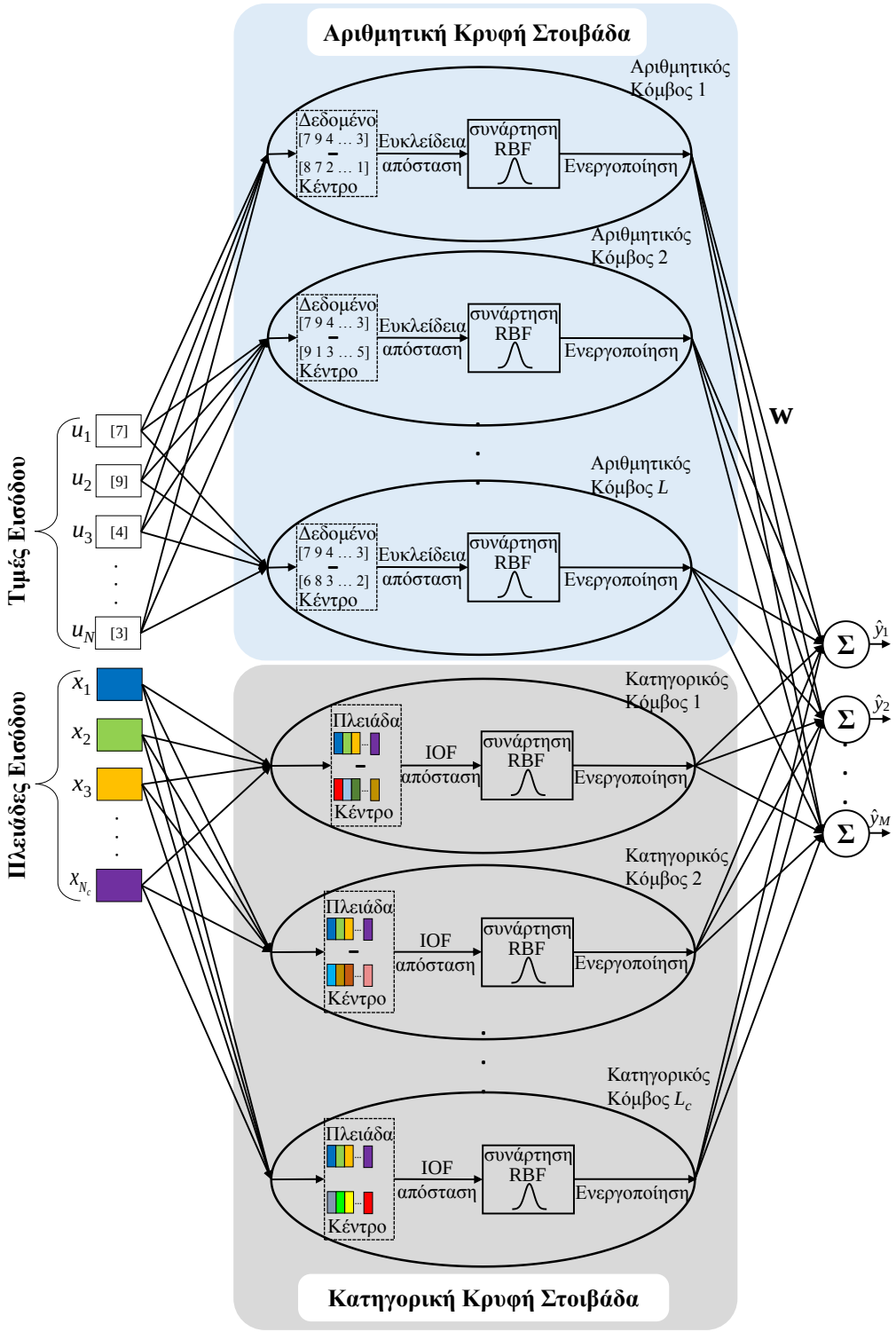
Η πρώτη στοιβάδα ενός δικτύου MRBF είναι η στοιβάδα εισόδου που αποτελείται από δύο είδη νευρώνων. Πιο συγκεκριμένα, η στοιβάδα εισόδου αποτελείται από N αριθμητικούς και N_c κατηγορικούς νευρώνες, οι οποίοι λαμβάνουν ως εισόδους αριθμητικές τιμές και κατηγορικές πλειάδες, αντίστοιχα, και τις διανέμουν στην κρυφή στοιβάδα.

Η κρυφή στοιβάδα αποτελείται επίσης, από δύο τύπους νευρώνων, τους αριθμητικούς και τους κατηγορικούς, όπου οι αριθμητικοί κρυφοί νευρώνες συνδέονται με τους αριθμητικούς νευρώνες εισόδου, και αντίστοιχα, οι κατηγορικοί με τους κατηγορικούς. Οι κρυφοί νευρώνες ανάλογα με τον τύπο τους, εκτελούν ένα μη γραμμικό μετασχηματισμό από τον χώρο εισόδου σε έναν νέο χώρο υψηλότερων διαστάσεων. Ο κάθε κρυφός νευρώνας αντιστοιχεί σε ένα κέντρο, όπου οι αριθμητικοί κρυφοί νευρώνες αντιστοιχούν σε αριθμητικά κέντρα και οι κατηγορικοί κρυφοί νευρώνες αντιστοιχούν σε κατηγορικά κέντρα.

Κάθε αριθμητικός νευρώνας l της κρυφής στοιβάδας χαρακτηρίζεται από ένα αριθμητικό κέντρο $\hat{\mathbf{u}}_l$ μεγέθους N. Μέσα σε κάθε αριθμητικό κρυφό νευρώνα, υπολογίζεται η ευκλείδεια απόσταση $\mu_l\left(\mathbf{u}(k)\right)$, με k=1,2,…,K, όπου K είναι τα συνολικά διαθέσιμα δεδομένα. Ο υπολογισμός της ευκλείδειας απόστασης αντιστοιχεί στη διαφορά μεταξύ του διανύσματος εισόδου $\mathbf{u}(k)$ και του αντίστοιχου αριθμητικού κέντρου $\hat{\mathbf{u}}_l$, και περιγράφεται από την εξίσωση (2.1). Στη συνέχεια, η ενεργότητα κάθε αριθμητικού νευρώνα l περνά από τη συνάρτηση ενεργοποίησης TPS (εξίσωση (2.4)), με αποτέλεσμα οι αποκρίσεις του αριθμητικού μέρους της κρυφής στοιβάδας να περιγράφονται από την εξίσωση (2.6). Έτσι,

κατασκευάζεται το αριθμητικό κομμάτι της κρυφής στοιβάδας, που απαρτίζεται από L αριθμητικούς νευρώνες.

Αντίστοιχα, κάθε κατηγορικός νευρώνας l του κατηγορικού μέρους της κρυφής στοιβάδας χαρακτηρίζεται από ένα κατηγορικό κέντρο πλειάδας $\hat{\mathbf{x}}_l$. Τα κέντρα πλειάδων



Σχήμα 4.2. Δομή ενός δικτύου MRBF για την ταυτόχρονη διαχείριση αριθμητικών και κατηγορικών μεταβλητών.

είναι πλειάδες μεγέθους N_c και περιέχουν κατηγορικές τιμές, οι οποίες προέρχονται από τις κατηγορίες που εμπεριέχονται στο κατηγορικό σύνολο δεδομένων **B**. Σε κάθε κατηγορικό κρυφό νευρώνα, λαμβάνει χώρα ένας υπολογισμός απόστασης, αποδίδοντας μια αριθμητική τιμή $D\left(\mathbf{x}(k),\hat{\mathbf{x}}_l\right)$ στην ανομοιότητα μεταξύ μιας κατηγορικής πλειάδας εισόδου $\mathbf{x}(k)$ και του αντίστοιχου κατηγορικού κέντρου $\hat{\mathbf{x}}_l$, χρησιμοποιώντας κάποιο κατάλληλο μέτρο. Το μέτρο που υιοθετήθηκε από αυτή την τεχνική είναι το μέτρο IOF εξίσωση (4.3), λόγω της καλής απόδοσης που παρουσιάζει σε συνδυασμό με νευρωνικά δίκτυα [101]. Στη συνέχεια, η ενεργότητα κάθε κατηγορικού νευρώνα l περνά από τη συνάρτηση ενεργοποίησης TPS και παράγει μια αριθμητική έξοδο $z_{c,l}$ (εξίσωση (4.10)) με αποτέλεσμα οι αποκρίσεις του κατηγορικού μέρους της κρυφής στοιβάδας να περιγράφονται από την εξίσωση (4.12).

Έτσι, συνολικά, οι αποκρίσεις της κρυφής στοιβάδας, δηλαδή των αριθμητικών και των κατηγορικών κέντρων, εκφράζονται από την παρακάτω σχέση:

$$\hat{\mathbf{Z}}(k)=\left[\mathbf{z}(k),\mathbf{z}_c(k)\right]=\left[\underbrace{z_1(k),z_2(k),\ldots,z_L(k)}_{\text{αριθμητικά κέντρα}},\underbrace{z_{c,1}(k),z_{c,2}(k),\ldots,z_{c,L_C}(k)}_{\text{κατηγορικά κέντρα}}\right] \tag{4.14}$$

Στη συνέχεια, οι αποκρίσεις της κρυφής στοιβάδας $\hat{\mathbf{Z}}(k)$ περνούν μέσα από μία διαδικασία μείωσης διαστατικότητας, με σκοπό τη μείωση του αριθμού των ενεργών κέντρων, καθώς είναι πιθανό να υπάρχουν γραμμικές συσχετίσεις μεταξύ των αριθμητικών και κατηγορικών κέντρων. Για το λόγο αυτό, είναι αναγκαία μία τέτοια διαδικασία, της οποίας η εφαρμογή στις αποκρίσεις της κρυφής στοιβάδας, αναγκάζει τον αριθμό των νέων μειωμένων αποκρίσεων της κρυφής στοιβάδας $\mathbf{Z}(k)$ να είναι μικρότερος, ή το πολύ ίσος, με τον αριθμό των δεδομένων εκπαίδευσης. Έπειτα, αυτές οι νέες αποκρίσεις της κρυφής στοιβάδας πολλαπλασιάζονται με τα αντίστοιχα συναπτικα βάρη $\mathbf{W}$ και καταλήγουν στη στοιβάδα εξόδου.

Η στοιβάδα εξόδου αποτελείται από έναν ή περισσότερους νευρώνες του ίδιου τύπου και συγκεκριμένα, περιλαμβάνει M κόμβους άθροισης, οι οποίοι αθροίζουν τα σταθμισμένα σήματα της κρυφής στοιβάδας. Στην περίπτωση που το δίκτυο προορίζεται για προβλήματα προσέγγισης συνάρτησης, η στοιβάδα εξόδου αποτελείται μόνο από έναν νευρώνα, οπότε το M ισούται με 1. Αντιθέτως, όταν το δίκτυο προορίζεται για προβλήματα ταξινόμησης, οι νευρώνες της στοιβάδας εξόδου είναι τόσοι όσες και οι κλάσεις εξόδου του προβλήματος. Έτσι, η προκύπτουσα αριθμητική έξοδος $\hat{y}_m(k)$ περιγράφεται από την παρακάτω εξίσωση και η διαδικασία πρόβλεψης της κατηγορίας εξόδου $\mathscr{C}(k)$ είναι η ίδια με αυτήν των δικτύων CRBF (κεφάλαιο 4.3, εξίσωση (4.13)):

$$\begin{aligned}&\hat{y}_m(k)=\mathbf{Z}(k)\cdot\mathbf{w}_m=\\&\sum_{l=1}^{L+L_C}w_{l,m}Z_l(k)=\sum_{l=1}^{L}w_{l,m}g\left(\mu_l\left(\mathbf{u}(k)\right)\right)+\sum_{l=1}^{L_C}w_{L+l,m}g\left(D\left(\mathbf{x}(k),\hat{\mathbf{x}}_l\right)\right)\end{aligned} \tag{4.15}$$

όπου $w_{l,m}$ είναι το συναπτικό βάρος που σταθμίζει το σήμα του l κρυφού νευρώνα και εισέρχεται στον m νευρώνα της στοιβάδας εξόδου, και m=1,2,…,M.

4.4.1 Αλγόριθμος εκπαίδευσης νευρωνικών δικτύων MRBF

Η εκπαίδευση των νευρωνικών δικτύων MRBF πραγματοποιείται με μία νέα τεχνική, η οποία ουσιαστικά αποτελεί συνδυασμό της μεθόδου των ασαφών μέσων και του αλγορίθμου CCS, και παρουσιάζεται στη συνέχεια.

Έστω ότι είναι διαθέσιμο ένα σύνολο K σταθμισμένων μικτών δεδομένων εισόδου-εξόδου, το οποίο περιέχει N αριθμητικές και N_c κατηγορικές εισόδους της μορφής $\mathbf{u}(k)=\left[u_1(k),u_2(k),\ldots,u_N(k)\right]^T$ και $\mathbf{x}(k)=\left[x_1(k),\mathrm{x}_2(k),\ldots,\mathrm{x}_{N_c}(k)\right]^T$, αντίστοιχα, όπου k=1,2,…,K, και M μεταβλητές εξόδου. Τα διαθέσιμα δεδομένα χωρίζονται με τυχαίο τρόπο σε τρία υποσύνολα: εκπαίδευσης, αξιολόγησης και ελέγχου, τα οποία περιέχουν K_{train}, K_{val} και K_{test} δεδομένα, αντίστοιχα.

Η εκπαίδευση ενός δικτύου MRBF χωρίζεται σε δύο στάδια: στο πρώτο στάδιο, όπου επιλέγονται τα αριθμητικά και τα κατηγορικά κέντρα ξεχωριστά, και στο δεύτερο στάδιο, όπου μειώνονται τα ενεργά κέντρα και προσδιορίζονται τα συναπτικά βάρη. Το πρώτο στάδιο εκπαίδευσης πραγματοποιείται συνδυάζοντας τους αλγορίθμους των ασαφών μέσων (FM ή SFM) [34] και CCS [101]. Συγκεκριμένα, για την επιλογή του αριθμού και των τοποθεσιών των αριθμητικών κέντρων χρησιμοποιείται ο αλγόριθμος FM (βλ. κεφάλαιο 2.2.2) και για τον προσδιορισμό του αριθμού και των περιεχομένων των κέντρων των πλειάδων, δηλαδή των κατηγορικών κέντρων, χρησιμοποιείται ο αλγόριθμος CCS (βλ. κεφάλαιο 4.3.1). Οι δύο αυτοί αλγόριθμοι διακρίνονται για την ταχύτητά τους και την ταυτόχρονη επιλογή της δομής και των παραμέτρων του δικτύου, καθώς χρησιμοποιούν μόνο ένα πέρασμα των δεδομένων.

Μετά την επιλογή του συνολικού αριθμού των αριθμητικών και κατηγορικών κέντρων και τον υπολογισμό των αποκρίσεων των αριθμητικών και κατηγορικών κρυφών νευρώνων, όπου συνολικά περιγράφονται από την εξίσωση (4.14), ακολουθεί το δεύτερο στάδιο εκπαίδευσης του δικτύου MRBF. Στο δεύτερο στάδιο, αρχικά, χρησιμοποιείται μία τεχνική μείωσης διαστατικότητας για τη μείωση του αριθμού των ενεργών κέντρων. Αυτό κρίνεται αναγκαίο, καθώς είναι πιθανόν να υπάρχουν γραμμικές συσχετίσεις μεταξύ των αριθμητικών και κατηγορικών μεταβλητών. Οι συσχετίσεις αυτές δεν μπορούν να γίνουν αντιληπτές κατά την επιλογή των κέντρων, καθώς αυτή γίνεται με δύο διαφορετικούς αλγόριθμους, ανάλογα με το είδος των μεταβλητών. Εδώ, πρέπει να σημειωθεί ότι οι αλγόριθμοι αυτοί, από μόνοι τους, μπορούν να επιλέγουν τα κατάλληλα κέντρα ώστε να μην υπάρχουν γραμμικές συσχετίσεις μεταξύ τους, αλλά όταν συνδυάζονται δεν ισχύει αυτό. Επίσης, οι αλγόριθμοι αυτοί θέτουν τους αριθμούς των επιλεγμένων αριθμητικών και κατηγορικών κέντρων να είναι μικρότεροι ή ίσοι με τον αριθμό των δεδομένων εκπαίδευσης, πράγμα το οποίο σημαίνει ότι όταν τα κέντρα αυτά συνδυάζονται, μπορεί το πλήθος τους να είναι, μεγαλύτερο από τον αριθμό των δεδομένων εκπαίδευσης.

Οι τεχνικές μείωσης διαστατικότητας έχουν σα στόχο να μειώσουν τις διαστάσεις των διαθέσιμων στοιχείων, προβάλλοντας το σύνολο των διανυσμάτων υψηλής διάστασης σε ένα νέο χώρο χαμηλότερης διάστασης, όπου οι νέες μεταβλητές που δημιουργούνται παρέχουν ισχυρή πληροφόρηση σχετική με τα χαρακτηριστικά των αρχικών μεταβλητών. Συγκεκριμένα, η τεχνική μείωσης διαστατικότητας που χρησιμοποιείται είναι η μέθοδος ανάλυσης κύριων συνιστωσών (Principal Component Analysis, PCA) [111].

Η μέθοδος PCA αποτελεί μια από τις απλούστερες και τις πιο διαδεδομένες μεθόδους ανάλυσης δεδομένων. Αυτή η μέθοδος στοχεύει στην ανεύρεση ενός μικρού πλήθους νέων μεταβλητών, οι οποίες είναι το αποτέλεσμα ενός γραμμικού συνδυασμού προερχόμενου από τις αρχικές μεταβλητές. Οι νέες μεταβλητές που ονομάζονται κύριες συνιστώσες (Principal Components, PCs) δημιουργούνται με περιστροφή των αρχικών αξόνων με τέτοιο τρόπο

ώστε η πρώτη κύρια συνιστώσα να ελαχιστοποιεί το άθροισμα των τετραγωνισμένων ορθογώνιων αποκλίσεων ανάμεσα σε αυτό και σε όλες τις μεταβλητές [241]. Ο συνολικός αριθμός των κύριων συνιστωσών είναι ίσος με τον αριθμό των αρχικών μεταβλητών που δεν συσχετίζονται μεταξύ τους γραμμικά.

Με μαθηματικούς όρους, η μέθοδος PCA είναι μια μέθοδος για την ανάλυση ενός πίνακα **Z** διάστασης ($K \times L_{all}$) και βαθμού (rank) r σε άθροισμα r πινάκων βαθμού 1 και ένα υπολειμματικό πίνακα **E**:

$$\mathbf{Z} = \mathbf{Q}_1 + \mathbf{Q}_2 + \ldots + \mathbf{Q}_r + \mathbf{E} \tag{4.16}$$

Όλοι οι πίνακες βαθμού 1 που προκύπτουν μπορούν να γραφτούν ως εξωτερικά γινόμενα μεταξύ δύο διανυσμάτων, του διανύσματος "score" $\mathbf{t}_j$, j=1,2,…,r διάστασης (K×1) και του διανύσματος "loading" $\mathbf{p}_j$, j=1,2,…,r διάστασης ($K \times L_{all}$):

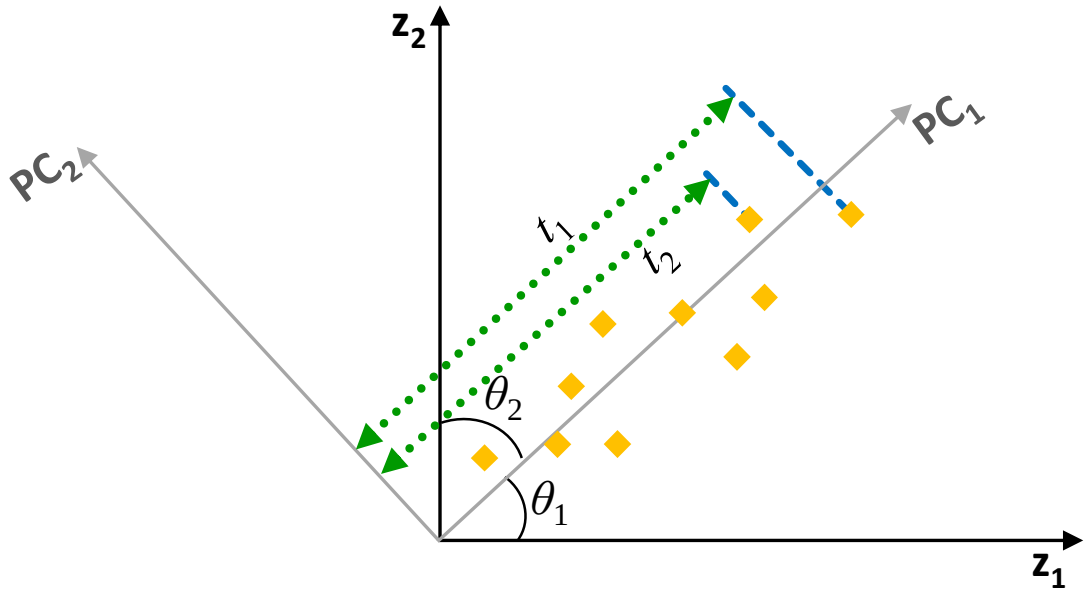
$$\mathbf{Z} = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \ldots + \mathbf{t}_r\mathbf{p}_r^T + \mathbf{E} \tag{4.17}$$

Τα διανύσματα loading $\mathbf{p}_j$, με j=1,2,…,r, είναι τα ιδιοδιανύσματα που σχετίζονται με την j-οστή ιδιοτιμή του πίνακα συνδιακύμανσης (covariance matrix) $\mathbf{Z}^T\mathbf{Z}$, όπου r είναι ο βαθμός του πίνακα **Z**. Ο βαθμός r ισούται με τον αριθμό των μη μηδενικών ιδιοτιμών του πίνακα συνδιακύμανσης, δηλαδή ισούται με τον αριθμό των μεταβλητών που δεν συσχετίζονται γραμμικά μεταξύ τους. Έτσι, η πληροφορία που περιέχεται στον πίνακα **Z**, μπορεί να αναπαραχθεί χρησιμοποιώντας μόνο τα διανύσματα score και loading των οποίων οι ιδιοτιμές είναι μη μηδενικές. Ως εκ τούτου, όταν όλες οι ιδιοτιμές του πίνακα συνδιακύμανσης είναι διαφορετικές του μηδενός, τότε ο βαθμός r ισούται με τον αριθμό των μεταβλητών L_{all} και άρα είναι απαραίτητα όλα τα διανύσματα score και loading για να αναπαραχθεί η πληροφορία που περιέχεται στον πίνακα **Z**.

Βέβαια, στην περίπτωση που τα δεδομένα περιέχουν θόρυβο, καμία από τις ιδιοτιμές δεν θα είναι ίση με μηδέν, διότι καμία μεταβλητή δεν θα είναι τελείως γραμμικά εξαρτημένη από κάποια άλλη. Ωστόσο, οι κύριες συνιστώσες που αντιστοιχούν σε μικρές ιδιοτιμές, μπορούν να θεωρηθούν ότι περιέχουν μόνο θόρυβο και όχι κάποια χρήσιμη πληροφορία και άρα είναι δυνατό να απορριφθούν και έτσι να μειωθεί η διάσταση του προβλήματος.

Η γεωμετρική ερμηνεία των διανυσμάτων "score" $\mathbf{t}_j$ και "loading" $\mathbf{p}_j$, j=1,2,..,r φαίνεται στο σχήμα 4.3. Συγκεκριμένα, το πρώτο διάνυσμα loading $\mathbf{p}_1 = \left[\cos(\theta_1), \cos(\theta_2)\right]^T$ περιέχει τα συνημίτονα των γωνιών θ_1 και θ_2, ενώ το διάνυσμα $\mathbf{t}_1 = \left[t_{1,}, t_{2,} \ldots, t_K\right]^T$, περιέχει τις αποστάσεις των προβολών των δεδομένων πάνω στην πρώτη κύρια συνιστώσα, από την αρχή των αξόνων. Η πρώτη κύρια συνιστώσα βρίσκεται στη διεύθυνση της μέγιστης διακύμανσης των δεδομένων, ενώ η δεύτερη κύρια συνιστώσα βρίσκεται κάθετη στη πρώτη και στη διεύθυνση της δεύτερης μέγιστης διακύμανσης.

Με βάση την προηγούμενη ανάλυση, η μέθοδος PCA εφαρμόζεται στις αποκρίσεις της κρυφής στοιβάδας $\hat{\mathbf{Z}}$ για το σύνολο εκπαίδευσης, προκειμένου να μειωθεί η διαστατικότητά τους. Οι νέες μειωμένες αποκρίσεις της κρυφής στοιβάδας $\mathbf{Z}$ μαζί με τα δεδομένα εξόδου $\mathbf{Y}_{train} = \left[\mathbf{y}(1),\ \mathbf{y}(2),\ \ldots,\ \mathbf{y}(K_{train})\right]^T$, που είναι σε αριθμητική μορφή υιοθετώντας την κωδικοποίηση "1 of C", για τον προσδιορισμό των συναπτικών βαρών



Σχήμα 4.3. Κύριες συνιστώσες (PCs) και διανύσματα score και loading για έναν πίνακα **Z** διάστασης (10×2): 10 παραδειγμάτων και 2 διαστάσεων.

$\mathbf{W} = \left[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_M\right]$, εφαρμόζοντας τη μέθοδο της γραμμικής παλινδρόμησης, όπως περιγράφεται από την εξίσωση (2.14). Στο σημείο αυτό, πρέπει να σημειωθεί ότι εφόσον η μέθοδος PCA μειώνει τον αριθμό των ενεργών κέντρων, συμβάλει στη μείωση του υπολογιστικού φορτίου, καθώς η μέθοδος των ελαχίστων τετραγώνων διαχειρίζεται λιγότερα δεδομένα. Έτσι, ολοκληρώνεται η διαδικασία της εκπαίδευσης του νευρωνικού δικτύου MRBF.

4.4.2 Επιλογή μοντέλου MRBF

Ένα νευρωνικό δίκτυο MRBF εκπαιδεύεται με βάση το σύνολο δεδομένων εκπαίδευσης. Ενώ, το σύνολο δεδομένων αξιολόγησης χρησιμοποιείται για τη ρύθμιση των παραμέτρων του μοντέλου, δηλαδή για την επιλογή του καταλληλότερου μοντέλου. Τελικά, η αξιοπιστία του τελικού επιλεγμένου μοντέλου ελέγχεται στο σύνολο δεδομένων ελέγχου.

Το τελικό μοντέλο επιλέγεται με βάση δύο παραμέτρους, τον αριθμό των ασαφών συνόλων s και το κατώφλι δ. Τα ασαφή σύνολα καθορίζουν τη διαμέριση του χώρου εισόδου που αφορά στις αριθμητικές μεταβλητές. Η κατανομή των κέντρων κατά μήκος της κατεύθυνσης της κάθε αριθμητικής μεταβλητής εισόδου είναι η ίδια, διότι χρησιμοποιείται ο ίδιος αριθμός ασαφών συνόλων για όλες τις αριθμητικές μεταβλητές, δηλαδή η ασαφής διαμέριση είναι ισομερής σε κάθε διάσταση. Ως εκ τούτου, η χρήση περισσότερων ασαφών συνόλων οδηγεί στην επιλογή περισσότερων κέντρων της κρυφής στοιβάδας και το αντίθετο. Αντίστοιχα, το κατώφλι δ είναι μια σημαντική παράμετρος, δεδομένου ότι καθορίζει τον αριθμό των κατηγορικών κέντρων και άρα το μέγεθος του κατηγορικού μέρους της κρυφής στοιβάδας του δικτύου. Μικρές τιμές του κατωφλίου δ οδηγούν σε μεγάλο αριθμό κατηγορικών κέντρων, ενώ αντίθετα, μεγάλες τιμές του κατωφλίου δ οδηγούν σε μικρό αριθμό κατηγορικών κέντρων.

Η κατάλληλη επιλογή των τιμών των ασαφών συνόλων s και του κατωφλίου δ, και κατά συνέπεια του μοντέλου MRBF, επιτυγχάνεται χρησιμοποιώντας μια διαδικασία

δοκιμής-σφάλματος στο υποσύνολο αξιολόγησης. Πιο συγκεκριμένα, εφαρμόζεται μία διαδικασία διεξοδικής αναζήτησης, η οποία βασίζεται σε μία μέθοδο αναζήτησης τύπου πλέγματος, δοκιμάζοντας πιθανούς συνδυασμούς τιμών αυτών των δύο παραμέτρων μεταξύ κατώτερων και μέγιστων ορίων. Τα όρια των ασαφών συνόλων καθορίζονται μεταξύ δύο ακέραιων αριθμών, τον ελάχιστο και τον μέγιστο αριθμό των ασαφών συνόλων, s_{min} και s_{max}, αντίστοιχα, χρησιμοποιώντας ένα σταθερό βήμα. Επίσης, όσον αφορά στην παράμετρο δ, χρησιμοποιώντας ένα σταθερό βήμα, το προτεινόμενο σύστημα δοκιμάζει τιμές μεταξύ των κατώτατων και των ανώτατων ορίων, δ_{min} και δ_{max}, αντίστοιχα. Το μέγεθος του βήματος που αφορά στο κατώφλι δ, μπορεί στη συνέχεια να μειωθεί για να εκτελεστεί μία πιο λεπτομερής αναζήτηση μεταξύ των τιμών του δ που παράγουν καλά αποτελέσματα.

Αφού κατασκευασθεί η κρυφή στοιβάδα του δικτύου MRBF, οι αποκρίσεις της περνούν μέσα από τη διαδικασία PCA. Ο αριθμός των παραγόμενων αποτελεσμάτων PCA είναι μικρότερος ή ίσος με τον αρχικό αριθμό των νευρώνων της κρυφής στοιβάδας. Τέλος, επιλέγεται το μοντέλο που αντιστοιχεί στις τιμές των ασαφών συνόλων s και του κατωφλίου δ, οι οποίες δίνουν το καλύτερο αποτέλεσμα στο υποσύνολο αξιολόγησης. Έτσι, το δίκτυο είναι έτοιμο να χρησιμοποιηθεί σε προβλέψεις για νέα σύνολα δεδομένων. Ο ψευδο-κώδικας για την προτεινόμενη διαδικασία εκπαίδευσης δίνεται στον αλγόριθμο 8.

Αλγόριθμος 8 – Διαδικασία εκπαίδευσης MRBF

Δεδομένα Εισόδου: $\{\{\mathbf{U}_{train}, \mathbf{X}_{train}\}, \mathbf{Y}_{train}\}, \{\{\mathbf{U}_{val}, \mathbf{X}_{val}\}, \mathbf{Y}_{val}\}$: Υποσύνολα εκπαίδευσης-αξιολόγησης, όπου $\mathbf{X}_{train} = [\mathbf{x}(1)\ \mathbf{x}(2)\ \ldots\ \mathbf{x}(K_{train})]^T$ είναι οι πλειάδες εισόδου εκπαίδευσης, $\{\mathbf{U}_{train}, \mathbf{Y}_{train}\}, \{\mathbf{U}_{val}, \mathbf{Y}_{val}\}$ είναι τα υποσύνολα εκπαίδευσης-αξιολόγησης αριθμητικών δεδομένων, και $\{\mathbf{X}_{train}, \mathbf{Y}_{train}\}, \{\mathbf{X}_{val}, \mathbf{Y}_{val}\}$ είναι τα υποσύνολα εκπαίδευσης-αξιολόγησης αριθμητικών δεδομένων

δ_{min}, δ_{max}: Ελάχιστη και μέγιστη τιμή κατωφλίου

s_{min}, s_{max}: Ελάχιστος και μέγιστος αριθμός των ασαφών συνόλων που διαμερίζεται το πεδίο ορισμού κάθε αριθμητικής μεταβλητής εισόδου

Δεδομένα Εξόδου: L, L_C: Αριθμός επιλεγμένων αριθμητικών κέντρων και κατηγορικών κέντρων πλειάδων, αντίστοιχα

$\hat{\mathbf{U}} = [\hat{\mathbf{u}}_1,\ \hat{\mathbf{u}}_2,\ \ldots,\ \hat{\mathbf{u}}_L]^T, \hat{\mathbf{X}} = [\hat{\mathbf{x}}_1,\ \hat{\mathbf{x}}_2,\ \ldots,\ \hat{\mathbf{x}}_{L_C}]^T$: Τοποθεσίες επιλεγμένων αριθμητικών κέντρων και κατηγορικών κέντρων πλειάδων, αντίστοιχα

$\mathbf{W}$: Συναπτικά βάρη επιλεγμένων κέντρων

1. **Για** $\delta = \delta_{min}$: δ_{max} **Εκτέλεσε:**
2. Πέρνα τα δ και $\mathbf{X}_{train}$ στον αλγόριθμο 6, για να υπολογιστούν ο αριθμός των κατηγορικών κέντρων πλειάδων L_C και τα περιεχόμενα $\hat{\mathbf{X}}$
3. **Για** $k = 1{:}K_{train}$ **Εκτέλεσε:**

4. Υπολόγισε τις αριθμητικές αποκρίσεις $\mathbf{z}_c(k)$ των πλειάδων εισόδου εκπαίδευσης χρησιμοποιώντας τον τύπο (4.10)
5. **Τέλος Για**
6. **Για** $s = s_{min}$: s_{max} **Εκτέλεσε:**
7. Δώσε τον αριθμό των ασαφών συνόλων s και το σύνολο δεδομένων εκπαίδευσης $\{\mathbf{U}_{train}, \mathbf{Y}_{train}\}$ στον αλγόριθμο 1, προκειμένου να υπολογιστεί ο συνολικός αριθμός των αριθμητικών κέντρων L και οι τοποθεσίες τους $\hat{\mathbf{U}}$
8. **Για** $k = 1{:}K_{train}$ **Εκτέλεσε:**
9. Υπολόγισε τις αποκρίσεις $\mathbf{z}(k)$ των αριθμητικών νευρώνων της κρυφής στοιβάδας χρησιμοποιώντας τον τύπο (2.6)
10. **Τέλος Για**
11. Μείωσε τις αποκρίσεις της κρυφής στοιβάδας $\hat{\mathbf{Z}}(k)$ χρησιμοποιώντας τη μέθοδο PCA
12. Υπολόγισε τα συναπτικά βάρη $\mathbf{W}$ εφαρμόζοντας τη μέθοδο των ελαχίστων τετραγώνων (εξίσωση (2.14)) στα $\mathbf{Z}(k)$ και $\mathbf{Y}_{train}$
13. Υπολόγισε τις προβλέψεις του μοντέλου στο υποσύνολο αξιολόγησης $\mathbf{Y}_{val}$ χρησιμοποιώντας τους τύπους (4.11)-(4.13)
14. **Τέλος Για**
15. **Τέλος Για**
16. Επέλεξε το μοντέλο που παράγει τα καλύτερα αποτελέσματα στο υποσύνολο αξιολόγησης $\{\{\mathbf{U}_{val}, \mathbf{X}_{val}\}, \mathbf{Y}_{val}\}$

4.4.3 Εφαρμογή δικτύων MRBF σε μικτά σύνολα δεδομένων

Το προτεινόμενο σχήμα αξιολογήθηκε σε μια ποικιλία από σύνολα μικτών δεδομένων τύπου benchmark, τα οποία περιλαμβάνουν είτε συνθετικά είτε πραγματικά δεδομένα και αφορούν τόσο σε προβλήματα προσέγγισης συνάρτησης όσο και σε προβλήματα ταξινόμησης. Αυτά τα σύνολα δεδομένων είναι διαθέσιμα στη βάση UCI Machine Learning Repository [190], εκτός από τα σύνολα δεδομένων GB Road Traffic Counts και Local police recorded crime data που μπορούν να βρεθούν στη βάση δεδομένων data.gov.uk [242]. Στον πίνακα 4.4 παρουσιάζονται όλα τα σύνολα δεδομένων που χρησιμοποιήθηκαν σε αυτή την εφαρμογή, μαζί με τους αριθμούς των παραδειγμάτων και των μεταβλητών εισόδου (κατηγορικών και αριθμητικών) και εξόδου (για τα σύνολα δεδομένων ταξινόμησης, ο αριθμός μεταβλητών εξόδου αντιστοιχεί στον αριθμό των κλάσεων). Μια σύντομη περιγραφή για κάθε σύνολο δεδομένων δίνεται παρακάτω, ενώ περισσότερες πληροφορίες μπορούν να βρεθούν στις βάσεις από τις οποίες ανακτήθηκαν.

- **Bike Sharing**: Με βάση το σύστημα capital bike sharing (CBS) στην Ουάσιγκτον των Ηνωμένων Πολιτειών της Αμερικής, τα αρχεία καταγραφής “Bike Sharing Hour” και “Bike Sharing Day” περιλαμβάνουν πληροφορίες για την ενοικίαση ποδηλάτων ανά ώρα και ανά ημέρα, αντίστοιχα [243]. Ο στόχος και των δύο συνόλων δεδομένων είναι να προβλεφθεί ο αριθμός των ποδηλάτων που ενοικιάστηκαν τα έτη 2011 και 2012,

λαμβάνοντας υπόψη πληροφορίες όπως οι καιρικές συνθήκες, το είδος της ημέρας, το είδος των χρηστών, κ.α.

- **GB Road Traffic Counts**: Αυτή η βάση δεδομένων δημιουργήθηκε από το Υπουργείο Μεταφορών και παρέχει στοιχεία για το επίπεδο των δρόμων της Μεγάλης Βρετανίας [244]. Στην παρούσα εφαρμογή χρησιμοποιήθηκε η βάση δεδομένων "raw count data major roads". Ο στόχος σε αυτό το σύνολο δεδομένων είναι να υπολογιστεί ο αριθμός των οχημάτων στο οδικό δίκτυο της Μεγάλης Βρετανίας τα έτη 2000-2015, χρησιμοποιώντας τις εξής εισόδους: την ώρα (Hour), την ώρα που πραγματοποιήθηκε η μέτρηση (dCount), τις συντεταγμένες ενός συγκεκριμένου σημείου μέτρησης (S Ref E, S Ref N), την τοποθεσία του γραφείου περιφέρειας (Region Name (GO)), το όνομα του δρόμου (Road), το είδος του δρόμου (RCategory).
- **Local police recorded crime data:** Αυτή η βάση δεδομένων περιέχει στοιχεία εγκλημάτων που έχουν καταγράψει οι αστυνομικές δυνάμεις των περιοχών της Αγγλίας και της Ουαλίας [245]. Ο στόχος αυτού του συνόλου δεδομένων είναι να προβλεφθεί ο συνολικός αριθμός των καταγεγραμμένων εγκλημάτων από τις αστυνομικές δυνάμεις τα έτη 2008-2012, χρησιμοποιώντας τις εξής εισόδους: πληροφορίες για τις επιθέσεις (offence code, offence group, offence subgroup, offence description, force name) και τη

Πίνακας 4.4. Επισκόπηση συνόλων δεδομένων τύπου benchmark

Σύνολο δεδομένων	**Συντομο-γραφία**	**Αριθμός αριθμητικών μεταβλητών εισόδου**	**Αριθμός κατηγορικών μεταβλητών εισόδου**	**Αριθμός μεταβλητών εξόδου**[β]	**Αριθμός δεδομένων**
ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ ΠΡΟΣΕΓΓΙΣΗΣ ΣΥΝΑΡΤΗΣΗΣ					
Bike Sharing (Day)	BSD	8	6	1	731
Bike Sharing (Hour)	BSH	8	7	1	17379
GB Road Traffic Counts	GBR	4	3	1	1048575
Local police recorded crime data	LPR	2	5	1	128845
Student Performance (Math course)	SPM	15	17	1	395
Student Performance (Portuguese language course)	SPP	15	17	1	649
ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ ΤΑΞΙΝΟΜΗΣΗΣ					
Annealing	ANN	6	32	5	898
Australian Credit Approval (Statlog)	ACA	6	8	2	690
Census Income	CIN	6	8	2	45222
Diabetic Retinopathy Debrecen	DRD	16	3	2	1151
Heart (Statlog)	HEA	6	7	2	270
QSAR biodegradation	QSA	38	3	2	1055

[β]Για τα σύνολα δεδομένων ταξινόμησης, ο αριθμός μεταβλητών εξόδου αντιστοιχεί στον αριθμό των κλάσεων

χρονική περίοδο (οικονομικό έτος και οικονομικό τετράμηνο).

- **Student Performance Data Set**: Αυτή η βάση δεδομένων περιέχει δύο σύνολα δεδομένων, το ‘‘Student Performance Math’’ και το ‘‘Student Performance Port’’, τα οποία περιέχουν τις επιδόσεις των μαθητών στη δευτεροβάθμια εκπαίδευση στα πορτογαλικά σχολεία σε δύο μαθήματα: τα μαθηματικά και τη πορτογαλική γλώσσα, αντίστοιχα [246].
- **Annealing**: Αυτό το σύνολο δεδομένων είναι επίσης γνωστό ως “Steel annealing data” και είναι ένα τυπικό παράδειγμα συνόλου δεδομένων υψηλών διαστάσεων, καθώς περιλαμβάνει 37 μεταβλητές εισόδου. Η έξοδος αποτελείται από 5 διαφορετικές κλάσεις.
- **Australian Credit Approval (Statlog):** Αυτό το σύνολο δεδομένων αφορά σε εφαρμογές των πιστωτικών καρτών. Για την προστασία του απορρήτου των δεδομένων, οι πληροφορίες δίνονται κωδικοποιημένες σε σύμβολα.
- **Census Income**: Αυτό το σύνολο δεδομένων είναι επίσης γνωστό ως “Adult”. Ο στόχος είναι να εκτιμηθεί αν το επίπεδο του ετήσιου εισοδήματος των πολιτών των ΗΠΑ είναι πάνω από 50K$, με βάση κάποια χαρακτηριστικά όπως το επίπεδο εκπαίδευσης, η οικογενειακή κατάσταση, το επάγγελμα, το φύλο, η ηλικία, κ.α.
- **Diabetic Retinopathy Debrecen:** Αυτό το σύνολο δεδομένων περιλαμβάνει χαρακτηριστικά γνωρίσματα που εξάγονται από το σύνολο δεδομένων εικόνας Messidor με στόχο να προβλεφθεί αν μια εικόνα περιέχει σημάδια διαβητικής αμφιβληστροειδοπάθειας ή όχι [247].
- **Heart (Statlog)**: Ο στόχος σε αυτό το σύνολο δεδομένων είναι να αναγνωρισθεί η παρουσία καρδιακών παθήσεων σε έναν ασθενή με βάση στοιχεία όπως η ηλικία, το φύλο, το είδος του πόνου στο στήθος, κ.α.
- **QSAR biodegradation Data Set**: Αυτό το σύνολο δεδομένων μελετά τις σχέσεις μεταξύ της χημικής δομής και της βιοδιάσπασης των μορίων [248]. Χρησιμοποιούνται 41 μοριακοί δείκτες προκειμένου να καθοριστεί αν μια χημική ουσία είναι βιοαποικοδομήσιμη ή όχι.

Για κάθε σύνολο δεδομένων, τα δεδομένα κατανεμήθηκαν με τυχαίο τρόπο σε τρία υποσύνολα εκπαίδευσης, αξιολόγησης και ελέγχου, χρησιμοποιώντας το 50% των δεδομένων για εκπαίδευση, το 25% για αξιολόγηση και το υπόλοιπο 25% για έλεγχο των νευρωνικών δικτύων. Στην περίπτωση των προβλημάτων κατηγοριοποίησης, δόθηκε ιδιαίτερη προσοχή στο να κατανεμηθούν τα δεδομένα που ανήκουν στις ίδιες κλάσεις, ομοιόμορφα μεταξύ των τριών υποσυνόλων σύμφωνα με την παραπάνω αναλογία.

Για λόγους σύγκρισης, χρησιμοποιήθηκαν οι εξής τεχνικές: τα δίκτυα MLP που εκπαιδεύτηκαν με τον αλγόριθμο Levenberg-Marquardt [4], οι μηχανές διανυσμάτων υποστήριξης αποφάσεων SVM με γκαουσιανή συνάρτηση πυρήνα [228], τα δέντρα ταξινόμησης και παλινδρόμησης (CART) [236] και οι μηχανές ακραίας μάθησης (extreme learning machine, ELM) [249] με πέντε διαφορετικές συναρτήσεις ενεργόποιησης. Όλες οι μέθοδοι που χρησιμοποιήθηκαν σε αυτή την εφαρμογή εκτός τις μηχανές SVM, εκτελέστηκαν 30 φορές για κάθε σύνολο δεδομένων. Επίσης, όλες οι μέθοδοι βελτιστοποιήθηκαν ξεχωριστά για κάθε σύνολο δεδομένων.

Για τα δίκτυα MLP, χρησιμοποιήθηκε μία αρχιτεκτονική δύο κρυφών στοιβάδων όπου εκτελέστηκε μια διαδικασία διεξοδικής αναζήτησης, δοκιμάζοντας όλους τους πιθανούς συνδυασμούς ανάμεσα στον αριθμό των νευρώνων της πρώτης κρυφής στοιβάδας που κυμαίνεται από 10 έως 40 και της δεύτερης κρυφής στοιβάδας που κυμαίνεται από 10 έως 30.

Οι ταξινομητές SVM χρησιμοποιήθηκαν για προβλήματα ταξινόμησης, ενώ για προβλήματα προσέγγισης συνάρτησης εφαρμόστηκε μια παραλλαγή των μηχανών SVM, οι

μηχανές SVM προσέγγισης συνάρτησης (SVM-regression, SVMr). Αμφότεροι οι ταξινομητές SVM και οι μηχανές SVMr κατασκευάστηκαν με γκαουσιανή συνάρτηση πυρήνα. Επίσης, οι παράμετροι των μηχανών SVM και SVMr βελτιστοποιήθηκαν με βάση μια διαδικασία διεξοδικής αναζήτησης πλέγματος, παρόμοια με αυτή που χρησιμοποιήθηκε στα δίκτυα MLP, βελτιστοποιώντας έτσι το εύρος των γκαουσιανών σ και τον συντελεστή ποινής C. Στην περίπτωση των συνόλων δεδομένων με περισσότερες από δύο κατηγορίες εξόδου, εφαρμόστηκε μία τεχνική ένας-εναντίον-όλων (one-against-all), όπως προτάθηκε από τον Vapnik [228].

Τα δέντρα CART βελτιστοποιήθηκαν με βάση τον αριθμό των παρατηρήσεων σε κάθε κλαδί και τον αριθμό των μεταβλητών εισόδου που επιλέγονται τυχαία. Για τις μηχανές ELM εφαρμόστηκαν οι εξής συναρτήσεις ενεργοποίησης: η συνάρτηση κατωφλίου (Hard-limit function, har), η συνάρτηση ακτινικής βάσης (radial basis function, rad), η σιγμοειδής συνάρτηση (sigmoidal function, sig), η ημιτονοειδής συνάρτηση (sine function, sin), η συνάρτηση τριγωνικής βάσης (triangular basis function, tri). Επίσης, οι μηχανές ELM βελτιστοποιήθηκαν με βάση των αριθμό των πυρήνων.

Σε όλες τις περιπτώσεις, οι παράμετροι όλων των μεθόδων που χρησιμοποιήθηκαν σε αυτή την εφαρμογή προσδιορίστηκαν χρησιμοποιώντας το σύνολο δεδομένων εκπαίδευσης, στη συνέχεια η επιλογή μοντέλου επιτεύχθηκε με βάση το σύνολο δεδομένων αξιολόγησης, και τελικά η απόδοση του επιλεγμένου μοντέλου ελέγχθηκε στο σύνολο ελέγχου.

4.4.4 Αποτελέσματα και συζήτηση

Τα αποτελέσματα για τα σύνολα δεδομένων προσέγγισης συνάρτησης και ταξινόμησης, συνοψίζονται στους πίνακες 4.5 και 4.6, αντίστοιχα. Στους συγκεκριμένους πίνακες απεικονίζονται για όλες τις μεθόδους στα σύνολα δεδομένων αξιολόγησης και ελέγχου, οι δείκτες σφάλματος RMSE και R^2 για τα σύνολα δεδομένων προσέγγισης συνάρτησης, οι δείκτες σφάλματος MCC και ακρίβεια% για τα σύνολα δεδομένων ταξινόμησης. Επιπλέον, απεικονίζονται για την προτεινόμενη μέθοδο η επιλεγμένη ασαφής διαμέριση και το επιλεγμένο κατώφλι δ, για τα δίκτυα MLP εμφανίζονται οι νευρώνες των δύο κρυφών στοιβάδων, για τις δύο μηχανές SVM φαίνονται ο συντελεστής ποινής C και το εύρος των γκαουσιανών πυρήνων σ, και για τη μέθοδο ELM αναφέρεται ο αριθμός των πυρήνων. Στους πίνακες απεικονίζεται το καλύτερο αποτέλεσμα και σε παρένθεση ο μέσος όρος και η τυπική απόκλιση των 30 αποτελεσμάτων που έδωσαν οι παραπάνω τεχνικές.

Φαίνεται ότι τα μοντέλα της προτεινόμενης μεθοδολογίας παράγουν μεγαλύτερη ακρίβεια από την άποψη της βέλτιστης και της μέσης λύσης. Πιο συγκεκριμένα, στα προβλήματα προσέγγισης συνάρτησης, όσον αφορά στον δείκτη RMSE και στα σύνολα αξιολόγησης και ελέγχου, τα δίκτυα MRBF υπερτερούν έναντι όλων των μεθόδων στα 5 από τα 6 σύνολα δεδομένων, με βάση το καλύτερο αποτέλεσμα και τον μέσο όρο. Μόνο τα δίκτυα MLP ξεπερνούν τα δίκτυα MRBF σε 1 σύνολο δεδομένων. Περίπου παρόμοια αποτελέσματα εμφανίζονται και με βάση τον δείκτη R^2, με τη διαφορά ότι τα δίκτυα MLP δίνουν λίγο καλύτερα αποτελέσματα και η τεχνική ELM-sig καταφέρνει και προβλέπει με την ίδια ακρίβεια που προβλέπουν τα δίκτυα MRBF, σε ένα σύνολο αξιολόγησης όσον αφορά την καλύτερη τιμή, σε δύο σύνολα αξιολόγησης όσον αφορά στη μέση τιμή, και σε δύο σύνολα ελέγχου όσον αφορά στη καλύτερη και στη μέση τιμή.

Στα προβλήματα ταξινόμησης, όσον αφορά στο σύνολο αξιολόγησης και στον δείκτη MCC, τα δίκτυα MRBF υπερτερούν έναντι όλων των μεθόδων στα 5 από τα 6 σύνολα δεδομένων, με βάση το καλύτερο αποτέλεσμα και στα 4 από τα 6 σύνολα δεδομένων με βάση

τον μέσο όρο. Οι τεχνικές CART, ELM-har και ELM-sig ξεπερνούν τα δίκτυα MRBF σε 1 σύνολο δεδομένων αξιολόγησης δίνοντας μεταξύ τους το ίδιο καλύτερο αποτέλεσμα, ενώ οι τεχνικές SVM και CART συνολικά ξεπερνούν τα δίκτυα MRBF σε 2 σύνολα δεδομένων αξιολόγησης με βάση τον μέσο όρο. Στα προβλήματα ταξινόμησης, όσον αφορά στο σύνολο ελέγχου και στον δείκτη MCC, τα δίκτυα MRBF υπερτερούν έναντι όλων των μεθόδων στα 4 από τα 6 σύνολα δεδομένων, με βάση το καλύτερο αποτέλεσμα και τον μέσο όρο. Οι τεχνικές CART, ELM-har και ELM-sig ξεπερνούν τα δίκτυα MRBF σε 1 σύνολο δεδομένων αξιολόγησης δίνοντας μεταξύ τους το ίδιο καλύτερο αποτέλεσμα και σε 1 ακόμη σύνολο δεδομένων δίνουν το καλύτερο αποτέλεσμα οι μηχανές SVM, ενώ οι τεχνικές SVM και CART συνολικά ξεπερνούν τα δίκτυα MRBF σε 2 σύνολα δεδομένων ελέγχου με βάση τον μέσο όρο. Τα ίδια ακριβώς αποτελέσματα εμφανίζονται και με βάση τον δείκτη ακρίβεια%.

Σε αυτό το σημείο, θα πρέπει να αναφερθεί ότι τα αποτελέσματα στα σύνολα αξιολόγησης παρατίθενται για λόγους πληρότητας. Ουσιαστικά, τα αποτελέσματα που δείχνουν αν ένα μοντέλο αποδίδει καλά ή όχι, αντιστοιχούν στο σύνολο ελέγχου, το οποίο είναι ένα ξεχωριστό σύνολο δεδομένων και δεν χρησιμοποιείται στη διαδικασία δημιουργίας ή επιλογής των μοντέλων.

Πίνακας 4.5. Αποτελέσματα για τα σύνολα δεδομένων προσέγγισης συνάρτησης – RMSE και R^2

Σύνολο δεδομένων	**Μέθοδος**	**RMSE συνόλου αξιολόγησης**	**RMSE συνόλου ελέγχου**	**R^2 συνόλου αξιολόγησης**	**R^2 συνόλου ελέγχου**	**Παράμετροι[α]**
BSD	MRBF	**5.84E+00**(**6.13E+00**±2.07E-01)	**9.29E+00**(**9.38E+00**±3.75E-01)	**1.00E+00**(**1.00E+00**±7.17E-07)	**1.00E+00**(**1.00E+00**±1.88E-06)	[32 0.95]
	MLP	1.40E+01(8.25E+01±3.67E+01)	9.76E+00(8.12E+01±4.03E+01)	**1.00E+00**(9.99E-01±1.14E-03)	**1.00E+00**(9.99E-01±1.21E-03)	[20 11]
	SVR	3.90E+02(3.90E+02±0.00E+00)	4.18E+02(4.18E+02±0.00E+00)	9.58E-01(9.58E-01±0.00E+00)	9.54E-01(9.54E-01±0.00E+00)	[1024 3.66]
	CART	2.14E+02(2.52E+02±3.13E+01)	2.24E+02(2.48E+02±1.81E+01)	9.87E-01(9.82E-01±4.67E-03)	9.87E-01(9.84E-01±2.42E-03)	[1 13]
	ELM-har	2.66E+02(2.33E+02±1.38E+01)	2.19E+02(2.20E+02±1.41E+01)	9.80E-01(9.85E-01±1.82E-03)	9.87E-01(9.87E-01±1.67E-03)	1900
	ELM-rad	3.80E+02(3.47E+02±1.53E+01)	3.62E+02(3.44E+02±2.30E+01)	9.60E-01(9.67E-01±2.95E-03)	9.65E-01(9.69E-01±4.31E-03)	1960
	ELM-sig	4.45E+01(3.61E+01±3.24E+00)	3.87E+01(3.36E+01±2.55E+00)	9.99E-01(**1.00E+00**±6.60E-05)	**1.00E+00**(**1.00E+00**±4.52E-05)	1950
	ELM-sin	3.06E+02(2.75E+02±1.64E+01)	2.76E+02(2.76E+02±1.92E+01)	9.74E-01(9.79E-01±2.49E-03)	9.80E-01(9.80E-01±2.87E-03)	1910
	ELM-tri	5.85E+02(5.24E+02±3.65E+01)	5.46E+02(5.03E+02±3.50E+01)	9.05E-01(9.24E-01±1.05E-02)	9.21E-01(9.32E-01±9.25E-03)	1910
BSH	MRBF	1.24E-01(1.26E-01±1.67E-03)	8.94E-02(9.08E-02±1.46E-03)	**1.00E+00**(**1.00E+00**±1.36E-08)	**1.00E+00**(**1.00E+00**±8.12E-09)	[33 0.16]
	MLP	**3.34E-02**(**1.18E-02**±9.16E-04)	**3.48E-02**(**1.56E-02**±1.26E-03)	**1.00E+00**(**1.00E+00**±8.10E-11)	**1.00E+00**(**1.00E+00**±1.58E-10)	[27 16]
	SVR	1.06E+01(1.06E+01±0.00E+00)	1.05E+01(1.05E+01±0.00E+00)	9.96E-01(9.96E-01±0.00E+00)	9.97E-01(9.97E-01±0.00E+00)	[1024 4.22]
	CART	5.73E+00(6.11E+00±3.43E-01)	5.56E+00(6.01E+00±3.06E-01)	9.99E-01(9.99E-01±1.39E-04)	9.99E-01(9.99E-01±1.16E-04)	[4 14]
	ELM-har	2.01E+01(1.97E+01±2.40E-01)	1.95E+01(1.98E+01±3.51E-01)	9.87E-01(9.88E-01±3.06E-04)	9.88E-01(9.88E-01±4.28E-04)	4600
	ELM-rad	2.59E+01(2.48E+01±5.03E-01)	2.54E+01(2.47E+01±5.88E-01)	9.78E-01(9.80E-01±8.12E-04)	9.80E-01(9.81E-01±8.84E-04)	4400
	ELM-sig	2.82E+00(2.71E+00±5.67E-02)	2.82E+00(2.70E+00±5.80E-02)	**1.00E+00**(**1.00E+00**±9.93E-06)	**1.00E+00**(**1.00E+00**±9.64E-06)	4800
	ELM-sin	1.87E+01(1.81E+01±4.06E-01)	1.86E+01(1.82E+01±3.62E-01)	9.89E-01(9.89E-01±4.73E-04)	9.89E-01(9.90E-01±4.01E-04)	4600
	ELM-tri	4.20E+01(4.01E+01±1.19E+00)	4.16E+01(4.00E+01±8.00E-01)	9.43E-01(9.48E-01±3.08E-03)	9.47E-01(9.51E-01±1.96E-03)	4000

Απεικονίζεται το καλύτερο αποτέλεσμα που βρέθηκε σύμφωνα με τον δείκτη RMSE στο σύνολο αξιολόγησης και ακολουθείται από τον μέσο όρο και την τυπική απόκλιση των 30 αποτελεσμάτων μέσα σε παρένθεση.

[α] Οι παράμετροι περιέχουν για τα MRBF: [ασαφής διαμέριση κατώφλι δ], για τα MLPs: [νευρώνες 1ης στοιβάδας νευρώνες 2ης στοιβάδας], για τα SVR: [συντελεστής ποινής C εύρος γκαουσιανών πυρήνων σ], για τα ELMs: αριθμός πυρήνων.

Πίνακας 4.5. Αποτελέσματα για τα σύνολα δεδομένων προσέγγισης συνάρτησης – RMSE και R^2 *(συνέχεια)*

Σύνολο δεδομένων	**Μέθοδος**	**RMSE συνόλου αξιολόγησης**	**RMSE συνόλου ελέγχου**	**R^2 συνόλου αξιολόγησης**	**R^2 συνόλου ελέγχου**	**Παράμετροι[α]**
GBR	MRBF	**3.06E+02**(**3.08E+02**±1.30E-01)	**3.08E+02**(**3.09E+02**±1.44E-01)	**9.25E-01**(**9.24E-01**±2.59E-04)	**9.24E-01**(**9.24E-01**±7.15E-03)	[19 0.04]
	MLP	5.64E+02(5.53E+02±1.73E+01)	5.61E+02(5.52E+02±1.40E+01)	8.00E-01(8.07E-01±1.19E-02)	8.03E-01(8.10E-01±9.60E-03)	[30 18]
	SVR	5.43E+02(5.43E+02±0.00E+00)	5.48E+02(5.48E+02±0.00E+00)	8.34E-01(8.34E-01±0.00E+00)	8.30E-01(8.30E-01±0.00E+00)	[891.44 2.745]
	CART	3.57E+02(3.63E+02±4.49E+00)	3.62E+02(3.70E+02±6.38E+00)	9.20E-01(9.17E-01±2.06E-03)	9.18E-01(9.14E-01±2.94E-03)	[10 6]
	ELM-har	5.86E+02(5.85E+02±2.21E+00)	5.80E+02(5.82E+02±2.43E+00)	7.84E-01(7.85E-01±1.62E-03)	7.89E-01(7.88E-01±1.78E-03)	15000
	ELM-rad	5.04E+02(5.03E+02±1.59E+00)	5.00E+02(4.99E+02±1.45E+00)	8.40E-01(8.41-01±1.01E-03)	8.43E-01(8.44E-01±9.06E-04)	10000
	ELM-sig	4.70E+02(4.66E+02±3.71E+00)	4.63E+02(4.63E+02±1.85E+00)	8.61E-01(8.63E-01±2.19E-03)	8.66E-01(8.66E-01±1.07E-03)	10000
	ELM-sin	4.65E+02(4.63E+02±3.02E+00)	4.65E+02(4.62E+02±4.03E+00)	8.64E-01(8.65E-01±1.76E-03)	8.64E-01(8.66E-01±2.33E-03)	10000
	ELM-tri	5.17E+02(5.15E+02±2.22E+00)	5.26E+02(5.18E+02±1.23E+00)	8.32E-01(8.33-01±1.44E-03)	8.26E-01(8.32E-01±7.97E-04)	15000
LPR	MRBF	**2.14E+02**(**2.16E+02**±4.40E-01)	**1.85E+02**(**1.87E+02**±9.56E-01)	**9.40E-01**(**9.37E-01**±2.59E-04)	**9.41E-01**(**9.40E-01**±6.21E-04)	[5 0.02]
	MLP	2.15E+02(2.27E+02±4.06E+01)	1.87E+02(2.28E+02±5.19E+01)	9.38E-01(9.28E-01±2.54E-02)	9.40E-01(9.06E+00±4.20E-02)	[23 20]
	SVR	2.87E+02(2.87E+02±0.00E+00)	2.90E+02(2.90E+02±0.00E+00)	8.73E-01(8.73E-01±0.00E+00)	8.10E-01(8.10E-01±0.00E+00)	[955.426 3.13]
	CART	2.18E+02(3.14E+02±6.78E+01)	2.04E+02(3.05E+02±9.24E+01)	9.35E-01(8.60E-01±6.46E-02)	9.28E-01(8.25E-01±1.14E-01)	[5 5]
	ELM-har	6.92E+02(6.84E+02±4.00E+00)	6.51E+02(6.53E+02±2.70E+00)	3.48E-01(3.62E-01±7.48E-03)	2.67E-01(2.64E-01±6.09E-03)	10000
	ELM-rad	5.75E+02(5.64E+02±8.79E+00)	5.52E+02(5.45E+02±6.95E+00)	5.49E-01(5.66E-01±1.35E-02)	4.74E-01(4.87E-01±1.31E-02)	20000
	ELM-sig	2.26E+02(2.23E+02±2.25E+00)	2.21E+02(2.22E+02±1.72E+00)	9.30E-01(9.32E-01±1.37E-03)	9.15E-01(9.15E-01±1.32E-03)	25000
	ELM-sin	2.85E+02(2.78E+02±4.89E+00)	2.74E+02(2.72E+02±3.88E+00)	8.89E-01(8.95E-01±3.69E-03)	8.70E-01(8.72E-01±3.63E-03)	25000
	ELM-tri	7.49E+02(7.44E+02±3.56E+00)	6.71E+02(6.69E+02±3.93E+00)	2.35E-01(2.46E-01±7.22E-03)	2.23E-01(2.26E-01±9.07E-03)	5000

Απεικονίζεται το καλύτερο αποτέλεσμα που βρέθηκε σύμφωνα με τον δείκτη RMSE στο σύνολο αξιολόγησης και ακολουθείται από τον μέσο όρο και την τυπική απόκλιση των 30 αποτελεσμάτων μέσα σε παρένθεση.

[α] Οι παράμετροι περιέχουν για τα MRBF: [ασαφής διαμέριση κατώφλι δ], για τα MLPs: [νευρώνες 1ης στοιβάδας νευρώνες 2ης στοιβάδας], για τα SVR: [συντελεστής ποινής C εύρος γκαουσιανών πυρήνων σ], για τα ELMs: αριθμός πυρήνων.

Πίνακας 4.5. Αποτελέσματα για τα σύνολα δεδομένων προσέγγισης συνάρτησης – RMSE και R^2 *(συνέχεια)*

Σύνολο δεδομένων	**Μέθοδος**	**RMSE συνόλου αξιολόγησης**	**RMSE συνόλου ελέγχου**	**R^2 συνόλου αξιολόγησης**	**R^2 συνόλου ελέγχου**	**Παράμετροι[α]**
SPM	MRBF	**1.78E+00(1.87E+00**±5.38E-02)	**2.22E+00(2.25E+00**±4.29E-02)	**8.36E-01(8.19E-01**±1.05E-02)	**8.12E-01(8.06E-01**±7.36E-03)	[5 0.91]
	MLP	3.04E+00(3.52E+00±3.01E-01)	2.99E+00(3.74E+00±3.27E-01)	5.20E-01(3.53E-01±1.13E-01)	6.58E-01(4.59E-01±9.35E-02)	[27 10]
	SVR	2.34E+00(2.34E+00±0.00E+00)	2.65E+00(5.65E+00±0.00E+00)	7.18E-02(7.18E-02±0.00E+00)	7.34E-02(7.34E-02±0.00E+00)	[0.016 7.14]
	CART	2.43E+00(2.48E+00±7.38E-02)	2.55E+00(2.53E+00±3.22E-02)	6.94E-01(6.80E-01±2.00E-02)	7.50E-01(7.54E-01±6.34E-03)	[20 31]
	ELM-har	2.50E+00(2.30E+00±9.34E-02)	2.75E+00(2.71E+00±9.22E-02)	6.76E-01(7.25E-01±2.23E-02)	7.11E-01(7.19E-01±1.90E-02)	1880
	ELM-rad	3.08E+00(2.88E+00±1.01E-01)	3.28E+00(3.49E+00±1.51E-01)	5.06E-01(5.67E-01±3.01E-02)	5.88E-01(5.34E-01±3.98E-02)	1930
	ELM-sig	2.40E+00(2.24E+00±8.46E-02)	2.50E+00(2.60E+00±1.08E-01)	7.01E-01(7.38E-01±1.97E-02)	7.60E-01(7.41E-01±2.16E-02)	1990
	ELM-sin	4.85E+00(4.41E+00±1.87E-01)	5.78E+00(5.53E+00±1.65E-01)	-2.19E-01(-1.28E-02±8.64E-02)	-2.83E-01(-1.73E-01±7.03E-02)	1990
	ELM-tri	3.54E+00(3.28E+00±1.14E-01)	3.88E+00(3.97E+00±1.69E-01)	3.49E-01(4.40E-01±3.88E-02)	4.22E-01(3.95E-01±5.06E-02)	1930
SPP	MRBF	**1.06E+00(1.11E+00**±3.18E-02)	**1.18E+00(1.18E+00**±2.90E-02)	**8.79E-01(8.66E-01**±7.68E-03)	**8.63E-01(8.63E-01**±6.79E-03)	[5 0.95]
	MLP	1.72E+00(2.21E+00±1.93E-01)	2.03E+00(2.24E+00±1.75E-01)	6.82E-01(4.69E-01±9.08E-02)	5.93E-01(5.00E-01±7.96E-02)	[21 10]
	SVR	1.12E+00(1.12E+00±0.00E+00)	1.48E+00(1.48E+00±0.00E+00)	8.65E-01(8.65E-01±0.00E+00)	7.84E-01(7.84E-01±0.00E+00)	[3.249 6.974]
	CART	1.10E+00(1.18E+00±3.76E-02)	1.21E+00(1.22E+00±6.23E-02)	8.70E-01(8.50E-01±9.62E-03)	8.56E-01(8.52E-01±1.54E-02)	[40 29]
	ELM-har	1.61E+00(1.45E+00±6.92E-02)	1.67E+00(1.67E+00±7.48E-02)	7.22E-01(7.74E-01±2.18E-02)	7.26E-01(7.24E-01±2.48E-02)	1990
	ELM-rad	2.11E+00(1.93E+00±9.57E-02)	2.28E+00(2.20E+00±8.18E-02)	5.19E-01(6.00E-01±3.98E-02)	4.88E-01(5.23E-01±3.52E-02)	1950
	ELM-sig	1.51E+00(1.40E+00±6.44E-02)	1.50E+00(1.44E+00±4.94E-02)	7.55E-01(7.89E-01±1.94E-02)	7.78E-01(7.95E-01±1.42E-02)	1980
	ELM-sin	4.18E+00(3.84E+00±1.66E-01)	4.20E+00(3.84E+00±1.89E-01)	-8.82E-01(-5.90E-01±1.40E-01)	-7.44E-01(-4.61E-01±1.45E-01)	1920
	ELM-tri	2.45E+00(2.29E+00±8.06E-02)	2.41E+00(2.50E+00±1.19E-01)	3.55E-01(4.33E-01±3.93E-02)	4.25E-01(3.84E-01±5.83E-02)	1980

Απεικονίζεται το καλύτερο αποτέλεσμα που βρέθηκε σύμφωνα με τον δείκτη RMSE στο σύνολο αξιολόγησης και ακολουθείται από τον μέσο όρο και την τυπική απόκλιση των 30 αποτελεσμάτων μέσα σε παρένθεση.

[α] Οι παράμετροι περιέχουν για τα MRBF: [ασαφής διαμέριση κατώφλι δ], για τα MLPs: [νευρώνες 1ης στοιβάδας νευρώνες 2ης στοιβάδας], για τα SVR: [συντελεστής ποινής C εύρος γκαουσιανών πυρήνων σ], για τα ELMs: αριθμός πυρήνων.

Πίνακας 4.6. Αποτελέσματα για τα σύνολα δεδομένων ταξινόμησης – MCC και ακρίβεια %

Σύνολο δεδομένων	**Μέθοδος**	**MCC συνόλου αξιολόγησης**	**MCC συνόλου ελέγχου**	**Ακρίβεια% συνόλου αξιολόγησης**	**Ακρίβεια% συνόλου ελέγχου**	**Παράμετροι[α]**
ANN	MRBF	9.89E-01(9.87E-01±1.28E-02)	**9.68E-01**(**9.55E-01**±1.55E-02)	9.98E+01(9.92E+01±5.52E-01)	**9.87E+01**(**9.81E+01**±6.32E-01)	[12 0.43]
	MLP	9.78E-01(9.53E-01±1.95E-02)	9.24E-01(9.43E-01±2.11E-02)	9.91E+01(9.81E+01±7.69E-01)	9.69E+01(9.77E+01±8.28E-01)	[16 13]
	SVM	9.78E-01(9.78E-01±0.00E+00)	9.33E-01(9.33E-01±0.00E+00)	9.91E+01(9.91E+01±0.00E+00)	9.73E+01(9.73E+01±0.00E+00)	[128 7.73]
	CART	**1.00E+00**(**1.00E+00**±1.13E-16)	9.58E-01(9.49E-01±7.71E-03)	**1.00E+02**(**1.00E+02**±0.00E+00)	9.82E+01(9.78E+01±3.33E-01)	[10 34]
	ELM-har	**1.00E+00**(9.64E-01±1.67E-02)	9.55E-01(9.51E-01±1.77E-02)	**1.00E+02**(9.86E+01±6.64E-01)	9.82E+01(9.80E+01±6.98E-01)	1990
	ELM-rad	9.88E-01(9.51E-01±2.39E-02)	8.98E-01(9.03E-01±1.95E-02)	9.95E+01(9.80E+01±9.38E-01)	9.60E+01(9.62E+01±7.66E-01)	1890
	ELM-sig	**1.00E+00**(9.61E-01±2.19E-02)	9.36E-01(9.16E-01±3.41E-02)	**1.00E+02**(9.84E+01±8.71E-01)	9.73E+01(9.66E+01±1.42E+00)	230
	ELM-sin	9.78E-01(9.13E-01±3.95E-02)	8.32E-01(8.57E-01±4.39E-02)	9.91E+01(9.65E+01±1.57E+00)	9.33E+01(9.42E+01±1.79E+00)	210
	ELM-tri	9.67E-01(9.37E-01±1.82E-02)	9.21E-01(9.05E-01±2.33E-02)	9.87E+01(9.75E+01±7.15E-01)	9.69E+01(9.63E+01±8.80E-01)	1840
ACA	MRBF	**7.68E-01**(**7.38E-01**±1.68E-02)	**8.13E-01**(**7.85E-01**±2.52E-02)	**8.84E+01**(**8.67E+01**±8.58E-01)	**9.08E+01**(**8.94E+01**±1.24E+00)	[5 0.23]
	MLP	7.46E-01(6.26E-01±4.30E-02)	6.83E-01(6.65E-01±2.43E-02)	8.73E+01(8.14E+01±2.14E+00)	8.44E+01(8.34E+01±1.19E+00)	[24 10]
	SVM	7.34E-01(7.34E-01±0.00E+00)	7.84E-01(7.84E-01±0.00E+00)	8.61E+01(8.61E+01±0.00E+00)	8.90E+01(8.90E+01±0.00E+00)	[8 5.098]
	CART	7.18E-01(7.12E-01±1.72E-02)	7.87E-01(7.84E-01±9.20E-03)	8.50E+01(8.46E+01±1.06E+00)	8.90E+01(8.88E+01±5.29E-01)	[150 13]
	ELM-har	7.23E-01(6.27E-01±4.52E-02)	7.56E-01(7.23E-01±4.91E-02)	8.61E+01(8.13E+01±2.26E+00)	8.79E+01(8.63E+01±2.44E+00)	60
	ELM-rad	7.31E-01(6.22E-01±5.84E-02)	6.83E-01(7.00E-01±5.28E-02)	8.67E+01(8.11E+01±2.87E+00)	8.44E+01(8.52E+01±2.58E+00)	50
	ELM-sig	7.74E-01(7.00E-01±2.85E-02)	7.91E-01(7.80E-01±2.86E-02)	8.83E+01(8.46E+01±1.42E+00)	8.96E+01(8.91E+01±1.43E+00)	40
	ELM-sin	7.11E-01(6.52E-01±3.70E-02)	6.72E-01(7.37E-01±4.27E-02)	8.55E+01(8.25E+01±1.83E+00)	8.38E+01(8.69E+01±2.13E+00)	60
	ELM-tri	6.64E-01(5.58E-01±5.09E-02)	6.95E-01(6.26E-01±4.70E-02)	8.32E+01(7.79E+01±2.53E+00)	8.50E+01(8.15E+01±2.34E+00)	90

Απεικονίζεται το καλύτερο αποτέλεσμα που βρέθηκε σύμφωνα με τον δείκτη RMSE στο σύνολο αξιολόγησης και ακολουθείται από τον μέσο όρο και την τυπική απόκλιση των 30 αποτελεσμάτων μέσα σε παρένθεση.

[α] Οι παράμετροι περιέχουν για τα MRBF: [ασαφής διαμέριση κατώφλι δ], για τα MLPs: [νευρώνες 1ης στοιβάδας νευρώνες 2ης στοιβάδας], για τα SVMs: [συντελεστής ποινής C εύρος γκαουσιανών πυρήνων σ], για τα ELMs: αριθμός πυρήνων.

Πίνακας 4.6. Αποτελέσματα για τα σύνολα δεδομένων ταξινόμησης – MCC και ακρίβεια % *(συνέχεια)*

Σύνολο δεδομένων	**Μέθοδος**	**MCC συνόλου αξιολόγησης**	**MCC συνόλου ελέγχου**	**Ακρίβεια% συνόλου αξιολόγησης**	**Ακρίβεια% συνόλου ελέγχου**	**Παράμετροι[α]**
CIN	MRBF	**5.90E-01**(**5.85E-01**±2.45E-03)	**5.92E-01**(**5.91E-01**±1.64E-03)	**8.55E+01**(**8.53E+01**±8.42E-02)	**8.56E+01**(**8.56E+01**±5.56E-02)	[27 0.21]
	MLP	5.66E-01(5.57E-01±6.07E-03)	5.65E-01(5.62E-01±4.07E-03)	8.46E+01(8.42E+01±2.14E-01)	8.45E+01(8.45E+01±1.27E-01)	[32 19]
	SVM	5.60E-01(5.60E-01±0.00E+00)	5.68E-01(5.68E-01±0.00E+00)	7.96E+01(7.96E+01±0.00E+00)	8.01E+01(8.01E+01±0.00E+00)	[0.574 4.14]
	CART	5.77E-01(5.64E-01±8.05E-03)	5.84E-01(5.78E-01±8.25E-03)	8.51E+01(8.45E+01±2.75E-01)	8.54E+01(8.50E+01±2.87E-01)	[150 9]
	ELM-har	5.40E-01(5.30E-01±5.26E-03)	5.46E-01(5.40E-01±4.91E-03)	8.36E+01(8.33E+01±1.78E-01)	8.39E+01(8.37E+01±1.62E-01)	2200
	ELM-rad	5.49E-01(5.42E-01±4.11E-03)	5.47E-01(5.53E-01±5.28E-03)	8.40E+01(8.37E+01±1.53E-01)	8.39E+01(8.41E+01±1.86E-01)	2000
	ELM-sig	5.68E-01(5.62E-01±3.53E-03)	5.70E-01(5.71E-01±4.63E-03)	8.46E+01(8.44E+01±1.26E-01)	8.47E+01(8.47E+01±1.68E-01)	1600
	ELM-sin	5.67E-01(5.55E-01±4.36E-03)	5.73E-01(5.65E-01±3.62E-03)	8.45E+01(8.41E+01±1.55E-01)	8.48E+01(8.45E+01±1.36E-01)	2000
	ELM-tri	5.42E-01(5.28E-01±5.82E-03)	5.38E-01(5.40E-01±5.15E-03)	8.36E+01(8.32E+01±2.02E-01)	8.36E+01(8.37E+01±1.77E-01)	2000
DRD	MRBF	**5.63E-01**(**5.06E-01**±3.20E-02)	**4.71E-01**(**4.74E-01**±4.40E-02)	**7.70E+01**(**7.47E+01**±8.20E-01)	**7.33E+01**(**7.33E+01**±1.64E+00)	[10 0.56]
	MLP	4.90E-01(4.47E-01±2.54E-02)	4.70E-01(4.20E-01±2.56E-02)	7.39E+01(7.17E+01±1.26E+00)	7.31E+01(7.10E+01±1.24E+00)	[31 23]
	SVM	5.20E-01(5.20E-01±0.00E+00)	4.36E-01(4.36E-01±0.00E+00)	7.49E+01(7.46E+01±0.00E+00)	7.13E+01(7.13E+01±0.00E+00)	[222.861 7.73]
	CART	3.02E-01(2.99E-01±1.23E-02)	2.61E-01(2.59E-01±8.34E-03)	6.38E+01(6.37E+01±3.65E-01)	6.26E+01(6.26E+01±2.79E-01)	[200 18]
	ELM-har	3.70E-01(2.74E-01±5.33E-02)	1.54E-01(2.02E-01±5.61E-02)	6.79E+01(6.32E+01±2.62E+00)	5.74E+01(6.01E+01±2.81E+00)	140
	ELM-rad	4.43E-01(3.90E-01±3.27E-02)	3.33E-01(3.06E-01±3.36E-02)	7.18E+01(6.87E+01±1.66E+00)	6.64E+01(6.50E+01±1.69E+00)	140
	ELM-sig	5.08E-01(4.36E-01±3.22E-02)	4.05E-01(3.95E-01±2.99E-02)	7.42E+01(7.09E+01±1.61E+00)	6.96E+01(6.93E+01±1.50E+00)	90
	ELM-sin	5.10E-01(4.27E-01±3.21E-02)	3.78E-01(4.00E-01±3.54E-02)	7.49E+01(7.06E+01±1.66E+00)	6.89E+01(6.97E+01±1.77E+00)	80
	ELM-tri	4.35E-01(3.52E-01±5.06E-02)	2.96E-01(2.58E-01±4.70E-02)	7.04E+01(6.67E+01±2.41E+00)	6.47E+01(6.27E+01±2.38E+00)	180

Απεικονίζεται το καλύτερο αποτέλεσμα που βρέθηκε σύμφωνα με τον δείκτη RMSE στο σύνολο αξιολόγησης και ακολουθείται από τον μέσο όρο και την τυπική απόκλιση των 30 αποτελεσμάτων μέσα σε παρένθεση.

[α] Οι παράμετροι περιέχουν για τα MRBF: [ασαφής διαμέριση κατώφλι δ], για τα MLPs: [νευρώνες 1ης στοιβάδας νευρώνες 2ης στοιβάδας], για τα SVMs: [συντελεστής ποινής C εύρος γκαουσιανών πυρήνων σ], για τα ELMs: αριθμός πυρήνων.

Πίνακας 4.6. Αποτελέσματα για τα σύνολα δεδομένων ταξινόμησης – MCC και ακρίβεια % *(συνέχεια)*

Σύνολο δεδομένων	**Μέθοδος**	**MCC συνόλου αξιολόγησης**	**MCC συνόλου ελέγχου**	**Ακρίβεια% συνόλου αξιολόγησης**	**Ακρίβεια% συνόλου ελέγχου**	**Παράμετροι[α]**
HEA	MRBF	**7.59E-01**(**6.88E-01**±4.62E-02)	**7.44E-01**(**7.65E-01**±3.57E-02)	**8.81E+01**(**8.46E+01**±2.26E+00)	**8.70E+01**(**8.81E+01**±1.76E+00)	[5 0.56]
	MLP	6.74E-01(5.88E-01±4.48E-02)	7.06E-01(6.35E-01±5.81E-02)	8.36E+01(7.94E+01±2.19E+00)	8.55E+01(8.21E+01±2.78E+00)	[34 21]
	SVM	6.65E-01(6.65E-01±0.00E+00)	6.82E-01(6.82E-01±0.00E+00)	8.21E+01(8.21E+01±0.00E+00)	8.41E+01(8.41E+01±0.00E+00)	[0.33 1.682]
	CART	7.28E-01(5.59E-01±1.06E-01)	5.85E-01(5.19E-01±8.82E-02)	8.66E+01(7.77E+01±5.13E+00)	7.97E+01(7.62E+01±4.48E+00)	[20 4]
	ELM-har	7.58E-01(5.98E-01±6.24E-02)	7.34E-01(6.60E-01±7.83E-02)	8.80E+01(7.99E+01±3.08E+00)	8.69E+01(8.32E+01±3.76E+00)	890
	ELM-rad	7.34E-01(5.74E-01±7.44E-02)	5.01E-01(5.60E-01±1.04E-01)	8.66E+01(7.87E+01±3.74E+00)	7.54E+01(7.84E+01±5.01E+00)	40
	ELM-sig	7.58E-01(6.42E-01±5.09E-02)	6.45E-01(6.55E-01±5.05E-02)	8.80E+01(8.22E+01±2.55E+00)	8.26E+01(8.29E+01±2.51E+00)	20
	ELM-sin	7.58E-01(6.05E-01±6.60E-02)	5.92E-01(5.48E-01±9.41E-02)	8.80E+01(8.02E+01±3.22E+00)	7.97E+01(7.78E+01±4.54E+00)	20
	ELM-tri	6.74E-01(5.72E-01±5.44E-02)	5.85E-01(6.09E-01±6.85E-02)	8.36E+01(7.86E+01±2.64E+00)	7.97E+01(8.07E+01±3.23E+00)	1120
QSA	MRBF	**8.05E-01**(7.93E-01±1.82E-02)	6.99E-01(7.09E-01±2.59E-02)	**9.13E+01**(8.94E+01±8.02E-01)	8.68E+01(8.72E+01±1.14E+00)	[19 1]
	MLP	7.89E-01(6.65E-01±5.69E-02)	6.37E-01(5.35E-01±5.04E-02)	8.92E+01(8.32E+01±2.76E+00)	8.13E+01(7.67E+01±2.47E+00)	[20 15]
	SVM	8.06E-01(**8.02E-01**±0.00E+00)	**7.23E-01**(**7.23E-01**±0.00E+00)	9.09E+01(**9.09E+01**±0.00E+00)	**8.75E+01**(**8.75E+01**±0.00E+00)	[32 6.96]
	CART	6.72E-01(6.26E-01±2.99E-02)	6.47E-01(6.17E-01±2.24E-02)	8.56E+01(8.35E+01±1.36E+00)	8.38E+01(8.27E+01±1.08E+00)	[5 40]
	ELM-har	6.74E-01(5.60E-01±5.90E-02)	6.05E-01(5.46E-01±4.20E-02)	8.56E+01(7.99E+01±2.72E+00)	8.26E+01(7.94E+01±1.86E+00)	200
	ELM-rad	7.67E-01(6.86E-01±3.66E-02)	5.90E-01(6.27E-01±3.65E-02)	8.97E+01(8.58E+01±1.69E+00)	8.19E+01(8.33E+01±1.67E+00)	120
	ELM-sig	7.77E-01(7.08E-01±2.44E-02)	6.79E-01(6.38E-01±2.81E-02)	9.01E+01(8.69E+01±1.12E+00)	8.57E+01(8.41E+01±1.21E+00)	90
	ELM-sin	7.47E-01(7.07E-01±2.78E-02)	6.71E-01(6.59E-01±3.32E-02)	8.86E+01(8.68E+01±1.30E+00)	8.57E+01(8.50E+01±1.51E+00)	110
	ELM-tri	7.24E-01(6.31E-01±4.67E-02)	6.58E-01(6.13E-01±4.28E-02)	8.75E+01(8.31E+01±2.21E+00)	8.49E+01(8.25E+01±1.91E+00)	160

Απεικονίζεται το καλύτερο αποτέλεσμα που βρέθηκε σύμφωνα με τον δείκτη RMSE στο σύνολο αξιολόγησης και ακολουθείται από τον μέσο όρο και την τυπική απόκλιση των 30 αποτελεσμάτων μέσα σε παρένθεση.

[α] Οι παράμετροι περιέχουν για τα MRBF: [ασαφής διαμέριση κατώφλι δ], για τα MLPs: [νευρώνες 1ης στοιβάδας νευρώνες 2ης στοιβάδας], για τα SVMs: [συντελεστής ποινής *C* εύρος γκαουσιανών πυρήνων *σ*], για τα ELMs: αριθμός πυρήνων

ΚΕΦΑΛΑΙΟ 5

ΕΦΑΡΜΟΓΕΣ ΣΤΗΝ ΕΞΕΛΙΞΗ ΤΟΥ ΠΑΓΚΟΣΜΙΟΥ ΙΣΤΟΥ

Αρχικά, στο παρόν κεφάλαιο γίνεται μία σύντομη εισαγωγή στον παγκόσμιο ιστό και στην εξέλιξή του σε σημασιολογικό ιστό. Στόχος αυτού του κεφαλαίου είναι η παρουσίαση ενός νέου ευφυούς συστήματος με την ικανότητα αυτόματης ταξινόμησης ιστοσελίδων με βάση το περιεχόμενό τους, το οποίο υιοθετεί τα νευρωνικά δίκτυα MRBF και στοιχεία από τον κλάδο της εξόρυξης γνώσης στον ιστό. Το προτεινόμενο σύστημα ονομάστηκε MRBF-WPC είναι ένα ολοκληρωμένο εργαλείο γενικής χρήσης για την ταξινόμηση ιστοσελίδων, το οποίο μπορεί να χρησιμοποιηθεί ως βάση για τον εμπλουτισμό του περιεχομένου του παγκόσμιου ιστού με μεταδεδομένα, με αποτέλεσμα την εξέλιξή του σε σημασιολογικό ιστό.

5.1 Εισαγωγή στον παγκόσμιο ιστό

Τα τελευταία χρόνια, το διαδίκτυο (internet) και ο παγκόσμιο ιστός (Word Wide Web, WWW) αποτελούν αναπόσπαστο κομμάτι του τρόπου ζωής των ανθρώπων, καθώς χρησιμοποιούνται ευρέως για την αναζήτηση πληροφοριών, την επικοινωνία, τη μάθηση, το ηλεκτρονικό εμπόριο κ.α. [250]. Εδώ θα πρέπει να σημειωθεί ότι, συχνά, οι έννοιες του διαδικτύου και του παγκόσμιου ιστού συγχέονται.

Το διαδίκτυο αποτελείται από ένα μεγάλο αριθμό διασυνδεδεμένων υπολογιστών, σχεδόν σε όλο τον πλανήτη Γη, για να επικοινωνούν μεταξύ τους. Το διαδίκτυο δημιουργήθηκε για πρώτη φορά το 1969 από το Υπουργείο Άμυνας των Ηνωμένων Πολιτειών της Αμερικής ως στρατιωτικό πρόγραμμα, με την ονομασία ARPANET.

Ο παγκόσμιος ιστός αποτελεί υπηρεσία του διαδικτύου και είναι ένα σύστημα διασυνδεδεμένων εγγράφων που εξυπηρετεί στην επικοινωνία των υπολογιστών και των ανθρώπων. Ουσιαστικά, ο παγκόσμιος ιστός είναι ένας τρόπος προβολής όλων των πληροφοριών που είναι διαθέσιμες στο διαδίκτυο. Ένα λογισμικό πλοήγησης, ή αλλιώς φυλλομετρητής (browser), δίνει τη δυνατότητα στους χρήστες να δουν το περιεχόμενο των ιστοσελίδων (web pages), όπως κείμενο, εικόνες, βίντεο και άλλα πολυμέσα και να πλοηγηθούν ανάμεσά τους μέσω των υπερσυνδέσμων (hyperlinks).

Οι ραγδαίες εξελίξεις στον παγκόσμιο ιστό, έχουν ως αποτέλεσμα τη μεγάλη αύξηση της διαθέσιμης πληροφορίας. Καθημερινά, προστίθενται εκατομμύρια νέες σελίδες, οι οποίες αυξάνουν τον όγκο της διαθέσιμης πληροφορίας ακόμα περισσότερο. Αυτό συνεπάγεται την ολοένα και πιο δύσκολη διαχείριση της πληροφορίας. Επιπλέον, το γεγονός ότι το μεγαλύτερο μέρος του περιεχομένου στον παγκόσμιο ιστό είναι κατανοητό από τους ανθρώπους, αλλά δεν είναι προσπελάσιμο από τις μηχανές, κάνει ακόμη πιο δύσκολη τη διαχείριση της πληροφορίας. Είναι φανερό, λοιπόν, ότι θα πρέπει να βρεθούν τεχνικές που θα μπορούν να γεφυρώσουν το χάσμα μεταξύ της γλώσσας που αντιλαμβάνεται ο άνθρωπος και της γλώσσας που καταλαβαίνουν οι μηχανές.

Μια προσπάθεια δημιουργίας κατάλληλων υποδομών για την αξιοποίηση των διαθέσιμων πληροφοριών από τους υπολογιστές, αποτελεί ο σημασιολογικός ιστός (Semantic Web), ο οποίος δεν αντικαθιστά τον τωρινό παγκόσμιο ιστό, αλλά τον εξελίσσει, δίνοντας δομή και νόημα στην πληροφορία για να είναι κατανοητή τόσο από τους ανθρώπους, όσο και από τους υπολογιστές. Οι ικανότητες του ιστού με την πάροδο των χρόνων εξελίσσονται και αναφέρονται συνήθως σε τρία στάδια: τον ιστό 1.0 (Web 1.0), τον ιστό 2.0 (Web 2.0) και τον ιστό 3.0 (Web 3.0). Στο επόμενο κεφάλαιο γίνεται μία ιστορική αναδρομή από την αρχή δημιουργίας του ιστού.

5.1.1 Ιστορική αναδρομή

Ο παγκόσμιος ιστός επινοήθηκε από τον Tim Burners-Lee [251] το 1989. Αρχικά, η ιδέα του προοριζόταν στη δημιουργία ενός πιο αποτελεσματικού συστήματος επικοινωνίας του Ευρωπαϊκού οργανισμού πυρηνικών Ερευνών (CERN) στην Ελβετία. Τελικά, ο Tim Berners-Lee συνειδητοποίησε ότι η ιδέα αυτή θα μπορούσε να εφαρμοστεί σε ολόκληρο τον κόσμο. Μαζί με τον Robert Cailliau πρότειναν το 1990 να χρησιμοποιηθεί το hypertext «ώστε διάφορα είδη πληροφοριών να είναι προσβάσιμα σαν ένα πλέγμα κόμβων όπου ο χρήστης μπορεί να περιηγηθεί σε αυτά». Έτσι, σχεδιάστηκε και δοκιμάστηκε η πρώτη υπηρεσία ιστού.

Η πρώτη γενιά του ιστού είναι ο ιστός 1.0 (Web 1.0), ο οποίος, σύμφωνα με τον Berners-Lee, θα μπορούσε να θεωρηθεί ως «διαδικτυακός τόπος μόνο για ανάγνωση» [252].

Ο πρώτος ιστός επέτρεπε στους χρήστες να αναζητούν πληροφορίες και να διαβάζουν τα αντίστοιχα έγγραφα. Η αλληλεπίδραση των χρηστών του διαδικτύου με τις ιστοσελίδες, ή η συνεισφορά περιεχομένου ήταν σχεδόν ελλιπείς. Η πρώτη γενιά του ιστού αναφέρεται στην εποχή των στατικών ιστοσελίδων και ο στόχος ήταν η δημιουργία μιας ηλεκτρονικής πλατφόρμας, ώστε να διατίθενται οι επιθυμητές πληροφορίες. Οι τεχνολογίες του ιστού 1.0 περιλαμβάνουν τρία βασικά πρωτόκολλα: α) το HTTP (HyperText Transfer Protocol) για τη διακίνηση των εγγράφων στον ιστό, β) τη γλώσσα σήμανσης (markup) HTML (HyperText Markup Language) για τον καθορισμό της δομής, του περιεχομένου και της παρουσίασης ενός εγγράφου, και γ) τα URIs (Uniform Resource Identifiers) για τον προσδιορισμό ονομάτων και διευθύνσεων πόρων στο διαδίκτυο [252]. Με την έννοια πόρος εννοείται οποιαδήποτε οντότητα του ιστού, όπως είναι μια ιστοσελίδα, ηλεκτρονικά αρχεία, ή ακόμα και ένα βιβλίο που δεν είναι άμεσα διαθέσιμο στο διαδίκτυο

Η έλλειψη αλληλεπίδρασης των χρηστών με τον ιστό 1.0 οδήγησε στη δημιουργία του ιστού δεύτερης γενιάς. Οι Tim O'Reilly και Dale Dougherty το 2004, στο συνέδριο που οργάνωσε η O'Reilly Media με θέμα την αναβάθμιση του παγκόσμιου ιστού, όρισαν τον ιστό δεύτερης γενιάς ως ιστό 2.0 (Web 2.0) [253] και ο Tim O'Reilly αναφέρθηκε στον ιστό 2.0 ως «την επιχειρηματική επανάσταση στη βιομηχανία των υπολογιστών, που προήλθε λόγω της θεώρησης του διαδικτύου ως πλατφόρμα και από την προσπάθεια κατανόησης των κανόνων που οδήγησαν στην επιτυχία της νέας αυτής πλατφόρμας». Ο ιστός 2.0 είναι γνωστός και ως ιστός «ανάγνωσης-εγγραφής» [252] και ο βασικός του στόχος ήταν η αλληλεπίδραση των χρηστών. Η νέα δυνατότητα του ιστού 2.0, για εισαγωγή περιεχομένου και αλληλεπίδραση με άλλους χρήστες, άλλαξε δραματικά το τοπίο του ιστού σε σύντομο χρονικό διάστημα. Χώροι όπως της δημοσιογραφίας, της πολιτικής και των επιχειρήσεων άλλαξαν, ενώ ταυτόχρονα οι άνθρωποι που κάνουν χρήση του παγκόσμιου ιστού δεν περιορίζονται πια στο να είναι ανεξάρτητοι αναγνώστες ή καταναλωτές πληροφοριών, αλλά έχουν τη δυνατότητα να συμβάλουν οι ίδιοι στην πληροφόρηση. Ο δεύτερος ιστός φαίνεται ότι αποτέλεσε μια ευπρόσδεκτη απάντηση σε μια απαίτηση των χρηστών του διαδικτύου να εμπλακούν περισσότερο σε σχέση με τις πληροφορίες που έχουν στη διάθεσή τους, όπως για παράδειγμα οι ιστότοποι κοινωνικής δικτύωσης (Twitter, YouTube, Facebook, κ.α).

Ο ιστός 2.0 περιέχει έναν τεράστιο όγκο δεδομένων σε φυσική γλώσσα, βάσεις δεδομένων, υπηρεσίες διαδικτύου (web services), κ.α. Αυτά τα δεδομένα μπορούν να αξιοποιηθούν από τους ανθρώπους, αλλά δεν είναι εφικτή η χρήση τους από τους υπολογιστές. Ο ιστός 2.0 μειονεκτεί όσον αφορά στην ανάλυση πληροφοριών, γι' αυτό τα τελευταία χρόνια γίνεται μια σημαντική προσπάθεια για την αυτόματη περιγραφή και επεξεργασία των δεδομένων που υπάρχουν στο διαδίκτυο. Η λύση σε αυτό το πρόβλημα είναι η εξέλιξη του δεύτερου ιστού στον ιστό 3.0 (Web 3.0), ο οποίος επεκτείνοντας τις αναφορές του Tim Berners-Lee, αντιστοιχεί σε έναν ιστό «ανάγνωσης-εγγραφής-εκτέλεσης» [254].

5.1.2 Το όραμα του σημασιολογικού ιστού

Ο Tim Berners-Lee όρισε τον ιστό τρίτης γενιάς ως σημασιολογικό ιστό (Semantic Web) [255] και τον οραματίστηκε ως εξής:

> *Ο σημασιολογικός ιστός θα φέρει νόημα στο περιεχόμενο των ιστοσελίδων, δημιουργώντας ένα περιβάλλον όπου οι πράκτορες λογισμικού που περιπλανώνται από σελίδα σε σελίδα εκτελούν προηγμένες διεργασίες για τους χρήστες [112].*

Η βασική ιδέα του σημασιολογικού ιστού είναι το περιεχόμενό του να είναι κατανοητό από τους υπολογιστές με σκοπό την αποτελεσματικότερη εύρεση, αυτοματοποίηση, ενσωμάτωση και επαναχρησιμοποίηση του σε διάφορες εφαρμογές. Ο σημασιολογικός ιστός δεν θα καταργήσει τον τωρινό ιστό, αλλά θα τον εξελίξει διορθώνοντας τις ατέλειές του, επιτρέποντας στους χρήστες να βρίσκουν, να μοιράζονται και να επεξεργάζονται πιο εύκολα τις διαθέσιμες πληροφορίες. Ο σημασιολογικός ιστός αποτελεί μια συλλογική προσπάθεια του World Wide Web Consortium (W3C) στην οποία συμμετέχει ένας μεγάλος αριθμός ερευνητών και συνεργατών.

Δύο βασικές έννοιες που παίζουν καθοριστικό ρόλο στην ανάπτυξη του σημασιολογικού ιστού είναι τα μεταδεδομένα και οι οντολογίες. Σύμφωνα με τον Tim Berner Lee, τα μεταδεδομένα κάνουν σαφή την πληροφορία που είναι αόριστη. Ουσιαστικά, τα μεταδεδομένα έχουν ως στόχο την περιγραφή πληροφοριών ή πόρων (resources). Τα μεταδεδομένα ορίζουν τις έννοιες των δεδομένων και τις συσχετίσεις μεταξύ τους, δηλαδή προσδίδουν μια σημασία στα δεδομένα, τα οποία επίσης ονομάζονται σημασιολογικά δεδομένα. Τα δεδομένα που περιγράφονται από τα μεταδεδομένα, μπορεί να έχουν δομημένη, αδόμητη ή ημιδομημένη μορφή. Τα δομημένα δεδομένα είναι ουσιαστικά "οργανωμένα δεδομένα", αφορούν δηλαδή σε δεδομένα που μπορούν να διαχωριστούν σε επιμέρους δεδομένα και να οργανωθούν σε ιεραρχικές δομές, όπως είναι για παράδειγμα τα δεδομένα πελατών μιας τράπεζας. Τα αδόμητα, ή μη δομημένα δεδομένα, αφορούν σε δεδομένα που δεν έχουν εγγενή δομή, όπως είναι για παράδειγμα ένα κείμενο. Τα ημιδομημένα δεδομένα βρίσκονται ανάμεσα στα δομημένα και στα μη δομημένα δεδομένα, δηλαδή διαθέτουν κάποιου είδους δομή, όπως είναι για παράδειγμα κάποιες ιστοσελίδες.

Οι μηχανές, χρησιμοποιώντας μεταδεδομένα γνωρίζουν τη σημασιολογία της πληροφορίας με αποτέλεσμα να μπορούν να την επεξεργάζονται με αποδοτικό τρόπο. Η οργάνωση των μεταδεδομένων οδηγεί στη δημιουργία οντολογιών. Ο όρος οντολογία προέρχεται από την επιστήμη της φιλοσοφίας, η οποία μελετά την ύπαρξη. Στη βιβλιογραφία υπάρχουν διάφοροι ορισμοί για την έννοια της οντολογίας. Ο Thomas R. Gruber ορίζει την οντολογία ως εξής:

> *Μια τυπική, σαφής προδιαγραφή μιας διαμοιρασμένης εννοιολογικής αναπαράστασης [256].*

Σύμφωνα με την W3C η οντολογία ορίζεται ως εξής:

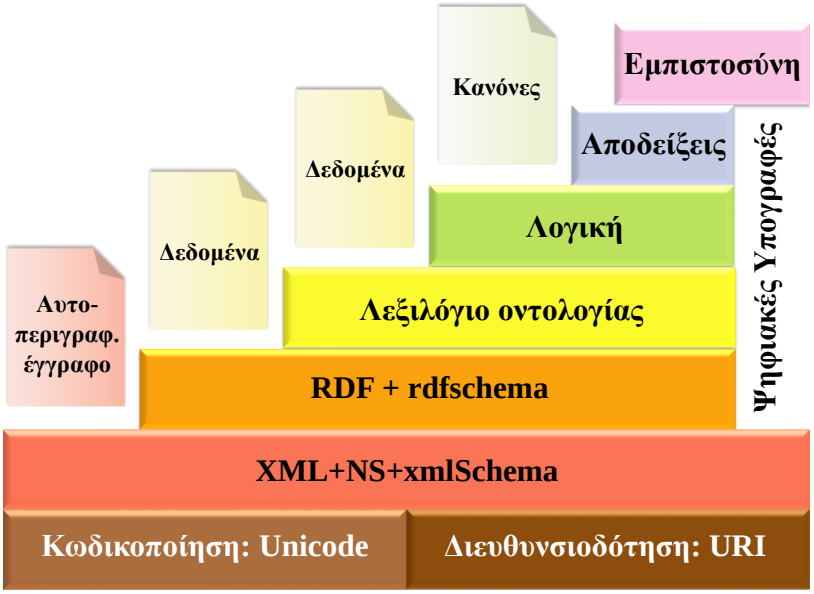
> *Μια οντολογία καθορίζει εκείνους τους όρους που χρησιμοποιούνται για την περιγραφή και την αναπαράσταση ενός πεδίου γνώσης. Περιλαμβάνει ορισμούς τόσο των βασικών εννοιών του πεδίου γνώσης όσο και των μεταξύ τους σχέσεων με τρόπο μηχανικά αναγνώσιμο. Η οντολογία κωδικοποιεί γνώση για ένα πεδίο όπως επίσης και γνώση που συνδέει πεδία. Κατά αυτό τον τρόπο, παράγει γνώση επαναχρησιμοποιήσιμη [257].*

5.1.3 Τα επίπεδα του σημασιολογικού ιστού

Ο Tim Berners-Lee πρότεινε μια πολυεπίπεδη αρχιτεκτονική για τον σημασιολογικό ιστό, που συχνά αντιπροσώπευε χρησιμοποιώντας ένα διάγραμμα, με πολλές παραλλαγές από τότε. Σύμφωνα με τον Berners-Lee, η αρχιτεκτονική του σημασιολογικού ιστού αποτελείται από μια σειρά από επίπεδα τεχνολογιών και γλωσσών προγραμματισμού, η οποία ονομάζεται στοίβα του σημασιολογικού ιστού (Semantic Web layercake / Semantic Web Stack). Η

τελευταία παρουσιάζεται στο σχήμα 5.1 [258] και παρακάτω δίνεται μια σύντομη περιγραφή για το κάθε επίπεδο του σημασιολογικού ιστού:

- **Επίπεδο κωδικοποίησης Unicode και διευθυνσιοδότησης URI**: Αυτό το επίπεδο εμπεριέχει την τωρινή μορφή του παγκόσμιου ιστού, όπου τα δεδομένα βρίσκονται σε φυσικό και λογικό επίπεδο. Στο φυσικό επίπεδο τα έγγραφα κωδικοποιούνται με βάση το πρότυπο Unicode, το οποίο είναι το σύστημα κωδικοποίησης χαρακτήρων που βασίζονται όλα τα σύγχρονα πρωτόκολλα της πληροφορικής και των τηλεπικοινωνιών. Επίσης, το επίπεδο αυτό, εμπεριέχει και το αναγνωριστικό URI που αποτελεί έναν μηχανισμό διευθυνσιοδότησης στον παγκόσμιο και στον σημασιολογικό ιστό. Ουσιαστικά, το URI περιέχει μια σειρά χαρακτήρων και χρησιμοποιείται για να ταυτοποιήσει με μοναδικό τρόπο ένα διαδικτυακό πόρο.
- **Επίπεδο XML**: Το επίπεδο αυτό αφορά στη λογική αναπαράσταση των εγγράφων του σημασιολογικού ιστού βάσει του προτύπου XML (eXtensible Markup Language), το οποίο είναι μία γλώσσα σήμανσης που επιτρέπει την εγγραφή δομημένων εγγράφων ιστού με χρήση του λεξιλογίου που έχει οριστεί από τον χρήστη. Σε αυτό το επίπεδο περιλαμβάνεται και η XML Schema, η οποία είναι μια περιγραφική γλώσσα για τον ορισμό της δομής των εγγράφων XML.
- **Επίπεδο RDF**: Αυτό το επίπεδο αφορά στο βασικό μοντέλο δηλώσεων του σημασιολογικού ιστού που αποτελεί το πρότυπο Resource Description Framework (RDF), το οποίο συμβάλει στην κατανόηση των δεδομένων από τους υπολογιστές, αναπαριστώντας το νόημα και περιγράφοντας τα μεταδεδομένα των πόρων. Ένα μοντέλο RDF περιγράφει τη δομημένη πληροφορία και εμπεριέχει τους πόρους, οι οποίοι είναι αντικείμενα προσπελάσιμα, ή μη, από τον παγκόσμιο ιστό και τα οποία προσδιορίζονται με ένα μοναδικό URI. Το βασικό δομικό στοιχείο ενός μοντέλου RDF αποτελεί η δήλωση (statement), ή τριπλέτα (triple), η οποία περιγράφει τη συσχέτιση μεταξύ δύο πόρων. Ένας πόρος περιέχει ιδιότητες (properties), οι οποίες αφορούν σε γνωρίσματά του, ή συσχετίσεις με άλλους πόρους, και αυτές οι ιδιότητες αναπαρίστανται με τριπλέτες



Σχήμα 5.1. Αρχιτεκτονική του σημασιολογικού ιστού σύμφωνα με τον Tim Berners-Lee [258].

της μορφής <υποκείμενο (subject), κατηγόρημα (predicate), αντικείμενο (object)>. Το υποκείμενο αναφέρεται σε έναν πόρο, το αντικείμενο αφορά στον πόρο με τον οποίο συσχετίζεται ο πόρος-υποκείμενο και το κατηγόρημα αφορά στο είδος της συσχέτισης των δύο πόρων. Σε αυτό το επίπεδο περιλαμβάνεται και η RDF Schema, η οποία βασίζεται στην RDF και παρέχει θεμελιώδη στοιχεία μοντελοποίησης με σκοπό την οργάνωση των αντικειμένων του ιστού σε ιεραρχίες.

- **Επίπεδο Οντολογίας (Ontology Layer)**: Στο επίπεδο αυτό ορίζεται το λεξιλόγιο των οντολογιών και οι διαδικτυακές γλώσσες ορισμού τους. Οι οντολογίες μπορούν να εγγραφούν κυρίως με τις γλώσσες OWL και RDFS.
- **Επίπεδο Λογικής (Logic Layer)**: Αυτό το επίπεδο χρησιμοποιεί λογικούς κανόνες και τις συσχετίσεις που υπάρχουν μεταξύ των δεδομένων με σκοπό την εξαγωγή συμπερασμάτων. Έτσι, οι υπολογιστές έχουν τη δυνατότητα λήψης αποφάσεων. Ουσιαστικά, αυτό το επίπεδο χρησιμοποιείται για να ενισχύσει περαιτέρω τις γλώσσες οντολογιών.
- **Επίπεδο Αποδείξεων (Proof Layer)**: Αυτό το επίπεδο δεν έχει αναπτυχθεί πλήρως και αφορά στην παροχή αποδείξεων σχετικά με την προέλευση των πληροφοριών.
- **Επίπεδο Εμπιστοσύνης (Trust Layer)**: Αυτό το επίπεδο, επίσης, δεν έχει αναπτυχθεί πλήρως, ενώ προκύπτει μέσω της χρήσης ψηφιακών υπογραφών (digital signatures) και άλλων ειδών γνώσης βάσει συστάσεων (recommendations) από αξιόπιστους πράκτορες, ή από οργανισμούς αξιολόγησης και πιστοποίησης και οργανισμούς καταναλωτών.

5.2 Εισαγωγή στην εξέλιξη του παγκόσμιου ιστού

Ο παγκόσμιος ιστός αποτελεί ένα τεράστιο επίτευγμα, το οποίο έχει διαπεράσει τις περισσότερες πλευρές της καθημερινής ζωής των ανθρώπων. Αυτό οφείλεται κυρίως στην απλότητα χρήσης που παρουσιάζει. Ο ιστός στη σημερινή του μορφή είναι προσιτός στους ανθρώπους μέσω των μηχανών αναζήτησης και των φυλλομετρητών, οι υπολογιστές όμως δεν μπορούν να κατανοήσουν το περιεχόμενο του παγκόσμιου ιστού. Ο άναρχος τρόπος εισαγωγής των ιστοσελίδων και η παντελής έλλειψη δομής και οργάνωσής τους έχει δημιουργήσει ένα χαοτικό σύνολο όπου η πρόσβαση στην πληροφορία είναι ιδιαίτερα δύσκολη για τους υπολογιστές. Η περιγραφή και οργάνωση αυτού του τεράστιου περιεχομένου είναι απαραίτητη για την αξιοποίηση του πλήρους δυναμικού του ως πηγή πληροφοριών.

Στη σύγχρονη βιβλιογραφία παρουσιάζεται έντονη ερευνητική δραστηριότητα ως προς την εξέλιξη του παγκόσμιου ιστού. Η επέκταση του παγκόσμιου ιστού είναι ο σημασιολογικός ιστός, όπου η πληροφορία είναι καλά ορισμένη δίνοντας τη δυνατότητα στους υπολογιστές να κατανοούν και να αξιοποιούν τις πληροφορίες [112]. Ουσιαστικά, ο σημασιολογικός ιστός αποτελεί μία ερευνητική προσπάθεια σε εξέλιξη, όπου οι βασικές προδιαγραφές έχουν δοθεί, αλλά ο τρόπος περιγραφής των πηγών, και οι υποστηριζόμενες υπηρεσίες βρίσκονται ακόμη υπό ανάπτυξη. Η υλοποίηση του σημασιολογικού ιστού εξαρτάται σε μεγάλο βαθμό, αρχικά από τον εμπλουτισμό της υπάρχουσας πληροφορίας με μεταδεδομένα και έπειτα από τις διαθέσιμες οντολογίες, επιτρέποντας στις μηχανές να "κατανοούν" τις διαθέσιμες πληροφορίες. Ωστόσο, ο μη αυτόματος σχολιασμός και η χειρωνακτική ανάπτυξη των οντολογιών είναι δύο γνωστά προβλήματα που πρέπει να

επιλυθούν προκειμένου ο παγκόσμιος ιστός να εξελιχθεί και να τύχει ευρείας αποδοχής από τους τωρινούς χρήστες.

Παρόλο που υπάρχουν διάφορα πρότυπα μεταδεδομένων στο μεγαλύτερο μέρος της «ζωής» του ιστού, οι άνθρωποι συνήθως δεν επιθυμούν να ξοδεύουν χρόνο προσθέτοντας δομημένα δεδομένα στο περιεχόμενο του ιστού, διότι είναι χρονοβόρα (και συχνά βαρετή) διαδικασία. Επίσης, οι περισσότεροι δημιουργοί περιεχομένου στον ιστό δεν γνωρίζουν πλήρως τα μεταδεδομένα και τη σημασία τους στον παγκόσμιο ιστό. Ακόμη και αν τη γνώριζαν, βέβαια, το κίνητρό τους για την προσθήκη δεδομένων που ίσως δεν είναι άμεσο όφελος για τον εαυτό τους, πιθανότατα δεν θα ήταν αρκετό.

Ο σημασιολογικός ιστός δεν ορίζει τον τρόπο με τον οποίο παράγονται ή καταχωρούνται οι σχολιασμοί. Η πιο τυπική προσέγγιση είναι αυτή της «εγγραφής με το χέρι», δηλαδή του μη αυτόματου σχολιασμού. Αυτός ο τρόπος, αν γίνει σωστά, οδηγεί σε αποτελέσματα υψηλής ποιότητας, ωστόσο είναι μια αργή και κουραστική διαδικασία. Είναι φανερό ότι οι σχολιασμοί που δημιουργούνται με αυτόν τον τρόπο περιορίζονται πάντοτε σε ένα μικρό μέρος του παγκόσμιου ιστού, επειδή πολλοί δημιουργοί ιστοσελίδων δεν καταβάλουν τις πρόσθετες προσπάθειες που είναι αναγκαίες για την καταχώρηση των σχολιασμών. Επίσης, ακόμα και όταν οι δημιουργοί των ιστοσελίδων καταχωρήσουν σχολιασμούς, συχνά δεν θα είναι ενημερωμένοι, διότι οι περισσότεροι δημιουργοί ιστοσελίδων «ξεχνούν» να τους ενημερώνουν όταν αλλάζει ένα έγγραφο. Ως εκ τούτου, οι αυτόματες ή ημιαυτόματες τεχνικές αποτελούν ελκυστική εναλλακτική λύση για τη δημιουργία σχολιασμού, με απώτερο σκοπό την επέκταση του παγκόσμιου ιστού.

Ο αυτοματοποιημένος σχολιασμός του περιεχομένου του ιστού είναι ένα σημαντικό βήμα στην προσπάθεια εξέλιξης του παγκόσμιου ιστού. Προς το παρόν, όμως, η διαθέσιμη τεχνολογία για την παροχή αυτόματου σχολιασμού δεν είναι ακόμα ώριμη. Δεδομένου ότι ο ιστός περιλαμβάνει πολλά διαφορετικά πεδία, είναι πολύ δύσκολο το έργο της επισήμανσης ολόκληρου του ιστού ταυτόχρονα.

Οι Satti et.al. [117] πρότειναν να διαχωριστεί το έργο του σχολιασμού ολόκληρου του ιστού σε κομμάτια που αντιπροσωπεύουν διαφορετικά πεδία. Με αυτόν τον τρόπο, το πρόβλημα του σχολιασμού ολόκληρου του ιστού μπορεί να αντιμετωπιστεί ευκολότερα. Το πρώτο βήμα για τον αυτόματο σχολιασμό στον ιστό είναι η ταξινόμηση αντικειμένων ιστού, συνήθως ιστοσελίδων [117].

5.2.1 Βασικές έννοιες

Οι ραγδαίες πρόοδοι στους τομείς των υπολογιστικών συστημάτων, του διαδικτύου και του παγκόσμιου ιστού δημιούργησαν έναν τεράστιο όγκο πληροφορίας, που στο μεγαλύτερο μέρος της είναι μη δομημένη και καθημερινά αυξάνεται συνεχώς. Η εξόρυξη δεδομένων (Data Mining) είναι η επιστήμη της εξαγωγής χρήσιμης πληροφορίας από σύνολα ή βάσεις δεδομένων μεγάλου μεγέθους [259]. Σε αντίθεση με τα δεδομένα που περιέχουν οι βάσεις δεδομένων, το περιεχόμενο των κειμένων είναι αδόμητο, με αποτέλεσμα η επεξεργασία τους να καθίσταται ένα ιδιαίτερα δύσκολο έργο για τους αλγορίθμους. Δυστυχώς, σε πολλές εφαρμογές, οι ηλεκτρονικές πληροφορίες είναι διαθέσιμες μόνο με τη μορφή κειμένων φυσικής γλώσσας και όχι με τη μορφή δομημένων βάσεων δεδομένων.

Για να μπορέσει μια μηχανή να επεξεργαστεί ένα κείμενο, απαιτούνται εξειδικευμένα εργαλεία με πιο ισχυρούς αλγόριθμους, σε σχέση με τις μεθόδους που χρησιμοποιούνται στα δομημένα δεδομένα. Επομένως, είναι αναγκαία η ανάπτυξη εργαλείων που να μπορούν να αναλύσουν μη δομημένα δεδομένα. Ένας μεγάλος αριθμός εργασιών στη βιβλιογραφία

ερευνά τη μοντελοποίηση τέτοιου είδους δεδομένων και με βάση αυτές τις έρευνες, αναπτύχθηκαν τεχνικές εξόρυξης κειμένου (Text Mining) [260]. Η εξόρυξη κειμένου είναι ένας νέος ερευνητικός τομέας, ο οποίος εξάγει πληροφορίες από κείμενα φυσικής γλώσσας. Πιο συγκεκριμένα, η εξόρυξη κειμένου αφορά στην ανακάλυψη χρήσιμων πληροφοριών από μη δομημένα, ή ημιδομημένα, κείμενα [261]. Με βάση αυτό, η εξόρυξη κειμένου είναι επίσης γνωστή ως εξόρυξη δεδομένων κειμένου (Text Data Mining) ή ανακάλυψη γνώσης από βάσεις δεδομένων κειμένων (Knowledge Discovery in Textual databases) [262].

Βασικά στοιχεία της εξόρυξης κειμένου αποτελούν η ανάκτηση πληροφορίας (Information Retrieval) και η εξαγωγή πληροφορίας (Information Extraction). Η ανάκτηση πληροφορίας είναι η επιστήμη της αναζήτησης πληροφοριών σε ένα έγγραφο, της αναζήτησης των εγγράφων που περιέχουν την επιθυμητή πληροφορία, καθώς και της αναζήτησης των μεταδεδομένων που περιγράφουν δεδομένα. Ο στόχος της εξαγωγής πληροφοριών είναι η εύρεση των επιθυμητών πληροφοριών σε κείμενα φυσικής γλώσσας και η αποθήκευση τους σε μορφή κατάλληλη για αυτόματη αναζήτηση και επεξεργασία. Η ανάκτηση πληροφορίας και η εξαγωγή πληροφορίας ανήκουν στην επιστήμη της επεξεργασίας φυσικής γλώσσας (Natural Language Processing). Η επεξεργασία φυσικής γλώσσας είναι το πεδίο της επιστήμης των υπολογιστών που ασχολείται με την εφαρμογή γλωσσικών και στατιστικών αλγορίθμων, με σκοπό την εξαγωγή νοήματος από το κείμενο με τρόπο που είναι παρόμοιος με τον τρόπο που αντιλαμβάνεται τη γλώσσα ο ανθρώπινος εγκέφαλος. Πιο απλά, η επεξεργασία φυσικής γλώσσας είναι ένα είδος γλωσσικής ανάλυσης που βοηθά μία μηχανή να «διαβάσει» ένα κείμενο.

Ένας από τους σημαντικότερους στόχους της εξόρυξης κειμένου είναι η ταξινόμηση εγγράφων. Ένας μεγάλος αριθμός ερευνητών έχει επικεντρωθεί στην ταξινόμηση κειμένων λόγω της εκρηκτικής αύξησης του παγκόσμιου ιστού. Πιο συγκεκριμένα, η έρευνα τα τελευταία χρόνια εστιάζεται στην ταξινόμηση ιστοσελίδων. Βέβαια, η ταξινόμηση ιστοσελίδων παρουσιάζει δυσκολίες σε σχέση με την παραδοσιακή ταξινόμηση κειμένου. Μία τέτοια δυσκολία είναι η μεγάλη διακύμανση του περιεχομένου των ιστοσελίδων και της ποιότητάς τους. Επίσης, άλλη μία δυσκολία είναι ότι οι ιστοσελίδες, εκτός από κείμενο, μπορεί να περιέχουν εικόνες, ήχο, βίντεο κ.α..

Ο τομέας της εξόρυξης γνώσης από το περιεχόμενο του ιστού (web content mining) έχει ως στόχο τη λύση του προβλήματος ταξινόμησης ιστοσελίδων. Η εξόρυξη γνώσης από το περιεχόμενο του ιστού ανήκει στον κλάδο της εξόρυξης γνώσης από τον ιστό (web mining), ο οποίος αποτελεί μία ευρεία ερευνητική περιοχή που αφορά στην επεξεργασία των δεδομένων του ιστού και στην ανακάλυψη γνώσης από αυτά. Η εξόρυξη γνώσης από τον ιστό χωρίζεται σε τρεις κύριους άξονες:

- την εξόρυξη γνώσης από τα δεδομένα δομής του ιστού (web structure mining): ασχολείται με τη δομή μεταξύ των ιστοσελίδων, δηλαδή την τοπολογία τους
- την εξόρυξη γνώσης από τα δεδομένα χρήσης του ιστού (web usage mining): ασχολείται με τη συμπεριφορά των χρηστών στο διαδίκτυο
- την εξόρυξη γνώσης από το περιεχόμενο του ιστού: ασχολείται με ασχολείται με το περιεχόμενο των ιστοσελίδων.

Η τελευταία, είναι ο πιο ευρέως διερευνώμενος άξονας της εξόρυξης γνώσης από τον ιστό, και ουσιαστικά, είναι η διαδικασία εξαγωγής γνώσης από μη δομημένα ή ημιδομημένα δεδομένα. Η εξόρυξη γνώσης από το περιεχόμενο του ιστού είναι στενά συνδεδεμένη με την εξόρυξη κειμένου, καθώς το μεγαλύτερο μέρος των ιστοσελίδων αποτελείται από κείμενο, και επίσης, το κείμενο αποτελεί τον κινητήριο μοχλό στη διακίνηση της πληροφορίας.

5.2.2 Ταξινόμηση ιστοσελίδων, προεπεξεργασία και αναπαράσταση κειμένου

Η ταξινόμηση ιστοσελίδων παίζει σημαντικό ρόλο στην εξέλιξη του παγκόσμιου ιστού, αν σκεφτεί κανείς ότι οργανώνοντας το περιεχόμενο των ιστοσελίδων σε κατηγορίες και επισημειώνοντάς το με ετικέτες, γίνεται ευκολότερη η διαχείριση της διαθέσιμης γνώσης. Η ταξινόμηση ιστοσελίδων είναι η διαδικασία εκχώρησης μιας ιστοσελίδας σε μία από τις προκαθορισμένες κατηγορίες. Η ταξινόμηση του κειμενικού περιεχομένου μιας ιστοσελίδας αφορά στην αναγνώριση των σημαντικότερων συστατικών στοιχείων του κειμένου της, με κύριο στόχο την ανάθεση της ιστοσελίδας σε προκαθορισμένες κατηγορίες. Τα στάδια από τα οποία πρέπει να περάσουν οι ιστοσελίδες για να πραγματοποιηθεί η ταξινόμησή τους, συνήθως είναι τα παρακάτω [263]:

1. Συλλογή: Συλλογή πληροφορίας από τον ιστό.
2. Εξαγωγή: Εξάγεται μόνο το κειμενικό περιεχόμενο, οι ετικέτες HTML και όλα τα υπόλοιπα στοιχεία αφαιρούνται.
3. Προεπεξεργασία: Αφαίρεση περιττών στοιχείων από το κειμενικό περιεχόμενο.
4. Μετασχηματισμός: Μετατροπή πληροφορίας σε κατάλληλα επεξεργάσιμη μορφή.
5. Εξόρυξη: Εφαρμογή τεχνικών για την παραγωγή κατάλληλου μοντέλου πρόβλεψης.
6. Επικύρωση: Έλεγχος του παραγόμενου μοντέλου.

Είναι ευρέως αποδεκτό ότι ο τρόπος με τον οποίο αναπαρίσταται η κειμενική πληροφορία επηρεάζει την ποιότητα των αποτελεσμάτων που μπορούν να επιτευχθούν. Η πιο συνηθισμένη μορφή αναπαράστασης κειμένων είναι η μέθοδος του «σάκου λέξεων» (bag of words), όπου κάθε κείμενο αποτελείται από όρους που αναπαριστούν αντιπροσωπευτικές λέξεις-κλειδιά. Πιο συγκεκριμένα, το κείμενο αναπαρίσταται ως ένας «σάκος» που περιέχει τις λέξεις του κειμένου, χωρίς όμως να ασχολείται με τη γραμματική ή το συντακτικό. Υπάρχουν και άλλες πιο εξεζητημένες προσεγγίσεις αναπαράστασης, αλλά πειραματικά αποτελέσματα [264, 265] έχουν δείξει ότι δεν παρουσιάζουν σημαντικά καλύτερη απόδοση.

Για την εκτέλεση της αυτόματης ταξινόμησης ιστοσελίδων, αφού συλλεγούν οι ιστοσελίδες και εξαχθεί το κειμενικό περιεχόμενο, το τελευταίο περνά από το στάδιο της προεπεξεργασίας. Ο κύριος στόχος της προεπεξεργασίας των δεδομένων είναι η μείωση των διαστάσεων του προβλήματος ελέγχοντας το μέγεθος του λεξιλογίου (τον αριθμό των διαφορετικών όρων). Σε ορισμένες περιπτώσεις, εκτός από τη μείωση της πολυπλοκότητας του προβλήματος, αυτή η προεπεξεργασία κάνει τα δεδομένα πιο ομοιόμορφα κατά τέτοιο τρόπο που βελτιώνει την απόδοση των αλγορίθμων. Στη συνέχεια, παρουσιάζονται τα βασικότερα βήματα προεπεξεργασίας των κειμένων.

1. Αφαίρεση της δομής των κειμένων: Αφαιρείται η μορφοποίηση των κειμένων και οι χαρακτήρες γίνονται όλοι κεφαλαίοι ή πεζοί.
2. Λημματοποίηση (Tokenization): Αφαιρούνται οι αριθμοί, τα σύμβολα, τα σημεία στίξης και μένουν μόνο οι λέξεις του κειμένου.
3. Αφαίρεση των κοινών λέξεων (stopwords): Αφαιρούνται οι λέξεις που δεν προσδίδουν χρήσιμη πληροφορία και δημιουργούν θόρυβο, όπως είναι οι λέξεις που συνήθως αποτελούνται από λιγότερους από 3 χαρακτήρες, όπως επίσης, τα άρθρα, οι σύνδεσμοι των προτάσεων κ.α..

4. Στελέχωση των λέξεων (Stemming): Αφαιρείται η κατάληξη των λέξεων με σκοπό την αναγωγή των λέξεων στη ρίζα τους. Ένας από τους πιο γνωστούς αλγορίθμους για τη διαδικασία αυτή είναι ο αλγόριθμος Porter [266].

Αφού εκτελείται η διαδικασία της προεπεξεργασίας, στη συνέχεια τα διαθέσιμα κείμενα μετατρέπονται σε άμεσα επεξεργάσιμα δεδομένα μέσω μιας διαδικασίας μετασχηματισμού. Το πιο γνωστό μοντέλο αναπαράστασης κειμένων που εφαρμόζεται για αλγορίθμους κατηγοριοποίησης είναι το μοντέλο διανυσματικού χώρου (Vector Space Model) [267, 268], το οποίο προέρχεται από τα συστήματα ανάκτησης πληροφορίας. Το μοντέλο διανυσματικού χώρου αναπαριστά κάθε κείμενο με ένα διάνυσμα χαρακτηριστικών, με βάση τους όρους που εξάγονται από τα κείμενα. Το διάνυσμα χαρακτηριστικών αποτελείται από έναν αριθμό διαστάσεων, όπου αυτός ο αριθμός είναι ίσος με το πλήθος των μοναδικών όρων που εξάγονται από τα κείμενα. Κάθε διάσταση-όρος σε κάθε διάνυσμα-κείμενο έχει μία τιμή, η οποία αντιστοιχεί στην εμφάνιση του όρου μέσα στο κείμενο (0 απουσία όρου, 1 παρουσία όρου). Σε αυτή τη τιμή μπορεί, επίσης, να αποδοθεί ένα βάρος. Αυτό το βάρος μπορεί να αντιστοιχεί για παράδειγμα στον αριθμό ή στη συχνότητα εμφάνισης του εκάστοτε όρου μέσα στο κάθε κείμενο. Στη βιβλιογραφία έχουν αναφερθεί αρκετές μέθοδοι απόδοσης βάρους.

5.3 Ανάπτυξη συστήματος για την ταξινόμηση ιστοσελίδων

Σε αυτή τη διδακτορική διατριβή αναπτύχθηκε ένα ευφυές σύστημα με βάση τα δίκτυα MRBF, το οποίο μπορεί να κατατάσσει αυτόματα μία ιστοσελίδα με βάση το περιεχόμενό της και ονομάστηκε MRBF-WPC (MRBF - Web Page Classification). Συγκεκριμένα, το σύστημα αυτό, για την επίλυση του προβλήματος ταξινόμησης ιστοσελίδων, χρησιμοποιεί ένα σχετικά μικρό αριθμό χαρακτηριστικών, από αριθμητικά και κατηγορικά δεδομένα. Επίσης, το σύστημα αυτό, μπορεί να εφαρμοστεί ως γενικής χρήσης εργαλείο ταξινόμησης ιστοσελίδων, δηλαδή μπορεί να ταξινομήσει μία σελίδα με βάση έχει τη δυνατότητα να ταξινομήσει μία ιστοσελίδα

Το προτεινόμενο σύστημα μετασχηματίζει τις ιστοσελίδες σε κατάλληλα επεξεργάσιμη μορφή με βάση τα στάδια που αναφέρονται στο κεφάλαιο 5.2.2: συλλογή, εξαγωγή, προεπεξεργασία, μετασχηματισμός, εξόρυξη, αξιολόγηση. Πιο συγκεκριμένα, το σύστημα MRBF-WPC για το στάδιο της προεπεξεργασίας ακολουθεί τα βήματα που αναφέρονται στο κεφάλαιο 5.2.2, ενώ για το στάδιο του μετασχηματισμού χρησιμοποιεί το μοντέλο διανυσματικού χώρου.

Για την εφαρμογή του μοντέλου διανυσματικού χώρου, αρχικά, εξάγονται οι όροι-λέξεις που εμπεριέχονται στα κείμενα. Έπειτα, κάθε όρος σε κάθε διάνυσμα-κείμενο παίρνει την τιμή 1 ή 0, ανάλογα με τον αν εμφανίζεται ή όχι ο όρος στο συγκεκριμένο κείμενο, αντίστοιχα. Σε αυτή την τιμή αποδίδεται ένα βάρος που είναι ίσο με τη συχνότητα εμφάνισης του κάθε όρου μέσα στο συγκεκριμένο κείμενο. Έτσι, δημιουργείται ένα σύνολο δεδομένων που περιέχει τις συχνότητες των όρων, με πλήθος ίσο με τον αριθμό των κειμένων, και μεταβλητές εισόδου, τις λέξεις των κειμένων. Έπειτα, σε αυτό το σύνολο δεδομένων εφαρμόζεται η μέθοδος ανάλυσης κύριων συνιστωσών (PCA), η οποία αναλύεται στο κεφάλαιο 4.4.1. Η μέθοδος αυτή εφαρμόζεται με σκοπό τη μείωση του αριθμού των μεταβλητών εισόδου. Στη συνέχεια, εκτός από τις κύριες συνιστώσες των συχνοτήτων εμφάνισης των λέξεων, προστίθενται επιπλέον, οι παρακάτω μεταβλητές εισόδου:

- μία αριθμητική μεταβλητή, που περιλαμβάνει τον αριθμό των λέξεων που περιέχει το κείμενο κάθε ιστοσελίδας,
- 10 κατηγορικές μεταβλητές, που περιλαμβάνουν τις πρώτες λέξεις που εμφανίζονται στο κείμενο κάθε ιστοσελίδας,
- 3 κατηγορικές μεταβλητές, που περιέχουν τις λέξεις με τη μεγαλύτερη συχνότητα εμφάνισης μέσα στο κείμενο κάθε ιστοσελίδας,
- 3 κατηγορικές μεταβλητές, που περιέχουν τις λέξεις με τη μικρότερη συχνότητα εμφάνισης μέσα στο κείμενο κάθε ιστοσελίδας.

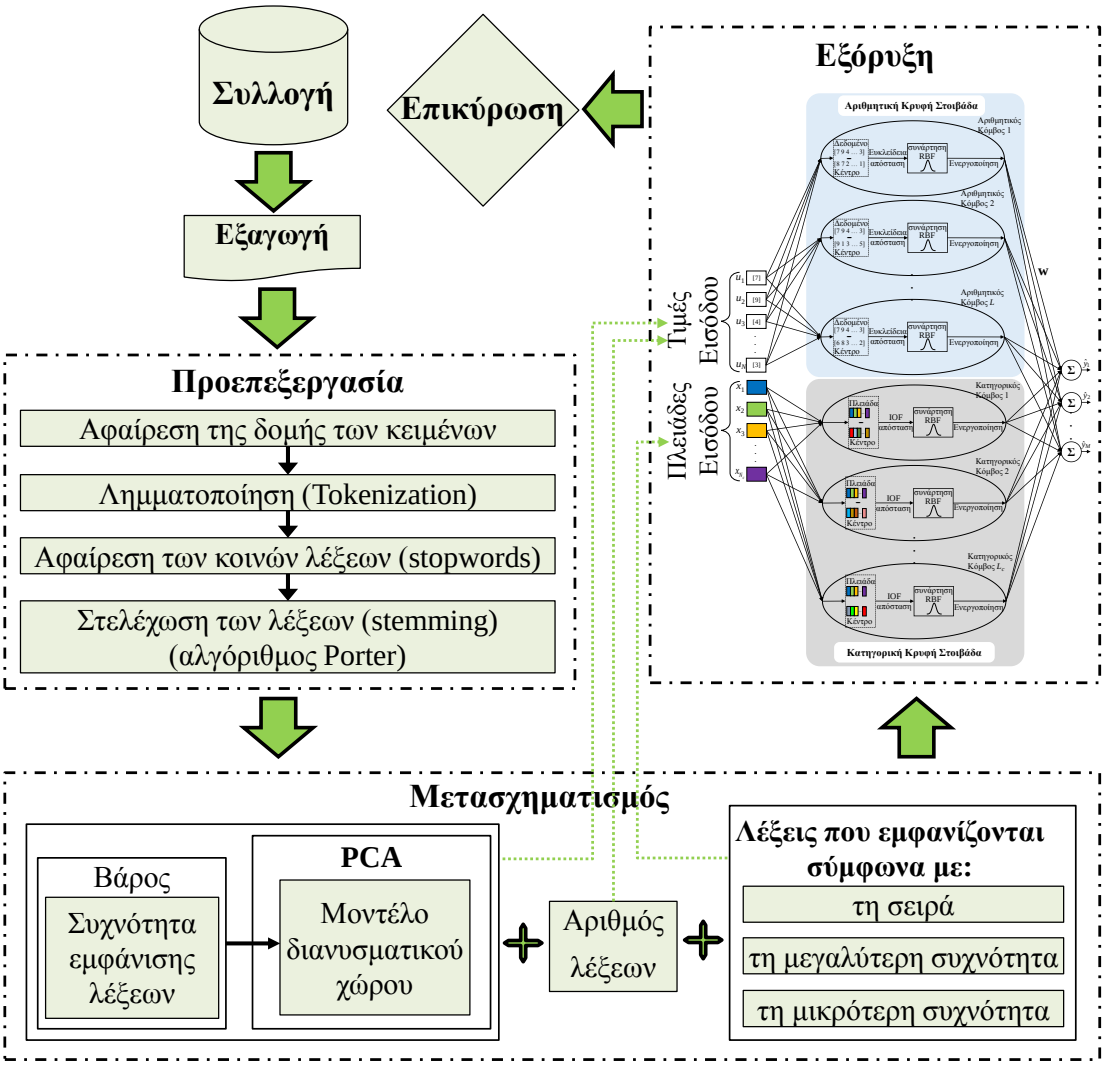
Έπειτα, αφού επιλεγούν τα χαρακτηριστικά και οι ιστοσελίδες μετασχηματιστούν σε κατάλληλα επεξεργάσιμη μορφή, το προτεινόμενο σύστημα προχωρά στο στάδιο της εξόρυξης. Αρχικά, οι ιστοσελίδες χωρίζονται με τυχαίο τρόπο σε 3 σύνολα δεδομένων, εκπαίδευσης, αξιολόγησης και ελέγχου, χρησιμοποιώντας το 50% των δεδομένων για εκπαίδευση, το 25% για αξιολόγηση και το υπόλοιπο 25% για έλεγχο των νευρωνικών δικτύων (όπως αναφέρεται στο κεφάλαιο 1.2.7). Στη συνέχεια, εφαρμόζονται τα δίκτυα MRBF, τα οποία λαμβάνουν ως εισόδους τις αριθμητικές και κατηγορικές μεταβλητές που αναλύθηκαν στην προηγούμενη παράγραφο και εκπαιδεύονται με βάση τον αλγόριθμο που παρουσιάστηκε στο κεφάλαιο 4.4.1 στο σύνολο εκπαίδευσης, ενώ το τελικό μοντέλο επιλέγεται με βάση το σύνολο αξιολόγησης. Τέλος, το στάδιο επικύρωσης περιλαμβάνει τον έλεγχο των δικτύων MRBF, όπου χρησιμοποιείται το σύνολο δεδομένων ελέγχου, το οποίο περιέχει άγνωστα δεδομένα για το δίκτυο MRBF. Μία επισκόπηση της λειτουργίας του προτεινόμενου συστήματος MRBF-WPC φαίνεται στο σχήμα 5.2.

5.3.1 Εφαρμογή του προτεινόμενου συστήματος στην ταξινόμηση ιστοσελίδων

Το προτεινόμενο σύστημα αυτόματης ταξινόμησης ιστοσελίδων εφαρμόστηκε στο σύνολο δεδομένων WebKB, που επίσης είναι γνωστό ως σύνολο δεδομένων 4 πανεπιστημίων (4 Universities Data Set) [121, 122]. Το σύνολο δεδομένων WebKB αναπτύχθηκε από την ομάδα εκμάθησης κειμένου CMU για το πρόγραμμα World Wide Knowledge Base (WebKB). Αυτό το σύνολο δεδομένων περιέχει 8145 ιστοσελίδες, οι οποίες συλλέχθηκαν από τμήματα πληροφορικής διαφόρων πανεπιστημίων το 1997. Οι ιστοσελίδες συλλέχθηκαν κυρίως από τέσσερα πανεπιστήμια (Cornell, Texas, Washington, Wisconsin) και κατατάχθηκαν χειροκίνητα σε επτά κατηγορίες:

- Course
- Department
- Faculty
- Project
- Staff
- Student
- other

Σύμφωνα με τους Nigam et al. [269], οι κατηγορίες “department” και “staff” απορρίφθηκαν επειδή υπήρχαν μόνο λίγες σελίδες από κάθε πανεπιστήμιο. Επίσης, οι σελίδες που αντιστοιχούσαν στην κατηγορία “other” διαγράφηκαν επειδή παρουσίαζαν μεγάλες διαφορές



Σχήμα 5.2. Σχηματική αναπαράσταση του συστήματος MRBF-WPC.

μεταξύ τους. Έτσι, χρησιμοποιήθηκαν οι κατηγορίες “course”, “faculty”, “project”, “student”, που συνολικά αντιστοιχούν σε 4199 σελίδες.

Για λόγους σύγκρισης, δοκιμάστηκε ένας αριθμός διαφορετικών τεχνικών ταξινόμησης, οι οποίες περιλαμβάνουν:

- τα νευρωνικά δίκτυα MLP με δύο κρυφές στοιβάδες, τα οποία εκπαιδεύτηκαν με τον αλγόριθμο Levenberg – Marquardt [2]
- τις μηχανές διανυσμάτων υποστήριξης SVM με γκαουσιανή συνάρτηση πυρήνα [228],
- τα δέντρα ταξινόμησης και παλινδρόμησης (CART) [236]

Οι παράμετροι των ταξινομητών, προσδιορίστηκαν χρησιμοποιώντας το σύνολο δεδομένων εκπαίδευσης και στη συνέχεια, με βάση το σύνολο δεδομένων αξιολόγησης, επιλέχτηκε το μοντέλο. Όλοι οι ταξινομητές εκτελέστηκαν 30 φορές και βελτιστοποιήθηκαν ξεχωριστά για κάθε σύνολο δεδομένων. Τα νευρωνικά δίκτυα MLP βελτιστοποιήθηκαν σύμφωνα με τον αριθμό των κόμβων των στοιβάδων. Οι παράμετροι των ταξινομητών SVM βελτιστοποιήθηκαν με βάση μια διαδικασία διεξοδικής αναζήτησης πλέγματος, όσον αφορά

στο εύρος των γκαουσιανών *σ* και στον συντελεστή ποινής *C*. Οι ταξινομητές που βασίζονται σε δέντρα βελτιστοποιήθηκαν με βάση τον αριθμό των δέντρων [238]. Τέλος, η απόδοση των ταξινομητών χαρακτηρίστηκε χρησιμοποιώντας τα κριτήρια σφάλματος MCC και ακρίβεια%, τα οποία περιγράφονται στο κεφάλαιο 1.2.8.

5.3.2 Αποτελέσματα και συζήτηση

Τα αποτελέσματα για όλες τις μεθόδους, με βάση τους δείκτες σφάλματος MCC και ακρίβεια % στα σύνολα αξιολόγησης και ελέγχου, συνοψίζονται στον πίνακα 5.1. Επίσης, στον πίνακα 5.1 με έντονους χαρακτήρες, δηλώνεται η καλύτερη απόδοση στα σύνολα αξιολόγησης και ελέγχου, για κάθε ένα από τα κριτήρια σφάλματος. Όλες οι τιμές που παρουσιάζονται στον πίνακα 5.1 αντιστοιχούν στο μοντέλο που επιτυγχάνει το καλύτερο αποτέλεσμα σε σχέση με τον δείκτη MCC στο σύνολο δεδομένων αξιολόγησης. Επίσης, για όλες τις τεχνικές που περιλαμβάνουν στοχαστικά στοιχεία, δηλαδή το προτεινόμενο σύστημα και τις τεχνικές MLP και CART, το καλύτερο αποτέλεσμα ακολουθείται από τον μέσο όρο και την τυπική απόκλιση των 30 εκτελέσεων μέσα σε παρένθεση.

Η σύγκριση της τεχνικής MRBF-WPC με τις υπόλοιπες μεθόδους, δείχνει ότι, με βάση τους δείκτες σφάλματος MCC και ακρίβεια%, το προτεινόμενο σύστημα υπερτερεί σε κάθε περίπτωση, είτε λαμβάνοντας υπόψη το καλύτερο αποτέλεσμα, είτε τον μέσο όρο των 20 τρεξιμάτων. Η υπεροχή αυτή, που περιλαμβάνει τόσο το σύνολο δεδομένων αξιολόγησης, όσο και το σύνολο ελέγχου, μπορεί να εξηγηθεί με βάση το ότι ο προτεινόμενος αλγόριθμος λαμβάνει ξεχωριστά υπόψη τη διαφορετική φύση των αριθμητικών και κατηγορικών δεδομένων, σε αντίθεση με τις υπόλοιπες μεθοδολογίες, είτε αυτές είναι εγγενώς σχεδιασμένες να χειρίζονται αριθμητικά δεδομένα και περιλαμβάνουν αυθαίρετη μετατροπή των κατηγορικών σε αριθμητικά (MLP και SVM), είτε το αντίθετο (δέντρα CART). Ως προς τη σύγκριση ανάμεσα στις υπόλοιπες μεθοδολογίες, όσον αφορά στους δείκτες σφάλματος MCC και ακρίβεια % τα δίκτυα MLP και οι μηχανές SVM δίνουν σχεδόν παρόμοια αποτελέσματα, με τα τελευταία να προβλέπουν λίγο καλύτερα στο σύνολο ελέγχου. Τέλος, τα

Πίνακας 5.1. Αποτελέσματα για το σύνολο δεδομένων WebKB – MCC και ακρίβεια%

Μέθοδος	**MCC συνόλου αξιολόγησης**	**MCC συνόλου ελέγχου**	**Ακρίβεια% συνόλου αξιολόγησης**	**Ακρίβεια% συνόλου ελέγχου**
MRBF-WPC	**0.80** (**0.80**±7.44×10^{-3})	**0.84** (**0.83**±2.00×10^{-2})	**85.3** (**85.1**±2.62×10^{-1})	**88.1** (**87.7**±3.98×10^{-1})
MLP	0.78 (0.78±5.92×10^{-3})	0.78 (0.77±9.63×10^{-3})	83.3 (82.9±4.15×10^{-1})	84.0 (83.7±6.61×10^{-1})
SVM	0.77	0.78	84.0	84.8
CART	0.61 (0.58±2.00×10^{-2})	0.59 (0.60±1.80×10^{-2})	73.8 (72.2±1.78×10^{0})	71.5 (73.7±1.19×10^{0})

Απεικονίζεται η εκτέλεση που βρέθηκε το καλύτερο αποτέλεσμα σύμφωνα με τον δείκτη MCC στο σύνολο αξιολόγησης για όλους τους ταξινομητές. Όλες οι μέθοδοι εκτελέστηκαν 20 φορές, εκτός τις μηχανές SVM. Για όλες τις τεχνικές, εκτός τις μηχανές SVM, το καλύτερο αποτέλεσμα ακολουθείται από τον μέσο όρο και την τυπική απόκλιση μέσα σε παρένθεση.

δέντρα CART κατατάσσονται στην τελευταία θέση με αρκετά μεγάλη διαφορά σε σχέση με τις άλλες τεχνικές που δοκιμάστηκαν, όσον αφορά στο σύνολο ελέγχου.

Μια καλύτερη εικόνα των αποτελεσμάτων του συστήματος MRBF-WPC, δίνουν οι πίνακες σύγχυσης του σχήματος 5.3. Αυτοί οι πίνακες παρουσιάζουν τα αποτελέσματα στα σύνολα δεδομένων αξιολόγησης και ελέγχου, που έδωσε το μοντέλο το οποίο βρήκε το καλύτερο MCC στο σύνολο αξιολόγησης. Παρατηρώντας τους πίνακες σύγχυσης, είναι εμφανές ότι το προτεινόμενο σύστημα, όχι μόνο κατορθώνει ένα ικανοποιητικό συνολικό ποσοστό επιτυχίας, αλλά επίσης προβλέπει με καλή απόδοση κάθε κατηγορία ιστοσελίδας ξεχωριστά. Συγκεκριμένα, στο σύνολο αξιολόγησης η προτεινόμενη μεθοδολογία προβλέπει καλύτερα την κατηγορία “Course” και ακολουθούν οι κατηγορίες “Faculty”, “Student” και “Project”. Όσον αφορά στο σύνολο ελέγχου το σύστημα MRBF-WPC προβλέπει καλύτερα την κατηγορία “Course” και ακολουθούν οι κατηγορίες “Faculty”, “Project” και “Student”.

Τέλος, στον πίνακα 5.2 φαίνονται οι βελτιστοποιημένες παράμετρο του μοντέλου που έδωσε το καλύτερο αποτέλεσμα σύμφωνα με τον δείκτη MCC στο σύνολο αξιολόγησης για όλους τους ταξινομητές.

Πίνακας 5.2. Αποτελέσματα για το σύνολο δεδομένων WebKB – βελτιστοποιημένες παράμετροι και υπολογιστικοί χρόνοι

Μέθοδος	**Βελτιστοποιημένες Παράμετροι**	**Τιμές**
MRBF-WPC	Ασαφής διαμέριση	35
	Κατώφλι δ	0.4
MLP	Αριθμός νευρώνων 1ης στοιβάδας	34
	Αριθμός νευρώνων 2ης στοιβάδας	30
SVM	Συντελεστής ποινής C	0.063
	Εύρος των γκαουσιανών πυρήνων σ	0.966
CART	Αριθμός παρατηρήσεων σε κάθε κλαδί	100
	Αριθμός μεταβλητών εισόδου	100

Απεικονίζονται οι παράμετροι της εκτέλεσης για την οποία βρέθηκε το καλύτερο αποτέλεσμα σύμφωνα με τον δείκτη MCC στο σύνολο αξιολόγησης για όλους τους ταξινομητές. Όλες οι μέθοδοι εκτελέστηκαν 20 φορές, εκτός τις μηχανές SVM.

Προβλέψεις \ Πραγματικές Τιμές	Course	Faculty	Project	Student	
Course	20.1% (195)	0.6% (6)	0.3% (3)	0.0% (0)	95.6% 4.4%
Faculty	0.3% (3)	21.9% (213)	1.6% (16)	1.5% (15)	86.2% 13.8%
Project	0.9% (9)	0.6% (6)	8.4% (82)	0.8% (8)	78.1% 21.9%
Student	1.3% (13)	4.5% (44)	2.1% (20)	34.8% (338)	81.4% 18.6%
	88.6% 11.4%	79.2% 20.8%	67.8% 32.2%	93.6% 6.4%	85.3% 14.7%

(α)

Προβλέψεις \ Πραγματικές Τιμές	Course	Faculty	Project	Student	
Course	21.0% (203)	0.7% (7)	0.8% (8)	0.2% (2)	92.3% 7.7%
Faculty	0.2% (2)	22.2% (215)	0.5% (5)	1.5% (15)	90.7% 9.3%
Project	0.5% (5)	0.9% (9)	9.6% (93)	0.2% (2)	85.3% 14.7%
Student	0.9% (9)	3.8% (37)	1.4% (14)	35.3% (342)	85.1% 14.9%
	92.7% 7.3%	80.2% 18.8%	77.5% 22.5%	94.7% 5.3%	88.1% 11.9%

(β)

Σχήμα 5.3. Πίνακες σύγχυσης για το μοντέλο που έδωσε το καλύτερο MCC στο σύνολο αξιολόγησης για το σύστημα MRBF-WPC στα σύνολα δεδομένων (α) αξιολόγησης και (β) ελέγχου.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Ο στόχος της παρούσας διδακτορικής διατριβής ήταν διττός: σε μεθοδολογικό επίπεδο στόχος ήταν η ανάπτυξη νέων καινοτόμων τεχνικών υπολογιστικής νοημοσύνης με τη δυνατότητα χειρισμού αριθμητικών, κατηγορικών και μικτών δεδομένων ως εισόδους, και βελτιωμένες ικανότητες πρόβλεψης. Σε επίπεδο εφαρμογής στόχος ήταν ο συνδυασμός αυτών των μεθόδων με στοιχεία από τον κλάδο της εξόρυξης γνώσης στον ιστό, για την ανάπτυξη ενός νέου ευφυούς συστήματος αυτόματης ταξινόμησης ιστοσελίδων, που μπορεί να χρησιμοποιηθεί ως βάση για την εξέλιξη της τρέχουσας μορφής του παγκόσμιου ιστού σε σημασιολογικό ιστό.

Η επίτευξη των παραπάνω στόχων που αφορούν στο μεθοδολογικό επίπεδο, πραγματοποιήθηκε αντλώντας από τον χώρο της υπολογιστικής νοημοσύνης και συγκεκριμένα τους τομείς των τεχνητών νευρωνικών δικτύων, του εξελικτικού υπολογισμού και της νοημοσύνης σμήνους, βασικά εργαλεία που περιλαμβάνουν τα δίκτυα RBF, την τεχνική ESA και τον αλγόριθμο PSO.

Σε ότι αφορά στον χειρισμό αριθμητικών εισόδων, στα πλαίσια της παρούσας διατριβής, αναπτύχθηκαν 3 νέοι αλγόριθμοι εκπαίδευσης νευρωνικών δικτύων RBF. Συνολικά, αυτοί οι αλγόριθμοι, παρουσιάζουν τα εξής πλεονεκτήματα σε σχέση με τις περισσότερες μεθόδους που έχουν προταθεί στη βιβλιογραφία:

- επιτυγχάνουν την πλήρη βελτιστοποίηση των δικτύων RBF
- επιτυγχάνουν μεγαλύτερη ακρίβεια πρόβλεψης
- επιτυγχάνουν συνεπέστερα αποτελέσματα όσον αφορά στην ακρίβεια πρόβλεψης
- δημιουργούν μοντέλα με απλούστερες δομές

Οι προτεινόμενοι αλγόριθμοι προσφέρουν πολλές επιλογές με αποτέλεσμα, αναλόγως με τις ανάγκες του εκάστοτε προβλήματος, η επιλογή και η εφαρμογή ενός από αυτούς τους αλγορίθμους να εγγυάται ικανοποιητικές λύσεις. Συγκεκριμένα, ο αλγόριθμος PSO-NSFM δημιουργεί μοντέλα δικτύων RBF που παρουσιάζουν απλούστερες δομές και ταυτόχρονα μεγαλύτερη ακρίβεια πρόβλεψης σε μικρότερους υπολογιστικούς χρόνους, έναντι καθιερωμένων τεχνικών μηχανικής μάθησης. Ο αλγόριθμος COOP-VW επιτρέπει τη χρήση συναρτήσεων βάσης μεταβλητού εύρους, οι οποίες αυξάνουν την ευελιξία των παραγόμενων μοντέλων RBF, με αποτέλεσμα την επίτευξη ακόμα μεγαλύτερης ακρίβειας πρόβλεψης, αλλά με σχετικά μικρή αύξηση του υπολογιστικού κόστους σε σχέση με τον αλγόριθμο PSO-NSFM. Ο αλγόριθμος ESA-NSFM δημιουργεί μοντέλα δικτύων RBF, τα οποία παρουσιάζουν αυξημένη ακρίβεια πρόβλεψης, σε σχέση με τις καθιερωμένες τεχνικές μηχανικής μάθησης, και ταυτόχρονα εύκολη επιλογή λειτουργικών παραμέτρων.

Οι προτεινόμενοι αλγόριθμοι αξιολογήθηκαν σε μια ποικιλία από διαφορετικά σύνολα δεδομένων τύπου benchmark, καλύπτοντας ένα ευρύ φάσμα από πραγματικά και συνθετικά σύνολα δεδομένων, που περιλαμβάνουν προβλήματα προσέγγισης συνάρτησης και ταξινόμησης. Τα σύνολα δεδομένων που χρησιμοποιήθηκαν, περιλαμβάνουν εφαρμογές από διαφορετικούς επιστημονικούς τομείς, συμπεριλαμβανομένης της ιατρικής, των κοινωνικών

επιστημών, της αεροναυπηγικής, της ενέργειας, της επιστήμης των υπολογιστών, κ.α. Η απόδοση των προτεινόμενων μεθόδων σε σύνολα δεδομένων με μεγάλη πολυπλοκότητα υποδηλώνει ότι θα μπορούσαν να χρησιμοποιηθούν στην επεξεργασία εικόνας, στην αναγνώριση ομιλίας, στη βιομετρία, στη βιοπληροφορική και άλλες εφαρμογές που συνήθως απαιτούν αυξημένη υπολογιστική ισχύ. Επίσης, στην συντριπτική πλειοψηφία των συνόλων δεδομένων που χρησιμοποιήθηκαν, οι προτεινόμενες μέθοδοι βρέθηκαν να αποδίδουν καλύτερα σε σχέση με καθιερωμένες μεθόδους από τον χώρο της μηχανικής μάθησης που περιλαμβάνουν, ανάμεσα σε άλλες, τα νευρωνικά δίκτυα MLP και τις μηχανές SVM.

Στα πλαίσια αυτής της διδακτορικής διατριβής, αναπτύχθηκαν επίσης 2 καινοτόμες αρχιτεκτονικές νευρωνικών δικτύων και οι αντίστοιχοι αλγόριθμοι εκπαίδευσής τους, με τη δυνατότητα χειρισμού κατηγορικών και μικτών δεδομένων. Οι αρχιτεκτονικές αυτές επεκτείνουν τις δυνατότητες των δικτύων RBF, τα οποία, ενώ αποτελούν μια άριστη μέθοδο προσέγγισης αριθμητικών δεδομένων, υστερούν στο χειρισμό κατηγορικών και μικτών δεδομένων.

Συγκεκριμένα, η νέα αρχιτεκτονική CRBF έχει την ικανότητα να χειρίζεται αποκλειστικά κατηγορικά δεδομένα, χρησιμοποιώντας κατηγορικές πλειάδες ως κέντρα RBF, ενώ εκπαιδεύεται με τον αλγόριθμο CSS. Αυτός ο αλγόριθμος παρουσιάζει τρία σημαντικά πλεονεκτήματα συμπεριλαμβανομένης της απλότητάς του, της ταχύτητάς του και της χρήσης μίας μόνο παραμέτρου για την επιλογή μοντέλου. Τα δίκτυα CRBF επεκτείνουν τις δυνατότητες των δικτύων RBF στο χειρισμό κατηγορικών δεδομένων, λαμβάνοντας υπόψη άμεσα την ιδιαίτερη φύση των τελευταίων, σε αντίθεση με τις περισσότερες μεθόδους στη βιβλιογραφία, οι οποίες χειρίζονται τις κατηγορικές μεταβλητές με έμμεσο τρόπο. Αντίστοιχα, η αρχιτεκτονική των νευρωνικών δικτύων MRBF που ουσιαστικά αποτελεί συνδυασμό των δικτύων RBF και CRBF, και εκπαιδεύεται με μία νέα μεθοδολογία, έχει την ικανότητα να χειρίζεται μικτά δεδομένα. Σε αντίθεση με τις μεθόδους που έχουν προταθεί στη βιβλιογραφία, οι περισσότερες εκ των οποίων αντιμετωπίζουν τα μικτά δεδομένα ως ένα ενιαίο είδος δεδομένων, μετατρέποντας τα κατηγορικά δεδομένα σε αριθμητικά, ή το αντίθετο, η αρχιτεκτονική δικτύων MRFB παρουσιάζει την ικανότητα να λαμβάνει υπόψη ξεχωριστά τη φύση των αριθμητικών και των κατηγορικών δεδομένων. Επίσης θα πρέπει να τονισθεί ότι τα δίκτυα CRBF και MRBF, λόγω των γρήγορων αλγορίθμων εκπαίδευσης που χρησιμοποιούν, μπορούν να δώσουν λύσεις σε δεδομένα μεγάλου όγκου (big data).

Οι προτεινόμενες 2 αρχιτεκτονικές νευρωνικών δικτύων αξιολογήθηκαν σε μια ποικιλία από διαφορετικά σύνολα δεδομένων τύπου benchmark, κάποια εκ των οποίων ανήκουν στην κατηγορία big data. Επίσης, τα σύνολα δεδομένων που χρησιμοποιήθηκαν περιλαμβάνουν εφαρμογές από διαφορετικούς επιστημονικούς τομείς, όπως πχ η μοριακή βιολογία, η ιατρική, οι κοινωνικές επιστήμες, η θεωρία παιγνίων, η επιστήμη της εγκληματολογίας, κ.α. Για λόγους σύγκρισης, χρησιμοποιήθηκαν αρκετά διαφορετικά σχήματα μάθησης που περιλαμβάνουν μηχανές ELM, κατηγοριοποιητές Bayes, δέντρα απόφασης κ.α. Τα αποτελέσματα έδειξαν ότι οι προτεινόμενες αρχιτεκτονικές νευρωνικών δικτύων παρουσιάζουν καλή ικανότητα γενίκευσης και μεγαλύτερη ικανότητα πρόβλεψής στη μεγάλη πλειοψηφία των περιπτώσεων, έναντι των μεθόδων που δοκιμάστηκαν.

Τα δίκτυα MRBF λόγω της ικανότητάς τους να χειρίζονται μικτά δεδομένα, υιοθετήθηκαν, σε συνδυασμό με στοιχεία από τον κλάδο της εξόρυξης γνώσης στον ιστό, για να δώσουν λύση σε ένα βασικό πρόβλημα που καθυστερεί την εξέλιξη του παγκόσμιου ιστού σε σημασιολογικό ιστό. Στα πλαίσια αυτής της διατριβής, αναπτύχθηκε το σχήμα MRBF-WPC, το οποίο είναι ένα νέο ευφυές σύστημα με την ικανότητα αυτόματης ταξινόμησης ιστοσελίδων με βάση το περιεχόμενό τους. Το σύστημα MRBF-WPC αποτελεί ένα εργαλείο γενικής χρήσης για την ταξινόμηση ιστοσελίδων, που μπορεί για παράδειγμα να ταξινομήσει

μία ιστοσελίδα ανάλογα με τη λειτουργία της, το είδος της, τη θεματική περιοχή της, κ.α. Έτσι, χρησιμοποιώντας το σύστημα αυτό, οι ιστοσελίδες μπορούν να οργανωθούν σε κατηγορίες, με αποτέλεσμα την ευκολότερη αναζήτηση και εύρεση της επιθυμητής πληροφορίας. Σε αντίθεση με τις περισσότερες μεθόδους στη βιβλιογραφία, οι οποίες εκμεταλλεύονται ένα μόνο είδος χαρακτηριστικών, πχ αριθμητικά μέσω συχνότητας εμφάνισης λέξεων, το σύστημα αυτό μπορεί να αποκωδικοποιήσει την πληροφορία που περιέχουν ποικίλα χαρακτηριστικά για την επιτυχή ταξινόμηση των ιστοσελίδων.

Επίσης, το σύστημα MRBF-WPC αποτελεί ένα ολοκληρωμένο εργαλείο που μπορεί να χρησιμοποιηθεί για τον αυτόματο εμπλουτισμό του περιεχομένου του ιστού με σχολιασμούς, καθώς η υλοποίηση του σημασιολογικού ιστού εξαρτάται σε μεγάλο βαθμό από τον εμπλουτισμό της υπάρχουσας πληροφορίας. Δεδομένου ότι ο τρόπος παραγωγής της τελευταίας αποτελεί ένα μεγάλο εμπόδιο για πολλούς και διάφορους λόγους, και οι περισσότερες μέθοδοι που έχουν προταθεί για αυτόν τον σκοπό στη βιβλιογραφία είναι μόνο ημι-αυτόματες, η πλήρης αυτοματοποίηση που προσφέρει η προτεινόμενη τεχνική, μπορεί να παίξει σημαντικό ρόλο. Στα πλαίσια αυτά, το σύστημα MRBF-WPC μπορεί να χρησιμοποιηθεί ως ένα πρώτο βήμα για την αυτόματη επισημείωση του ιστού, καθώς το έργο της επισήμανσης ολόκληρου του ιστού ταυτόχρονα είναι δύσκολο, δεδομένου ότι η διαθέσιμη τεχνολογία για την αυτόματη επισημείωση δεν είναι ακόμα ώριμη.

Το σύστημα MRBF-WPC συγκρίθηκε με ποικίλες μεθόδους, συμπεριλαμβανομένων διάφορων κλασσικών τεχνικών μηχανικής μάθησης και τα αποτελέσματα έδειξαν ότι έχει την ικανότητα να ταξινομεί με μεγαλύτερη ακρίβεια τις ιστοσελίδες, ως προς το συνολικό ποσοστό επιτυχίας, αλλά και για κάθε κατηγορία ιστοσελίδας ξεχωριστά, σε σχέση με όλες τις τεχνικές που δοκιμάστηκαν.

Τέλος, αντικείμενο μελλοντικής έρευνας θα μπορούσε να αποτελέσει η εφαρμογή αλγορίθμων ευφυΐας σμήνους σε συνδυασμό με τη μέθοδο NSFM για την εκπαίδευση νευρωνικών δικτύων MRBF αλλά και η προσαρμογή των τεχνικών βελτιστοποίησης για την εκπαίδευση άλλων αρχιτεκτονικών δικτύων, όπως deep neural networks. Άλλη μία πιθανή μελλοντική επέκταση είναι η εφαρμογή των μεθόδων υπολογιστικής νοημοσύνης, οι οποίες αναπτύχθηκαν στα πλαίσια της παρούσας διδακτορικής διατριβής, στην επίλυση προβλημάτων πολύ μεγάλου όγκου δεδομένων χρησιμοποιώντας τεχνικές παράλληλου υπολογισμού, λαμβάνοντας υπόψη ότι οι μεθοδολογίες εκπαίδευσης δικτύων RBF έχουν το σημαντικό πλεονέκτημα ότι δεν απαιτούν να φορτωθούν όλα τα δεδομένα στη μνήμη του υπολογιστικού συστήματος. Ακόμη, μπορούν να ληφθούν υπόψη, ειδικά στις εφαρμογές στον παγκόσμιο ιστό, ετερογενείς πληροφορίες, όπως εικόνες, ήχος κ.λπ. με την εξαγωγή κατάλληλων χαρακτηριστικών που μπορούν να ενσωματωθούν στα δεδομένα εκπαίδευσης των δικτύων. Με τον τρόπο αυτό, η επίλυση του προβλήματος της σημασιολογικής ετερογένειας θα μπορούσε να οδηγήσει ακόμα ένα βήμα πιο κοντά στην εξέλιξη του παγκόσμιου ιστού σε σημασιολογικό ιστό.

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] A. Engelbrecht, *Computational Intelligence: An Introduction*, 2nd ed. Chichester: John Wiley &Sons, Ltd, 2007.

[2] S. Haykin, *Neural Networks: A comprehensive foundation*, second ed. Upper Saddle River, NJ: Prentice Hall, 1999.

[3] M. E. Aydin and T. C. Fogarty, "A Distributed Evolutionary Simulated Annealing Algorithm for Combinatorial Optimisation Problems," *Journal of Heuristics*, vol. 10, pp. 269-292, 2004.

[4] M. T. Hagan and M. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, pp. 989–993, 1994.

[5] M. J. D. Powell, "Radial basis functions for multi-variable interpolation: A review," in *IMA Conference on Algorithms for the Approximation of Functions and Data*, RMCS, Shrivenham, England, 1985, pp. 143-167.

[6] M. J. D. Powell, "The theory of radial basis function approximation in 1990," in *Advances in Numerical Analysis*. vol. 2, W. A. Light, Ed., ed Oxford: Oxford University Press, 1992, pp. 105-210.

[7] C. A. Michelli, "Interpolation of scattered data: distance matrices and conditionally positive definite functions," *Constructive Approximation*, vol. 2, pp. 11-22, 1986.

[8] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proceedings of the IEEE*, vol. 78, pp. 1481 - 1497, 1990.

[9] C. Darken and J. Moody, "Fast Adaptive K-Means Clustering: Some Empirical Results," in *IEEE INNS International Joint Conference On Neural Networks*, San Diego, CA, 1990, pp. 233-238.

[10] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," in *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967, p. 281.

[11] J. Moody and C. J. Darken, "Fast learning in networks of locally-tuned processing units," *Neural Computation*, vol. 1, pp. 281-294, 1989.

[12] S. Chen, C.F.N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Transactions on Neural Networks*, vol. 2, pp. 302 - 309, 1991.

[13] X. Hong, "A Fast Identification Algorithm for Box–Cox Transformation Based Radial Basis Function Neural Network," *IEEE Transactions on Neural Networks*, vol. 17, pp. 1064 - 1069, 2006.

[14] S. Chen, X. Hong, B. L. Luk, and C. J. Harris, "Construction of Tunable Radial Basis Function Networks Using Orthogonal Forward Selection," *IEEE Transactions on Systems, Man, and Cybernetics Part B*, vol. 39, pp. 457 - 466, 2009.

[15] S. Chen, X. X. Wang, X. Hong, and C. J. Harris, "Kernel Classifier Construction Using Orthogonal Forward Selection and Boosting With Fisher Ratio Class Separability Measure," *IEEE Transactions on Neural Networks*, vol. 17, pp. 1652 - 1656, 2006.

[16] Y. W. Wong, K. P. Seng, and L. M. Ang, "Radial Basis Function Neural Network With Incremental Learning for Face Recognition," *IEEE Transactions on Systems, Man, and Cybernetics Part B*, vol. 41, pp. 940-949, 2011.

[17] M. Hou and X. Han, "Constructive Approximation to Multivariate Function by Decay RBF Neural Network," *IEEE Transactions on Neural Networks*, vol. 21, pp. 1517 - 1523, 2010.

[18] L. Yingwei, N. Sundararajan, and P. Saratchandran, "A Sequential Learning Scheme for Function Approximation Using Minimal Radial Basis Function Neural Networks," *Neural Computation*, vol. 19, pp. 461-478, 1997.

[19] A. Leonardis and H. Bischof, "An efficient MDL-based construction of RBF networks," *Neural Networks*, vol. 11, pp. 963–973, 1998.

[20] H. Sarimveis, A. Alexandridis, G. Tsekouras, and G. Bafas, "A fast and efficient algorithm for training radial basis function neural networks based on a fuzzy partition of the input space," *Industrial and Engineering Chemistry Research*, vol. 41, pp. 751-759, 2002.

[21] W. Pedrycz, V. Loia, and S. Senatore, "Fuzzy Clustering With Viewpoints," *IEEE Transactions on Fuzzy Systems*, vol. 18, pp. 274-284, 2010.

[22] U. Maulik, S. Bandyopadhyay, and I. Saha, "Integrating Clustering and Supervised Learning for Categorical Data Analysis," *IEEE Transactions on Systems, Man, and Cybernetics Part A*, vol. 40, pp. 664-675, 2010.

[23] X. X. Zhang, H. X. Li, and C. K. Qi, "Spatially Constrained Fuzzy-Clustering-Based Sensor Placement for Spatiotemporal Fuzzy-Control System," *IEEE Transactions on Fuzzy Systems*, vol. 18, pp. 946-957, 2010.

[24] A. Alexandridis, E. Chondrodima, E. Efthimiou, G. Papadakis, F. Vallianatos, and D. Triantis, "Large Earthquake Occurrence Estimation Based on Radial Basis Function Neural Networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, pp. 5443-5453, 2014.

[25] P. Petsagourakis, E. Chondrodima, A. Alexandridis, and H. Sarimveis, "Reinforcement learning algorithm for the process control using online radial basis function neural networks," in *10th Greek scientific conference on Chemical Engineering*, Patra, Greece, 2015.

[26] K. Ninos, C. Giannakakis, I. Kompogiannis, I. Stavrakas, and A. Alexandridis, "Nonlinear Control of a DC-Motor Based on Radial Basis Function Neural Networks," in *IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA) Proceedings*, Istanbul, Turkey, 2011.

[27] P. Patrinos, A. Alexandridis, K. Ninos, and H. Sarimveis, "Variable selection in nonlinear modeling based on RBF networks and evolutionary computation," *International Journal of Neural Systems*, vol. 20, pp. 365-379, 2010.

[28] A. Alexandridis, P. Patrinos, H. Sarimveis, and G. Tsekouras, "A two-stage evolutionary algorithm for variable selection in the development of RBF neural network models," *Chemometrics and Intelligent Laboratory Systems*, vol. 75, pp. 149-162, 2005.

[29] A. Alexandridis, E. Chondrodima, K. Moutzouris, and D. Triantis, "A neural network approach for the prediction of the refractive index based on experimental data," *Journal of Materials Science*, vol. 47, pp. 883-891, 2012.

[30] A. Alexandridis, D. Triantis, E. Chondrodima, C. Stergiopoulos, G. Hloupis, I. Stavrakas, and K. Ninos, "An adaptive soft-sensor for non-destructive cement-based

material testing, through the use of RBF networks," in *2012 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, 2012, pp. 51-56.

[31] A. Afantitis, G. Melagraki, K. Makridima, A. Alexandridis, H. Sarimveis, and O. Iglessi-Markopoulou, "Prediction of high weight polymers glass transition temperature using RBF neural networks," *Journal of Molecular Structure: THEOCHEM*, vol. 716, pp. 193–198, 2005.

[32] E. Chondrodima, A. Alexandridis, N. Pelekis, and T. Y., "Prediction in Activity-Travel Diaries using Neural Networks," in *Fourth DATA SIM Workshop*, Piraeus, Greece, 2013.

[33] E. Chondrodima, "Prediction in Activity-Travel Diaries using Neural Networks," Master Thesis, Department of Informatics, University of Piraeus, Piraeus, Greece, 2013.

[34] A. Alexandridis, H. Sarimveis, and G. Bafas, "A new algorithm for online structure and parameter adaptation of RBF networks," *Neural Networks*, vol. 16, pp. 1003-1017, 2003.

[35] A. Alexandridis, H. Sarimveis, and K. Ninos, "A radial basis function network training algorithm using a non-symmetric partition of the input space - Application to a model predictive control configuration," *Advances in Engineering Software*, vol. 42, pp. 830-837, 2011.

[36] J. A. Leonard and M. A. Kramer, "Radial basis function networks for classifying process faults," *IEEE Control Systems*, pp. 31–38, 1991.

[37] M. Orr, "Optimising the widths of radial basis functions," in *5th Brazilian Symposium on Neural Networks*, Brazil, 1998, pp. 26-29.

[38] W. Yao, X. Chen, M. v. Tooren, and Y. Wei, "Euclidean Distance and Second Derivative Based Widths Optimization of Radial Basis Function Neural Networks," in *IEEE International Joint Conference on Neural Networks (IJCNN)*, Barcelona, 2010, pp. 1-8.

[39] W. Yao, X. Chen, Y. Zhao, and M. v. Tooren, "Concurrent Subspace Width Optimization Method for RBF Neural Network Modeling," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, pp. 247-259, 2012.

[40] S. F. Su, C. C. Chuang, C. W. Tao, J. J. Jeng, and C. C. Hsiao, "Radial Basis Function Networks With Linear Interval Regression Weights for Symbolic Interval Data," *IEEE Transactions on Systems, Man, and Cybernetics Part B*, vol. 42, pp. 69-80, 2012.

[41] D. Zhang, L. F. Deng, K. Y. Cai, and A. So, "Fuzzy Nonlinear Regression With Fuzzified Radial Basis Function Network," *IEEE Transactions on Fuzzy Systems*, vol. 13, pp. 742-760, 2005.

[42] R. Langari, L. Wang, and J. Yen, "Radial basis function networks, regression weights, and the expectation-maximization algorithm," *IEEE Transactions on Systems, Man, and Cybernetics Part A*, vol. 27, pp. 613 - 623, 1997.

[43] C. Andrieu, N. d. Freitas, and A. Doucet, "Robust Full Bayesian Learning for Radial Basis Networks," *Neural Computation*, vol. 13, pp. 2359–2407, 2001.

[44] H. X. Huan, D. T. T. Hien, and H. H. Tue, "Efficient Algorithm for Training Interpolation RBF Networks with Equally Spaced Nodes," *IEEE Transactions on Neural Networks*, vol. 22, pp. 982-988, 2011.

[45] T. Xie, H. Yu, J. Hewlett, P. Rózycki, and B. Wilamowski, "Fast and Efficient Second-Order Method for Training Radial Basis Function Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, pp. 609-619, 2012.

[46] D. B. Fogel, L. J. Fogel, and V. W. Porto, "Evolutionary programming for training neural networks," in *International Joint Conference on Neural Networks (IJCNN)*, San Diego, CA, USA 1990, pp. 601-605.

[47] R. Bisoi and P. K. Dash, "A hybrid evolutionary dynamic neural network for stock market trend analysis and prediction using unscented Kalman filter," *Applied Soft Computing*, vol. 19, pp. 41-56, 2014.

[48] J. P. Donate, P. Cortez, G. Gutiérrez Sánchez, and A. Sanchis de Miguel, "Time series forecasting using a weighted cross-validation evolutionary artificial neural network ensemble," *Neurocomputing*, vol. 109, pp. 27-32, 2013.

[49] J. P. Donate, X. Li, G. Gutiérrez Sánchez, and A. Sanchis de Miguel, "Time series forecasting by evolving artificial neural networks with genetic algorithms, differential evolution and estimation of distribution algorithm," *Neural Computing and Applications*, vol. 22, pp. 11-20, 2013.

[50] S. C. Tan, J. Watada, Z. Ibrahim, and M. Khalid, "Evolutionary Fuzzy ARTMAP Neural Networks for Classification of Semiconductor Defects," *IEEE Transactions on Neural Networks and Learning Systems*, 2014.

[51] H. Sarimveis, A. Alexandridis, S. Mazarakis, and G. Bafas, "A new algorithm for developing dynamic radial basis function neural network models based on genetic algorithms," *Computers and Chemical Engineering*, vol. 28, pp. 209-217, 2004.

[52] P. A. Gutiérrez, C. Hervás-Martínez, and F. J. Martínez-Estudillo, "Logistic Regression by Means of Evolutionary Radial Basis Function Neural Networks," *IEEE Transactions on Neural Networks*, vol. 22, pp. 246-263, 2011.

[53] J. González, I. Rojas, J. Ortega, H. Pomares, F. J. Fernández, and A. F. Díaz, "Multiobjective Evolutionary Optimization of the Size, Shape, and Position Parameters of Radial Basis Function Networks for Function Approximation," *IEEE Transactions on Neural Networks*, vol. 14, pp. 1478-1495, 2003.

[54] O. Buchtala, M. Klimek, and B. Sick, "Evolutionary Optimization of Radial Basis Function Classifiers for Data Mining Applications," *IEEE Transactions on Systems, Man, and Cybernetics Part B*, vol. 35, pp. 928-947, 2005.

[55] F. Fernández-Navarro, C. Hervás-Martínez, R. Ruiz, and J. C. Riquelme, "Evolutionary Generalized Radial Basis Function neural networks for improving prediction accuracy in gene classification using feature selection," *Applied Soft Computing*, vol. 12, pp. 1787-1800, 2012.

[56] W. Huang, S.-K. Oh, and W. Pedrycz, "Design of hybrid radial basis function neural networks (HRBFNNs) realized with the aid of hybridization of fuzzy clustering method (FCM) and polynomial neural networks (PNNs)," *Neural Networks*, vol. 60, pp. 166-181, 2014.

[57] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *IEEE International Conference on Neural Networks*, 1995, pp. 1942-1948.

[58] F. Valdez, P. Melin, and O. Castillo, "Evolutionary method combining Particle Swarm Optimisation and Genetic Algorithms using fuzzy logic for parameter adaptation and aggregation: the case neural network optimisation for face recognition.," *International Journal of Artificial Intelligence*, vol. 2, pp. 77-102, 2010.

[59] F. Valdez, P. Melin, and O. Castillo, "An improved evolutionary method with fuzzy logic for combining Particle Swarm Optimization and Genetic Algorithms.," *Applied Soft Computing*, vol. 11, pp. 2625-2632, 2011.

[60] S. Kiranyaz, T. Ince, A. Yildirim, and M. Gabbouj, "Evolutionary artificial neural networks by multi-dimensional particle swarm optimization," *Neural Networks,* vol. 22, pp. 1448-1462, 2009.

[61] R. K. Agrawal and N. G. Bawane, "Multiobjective PSO based adaption of neural network topology for pixel classification in satellite imagery," *Applied Soft Computing,* vol. 28, pp. 217-225, 2015.

[62] S. Chen, X. Hong, and C. J. Harris, "Particle Swarm Optimization Aided Orthogonal Forward Regression for Unified Data Modeling," *IEEE Transactions on Evolutionary Computation,* vol. 14, pp. 477-499, 2010.

[63] D. S. Huang and J. X. Du, "A Constructive Hybrid Structure Optimization Methodology for Radial Basis Probabilistic Neural Networks," *IEEE Transactions on Neural Networks,* vol. 19, pp. 2099-2115, 2008.

[64] S. K. Oh, W. D. Kim, W. Pedrycz, and S. C. Joo, "Design of K-means clustering-based polynomial radial basis function neural networks (pRBFNNs) realized with the aid of particle swarm optimization and differential evolution," *Neurocomputing,* vol. 78, pp. 121–132, 2012.

[65] S. K. Oh, W. D. Kim, W. Pedrycz, and B. J. Park, "Polynomial-based radial basis function neural networks (P-RBF NNs) realized with the aid of particle swarm optimization," *Fuzzy Sets and Systems,* vol. 163, pp. 54-77, 2011.

[66] V. Fathi and G. A. Montazer, "An improvement in RBF learning algorithm based on PSO for real time applications," *Neurocomputing,* vol. 111, pp. 169-176, 2013.

[67] A. Alexandridis, E. Chondrodima, and H. Sarimveis, "Radial Basis Function network training using a non-symmetric partition of the input space and Particle Swarm Optimization," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 24, pp. 219-230, 2013.

[68] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing,* vol. 10, pp. 293-302, Jul 2002.

[69] B. L. Sturm, "A survey of evaluation in music genre recognition," in *10th International Workshop on Adaptive Multimedia Retrieval*, Copenhagen, Denmark 2012.

[70] S. Ono, H. Maeda, K. Sakimoto, and S. Nakayama, "User-system cooperative evolutionary computation for both quantitative and qualitative objective optimization in image processing filter design," *Applied Soft Computing,* vol. 15, pp. 203-218, 2014.

[71] W. Zhao, S. Alam, and H. A. Abbass, "MOCCA-II: A multi-objective co-operative co-evolutionary algorithm," *Applied Soft Computing,* vol. 23, pp. 407-416, 2014.

[72] F. B. de Oliveira, R. Enayatifar, H. J. Sadaei, F. G. Guimarães, and J. Y. Potvin, "A cooperative coevolutionary algorithm for the Multi-Depot Vehicle Routing Problem," *Expert Systems with Applications,* vol. 43, pp. 117-130, 2016.

[73] N. Garcia-Pedrajas, C. Hervas-Martinez, and J. Munoz-Perez, "COVNET: a cooperative coevolutionary model for evolving artificial neural networks," *IEEE Transactions on Neural Networks,* vol. 14, pp. 575-596, 2003.

[74] R. Chandra, M. Frean, and M. Zhang, "Crossover-based local search in cooperative co-evolutionary feedforward neural networks," *Applied Soft Computing,* vol. 12, pp. 2924-2932, 2012.

[75] R. Chandra and M. Zhang, "Cooperative coevolution of Elman recurrent neural networks for chaotic time series prediction," *Neurocomputing,* vol. 86, pp. 116-123, 2012.

[76] L. Zhao and F. Qian, "Tuning the structure and parameters of a neural network using cooperative binary-real particle swarm optimization," *Expert Systems with Applications,* vol. 38, pp. 4972-4977, 2011.

[77] M. D. Pérez-Godoy, A. J. Rivera, C. J. Carmona, and M. J. del Jesus, "Training algorithms for Radial Basis Function Networks to tackle learning processes with imbalanced data-sets," *Applied Soft Computing,* vol. 25, pp. 26-39, 2014.

[78] M. D. Pérez-Godoy, A. J. Rivera, F. J. Berlanga, and M. J. Del Jesus, "CO2RBFN: an evolutionary cooperative–competitive RBFN design algorithm for classification problems," *Soft Computing,* vol. 14, pp. 953-971, 2010.

[79] A. Alexandridis, E. Chondrodima, and H. Sarimveis, "Cooperative learning for radial basis function networks using particle swarm optimization," *Applied Soft Computing,* vol. 49, pp. 485-497, 2016.

[80] F. van den Bergh and A. P. Engelbrecht, "A Cooperative approach to particle swarm optimization," *IEEE Transactions on Evolutionary Computation,* vol. 8, pp. 225-239, 2004.

[81] M. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *Journal of Chemical Physics,* vol. 21, pp. 1087-1092, 1953.

[82] S. Kirkpatrick, C. D. Gelatti, and M. P. Vecchi, "Optimization by simulated annealing," *Science & Education,* vol. 220, pp. 671-680, 1983.

[83] A. Alexandridis and E. Chondrodima, "A medical diagnostic tool based on radial basis function classifiers and evolutionary simulated annealing," *Journal of Biomedical Informatics,* vol. 49, pp. 61-72, 2014.

[84] K. Pearson, "X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *Philosophical Magazine Series 5,* vol. 50, pp. 157-175, 1900.

[85] S. Boriah, V. Chandola, and V. Kumar, "Similarity Measures for Categorical Data: A Comparative Evaluation," in *8th SIAM International Conference on Data Mining,* 2008, pp. 243-254.

[86] C. Wang, X. Dong, F. Zhou, L. Cao, and C.-H. Chi, "Coupled Attribute Similarity Learning on Categorical Data," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 26, pp. 781-797, 2015.

[87] H. Jia, Y.-M. Cheung, and J. Liu, "A New Distance Metric for Unsupervised Learning of Categorical Data," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 27, pp. 1065-1079, 2016.

[88] C. Stanfill and D. Waltz, "Towards memory-based reasoning," *Communications of the ACM,* vol. 29, pp. 1213-1228, 1986.

[89] S. Cost and S. Salzberg, "A weighted nearest neighbor algorithm for learning with symbolic features," *Machine Learning,* vol. 10, pp. 57-78, 1993.

[90] D. R. Wilson and T. R. Martinez, "Improved heterogeneous distance functions," *Journal of Artificial Intelligence Research,* vol. 6, pp. 1-34, 1997.

[91] C.-C. Hsu and S.-H. Lin, "Visualized Analysis of Mixed Numeric and Categorical Data via Extended Self-Organizing Map," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 23, pp. 72-86, 2012.

[92] P. Andritsos, P. Tsaparas, R. e. J. Miller, and K. C. Sevcik, "LIMBO: Scalable Clustering of Categorical Data," in *4th Annual MITACS Conference*, Ottawa, Ontario, 2003.

[93] Y. Qian, F. Li, J. Liang, B. Liu, and C. Dang, "Space Structure and Clustering of Categorical Data," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 27, pp. 2047-2059, 2016.

[94] R. Ilin, "Unsupervised Learning of Categorical Data With Competing Models," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 23, pp. 1726-1737, 2012.

[95] T. Xiong, S. Wang, Q. Jiang, and J. Z. Huang, "A Novel Variable-order Markov Model for Clustering Categorical Sequences," *IEEE Transactions on Knowledge and Data Engineering,* vol. 26, pp. 2339-2353, 2014.

[96] L. Chen, S. Wang, K. Wang, and J. Zhu, "Soft subspace clustering of categorical data with probabilistic distance," *Pattern Recognition,* vol. 51, pp. 322-332, 2016.

[97] L. Bai and J. Liang, "The k-modes type clustering plus between-cluster information for categorical data," *Neurocomputing,* vol. 133, pp. 111-121, 2014.

[98] V. Cheng, C.-H. Li, J. T. Kwok, and C.-K. Li, "Dissimilarity learning for nominal data," *Pattern Recognition,* vol. 37, pp. 1471-1477, 2004.

[99] L. Chen and G. Guo, "Nearest neighbor classification of categorical data by attributes weighting," *Expert Systems with Applications,* vol. 42, pp. 3142-3149, 2015.

[100] F. Cao, J. Z. Huang, and J. Liang, "Trend analysis of categorical data streams with a concept change method," *Information Sciences,* vol. 276, pp. 160-173, 2014.

[101] A. Alexandridis, E. Chondrodima, N. Giannopoulos, and H. Sarimveis, "A Fast and Efficient Method for Training Categorical Radial Basis Function Networks," *IEEE Transactions on Neural Networks and Learning Systems,* In Press. DOI:10.1109/TNNLS.2016.2598722.

[102] J. Ji, T. Bai, C. Zhou, C. Ma, and Z. Wang, "An improved k-prototypes clustering algorithm for mixed numeric and categorical data," *Neurocomputing,* vol. 120, pp. 590-596, 2013.

[103] C.-C. Hsu and Y.-C. Chen, "Mining of mixed data with application to catalog marketing," *Expert Systems with Applications,* vol. 32, pp. 12-23, 2007.

[104] C.-C. Hsu, S.-H. Lin, and W.-S. Tai, "Apply extended self-organizing map to cluster and classify mixed-type data," *Neurocomputing,* vol. 74, pp. 3832-3842, 2011.

[105] J. C. Bezdek, J. Keller, and R. Krisnapuram, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. Boston: Kluwer Academic Publishers, 1999.

[106] Z. Zheng, M. Gong, J. Ma, L. Jiao, and Q. Wu, "Unsupervised evolutionary clustering algorithm for mixed type data," in *IEEE Congress on Evolutionary Computation,* 2010, pp. 1-8.

[107] S. P. Chatzis, "A fuzzy c-means-type algorithm for clustering of data with mixed numeric and categorical attributes employing a probabilistic dissimilarity functional," *Expert Systems with Applications,* vol. 38, pp. 8684-8689, 2011.

[108] G. David and A. Averbuch, "SpectralCAT: Categorical spectral clustering of numerical and nominal data," *Pattern Recognition,* vol. 45, pp. 416-433, 2012.

[109] D. Lam, M. Wei, and D. Wunsch, "Clustering Data of Mixed Categorical and Numerical Type With Unsupervised Feature Learning," *IEEE Access,* vol. 3, pp. 1605-1613, 2015.

[110] E. Chondrodima, A. Alexandridis, and H. Sarimveis, "A new algorithm for training radial basis function networks with mixed inputs," Submitted.

[111] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology,* vol. 24, pp. 417-441, 1933.

[112] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Scientific american*, vol. 284, pp. 28-37, 2001.

[113] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan, "GATE: an architecture for development of robust HLT applications," in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, Philadelphia, Pennsylvania, 2002, pp. 168-175.

[114] N. F. Noy, N. H. Shah, P. L. Whetzel, B. Dai, M. Dorf, N. Griffith, C. Jonquet, D. L. Rubin, M.-A. Storey, C. G. Chute, and M. A. Musen, "BioPortal: ontologies and integrated data resources at the click of a mouse," *Nucleic Acids Research*, vol. 37, pp. W170-W173, 2009.

[115] G. K. Savova, J. J. Masanz, P. V. Ogren, J. Zheng, S. Sohn, K. C. Kipper-Schuler, and C. G. Chute, "Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications," *Journal of the American Medical Informatics Association*, vol. 17, pp. 507-513, 2010.

[116] S. Price, "A review of the state of the art in Machine Learning on the Semantic Web," in *Proceedings of 2003 UK Workshop on Computational Intelligence*, 2004, pp. 292-299.

[117] A. Satti, N. Cercone, and V. Keselj, "Experiments in Web Page Classification for Semantic Web," in *Workshop on Web-based Support Systems*, 2004, pp. 137-141.

[118] X. Qi and B. D. Davison, "Web page classification: Features and algorithms," *ACM Comput. Surv.*, vol. 41, pp. 1-31, 2009.

[119] B. Altınel, M. Can Ganiz, and B. Diri, "A corpus-based semantic kernel for text classification by using meaning values of terms," *Engineering Applications of Artificial Intelligence*, vol. 43, pp. 54-66, 2015.

[120] E. Chondrodima, A. Alexandridis, and H. Sarimveis, "A web page classification tool for Semantic Web development," Under Preparation.

[121] *The 4 Universities Data Set*. [Online] Available: www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-20/www/data/

[122] M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam, and S. Slattery, "Learning to extract symbolic knowledge from the World Wide Web," in *Proceedings of the fifteenth national/tenth conference on Artificial intelligence/Innovative applications of artificial intelligence*, Madison, Wisconsin, USA, 1998, pp. 509-516.

[123] A. M. Turing, "Computing Machinery and Intelligence," *Mind*, vol. LIX, pp. 433-460, 1950.

[124] A. Barr and E. A. Feigenbaum, *The Handbook of Artificial Intelligence* vol. 2: Pitman Books LTd., 1983.

[125] J. C. Bezdek, "What is computational intelligence?," in *Computational Intelligence: Imitating Life*, J. Zurada, R. Marks, and C. Robinson, Eds., ed Piscataway: IEEE Press, 1994.

[126] J. Heaton, *Introduction to Neural Networks with Java*, 2 ed.: Heaton Research, Inc., 2008.

[127] M. A. Arbib, *Brains, Machines, and Mathematics*, 2 ed. New York: Springer-Verlag, 1987.

[128] S. Ramón y Cajál, *Histologie du Systéms Nerveux de l'homme et des vertébrés*. Paris: Maloine, 1911.

[129] G. M. Shepherd and C. Koch, "Introduction to synaptic circuits," in *The Synaptic Organization of the Brain*, G. M. Shepherd, Ed., ed New York: Oxford University Press, 1990, pp. 3-31.

[130] W. S. McCulloch and W. Pitts, "A Logical Calculus of Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics*, vol. 5, pp. 115-133, 1943.

[131] D. O. Hebb, *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley, 1949.

[132] M. L. Minsky, "Neural Nets and the Brain Model Problem," Doctoral Dissertation, Princeton University, Princeton, NJ, 1954.

[133] F. Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain," *Psychological Review*, vol. 65, pp. 386-408, 1958.

[134] B. Widrow and M. E. Hoff, "Adaptive Switching Circuits," *IRE WESCON Convention Record*, pp. 96-104, 1960.

[135] N. J. Nilsson, *Learning Machines: Foundations of Trainable Pattern Classifiers*. New York: McGraw Hill, 1965.

[136] M. Minsky and S. Papert, *Perceptrons*. Cambridge, MA: MIT Press, 1969.

[137] K. Fukushima and S. Miyaka, "Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position," *Biological Cybernetics*, vol. 36, pp. 193-202, 1980.

[138] T. Kohonen, *Associative Memory: A System-Theoretical Approach*. Berlin: Springer-Verlag, 1977.

[139] T. Kohonen, "A Simple Paradigm for the Self-organized Formation of Structured Feature Maps," in *Competition and Cooperation in Neural Nets*. vol. 45, S. Amari and M. Arbib, Eds., ed Berlin: Springer-Verlag, 1982.

[140] T. Kohonen, *Self-Organization and Associative Memory*. Berlin: Springer Verlag, 1984.

[141] T. Kohonen, "The Neural Phonetic Typewriter," *IEEE Computer*, vol. 27, pp. 11-22, 1988.

[142] J. A. Anderson, J. W. Silverstein, S. A. Rite, and R. S. Jones, "Distinctive Features, Categorical Perception, and Probability Learning: Some Applications of a Neural Model," *Psychological Review*, vol. 84, pp. 413-451, 1977.

[143] S. Grossberg, "Classical and Instrumental Learning by Neural Networks," in *Progress in Theoretical Biology*. vol. 3, ed New York: Academic Press, 1977, pp. 51-141.

[144] S. Grossberg, *Studies of Mind and Brain: Neural Principles of Learning Perception, Development, Cognition, and Motor Control*. Boston: Reidel Press, 1982.

[145] J. J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," in *Proceedings of the National Academy of Sciences*, USA, 1982, pp. 2554-58.

[146] J. J. Hopfield, "Neurons with Graded Response Have Collective Computational Properties Like Those of Two State Neurons," in *Proceedings of the National Academy of Sciences*, USA, 1984, pp. 3088-3092.

[147] D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* vol. 1. Cambridge, MA: MIT Press, 1986.

[148] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel distributed processing: explorations in the microstructure of cognition*. vol. 1, D. E. Rumelhart and J. L. McClelland, Eds., ed: MIT Press, 1986, pp. 318-362.

[149] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Representations of back-Propagation errors," *Nature (London)* vol. 323, pp. 533–536, 1986.

[150] P. J. Werbos, "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences," Doctoral Dissertation, Applied Mathematics, Harvard University, MA, 1974.

[151] S. E. Dreyfus, "Artificial Neural Networks, Back Propagation and the Kelley-Bryson Gradient Procedure," *Journal of Guidance, Control, and Dynamics,* vol. 13, pp. 926-928, 1990.

[152] A. E. Bryson and Y. C. Ho, *Applied Optimal Control*. Waltham, MA: Blaisdell, 1969.

[153] A. E. Bryson, "A Gradient Method for Optimizing Multi-Stage Allocation Processes," in *Proceedings of the Harvard University Symposium on Digital Computers and Their Applications*, 1961.

[154] H. J. Kelley, "Gradient Theory of Optimal Flight Path," *ARS Journal,* vol. 30, pp. 947-954, 1960.

[155] G. E. Hinton, "Connectionist learning procedures," *Artificial Intelligence,* vol. 40, pp. 185-234, 1989.

[156] D. B. Parker, "Optimal algorithms for adaptive networks: second order back propagation, second order direct propagation, and second order Hebbian learning," in *IEEE First International Conference on Neural Networks*, San Diego, CA, 1987, pp. 593-600.

[157] R. J. Williams and J. Peng, "An efficient gradient-based algorithm for on-line training of recurrent network trajectories," *Neural Computation,* vol. 2, pp. 490-501, 1990.

[158] H. Ritter, "Self-organising feature maps: Kohonen maps," in *The Hanbook of Brain Theory and Neural Networks*, M. A. Arbib, Ed., ed Cambridge, MA: MIT Press, 1995, pp. 846-851.

[159] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics,* vol. 43, pp. 59-69, 1982.

[160] T. Kohonen, "The Self-Organizing Map," in *Proceedings of the IEEE,* 1990, pp. 1464-1480.

[161] T. Kohonen, "Exploration of very large databases by self-organising maps," in *1997 International Conference on Neural Networks*, Houston, 1997, pp. PL1-PL6.

[162] S. P. Luttrell, "Self-organization: A derivation from first principles of a class of learning algorithms," in *IEEE Conference on Neural Networks*, Washington, DC, 1989, pp. 495-498.

[163] S. P. Luttrell, "Code vector density in topographic mappings: Scalar case," *IEEE Transactions on Neural Networks,* vol. 2, 1991.

[164] J. M. Mendel and R. W. McLaren, "Reinforcement learning control and pattern recognition systems," in *Adaptive, Learning and Pattern Recognition Systems: Theory and Applications*. vol. 66, J. M. Mendel and K. S. Fu, Eds., ed New York: Academic Press, 1970, pp. 287-318.

[165] Y. Lean, W. Shouyang, and K. K. Lai, "An integrated data preparation scheme for neural network data analysis," *IEEE Transactions on Knowledge and Data Engineering,* vol. 18, pp. 217-230, 2006.

[166] B. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," *Biochimica et Biophysica Acta - Protein Structure,* vol. 405, pp. 442-451, 1975.

[167] J. Gorodkin, "Comparing two K-category assignments by a K-category correlation coefficient," *Computational Biology and Chemistry,* vol. 28, pp. 367-374, 2004.

[168] G. Jurman and C. Furlanello, "A unifying view for performance measures in multi-class prediction," *PLoS One,* vol. 7, p. e41882, 2012.

[169] L. A. Zadeh, "Fuzzy sets," *Information and Control,* vol. 8, pp. 338-353, 1965.

[170] A. S. Fraser, "Simulation of Genetic Systems by Automatic Digital Computers I. Introduction," *Australian Journal of Biological Sciences,* vol. 10, pp. 484-491, 1957.

[171] A. S. Fraser, "Simulation of Genetic Systems by Automatic Digital Computers II. Effects of Linkage on Rates of Advance Under Selection," *Australian Journal of Biological Sciences,* vol. 10, pp. 492-500, 1957.

[172] H. J. Bremermann, "Optimization through evolution and recombination," in *Self-Organizing Systems,* 1962, pp. 93-106.

[173] J. Reed, R. Toombs, and N. A. Barricelli, "Simulation of biological evolution and machine learning," *Journal of Theoretical Biology,* vol. 17, pp. 319-342, 1967.

[174] I. Rechenberg, "Cybernetic Solution Path of an Experimental Problem," Technical report, Ministery of Aviation, 1965.

[175] I. Rechenberg, *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Verlag, Stuttgart Frommann-Holzboog, 1973.

[176] H. P. Schwefel, "Evolutionsstrategie und numerische Optimierung," PhD thesis, Technical University Berlin, 1975.

[177] H. P. Schwefel, *Numerical Optimization of Computer Models*. Chichester: John Wiley & Sons, 1981.

[178] L. J. Fogel, A. Owens, and M. Walsh, *Artificial Intelligence through Simulated Evolution*: John Wiley & Sons, 1966.

[179] J. H. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press, 1975.

[180] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man, and Cybernetics Part B,* vol. 26, pp. 29-41, 1996.

[181] D. Karaboga, B. Gorkemli, C. Ozturk, and N. Karaboga, "A comprehensive survey: artificial bee colony (ABC) algorithm and applications," *Artificial Intelligence Review,* vol. 42, pp. 21-57, 2014.

[182] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in *IEEE International Conference on Evolutionary Computation Proceedings*, 1998, pp. 69-73.

[183] R. Eberhart and Y. Shi, "Particle swarm optimization: developments, applications and resources," in *Proceedings of the 2001 Congress on Evolutionary Computation*, 2001, pp. 81-86.

[184] T. M. Cover, "Geometrical and Statistical properties of systems of linear inequalities with applications in pattern recognition," *IEEE Transactions on Electronic Computers,* vol. EC-14, pp. 326–334, 1965.

[185] H. N. Mhaskar, "Neural networks for optimal approximation of smooth and analytic functions," *Neural Computation,* vol. 8, pp. 1731-1742, 1996.

[186] P. Niyogi and F. Girosi, "On the relationship between generalization errors hypothesis complexity and sample complexity for radial basis functions," *Neural Computation,* vol. 8, pp. 819-842, 1996.

[187] M. R. Garrey and D. S. Johnson, *Computers and Intractability: A guide to the theory of NP-completeness*: W. H. Freeman, 1979.

[188] P. Geladi and B. R. Kowalski, "Partial Least Squares Regression A Tutorial," *Analytica Chimica Acta,* vol. 185, pp. 1-17, 1986.

[189] J. Nie, "Fuzzy control of Multivariable Nonlinear Servomechanisms with Explicit Decoupling Scheme," *IEEE Transactions on Fuzzy Systems,* vol. 5, pp. 304-311, 1997.

[190] A. Asuncion and D. J. Newman, *UCI Machine Learning Repository, University of California, Irvine, CA [Online]. Available:* http://archive.ics.uci.edu/ml/, 2007.

[191] G. E. P. Box and G. M. Jenkins, *Time series analysis: Forecasting and control*. San. Francisco: Holden-Day, 1970.

[192] K. Kilic, B. A. Sproule, I. B. Türksen, and C. A. Naranjo, "A fuzzy system modeling algorithm for data analysis and approximate reasoning," *Robotics and Autonomous Systems,* vol. 49, 2004.

[193] J. Friedman, "Multivariate adaptive regression splines," *Annals of Statistics,* vol. 19, pp. 1–67, 1991.

[194] Z. Shi and M. Han, "Support Vector Echo-State Machine for Chaotic Time-Series Prediction," *IEEE Transactions on Neural Networks,* vol. 18, pp. 359-372, 2007.

[195] C. C. Holmes and B. K. Mallick, "Bayesian Wavelet Networks for Nonparametric Regression," *IEEE Transactions on Neural Networks,* vol. 11, pp. 27-35, 2000.

[196] C. K. Loo, M. Rajeswari, and M. V. C. Rao, "Novel Direct and Self-Regulating Approaches to Determine Optimum Growing Multi-Experts Network Structure," *IEEE Transactions on Neural Networks,* vol. 15, pp. 1378-1395, 2004.

[197] T. Samad, "Backpropagation with expected source values," *Neural Networks,* vol. 4, pp. 615–618, 1991.

[198] B. L. Sturm, "An analysis of the GTZAN music genre dataset," in *Proceedings of the 2nd International ACM workshop on Music information retrieval with user-centered and multimodal strategies*, 2012, pp. 7-12.

[199] L. R. Rabiner and B.-H. Juang, *Fundamentals of speech recognition* vol. 14: PTR Prentice Hall Englewood Cliffs, 1993.

[200] M. Potter and K. De Jong, "A cooperative coevolutionary approach to function optimization," in *Parallel Problem Solving from Nature — PPSN III*. vol. 866, Y. Davidor, H.-P. Schwefel, and R. Männer, Eds., ed: Springer Berlin Heidelberg, 1994, pp. 249-257.

[201] F. van den Bergh and A. P. Engelbrecht, "Cooperative learning in neural networks using particle swarm optimizers," *South African Computer Journal,* vol. 26, pp. 84-90, 2000.

[202] A. Tsanas and A. Xifara, "Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools," *Energy and Buildings,* vol. 49, pp. 560-567, 2012.

[203] A. Tsanas, M. Little, P. McSharry, and L. Ramig, "Accurate Telemonitoring of Parkinson's Disease Progression by Noninvasive Speech Tests," *IEEE Transactions on Biomedical Engineering,* vol. 57, pp. 884-893, 2010.

[204] P. B. Silva, A. S. Marçal, and R. A. da Silva, "Evaluation of Features for Leaf Discrimination," in *Image Analysis and Recognition*. vol. 7950, M. Kamel and A. Campilho, Eds., ed: Springer Berlin Heidelberg, 2013, pp. 197-204.

[205] I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective," *Artificial Intelligence in Medicine,* vol. 23, pp. 89-109, 2001.

[206] R. Dua, D. G. Beetner, W. V. Stoecker, and D. C. Wunsch, "Detection of basal cell carcinoma using electrical impedance and neural networks," *IEEE Transactions on Biomedical Engineering,* vol. 51, pp. 66-71, 2004.

[207] H. Q. Wang, H. S. Wong, H. Zhu, and T. T. Yip, "A neural network-based biomarker association information extraction approach for cancer classification," *Journal of Biomedical Informatics,* vol. 42, pp. 654-666, 2009.

[208] A. Statnikov, I. Tsamardinos, Y. Dosbayev, and C. F. Aliferis, "GEMS: A system for automated cancer diagnosis and biomarker discovery from microarray gene expression data," *International Journal of Medical Informatics,* vol. 74, pp. 491-503, 2005.

[209] J. Nahar, T. Imam, K. S. Tickle, and Y.-P. Phoebe Chen, "Computational intelligence for heart disease diagnosis: A medical knowledge driven approach," *Expert Systems with Applications,* vol. 40, pp. 96-104, 2013.

[210] M. A. Mazurowski, P. A. Habas, J. M. Zurada, J. Y. Lo, J. A. Baker, and T. G. D, "Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance," *Neural Networks,* vol. 21 pp. 427-436, 2008.

[211] Y. W. Peng, Zhiqing and J. Jiang, "A novel feature selection approach for biomedical data classification," *Journal of Biomedical Informatics,* vol. 43, pp. 15-23, 2010.

[212] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," *Journal of Biomedical Informatics,* vol. 35, pp. 352-359, 2002.

[213] J. Lakoumentas, J. Drakos, M. Karakantza, G. C. Nikiforidis, and G. C. Sakellaropoulos, "Bayesian clustering of flow cytometry data for the diagnosis of B-Chronic Lymphocytic Leukemia," *Journal of Biomedical Informatics,* vol. 42, pp. 251-261, 2009.

[214] D. Zhao and C. Weng, "Combining PubMed knowledge and EHR data to develop a weighted bayesian network for pancreatic cancer prediction," *Journal of Biomedical Informatics,* vol. 44, pp. 859-868, 2011.

[215] S. Tortajada, E. Fuster-Garcia, J. Vicente, P. Wesseling, F. A. Howe, M. Julià-Sapé, A. P. Candiota, D. Monleón, A. Moreno-Torres, J. Pujol, J. R. Griffiths, A. Wright, A. C. Peet, M. C. Martínez-Bisbal, B. Celda, C. Arús, M. Robles, and J. M. García-Gómez, "Incremental Gaussian Discriminant Analysis based on Graybill and Deal weighted combination of estimators for brain tumour diagnosis," *Journal of Biomedical Informatics,* vol. 44, pp. 677-687, 2011.

[216] M. R. Akbarzadeh-T and M. Moshtagh-Khorasani, "A hierarchical fuzzy rule-based approach to aphasia diagnosis," *Journal of Biomedical Informatics,* vol. 40, pp. 465-475, 2007.

[217] C. S. Son, Y. N. Kim, H. S. Kim, H. S. Park, and M. S. Kim, "Decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches " *Journal of Biomedical Informatics,* vol. 45, pp. 999-1008, 2012.

[218] T. F. Cova, J. L. Pereira, and A. A. Pais, "Is standard multivariate analysis sufficient in clinical and epidemiological studies?," *Journal of Biomedical Informatics,* vol. 46, pp. 75-86, 2013.

[219] G. Sateesh Babu and S. Suresh, "Parkinson's disease prediction using gene expression - A projection based learning meta-cognitive neural classifier approach," *Expert Systems with Applications,* vol. 40, pp. 1519-1529, 2013.

[220] S. Ghosh-Dastidar, H. Adeli, and N. Dadmehr, "Principal Component Analysis-Enhanced Cosine Radial Basis Function Neural Network for Robust Epilepsy and Seizure Detection," *IEEE Transactions on Biomedical Engineering,* vol. 55, pp. 512-518, 2008.

[221] C. H. Wei, Y. Li, and P. J. Huang, "Mammogram retrieval through machine learning within BI-RADS standards," *Journal of Biomedical Informatics,* vol. 44, pp. 607-614, 2011.

[222] I. Ilhan and G. Tezel, "A genetic algorithm-support vector machine method with parameter optimization for selecting the tag SNPs," *Journal of Biomedical Informatics,* vol. 46, pp. 328-340, 2013.

[223] A. Lambrou, H. Papadopoulos, and A. Gammerman, "Reliable Confidence Measures for Medical Diagnosis With Evolutionary Algorithms," *IEEE Transactions on Information Technology in Biomedicine,* vol. 15, pp. 93-99, 2011.

[224] M. Esposito, I. De Falco, and G. De Pietro, "An evolutionary-fuzzy DSS for assessing health status in multiple sclerosis disease," *International Journal of Medical Informatics,* vol. 80, pp. e245-e254, 2011.

[225] E. Avci, "A new intelligent diagnosis system for the heart valve diseases by using genetic-SVM classifier," *Expert Systems with Applications,* vol. 36, pp. 10618-10626, 2009.

[226] H. Sarimveis, P. Doganis, and A. Alexandridis, "A classification technique based on radial basis function neural networks," *Advances in Engineering Software,* vol. 37, pp. 218-221, 2006.

[227] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A Training Algorithm for Optimal Margin Classifiers," in *5th Annual ACM Workshop on Computational Learning Theory,* 1992, pp. 144-152.

[228] V. Vapnik, *The Nature of Statistical Learning Theory*. London: Springer-Verlag, 1995.

[229] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," ed: Department of Computer Science, National Taiwan University, 2003.

[230] T. Di Noia, V. C. Ostuni, F. Pesce, G. Binetti, D. Naso, F. P. Schena, and E. Di Sciascio, "An end stage kidney disease predictor based on an artificial neural networks ensemble," *Expert Systems with Applications,* vol. 40, pp. 4438-4445, 2013.

[231] J. Luengo, S. García, and F. Herrera, "A study on the use of imputation methods for experimentation with Radial Basis Function Network classifiers handling missing attribute values: The good synergy between RBFNs and EventCovering method," *Neural Networks,* vol. 23, pp. 406-418, 2010.

[232] K. S. Jones, "A statistical interpretation of term specificity and its application in retrieval," in *Document retrieval systems,* W. Peter, Ed., ed: Taylor Graham Publishing, 1988, pp. 132-142.

[233] C. Renjifo, D. Barsic, C. Carmen, K. Norman, and G. S. Peacock, "Improving radial basis function kernel classification through incremental learning and automatic parameter selection," *Neurocomputing,* vol. 72, pp. 3-14, 2008.

[234] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms* vol. 6. Cambridge: MIT Press, 2001.

[235] G. H. John and P. Langley, "Estimating Continuous Distributions in Bayesian Classifiers," in *Eleventh Conference on Uncertainty in Artificial Intelligence,* San Mateo, 1995, pp. 338-345.

[236] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, "Classification and Regression Trees," in *Wadsworth International Group,* ed Belmont, California, 1984.

[237] L. Breiman, "Random Forests," *Machine Learning,* vol. 45, pp. 5-32, 2001.

[238] A. Liaw and W. Matthew, "Classification and regression by random Forest," *R news*, vol. 2, pp. 18-22, 2002.

[239] M. Marbac, C. Biernacki, and V. Vandewalle, "Model-based clustering of Gaussian copulas for mixed data," *Communications in Statistics - Theory and Methods*, 2017, In Press.

[240] D. McParland and I. C. Gormley, "Model based clustering for mixed data: clustMD," *Advances in Data Analysis and Classification*, vol. 10, pp. 155-169, 2016.

[241] D. Dong and T. J. McAvoy, "Nonlinear principal component analysis—Based on principal curves and neural networks," *Computers & Chemical Engineering*, vol. 20, pp. 65-78, 1996.

[242] Open Government Licence. [Online]. Available: https://data.gov.uk/.

[243] H. Fanaee-T and J. Gama, "Event labeling combining ensemble detectors and background knowledge," *Progress in Artificial Intelligence*, vol. 2, pp. 113-127, 2014.

[244] GB Road Traffic Counts. Open Government Licence. [Online]. Available: https://data.gov.uk/dataset/gb-road-traffic-counts

[245] Local police recorded crime data. Open Government Licence. [Online]. Available: https://data.gov.uk/dataset/local-police-recorded-crime-data

[246] P. Cortez and A. Silva, "Using Data Mining to Predict Secondary School Student Performance," in *Proceedings of 5th Future Business Technology Conference*, Porto, Portugal, 2008, pp. 5-12.

[247] B. Antal and A. Hajdu, "An ensemble-based system for automatic screening of diabetic retinopathy," *Knowledge-Based Systems*, vol. 60, pp. 20-27, 2014.

[248] K. Mansouri, T. Ringsted, D. Ballabio, R. Todeschini, and V. Consonni, "Quantitative Structure–Activity Relationship Models for Ready Biodegradability of Chemicals," *Journal of Chemical Information and Modeling*, vol. 53, pp. 867-878, 2013.

[249] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, pp. 489-501, 2006.

[250] V. Pornsakulvanich, P. Haridakis, and A. M. Rubin, "The influence of dispositions and Internet motivation on online communication satisfaction and relationship closeness," *Computers in Human Behavior*, vol. 24, pp. 2292-2310, 2008.

[251] P. Marks, "Interview: Tim Berners-Lee," *New Scientist*, vol. 202, p. 29, 2009.

[252] R. A. Sindhu and R. M. Chezian, "The Movement of Web from Web 0.0 to Web 5.0: A Comparative Study," *International Journal of Multidisciplinary Research and Development*, vol. 3, pp. 176-179, 2016.

[253] M. Pieri and D. Diamantini, "An E-learning Web 2.0 Experience," *Procedia - Social and Behavioral Sciences*, vol. 116, pp. 1217-1221, 2014.

[254] A. Banu, S. S. Fatima, and K. U. R. Khan, "Building OWL Ontology: LMSO-Library Management System Ontology," in *Advances in Computing and Information Technology*, ed: Springer, 2013, pp. 521-530.

[255] G. Kuck, "Tim Berners-Lee's Semantic Web," *SA Journal of Information Management*, vol. 6, 2004.

[256] T. R. Gruber, "A Translation Approach to Portable Ontology Specifications," *Knowledge Acquisition*, vol. 5, pp. 199-220, 1993.

[257] G. Solomou, A. Kouneli, and A. Kameas, "Using ontologies for modeling knowledge domains in distance learning," in *6th International Conference on Open and Distance Learning*, Loutraki, Greece, 2011.

[258] T. Berners-Lee, "Semantic web-xml2000," http://www.w3.org/2000/Talks/1206-xml2k-tbl/, 2000.

[259] D. J. Hand, P. Smyth, and H. Mannila, *Principles of data mining*: MIT Press, 2001.

[260] M. Rajman and R. Besançon, "Text Mining - Knowledge extraction from unstructured textual data," in *6th Conference of International Federation of Classification Societies*, Rome, 1998, pp. 473-480.

[261] R. J. Mooney and U. Y. Nahm, "Text mining with information extraction," in *Multilingualism and electronic language management : proceedings of the 4th International MIDP Colloquium*. vol. 4, W. Daelemans, T. du Plessis, C. Snyman, and L. Teck, Eds., ed Bloemfontein, South Africa: Van Schaik Pub., 2003, pp. 141-160.

[262] S. V. Gaikwad, A. Chaugule, and P. Patil, "Text mining methods and techniques," *International Journal of Computer Applications*, vol. 85, pp. 42-45, 2014.

[263] L. Jena and N. K. Kamila, "Data Extraction and Web page Categorization using Text Mining," *International Journal of Application or Innovation in Engineering & Management*, vol. 2, pp. 67-76, 2013.

[264] S. Dumais, J. Platt, D. Heckerman, and M. Sahami, "Inductive learning algorithms and representations for text categorization," in *Proceedings of the seventh international conference on Information and knowledge management*, Bethesda, Maryland, USA, 1998, pp. 148-155.

[265] C. Apté, F. Damerau, and S. M. Weiss, "Automated learning of decision rules for text categorization," *ACM Transactions on Information Systems*, vol. 12, pp. 233-251, 1994.

[266] M. F. Porter, "An algorithm for suffix stripping," in *Readings in information retrieval*, J. Karen Sparck and W. Peter, Eds., ed: Morgan Kaufmann Publishers Inc., 1997, pp. 313-316.

[267] L. Guthrie, E. Walker, and J. Guthrie, "Document classification by machine: theory and practice," in *Proceedings of the 15th conference on Computational linguistics - Volume 2*, Kyoto, Japan, 1994, pp. 1059-1063.

[268] G. Salton, A. Wong, and C. S. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, vol. 18, pp. 613-620, 1975.

[269] K. Nigam, A. K. Mccallum, S. Thrun, and T. Mitchell, "Text Classification from Labeled and Unlabeled Documents using EM," *Machine Learning*, vol. 39, pp. 103-134, 2000.