



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

**Υβριδικό Πιθανοτικό Δίκτυο Μη Επιβλεπόμενης Μάθησης
για την Ανάλυση Ακολουθιών Συμβόλων και Βιολογικών
Αλληλουχιών**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Χρήστος Ι. Φερλές

Αθήνα, Μάρτιος 2014



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Υβριδικό Πιθανοτικό Δίκτυο Μη Επιβλεπόμενης Μάθησης για την Ανάλυση Ακολουθιών Συμβόλων και Βιολογικών Αλληλουχιών

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Χρήστος Ι. Φερλές

Συμβουλευτική Επιτροπή: Ανδρέας-Γεώργιος Σταφυλοπάτης
Στέφανος Κόλλιας
Παναγιώτης Τσανάκας

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την 7^η Μαρτίου 2014.

.....
Α.-Γ. Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

.....
Στ. Κόλλιας
Καθηγητής Ε.Μ.Π.

.....
Π. Τσανάκας
Καθηγητής Ε.Μ.Π.

.....
Κ. Νικήτα
Καθηγήτρια Ε.Μ.Π.

.....
Γ. Στάμου
Επίκουρος Καθηγητής Ε.Μ.Π.

.....
Γ. Ματσόπουλος
Επίκουρος Καθηγητής Ε.Μ.Π.

.....
Κ. Μπλέκας
Επίκουρος Καθηγητής
Πανεπιστημίου Ιωαννίνων

Αθήνα, Μάρτιος 2014

.....
Χρήστος Ι. Φερλές

Υποψήφιος Διδάκτωρ Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Χρήστος Ι. Φερλές, 2014.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

ΠΕΡΙΛΗΨΗ

Υβριδικά δίκτυα μηχανικής μάθησης μπορούν να σχηματιστούν από τη συνένωση τεχνικών μη επιβλεπόμενης μάθησης και πιθανοτικών προσεγγίσεων, σε μία προσπάθεια να συνδυαστούν τα πλεονεκτήματά τους και να ξεπεραστούν ορισμένοι από τους περιορισμούς τους. Η παρούσα διδακτορική διατριβή εισαγάγει μία τέτοια υβριδική προσέγγιση η οποία συνδυάζει τους αυτο-οργανούμενους χάρτες - Self-Organizing Maps (SOMs) με τα κρυφά μοντέλα Markov - Hidden Markov Models (HMMs). Ο αυτο-οργανούμενος χάρτης κρυφών μοντέλων Markov - Self-Organizing Hidden Markov Model Map (SOHMMM) είναι μία τομή ανάμεσα στις θεωρητικές βάσεις και στις αλγοριθμικές πραγματώσεις των δύο συστατικών στοιχείων του. Αμφότερες οι αρχιτεκτονικές των δύο συνιστωσών του συγχωνεύονται. Ο λειτουργικός πυρήνας του SOHMMM αποτελείται από ένα νέο ενοποιημένο SOM-HMM αλγόριθμο. Η ένωση και η συνέργια των μεθοδολογιών της μη επιβλεπόμενης εκπαίδευσης του SOM και του δυναμικού προγραμματισμού του HMM παράγουν μία κλάση στοχαστικών αλγορίθμων μη επιβλεπόμενης μάθησης, οι οποίοι ενσωματώνονται πλήρως με το SOHMMM. Τα κύρια πλεονεκτήματα του προτεινόμενου συστήματος είναι η ικανότητα του να εκμεταλλεύεται την ενυπάρχουσα αλλά αφανή πληροφορία που κρύβεται στις χωροχρονικές συσχετίσεις των στοιχείων των δεδομένων, καθώς και το γεγονός ότι απαιτεί ελάχιστη, ή ακόμη και καθόλου, εκ των προτέρων γνώση που να σχετίζεται με το εκάστοτε υπό εξέταση πρόβλημα μοντελοποίησης.

Σε μία πιο πραγματιστική βάση, μπορεί να υποστηριχθεί ότι οι αντίστοιχες αρχιτεκτονικές και οι επιμέρους μεθοδολογίες μάθησης συνενώνονται σε μία προσπάθεια να ικανοποιηθούν οι αυξανόμενες απαιτήσεις που πηγάζουν από τα DNA, RNA και πρωτεϊνικά μόρια. Η αντιμετώπιση πολλών προβλημάτων ανάλυσης βιολογικών αλληλουχιών επιτυγχάνεται μέσω του προκύπτοντος αυτόματου μηχανισμού μάθησης από ανεπεξέργαστα δεδομένα. Εξαιτίας του γεγονότος ότι το SOHMMM μπορεί να φέρει εις πέρας αναλύσεις ακολουθιών και αλληλουχιών, απαιτώντας ελάχιστη ή μηδενική εκ των προτέρων γνώση, μπορεί να έχει μία σειρά εφαρμογών στην ομαδοποίηση, στην μείωση διαστατικότητας και στην οπτικοποίηση συστάδων ακολουθιών μεγάλης κλίμακας, και επιπλέον, υπό συγκεκριμένες προϋποθέσεις, στην αναζήτηση και στην κατηγοριοποίηση τους. Τρεις εκτενείς σειρές πειραμάτων, βασιζόμενες σε τεχνητά σύνολα ακολουθιών, στην πρωτεϊνική οικογένεια

των σφαιρινών και στα splice junctions αλληλουχιών γονιδίων, επιδεικνύουν τα χαρακτηριστικά και τις δυνατότητες του SOHMMM.

Λέξεις Κλειδιά

Μηχανική Μάθηση, Νευρωνικά Δίκτυα, Υπολογιστική Νοημοσύνη, Πιθανοτικοί Αλγόριθμοι Μη Επιβλεπόμενης Μάθησης, Αυτο-Οργανούμενοι Χάρτες, Κρυφά Μοντέλα Markov, Αυτόνομη, Χωροχρονική, Ομαδοποίηση, Μείωση Διαστατικότητας, Οπτικοποίηση Δεδομένων Υψηλών Διαστάσεων, Μη Γραμμική Προβολή, Κατηγοριοποίηση, Μη Αριθμητικές/Συμβολικές Ακολουθίες, Βιοπληροφορική, Υπολογιστική Μοριακή Βιολογία, DNA, RNA, Γονίδια, Πρωτεΐνες, Βιολογικά Μόρια.

ABSTRACT

Hybrid machine learning networks can be formed by merging unsupervised learning/self-organizing techniques and probabilistic approaches, in an effort to combine their advantages and overcome certain of their limitations. The present PhD thesis introduces such a hybrid approach that combines the Self-Organizing Map (SOM) and the Hidden Markov Model (HMM). The Self-Organizing Hidden Markov Model Map (SOHMMM) establishes a cross-section between the theoretic foundations and algorithmic realizations of its two constituents. Both components' corresponding architectures are fused. The SOHMMM's functional core consists of a novel unified SOM-HMM algorithm. The fusion and synergy of the SOM unsupervised training and the HMM dynamic programming methodologies bring forth a class of stochastic unsupervised learning algorithms, which are fully integrated into the SOHMMM. The principal advantages of the proposed system are its capability to exploit latent information hidden in the spatiotemporal correlations of data elements, and the fact that it necessitates little, or even no, additional prior information and domain knowledge on the modeling problem at hand.

On a more pragmatic basis, the respective architectures and learning methodologies are merged in an attempt to meet the increasing requirements imposed by the properties of DeoxyriboNucleic Acid (DNA), RiboNucleic Acid (RNA), and protein chain molecules. Addressing many biological sequence analysis problems is achieved through the resulting automatic raw sequence data learning mechanism. Since the SOHMMM carries out probabilistic sequence analysis with little or no prior knowledge, it can have a variety of applications in clustering, dimensionality reduction and visualization of large-scale sequence spaces, and also, in sequence discrimination, search and classification. Three series of experiments based on artificial sequence data, the globin protein family and splice junction gene sequences demonstrate the SOHMMM's characteristics and capabilities.

Keywords

Machine Learning, Neural Networks, Computational Intelligence, Probabilistic Unsupervised Learning Algorithms, Self-Organizing Maps (SOMs), Hidden Markov Models (HMMs), Autonomous, Spatiotemporal, Clustering, Dimensionality Reduction,

High-Dimensional Data Visualization, Nonlinear Projection, Classification, Non-Numerical/Symbolic Sequences, Bioinformatics, Computational Molecular Biology, DNA, RNA, Genes, Proteins, Biological Molecules.

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα διδακτορική διατριβή εκπονήθηκε εν μέρει στο εργαστήριο Ευφών Υπολογιστικών Συστημάτων και εν μέρει σε περιοχές του θαλάσσιου συνόρου μεταξύ αμνησίας και αδημονίας.

Η έρευνα στο πεδίο της μηχανικής μάθησης καθώς και η επακόλουθη εκπόνηση της διατριβής έγινε υπό την επίβλεψη του καθηγητή κ. Ανδρέα-Γεωργίου Σταφυλοπάτη. Τον ευχαριστώ εκ βαθέων για την απόλυτη και ουσιαστική ελευθερία που μου παρείχε η ακαδημαϊκή πτυχή της οποίας αποτέλεσε μόνο ένα τμήμα της συνολικής. Επίσης τον ευχαριστώ θερμά για την πολύπλευρη επιστημονική καθοδήγηση, τις ποικίλες ευκαιρίες, τις καίριες συμβουλές και την πραγματική στήριξη που ανιδιοτελώς μου προσέφερε σε κάθε στάδιο αλλά και ανά πάσα χρονική στιγμή.

Επίσης, θα ήθελα να εκφράσω τις ευχαριστίες μου στον καθηγητή κ. Στ. Κόλλια, για την υποστήριξη που μου παρείχε σε κρίσιμες συγκυρίες, και τον καθηγητή κ. Π. Τσανάκα, μέλη της τριμελούς συμβουλευτικής επιτροπής. Τέλος, ευχαριστώ τα μέλη της εξεταστικής επιτροπής, την καθηγήτρια κα Κ. Νικήτα, τον επίκουρο καθηγητή κ. Γ. Στάμου, τον επίκουρο καθηγητή κ. Γ. Ματσόπουλο και –ιδιαιτέρως- τον επίκουρο καθηγητή του Πανεπιστημίου Ιωαννίνων κ. Κ. Μπλέκα, για την τιμή που μου έκαναν να συμμετάσχουν στην αξιολόγηση αυτής της διατριβής.

Χρήστος Φερλές
Αφιάρτης, Μάρτιος 2014

ΠΕΡΙΕΧΟΜΕΝΑ

Κατάλογος Σχημάτων	13
Κατάλογος Πινάκων	15
1 Εισαγωγικό Σημείωμα	17
2 Σχετικές Προσεγγίσεις	23
3 Μαθηματικό Υπόβαθρο και Προσπαιτούμενα	31
3.1 Αλυσίδα Markov	31
3.2 Παρατηρήσιμη Στοχαστική Διαδικασία	32
3.3 Ανεξαρτησία Υπό Συνθήκη	33
3.4 Αλγόριθμος Forward-Backward	33
3.5 Συμπερασμός Bayes και Εκτίμηση Παραμέτρων	37
3.6 Gradient Descent	40
3.7 Εκθετική Επαναπαραμετροποίηση	41
3.8 Παράγωγοι της Πιθανοφάνειας	43
4 Το Πρότυπο SOHMMM	59
4.1 Γενική Θεώρηση	59
4.2 Αλγόριθμοι Εκπαίδευσης για Αυτο-Οργανούμενα Δίκτυα	62
4.3 Στοχαστικός Αλγόριθμος Εκπαίδευσης του SOHMMM	63
4.3.1 Διαδικασία ανταγωνισμού	64
4.3.2 Διαδικασία Συνεργασίας	65
4.3.3 Διαδικασία Προσαρμογής	66
4.4 Τροποποιημένος Αλγόριθμος Εκπαίδευσης του SOHMMM	68
4.5 Αλγοριθμικά Ζητήματα Υλοποίησης του SOHMMM	71
4.5.1 Αρχιτεκτονική των HMM Νευρώνων	72
4.5.2 Αρχικοποίηση	74
4.5.3 Σπανίως Παρατηρήσιμα Σύμβολα	75
4.5.4 Διφορούμενα Σύμβολα	75

4.5.5	Μετατροπή Κλίμακας	76
4.5.6	Τύπος και Μορφή του Πλέγματος Νευρώνων.....	81
4.5.7	Συνάρτηση Τοπολογικής Γειτονιάς	82
4.5.8	Προσδιορισμός του Ρυθμού Μάθησης	82
4.5.9	Περιορισμένος Αριθμός Ακολουθιών Παρατηρήσεων.....	83
4.5.10	Χαρακτηρισμός Περιοχών του Χάρτη.....	84
4.5.11	Προσέγγιση Πυκνότητας	84
4.5.12	Επιβάλλοντας Αναπαραστάσεις σε Περιοχές του Χάρτη.....	85
4.5.13	Σπάνιες Περιπτώσεις.....	85
4.5.14	On-Line και Batch Μάθηση.....	85
4.5.15	Υπολογιστική Πολυπλοκότητα.....	87
4.5.16	Παραδείγματα Αλγορίθμων	87
4.6	Τεχνικές Οπτικοποίησης και Γραφικής Απεικόνισης.....	91
4.6.1	Απεικόνιση Πυκνότητας Αλληλουχιών	92
4.6.2	Προβολή Πιθανοφάνειας Αλληλουχίας.....	94
5	Πειράματα και Εφαρμογές	97
5.1	Τεχνητές Ακολουθίες Παρατηρήσεων.....	97
5.2	Πρωτεϊνική Οικογένεια Σφαιρινών.....	103
5.2.1	Βασική Διερεύνηση	103
5.2.2	Περιορισμένος Αριθμός Δεδομένων Εκπαίδευσης.....	115
5.2.3	Συγκριτικά Αποτελέσματα.....	118
5.3	Splice Junctions Ακολουθιών Γονιδίων.....	121
6	Επίλογος	131
	Βιβλιογραφία	135

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

4 Το Πρότυπο SOHMMM

- 4.1 Χαρακτηριστικό παράδειγμα πλέγματος SOHMMM δύο διαστάσεων.
- 4.2 Παράδειγμα τοπολογικής γειτονιάς για διαφορετικά στιγμιότυπα του διακριτού χρόνου ($y_1 < y_2$).

5 Πειράματα και Εφαρμογές

- 5.1 Χαρακτηριστικά πλέγματα SOHMMM όπου κάθε HMM νευρώνας αναπαρίσταται από ένα εξάγωνο. Ένα τυχαία αρχικοποιημένο (ανεκπαίδευτο) SOHMMM εμφανίζεται στην (a). Το ίδιο SOHMMM μετά το πέρας του στοχαστικού αλγορίθμου μάθησης παρουσιάζεται στην (b). Διαφορετικές αποχρώσεις του γκρι χρησιμοποιούνται για τη διάκριση των HMMs που περιγράφουν ακολουθίες παρατηρήσεων προερχόμενες από τα τρία κύρια σύνολα δεδομένων. Στην εικόνα (c) νευρώνες που αναπαριστούν ακολουθίες παρατηρήσεων οι οποίες ανήκουν σε καθένα από τα τέσσερα διαφορετικά παράγωγα σύνολα δεδομένων διαχωρίζονται με τη χρήση κάθετων, διαγώνιων και οριζόντιων διαγραμμίσεων. Μία ενδεικτική συστοιχία SOHMMM μετά την ολοκλήρωση της αντίστοιχης διαδικασίας μη επιβλεπόμενης μάθησης που βασίστηκε στο περιορισμένο σύνολο εκπαίδευσης απεικονίζεται στην (d).
- 5.2 (a) Ενδεικτικό παράδειγμα μίας συστοιχίας SOHMMM μετά το πέρας της διαδικασίας προσαρμογής που χρησιμοποιεί το πλήρες σύνολο των 560 σφαιρινών. (b) Εναλλακτικό παράδειγμα ενός SOHMMM, μετά την ολοκλήρωση του αλγορίθμου μη επιβλεπόμενης μάθησης, για το περιορισμένο σύνολο εκπαίδευσης των 225 σφαιρινών.
- 5.3 Γραφικές παραστάσεις ποσοστιαίου σφάλματος-πλήθους εποχών του SOHMMM, όπως αυτές προκύπτουν για το σύνολο δεδομένων των σφαιρινών. Η γραφική παράσταση (a) αφορά τις πρώτες 5 εποχές εκπαίδευσης από τις συνολικά 50 που χρησιμοποιήθηκαν.
- 5.4 Μέσες τιμές ποσοστιαίων σφαλμάτων που παράγονται από πλέγματα SOHMMM διαφόρων διαστάσεων.

- 5.5 Γραφική παράσταση ποσοστιαίου σφάλματος-μεγέθους χώρου κρυφών καταστάσεων των HMM νευρώνων ενός 6x8 SOHMMM, η οποία αντιστοιχεί στο σύνολο δεδομένων των σφαιρινών.
- 5.6 SDDD βασιζόμενο στην πλήρη πρωτεϊνική οικογένεια των σφαιρινών. Κάθε εξάγωνο, ανεξαρτήτως μεγέθους, αναπαριστά έναν ξεχωριστό νευρώνα SOHMMM.
- 5.7 Η υλοποιηθείσα SDDD απεικόνιση εμπλουτισμένη με πληροφορία ανάθεσης κατηγοριών προερχόμενης από εκ των υστέρων χαρακτηρισμό των συστάδων.
- 5.8 Χαρακτηριστική SLP απεικόνιση που αντιστοιχεί σε μία α -σφαιρίνη. Κάθε εξάγωνο, ανεξαρτήτως απόχρωσης, αντιστοιχεί σε έναν ξεχωριστό κόμβο SOHMMM.
- 5.9 Ενδεικτική SLP βασιζόμενη σε μυοσφαιρίνη.
- 5.10 Παράδειγμα μίας SLP γραφικής αναπαράστασης που αντιστοιχεί σε μία β -σφαιρίνη.
- 5.11 Χαρακτηριστικό αποτέλεσμα μίας SLP που αφορά σε μία πρωτεϊνική αλυσίδα κυτοσφαιρίνης.
- 5.12 Γραφικές παραστάσεις ποσοστιαίου σφάλματος-πλήθους εποχών του SOHMMM, όπως αυτές προκύπτουν για το σύνολο δεδομένων των splice junctions. Η γραφική παράσταση (a) αφορά τις πρώτες 10 εποχές εκπαίδευσης από τις συνολικά 100 που χρησιμοποιήθηκαν.
- 5.13 Μέσες τιμές ποσοστιαίων σφαλμάτων που παράγονται από πλέγματα SOHMMM διαφόρων διαστάσεων.
- 5.14 Γραφική παράσταση ποσοστιαίου σφάλματος-μεγέθους χώρου κρυφών καταστάσεων των HMM νευρώνων ενός 11x8 SOHMMM, η οποία αντιστοιχεί στο σύνολο δεδομένων των splice sites.
- 5.15 SDDD που αντιστοιχεί στο πρόβλημα προσδιορισμού των splice junctions.
- 5.16 Ενδεικτική SLP απεικόνιση μίας αλληλουχίας που περιέχει ένα όριο εξωνίου-εσωνίου.
- 5.17 Χαρακτηριστικό παράδειγμα μίας SLP γραφικής αναπαράστασης που αφορά σε ένα acceptor splice site.

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

5 Πειράματα και Εφαρμογές

- 5.1 Μέσες Τιμές Ποσοστιαίων Ακριβειών: Αποτελέσματα κατηγοριοποίησης της οικογένειας των σφαιρινών που επιτυγχάνονται από διαφορετικές προσεγγίσεις όπως SOHMMM, σχεσιακά SOMs επιβλεπόμενης μάθησης, (σχεσιακές) NG παραλλαγές και τεχνικές k – μέσων .
- 5.2 Μέσες Τιμές Ποσοστιαίων Σφαλμάτων: Οι επιδόσεις του SOHMMM και διαφόρων άλλων αλγορίθμων μηχανικής μάθησης στο πρόβλημα καθορισμού των splice junctions.

1. ΕΙΣΑΓΩΓΙΚΟ ΣΗΜΕΙΩΜΑ

Βασική προϋπόθεση της μηχανικής μάθησης είναι ότι η διαθέσιμη πληροφορία μπορεί να αναπαρασταθεί από σύνολα αριθμητικών και συμβολικών ποσοτήτων. Με την ενσωμάτωση και την επεξεργασία των προκύπτων αριθμητικών και συμβολικών δεδομένων μπορούν να επιτελεστούν πλείστες εργασίες, όπως μοντελοποίηση και ανάλυση, που σχετίζονται άμεσα ή έμμεσα με τους αντίστοιχους χώρους προβλημάτων. Γενικώς, είναι κοινή πρακτική να κωδικοποιείται ρητά κάθε κομμάτι σημαντικής ή χρήσιμης πληροφορίας, πράγμα που περιλαμβάνει τις όποιες εγγενείς χρονικές και χωρικές εξαρτήσεις. Αυτό έχει ως αποτέλεσμα ένας διόλου ευκαταφρόνητος αριθμός προσεγγίσεων μηχανικής μάθησης να θεωρεί ότι η διάταξη των στοιχείων ενός εκάστοτε συνόλου δεδομένων δεν περικλείει κανενός είδους πληροφορία (ήτοι να θεωρεί τη σχετική θέση και διάταξη των στοιχείων κάθε διανύσματος δεδομένων αυθαίρετες). Σε συγκεκριμένες περιπτώσεις, ωστόσο, αυτό δεν είναι (απόλυτα) ακριβές: η χωρική διευθέτηση των στοιχείων των διανυσμάτων (δηλαδή το ποιο στοιχείο έρχεται πρώτο, δεύτερο και ούτω καθεξής), ή ισοδύναμα, η χρονική διάταξη τους (δηλαδή το ποια στοιχεία προηγούνται και ποια έπονται) είναι σημαντική, και επιπλέον, αποτελεί μία πηγή επιπρόσθετων πληροφοριών. Κατά συνέπεια τουλάχιστον οι μεθοδολογίες μηχανικής μάθησης που χρησιμοποιούνται για ανάλυση βιολογικών ακολουθιών και για μοντελοποίηση φωνής, οφείλουν να είναι σε θέση να προσπελάσουν και να αξιοποιήσουν τέτοιου τύπου ενυπάρχουσα αλλά αφανή πληροφορία.

Υπό τη θεώρηση μηδενικής ή ανύπαρκτης εκ των προτέρων γνώσης, η μόνη διαθέσιμη πληροφορία που απομένει είναι αυτά καθαυτά τα διανύσματα και αυτές καθαυτές οι ακολουθίες (δηλαδή οι τιμές των στοιχείων τους καθώς και τα μεγέθη/μήκη τους) που περιγράφουν τους υπό εξέταση χώρους προβλημάτων. Υπό αυτές τις συνθήκες, τεχνικές (ημι)επιβλεπόμενης ή ενισχυτικής μάθησης καθίστανται πρακτικά ανεφάρμοστες. Τα μόνα διαθέσιμα μέσα για τη διεξαγωγή πρωτογενούς διερευνητικής ανάλυσης δεδομένων (κάτι το οποίο είναι σύνηθες και προφανώς ζητούμενο κατά τα πρώιμα στάδια έρευνας) αποδεικνύεται ότι είναι οι μεθοδολογίες μη επιβλεπόμενης μάθησης. Κατά συνέπεια, λαμβάνοντας υπόψη τα χαρακτηριστικά και την ποιότητα της παρεχόμενης πληροφορίας, μπορεί κανείς να διαπιστώσει ότι η γκάμα των εφαρμογών αυτών των μεθοδολογιών είναι σαφώς διευρυμένη.

Όλη η ζωή σε αυτόν τον πλανήτη βασίζεται σε αλληλουχίες δεσοξυριβονουκλεϊνικού οξέος - DeoxyriboNucleic Acid (DNA), ριβοζονουκλεϊνικού οξέος - RiboNucleic Acid (RNA) και πρωτεϊνών· τους τρεις θεμελιώδεις τύπους βιολογικών μορίων. DNA, RNA και πρωτεΐνες είναι παραδείγματα αλληλουχιών που είναι γραμμένες στο αλφάβητο των τεσσάρων νουκλεοτιδίων ή στο αλφάβητο των είκοσι αμινοξέων. Η έλευση καινοτόμων και αποδοτικών πειραματικών μεθοδολογιών, κυρίως των τεχνικών του προσδιορισμού της νουκλεοτιδικής αλληλουχίας στα γονιδιώματα, των μικροσυστοιχιών DNA και της φασματομετρίας μάζας, έχουν οδηγήσει σε μία εκθετική αύξηση των γραμμικών περιγραφών βιολογικών μορίων πρωτεϊνών, DNA και RNA. Αυτή η αύξηση συνεπάγεται ακόμη περισσότερες απαιτήσεις για αυτοματοποιημένη επεξεργασία και ανάλυση. Συνολικά, αυτές οι τεχνολογίες υψηλής ρυθμαπόδοσης είναι σε θέση να παράγουν ταχέως terabytes δεδομένων τα οποία υπερβαίνουν τις δυνατότητες συμβατικών βιολογικών προσεγγίσεων. Για να ανταποκριθούν στις νέες αυτές συνθήκες οι επιστήμονες χρησιμοποιούν αλγόριθμους, στατιστικές μεθόδους και άλλες μαθηματικές τεχνικές για να αποκρυπτογραφήσουν τη γλώσσα του DNA [1]. Συνεπώς, η ανάγκη υπολογιστικών/στατιστικών τεχνικών καθώς και μεθόδων μηχανικής μάθησης, για την ποιοτική και ποσοτική ανάλυση βιολογικών μορίων, είναι σήμερα εντονότερη από ποτέ.

Αν και ο σκοπός της ανάλυσης βιολογικών αλληλουχιών συνήθως είναι η μελέτη μίας συγκεκριμένης αλληλουχίας, της μοριακής δομής και λειτουργίας της, η ανάλυση συχνά επεκτείνεται στη μελέτη συλλογών αλληλουχιών που απαρτίζονται από διαφορετικές εκδοχές ξεχωριστών ειδών, ή και ακόμη, στην περίπτωση πολυμορφισμών από διαφορετικές εκδοχές του ίδιου είδους. Συνεπώς, η ανάλυση ακολουθιακών προτύπων οφείλει να λαμβάνει υπόψη ότι οι βιολογικές αλληλουχίες είναι ποικιλόμορφες. Οι εν λόγω παρατηρούμενες διακυμάνσεις εν μέρει είναι αποτέλεσμα τυχαίων συμβάντων που ενισχύονται από τους μηχανισμούς της εξέλιξης. Από τη στιγμή που αλληλουχίες νουκλεοτιδίων ή αμινοξέων με μία δεδομένη δομή και λειτουργία, εν γένει, διαφέρουν και είναι ασαφείς, τα μοντέλα που θα επιστρατευτούν για να τις χειριστούν πρέπει να είναι πιθανοτικά [2].

Παρότι συμβατικοί υπολογιστικοί αλγόριθμοι και συνήθειες στατιστικές τεχνικές έχουν φανεί χρήσιμες, αποδεικνύονται όλο και περισσότερο αναποτελεσματικές στην αντιμετώπιση των πλέον ενδιαφερόντων προβλημάτων ανάλυσης ακολουθιών. Αυτό συμβαίνει εξαιτίας της εγγενούς πολυπλοκότητας των βιολογικών συστημάτων, που

οφείλεται στις σύνθετες βιολογικές διεργασίες και στην έλλειψη μίας ολοκληρωμένης και συνεπούς θεωρίας αναφορικά με την οργάνωση και τη λειτουργία της ζωής σε μοριακό επίπεδο. Από την άλλη πλευρά, νευρωνικά δίκτυα - Neural Networks (NNs), κρυφά μοντέλα Markov - Hidden Markov Models (HMMs), δίκτυα Bayes και άλλες παρεμφερείς προσεγγίσεις μηχανικής μάθησης ταιριάζουν ιδανικά σε πεδία που χαρακτηρίζονται από την ύπαρξη μεγάλου όγκου (πιθανώς ελλιπών, κατακερματισμένων και αλλοιωμένων από θόρυβο) δεδομένων, πολύπλοκων δομών, μεταβαλλόμενων προτύπων και απουσίας γενικών θεωριών [2]. Η θεμελιώδης ιδέα πίσω από αυτές τις προσεγγίσεις είναι η με αυτόματο τρόπο απόκτηση γνώσης από τα δεδομένα βάσει μίας διαδικασίας προσαρμογής, ρύθμισης παραμέτρων, μη γραμμικής παρεμβολής, παλινδρόμησης και εκπαίδευσης από παραδείγματα.

Μία από τις μεγαλύτερες προκλήσεις που αντιμετωπίζει κανείς κατά την εκτεταμένη εφαρμογή ήδη υφιστάμενων συστημάτων μηχανικής μάθησης προέρχεται από τον αλγοριθμικό έλεγχο. Η πλειοψηφία των σύγχρονων αλγορίθμων μάθησης αξιούνουν ρύθμιση παραμέτρων εκ νέου και εξατομικευμένα για κάθε ξεχωριστό πρόβλημα που καλούνται να αντιμετωπίσουν. Οι περιορισμοί των σύγχρονων συστημάτων μηχανικής μάθησης εν συγκρίσει με τα βιολογικά συστήματα είναι ένα γεγονός που έχει ήδη επισημανθεί [3]. Τα βιολογικά συστήματα μάθησης έχουν τη δυνατότητα να εκπαιδεύονται αυτόνομα, στα πλαίσια ενός συνεχώς μεταβαλλόμενου και αβέβαιου κόσμου. Η ιδιότητα τους αυτή περικλείει τις ικανότητες δημιουργίας της δικής τους επίβλεψης, επιλογής των προτύπων εκπαίδευσης που φέρουν την περισσότερη πληροφορία, παραγωγής της δικής τους συνάρτησης κόστους, και αποτίμησης των επιδόσεών τους· με άλλα λόγια ένα είδος κλιμακωτής οργάνωσης της διαδικασιών προσαρμογής και εκπαίδευσης. Εν αντιθέσει, κυριολεκτικά όλες οι σύγχρονες προσεγγίσεις μηχανικής μάθησης ουσιαστικά απαιτούν έναν ανθρώπινο επιβλέποντα που να σχεδιάζει την αρχιτεκτονική του συστήματος, να σταχυολογεί τα πρότυπα εκπαίδευσης, να καθορίζει τη μορφή της αναπαράστασης των προτύπων εκπαίδευσης, να επιλέγει τον αλγόριθμο μάθησης, να ορίζει τις παραμέτρους προσαρμογής, να επιλέγει το πότε θα τερματίζεται η διαδικασία μάθησης, να αποφασίζει με ποιο τρόπο θα αξιολογούνται οι επιδόσεις και η αποδοτικότητα του αλγορίθμου. Αυτή η ισχυρή εξάρτηση από τον ανθρώπινο παράγοντα επίβλεψης επιβραδύνει σε μεγάλο βαθμό την ανάπτυξη πανταχού παρόντων αυτόνομων τεχνητών συστημάτων μηχανικής μάθησης.

Η διερευνητική ανάλυση δεδομένων αντιπροσωπεύει μία πρωτογενή εξερεύνηση υπό συνθήκες (σχεδόν) πλήρους έλλειψης εκ των προτέρων γνώσης. Ο αντικειμενικός στόχος της είναι η παραγωγή αφαιρετικών περιγραφών και περιεκτικών αναπαραστάσεων μεγάλων συνόλων δεδομένων. Ο σκοπός της ομαδοποίησης [4], [5] είναι η διαμέριση ενός πεπερασμένου συνόλου δεδομένων σε ένα πεπερασμένο σύνολο συστάδων, κατά τέτοιο τρόπο ώστε η ομοιότητα μεταξύ των στοιχείων μίας συστάδας να είναι μεγαλύτερη της ομοιότητας μεταξύ των στοιχείων που ανήκουν σε διαφορετικές συστάδες. Η ομαδοποίηση είναι μία καθιερωμένη μέθοδος για την αυτόματη δημιουργία ανώτερων αφαιρέσεων (συμβολισμών) από ανεπεξέργαστα δεδομένα, μία άλλη εναλλακτική μέθοδος είναι η προβολή δεδομένων υψηλής διάστασης ως σημεία μίας απεικόνισης χαμηλής διάστασης. Αυτές οι απεικονίσεις αναπαριστούν τα δεδομένα εισόδου σε ένα χώρο χαμηλότερης διάστασης με τέτοιο τρόπο ούτως ώστε οι συστάδες και οι μετρικές σχέσεις των δεδομένων στοιχείων να παρουσιάζονται όσο πιο πιστά γίνεται. Ο αυτο-οργανούμενος χάρτης - Self-Organizing Map (SOM) είναι ένας συνδυασμός αυτών των δύο μεθόδων πράγμα που σημαίνει ότι μπορεί να χρησιμοποιηθεί παράλληλα για ομαδοποίηση δεδομένων, μείωση διαστατικότητας και οπτικοποίηση συστάδων [6]-[12].

Τα HMMs εδράζονται στα σταθερά θεμέλια των πιθανοτήτων και της στατιστικής, ενώ συνοδεύονται από θεωρητικά αποδεδειγμένους αλγορίθμους μάθησης [13], [2]. Είναι στοχαστικές διαδικασίες που παράγονται από δύο συνυφασμένους πιθανοτικούς μηχανισμούς. Αυτοί είναι μία (κρυφή) υποκείμενη αλυσίδα Markov πεπερασμένου πλήθους καταστάσεων και ένα (παρατηρήσιμο) υπερκείμενο σύνολο συναρτήσεων πιθανότητας που συσχετίζονται με τις εν λόγω καταστάσεις. Η αρχιτεκτονική και οι αλγόριθμοι ρύθμισης των HMMs αποδεικνύονται ιδιαίτερα πρακτικοί για την αξιοποίηση και την αποκρυπτογράφηση της ενυπάρχουσας πληροφορίας που βρίσκεται κρυμμένη στη χωρική δομή, στις χρονικές συσχετίσεις και στην γραμμική ακολουθιακή φύση μορίων αλληλουχιών. Τα HMMs είναι χρήσιμα εργαλεία για την ανάλυση ακολουθιακών μοτίβων από μονομερή, επιδεικνύοντας παράλληλα ευελιξία και προσαρμοστικότητα. Τα HMMs, όταν χρησιμοποιούνται σωστά, λειτουργούν πολύ καλά στην πράξη, αυτό έχει ως αποτέλεσμα η γκάμα των εφαρμογών τους [14], [1] να περιλαμβάνει αναγνώριση και ανάλυση γονιδίων, ανακάλυψη (περιοδικών) μοτίβων και προτύπων, πολλαπλές ευθυγραμμίσεις και ταιριάσματα. Επίσης, παρέχουν τη δυνατότητα επεξεργασίας και ανάλυσης βιολογικών αλληλουχιών υψηλής διάστασης

και διαφορετικών μηκών, που είναι γραμμένες στα συμβολικά/μη αριθμητικά αλφάβητα των νουκλεοτιδίων και των αμινοξέων.

Ο αυτο-οργανούμενος χάρτης κρυφών μοντέλων Markov - Self-Organizing Hidden Markov Model Map (SOHMMM) συνιστά ένα παράδειγμα νευρωνικού δικτύου μηχανικής μάθησης που μπορεί να καταταχθεί στην ευρέως νοούμενη κλάση των πιθανοτικών μοντέλων μη επιβλεπόμενης μάθησης. Ουσιαστικά το SOHMMM είναι μία διασταύρωση μεταξύ SOM και HMM σε επίπεδο θεωρητικής θεμελίωσης και αλγοριθμικής πραγμάτωσης. Αμφότερες οι δύο αρχιτεκτονικές συγχωνεύονται σε ένα ενιαίο και αδιαίρετο σχήμα. Ο λειτουργικός πυρήνας του SOHMMM αποτελείται από έναν ενοποιημένο υβριδικό SOM-HMM αλγόριθμο. Συγκεκριμένα το προτεινόμενο μοντέλο είναι εξοπλισμένο με έναν στοχαστικό αλγόριθμο εκπαίδευσης που αποτελεί συνένωση της μη επιβλεπόμενης μάθησης του SOM και της forward-backward μεθόδου δυναμικού προγραμματισμού του HMM. Από τη στιγμή που το SOHMMM είναι σε θέση να κάνει αναλύσεις δεδομένων ακολουθιών, απαιτώντας ελάχιστη ή μηδενική εκ των προτέρων γνώση, μπορεί να έχει μία σειρά εφαρμογών στην ομαδοποίηση, στην μείωση διαστατικότητας και στην οπτικοποίηση συστάδων ακολουθιών μεγάλης κλίμακας, και επιπλέον, υπό συγκεκριμένες προϋποθέσεις, στην αναζήτηση και στην κατηγοριοποίηση τους.

Στις [15]-[17] περιέχεται μία συνοπτική παρουσίαση του θεωρητικού σκέλους του SOHMMM ακολουθούμενη από ορισμένες αδρές πειραματικές διερευνήσεις. Συγκεκριμένα, η πειραματική μεθοδολογία που ακολουθήθηκε (η οποία βασίζεται σε πρωτεϊνικές αλληλουχίες [15], [17] και σε τεχνητά σύνολα ακολουθιών [16]) συνιστά μία πρωταρχική απόπειρα μελέτης και ανάλυσης χαρακτηριστικών πτυχών και ιδιοτήτων του SOHMMM. Περαιτέρω, στις [18], [19] περιέχονται εμβριθείς και λεπτομερείς αναλύσεις του SOHMMM, που ξεκινούν από τις πλέον θεμελιώδεις έννοιες και καταλήγουν σε πιο σύνθετα ζητήματα. Μία σημαντική διαφοροποίηση που υιοθετείται στην [18] έχει να κάνει με έναν τροποποιημένο αλγόριθμο προσαρμογής του SOHMMM (που βασίζεται σε μία κατάλληλα κατασκευασμένη συνάρτηση ενέργειας), ο οποίος έχει ένα πιο στιβαρό μαθηματικό υπόβαθρο και μπορεί επίσης να αποδειχθεί αναλυτικά. Παρομοίως, ένα καθοριστικό νέο στοιχείο της [19] αφορά στην αναβάθμιση του στοχαστικού αλγόριθμου μη επιβλεπόμενης μάθησης του SOHMMM με μία διαδικασία μετατροπής κλίμακας, γεγονός που τον καθιστά ικανό να χειρίζεται και να επεξεργάζεται ακολουθίες αυθαίρετα μεγάλου μήκους. Αμφότεροι αυτοί οι δύο αλγόριθμοι είναι σαφείς σημαντικές βελτιώσεις των πρόγονων εκδοχών τους [15]-[17].

Επιπρόσθετα, σε αυτές τις δύο πιο πρόσφατες μελέτες, ζητήματα που άπτονται της υλοποίησης και εφαρμογής του αλγορίθμου SOHMMM αντιμετωπίζονται με συστηματικό και μεθοδικό τρόπο. Παράλληλα, διεξάγονται ενδελεχή και εκτενή πειράματα έτσι όλα τα προκύπτοντα ποιοτικά συμπεράσματα και ποσοτικά αποτελέσματα αναφέρονται στα χαρακτηριστικά, στις επιδόσεις και στις δυνατότητες των δύο βελτιωμένων τροποποιήσεων του αλγορίθμου SOHMMM. Τέλος, στην [20] προτείνονται τεχνικές οι οποίες συνιστούν επέκταση και διεύρυνση του φάσματος λειτουργίας του SOHMMM εξοπλίζοντας το με τη δυνατότητα να υλοποιεί οπτικοποιήσεις και να παράγει μη γραμμικές αναπαραστάσεις.

2. ΣΧΕΤΙΚΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ

Στην βιβλιογραφία απαντώνται διάφορες προσεγγίσεις που επεκτείνουν το SOM με επαναληπτικές συνδέσεις οι οποίες επιτρέπουν την αναδρομική επεξεργασία ακολουθιών. Συχνά, αυτά τα μοντέλα βασίζονται στις ίδιες αρχές ανταγωνισμού και συνεργασίας που διέπουν το σχηματισμού του SOM.

Το sequential activation retention δίκτυο είναι μία αρχιτεκτονική χάρτη χαρακτηριστικών που εξοπλίζει το SOM με τη δυνατότητα δημιουργίας κατανομημένων προτύπων απόκρισης (αναπαραστάσεων) για διαφορετικές ακολουθίες διανυσμάτων. Έτσι, πολύπλοκες δομές δεδομένων αναπαριστώνται ως πολύπλοκα πρότυπα που ενσωματώνουν περισσότερους του ενός ενεργοποιημένους νευρώνες. Ο extended Kohonen feature map [21] αναβαθμίζει το SOM με εσωτερικές συνδέσεις μεταξύ καθενός νευρώνα με όλους τους άλλους νευρώνες. Το αποτέλεσμα είναι με κάθε νευρώνα να συσχετίζονται δύο ενεργοποιήσεις μία εξωτερική και μία εσωτερική.

Ο temporal Kohonen map της [22] επεκτείνει το SOM με αναδρομικές αυτο-συνδέσεις στους νευρώνες, τέτοιες ώστε αυτοί να συμπεριφέρονται ως ολοκληρωτές. Έτσι το προκύπτον μοντέλο είναι σε θέση να δίνει έννοια σε μοτίβα που εμφανίζονται σε ακολουθίες. Το recurrent SOM όπως ορίζεται στις [23], [24] επιστρατεύει μία παρόμοια δυναμική. Ωστόσο, η επαναληπτική δομή, μετακινώντας τους ολοκληρωτές από τις μονάδες εξόδου στις μονάδες εισόδου, επιτρέπει στο χάρτη να αποθηκεύει πληροφορία σχετική με την αλλαγή του μέτρου και της κατεύθυνσης του διανύσματος εισόδου. Φυσιολογικά, αυτή η διαδικασία αποθηκεύει περισσότερη πληροφορία από ότι οι σταθμισμένες αποστάσεις του temporal Kohonen map. Μολαταύτα, ένας περιορισμός των νευρώνων ολοκληρωτών είναι ότι οι χρονικά απομακρυσμένες πληροφορίες τείνουν να εξασθενούν εκθετικά. Το recursive SOM [25], είναι ένα τροποποιημένο SOM όπου η ανάδραση χρησιμοποιείται για την αναπαράσταση του χρόνου, και έχει προταθεί ως ένας μηχανισμός για την πρόβλεψη ακολουθιών. Η αναπαράσταση του χρόνου η οποία είναι ρητή και αυτο-αναφερόμενη υλοποιείται μέσω ομογενών συνδέσεων προσο-τροφοδότησης και οπισθο-τροφοδότησης, οι οποίες περιγράφουν τις ενεργοποιήσεις. Με αυτόν τον τρόπο, τμήματα της δοθείσας ακολουθίας επεξεργάζονται αναδρομικά βάσει των ήδη σχηματισμένων τοπικών αναπαραστάσεων. Το merge SOM context [26] αναφέρεται στη συνένωση αυθαίρετων τοπολογιών πλεγμάτων με μία αρχιτεκτονική ανεκτική στο θόρυβο η οποία υλοποιεί

μία συσχέτιση με τον προηγούμενο νικητή νευρώνα, παρέχοντας έτσι μεγαλύτερη ευελιξία και χωρητικότητα εξαιτίας της ρητής αναπαράστασης αυτής της συνάφειας.

Το SOM για δομημένα δεδομένα [27], [28] είναι ένα γενικευμένο μοντέλο μη επιβλεπόμενης μάθησης που χρησιμοποιείται για την επεξεργασία αντικειμένων με δομή, τουτέστιν κατευθυνόμενους ακυκλικούς γράφους - Directed Acyclic Graphs (DAGs) με ανάθεση πληροφορίας κατηγορίας. Προτείνεται ένας αναδρομικός μηχανισμός ικανός να αξιοποιεί την πληροφορία που βρίσκεται αποτυπωμένη σε κάθε κόμβο ενός εκάστοτε DAG, καθώς και την πληροφορία που είναι κωδικοποιημένη στην τοπολογία του DAG. Έτσι, το εν λόγω μοντέλο είναι σε θέση να χειρίζεται δεδομένα που χαρακτηρίζονται από περίπλοκες δομές, μεταβλητά μήκη και αυξημένη πολυπλοκότητα. Στην [29] παρέχεται μία λύση στο πρόβλημα ότι ορισμένοι DAGs δεν διακρίνονται από το SOM για δομημένα δεδομένα. Η προτεινόμενη προσέγγιση είναι αποτελεσματική για τη διάκριση της απεικόνισης κορυφών και υπο-δομών βάσει του πλαισίου εντός του οποίου απαντώνται εντός του δέντρου. Μία εμπειρική συγκριτική αξιολόγηση αυτών των αναδρομικών μοντέλων παρουσιάζεται στην [30].

Στην [31] η προσοχή εστιάζεται στα δίκτυα μη επιβλεπόμενης μάθησης που χρησιμοποιούν επαναληπτική/αναδρομική δυναμική (για την επεξεργασία σύνθετων δομών), επίσης, εξάγεται ένα γενικό πλαίσιο εργασίας για τη μοντελοποίηση αναδρομικών SOMs. Το γενικό αυτό πλαίσιο (με ελάχιστες διαφοροποιήσεις σε ορισμένα σημεία της διατύπωσης) καλύπτει τις [21]-[35] καθώς και το κανονικό SOM. Αυτές οι μέθοδοι μοιράζονται την ίδια βασική αναδρομική δυναμική αλλά διαφέρουν στον εσωτερικό τρόπο αναπαράστασης των δομών. Επιπρόσθετα, παρέχεται μία ομοιόμορφη προτυποποίηση που επιτρέπει την απευθείας διερεύνηση ενδεχόμενων αλγορίθμων μάθησης καθώς και θεωρητικών ιδιοτήτων διαφόρων σημαντικών προσεγγίσεων που συνδυάζουν το SOM με επαναληπτικές τεχνικές.

Το μοντέλο self-organizing mixture [36] εισάγει έναν αλγόριθμο μεγιστοποίησης προσδοκίας - Expectation-Maximization (E-M) ο οποίος παράγει χάρτες που διατηρούν την τοπολογία τους και που βασίζονται σε μικτά πιθανοτικά μοντέλα. Τα μικτά πιθανοτικά μοντέλα είναι πυκνότητες (ή κατανομές) πιθανοτήτων που μπορούν να γραφούν ως σταθμισμένα αθροίσματα άλλων συνιστωσών πιθανοτήτων. Ένα αυτο-οργανούμενο μικτό μοντέλο μπορεί να μεταφραστεί ως ένα μοντέλο που υποθέτει ότι υπάρχουν πολλαπλές πηγές δημιουργίας των δεδομένων εισόδου, κάθε τέτοια πηγή επιλέγεται για τη δημιουργία δεδομένων με πιθανότητα ίση με το συντελεστή στάθμισης (που συνήθως είναι μία κανονικοποιημένη συνάρτηση γειννίας). Το

δίκτυο που περιγράφεται στην [37] είναι μία ειδική περίπτωση αυτού της [36], αφού προτείνει ένα πιθανοτικό Bernoulli SOM ικανό να χειρίζεται αποκλειστικά δυαδικά δεδομένα. Μία άλλη μέθοδος για τη δημιουργία τοπογραφικών απεικονίσεων είναι η [38]. Ουσιαστικά αυτή η προσέγγιση εφαρμόζει ντετερμινιστική ανόπτηση σε κατάλληλες συναρτήσεις κόστους για την παραγωγή των αντίστοιχων χαρτογραφήσεων.

Ο Generative Topographic Map (GTM) [39] συχνά παρουσιάζεται ως ένα πιθανοτικό ανάλογο του SOM. Ωστόσο, επιτυγχάνει την τοπογραφική οργάνωση με εντελώς διαφορετικό τρόπο από το SOM. Το GTM οικοδομείται επάνω σε ένα περιορισμένο (σχετιζόμενο με κρυφές συντεταγμένες) πιθανοτικό μικτό μοντέλο που υπακούει σε συγκεκριμένους τοπολογικούς περιορισμούς, και το οποίο προσαρμόζεται στα δεδομένα που δίδονται. Οι μικτές υπομονάδες παραμετροποιούνται ως ένας γραμμικός συνδυασμός από ομαλές μη γραμμικές συναρτήσεις βάσης που εδράζονται στις σταθερές θέσεις των υπομονάδων στον υποκείμενο χώρο. Εξαιτίας αυτής της απεικόνισης, γειτονικές υπομονάδες στον υποκείμενο χώρο θα έχουν παρεμφερείς παραμέτρους. Οι εν λόγω παράμετροι του GTM καθορίζονται από ένα αλγόριθμο E-M που βελτιστοποιεί την πιθανοφάνεια. Αρχικά, το GTM περιελάμβανε μικτές υπομονάδες Gauss, γεγονός που το καθιστούσε κατάλληλο για την ανάλυση συνεχών δεδομένων. Ένα συγκεκριμένο GTM αναπτύχθηκε αργότερα για την επεξεργασία δυαδικών δεδομένων, το οποίο υιοθετεί μία προσέγγιση της διωνυμικής πιθανοφάνειας [40]. Το GTM στην [41] χρησιμοποιεί μικτά μοντέλα που ανήκουν στην οικογένεια των εκθετικών κατανομών. Το προκύπτον τροποποιημένο μοντέλο δύναται να εφαρμοστεί απευθείας σε διακριτά πολυμεταβλητά δεδομένα. Ένα GTM για την οπτικοποίηση διακριτών ακολουθιακών δεδομένων παρουσιάζεται στην [42]. Μία περαιτέρω επέκταση του GTM για την οπτικοποίηση δεδομένων δενδρικής δομής προτείνεται στην [43]. Είναι σημαντικό να σημειώσουμε ότι σε όλες τις παραλλαγές του GTM οφείλουμε να προσδιορίσουμε εκ των προτέρων ορισμένες δομικές παραμέτρους. Κατ' ελάχιστον, πρέπει να καθοριστούν το πλήθος, ο τύπος, η μορφή των μικτών υπομονάδων και των μη γραμμικών συναρτήσεων βάσης. Επίσης, ένας βασικός περιορισμός είναι ότι δεν φαίνεται να υπάρχει μία μεθοδολογία για την τροποποίηση του GTM με άλλα μικτά μοντέλα πλην της οικογένειας των εκθετικών κατανομών.

Στην [44] διερευνάται η σχέση μεταξύ μικτών μοντέλων και SOMs. Παρουσιάζεται μία μεθοδολογία που εξοπλίζει με δυνατότητες αυτο-οργάνωσης ένα μικτό μοντέλο.

Επίσης, αποσαφηνίζεται ο τρόπος με τον οποίο το προκύπτον κριτήριο βελτιστοποίησης μπορεί να μεταφραστεί με όρους στατιστικής ανάλυσης. Αντιστρόφως, το SOM αντιστοιχίζεται με ένα μικτό μοντέλο που έχει επιπλέον έναν όρο ομαλοποίησης (ελαστικό δίκτυο). Ο απώτερος σκοπός της γενικευμένης ταξινομίας [45] είναι να παρέχει ένα μη γραμμικό πλαίσιο εργασίας για την περιγραφή χωροχρονικών δικτύων μη επιβλεπόμενης μάθησης, καθιστώντας έτσι ευκολότερη τη σύγκριση και την αντιπαραβολή των λειτουργικών τους χαρακτηριστικών καθώς και των αναπαραστάσεων που επιτυγχάνουν. Από περιγραφικής σκοπιάς, η προτεινόμενη προσέγγιση είναι σαφώς ευρύτερη και περιεκτικότερη συγκριτικά με προηγούμενες απόπειρες. Επίσης, καταστρώνεται μία στρατηγική για τη δημιουργία ενός ενοποιημένου μαθηματικού εργαλείου για την ανάλυση, ίσως ακόμη και για το σχεδιασμό, χωροχρονικών δικτύων μη επιβλεπόμενης μάθησης.

Μέχρι στιγμής, έχει γίνει μία επισκόπηση ενός αριθμού διαδεδομένων μοντέλων του ερευνητικού πεδίου των χωροχρονικών αυτο-οργανούμενων δικτύων μη επιβλεπόμενης μάθησης. Η πλειονότητα των παραλλαγών του SOM [21]-[38] είναι σε θέση να επεξεργάζεται ακολουθίες ή δένδρα δεδομένων με αριθμητική κωδικοποίηση. Αυτές οι μεθοδολογίες, που μοντελοποιούν χωροχρονική πληροφορία, προϋποθέτουν ότι οι δεδομένες ακολουθίες αποτυπώνονται σε κάποια αριθμητική μορφή, δηλαδή προϋποθέτουν ότι αυτές οι ακολουθίες απαρτίζονται από συνεχείς/πραγματικές ή από διακριτές/δυναδικές τιμές. Αυτά τα μοντέλα είναι ικανά να χειριστούν και να αναλύσουν ακολουθίες που αποτελούνται από συμβολικά στοιχεία μόνο κατόπιν της παρεμβολής ενός ενδιάμεσου σταδίου προεπεξεργασίας (για το μετασχηματισμό των ακολουθιών συμβόλων σε κάποιο κατάλληλο αριθμητικό χώρο).

Συχνά, τέτοιες ακολουθίες συμβόλων δεν δύναται να αναπαρασταθούν αποδοτικά με τη μορφή αριθμητικών διανυσμάτων. Για να ξεπεραστεί αυτός ο περιορισμός, ορισμένες προσεγγίσεις χρησιμοποιούν ορθογώνια κωδικοποίηση (για παράδειγμα οι [26], [37], [46]). Αυτού του τύπου η αναπαράσταση έχει το πλεονέκτημα ότι δεν εισαγάγει (ανύπαρκτες κατά τα άλλα) αλγεβρικές συσχετίσεις μεταξύ των συμβόλων. Από την άλλη πλευρά, αυτό το σχήμα αραιής κωδικοποίησης πάσχει από το μειονέκτημα της μεγάλης αύξησης του μήκους των ακολουθιών που προέρχονται από βιολογικές αλληλουχίες, ακόμη και κατά είκοσι φορές (όπως για παράδειγμα στην περίπτωση πρωτεϊνών). Το αποτέλεσμα είναι οι υπάρχουσες παραλλαγές του SOM (που χρησιμοποιούν ορθογώνια κωδικοποίηση) να βρίσκουν εφαρμογές σε προβλήματα μικρής κλίμακας, αφού, στην πράξη, οι βιολογικές αλληλουχίες

συνοδεύονται από αυξημένες απαιτήσεις υπολογισμού και μοντελοποίησης. Οι ίδιες παρατηρήσεις ισχύουν και για την πλειοψηφία των GTMs, καθώς αυτά παρουσιάζουν ανάλογα προβλήματα. Με μόνη ίσως εξαίρεση την [42], και αυτές οι τεχνικές δεν είναι σχεδιασμένες για τον απευθείας χειρισμό μη αριθμητικών ακολουθιών μεταβλητών μεγεθών και μεγάλου μήκους.

Οι τεχνικές που προτείνονται στις [47], [48] συνδυάζουν την ιδέα του γενικευμένου ενδιάμεσου και του batch υπολογισμού του SOM. Παρότι είναι πιο πολύπλοκες από τα διανυσματικά ανάλογα τους, είναι γενικής χρήσης. Η λειτουργία τους βασίζεται σε δεδομένα που περιγράφονται από προϋπολογισμένες τιμές (αν)ομοιοτήτων, όπως είναι για παράδειγμα σταθμισμένες αποστάσεις Levenshtein ή αποστάσεις τοπικών χαρακτηριστικών. Ένα μειονέκτημα του Dissimilarity SOM (DSOM) έχει να κάνει με το υπολογιστικό κόστος της εκπαίδευσης, το οποίο ισούται με το τετράγωνο του αριθμού των προτύπων εισόδου αντί να είναι γραμμικό όπως συμβαίνει στο κανονικό SOM. Στην [49] προτείνονται διάφορες τροποποιήσεις στο βασικό DSOM αλγόριθμο, που οδηγούν σε κατά μέσο όρο ταχύτερες υλοποιήσεις. Ωστόσο, η τετραγωνική φύση του αλγορίθμου δεν δύναται να αποφευχθεί, γιατί ουσιαστικά όλα τα πρότυπα εισόδου που αναπαριστώνται από τιμές (αν)ομοιοτήτων περιγράφονται εγγενώς από ένα τετραγωνικό πλήθος αποστάσεων. Ένα ακόμη μειονέκτημα του DSOM, και εν γένει της ομαδοποίησης που βασίζεται σε γενικευμένους ενδιάμεσους, είναι ότι περιορίζει τις πιθανές θέσεις των νευρώνων στο να ταυτίζονται με αυτές του χώρου εισόδου, πράγμα που συνεπάγεται σοβαρή απώλεια ακρίβειας. Από τη στιγμή που η ανανέωση που λαμβάνει χώρα δεν είναι συνεχής και μπορούν να υλοποιηθούν μόνο διακριτά βήματα προσαρμογής, τα αποτελέσματα είναι συνήθως υποδεέστερα σε σύγκριση με αυτά που λαμβάνονται από το κανονικό SOM σε αμιγώς Ευκλείδειο χώρο. Το πρόβλημα αυτό έχει αναγνωριστεί στην [50] όπου και ακολουθείται μία διαφορετική προσέγγιση.

Εν συγκρίσει με άλλες τεχνικές μηχανικής μάθησης (όπως αυτές των επαναληπτικών, αναδρομικών ή μικτών μοντέλων) τα HMMs είναι τα πλέον εύχρηστα και αποδοτικά εργαλεία για την ανάλυση ακολουθιών [14], [13]. Εάν και η θεωρία των HMMs μπορεί να έχει αξιόλογο αντίκτυπο στη διερευνητική ανάλυση δεδομένων, στις περισσότερες των περιπτώσεων η συμβολή των HMMs παραμένει περιορισμένη και αποσπασματική. Με σκοπό την αντιμετώπιση αυτού του φαινομένου, έχουν γίνει προσπάθειες συνδυασμού HMMs και NNs για τη δημιουργία υβριδικών μοντέλων που να ενσωματώνουν τη δυνατότητα προσαρμογής των NNs με τις ιδιότητες

χωρική/χρονική αναπαράσταση των HMMs. Σε πολλές από αυτές τις υβριδικές τεχνικές τα νευρωνικά δίκτυα χρησιμοποιούνται ως εμπρόσθιοι επεξεργαστές για την εξαγωγή χαρακτηριστικών και τα HMMs αξιοποιούνται, στη συνέχεια, σε ανώτερα στάδια επεξεργασίας. Εάν και αντίστροφες προσεγγίσεις είναι σπάνιες, εντούτοις υπάρχουν. Σχεδόν πάντα οι HMM και NN υπομονάδες εκπαιδεύονται ξεχωριστά και ανεξάρτητα. Αντιθέτως, στην [51] προτείνεται μία κλάση υβριδικών αρχιτεκτονικών όπου οι HMM και NN υπομονάδες είναι ενοποιημένες και αδιαίρετες. Σε αυτές τις αρχιτεκτονικές η NN υπομονάδα χρησιμοποιείται για την επαναπαραμετροποίηση και ρύθμιση του HMM τμήματος.

Στην βιβλιογραφία, μπορούν να βρεθούν ορισμένες μέθοδοι που βασίζονται σε σενάρια συνένωσης SOM και HMM υπομονάδων (για παράδειγμα [52]-[60]). Οι διαδικασίες εκπαίδευσης των δύο αυτών μονάδων είναι σχεδόν πάντα ασυσχέτιστες και λαμβάνουν χώρα ανεξάρτητα η μία της άλλης. Όλες οι μέθοδοι μπορούν να διαχωριστούν σε δύο κατηγορίες. Στην πρώτη ανήκουν προσεγγίσεις όπου το SOM χρησιμοποιείται ως εμπρόσθιος επεξεργαστής (για διανυσματική κβάντιση, προεπεξεργασία ή εξαγωγή χαρακτηριστικών) που τροφοδοτεί τις εισόδους του HMM, ενώ στη δεύτερη κλάση ανήκουν μέθοδοι που συνδέουν το SOM με τις εξόδους του HMM. Σε αυτήν την περίπτωση τα SOMs σχετίζονται με διαφορετικές καταστάσεις (ή διαφορετικές ομάδες καταστάσεων), και στη συνέχεια, συνδέονται μέσω ενός ή περισσοτέρων κρυφών επιπέδων (κατά πολυεπίπεδο τρόπο). Από όλες τις παραπάνω μεθόδους οι δύο σχετικά πιο εξελιγμένες είναι αυτές που περιγράφονται στις [58], [59]. Αμφότερες βασίζονται σε συνεχείς συναρτήσεις πυκνότητας πιθανότητας και άρα μπορούν να επεξεργάζονται αποκλειστικά αριθμητικές ακολουθίες. Σύμφωνα με την πρώτη προσέγγιση [58] κάθε HMM απαρτίζεται από ορισμένους κόμβους του SOM πλέγματος (δεν προτείνεται μία ένα-προς-ένα αντιστοίχιση μεταξύ ενός πλήρους HMM και ενός μοναδικού SOM κόμβου). Η εν λόγω μέθοδος δεν συνοδεύεται από ένα κατάλληλο και συνεπές σχήμα μη επιβλεπόμενης μάθησης. Ορισμένα ζητήματα ρύθμισης αντιμετωπίζονται από έναν batch αλγόριθμο προσαρμογής που είναι κατάλληλος για HMMs με μικρούς χώρους διαστάσεων, ενώ οι παράμετροι μετάβασης δεν ρυθμίζονται σε κανένα σημείο. Ο αλγόριθμος εκπαίδευσης της δεύτερης προσέγγισης [59] βασίζεται στον αλγόριθμο Baum-Welch [61]. Ομοίως και σε αυτήν την περίπτωση εξάγονται ορισμένες σχέσεις batch προσαρμογής, οι οποίες αποκλίνουν από τη βασική φιλοσοφία του αλγορίθμου του Kohonen. Ουσιαστικά το ζήτημα που προκύπτει είναι ότι γίνονται ορισμένες υποθέσεις οι οποίες δεν αποδεικνύονται

αναλυτικά. Οι όροι της αντίστροφης πιθανοφάνειας αντικαθίστανται από μία σειρά σταθμιζομένων συντελεστών, πράγμα το οποίο είναι αβάσιμο, αφού έχει αποδειχθεί στη βιβλιογραφία ότι στην περίπτωση πολλαπλών ακολουθιών παρατηρήσεων η σωστή τροποποίηση του αλγορίθμου Baum-Welch είναι η στάθμιση της συμβολής κάθε ακολουθίας με την αντίστροφη τιμή της πιθανοφάνειας που της αντιστοιχεί [13].

Μία μέθοδος ομαδοποίησης έχει επιτευχθεί εκπαιδευοντας παράλληλα έναν αριθμό HMMs και χρησιμοποιώντας μία μορφή ανταγωνιστικής/μη επιβλεπόμενης μάθησης για την κατασκευή ενός σύνθετου HMM [62]. Σύμφωνα με αυτήν την προσέγγιση, ο αλγόριθμος Baum-Welch χρησιμοποιείται στην πλέον ακέραια μορφή του για να διαχωρίσει με αυτόματο τρόπο αλληλουχίες μίας συγκεκριμένης οικογένειας πρωτεϊνών σε κατάλληλες συστάδες (υποκατηγορίες).

3. ΜΑΘΗΜΑΤΙΚΟ ΥΠΟΒΑΘΡΟ ΚΑΙ ΠΡΟΑΠΑΙΤΟΥΜΕΝΑ

3.1. Αλυσίδα Markov

Έστω ένα σύνολο δειγμάτων, ή αλλιώς αλφάβητο, $S = \{s_1, s_2, \dots, s_N\}$ και μία ακολουθία διακριτών τυχαίων μεταβλητών $q_t, t \in \mathbb{N}^*$ που λαμβάνουν τιμές στο S . Το σύνολο S συχνά καλείται χώρος καταστάσεων ενώ τα σύμβολα του αλφάβητου s_n αποκαλούνται καταστάσεις. Μία ακολουθία τυχαίων μεταβλητών $\{q_t\}_{t=1}^{\infty}$ ονομάζεται αλυσίδα Markov, εάν για κάθε $t > 1$ και $s_g, \dots, s_i, s_j \in S$ ισχύει

$$P(q_t = s_j | q_{t-1} = s_i, \dots, q_1 = s_g) = P(q_t = s_j | q_{t-1} = s_i). \quad (3.1)$$

Η βασική ιδέα πίσω από την αλυσίδα Markov είναι ότι δοθέντος ενός παρόντος γεγονότος $q_t = s_j$ η υπό συνθήκη πιθανότητα (δεσμευμένη) πραγματοποίησης αυτού με την προϋπόθεση του παρελθόντος ιστορικού γεγονότων $q_{t-1} = s_i, q_{t-2} = s_h, \dots, q_1 = s_g$ εξαρτάται αποκλειστικά από το άμεσο παρελθόν γεγονός $q_{t-1} = s_i$ και όχι το οποιοδήποτε απώτερο παρελθόν $q_{t-2} = s_h, \dots, q_1 = s_g$. Η συνθήκη αυτή είναι γνωστή ως ιδιότητα Markov, ή αλλιώς ως έλλειψη μνήμης.

Θεωρούμε μία αλυσίδα Markov $\{q_t\}_{t=1}^{\infty}$. Εάν $q_t = s_j$ τότε λέμε ότι η αλυσίδα βρίσκεται στην κατάσταση s_j τη χρονική στιγμή t , ή ισοδύναμα, ότι η αλυσίδα επισκέπτεται την κατάσταση s_j τη χρονική στιγμή t . Υποθέτουμε ότι οι δεσμευμένες πιθανότητες (για $t > 1$)

$$a_{ij} = P(q_t = s_j | q_{t-1} = s_i), \quad 1 \leq i \leq N, \quad 1 \leq j \leq N \quad (3.2)$$

δεν μεταβάλλονται με το χρόνο, δηλαδή ότι είναι ανεξάρτητες της παραμέτρου t και καλούνται (στάσιμες) πιθανότητες μετάβασης “ενός βήματος”. Κάθε τέτοια αλυσίδα Markov με στάσιμες πιθανότητες μετάβασης ονομάζεται ομογενής. Για τις παραμέτρους μετάβασης μεταξύ καταστάσεων ισχύουν οι ακόλουθες συνθήκες

$$a_{ij} \geq 0, 1 \leq i \leq N, 1 \leq j \leq N \quad (3.3)$$

$$\sum_{j=1}^N a_{ij} = 1, 1 \leq i \leq N \quad (3.4)$$

αφού ως δεσμευμένες πιθανότητες υπακούουν στους συνήθεις στοχαστικούς περιορισμούς. Κατά συνέπεια, ένας πίνακας πιθανοτήτων μετάβασης $A = \{a_{ij}\}$ με αυτές τις ιδιότητες καλείται στοχαστικός πίνακας.

Τη χρονική στιγμή $t=1$ η κατάσταση q_1 καθορίζεται βάσει της κατανομής πιθανότητας αρχικής κατάστασης

$$\pi_j = P(q_1 = s_j), 1 \leq j \leq N. \quad (3.5)$$

Ομοίως το διάνυσμα $\pi = \{\pi_j\}$ ικανοποιεί τους συνήθεις στοχαστικούς περιορισμούς

$$\pi_j \geq 0, 1 \leq j \leq N \quad (3.6)$$

$$\sum_{j=1}^N \pi_j = 1. \quad (3.7)$$

3.2. Παρατηρήσιμη Στοχαστική Διαδικασία

Έστω μία στοχαστική διαδικασία $\{Y_t\}_{t=1}^{\infty}$ η οποία λαμβάνει τιμές σε ένα πεπερασμένο σύνολο δειγμάτων $V = \{v_1, v_2, \dots, v_M\}$, όπου στη γενική περίπτωση το μέτρο M μπορεί να είναι διαφορετικό από το μέτρο N . Οι διαδικασίες $\{Y_t\}_{t=1}^{\infty}$ και $\{q_t\}_{t=1}^{\infty}$ σχετίζονται για κάθε $t \geq 1$ μέσω των κατανομών δεσμευμένης πιθανότητας

$$b_j(k) = P(Y_t = v_k | q_t = s_j) = P(v_k \text{ at } t | q_t = s_j), 1 \leq j \leq N, 1 \leq k \leq M. \quad (3.8)$$

Εφεξής με $B = \{b_j(k)\}$ συμβολίζεται ο πίνακας πιθανοτήτων εμφάνισης, ο οποίος είναι επίσης ένας στοχαστικός πίνακας υπό την έννοια ότι

$$b_j(k) \geq 0, 1 \leq j \leq N, 1 \leq k \leq M \quad (3.9)$$

$$\sum_{k=1}^M b_j(k) = 1, 1 \leq j \leq N. \quad (3.10)$$

3.3. Ανεξαρτησία Υπό Συνθήκη

Έστω ότι $O = o_1 o_2 \dots o_T$ είναι μία ακολουθία παρατηρήσεων όπου κάθε παρατήρηση o_t λαμβάνει μία συγκεκριμένη τιμή από το σύνολο V , ενώ με T ορίζουμε το πλήθος των παρατηρήσεων στην ακολουθία. Θεωρούμε μία δεδομένη ακολουθία καταστάσεων μήκους T : $Q = q_1 q_2 \dots q_T$. Η δεσμευμένη πιθανότητα της ακολουθίας παρατηρήσεων O δεδομένης της ακολουθίας καταστάσεων Q είναι

$$P(O | Q, B) = \prod_{t=1}^T P(o_t | q_t, B) = \prod_{t=1}^T b_{q_t}(o_t). \quad (3.11)$$

Με άλλα λόγια τα παρατηρήσιμα σύμβολα είναι υπό συνθήκη ανεξάρτητα δοθείσης της ακολουθίας των παρατηρήσεων.

3.4. Αλγόριθμος Forward-Backward

Η στοχαστική διαδικασία $\{q_t\}_{t=1}^{\infty}$ είναι αλυσίδα Markov με στάσιμες πιθανότητες μετάβασης, κατά συνέπεια η πιθανότητα μίας ακολουθίας καταστάσεων Q είναι

$$P(Q | A, \pi) = \pi_{q_1} \prod_{s=1}^{T-1} a_{q_s, q_{s+1}}. \quad (3.12)$$

Υπό αυτές τις προϋποθέσεις η συνδυασμένη πιθανότητα της O και της Q είναι

$$P(O, Q | A, B, \pi) = P(O | Q, A, B, \pi) P(Q | A, B, \pi) = P(O | Q, B) P(Q | A, \pi). \quad (3.13)$$

Στην παραπάνω σχέση η δεύτερη ισότητα προκύπτει από το γεγονός ότι η ακολουθία παρατηρήσεων O είναι ανεξάρτητη της μήτρας πιθανοτήτων μετάβασης και της

κατανομής πιθανότητας αρχικής κατάστασης, και επιπλέον, από το γεγονός ότι η ακολουθία καταστάσεων Q είναι ανεξάρτητη της μήτρας πιθανοτήτων εμφάνισης.

Ακολουθώς μπορούμε να υπολογίσουμε την πιθανότητα μίας ακολουθίας παρατηρήσεων O κάνοντας χρήση του γενικευμένου προσθετικού θεωρήματος (ή αλλιώς του θεωρήματος ολικών πιθανοτήτων) αθροίζοντας τις πιθανότητες για όλες τις δυνατές ακολουθίες καταστάσεων (οι οποίες προφανώς συνιστούν ασυμβίβαστα ενδεχόμενα). Έτσι προκύπτει

$$P(O | A, B, \pi) = \sum_{\text{all } Q} P(O, Q | A, B, \pi) = \sum_{\text{all } q_1, q_2, \dots, q_T} \left[\prod_{t=1}^T b_{q_t}(o_t) \pi_{q_1} \prod_{s=1}^{T-1} a_{q_s, q_{s+1}} \right] = \sum_{\text{all } q_1, q_2, \dots, q_T} \pi_{q_1} b_{q_1}(o_1) a_{q_1, q_2} b_{q_2}(o_2) \dots a_{q_{T-1}, q_T} b_{q_T}(o_T). \quad (3.14)$$

Σε αυτή τη σχέση η δεύτερη ισότητα είναι άμεση απόρροια των (3.11)-(3.13).

Είναι χρήσιμο να επισημάνουμε ότι σε συγκεκριμένες περιπτώσεις ως δείκτες των πιθανοτήτων μετάβασης, εμφάνισης και αρχικής κατάστασης χρησιμοποιούνται οι q_i και o_i αντί για τους i , j και k . Αυτός ο εναλλακτικός συμβολισμός χρησιμοποιείται οπουδήποτε οι ακριβείς τιμές των καταστάσεων και των παρατηρήσιμων σύμβολων είναι επουσιώδεις και δεν απαιτούνται, παρότι βεβαίως οι τιμές τους θεωρούνται δεδομένες και γνωστές.

Είναι πλέον σαφές ότι ένα HMM συνιστά μία διπλά ενσωματωμένη στοχαστική διαδικασία με μία υποκείμενη στοχαστική διεργασία που δεν είναι παρατηρήσιμη (είναι κρυφή κατά κάποιο τρόπο) και η οποία αποκαλύπτεται εμμέσως μέσω ενός άλλου συνόλου στοχαστικών διεργασιών που παράγουν τις ακολουθίες παρατηρήσεων. Συνολικά, μία πλήρης περιγραφή ενός HMM απαιτεί τον προσδιορισμό του χώρου καταστάσεων, του συνόλου δειγμάτων (συμπεριλαμβανομένων των παρατηρήσιμων συμβόλων), της μήτρας πιθανοτήτων μετάβασης A , της μήτρας πιθανοτήτων εμφάνισης B και της κατανομής πιθανότητας αρχικής κατάστασης π . Εφεξής για λόγους συντομίας για κάθε ξεχωριστό HMM θα χρησιμοποιούμε τον συμβολισμό

$$\lambda = (A, B, \pi). \quad (3.15)$$

Ένα κατάλληλα σχεδιασμένο HMM λ μπορεί να χρησιμοποιηθεί ως γεννήτρια μίας ακολουθίας παρατηρήσεων $O = o_1 o_2 \dots o_T$, με τον τρόπο που περιγράφεται στη συνέχεια:

- 1) επιλέγουμε μία αρχική κατάσταση $q_1 = s_i$ με τυχαίο τρόπο βάσει της κατανομής πιθανότητας αρχικής κατάστασης π ,
- 2) θέτουμε $t = 1$,
- 3) επιλέγουμε ένα παρατηρήσιμο σύμβολο $o_t = v_k$ με τυχαίο τρόπο βάσει της κατανομής πιθανότητας εμφάνισης b_i που αντιστοιχεί στην παρούσα κατάσταση s_i ,
- 4) μεταβαίνουμε στην επόμενη κατάσταση $q_{t+1} = s_j$ με τυχαίο τρόπο βάσει της κατανομής πιθανότητας μετάβασης a_i από την παρούσα κατάσταση s_i σε μία νέα κατάσταση,
- 5) θέτουμε $s_i = s_j$ και $t = t + 1$,
- 6) εφόσον ισχύει ότι $t \leq T$ επιστρέφουμε στο βήμα 3, αλλιώς ολοκληρώνεται η διαδικασία.

Η ανωτέρω διαδικασία μπορεί επίσης να χρησιμοποιηθεί για να μελετηθεί το πως μία δεδομένη ακολουθία παρατηρήσεων παρήχθη από κάποιο HMM.

Μία διαδικασία που συστηματοποιεί και επιταχύνει τους απαιτούμενους υπολογισμούς για ένα HMM δίνεται από τον αλγόριθμο forward-backward. Αυτός ο αλγόριθμος παρέχει μία υπολογιστικά εύρωστη λύση στο πρόβλημα της αποτίμησης (εύρεσης) για ένα πλήρως συνεκτικό κανονικό HMM. Επιπροσθέτως, αυτός ο αλγόριθμος παρέχει τα βασικά εργαλεία για την κατασκευή άλλων αλγορίθμων που είναι απαραίτητοι για την εκπαίδευση ενός HMM. Συγκεκριμένα, το πρόβλημα της εκπαίδευσης ενός HMM μπορεί να οριστεί και να διατυπωθεί με βάση τις ποσότητες που εισάγονται από τη διαδικασία forward-backward.

Θεωρούμε τη forward μεταβλητή $\alpha_t(i)$ που ορίζεται ως η ταυτόχρονη πιθανότητα εμφάνισης της ακολουθίας παρατηρήσεων μέχρι τη χρονική στιγμή $t \leq T$ και της αλυσίδας Markov να βρίσκεται στην κατάσταση s_i τη χρονική στιγμή t , δοθέντος του μοντέλου λ

$$\alpha_t(i) = P(o_1 o_2 \dots o_t, q_t = s_i | \lambda). \quad (3.16)$$

Μία αναδρομική λύση για τη $\alpha_t(i)$ είναι η ακόλουθη

/* αρχικοποίηση */

$$\alpha_1(i) = \pi_i b_i(o_1), 1 \leq i \leq N \quad (3.17)$$

/* αναδρομή */

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(o_{t+1}), 1 \leq t \leq T-1, 1 \leq j \leq N. \quad (3.18)$$

Κατά παρόμοιο τρόπο, θεωρούμε τη backward μεταβλητή $\beta_t(i)$ που ορίζεται ως η πιθανότητα εμφάνισης της ακολουθίας παρατηρήσεων από τη χρονική στιγμή $t+1$ μέχρι το τέλος T υπό την προϋπόθεση ότι η αλυσίδα Markov βρίσκεται στην κατάσταση s_i τη χρονική στιγμή t , δοθέντος του μοντέλου λ

$$\beta_t(i) = P(o_{t+1} o_{t+2} \dots o_T | q_t = s_i, \lambda). \quad (3.19)$$

Ομοίως με πριν, μπορούμε να λύσουμε αναδρομικά ως εξής

/* αρχικοποίηση */

$$\beta_T(i) = 1, 1 \leq i \leq N \quad (3.20)$$

/* αναδρομή */

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), t = T-1, T-2, \dots, 1, 1 \leq i \leq N. \quad (3.21)$$

Ο αλγόριθμος forward-backward που περιγράφηκε προηγούμενα, επιτρέπει την αποτίμηση της πιθανότητας $P(O | \lambda)$ κατά τέτοιο τρόπο ώστε οι υπολογιστικές απαιτήσεις να είναι γραμμικώς ανάλογες ως προς το μήκος της ακολουθίας παρατηρήσεων

$$P(O | \lambda) = \sum_{j=1}^N \alpha_t(j) \beta_t(j), 1 \leq t \leq T. \quad (3.22)$$

Αξίζει να σημειωθεί ότι για κάθε δυνατή τιμή του t προκύπτει ένας αντίστοιχος τρόπος υπολογισμού της $P(O | \lambda)$. Για παράδειγμα, για $t = T$ έχουμε

$$P(O | \lambda) = \sum_{j=1}^N \alpha_T(j) \quad (3.23)$$

αφού βάσει σύμβασης ισχύει ότι: $\beta_T(j) = 1, 1 \leq j \leq N$.

3.5. Συμπερασμός Bayes και Εκτίμηση Παραμέτρων

Το παράδειγμα του Bayes παρέχει μία ισχυρή προσέγγιση για το χειρισμό και την ποσοτικοποίηση της αβεβαιότητας που περιβάλλει την επιλογή του διάνυσματος παραμέτρων ενός μοντέλου. Σε αυτή τη φάση θα εστιάσουμε στο είδος του συμπερασμού που μας ενδιαφέρει πρώτιστα. Αυτό είναι η εξαγωγή ενός παραμετροποιημένου HMM λ από μία συλλογή δεδομένων O , μία εκτενέστερη ανάλυση περιέχεται στην [2]. Από το θεώρημα του Bayes άμεσα προκύπτει ότι

$$P(\lambda | O) = P(O | \lambda)P(\lambda)/P(O). \quad (3.24)$$

Η εκ των προτέρων (prior) πιθανότητα $P(\lambda)$ αντιπροσωπεύει την εκτίμηση της πιθανότητας ότι το HMM λ είναι σωστό προτού λάβουμε οποιαδήποτε γνώση των δεδομένων O , ισοδύναμα μπορούμε να πούμε ότι η $P(\lambda)$ αντιπροσωπεύει τη συνάρτηση πυκνότητας πιθανότητας η οποία αναφέρεται στην πληροφορία σχετικά με το διάνυσμα παραμέτρων πριν γίνουν οποιοσδήποτε παρατηρήσεις στο περιβάλλον. Η εκ των υστέρων (posterior) πιθανότητα $P(\lambda | O)$ αντιπροσωπεύει την αναθεωρημένη εκτίμηση της πιθανότητας ότι το HMM λ είναι σωστό δοθέντος ότι έχουμε λάβει γνώση, δηλαδή ότι έχουμε παρατηρήσει, τη συλλογή δεδομένων O , ισοδύναμα μπορούμε να θεωρήσουμε ότι η $P(\lambda | O)$ αντιπροσωπεύει την υπό συνθήκη συνάρτηση πυκνότητας πιθανότητας η οποία αναφέρεται στο διάνυσμα παραμέτρων αφού ολοκληρωθεί η παρατήρηση του περιβάλλοντος. Ο όρος $P(O | \lambda)$ συχνά αναφέρεται

ως πιθανοφάνεια και αντιπροσωπεύει την υπό συνθήκη συνάρτηση πυκνότητας πιθανότητας που αναφέρεται στην παρατήρηση των δεδομένων O (ή της απόκρισης του περιβάλλοντος) με δεδομένο το διάνυσμα παραμέτρων του HMM λ . Η μαρτυρία (evidence) $P(O)$ αναφέρεται στα στατιστικά χαρακτηριστικά της συλλογής δεδομένων O και δεν σχετίζεται με το HMM, ουσιαστικά αντιπροσωπεύει τη συνάρτηση πυκνότητας πιθανότητας η οποία αναφέρεται στην πληροφορία που περιέχεται στην απόκριση του περιβάλλοντος.

Επιπλέον, αναφορικά με την εκτίμηση του διανύσματος παραμέτρων, η $P(O)$ στον παρονομαστή της δεξιάς πλευράς της (3.24) παίζει απλώς το ρόλο μίας σταθεράς κανονικοποίησης η οποία δεν εξαρτάται από τις παραμέτρους του HMM και κατά συνέπεια δεν επηρεάζει σε κανένα σημείο τη διαδικασία της εκτίμησης. Συνεπώς από τη στιγμή που η εκ των προτέρων πιθανότητα $P(\lambda)$ και η πιθανοφάνεια $P(O|\lambda)$ είναι ορισμένες και σαφείς η αρχική διαδικασία μοντελοποίησης ολοκληρώνεται

$$P(\lambda|O) \sim P(O|\lambda)P(\lambda) \quad (3.25)$$

η πυκνότητα της εκ των υστέρων πιθανότητας του παραμετροποιημένου HMM λ είναι ανάλογη προς το γινόμενο της συνάρτησης πιθανοφάνειας επί την εκ των προτέρων πιθανότητα. Το μόνο που απομένει είναι η εφαρμογή της θεωρίας πιθανοτήτων.

Ο συνηθέστερος αντικειμενικός σκοπός είναι η εύρεση ή η προσέγγιση του βέλτιστου HMM από μία ομοταξία, τουτέστιν να βρεθεί το σύνολο των παραμέτρων που μεγιστοποιεί την εκ των υστέρων πιθανότητα $P(\lambda|O)$. Η διαδικασία αυτή αναφέρεται ως μέγιστη a posteriori εκτίμηση - Maximum A Posteriori (MAP) estimation και μπορεί να διατυπωθεί ως εξής

$$\lambda_{MAP} = \arg \max_{\lambda} \{P(\lambda|O)\}. \quad (3.26)$$

Από σκοπιάς βελτιστοποίησης η εκ των προτέρων πιθανότητα $P(\lambda)$ παίζει το ρόλο ενός ρυθμιστή, δηλαδή ενός επιπλέον όρου για την επιβολή επιπρόσθετων περιορισμών, όπως για παράδειγμα η εξομάλυνση. Ωστόσο, εάν η εκ των προτέρων πιθανότητα $P(\lambda)$ είναι ομοιόμορφη ως προς τη θεωρούμενη ομοταξία HMMs, τότε το πρόβλημα απλοποιείται στην εύρεση της μέγιστης πιθανοφάνειας $P(O|\lambda)$ μεταξύ των

HMMs της ομοταξίας. Η συνάρτηση πιθανοφάνειας, εξεταζόμενη από μόνη της, παρέχει τη βάση για την εκτίμηση μέγιστης πιθανοφάνειας - Maximum Likelihood (ML) estimation, όπως υποδεικνύει η

$$\lambda_{ML} = \arg \max_{\lambda} \{P(O | \lambda)\}. \quad (3.27)$$

Για μία πιο εμβριθή εκτίμηση του συνόλου των παραμέτρων, ωστόσο καταφεύγουμε στη MAP εκτίμηση. Το παράδειγμα του Bayes για την εκτίμηση παραμέτρων, το οποίο εδράζεται στο θεώρημα του Bayes, όπως παρουσιάζεται στην (3.24) και αποτυπώνεται στη MAP εκτίμηση (3.26), αξιοποιεί όλη την πιθανή (ενδεχόμενα εικάσιμη) πληροφορία σχετικά με το σύνολο παραμέτρων. Εν αντιθέσει, η ML εκτίμηση κινείται στις παρυφές του παραδείγματος Bayes, αγνοώντας την εκ των προτέρων πιθανότητα. Φυσικά, η πρόκληση όσον αφορά την εφαρμογή της διαδικασίας εκτίμησης MAP έγκειται στον τρόπο με τον οποίο θα καθορίσουμε κατάλληλα την εκ των προτέρων πιθανότητα, πράγμα το οποίο καθιστά τη διαδικασία MAP πιο απαιτητική από υπολογιστικής σκοπιάς συγκριτικά με τη ML. Μάλιστα σε ορισμένες περιπτώσεις η εύρεση ή ακόμη και η εκτίμηση της εκ των προτέρων πιθανότητας είναι ανέφικτη, γεγονός που καθιστά τη διαδικασία ML μοναδική επιλογή.

Για τεχνικούς λόγους οι επιμέρους πιθανότητες μπορεί να λάβουν πολύ μικρές τιμές. Έτσι από υπολογιστικής σκοπιάς, συνήθως είναι βολικότερο (ή σε κάποιες περιπτώσεις αναπόφευκτο) να δουλεύουμε με τους αντίστοιχους λογάριθμους

$$\log P(\lambda | O) = \log P(O | \lambda) + \log P(\lambda) - \log P(O). \quad (3.28)$$

Αυτό μπορούμε να το κάνουμε διότι ο λογάριθμος είναι μία γνησίως αύξουσα συνάρτηση του ορίσματός της. Κατά συνέπεια μπορούμε να εκφράσουμε τη ML εκτίμηση στην επιθυμητή μορφή γράφοντας

$$\lambda_{ML} = \arg \max_{\lambda} \{\log P(O | \lambda)\}. \quad (3.29)$$

Προφανώς αντίστοιχη διατύπωση ισχύει για τη MAP εκτίμηση.

3.6. Gradient Descent

Στα περισσότερα ενδιαφέροντα μοντέλα, η συνάρτηση που βελτιστοποιείται είναι σύνθετη και πολύπλοκη με αποτέλεσμα να μην επιδέχεται αναλυτική λύση. Έτσι, κανείς πρέπει να καταφύγει σε επαναληπτικές και ενδεχομένως στοχαστικές μεθόδους όπως η κατάβαση κλίσης - gradient descent [63], [64] ή η προσομοιωμένη απόπτωση - simulated annealing, συχνά επίσης πρέπει να αρκεστεί σε προσεγγιστικές ή υποβέλτιστες λύσεις. Τέτοιες και παρεμφερείς διαδικασίες έχειδειχθεί ότι παράγουν λύσεις συγκρίσιμες με αυτές άλλων εκτιμητριών διαδικασιών.

Βάσει όσων αναφέρθηκαν προηγούμενα, ενδιαφερόμαστε για την εκτίμηση παραμέτρων, δηλαδή για την εύρεση του καλύτερου δυνατού HMM $\lambda = \lambda(x)$ το οποίο να ελαχιστοποιεί μία δεδομένη συνάρτηση κόστους $f(x)$. Στην περίπτωση που η αντικειμενική συνάρτηση κόστους $f(x)$ είναι διαφορίσιμη, μπορεί κανείς να προσπαθήσει να βρει (ή να προσεγγίσει) τα ελάχιστα της χρησιμοποιώντας μία από τις παλαιότερες και κλασικότερες μεθόδους βελτιστοποίησης, την gradient descent. Όπως υποδηλώνει και το όνομα της, η gradient descent είναι μία επαναληπτική διαδικασία, όπου η παράμετρος που μας ενδιαφέρει, στην προκείμενη περίπτωση η x , ρυθμίζεται σύμφωνα με τον κανόνα

$$x^{(next)} = x^{(now)} - \eta \nabla f(x) \Big|_{x=x^{(now)}} \quad (3.30)$$

όπου το εύρος του βήματος ή αλλιώς ο ρυθμός μάθησης η δύναται να παραμένει σταθερός ή να αναπροσαρμόζεται κατά τη διάρκεια της διαδικασίας μάθησης. Επιπλέον, η σύγκλιση της μεθόδου gradient descent μπορεί να εξασφαλιστεί όταν ο ρυθμός μάθησης η ικανοποιεί συγκεκριμένες προδιαγραφές [65], [66].

Εάν και η βασική αρχή της gradient descent είναι απλή, σε σύνθετα παραμετροποιημένα μοντέλα μπορεί να οδηγήσει σε διαφορετικές υλοποιήσεις ανάλογα με το πως ακριβώς υπολογίζεται η κλίση. Σε μοντέλα γράφων, αυτό συχνά απαιτεί τη διάδοση της πληροφορίας “προς τα πίσω”. Μία από αυτές τις περιπτώσεις είναι ο αλγόριθμος forward-backward των HMMs. Προφανώς το αποτέλεσμα μίας διαδικασίας gradient descent εξαρτάται από την αρχική εκτίμηση ή αλλιώς την αρχικοποίηση. Επιπροσθέτως εάν η συνάρτηση που βελτιστοποιείται έχει ένα πολύπλοκο σύνολο τιμών, η gradient descent εν γένει καταλήγει σε τοπικά ελάχιστα

παρά σε ολικά. Επομένως, εφόσον αυτό είναι εφικτό, ενδείκνυται να εκτελείται η διαδικασία έναν ικανοποιητικό αριθμό φορών με άλλες κάθε φορά αρχικές τιμές και διαφορετικούς ρυθμούς μάθησης.

3.7. Εκθετική Επαναπαραμετροποίηση

Ένας ιδιαίτερος χρήσιμος μετασχηματισμός του συνόλου των παραμέτρων ενός HMM δύναται να εξαχθεί απευθείας με τη χρήση εκθετικών παραγόντων. Επαναπαραμετροποιούμε το HMM χρησιμοποιώντας κανονικοποιημένες εκθετικές συναρτήσεις της μορφής

$$a_{ij} = e^{w_{ij}} / \sum_{l=1}^N e^{w_{il}}, 1 \leq i \leq N, 1 \leq j \leq N \quad (3.31)$$

$$b_j(t) = e^{r_{jt}} / \sum_{l=1}^M e^{r_{jl}}, 1 \leq j \leq N, 1 \leq t \leq M \quad (3.32)$$

$$\pi_j = e^{u_j} / \sum_{l=1}^N e^{u_l}, 1 \leq j \leq N \quad (3.33)$$

όπου οι μεταβλητές w_{ij} , r_{jt} και u_j συνιστούν τις νέες παραμέτρους [67]. Αυτή η επαναπαραμετροποίηση έχει τρία κύρια πλεονεκτήματα. Πρώτον, οποιαδήποτε τροποποίηση ή αλλαγή στις νέες μεταβλητές αυτόματα ικανοποιεί τους συνήθεις στοχαστικούς περιορισμούς για τις κατανομές πιθανότητας μετάβασης, εμφάνισης και αρχικής κατάστασης. Δεύτερον, οι επιμέρους πιθανότητες των τριών αυτών κατανομών δεν μηδενίζονται ποτέ, παρότι δεν αποτρέπονται από το να λάβουν εξαιρετικά μικρές τιμές. Τρίτον, στην περίπτωση που υπάρχει μόνο μία ακολουθία εκπαίδευσης, ή ισοδύναμα στην περίπτωση μάθησης on-line, αποτρέπονται έντονες ασυνέχειες και απότομα άλματα κατά τη διερεύνηση του χώρου των παραμέτρων. Εξαιτίας αυτών των προτερημάτων η εν λόγω επαναπαραμετροποίηση ξεπερνά τους περιορισμούς του αλγορίθμου Baum-Welch [61] καθώς και άλλων παρεμφερών E-M μεθόδων.

Γενικώς, οι αλγόριθμοι που βασίζονται σε αυτή την επαναπαραμετροποίηση αποδεικνύονται χρήσιμοι και επωφελείς σε καταστάσεις που η εξομάλυνση ή/και η

μάθηση on-line καθίστανται κρίσιμης σημασίας. Αυτές οι καταστάσεις μπορούν να αφορούν:

- 1) μοντέλα μεγάλης κλίμακας, με μεγάλο αριθμό παραμέτρων, σχετικώς αραιά και ανεπαρκή δεδομένα, τα οποία μπορεί να είναι επιρρεπή σε υπερεκπαίδευση, παγίδευση σε τοπικά βέλτιστα και άλλες παθολογικές συμπεριφορές όταν εκπαιδεύονται με ασυνεχείς ή batch αλγορίθμους
- 2) αναλογικές υλοποιήσεις όπου μόνο συνεχείς κανόνες μάθησης δύνανται να επιτευχθούν
- 3) όλες τις περιπτώσεις στις οποίες η αποθήκευση δεδομένων για μάθηση batch δεν είναι επιθυμητή ή εφικτή.

Όπως αναφέρθηκε προηγούμενα, σε συγκεκριμένες περιπτώσεις ο αλγόριθμος Baum-Welch είναι προβληματικός [13]. Για παράδειγμα στην περίπτωση μίας μεμονωμένης ακολουθίας εκπαίδευσης ο αλγόριθμος, σε κάθε επανάληψη, επαναρχικοποιεί κάθε παράμετρο μετάβασης και εμφάνισης με την τιμή της αναμενόμενης συχνότητας της λαμβάνοντας υπόψη το τρέχον μοντέλο και το σύνολο εκπαίδευσης. Επομένως, είναι σαφές ότι ο αλγόριθμος Baum-Welch μπορεί να οδηγήσει σε απότομα άλματα και έντονες ασυνέχειες κατά την εξερεύνηση του χώρου παραμέτρων, πράγμα που καθιστά τη διαδικασία ακατάλληλη για μάθηση on-line. Ένα άλλο πρόβλημα του αλγορίθμου Baum-Welch είναι ότι οι μηδενικές πιθανότητες, ή αυτές που μηδενίζονται σε κάποια φάση, δεν μπορούν να αλλάξουν τιμή. Από τη στιγμή που μία πιθανότητα μετάβασης ή παρατήρησης μηδενιστεί παραμένει μηδέν για πάντα χωρίς να είναι σε θέση ο αλγόριθμος να την επαναφέρει σε κάποια άλλη τιμή. Μία τέτοια κατάσταση είναι φυσικά ανεπιθύμητη και συνήθως αποτρέπεται επιβάλλοντας τεχνητούς περιορισμούς όπως το ότι καμία παράμετρος δεν μπορεί να είναι μικρότερη από μία σταθερή μικρή τιμή κατωφλίου.

Είναι σαφές ότι οι νέες μεταβλητές w_{ij} , r_{jt} , u_j μπορούν να οργανωθούν σε μία σειρά από πίνακες, τους $W = \{w_{ij}\}$, $R = \{r_{jt}\}$, $U = \{u_j\}$. Αντιστοίχως, μία πλήρης περιγραφή ενός HMM απαιτεί τον προσδιορισμό του χώρου καταστάσεων, του συνόλου δειγμάτων (συμπεριλαμβανομένων των παρατηρήσιμων συμβόλων) και των στοχαστικών μήτρων W , R , U . Εφεξής για λόγους συντομίας για κάθε ξεχωριστό HMM θα χρησιμοποιούμε τον συμβολισμό

$$\lambda = (W, R, U). \quad (3.34)$$

3.8. Παράγωγοι της Πιθανοφάνειας

Στην συνέχεια διατυπώνονται και αποδεικνύονται αναλυτικά οι μαθηματικές σχέσεις που αντιστοιχούν στις μερικές παραγώγους της πιθανοφάνειας $P(O|\lambda)$ ως προς τις παραμέτρους του υπό εξέταση HMM. Δυνητικά αυτές οι εξισώσεις μπορούν να βρουν μία πλειάδα εφαρμογών σε μοντέλα που περιέχουν ή αποτελούνται από HMMs για τα οποία απαιτείται κατάλληλη εκπαίδευση και ρύθμιση.

Σε αυτό το σημείο για να περιοριστούν, κατά το δυνατόν, οι παραπομπές επαναδιατυπώνουμε τους τύπους αναδρομικού υπολογισμού των forward και backward μεταβλητών

$$\alpha_1(j) = \pi_j b_j(o_1), 1 \leq j \leq N \quad (3.35)$$

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(o_{t+1}), 1 \leq t \leq T-1, 1 \leq j \leq N \quad (3.36)$$

$$\beta_T(j) = 1, 1 \leq j \leq N \quad (3.37)$$

$$\beta_t(j) = \sum_{i=1}^N a_{ji} b_i(o_{t+1}) \beta_{t+1}(i), t = T-1, T-2, \dots, 1, 1 \leq j \leq N. \quad (3.38)$$

Στην ανάλυση του προβλήματος της αποτίμησης της $P(O|\lambda)$ με χρήση των forward και backward μεταβλητών, δείξαμε ότι υπάρχουν T ισοδύναμοι τρόποι υπολογισμού της πιθανοφάνειας

$$P(O|\lambda) = \sum_{j=1}^N \alpha_t(j) \beta_t(j), 1 \leq t \leq T. \quad (3.39)$$

Για $t = T$ ο τύπος του υπολογισμού παίρνει τη μορφή

$$P(O|\lambda) = \sum_{j=1}^N \alpha_T(j) \beta_T(j). \quad (3.40)$$

Στη συνέχεια χρησιμοποιώντας την αναδρομική σχέση (3.36) για την $\alpha_T(j)$ έχουμε

$$P(O | \lambda) = \sum_{j=1}^N \sum_{i=1}^N \alpha_{T-1}(i) \mathbf{a}_{ij} b_j(o_T) \beta_T(j). \quad (3.41)$$

Λήμμα 1

$$\frac{\partial P(O | \lambda)}{\partial \mathbf{a}_{rs}} = \sum_{l=1}^{T-1} \alpha_l(r) b_s(o_{l+1}) \beta_{l+1}(s). \quad (3.42)$$

Απόδειξη: αρχικά βρίσκουμε τη μερική παράγωγο της (3.41) ως προς \mathbf{a}_{rs} , το αποτέλεσμα είναι

$$\frac{\partial P(O | \lambda)}{\partial \mathbf{a}_{rs}} = \alpha_{T-1}(r) b_s(o_T) \beta_T(s) + \sum_{j=1}^N \sum_{i=1}^N \frac{\partial \alpha_{T-1}(i)}{\partial \mathbf{a}_{rs}} \mathbf{a}_{ij} b_j(o_T) \beta_T(j). \quad (3.43)$$

Εντεύθεν, χρειάζεται να υπολογίσουμε μόνο το δεύτερο όρο στο δεξί μέλος της σχέσης. Αυτό επιτυγχάνεται χρησιμοποιώντας ακόμη μία φορά την αναδρομική σχέση (3.36)

$$\frac{\partial \alpha_{T-1}(i)}{\partial \mathbf{a}_{rs}} = \frac{\partial}{\partial \mathbf{a}_{rs}} \left[\sum_{l=1}^N \alpha_{T-2}(l) \mathbf{a}_{li} \right] b_i(o_{T-1}). \quad (3.44)$$

Εκτελώντας την παραγωγή λαμβάνουμε

$$\frac{\partial \alpha_{T-1}(s)}{\partial \mathbf{a}_{rs}} = \alpha_{T-2}(r) b_s(o_{T-1}) + \left[\sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial \mathbf{a}_{rs}} \mathbf{a}_{ls} \right] b_s(o_{T-1}) \quad (3.45)$$

$$\frac{\partial \alpha_{T-1}(i)}{\partial \mathbf{a}_{rs}} = \left[\sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial \mathbf{a}_{rs}} \mathbf{a}_{li} \right] b_i(o_{T-1}), \quad i \neq s. \quad (3.46)$$

Εισάγουμε αυτές τις μερικές παραγώγους στο δεύτερο όρο της (3.43) και το αποτέλεσμα είναι

$$\begin{aligned}
 \sum_{j=1}^N \sum_{i=1}^N \frac{\partial \alpha_{T-1}(i)}{\partial \mathbf{a}_{rs}} \mathbf{a}_{ij} b_j(o_T) \beta_T(j) &= \sum_{j=1}^N \alpha_{T-2}(r) b_s(o_{T-1}) \mathbf{a}_{sj} b_j(o_T) \beta_T(j) + \\
 &\sum_{j=1}^N \left[\sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial \mathbf{a}_{rs}} \mathbf{a}_{ls} \right] b_s(o_{T-1}) \mathbf{a}_{sj} b_j(o_T) \beta_T(j) + \\
 &\sum_{j=1}^N \sum_{i=1, i \neq s}^N \left[\sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial \mathbf{a}_{rs}} \mathbf{a}_{li} \right] b_i(o_{T-1}) \mathbf{a}_{ij} b_j(o_T) \beta_T(j). \quad (3.47)
 \end{aligned}$$

Συνενώνουμε τους δύο τελευταίους όρους στο δεξί μέλος της (3.47) και έχουμε

$$\begin{aligned}
 \sum_{j=1}^N \sum_{i=1}^N \frac{\partial \alpha_{T-1}(i)}{\partial \mathbf{a}_{rs}} \mathbf{a}_{ij} b_j(o_T) \beta_T(j) &= \sum_{j=1}^N \alpha_{T-2}(r) b_s(o_{T-1}) \mathbf{a}_{sj} b_j(o_T) \beta_T(j) + \\
 &\sum_{j=1}^N \sum_{i=1}^N \sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial \mathbf{a}_{rs}} \mathbf{a}_{li} b_i(o_{T-1}) \mathbf{a}_{ij} b_j(o_T) \beta_T(j). \quad (3.48)
 \end{aligned}$$

Ακολούθως χρειάζεται να υπολογιστούν οι δύο όροι στο δεξί μέλος της (3.48). Για τον πρώτο όρο ισχύει

$$\sum_{j=1}^N \alpha_{T-2}(r) b_s(o_{T-1}) \mathbf{a}_{sj} b_j(o_T) \beta_T(j) = \alpha_{T-2}(r) b_s(o_{T-1}) \sum_{j=1}^N \mathbf{a}_{sj} b_j(o_T) \beta_T(j). \quad (3.49)$$

Χρησιμοποιώντας σε αυτό το στάδιο την αναδρομική σχέση (3.38) προκύπτει ότι

$$\alpha_{T-2}(r) b_s(o_{T-1}) \sum_{j=1}^N \mathbf{a}_{sj} b_j(o_T) \beta_T(j) = \alpha_{T-2}(r) b_s(o_{T-1}) \beta_{T-1}(s). \quad (3.50)$$

Για το δεύτερο όρο στο δεξί μέλος της (3.48) εναλλάσσοντας τη σειρά των αθροισμάτων λαμβάνουμε

$$\begin{aligned}
 \sum_{j=1}^N \sum_{i=1}^N \sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial \mathbf{a}_{rs}} \mathbf{a}_{li} b_i(o_{T-1}) \mathbf{a}_{ij} b_j(o_T) \beta_T(j) &= \\
 \sum_{i=1}^N \sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial \mathbf{a}_{rs}} \mathbf{a}_{li} b_i(o_{T-1}) \sum_{j=1}^N \mathbf{a}_{ij} b_j(o_T) \beta_T(j). \quad (3.51)
 \end{aligned}$$

Με τη χρήση της αναδρομικής σχέσης (3.38) έχουμε ότι

$$\sum_{i=1}^N \sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial \mathbf{a}_{rs}} \mathbf{a}_{li} b_i(o_{T-1}) \sum_{j=1}^N \mathbf{a}_{ij} b_j(o_T) \beta_T(j) = \sum_{i=1}^N \sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial \mathbf{a}_{rs}} \mathbf{a}_{li} b_i(o_{T-1}) \beta_{T-1}(i). \quad (3.52)$$

Με τη βοήθεια των σχέσεων (3.47)-(3.52) μπορούμε να ξαναγράψουμε το δεξί μέλος της (3.43) ως εξής

$$\begin{aligned} \frac{\partial P(O|\lambda)}{\partial \mathbf{a}_{rs}} &= \alpha_{T-1}(r) b_s(o_T) \beta_T(s) + \alpha_{T-2}(r) b_s(o_{T-1}) \beta_{T-1}(s) + \\ &\quad \sum_{i=1}^N \sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial \mathbf{a}_{rs}} \mathbf{a}_{li} b_i(o_{T-1}) \beta_{T-1}(i). \end{aligned} \quad (3.53)$$

Ωστόσο, ένας μπορεί να παρατηρήσει ότι το παραπάνω διπλό άθροισμα έχει την ίδια ακριβώς μορφή με αυτή του διπλού αθροίσματος στην (3.43). Εφεξής, μπορούμε να συνεχίσουμε παραγωγίζοντας τα $\frac{\partial \alpha_{T-2}(l)}{\partial \mathbf{a}_{rs}}$ και οργανώνοντας τις προκύπτουσες

σχέσεις με τον ίδιο τρόπο που παρουσιάστηκε προηγούμενα. Εξαιτίας του γεγονότος ότι η μεταβλητή $\alpha_1(l)$ δεν περιλαμβάνει κανένα παράγοντα \mathbf{a}_{rs} , διαπιστώνουμε ότι η παραγωγή σταματάει στο σημείο που εμφανίζεται ο παράγοντας $\frac{\partial \alpha_1(l)}{\partial \mathbf{a}_{rs}}$, μιας και για

αυτόν ισχύει ότι $\frac{\partial \alpha_1(l)}{\partial \mathbf{a}_{rs}} = 0$. Συνολικά επομένως θα έχουμε

$$\frac{\partial P(O|\lambda)}{\partial \mathbf{a}_{rs}} = \alpha_{T-1}(r) b_s(o_T) \beta_T(s) + \alpha_{T-2}(r) b_s(o_{T-1}) \beta_{T-1}(s) + \dots + \alpha_1(r) b_s(o_2) \beta_2(s). \quad (3.54)$$

Συνεπώς η σχέση (3.42) αποδείχθηκε.

Θεώρημα 1

$$\frac{\partial P(O | \lambda)}{\partial w_{ij}} = \mathbf{a}_{ij} \sum_{l=1}^{T-1} [\alpha_l(i) b_j(o_{l+1}) \beta_{l+1}(j) - \alpha_l(i) \beta_l(i)]. \quad (3.55)$$

Απόδειξη: εφαρμόζοντας τον κανόνα της αλυσίδας στο αριστερό μέλος της (3.55) έχουμε

$$\frac{\partial P(O | \lambda)}{\partial w_{ij}} = \sum_{z=1}^N \frac{\partial P(O | \lambda)}{\partial \mathbf{a}_{iz}} \frac{\partial \mathbf{a}_{iz}}{\partial w_{ij}} = \frac{\partial P(O | \lambda)}{\partial \mathbf{a}_{ij}} \frac{\partial \mathbf{a}_{ij}}{\partial w_{ij}} + \sum_{z=1, z \neq j}^N \frac{\partial P(O | \lambda)}{\partial \mathbf{a}_{iz}} \frac{\partial \mathbf{a}_{iz}}{\partial w_{ij}}. \quad (3.56)$$

Βασιζόμενοι στις κανονικοποιημένες εκθετικές συναρτήσεις μετασχηματισμού των παραμέτρων μετάβασης \mathbf{a}_{ij} μπορούμε να υπολογίσουμε τις αντίστοιχες μερικές παραγώγους ως εξής

$$\begin{aligned} \frac{\partial \mathbf{a}_{ij}}{\partial w_{ij}} &= \left(\frac{\partial e^{w_{ij}}}{\partial w_{ij}} \sum_{l=1}^N e^{w_{il}} - e^{w_{ij}} \frac{\partial \sum_{l=1}^N e^{w_{il}}}{\partial w_{ij}} \right) \left(\sum_{l=1}^N e^{w_{il}} \right)^{-2} = \\ &= \left(e^{w_{ij}} \sum_{l=1}^N e^{w_{il}} - e^{w_{ij}} e^{w_{ij}} \right) \left(\sum_{l=1}^N e^{w_{il}} \right)^{-2} = \mathbf{a}_{ij} (1 - \mathbf{a}_{ij}) \end{aligned} \quad (3.57)$$

$$\frac{\partial \mathbf{a}_{ij}}{\partial w_{ik}} = -e^{w_{ij}} \frac{\partial \sum_{l=1}^N e^{w_{il}}}{\partial w_{ik}} \left(\sum_{l=1}^N e^{w_{il}} \right)^{-2} = -e^{w_{ij}} e^{w_{ik}} \left(\sum_{l=1}^N e^{w_{il}} \right)^{-2} = -\mathbf{a}_{ij} \mathbf{a}_{ik}, \quad k \neq j. \quad (3.58)$$

Χρησιμοποιώντας αυτές τις εκφράσεις και αξιοποιώντας το πρώτο λήμμα (3.42), το δεξί μέλος της σχέσης (3.56) γίνεται

$$\begin{aligned} \frac{\partial P(O | \lambda)}{\partial w_{ij}} &= \mathbf{a}_{ij} (1 - \mathbf{a}_{ij}) \sum_{l=1}^{T-1} \alpha_l(i) b_j(o_{l+1}) \beta_{l+1}(j) - \sum_{z=1, z \neq j}^N \mathbf{a}_{iz} \mathbf{a}_{ij} \sum_{l=1}^{T-1} \alpha_l(i) b_z(o_{l+1}) \beta_{l+1}(z) = \\ &= \mathbf{a}_{ij} \sum_{l=1}^{T-1} \alpha_l(i) b_j(o_{l+1}) \beta_{l+1}(j) - \mathbf{a}_{ij} \left[\mathbf{a}_{ij} \sum_{l=1}^{T-1} \alpha_l(i) b_j(o_{l+1}) \beta_{l+1}(j) \right] - \\ &= \mathbf{a}_{ij} \sum_{z=1, z \neq j}^N \mathbf{a}_{iz} \sum_{l=1}^{T-1} \alpha_l(i) b_z(o_{l+1}) \beta_{l+1}(z). \end{aligned} \quad (3.59)$$

Συνενώνουμε τους δύο τελευταίους όρους στο δεξιότερο μέλος της (3.59) και έχουμε

$$\frac{\partial P(O | \lambda)}{\partial w_{ij}} = \mathbf{a}_{ij} \left[\sum_{l=1}^{T-1} \alpha_l(i) b_j(o_{l+1}) \beta_{l+1}(j) - \sum_{z=1}^N \mathbf{a}_{iz} \sum_{l=1}^{T-1} \alpha_l(i) b_z(o_{l+1}) \beta_{l+1}(z) \right]. \quad (3.60)$$

Εναλλάσσοντας τη σειρά των αθροισμάτων στο δεύτερο όρο του δεξιού μέλους της (3.60) προκύπτει ότι

$$\sum_{z=1}^N \mathbf{a}_{iz} \sum_{l=1}^{T-1} \alpha_l(i) b_z(o_{l+1}) \beta_{l+1}(z) = \sum_{l=1}^{T-1} \alpha_l(i) \sum_{z=1}^N \mathbf{a}_{iz} b_z(o_{l+1}) \beta_{l+1}(z). \quad (3.61)$$

Χρησιμοποιώντας σε αυτό το στάδιο την αναδρομική σχέση (3.38) προκύπτει ότι

$$\sum_{z=1}^N \mathbf{a}_{iz} \sum_{l=1}^{T-1} \alpha_l(i) b_z(o_{l+1}) \beta_{l+1}(z) = \sum_{l=1}^{T-1} \alpha_l(i) \beta_l(i). \quad (3.62)$$

Εισάγοντας αυτή τη σχέση στην ισότητα (3.60) συνεπάγεται ότι

$$\frac{\partial P(O | \lambda)}{\partial w_{ij}} = \mathbf{a}_{ij} \left[\sum_{l=1}^{T-1} \alpha_l(i) b_j(o_{l+1}) \beta_{l+1}(j) - \sum_{l=1}^{T-1} \alpha_l(i) \beta_l(i) \right]. \quad (3.63)$$

Συνεπώς η σχέση (3.55) αποδείχθηκε.

Αρχικά, ορίζουμε τη συνάρτηση δείκτη ως

$$I\{o_l = t | \lambda\} = \begin{cases} 1, & o_l = v_t \\ 0, & \text{otherwise.} \end{cases} \quad (3.64)$$

Λήμμα 2

$$\frac{\partial P(O | \lambda)}{\partial b_x(y)} = \frac{1}{b_x(y)} \sum_{l=1}^T I\{o_l = y | \lambda\} \alpha_l(x) \beta_l(x). \quad (3.65)$$

Απόδειξη: αρχικά βρίσκουμε τη μερική παράγωγο της (3.41) ως προς $b_x(y)$, το αποτέλεσμα είναι

$$\frac{\partial P(O | \lambda)}{\partial b_x(y)} = I\{o_T = y | \lambda\} \sum_{i=1}^N \alpha_{T-1}(i) \mathbf{a}_{ix} \beta_T(x) + \sum_{j=1}^N \sum_{i=1}^N \frac{\partial \alpha_{T-1}(i)}{\partial b_x(y)} \mathbf{a}_{ij} b_j(o_T) \beta_T(j). \quad (3.66)$$

Με την εφαρμογή του αναδρομικού τύπου (3.36) στον πρώτο όρο του δεξιού μέλους της σχέσης προκύπτει

$$\frac{\partial P(O | \lambda)}{\partial b_x(y)} = \frac{1}{b_x(o_T)} I\{o_T = y | \lambda\} \alpha_T(x) \beta_T(x) + \sum_{j=1}^N \sum_{i=1}^N \frac{\partial \alpha_{T-1}(i)}{\partial b_x(y)} \mathbf{a}_{ij} b_j(o_T) \beta_T(j). \quad (3.67)$$

Εντεύθεν, χρειάζεται να υπολογίσουμε μόνο το δεύτερο όρο στο δεξί μέλος της σχέσης. Αυτό επιτυγχάνεται χρησιμοποιώντας ακόμη μία φορά την αναδρομική σχέση (3.36)

$$\frac{\partial \alpha_{T-1}(i)}{\partial b_x(y)} = \frac{\partial}{\partial b_x(y)} \sum_{l=1}^N \alpha_{T-2}(l) \mathbf{a}_{li} b_l(o_{T-1}). \quad (3.68)$$

Εκτελώντας την παραγωγή λαμβάνουμε

$$\frac{\partial \alpha_{T-1}(x)}{\partial b_x(y)} = I\{o_{T-1} = y | \lambda\} \sum_{l=1}^N \alpha_{T-2}(l) \mathbf{a}_{lx} + \left[\sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial b_x(y)} \mathbf{a}_{lx} \right] b_x(o_{T-1}) \quad (3.69)$$

$$\frac{\partial \alpha_{T-1}(i)}{\partial b_x(y)} = \left[\sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial b_x(y)} \mathbf{a}_{li} \right] b_l(o_{T-1}), \quad i \neq x. \quad (3.70)$$

Εισάγουμε αυτές τις μερικές παραγώγους στο δεύτερο όρο της (3.67) και το αποτέλεσμα είναι

$$\sum_{j=1}^N \sum_{i=1}^N \frac{\partial \alpha_{T-1}(i)}{\partial b_x(y)} \mathbf{a}_{ij} b_j(o_T) \beta_T(j) = \sum_{j=1}^N I\{o_{T-1} = y | \lambda\} \sum_{l=1}^N \alpha_{T-2}(l) \mathbf{a}_{lx} \mathbf{a}_{xj} b_j(o_T) \beta_T(j) +$$

$$\begin{aligned} & \sum_{j=1}^N \left[\sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial b_x(y)} \mathbf{a}_{lx} \right] b_x(o_{T-1}) \mathbf{a}_{xj} b_j(o_T) \beta_T(j) + \\ & \sum_{j=1}^N \sum_{i=1, i \neq x}^N \left[\sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial b_x(y)} \mathbf{a}_{li} \right] b_i(o_{T-1}) \mathbf{a}_{ij} b_j(o_T) \beta_T(j). \quad (3.71) \end{aligned}$$

Συνενώνουμε τους δύο τελευταίους όρους στο δεξί μέλος της (3.71) και έχουμε

$$\begin{aligned} \sum_{j=1}^N \sum_{i=1}^N \frac{\partial \alpha_{T-1}(i)}{\partial b_x(y)} \mathbf{a}_{ij} b_j(o_T) \beta_T(j) &= \sum_{j=1}^N I\{o_{T-1} = y | \lambda\} \sum_{l=1}^N \alpha_{T-2}(l) \mathbf{a}_{lx} \mathbf{a}_{xj} b_j(o_T) \beta_T(j) + \\ & \sum_{j=1}^N \sum_{i=1}^N \sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial b_x(y)} \mathbf{a}_{li} b_i(o_{T-1}) \mathbf{a}_{ij} b_j(o_T) \beta_T(j). \quad (3.72) \end{aligned}$$

Ακολούθως χρειάζεται να υπολογιστούν οι δύο όροι στο δεξί μέλος της (3.72). Για τον πρώτο όρο ισχύει

$$\begin{aligned} \sum_{j=1}^N I\{o_{T-1} = y | \lambda\} \sum_{l=1}^N \alpha_{T-2}(l) \mathbf{a}_{lx} \mathbf{a}_{xj} b_j(o_T) \beta_T(j) &= \\ \sum_{l=1}^N I\{o_{T-1} = y | \lambda\} \alpha_{T-2}(l) \mathbf{a}_{lx} \sum_{j=1}^N \mathbf{a}_{xj} b_j(o_T) \beta_T(j). \quad (3.73) \end{aligned}$$

Χρησιμοποιώντας σε αυτό το στάδιο την αναδρομική σχέση (3.38) προκύπτει ότι

$$\begin{aligned} \sum_{l=1}^N I\{o_{T-1} = y | \lambda\} \alpha_{T-2}(l) \mathbf{a}_{lx} \sum_{j=1}^N \mathbf{a}_{xj} b_j(o_T) \beta_T(j) &= \\ \sum_{l=1}^N I\{o_{T-1} = y | \lambda\} \alpha_{T-2}(l) \mathbf{a}_{lx} \beta_{T-1}(x). \quad (3.74) \end{aligned}$$

Ενώ εφαρμόζοντας τον αναδρομικό τύπο (3.36) έχουμε ότι

$$\sum_{l=1}^N I\{o_{T-1} = y | \lambda\} \alpha_{T-2}(l) \mathbf{a}_{lx} \beta_{T-1}(x) = \frac{1}{b_x(o_{T-1})} I\{o_{T-1} = y | \lambda\} \alpha_{T-1}(x) \beta_{T-1}(x). \quad (3.75)$$

Για το δεύτερο όρο στο δεξί μέλος της (3.72) εναλλάσσοντας τη σειρά των αθροισμάτων λαμβάνουμε

$$\begin{aligned} \sum_{j=1}^N \sum_{i=1}^N \sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial b_x(y)} \mathbf{a}_{li} b_i(o_{T-1}) \mathbf{a}_{ij} b_j(o_T) \beta_T(j) = \\ \sum_{i=1}^N \sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial b_x(y)} \mathbf{a}_{li} b_i(o_{T-1}) \sum_{j=1}^N \mathbf{a}_{ij} b_j(o_T) \beta_T(j). \end{aligned} \quad (3.76)$$

Με τη χρήση της αναδρομικής σχέσης (3.38) έχουμε ότι

$$\begin{aligned} \sum_{i=1}^N \sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial b_x(y)} \mathbf{a}_{li} b_i(o_{T-1}) \sum_{j=1}^N \mathbf{a}_{ij} b_j(o_T) \beta_T(j) = \\ \sum_{i=1}^N \sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial b_x(y)} \mathbf{a}_{li} b_i(o_{T-1}) \beta_{T-1}(i). \end{aligned} \quad (3.77)$$

Με τη βοήθεια των σχέσεων (3.72)-(3.77) μπορούμε να ξαναγράψουμε το δεξί μέλος της (3.67) ως εξής

$$\begin{aligned} \frac{\partial P(O | \lambda)}{\partial b_x(y)} = \frac{1}{b_x(o_T)} I\{o_T = y | \lambda\} \alpha_T(x) \beta_T(x) + \frac{1}{b_x(o_{T-1})} I\{o_{T-1} = y | \lambda\} \alpha_{T-1}(x) \beta_{T-1}(x) + \\ \sum_{i=1}^N \sum_{l=1}^N \frac{\partial \alpha_{T-2}(l)}{\partial b_x(y)} \mathbf{a}_{li} b_i(o_{T-1}) \beta_{T-1}(i). \end{aligned} \quad (3.78)$$

Ωστόσο, ένας μπορεί να παρατηρήσει ότι το παραπάνω διπλό άθροισμα έχει την ίδια ακριβώς μορφή με αυτή του διπλού αθροίσματος στην (3.67). Εφεξής, μπορούμε να συνεχίσουμε παραγωγίζοντας τα $\frac{\partial \alpha_{T-2}(l)}{\partial b_x(y)}$ και οργανώνοντας τις προκύπτουσες σχέσεις με τον ίδιο τρόπο που παρουσιάστηκε προηγουμένα. Βασιζόμενοι στη σχέση (3.35) μπορούμε να δείξουμε ότι

$$\sum_{i=1}^N \sum_{l=1}^N \frac{\partial \alpha_1(l)}{\partial b_x(y)} \mathbf{a}_{li} b_i(o_2) \beta_2(i) = \sum_{i=1}^N \sum_{l=1}^N \pi_l \frac{\partial b_l(o_1)}{\partial b_x(y)} \mathbf{a}_{li} b_i(o_2) \beta_2(i) =$$

$$\sum_{i=1}^N \pi_x \frac{\partial b_x(o_1)}{\partial b_x(y)} \mathbf{a}_{xi} b_i(o_2) \beta_2(i) = I\{o_1 = y | \lambda\} \pi_x \sum_{i=1}^N \mathbf{a}_{xi} b_i(o_2) \beta_2(i). \quad (3.79)$$

Χρησιμοποιώντας σε αυτό το στάδιο την αναδρομική σχέση (3.38) προκύπτει ότι

$$I\{o_1 = y | \lambda\} \pi_x \sum_{i=1}^N \mathbf{a}_{xi} b_i(o_2) \beta_2(i) = I\{o_1 = y | \lambda\} \pi_x \beta_1(x). \quad (3.80)$$

Ενώ εφαρμόζοντας τον αναδρομικό τύπο (3.36) έχουμε ότι

$$I\{o_1 = y | \lambda\} \pi_x \beta_1(x) = \frac{1}{b_x(o_1)} I\{o_1 = y | \lambda\} \alpha_1(x) \beta_1(x). \quad (3.81)$$

Κατά συνέπεια, εξετάζοντας τις (3.78)-(3.81) διαπιστώνουμε ότι η παραγωγή τερματίζει στο σημείο που εμφανίζεται ο όρος $\frac{\partial \alpha_1(l)}{\partial b_x(y)}$, μιας και ο $\frac{\partial \alpha_1(l)}{\partial b_x(y)}$ είναι ο τελευταίος όρος που ενδέχεται να περιέχει έναν παράγοντα $b_x(y)$. Συνολικά επομένως θα έχουμε

$$\frac{\partial P(O | \lambda)}{\partial b_x(y)} = \sum_{l=1}^T \frac{1}{b_x(o_l)} I\{o_l = y | \lambda\} \alpha_l(x) \beta_l(x). \quad (3.82)$$

Εξαιτίας του γεγονότος ότι περιλαμβάνεται η συνάρτηση $I\{o_l = y | \lambda\}$, οι μόνοι μη μηδενικοί όροι του παραπάνω αθροίσματος είναι εκείνοι για τους οποίους ισχύει ο κανόνας $o_l = v_y$. Άρα μία ισοδύναμη έκφραση της σχέσης (3.82) είναι η

$$\frac{\partial P(O | \lambda)}{\partial b_x(y)} = \sum_{l=1}^T \frac{1}{b_x(y)} I\{o_l = y | \lambda\} \alpha_l(x) \beta_l(x). \quad (3.83)$$

Συνεπώς η σχέση (3.65) αποδείχθηκε.

Θεώρημα 2

$$\frac{\partial P(O | \lambda)}{\partial r_{jt}} = \sum_{l=1}^T [I\{o_l = t | \lambda\} \alpha_l(j) \beta_l(j) - b_j(t) \alpha_l(j) \beta_l(j)]. \quad (3.84)$$

Απόδειξη: εφαρμόζοντας τον κανόνα της αλυσίδας στο αριστερό μέλος της (3.84) έχουμε

$$\frac{\partial P(O | \lambda)}{\partial r_{jt}} = \sum_{z=1}^M \frac{\partial P(O | \lambda)}{\partial b_j(z)} \frac{\partial b_j(z)}{\partial r_{jt}} = \frac{\partial P(O | \lambda)}{\partial b_j(t)} \frac{\partial b_j(t)}{\partial r_{jt}} + \sum_{z=1, z \neq t}^M \frac{\partial P(O | \lambda)}{\partial b_j(z)} \frac{\partial b_j(z)}{\partial r_{jt}}. \quad (3.85)$$

Βασιζόμενοι στις κανονικοποιημένες εκθετικές συναρτήσεις μετασχηματισμού των παραμέτρων εμφάνισης $b_j(t)$ μπορούμε να υπολογίσουμε τις αντίστοιχες μερικές παραγώγους εφαρμόζοντας ακριβώς την ίδια διαδικασία με αυτήν που ακολουθήθηκε προηγούμενα. Οι τελικές εκφράσεις είναι

$$\frac{\partial b_j(t)}{\partial r_{jt}} = b_j(t)(1 - b_j(t)) \quad (3.86)$$

$$\frac{\partial b_j(t)}{\partial r_{jk}} = -b_j(t)b_j(k), \quad k \neq t. \quad (3.87)$$

Χρησιμοποιώντας αυτές τις εκφράσεις και αξιοποιώντας το δεύτερο λήμμα (3.65), το δεξί μέλος της σχέσης (3.85) γίνεται

$$\begin{aligned} \frac{\partial P(O | \lambda)}{\partial r_{jt}} &= b_j(t)(1 - b_j(t)) \frac{1}{b_j(t)} \sum_{l=1}^T I\{o_l = t | \lambda\} \alpha_l(j) \beta_l(j) - \\ &\sum_{z=1, z \neq t}^M b_j(z)b_j(t) \frac{1}{b_j(z)} \sum_{l=1}^T I\{o_l = z | \lambda\} \alpha_l(j) \beta_l(j) = \\ &\sum_{l=1}^T I\{o_l = t | \lambda\} \alpha_l(j) \beta_l(j) - b_j(t) \sum_{l=1}^T I\{o_l = t | \lambda\} \alpha_l(j) \beta_l(j) - \\ &\sum_{z=1, z \neq t}^M b_j(t) \sum_{l=1}^T I\{o_l = z | \lambda\} \alpha_l(j) \beta_l(j). \end{aligned} \quad (3.88)$$

Συνενώνουμε τους δύο τελευταίους όρους στο δεξιότερο μέλος της (3.88) και έχουμε

$$\frac{\partial P(O | \lambda)}{\partial r_{jt}} = \sum_{l=1}^T I\{o_l = t | \lambda\} \alpha_l(j) \beta_l(j) - \sum_{z=1}^M b_j(t) \sum_{l=1}^T I\{o_l = z | \lambda\} \alpha_l(j) \beta_l(j). \quad (3.89)$$

Εναλλάσσοντας τη σειρά των αθροισμάτων στο δεύτερο όρο του δεξιού μέλους της (3.89) προκύπτει ότι

$$\sum_{z=1}^M b_j(t) \sum_{l=1}^T I\{o_l = z | \lambda\} \alpha_l(j) \beta_l(j) = b_j(t) \sum_{l=1}^T \alpha_l(j) \beta_l(j) \sum_{z=1}^M I\{o_l = z | \lambda\}. \quad (3.90)$$

Κάθε σύμβολο o_l μίας ακολουθίας παρατηρήσεων O , έχει μία δεδομένη τιμή από το σύνολο δειγμάτων V (συγκεκριμένα λαμβάνει μία εκ των M διαφορετικών v_m τιμών), κατά συνέπεια το αποτέλεσμα της άθροισης των συναρτήσεων $I\{o_l = z | \lambda\}$ είναι

$$\sum_{z=1}^M I\{o_l = z | \lambda\} = 1. \quad (3.91)$$

Χρησιμοποιώντας την ισότητα (3.91) στη σχέση (3.90) προκύπτει ότι

$$\sum_{z=1}^M b_j(t) \sum_{l=1}^T I\{o_l = z | \lambda\} \alpha_l(j) \beta_l(j) = b_j(t) \sum_{l=1}^T \alpha_l(j) \beta_l(j). \quad (3.92)$$

Εισάγοντας αυτή την ισότητα στη σχέση (3.89) συνεπάγεται ότι

$$\frac{\partial P(O | \lambda)}{\partial r_{jt}} = \sum_{l=1}^T I\{o_l = t | \lambda\} \alpha_l(j) \beta_l(j) - b_j(t) \sum_{l=1}^T \alpha_l(j) \beta_l(j). \quad (3.93)$$

Συνεπώς η σχέση (3.84) αποδείχθηκε.

Λήμμα 3

$$\frac{\partial P(O | \lambda)}{\partial \pi_r} = b_r(o_1) \beta_1(r). \quad (3.94)$$

Απόδειξη: από τον τύπο υπολογισμού της πιθανοφάνειας (3.39) για $t = 1$ έχουμε ότι

$$P(O | \lambda) = \sum_{j=1}^N \alpha_1(j) \beta_1(j). \quad (3.95)$$

Υπολογίζουμε τη μερική παράγωγο της (3.95) ως προς π_r , το αποτέλεσμα είναι

$$\frac{\partial P(O | \lambda)}{\partial \pi_r} = \sum_{j=1}^N \frac{\partial \alpha_1(j)}{\partial \pi_r} \beta_1(j). \quad (3.96)$$

Εντεύθεν, χρειάζεται να υπολογίσουμε τον όρο στο δεξί μέλος της σχέσης. Αυτό επιτυγχάνεται χρησιμοποιώντας την αναδρομική σχέση (3.35)

$$\frac{\partial \alpha_1(j)}{\partial \pi_r} = b_j(o_1) \frac{\partial \pi_j}{\partial \pi_r}. \quad (3.97)$$

Εισάγοντας αυτή τη σχέση στη (3.96) και λαμβάνοντας υπόψη ότι $\frac{\partial \pi_j}{\partial \pi_r} = 0$ όταν

$r \neq j$, έχουμε

$$\frac{\partial P(O | \lambda)}{\partial \pi_r} = \sum_{j=1}^N b_j(o_1) \frac{\partial \pi_j}{\partial \pi_r} \beta_1(j) = b_r(o_1) \beta_1(r). \quad (3.98)$$

Συνεπώς η σχέση (3.94) αποδείχθηκε.

Θεώρημα 3

$$\frac{\partial P(O | \lambda)}{\partial u_j} = \pi_j [b_j(o_1) \beta_1(j) - P(O | \lambda)]. \quad (3.99)$$

Απόδειξη: εφαρμόζοντας τον κανόνα της αλυσίδας στο αριστερό μέλος της (3.99) έχουμε

$$\frac{\partial P(O | \lambda)}{\partial u_j} = \sum_{z=1}^N \frac{\partial P(O | \lambda)}{\partial \pi_z} \frac{\partial \pi_z}{\partial u_j} = \frac{\partial P(O | \lambda)}{\partial \pi_j} \frac{\partial \pi_j}{\partial u_j} + \sum_{z=1, z \neq j}^N \frac{\partial P(O | \lambda)}{\partial \pi_z} \frac{\partial \pi_z}{\partial u_j}. \quad (3.100)$$

Βασιζόμενοι στις κανονικοποιημένες εκθετικές συναρτήσεις μετασχηματισμού των παραμέτρων π_j μπορούμε να υπολογίσουμε τις αντίστοιχες μερικές παραγώγους εφαρμόζοντας ακριβώς την ίδια διαδικασία με αυτήν που ακολουθήθηκε προηγουμένα. Οι τελικές εκφράσεις είναι

$$\frac{\partial \pi_j}{\partial u_j} = \pi_j(1 - \pi_j) \quad (3.101)$$

$$\frac{\partial \pi_j}{\partial u_k} = -\pi_j \pi_k, \quad k \neq j. \quad (3.102)$$

Χρησιμοποιώντας αυτές τις εκφράσεις και αξιοποιώντας το τρίτο λήμμα (3.94), το δεξί μέλος της σχέσης (3.100) γίνεται

$$\begin{aligned} \frac{\partial P(O | \lambda)}{\partial u_j} &= \pi_j(1 - \pi_j)b_j(o_1)\beta_1(j) - \sum_{z=1, z \neq j}^N \pi_z \pi_j b_z(o_1)\beta_1(z) = \\ &= \pi_j b_j(o_1)\beta_1(j) - \pi_j \pi_j b_j(o_1)\beta_1(j) - \pi_j \sum_{z=1, z \neq j}^N \pi_z b_z(o_1)\beta_1(z). \end{aligned} \quad (3.103)$$

Συνενώνουμε τους δύο τελευταίους όρους στο δεξιότερο μέλος της (3.103) και έχουμε

$$\frac{\partial P(O | \lambda)}{\partial u_j} = \pi_j b_j(o_1)\beta_1(j) - \pi_j \sum_{z=1}^N \pi_z b_z(o_1)\beta_1(z). \quad (3.104)$$

Εφαρμόζοντας τον αναδρομικό τύπο (3.35) έχουμε ότι

$$\sum_{z=1}^N \pi_z b_z(o_1) \beta_1(z) = \sum_{z=1}^N \alpha_1(z) \beta_1(z). \quad (3.105)$$

Σύμφωνα με τη σχέση (3.39) ισχύει ότι

$$\sum_{z=1}^N \alpha_1(z) \beta_1(z) = P(O | \lambda). \quad (3.106)$$

Χρησιμοποιώντας τις (3.105) και (3.106) στο δεξί μέλος της σχέσης (3.104) προκύπτει ότι

$$\frac{\partial P(O | \lambda)}{\partial u_j} = \pi_j b_j(o_1) \beta_1(j) - \pi_j P(O | \lambda). \quad (3.107)$$

Συνεπώς η σχέση (3.99) αποδείχθηκε.

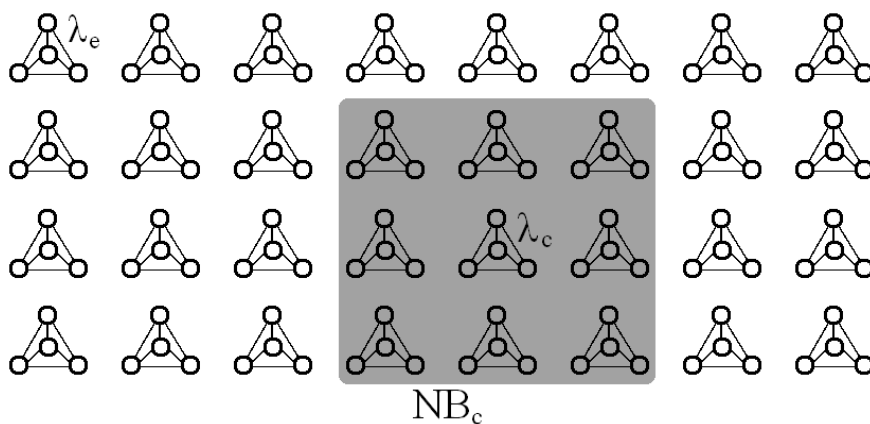
4. ΤΟ ΠΡΟΤΥΠΟ SOHMMM

4.1. Γενική Θεώρηση

Μελέτες που έχουν διεξαχθεί κατά τη διάρκεια πολλών ετών από ένα μεγάλο αριθμό ερευνητών παρέχουν βάσιμες ενδείξεις περί του γεγονότος ότι τα καλύτερα αποτελέσματα αυτο-οργάνωσης επιτυγχάνονται όταν οι ακόλουθες δύο μερικές (εν δυνάμει συμπληρωματικές) διεργασίες υλοποιούνται στις απλούστερες και καθαρότερες μορφές τους [6]:

- 1) εύρεση εκείνου του νευρώνα που εμφανίζει τη βέλτιστη ταύτιση με το πρότυπο εισόδου, ή ισοδύναμα αποκωδικοποίηση εκείνου του νευρώνα που περιγράφει βέλτιστα το διαθέσιμο δείγμα του χώρου εισόδου (αυτός ο νευρώνας συχνά αποκαλείται “νικητής”),
- 2) προσαρμοστική αναβάθμιση και βελτίωση της ταύτισης στην γειτονιά νευρώνων που εντοπίζονται πέριξ του νευρώνα “νικητή”, πράγμα που θα έχει ως αποτέλεσμα η απόκριση των εν λόγω νευρώνων στην επόμενη εφαρμογή ενός παρόμοιου προτύπου να είναι εντονότερη και υψηλότερη.

Φορμαλιστικά το SOHMMM δύναται να περιγραφεί ως μία μη γραμμική, διατεταγμένη, ομαλή χαρτογράφηση ακολουθιών παρατηρήσεων επάνω στα στοιχεία μίας κανονικής, χαμηλής διάστασης συστοιχίας HMMs. Η εν λόγω απεικόνιση πραγματοποιείται με τέτοιο τρόπο ώστε να ομοιάζει κατά το μέγιστο δυνατόν με τις δύο προαναφερθείσες διαδικασίες.



Εικόνα 4.1. Χαρακτηριστικό παράδειγμα πλέγματος SOHMMM δύο διαστάσεων.

Εν αρχή, κάθε στοιχείο e του πλέγματος SOHMMM συσχετίζεται με ένα HMM λ_e . Θεωρώντας τις υπό συνθήκη πιθανότητες μίας ακολουθίας παρατηρήσεων O δεδομένου καθενός HMM λ_e (πιθανοφάνειες), η εικόνα της ακολουθίας εισόδου O επάνω στο πλέγμα του SOHMMM ορίζεται να είναι το στοιχείο της συστοιχίας με τη βέλτιστη τιμή μίας κατάλληλης συνάρτησης διάκρισης, κόστους ή ενέργειας. Η αναφερόμενη συνάρτηση ενδείκνυται να συσχετίζεται ισχυρά με τις πιθανοφάνειες $P(O|\lambda_e)$ των ακολουθιών από παρατηρήσιμα σύμβολα και δυνητικά μπορεί να εξαρτάται από τα τοπολογικά χαρακτηριστικά του χάρτη SOHMMM.

Σε δεύτερο χρόνο, στόχος είναι να προσδιοριστεί και ακολούθως να προσαρμοστεί κάθε ξεχωριστό λ_e κατά τέτοιο τρόπο έτσι ώστε η επιτευχθείσα απεικόνιση να είναι περιγραφική, διατεταγμένη και αντιπροσωπευτική της κατανομής των ακολουθιών παρατηρήσεων. Θεωρούμε την εικόνα 4.1 όπου εμφανίζεται μία δισδιάστατη διατεταγμένη συστοιχία κόμβων κάθε ένας εκ των οποίων αντιστοιχίζεται με ένα HMM λ_e . Επιπλέον, θεωρούμε τη γειτονιά NB_c πέριξ του μοντέλου λ_c που ταυτίζεται με την O (δηλαδή του HMM που επιστρέφει τη βέλτιστη τιμή για τη χρησιμοποιούμενη συνάρτηση διάκρισης). Στην προκειμένη περίπτωση η NB_c περιλαμβάνει όλους τους νευρώνες εντός μίας συγκεκριμένης απόστασης από το νευρώνα c . Το επόμενο στάδιο είναι η ρύθμιση των παραμέτρων όλων των HMMs που εμπίπτουν στη γειτονιά NB_c , έτσι ώστε να βελτιστοποιηθούν οι αντίστοιχες συναρτήσεις κόστους ή ενέργειας, δηλαδή με άλλα λόγια να αντληθεί κάποια (επιπρόσθετη) γνώση από τη δεδομένη ακολουθία παρατηρήσεων. Στην πραγματικότητα γίνεται προσπάθεια να βελτιστοποιηθούν οι παράμετροι καθενός $\lambda_e \in NB_c$ με σκοπό την προσαρμογή του SOHMMM στην ακολουθία παρατηρήσιμων συμβόλων και μακροπρόθεσμα στο σύνολο των προτύπων εισόδου. Αυτοί οι στόχοι επιτυγχάνονται με τους στοχαστικούς αλγορίθμους του SOHMMM που θα περιγραφούν αργότερα.

Στο SOHMMM οι HMM νευρώνες τοποθετούνται στους κόμβους ενός πλέγματος το οποίο είναι συνήθως δισδιάστατο. Χάρτες μεγαλύτερης διάστασης είναι δυνατό να κατασκευαστούν, αλλά δεν χρησιμοποιούνται ιδιαίτερα. Οι νευρώνες συντονίζονται επιλεκτικά σε διάφορες ακολουθίες παρατηρήσεων ή κλάσεις ακολουθιών παρατηρήσιμων συμβόλων, κατά την πορεία μίας διαδικασίας ανταγωνιστικής μάθησης. Οι νευρώνες που συντονίζονται κατά αυτόν τον τρόπο παράλληλα

διατάσσονται ο ένας σε σχέση με τον άλλο με αποτέλεσμα να δημιουργείται ένα λογικό σύστημα συντεταγμένων για διαφορετικά πρότυπα εισόδου επάνω στο πλέγμα. Συνεπώς, το SOHMMM χαρακτηρίζεται από το σχηματισμό ενός τοπογραφικού χάρτη κατασκευασμένου βάσει των ακολουθιών παρατηρήσεων στον οποίο οι χωρικές θέσεις (ή αλλιώς οι συντεταγμένες) των νευρώνων στο πλέγμα είναι ενδεικτικές των εσωτερικών στατιστικών χαρακτηριστικών που ενυπάρχουν στις διαθέσιμες ακολουθίες παρατηρήσεων. Όπως αναφέρθηκε προηγούμενα το SOHMMM ορίζεται από μία συστοιχία HMMs που αντιπροσωπεύουν διαφορετικά συντονισμένους νευρώνες οι οποίοι επεξεργάζονται τις φέρουσες πληροφορία ακολουθίες. Κατά συνέπεια, οι νευρώνες μετασχηματίζουν το πρότυπα εισόδου σε μία κωδικοποιημένη κατά θέση κατανομή πιθανοτήτων.

Στην πραγματικότητα το SOHMMM όπως άλλωστε και το μοντέλο Kohonen (στο οποίο έχουμε βασιστεί σε μεγάλο βαθμό) ανήκει στην κατηγορία των αλγορίθμων της (διανυσματικής) κβάντισης. Μεταξύ άλλων, παράγει μία τοπολογική αντιστοίχιση η οποία τοποθετεί με βέλτιστο τρόπο ένα σταθερό αριθμό HMMs σε χώρους εισόδου υψηλότερης διάστασης, διευκολύνοντας έτσι τη συμπίεση των δεδομένων. Δεδομένου ότι το SOHMMM με αυτόν τον τρόπο συμπίεζει πληροφορία διατηρώντας παράλληλα τις σημαντικότερες τοπολογικές και στατιστικές συσχετίσεις των δεδομένων εισόδου, μπορεί επίσης να θεωρηθεί ότι παράγει κάποιο είδος αφαίρεσης. Επίσης, το SOHMMM όπως άλλωστε και το πρόγονο SOM μοντέλο δίνει τη δυνατότητα ομαδοποίησης και οπτικοποίησης των συστάδων προτύπων εισόδου που έχουν ανακαλυφθεί. Στη βασική του μορφή παράγει ένα γράφο της κατανομής πιθανότητας των ακολουθιών παρατηρήσεων μετατρέποντας τις μη γραμμικές (στατιστικές) σχέσεις μεταξύ των δεδομένων ακολουθιών σε απλές γεωμετρικές σχέσεις των απεικονίσεων τους σε ένα χώρο μικρής διάστασης. Το σύνολο αυτών των δυνατοτήτων (ομαδοποίηση, οπτικοποίηση συστάδων, μείωση διαστατικότητας, αφαίρεση) μπορούν να βρουν εφαρμογές σε μία σειρά σύνθετων εργασιών.

Συγκεφαλαιώνοντας, υπάρχουν τρεις σημαντικές συμπληρωματικές διαδικασίες που εμπλέκονται στο σχηματισμό του SOHMMM.

- 1) Ανταγωνισμός: για κάθε ακολουθία παρατηρήσεων οι HMM νευρώνες του δικτύου υπολογίζουν τις αντίστοιχες τιμές μίας συνάρτησης διάκρισης ή ενέργειας. Αυτή η συνάρτηση παρέχει τη βάση για τον ανταγωνισμό των νευρώνων. Ο νευρώνας που αντιστοιχεί στη βέλτιστη τιμή αυτής της συνάρτησης θεωρείται νικητής της διαδικασίας του ανταγωνισμού.

- 2) Συνεργασία: ο νικητής νευρώνας καθορίζει τη χωρική θέση και τα τοπολογικά χαρακτηριστικά μίας γειτονιάς διεγερμένων HMM νευρώνων, παρέχοντας έτσι τη βάση για τη συνεργασία μεταξύ τέτοιων γειτνιαζόντων νευρώνων.
- 3) Προσαρμογή: αυτός ο μηχανισμός επιτρέπει στους διεγερμένους HMM νευρώνες να βελτιώνουν τις τιμές της συνάρτησης διάκρισης/ενέργειας, σε σχέση με τη δεδομένη ακολουθία παρατηρήσεων, μέσω κατάλληλων προσαρμογών που εφαρμόζονται στις παραμέτρους τους. Προφανώς οι τροποποιήσεις και οι ρυθμίσεις των παραμέτρων πρέπει να είναι τέτοιες ώστε οι αποκρίσεις των γειτνιαζόντων HMM νευρώνων (και κατά πρώτιστο λόγο του νευρώνα νικητή) στην χρονικά επόμενη εφαρμογή ενός παρόμοιου προτύπου εισόδου να είναι βελτιωμένες.

Είναι σαφές ότι εάν οι δύο τελευταίες διαδικασίες (δηλαδή αυτή της συνεργασίας και αυτή της προσαρμογής ενοποιηθούν εννοιολογικά) τότε υπάρχει πλήρης συμφωνία με το πρωτόκολλο διαδικασιών της κλάσης των αλγορίθμων αυτο-οργάνωσης που παρουσιάστηκε στην αρχή.

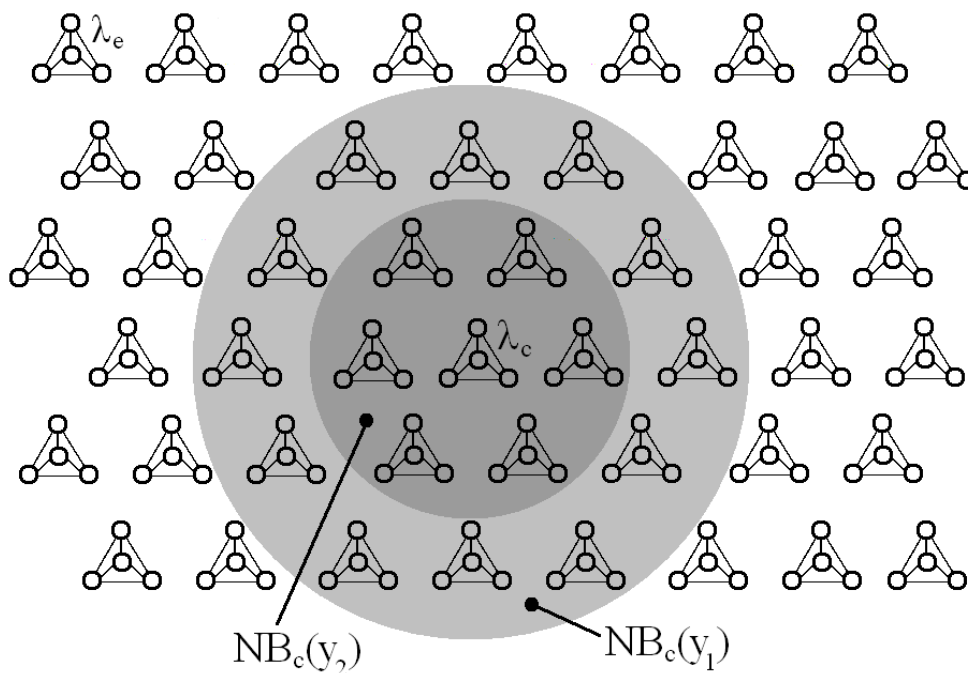
4.2. *Αλγόριθμοι Εκπαίδευσης για Αυτο-Οργανούμενα Δίκτυα*

Μία πλειάδα αλγορίθμων εκπαίδευσης έχει προταθεί για τα αυτο-οργανούμενα μοντέλα και ειδικότερα για τους αυτο-οργανούμενους χάρτες. Συγκεκριμένα για τα SOMs, η μεγάλη πλειοψηφία των μεθοδολογιών εκπαίδευσης βασίζεται κατ' ουσίαν στον αλγόριθμο μάθησης του Kohonen (μία μορφή μάθησης Hebb). Παρόλο που έχει διαδραματίσει ένα σημαντικό ρόλο στο χώρο της αυτο-οργανούμενης μάθησης και παρά την αδιαμφισβήτητη επιτυχία του, πράγμα που μεταξύ άλλων οφείλεται στην απλότητα και την αποδοτικότητα του, ο κανόνας του Kohonen και οι παραλλαγές αυτού παραμένουν, μέχρι ενός σημείου, ευρετικές μέθοδοι που στερούνται μαθηματικής απόδειξης στη γενική περίπτωση. Αντιθέτως, υπάρχουν προσεγγίσεις που διαφοροποιώντας την έννοια του νευρώνα νικητή και ορίζοντας μία κατάλληλη συνάρτηση ενέργειας παράγουν αλγορίθμους εκπαίδευσης οι οποίοι αντιστοιχούν σε κάποιας μορφής παραγωγή της αυτής της συνάρτησης ενέργειας. Κατά συνέπεια για αυτή τη δεύτερη γκάμα μεθοδολογιών μάθησης δικαιολογημένα υποστηρίζεται ότι εδράζονται σε ένα πιο στέρεο μαθηματικό υπόβαθρο, τουλάχιστον εν συγκρίσει με τις παραλλαγές του κανόνα Kohonen.

Εν πάση περιπτώσει, για το SOHMMM προτείνουμε δύο κλάσεις αλγορίθμων εκπαίδευσης, μία που βασίζεται στη φιλοσοφία και την αντίληψη του αλγορίθμου μάθησης του Kohonen, και μία που ακολουθεί τη μαθηματικά πιο στιβαρή προσέγγιση και λογική της συνάρτησης ενέργειας και του τροποποιημένου νικητή νευρώνα.

4.3. Στοχαστικός Αλγόριθμος Εκπαίδευσης του SOHMMM

Θεωρούμε ότι $O = o_1 o_2 \dots o_T$ είναι μία ακολουθία παρατηρήσεων όπου κάθε παρατήρηση o_t λαμβάνει μία συγκεκριμένη τιμή από το σύνολο συμβόλων $F = \{f_1, f_2, \dots, f_G\}$, ενώ με T ορίζουμε το πλήθος των παρατηρήσεων στην ακολουθία. Επιπλέον, θεωρούμε ότι A είναι μία κλάση HMMs τέτοια ώστε τα αντίστοιχα σύνολα δειγμάτων τους $V = \{v_1, v_2, \dots, v_M\}$ να είναι υπεрсύνολα του συνόλου συμβόλων F , δηλαδή για κάθε HMM να ισχύει $V \supseteq F$. Για να οριστεί ο αλγόριθμος μάθησης δεν απαιτείται καμία επιπλέον θεώρηση. Μάλιστα στη γενική περίπτωση δύο διακριτά HMMs $\lambda_e, \lambda_c \in A$ δύναται να διαφέρουν στους χώρους καταστάσεων, στα σύνολα δειγμάτων ή/και στις στοχαστικές μήτρες (τόσο ως προς τις αντίστοιχες διαστάσεις όσο και ως προς τις τιμές των αντίστοιχων στοιχείων).



Εικόνα 4.2. Παράδειγμα τοπολογικής γειτονιάς για διαφορετικά στιγμιότυπα του διακριτού χρόνου ($y_1 < y_2$).

4.3.1. Διαδικασία Ανταγωνισμού

Χωρίς βλάβη της γενικότητας θεωρούμε ένα δισδιάστατο πλέγμα SOHMMM (όπως αυτό που φαίνεται στην εικόνα 4.2) που ορίζει μία χαρτογράφηση του χώρου προτύπων, ο οποίος συνίσταται από ακολουθίες παρατήρησης, επάνω σε μία νευρωνική συστοιχία δύο διαστάσεων, αποτελούμενης από HMMs. Με κάθε νευρώνα e συσχετίζουμε ένα HMM $\lambda_e \in \mathcal{A}$ το οποίο ονομάζεται HMM αναφοράς. Το συνολικό πλήθος των νευρώνων άρα και των HMMs αναφοράς είναι E .

Υποθέτουμε ότι ταυτόχρονα σε όλα τα HMMs ανατίθεται μία πανομοιότυπη ακολουθία παρατηρήσεων O . Ακολούθως, μπορούμε να φανταστούμε ότι με τη χρήση μηχανισμών παράλληλου υπολογισμού, η ακολουθία O συγκρίνεται με όλα τα λ_e και η τοποθεσία του νευρώνα που παρουσιάζει βέλτιστη ταύτιση (ή συντονισμό), εξετάζοντας την πιθανοφάνεια, ορίζεται ως η θέση (ή απεικόνιση) της απόκρισης. Ουσιαστικά σε αυτό το στάδιο η ακριβής τιμή της απόκρισης δεν είναι τόσο σημαντική, αυτό που πρώτιστα ενδιαφέρει είναι το ότι η δεδομένη ακολουθία O απεικονίζεται σε αυτή τη θέση, όπως συμβαίνει στην περίπτωση ενός συνόλου αποκωδικοποιητών. Κατά συνέπεια, μπορούμε να ορίσουμε το HMM βέλτιστης ενεργοποίησης, δηλαδή το HMM νικητή (που συμβολίζεται με το δείκτη c), ως το HMM που παράγει την μέγιστη τιμή πιθανοφάνειας. Η εύρεση αυτού γίνεται χρησιμοποιώντας τον τύπο

$$c = \arg \max_e \{\log P(O | \lambda_e)\}. \quad (4.1)$$

Συνεπώς, μπορούμε να υποστηρίξουμε ότι το SOHMMM υλοποιεί μία μη γραμμική προβολή του χώρου των ακολουθιών παρατηρήσεων επάνω σε ένα δισδιάστατο νευρωνικό επίπεδο. Ένας χώρος εισόδου πρότυπων ενεργοποίησης αντιστοιχίζεται σε ένα διακριτό χώρο εξόδου νευρώνων μέσω μίας διαδικασίας ανταγωνισμού μεταξύ των HMMs του SOHMMM. Προφανώς ανάλογα με την εφαρμογή, η απόκριση του δικτύου μπορεί να είναι είτε ο δείκτης του βέλτιστα συντονισμένου νευρώνα (άρα και η θέση του στο πλέγμα SOHMMM) είτε το σύνολο παραμέτρων του HMM νικητή (άρα και η στατιστική περιγραφή του συγκεκριμένου υποχώρου).

4.3.2. Διαδικασία Συνεργασίας

Ο νευρώνας νικητής εντοπίζεται στο κέντρο μίας τοπολογικής γειτονιάς συνεργαζόμενων HMMs. Ένας νευρώνας που ενεργοποιείται τείνει να διεγείρει τους νευρώνες που βρίσκονται στην άμεση γειτονιά του περισσότερο από τους νευρώνες που είναι πιο απομακρυσμένοι. Αυτές οι παρατηρήσεις οδηγούν στο να ορίσουμε μία τοπολογική γειτονιά γύρω από το HMM νικητή που να φθίνει ομαλά με την αύξηση της απόστασης από αυτόν. Συγκεκριμένα έστω ότι με h_{ce} συμβολίζουμε την τοπολογική γειτονιά που είναι κεντραρισμένη γύρω από το HMM νικητή λ_c και περικλείει ένα σύνολο διεγερμένων συνεργαζόμενων HMMs, ένα τυπικό δείγμα των οποίων συμβολίζεται ως λ_e . Είναι απαραίτητο η συνάρτηση της τοπολογικής γειτονιάς να ικανοποιεί ορισμένες προδιαγραφές:

- 1) να έχει μοναδικό μέγιστο και να αποκτά αυτή τη μέγιστη τιμή της στο νευρώνα νικητή λ_c ,
- 2) να είναι συμμετρική γύρω από το HMM νικητή λ_c ,
- 3) το εύρος της τοπολογικής γειτονιάς να μειώνεται μονότονα με την αύξηση της απόστασης (στο πλέγμα SOHMMM) μεταξύ λ_c και λ_e .

Ο ρόλος που διαδραματίζει η συνάρτηση τοπολογικής γειτονιάς είναι κεντρικός καθώς δρα ως ένας πυρήνας εξομάλυνσης. Η μορφή και το εύρος της h_{ce} καθορίζουν την ελαστικότητα του δικτύωματος που θα προσαρμοστεί στα δεδομένα εισόδου.

Μία ευρέως χρησιμοποιούμενη, συνεχής, ομαλή, αναλλοίωτη μετατόπισης συνάρτηση τοπολογικής γειτονιάς είναι η Gauss

$$h_{ce}(y) = \exp\left(-\|\delta_c - \delta_e\|^2 / 2\sigma^2(y)\right) \quad (4.2)$$

όπου $y \geq 0$ είναι ένας ακέραιος που αντιπροσωπεύει το διακριτό χρόνο, τα $\delta_c, \delta_e \in \mathfrak{R}^2$ είναι τα διανύσματα θέσης των HMMs λ_c και λ_e στο πλέγμα SOHMMM, η παράμετρος $\sigma(y)$ είναι το πλάτος της τοπολογικής γειτονιάς και καθορίζει το βαθμό στον οποίο οι διεγερμένοι νευρώνες στην κοντινή περιοχή του νικητή νευρώνα συμμετέχουν στη διαδικασία μάθησης. Μία τέτοιου τύπου συνάρτηση Gauss ικανοποιεί πλήρως τις τρεις παραπάνω απαιτήσεις. Ένα άλλο χαρακτηριστικό του SOHMMM (όπως εξάλλου και του πρόγονου αλγορίθμου SOM) είναι ότι το εύρος της

τοπολογικής γειτονιάς επιτρέπεται να συρρικνώνεται καθώς ο χρόνος περνάει. Αυτή η απαίτηση ικανοποιείται επιβάλλοντας το πλάτος $\sigma(y)$ να μειώνεται μονότονα με την πάροδο του χρόνου. Έτσι καθώς αυξάνεται ο διακριτός χρόνος y (δηλαδή ο αριθμός των επαναλήψεων) και το πλάτος $\sigma(y)$ μειώνεται μονότονα, το εύρος της $h_{ce}(y)$ συρρικνώνεται ανάλογα. Ωστόσο, πρέπει να επισημάνουμε ότι η συνάρτηση τοπολογικής γειτονιάς θα έχει και πάλι τιμή μονάδας για το HMM νικητή, ανεξαρτήτως χρόνου.

Στη βιβλιογραφία εμφανίζεται συχνά και μία άλλη απλούστερη συνάρτηση τοπολογικής γειτονιάς η οποία επίσης μπορεί να χρησιμοποιηθεί στην περίπτωση του SOHMMM. Βάσει αυτής της προσέγγισης η τοπολογική γειτονιά αναφέρεται σε ένα σύνολο HMMs που γειτνιάζουν με το νευρώνα νικητή λ_c (εικόνα 4.2). Το σύνολο αυτό συμβολίζεται με NB_c , και ισχύει ότι

$$h_{ce}(y) = \begin{cases} 1, & \lambda_c \in NB_c(y) \\ 0, & \lambda_c \notin NB_c(y). \end{cases} \quad (4.3)$$

Είναι ομοίως κοινή πρακτική το εύρος του $NB_c(y)$ να μειώνεται μονότονα με την πάροδο του χρόνου. Ειδικώς στις περιπτώσεις που το μέγεθος του δικτύου SOHMMM δεν είναι μεγάλο συνηθίζεται η χρήση του συγκεκριμένου τύπου τοπολογικής γειτονιάς.

4.3.3. Διαδικασία Προσαρμογής

Κατά τη διάρκεια της εκπαίδευσης ή αλλιώς της διαδικασίας σχηματισμού της μη γραμμικής προβολής του SOHMMM, εκείνα τα HMMs που εντοπίζονται στην τοπολογική γειτονιά του βέλτιστα συντονισμένου HMM θα ενεργοποιηθούν ούτως ώστε να μάθουν κάτι από τη δοθείσα ακολουθία παρατηρήσεων O . Ως αποτέλεσμα της ενημέρωσης των παραμέτρων τους που λαμβάνει χώρα τοπικά στη γειτονιά, μετά από επαναλαμβανόμενες παρουσιάσεις των δεδομένων εκπαίδευσης, τα HMMs αναφοράς τείνουν να ακολουθούν την κατανομή των ακολουθιών παρατηρήσεων. Συνεπώς, ο αλγόριθμος οδηγεί σε μία τοπολογική διάταξη υπό την έννοια ότι οι νευρώνες που είναι γειτονικοί στο πλέγμα SOHMMM θα τείνουν να περιγράψουν ακολουθίες παρατηρήσεων που συσχετίζονται στατιστικά. Παράλληλα, το σύνολο των

επιμέρους αναπαραστάσεων του χώρου των προτύπων εισόδου, με το πέρας της διαδικασίας προσαρμογής, καταλήγει να περιγράφει και να συνοψίζει τις μη γραμμικές στατιστικές σχέσεις μεταξύ των δεδομένων ακολουθιών με έναν ενιαίο και περιεκτικό τρόπο.

Για την πραγματοποίηση των παραπάνω απαιτείται μία συστηματική διαδικασία προσαρμογής των παραμέτρων των επιμέρους HMMs. Η ρύθμιση των λ_e έτσι ώστε να μεγιστοποιηθούν οι αντίστοιχες πιθανοφάνειες $\log P(O | \lambda_e)$ (διαδικασία που παραπέμπει στη μεθοδολογία της ML εκτίμησης), μπορεί να επιτευχθεί συγχωνεύοντας τον επαναληπτικό αλγόριθμο εκπαίδευσης του SOM [6] με τη μέθοδο gradient descent για εκθετικά επαναπαραμετροποιημένα HMMs.

Αρχικά, για να εξαχθεί ο υβριδικός αυτός αλγόριθμος μάθησης, εφαρμόζουμε τη μέθοδο gradient descent για τη μεγιστοποίηση της πιθανοφάνειας $\log P(O | \lambda_e)$, θεωρώντας ότι η ελεύθερη παράμετρος που θέλουμε να εκτιμήσουμε είναι η x . Το αποτέλεσμα αυτής της διαδικασίας είναι

$$x^{(next)} = x^{(now)} + \eta \left. \frac{\partial \log P(O | \lambda)}{\partial x} \right|_{x=x^{(now)}} = x^{(now)} + \eta P(O | \lambda)^{-1} \left. \frac{\partial P(O | \lambda)}{\partial x} \right|_{x=x^{(now)}}. \quad (4.4)$$

Παρατηρούμε ότι η συνεισφορά της ακολουθίας παρατηρήσεων O σταθμίζεται από την αντίστροφη τιμή της πιθανοφάνειας της. Αυτό μπορεί να φαντάζει παράδοξο σε πρώτη ανάγνωση, αλλά μπορεί να καταστεί εύκολα κατανοητό. Εάν το HMM παράγει μία μικρή τιμή πιθανοφάνειας για μία δεδομένη ακολουθία, τότε η απόδοση του για την συγκεκριμένη ακολουθία παρατηρήσεων δεν είναι ικανοποιητική και κατά συνέπεια το σήμα σφάλματος πρέπει να είναι υψηλό.

Στην συνέχεια, ο επιθυμητός αλγόριθμος για τη ρύθμιση των παραμέτρων καθενός HMM δύναται να κατασκευαστεί συνδυάζοντας τα τρία αποδειχθέντα θεωρήματα με την εξίσωση (4.4). Ο προκύπτων on-line στοχαστικός αλγόριθμος εκπαίδευσης του SOHMMM είναι μία επαναληπτική διαδικασία, κατά την οποία οι παράμετροι ενδιαφέροντος, τουτέστιν οι $W^{(e)}$, $R^{(e)}$ και $U^{(e)}$, προσαρμόζονται βάσει των κάτωθι κανόνων μάθησης

$$w_{ij}^{(e)}(y+1) = w_{ij}^{(e)}(y) + \eta(y) h_{ce}(y).$$

$$\left[P(O | \lambda)^{-1} \mathbf{a}_{ij} \sum_{l=1}^{T-1} [\alpha_l(i) b_j(o_{l+1}) \beta_{l+1}(j) - \alpha_l(i) \beta_l(i)] \right]_{\lambda_e, y},$$

$$1 \leq i \leq N, 1 \leq j \leq N \quad (4.5)$$

$$r_{ji}^{(e)}(y+1) = r_{ji}^{(e)}(y) + \eta(y) h_{ce}(y) \cdot$$

$$\left[P(O | \lambda)^{-1} \sum_{l=1}^T [I\{o_l = t | \lambda\} \alpha_l(j) \beta_l(j) - b_j(t) \alpha_l(j) \beta_l(j)] \right]_{\lambda_e, y},$$

$$1 \leq i \leq N, 1 \leq t \leq M \quad (4.6)$$

$$u_j^{(e)}(y+1) = u_j^{(e)}(y) + \eta(y) h_{ce}(y) \left[P(O | \lambda)^{-1} \pi_j b_j(o_1) \beta_1(j) - \pi_j \right]_{\lambda_e, y},$$

$$1 \leq i \leq N. \quad (4.7)$$

Υπενθυμίζουμε ότι στις παραπάνω εξισώσεις η ακέραια παράμετρος $y \geq 0$ αντιπροσωπεύει το διακριτό χρόνο και ο δείκτης e συμβολίζει κάθε ξεχωριστό HMM αναφοράς ($1 \leq e \leq E$ όπου E είναι το συνολικό πλήθος HMMs στη θεωρούμενη συστοιχία SOHMMM). Η συνάρτηση $\eta(y)$ αντιστοιχεί στο ρυθμό μάθησης, ένα βαθμωτό μέγεθος για το οποίο ισχύει $0 < \eta(y) < 1$. Συνήθως, ο ρυθμός μάθησης μειώνεται μονότονα ως προς το χρόνο.

4.4. Τροποποιημένος Αλγόριθμος Εκπαίδευσης του SOHMMM

Ένας εναλλακτικός μηχανισμός μάθησης μπορεί να κατασκευαστεί στη βάση μίας έκφρασης ενέργειας, η οποία είναι συνδυασμένη συνάρτηση της πιθανοφάνειας κάθε ξεχωριστής ακολουθίας παρατηρήσεων και της τοπολογικής γειτονιάς του SOHMMM. Όπως αναφέρθηκε και προηγούμενα, ο προκύπτων αλγόριθμος αντιστοιχεί στην εφαρμογή της μεθόδου gradient descent για τη βελτιστοποίηση αυτής της συνάρτησης ενέργειας, γεγονός που μαρτυρεί ένα πιο στέρεο μαθηματικό υπόβαθρο, τουλάχιστον εν συγκρίσει με την ευρετική εκδοχή που παρουσιάστηκε προηγούμενα. Η επιλογή που κάνουμε για την συνάρτηση ενέργειας του SOHMMM είναι η ακόλουθη

$$\mathcal{E} = \left\langle \max_d \sum_{e=1}^E h_{de} P(O | \lambda_e) \right\rangle \quad (4.8)$$

όπου ο συμβολισμός $\langle \dots \rangle$ υποδηλώνει την αναμενόμενη ή μέση τιμή όπως αυτή υπολογίζεται από την κατανομή των δεδομένων ακολουθιών παρατηρήσεων. Υπενθυμίζουμε ότι η παράμετρος h_{de} αναφέρεται στη συνάρτηση τοπολογικής γειτονιάς θεωρώντας ως HMM βέλτιστου συντονισμού το λ_d . Κατ' ουσίαν, η συνάρτηση ενέργειας αντιπροσωπεύει την αναμενόμενη τιμή των μέγιστων σταθμισμένων πιθανοφανειών (των νευρώνων του SOHMMM) όπως αυτή εκτιμάται από την κατανομή των δεδομένων ακολουθιών. Συνεπώς, ως ένα βαθμό, μετράει την προσαρμοστικότητα και την αποδοτικότητα του SOHMMM ως προς τη δυνατότητα περιγραφής και μοντελοποίησης μίας συλλογής από ακολουθίες παρατηρήσεων. Επίσης, τέτοιες (και παρεμφερείς) συναρτήσεις ενέργειας είναι συνεχείς για πεπερασμένα σύνολα προτύπων εισόδου. Μία ισοδύναμη διατύπωση της συνάρτησης ενέργειας (4.8) είναι η κάτωθι

$$\mathcal{E} = \left\langle \sum_{d=1}^E K(O, \lambda_d) \sum_{e=1}^E h_{de} P(O | \lambda_e) \right\rangle \quad (4.9)$$

όπου

$$K(O, \lambda_c) = \begin{cases} 1, & c = \arg \max_d \sum_{e=1}^E h_{de} P(O | \lambda_e) \\ 0, & \text{otherwise.} \end{cases} \quad (4.10)$$

Αμφότερες οι εκφράσεις ενέργειας για το SOHMMM είναι σε αντιστοιχία με την παραλλαγή που έχει προταθεί στις [68], [69], όπου επίσης γίνεται μία τροποποίηση στην διαδικασία καθορισμού του νικητή νευρώνα.

Θεωρώντας ως συνάρτηση προς βελτιστοποίηση την προτεινόμενη συνάρτηση ενέργειας είμαστε σε θέση να εφαρμόσουμε απευθείας τη μέθοδο gradient descent. Δοθείσας μίας ακολουθίας παρατηρήσεων O , η συνάρτηση ενέργειας επιφέρει τη δειγματική συνάρτηση

$$\mathcal{E}(O) = \sum_{d=1}^E K(O, \lambda_d) \sum_{e=1}^E h_{de} P(O | \lambda_e). \quad (4.11)$$

Συνεπώς κάθε on-line επανάληψη του αλγορίθμου μάθησης αντιστοιχίζεται σε ένα βήμα της μεθόδου gradient descent όπως αυτή ισχύει για τη δειγματική συνάρτηση. Έστω ότι η ελεύθερη παράμετρος που θέλουμε να εκτιμήσουμε είναι η x , το αποτέλεσμα της διαδικασίας θα έχει τη μορφή

$$x^{(next)} = x^{(now)} + \eta \left. \frac{\partial \mathcal{E}(O)}{\partial x} \right|_{x=x^{(now)}}. \quad (4.12)$$

Εκτελώντας τη μερική παραγωγήση λαμβάνουμε

$$\frac{\partial \mathcal{E}(O)}{\partial x} = \sum_{d=1}^E \frac{\partial K(O, \lambda_d)}{\partial x} \sum_{e=1}^E h_{de} P(O | \lambda_e) + \sum_{d=1}^E K(O, \lambda_d) \sum_{e=1}^E h_{de} \frac{\partial P(O | \lambda_e)}{\partial x}. \quad (4.13)$$

Ο πρώτος όρος στο δεξί μέλος μηδενίζεται (μία λεπτομερής επεξήγηση περιέχεται στην [31]), ο δεύτερος όρος στο δεξί μέλος αποτελεί τον επιθυμητό κανόνα μάθησης

$$x^{(next)} = x^{(now)} + \eta \sum_{d=1}^E K(O, \lambda_d) \sum_{e=1}^E h_{de} \left. \frac{\partial P(O | \lambda_e)}{\partial x} \right|_{x=x^{(now)}}. \quad (4.14)$$

Ένα επιπρόσθετο θεωρητικό όφελος που προκύπτει από την εφαρμογή της μεθόδου gradient descent για την μεγιστοποίηση της συνάρτησης ενέργειας, είναι ότι γνωρίζουμε εκ των προτέρων τι ακριβώς βελτιστοποιεί ο κανόνας μάθησης του SOHMMM. Αυτό μεταφράζεται στο πρακτικό πλεονέκτημα του ότι διαθέτουμε ένα γενικό μέτρο της απόδοσης της διαδικασίας εκπαίδευσης.

Ο αλγόριθμος που μεγιστοποιεί τη δειγματική συνάρτηση, μέσω της προσαρμογής των παραμέτρων καθενός HMM, δύναται να εξαχθεί συνδυάζοντας τα τρία αποδειχθέντα θεωρήματα με την εξίσωση (4.14). Ο προκύπτον on-line τροποποιημένος στοχαστικός αλγόριθμος εκπαίδευσης του SOHMMM, ο οποίος εισάγει μία παραλλαγή στην διαδικασία καθορισμού του νικητή νευρώνα, είναι μία επαναληπτική διαδικασία κατά την οποία ρυθμίζονται οι παράμετροι ενδιαφέροντος. Κατά συνέπεια, μπορούμε να ορίσουμε το HMM βέλτιστης ενεργοποίησης ως το νευρώνα που παράγει την

μέγιστη τιμή του αθροίσματος σταθμισμένων πιθανοφανειών. Συμβολίζεται με το δείκτη c και η εύρεση αυτού γίνεται χρησιμοποιώντας τον τύπο

$$c = \arg \max_d \sum_{e=1}^E h_{de}(y) P(O | \lambda_e) \Big|_y. \quad (4.15)$$

Οι παράμετροι των HMMs του πλέγματος SOHMMM, τουτέστιν οι $W^{(e)}$, $R^{(e)}$ και $U^{(e)}$, προσαρμόζονται βάσει των κάτωθι τροποποιημένων κανόνων μάθησης

$$w_{ij}^{(e)}(y+1) = w_{ij}^{(e)}(y) + \eta(y) h_{ce}(y) \left[a_{ij} \sum_{l=1}^{T-1} [\alpha_l(i) b_j(o_{l+1}) \beta_{l+1}(j) - \alpha_l(i) \beta_l(i)] \Big|_{\lambda_e, y} \right],$$

$$1 \leq i \leq N, 1 \leq j \leq N \quad (4.16)$$

$$r_{jt}^{(e)}(y+1) = r_{jt}^{(e)}(y) + \eta(y) h_{ce}(y) \left[\sum_{l=1}^T [I\{o_l = t | \lambda\} \alpha_l(j) \beta_l(j) - b_j(t) \alpha_l(j) \beta_l(j)] \Big|_{\lambda_e, y} \right],$$

$$1 \leq i \leq N, 1 \leq t \leq M \quad (4.17)$$

$$u_j^{(e)}(y+1) = u_j^{(e)}(y) + \eta(y) h_{ce}(y) \left[[\pi_j b_j(o_1) \beta_1(j) - \pi_j P(O | \lambda)] \Big|_{\lambda_e, y} \right], 1 \leq j \leq N. \quad (4.18)$$

Υπενθυμίζουμε ότι στις παραπάνω εξισώσεις η ακέραια παράμετρος $y \geq 0$ αντιπροσωπεύει το διακριτό χρόνο.

4.5. Αλγοριθμικά Ζητήματα Υλοποίησης του SOHMMM

Τόσο το SOM (και οι παραλλαγές αυτού) όσο και τα HMMs έχουν μελετηθεί και αναλυθεί σε βάθος, ενώ παράλληλα, έχουν βρει πολλαπλές (πρακτικές) εφαρμογές. Επειδή ακριβώς το SOHMMM είναι ένα υβρίδιο του SOM και των HMMs είμαστε σε θέση, ως ένα βαθμό και τηρουμένων των αναλογιών, να δανειστούμε γνώση και να μεταφέρουμε τεχνογνωσία που έχουν αποκτηθεί από αυτά τα δύο ερευνητικά πεδία.

Στις προηγούμενες ενότητες έγινε μία παρουσίαση του SOHMMM στην πιο ευρεία και πλέον γενικευμένη προτυποποίηση του. Είναι αναμενόμενο ότι μία τέτοια ανάλυση μπορεί να αποτελέσει τη βάση για μία πλειάδα παραλλαγών στον αλγόριθμο, στη δομή

και στην αρχιτεκτονική του SOHMMM. Οι ενότητες που ακολουθούν περιέχουν μία διεξοδική και συνάμα εμπειριστατωμένη περιγραφή του SOHMMM από αλγοριθμικής σκοπιάς. Ακόμη, διερευνώνται και αναλύονται διάφορα κρίσιμα ζητήματα υλοποίησης και πρακτικής εφαρμογής αυτού.

4.5.1. Αρχιτεκτονική των HMM Νευρώνων

Στις περιπτώσεις κατά τις οποίες ο χώρος των καταστάσεων της αλυσίδας Markov (καθενός HMM) δεν έχει κάποια φυσική ή βιολογική ερμηνεία, παρουσιάζεται ένα βασικό πρόβλημα. Αυτό συνίσταται στο γεγονός ότι διαφορετικές τοπολογίες και αρχιτεκτονικές δύνανται να χρησιμοποιηθούν για τη μοντελοποίηση της ίδιας ακριβώς συλλογής ακολουθιών παρατηρήσεων. Το εύλογο ερώτημα που τίθεται είναι το κατά πόσο η HMM αρχιτεκτονική μπορεί να εξαχθεί και στην συνέχεια να ρυθμιστεί από τα πρότυπα εισόδου. Αυτό και παρεμφερή ερωτήματα δεν φαίνεται να έχουν μελετηθεί επαρκώς και πολύ περισσότερο δεν έχουν προταθεί καθολικές λύσεις [13].

Ωστόσο, έχουν υπάρξει κάποιες απόπειρες ανάπτυξης αλγορίθμων για την προσαρμογή HMM αρχιτεκτονικών [70], [71], [62]. Η βασική ιδέα στην [70] είναι ξεκινώντας με το πλέον πολύπλοκο μοντέλο, δηλαδή με ένα HMM που αντιστοιχίζει μία κρυφή κατάσταση σε κάθε παρατηρήσιμο σύμβολο, να συγχωνεύουμε επαναληπτικά τις καταστάσεις. Από την άλλη πλευρά, στην [71] το σημείο εκκίνησης είναι ένα μικρό, πλήρως συνδεδεμένο HMM. Ο αλγόριθμος, σε αυτή την περίπτωση, εξελίσσεται επαναληπτικά διαγράφοντας εκείνες τις μεταβάσεις που παρουσιάζουν πολύ μικρή πιθανότητα και διπλασιάζοντας τις καταστάσεις με το μεγαλύτερο αριθμό συνδέσεων. Τέλος, στην [62] παρουσιάζεται ένας αλγόριθμος για τη δυναμική προσαρμογή του μήκους κάθε HMM κατά τη διάρκεια της εκπαίδευσης. Η ιδέα είναι να προσθέτουμε και να αφαιρούμε καταστάσεις όποτε αυτό απαιτείται, σεβόμενοι πάντα το συνολικό μοτίβο συνδέσεων. Ο αντίκτυπος και η πρακτική εφαρμογή αυτών των τεχνικών στο ερευνητικό πεδίο των HMMs είναι περιορισμένες εξαιτίας των μειονεκτημάτων τους. Ειδικότερα:

- 1) όλες οι μέθοδοι βρίσκουν εφαρμογή μόνο σε ένα μεμονωμένο HMM το οποίο μπορεί να μοντελοποιεί ακριβώς μία κλάση δεδομένων ακολουθιών, επεκτάσεις για τις περιπτώσεις συστοιχιών από HMMs ή για την περίπτωση μοντελοποίησης δύο ή περισσότερων κατηγοριών ακολουθιών παρατηρήσεων δεν προτείνονται σε καμία περίπτωση,

- 2) σε όλες τις προσεγγίσεις αναφέρονται καλά αποτελέσματα για προβλήματα χαμηλής διάστασης που απαιτούν HMMs με μικρούς χώρους καταστάσεων, αντιθέτως στη γενική περίπτωση όπου το πλήθος των πιθανών καταστάσεων είναι υψηλό, οι υπό εξέταση μέθοδοι επιδεικνύουν υψηλό υπολογιστικό κόστος πράγμα που τις καθιστά αργές και συχνά ανεφάρμοστες,
- 3) η επιλογή του ποιες καταστάσεις θα συγχωνευτούν, προστεθούν, διπλασιαστούν, διαγραφούν, όπως επίσης και η επιλογή των κριτηρίων τερματισμού πηγάζουν από υποθέσεις και καθορίζονται από κατώφλια που βασίζονται κατά μεγάλο μέρος σε εκ των προτέρων πληροφορία και γνώση πεδίου,
- 4) η τεχνική που προτείνεται στην [62] βρίσκει εφαρμογή αποκλειστικά σε ένα συγκεκριμένο τύπο αρχιτεκτονικής, συγκεκριμένα τα left-to-right μοντέλα, επιπροσθέτως δεν έχει επαληθευτεί ότι μπορεί πάντα να συγκλίνει σε ένα ευσταθές μέγεθος δικτύου.

Στην περίπτωση του SOHMMM εσκεμμένα έχουμε αποφύγει τη χρήση τεχνικών προεπεξεργασίας. Οι βασικοί λόγοι για αυτή την απόφαση ήταν κυρίως τέσσερις. Πρώτον, η προεπεξεργασία εισαγάγει επιπρόσθετο υπολογιστικό κόστος, γεγονός που σε πρακτικές εφαρμογές μεταφράζεται σε αργές εκτελέσεις. Ειδικότερα, στην περίπτωση του SOHMMM, εάν αντί της εκτίμησης μίας συγκεκριμένης παραμέτρου (δηλαδή του μέτρου του χώρου καταστάσεων) χρησιμοποιούσαμε ένα στάδιο προεπεξεργασίας (όπως στις [70], [71]) το υπεισερχόμενο υπολογιστικό κόστος θα ήταν πολλαπλάσιο του αλγορίθμου εκπαίδευσης αυτού καθ' αυτού. Δεύτερον, όπως φάνηκε από τα προηγούμενα, υπάρχει μία σειρά μεθοδολογιών για την αξιοποίηση εκ των προτέρων πληροφορίας κατά το σχεδιασμό HMMs. Αυτές οι μέθοδοι εξαρτώνται σε μεγάλο βαθμό από το πεδίο εφαρμογής καθώς βασίζονται σε κατά περίπτωση υποθέσεις και εκτιμήσεις, με αποτέλεσμα συχνά να παράγουν αμφισβητούμενα αποτελέσματα. Η ενσωμάτωση μίας τέτοιας μεθόδου θα συνιστούσε σαφή παρέκκλιση από τη μη επιβλεπόμενη προσέγγιση του SOHMMM. Επί παραδείγματι, εάν χρησιμοποιούσαμε μία μεθοδολογία ρύθμισης των HMM αρχιτεκτονικών αξιοποιώντας εκ των προτέρων γνώση (κάτι το οποίο εύλογα παραπέμπει σε στρατηγικές επιβλεπόμενης μάθησης), θα οδηγούσε στο οξύμωρο ένας αλγόριθμος μη επιβλεπόμενης εκπαίδευσης να περιέχει μία ημι-επιβλεπόμενη μονάδα επεξεργασίας. Τρίτον, η απώλεια πληροφορίας, σε ένα στάδιο προεπεξεργασίας (όπως για παράδειγμα είναι η επιλογή χαρακτηριστικών, οι αριθμητικές κωδικοποιήσεις, οι τεχνικές ολισθαίνοντος παραθύρου), είναι τις περισσότερες φορές αναπόφευκτη [2],

[4], [31]. Τέταρτον, επιλέγουμε τυχαίες τιμές αρχικοποίησης για όλες τις παραμέτρους (σεβόμενοι τους συνήθεις στοχαστικούς περιορισμούς) για να αποδείξουμε ότι εκκινώντας από μία αυθαίρετη αρχική κατάσταση τα HMMs μακροπρόθεσμα προσαρμόζονται στην κατανομή των ακολουθιών παρατηρήσεων, πάντα εν σχέση με τη θέση τους στο πλέγμα SOHMMM. Με άλλα λόγια, τυχαία αρχικοποιημένα και αρχικά ανοργάνωτα HMMs στο τέλος ρυθμίζονται και αποκτούν διάταξη, κάτι το οποίο είναι ένα αναμενόμενο επακόλουθο της αυτο-οργάνωσης.

Συμπερασματικά, όπως έχει φανεί και από τις [70], [71], [62], δεν υπάρχει μία γενικής χρήσης, υπολογιστικά εύρωστη, θεωρητικά σωστή μέθοδος για τον καθορισμό του μεγέθους του χώρου καταστάσεων ενός μεμονωμένου HMM. Επιπλέον, είναι επιβεβλημένη η χρήση εκ των προτέρων γνώσης και επιπρόσθετης πληροφορίας πράγμα το οποίο καθιστά όλες τις τεχνικές ισχυρά εξαρτώμενες από το πεδίο εφαρμογής. Κατά συνέπεια, η απόφαση περί του πλήθους των καταστάσεων πρέπει να στηριχθεί στα πρότυπα εισόδου του εκάστοτε αλγορίθμου [72]. Μία προσέγγιση που βρίσκει εφαρμογή αποκλειστικά στα left-to-right μοντέλα είναι να καθορίζεται αυθαίρετα το πλήθος των καταστάσεων ως περίπου ίσο με τη μέση τιμή του πλήθους των παρατηρήσιμων συμβόλων όπως αυτή υπολογίζεται στο σύνολο των δεδομένων ακολουθιών. Από την άλλη πλευρά, στη γενική περίπτωση, ευρετικές μέθοδοι και διαδικασίες εκτίμησης παραμέτρων βασιζόμενες σε δοκιμές είναι οι μόνοι τρόποι ρύθμισης των HMMs άρα και κατάρτισης των αντιστοίχων αρχιτεκτονικών χαρακτηριστικών του αλγορίθμου SOHMMM.

4.5.2. Αρχικοποίηση

Με το πέρας του αλγορίθμου εκπαίδευσης του SOHMMM οι εξισώσεις μάθησης αναθέτουν στις παραμέτρους των HMMs τιμές που αντιστοιχούν σε ένα τοπικό μέγιστο της συνάρτησης διάκρισης ή της συνάρτησης ενέργειας. Ένας καίριο ζήτημα που αναδύεται είναι το κατά πόσο και το πως μπορούμε να πετύχουμε αρχικές εκτιμήσεις για αυτές τις παραμέτρους έτσι ώστε τα εντοπιζόμενα μέγιστα να είναι τα ολικά μέγιστα. Εάν και παρεμφερή ζητήματα είναι κεντρικά και κρίσιμα στη θεωρία βελτιστοποίησης μέχρι σήμερα δεν έχει δοθεί μία σαφής απάντηση.

Τυχαίες αρχικοποιήσεις για τα στοιχεία των στοχαστικών μήτρων W , R , U (και κατά συνέπεια για τα στοιχεία των πινάκων πιθανοτήτων μετάβασης A , εμφάνισης B και της κατανομής πιθανοτήτων αρχικής κατάστασης π όπως αυτά υπολογίζονται από

την εκθετική επαναπαραμετροποίηση) επαρκούν για να παρέχουν χρήσιμες εκτιμήσεις για αυτές τις παραμέτρους, στην πλειοψηφία των περιπτώσεων. Για παράδειγμα, η αρχικοποίηση των στοιχείων των στοχαστικών μήτρων βάσει της ομοιόμορφης κατανομής αποδεικνύεται στην πράξη περισσότερο από επαρκής.

4.5.3. Σπανίως Παρατηρήσιμα Σύμβολα

Ένα δεδομένο, το οποίο έχει επαληθευτεί και πειραματικά, είναι ότι ορισμένες παράμετροι ενός HMM πρέπει να οριοθετούνται ώστε να αποτρέπονται από το να λάβουν πολύ μικρές τιμές. Επί παραδείγματι, ο περιορισμός ότι η παράμετρος $b_j(k)$ πρέπει να είναι μεγαλύτερη μίας δεδομένης ελάχιστη τιμής, είναι απαραίτητος για να διασφαλιστεί ότι ακόμη και εάν το k -οστό σύμβολο παρατήρησης δεν εμφανιζόταν ποτέ στην κατάσταση j του συνόλου παρατηρήσεων εκπαίδευσης, θα υπάρχει πάντοτε μία πεπερασμένη πιθανότητα εμφάνισης του κατά την αποτίμηση μίας άγνωστης ακολουθίας παρατηρήσεων. Μάλιστα, πειραματικά αυτή η ελάχιστη τιμή έχει εντοπιστεί ότι πρέπει να βρίσκεται στην περιοχή $(10^{-10}, 10^{-3})$. Γενικότερα πάντως, πρέπει να έχουμε υπόψη ότι οι αποκλίσεις των παραμέτρων εμφάνισης $b_j(k)$ εμφανίζονται να είναι μεγαλύτερης σημαντικότητας και επίδρασης για τη συμπεριφορά ενός HMM απ' ό,τι αυτές των παραμέτρων a_{ij} και π_j .

4.5.4. Διφορούμενα Σύμβολα

Εξαιτίας του γεγονότος ότι οι τεχνικές προσδιορισμού ακολουθιών δεν είναι τέλειες, ασαφή σύμβολα μπορούν να παρουσιαστούν. Για παράδειγμα, το σύμβολο X μπορεί να αντιπροσωπεύει οποιοδήποτε από τα νουκλεοτίδια A ή C ή G ή T σε μία ακολουθία DNA, παρομοίως το σύμβολο B μπορεί να αναπαριστά την ασπαραγίνη ή το ασπαργανικό οξύ σε μία ακολουθία πρωτεϊνών. Τέτοιου τύπου διφορούμενα σύμβολα μπορούμε να τα χειριστούμε με διάφορους τρόπους στην περίπτωση των HMMs. Η πιο διαδεδομένη πρακτική είναι να ακολουθείται η “benefit of the doubt” προσέγγιση σύμφωνα με την οποία το διφορούμενο σύμβολο αντικαθίσταται από την πιο πιθανή εναλλακτική επιλογή. Μία δεύτερη προσέγγιση έχει να κάνει με την κατασκευή HMMs των οποίων τα σύνολα δειγμάτων $V = \{v_1, v_2, \dots, v_M\}$ περιέχουν και τα ασαφή σύμβολα. Τέλος, εάν ο αριθμός των διφορούμενων συμβόλων σε μία ακολουθία παρατηρήσεων είναι ιδιαίτερα μικρός, τότε απλά μπορεί να αγνοηθεί

πλήρως η ύπαρξη αυτών. Ωστόσο, ιδιαίτερη προσοχή απαιτεί η αντίστροφη κατάσταση κατά την οποία το ποσοστό ασαφών συμβόλων σε μία ακολουθία παρατηρήσεων είναι ασυνήθιστα υψηλό καθώς σε αυτή τη περίπτωση παρατηρούνται λανθασμένες αποκρίσεις. Σε αυτή τη περίπτωση εφόσον υπάρχει ικανοποιητικός αριθμός δεδομένων ακολουθιών με χαμηλά ποσοστά διαφορούμενων συμβόλων είναι καλύτερα οι όποιες προβληματικές ακολουθίες απλά να απορρίπτονται.

4.5.5. Μετατροπή Κλίμακας

Τυπικά οι πιθανότητες $P(o_1 o_2 \dots o_t | \lambda)$ λαμβάνουν πολύ μικρές τιμές, αφού ισούνται με το γινόμενο πολλών πιθανοτήτων μετάβασης και εμφάνισης καθεμία εκ των οποίων είναι μικρότερη της μονάδας. Ένας μπορεί να παρατηρήσει ότι καθώς το t αρχίζει να αυξάνει (δηλαδή να λαμβάνει τιμές μεγαλύτερες από 10) κάθε $a_t(i)$ όρος αρχίζει να συγκλίνει εκθετικά στο μηδέν. Για αρκούντως μεγάλες τιμές του t (δηλαδή για τιμές ίσες και μεγαλύτερες του 100) το αριθμητικό εύρος τιμών για τον υπολογισμό του $a_t(i)$ ξεπερνάει την αριθμητική ακρίβεια ουσιαστικά οποιουδήποτε υπολογιστή, ακόμη και στην περίπτωση αριθμητικής αυξημένης ακρίβειας (διπλής ακρίβειας). Συνεπώς, ο υπολογισμός του $a_t(i)$ καθώς το t αυξάνει, είναι πέραν των δυνατοτήτων πρακτικά οποιουδήποτε ηλεκτρονικού υπολογιστή ακόμη και σε διπλή ακρίβεια. Μία παρεμφερής παρατήρηση μπορεί να γίνει και για τις backward μεταβλητές $\beta_t(i)$, καθώς το t μειώνεται τείνουν στο μηδέν με εκθετικό τρόπο. Γενικότερα, κατά την υλοποίηση του αλγορίθμου SOHMMM και συγκεκριμένα της forward-backward διαδικασίας τα προκύπτοντα ζητήματα αριθμητικής ακρίβειας χρήζουν κατάλληλου χειρισμού. Αυτό που απαιτείται είναι μία διαδικασία μετατροπής κλίμακας βάσει της οποίας οι forward και backward μεταβλητές αλλάζουν περιοχές τιμών κατά τον αναδρομικό υπολογισμό τους (σταθμιζόμενες από έναν κατάλληλο συντελεστή που να συναρτάται του t αλλά να είναι ανεξάρτητος του i) ώστε να αποφευχθεί η υποχείλιση (underflow). Η μετατροπή κλίμακας των forward και backward μεταβλητών ορίζεται με συμπληρωματικό τρόπο έτσι ώστε οι εξισώσεις μάθησης να παραμένουν κατά το δυνατόν αμετάβλητες ως προς αυτό το μετασχηματισμό.

Οι ακριβείς σχέσεις που δίνουν τις υπό κλίμακα forward και backward μεταβλητές στηρίζονται στην προσέγγιση που περιγράφεται στην [72]. Σύμφωνα με αυτή οι υπό συνθήκη μεταβλητές ορίζονται ως

$$\hat{\alpha}_t(i) = \frac{P(o_1 o_2 \dots o_t, q_t = s_i | \lambda)}{P(o_1 o_2 \dots o_t | \lambda)} \quad (4.19)$$

$$\hat{\beta}_t(i) = \frac{P(o_{t+1} o_{t+2} \dots o_T | q_t = s_i, \lambda)}{P(o_{t+1} o_{t+2} \dots o_T | o_1 o_2 \dots o_t, \lambda)}. \quad (4.20)$$

Μία αναδρομική λύση για την υπό κλίμακα forward μεταβλητή $\hat{\alpha}_t(i)$ είναι η ακόλουθη

$$\tilde{\alpha}_1(i) = \pi_i b_i(o_1), 1 \leq i \leq N \quad (4.21)$$

$$\hat{\alpha}_1(i) = sc_1 \tilde{\alpha}_1(i), 1 \leq i \leq N \quad (4.22)$$

$$\tilde{\alpha}_t(i) = \left[\sum_{j=1}^N \hat{\alpha}_{t-1}(j) \mathbf{a}_{ji} \right] b_i(o_t), 2 \leq t \leq T, 1 \leq i \leq N \quad (4.23)$$

$$\hat{\alpha}_t(i) = sc_t \tilde{\alpha}_t(i), 2 \leq t \leq T, 1 \leq i \leq N. \quad (4.24)$$

Οι συντελεστές κλίμακας sc_t ορίζονται ως

$$sc_t = \left[\sum_{i=1}^N \tilde{\alpha}_t(i) \right]^{-1}, 1 \leq t \leq T. \quad (4.25)$$

Αυτό που έπεται είναι ο υπολογισμός της υπό κλίμακα backward μεταβλητής. Η μόνη διαφορά είναι ότι στην προκειμένη περίπτωση χρησιμοποιούνται οι ίδιοι παράγοντες κλίμακας με προηγούμενα. Από τη στιγμή που κάθε συντελεστής κλίμακας επαναφέρει το αριθμητικό μέγεθος των forward μεταβλητών και αφού οι backward μεταβλητές είναι συγκρίσιμες με αυτές, η χρήση του ίδιου παράγοντα κλίμακας είναι μία αποδοτική μέθοδος περιορισμού του επιπρόσθετου υπολογιστικού κόστους. Μία αναδρομική λύση για την υπό κλίμακα backward μεταβλητή $\hat{\beta}_t(i)$ είναι η ακόλουθη

$$\tilde{\beta}_T(i) = 1, 1 \leq i \leq N \quad (4.26)$$

$$\hat{\beta}_T(i) = sc_T \tilde{\beta}_T(i), 1 \leq i \leq N \quad (4.27)$$

$$\tilde{\beta}_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \hat{\beta}_{t+1}(j), t = T-1, T-2, \dots, 1, 1 \leq i \leq N \quad (4.28)$$

$$\hat{\beta}_t(i) = sc_t \tilde{\beta}_t(i), t = T-1, T-2, \dots, 1, 1 \leq i \leq N. \quad (4.29)$$

Πρέπει να καταστεί σαφές ότι η διαδικασία μετατροπής κλίμακας δεν είναι απαραίτητο να εφαρμόζεται σε κάθε χρονική στιγμή t , αλλά μπορεί να χρησιμοποιείται οποτεδήποτε αποφασίζεται ή κρίνεται απαραίτητο (π.χ. για την αποτροπή υποχειλίσισης κατά τους υπολογισμούς). Εάν η αλλαγή κλίμακας δεν εφαρμοστεί κάποια δεδομένη χρονική στιγμή t , ο αντίστοιχος συντελεστής κλίμακας sc_t ορίζεται ίσος με τη μονάδα πράγμα που μεταξύ άλλων ικανοποιεί όλες τις απαιτήσεις που έχουν περιγραφεί προηγούμενα.

Με χρήση της επαγωγικής μεθόδου μπορεί να αποδειχθεί ότι

$$\hat{\alpha}_t(i) = \left[\prod_{\tau=1}^t sc_\tau \right] \alpha_t(i), 1 \leq t \leq T, 1 \leq i \leq N \quad (4.30)$$

$$\hat{\beta}_t(i) = \left[\prod_{\tau=t}^T sc_\tau \right] \beta_t(i), 1 \leq t \leq T, 1 \leq i \leq N. \quad (4.31)$$

όπου οι $\alpha_t(i)$ και $\beta_t(i)$ είναι οι forward και backward μεταβλητές όπως αυτές υπολογίζονται βάσει του αλγορίθμου forward-backward.

Η πρώτη σημαντική αλλαγή που επέρχεται εξαιτίας της αλλαγής κλίμακας αφορά στη διαδικασία υπολογισμού της πιθανοφάνειας. Αρχικά, αξιοποιώντας την κλασσική έκφραση αποτίμησης της $P(O | \lambda)$ για $t = T$ καθώς και τη σχέση (4.30) προκύπτει ότι

$$P(O | \lambda) \prod_{\tau=1}^T sc_\tau = \sum_{j=1}^N \alpha_T(j) \prod_{\tau=1}^T sc_\tau = \sum_{j=1}^N \hat{\alpha}_T(j). \quad (4.32)$$

Επίσης συνδυάζοντας τις (4.24) και (4.25) έχουμε ότι

$$\sum_{j=1}^N \hat{\alpha}_T(j) = \sum_{j=1}^N sc_T \tilde{\alpha}_T(j) = \sum_{j=1}^N \left[\sum_{i=1}^N \tilde{\alpha}_T(i) \right]^{-1} \tilde{\alpha}_T(j) = 1. \quad (4.33)$$

Συνεπώς, η πιθανοφάνεια με χρήση των υπό κλίμακα forward και backward μεταβλητών μπορεί να υπολογιστεί χρησιμοποιώντας τις σχέσεις (4.32) και (4.33)

$$P(O | \lambda) = \left(\prod_{t=1}^T sc_t \right)^{-1} \quad (4.34)$$

ή ισοδύναμα

$$\log P(O | \lambda) = \log \left(\prod_{t=1}^T sc_t \right)^{-1} = - \sum_{t=1}^T \log(sc_t). \quad (4.35)$$

Έτσι μπορούμε να υπολογίσουμε το λογάριθμο της πιθανοφάνειας αντί την τιμή της πιθανοφάνειας αυτής καθαυτής, μιας και όπως έχει αναλυθεί προηγούμενα ο δεύτερος υπολογισμός ξεπερνάει τα όρια ακρίβειας αριθμητικών υπολογισμών οποιουδήποτε μηχανήματος.

Η δεύτερη σημαντική τροποποίηση που επέρχεται στο SOHMMM εξαιτίας της μετατροπής κλίμακας αφορά στους στοχαστικούς κανόνες μη επιβλεπόμενης μάθησης. Ένας αλγόριθμος για την προσαρμογή των παραμέτρων του SOHMMM που ταυτοχρόνως να διατηρεί τους απαιτούμενους υπολογισμούς εντός ορίων αριθμητικής ακρίβειας, δύναται να κατασκευαστεί συνδυάζοντας τις εξισώσεις εκπαίδευσης (4.5)-(4.7) με τις ακόλουθες τρεις προτάσεις. Στην συνέχεια διατυπώνονται και αποδεικνύονται αναλυτικά οι μαθηματικές σχέσεις που αντιστοιχούν στους κανόνες μάθησης του SOHMMM για την περίπτωση που εφαρμόζεται η διαδικασία μετατροπής κλίμακας. Ουσιαστικά οι εν λόγω σχέσεις συνιστούν επεκτάσεις των τριών θεωρημάτων πάνω στα οποία βασίζονται οι αλγόριθμοι εκπαίδευσης του SOHMMM, και πρακτικά, παρέχουν τρόπους υπολογισμού που στηρίζονται εξ ολοκλήρου στις υπό κλίμακα ποσότητες (συντελεστές και μεταβλητές).

Πρόταση 1

$$\sum_{l=1}^{T-1} \left[\hat{\alpha}_l(i) b_j(o_{l+1}) \hat{\beta}_{l+1}(j) - sc_l^{-1} \hat{\alpha}_l(i) \hat{\beta}_l(i) \right] =$$

$$P(O | \lambda)^{-1} \sum_{l=1}^{T-1} \left[\alpha_l(i) b_j(o_{l+1}) \beta_{l+1}(j) - \alpha_l(i) \beta_l(i) \right]. \quad (4.36)$$

Απόδειξη: αντικαθιστούμε τις σχέσεις (4.30) και (4.31) στο αριστερό μέλος της εξίσωσης (4.36) και έχουμε

$$\sum_{l=1}^{T-1} \left[\hat{\alpha}_l(i) b_j(o_{l+1}) \hat{\beta}_{l+1}(j) - sc_l^{-1} \hat{\alpha}_l(i) \hat{\beta}_l(i) \right] =$$

$$\sum_{l=1}^{T-1} \left\{ \left[\prod_{\tau=1}^l sc_\tau \right] \alpha_l(i) b_j(o_{l+1}) \left[\prod_{\tau=l+1}^T sc_\tau \right] \beta_{l+1}(j) - sc_l^{-1} \left[\prod_{\tau=1}^l sc_\tau \right] \alpha_l(i) \left[\prod_{\tau=l}^T sc_\tau \right] \beta_l(i) \right\} =$$

$$\sum_{l=1}^{T-1} \left\{ \left[\prod_{\tau=1}^T sc_\tau \right] \alpha_l(i) b_j(o_{l+1}) \beta_{l+1}(j) - \left[\prod_{\tau=1}^T sc_\tau \right] \alpha_l(i) \beta_l(i) \right\} =$$

$$\sum_{l=1}^{T-1} P(O | \lambda)^{-1} \left[\alpha_l(i) b_j(o_{l+1}) \beta_{l+1}(j) - \alpha_l(i) \beta_l(i) \right]. \quad (4.37)$$

Η τελευταία ισότητα είναι άμεση απόρροια της (4.34). Συνεπώς η σχέση (4.36) αποδείχθηκε.

Πρόταση 2

$$\sum_{l=1}^T \left[I\{o_l = t | \lambda\} sc_l^{-1} \hat{\alpha}_l(j) \hat{\beta}_l(j) - b_j(t) sc_l^{-1} \hat{\alpha}_l(j) \hat{\beta}_l(j) \right] =$$

$$P(O | \lambda)^{-1} \sum_{l=1}^T \left[I\{o_l = t | \lambda\} \alpha_l(j) \beta_l(j) - b_j(t) \alpha_l(j) \beta_l(j) \right]. \quad (4.38)$$

Απόδειξη: αντικαθιστούμε τις σχέσεις (4.30) και (4.31) στο αριστερό μέλος της εξίσωσης (4.38) και έχουμε

$$\begin{aligned}
 \sum_{l=1}^T \left[I\{o_l = t \mid \lambda\} sc_l^{-1} \hat{\alpha}_l(j) \hat{\beta}_l(j) - b_j(t) sc_l^{-1} \hat{\alpha}_l(j) \hat{\beta}_l(j) \right] = \\
 \sum_{l=1}^T \left\{ I\{o_l = t \mid \lambda\} sc_l^{-1} \left[\prod_{\tau=1}^l sc_\tau \right] \alpha_l(j) \left[\prod_{\tau=l}^T sc_\tau \right] \beta_l(j) - \right. \\
 \left. b_j(t) sc_l^{-1} \left[\prod_{\tau=1}^l sc_\tau \right] \alpha_l(j) \left[\prod_{\tau=l}^T sc_\tau \right] \beta_l(j) \right\} = \\
 \sum_{l=1}^T \left\{ I\{o_l = t \mid \lambda\} \left[\prod_{\tau=1}^T sc_\tau \right] \alpha_l(j) \beta_l(j) - b_j(t) \left[\prod_{\tau=1}^T sc_\tau \right] \alpha_l(j) \beta_l(j) \right\} = \\
 \sum_{l=1}^T P(O \mid \lambda)^{-1} \left[I\{o_l = t \mid \lambda\} \alpha_l(j) \beta_l(j) - b_j(t) \alpha_l(j) \beta_l(j) \right]. \quad (4.39)
 \end{aligned}$$

Η τελευταία ισότητα είναι άμεση απόρροια της (4.34). Συνεπώς η σχέση (4.38) αποδείχθηκε.

Πρόταση 3

$$b_j(o_1) \hat{\beta}_1(j) = P(O \mid \lambda)^{-1} b_j(o_1) \beta_1(j). \quad (4.40)$$

Απόδειξη: αντικαθιστούμε τη σχέση (4.31) στο αριστερό μέλος της εξίσωσης (4.40) και έχουμε

$$b_j(o_1) \hat{\beta}_1(j) = b_j(o_1) \left[\prod_{\tau=1}^T sc_\tau \right] \beta_1(j) = P(O \mid \lambda)^{-1} b_j(o_1) \beta_1(j). \quad (4.41)$$

Η τελευταία ισότητα είναι άμεση απόρροια της (4.34). Συνεπώς η σχέση (4.40) αποδείχθηκε.

4.5.6. Τύπος και Μορφή του Πλέγματος Νευρώνων

Για λόγους οπτικής παρατήρησης πρέπει να προτιμάται το εξαγωνικό πλέγμα, καθώς δεν ευνοεί τόσο πολύ τις οριζόντιες και κάθετες κατευθύνσεις όσο το ορθογωνικό. Οι ακμές της συστοιχίας νευρώνων ενδείκνυται να προσομοιάζουν αυτές ενός ορθογωνίου παρά αυτές ενός τετραγώνου, καθώς το ελαστικό δίκτυο που σχηματίζεται από τα HMMs πρέπει να προσαρμοστεί και να σταθεροποιηθεί κατά

μήκος αυτών των δύο διαστάσεων. Ένας μπορεί να παρατηρήσει ότι εάν η συστοιχία νευρώνων ήταν κυκλική δεν θα παρουσίαζε κανένα προσανατολισμό στο χώρο των ακολουθιών παρατηρήσεων, συνεπώς οποιαδήποτε επιμήκης μορφή είναι προτιμητέα.

4.5.7. Συνάρτηση Τοπολογικής Γειτονιάς

Η επιλογή του μεγέθους της τοπολογικής γειτονιάς είναι ένα ζήτημα που χρίζει κάποιας προσοχής. Εάν το αρχικό εύρος της τοπολογικής γειτονιάς είναι μικρό τότε οι νευρώνες του SOHMMM δεν αποκτούν μία συνολική διάταξη. Αντί αυτού εμφανίζονται διατμήσεις, που θυμίζουν μωσαϊκό, μεταξύ των οποίων παρατηρούνται ασυνέχειες. Το φαινόμενο αυτό μπορεί να αποφευχθεί ξεκινώντας με μία διευρυμένη τοπολογική γειτονιά που να συρρικνώνεται με την πάροδο του χρόνου. Το αρχικό εύρος της τοπολογικής γειτονιάς μπορεί να είναι ακόμη μεγαλύτερο και από το ήμισυ της διαμέτρου του δικτύου. Κατά τη διάρκεια της αδρής προσαρμογής του δικτύου, οπότε επιτυγχάνεται η συνολική διάταξη, και ενώ ο ρυθμός μάθησης λαμβάνει αρκετά υψηλές τιμές, η ακτίνα της τοπολογικής γειτονιάς μπορεί να καταλήξει να συμπεριλαμβάνει μόνο τους άμεσα γειτονικούς νευρώνες του λ_c . Σε όλη τη διάρκεια της λεπτομερούς προσαρμογής, δηλαδή του σταδίου της ρύθμισης, η τοπολογική γειτονιά μπορεί ακόμη να απαρτίζεται από τους κοντινότερους γείτονες του νικητή νευρώνα. Το τελικό συμπέρασμα είναι ότι εφόσον η αυτο-οργανούμενη διαδικασία ξεκινάει με μία εκτενή τοπολογική γειτονιά (δηλαδή με μεγάλη ακτίνα για το $NB_c(0)$, ή υψηλή τυπική απόκλιση για την $h_{ce}(0)$, πιθανώς και ίδιας τάξης μεγέθους με το ήμισυ της μεγαλύτερης διάστασης του δικτύου) συνήθως δεν υπάρχει κίνδυνος δημιουργίας ασταθών σχηματισμών στο πλέγμα SOHMMM.

4.5.8. Προσδιορισμός του Ρυθμού Μάθησης

Το χρονικό διάστημα που απαιτείται για να επιτευχθεί μία χοντρική διάταξη στο SOHMMM συνήθως είναι σχετικά μικρό. Η πλειοψηφία του υπολογιστικού φόρτου αφιερώνεται στο στάδιο της λεπτομερούς προσαρμογής (ρύθμισης), κάτι το οποίο είναι απαραίτητο για την επίτευξη μίας αρκούντως καλής στατιστικής ισορροπίας. Δεν είναι πλήρως ξεκάθαρο το πως μπορεί να βελτιστοποιηθεί ο ρυθμός μάθησης κατά το πρώτο στάδιο, μιας και επειδή το εύρος της τοπολογικής γειτονιάς αλλάζει επίσης, η κατάσταση είναι αρκετά περίπλοκη. Αντιθέτως, κατά τη διάρκεια του σταδίου της ρύθμισης, οπότε και το εύρος της τοπολογικής γειτονιάς παραμένει σταθερό, είναι

εφικτό να καθορίσουμε ένα σχεδόν βέλτιστο κανόνα για την ακολουθία τιμών του ρυθμού μάθησης.

Έτσι αρχικά ο ρυθμός μάθησης πρέπει να έχει υψηλές τιμές, πλησίον της μονάδας, και στη συνέχεια να μειώνεται μονότονα. Ο ακριβής τύπος της συνάρτησης $\eta(y)$ δεν είναι καίριας σημασίας. Η συνάρτηση του ρυθμού μάθησης μπορεί να είναι γραμμική, εκθετική ή αντιστρόφως ανάλογη της y . Μετά το πέρας του σταδίου της διάταξης ο ρυθμός μάθησης πρέπει να λαμβάνει τιμές μικρότερες της μονάδας (π.χ. ίσες και μικρότερες από 0.01) για ένα μεγάλο χρονικό διάστημα. Ούτε και σε αυτό το στάδιο είναι κρίσιμο το εάν ο ρυθμός μάθησης μειώνεται εκθετικά ή αντιστρόφως ανάλογα με την y . Στην περίπτωση ωστόσο μεγάλων χαρτών μπορεί ένα από τα ζητούμενα να είναι η ελαχιστοποίηση του συνολικού χρόνου εκπαίδευσης. Στην προκειμένη περίπτωση η επιλογή ενός βέλτιστο κανόνα για τον $\eta(y)$ είναι καθοριστικής σημασίας. Συνήθως τέτοιοι βέλτιστοι κανόνες ανήκουν στην οικογένεια συναρτήσεων που είναι αντιστρόφως ανάλογες της y . Αποδοτικές επιλογές για τις παραμέτρους τέτοιων συναρτήσεων βασίζονται σε πειραματικές διερευνήσεις που έχουν διεξαχθεί. Ένας ημι-εμπειρικός κανόνας, που θεωρεί τον ρυθμό μάθησης κοινό για όλους τους νευρώνες, ορίζει ως κατά μέσο όρο βέλτιστο ρυθμό μάθησης τον

$$\eta(y) = \frac{\Gamma}{y + Z} \quad (4.42)$$

όπου Γ και Z είναι καταλλήλως επιλεγμένες σταθερές.

4.5.9. Περιορισμένος Αριθμός Ακολουθιών Παρατηρήσεων

Στην περίπτωση που διατίθεται μικρός αριθμός δεδομένων ακολουθιών είναι εύλογο ότι αυτές πρέπει να ανακυκλωθούν τον απαραίτητο αριθμό φορών ούτως ώστε να ολοκληρωθεί το απαιτούμενο πλήθος εποχών εκπαίδευσης του αλγορίθμου SOHMMM. Για αυτήν την ανακύκλωση δειγμάτων υπάρχουν διάφορες εναλλακτικές: τα πρότυπα εισόδου δύναται να χρησιμοποιούνται βάσει μίας κυκλικής ή άλλης μετάθεσης, ή ακόμη και να επιλέγονται τυχαία με επανατοποθέτηση από το σύνολο εκπαίδευσης. Η πρακτική εμπειρία έχει δείξει ότι η κυκλική μετάθεση αποδίδει εξίσου καλά με τις υπόλοιπες μαθηματικά πιο δικαιολογημένες μεθόδους.

4.5.10. Χαρακτηρισμός Περιοχών του Χάρτη

Όταν έχει παρουσιαστεί ένας ικανοποιητικός αριθμός ακολουθιών παρατηρήσεων και τα HMMs του πλέγματος, με την εφαρμογή του αλγορίθμου μάθησης SOHMMM, έχουν συγκλίνει σε πρακτικά στάσιμες τιμές, το επόμενο βήμα είναι ο χαρακτηρισμός του χάρτη. Στόχος μίας τέτοιας διαδικασίας είναι ο εντοπισμός των απεικονίσεων διαφορετικών δεδομένων ακολουθιών επί του πλέγματος SOHMMM. Ο τρόπος με τον οποίο μπορούμε να απεικονίσουμε και να αναγνωρίσουμε ακολουθίες παρατήρησης είναι εύλογος. Σε πρώτη φάση εφαρμόζουμε στο SOHMMM ακολουθίες παρατηρήσεων που εκ των προτέρων έχουμε αναλύσει και των οποίων γνωρίζουμε την κατηγοριοποίηση. Σε δεύτερο χρόνο εξετάζουμε σε ποιες θέσεις του χάρτη υπάρχει βέλτιστη ταύτιση (συντονισμός) για αυτά τα πρότυπα εισόδου και χαρακτηρίζουμε τους εμπλεκόμενους HMM νευρώνες αναλόγως των κατηγοριών των αντίστοιχων προτύπων. Έτσι με αυτόν τον τρόπο προσδιορίζονται και ακολούθως χαρακτηρίζονται περιοχές του χάρτη SOHMMM. Από τη στιγμή που το SOHMMM θεωρείται συνεχές κατά μήκος της υποθετικής ελαστικής επιφάνειας που έχει προσαρμοστεί στα πρότυπα εισόδου, είναι δυνατό άγνωστες ακολουθίες παρατήρησης να προσεγγίζονται βάσει των πλησιέστερων ακολουθιών αναφοράς, σε μία διαδικασία που παραπέμπει σε αυτή της διανυσματικής κβάντισης.

4.5.11. Προσέγγιση Πυκνότητας

Η χαρτογράφηση που επιτυγχάνει το SOHMMM αντικατοπτρίζει, ως ένα βαθμό, τις διακυμάνσεις των στατιστικών στοιχείων της κατανομής εισόδου. Οι περιοχές του χώρου εισόδου από τις οποίες αντλούνται δείγματα ακολουθιών παρακολουθήσεων με υψηλή πιθανότητα εμφάνισης αντιστοιχίζονται σε μεγαλύτερα πεδία του χώρου εισόδου και, κατά συνέπεια, με καλύτερη ανάλυση, από ότι οι περιοχές του χώρου εισόδου από τις οποίες αντλούνται δείγματα ακολουθιών με χαμηλή πιθανότητα εμφάνισης. Σαν γενικό κανόνα (ο οποίος έχει επιβεβαιωθεί στην περίπτωση του SOM από πλείστες προσομοιώσεις) μπορούμε να πούμε ότι η απεικόνιση που υπολογίζει ο αλγόριθμος τείνει να υπερ-αντιπροσωπεύει τις περιοχές χαμηλής πυκνότητας εισόδου και να υπο-αντιπροσωπεύει τις περιοχές υψηλής πυκνότητας εισόδου. Με άλλα λόγια, ο αλγόριθμος αποτυγχάνει να παρέχει μία πιστή αναπαράσταση της κατανομής πιθανοτήτων του χώρου εισόδου.

4.5.12. Επιβάλλοντας Αναπαραστάσεις σε Περιοχές του Χάρτη

Ορισμένες φορές και ειδικά όταν το SOHMMM χρησιμοποιείται για την παρακολούθηση και τον έλεγχο πειραματικών δεδομένων, μπορεί να είναι επιθυμητό να απεικονίζονται οι κανονικές περιπτώσεις σε μία συγκεκριμένη θέση (στο κέντρο για παράδειγμα) του χάρτη. Ο τρόπος για να αντιστοιχίσουμε συγκεκριμένες δεδομένες ακολουθίες σε επιθυμητές περιοχές, είναι να χρησιμοποιήσουμε τις εν λόγω ακολουθίες παρατηρήσεων ως πρότυπα εκπαίδευσης των HMMs που εδράζονται σε αυτές τις περιοχές και στη συνέχεια να διατηρούμε το ρυθμό μάθησης χαμηλό σε αυτές τις θέσεις κατά τη διαδικασία της προσαρμογής.

4.5.13. Σπάνιες Περιπτώσεις

Σε πολλά πρακτικά προβλήματα, βαρύνουσας σημασίας περιπτώσεις μπορεί να παρουσιάζονται με μικρή στατιστική συχνότητα, με αποτέλεσμα να μην καταλαμβάνουν καμία περιοχή στο πλέγμα SOHMMM και άρα να μην αναπαρίστανται από το δίκτυο. Τέτοιες αξιοσημείωτες περιπτώσεις δύναται να τονιστούν κατά τη διάρκεια της διαδικασίας προσαρμογής θέτοντας ψηλότερες τιμές στις παραμέτρους η και h_{ce} των αντιστοίχων δειγμάτων, ή αυξάνοντας τη συχνότητα χρησιμοποίησης αυτών των σπάνιων ακολουθιών παρατηρήσεων στη διαδικασία μη επιβλεπόμενης μάθησης του SOHMMM. Η τελική επιλογή της κατάλληλης διαδικασίας επισήμανσης αυτών των περιπτώσεων ενδείκνυται να γίνει σε συνεργασία με τους τελικούς χρήστες των χαρτών ή με τους εμπειρογνώμονες των αντιστοίχων ερευνητικών περιοχών.

Από την άλλη πλευρά, εάν κάποιο υποσύνολο προτύπων εισόδου είναι εξαιρετικά δύσκολο να μοντελοποιηθεί, το πιο πιθανό είναι ότι αντιπροσωπεύει μία ασυνήθιστη περίπτωση ή απλά μία λανθασμένη. Σε αμφότερες τις περιπτώσεις είναι χρήσιμο να απορρίπτονται ακολουθίες παρατηρήσεων που αποκλίνουν καταφανώς της συνολικής εικόνας.

4.5.14. On-Line και Batch Μάθηση

Αξίζει να επισημανθεί ότι αμφότεροι οι αλγόριθμοι μάθησης του SOHMMM είναι καθαρά on-line, και ως τέτοιοι είναι πλησιέστερα στη Bayes φιλοσοφία της άμεσης αναπροσαρμογής μίας εκτίμησης ενόσω γίνονται διαθέσιμα νέα δεδομένα. Μεγαλύτερης σημασίας ωστόσο είναι το γεγονός ότι η απευθείας προσαρμογή ενός μοντέλου αμέσως μόλις παρουσιάζεται ένα πρότυπο εισόδου (δηλαδή σε βάση

παράδειγμα προς παράδειγμα), εισάγει ένα βαθμό στοχαστικότητας ο οποίος μπορεί να αποδειχθεί χρήσιμος κατά την εξερεύνηση του χώρου των πιθανών λύσεων και επίσης μπορεί να αποτρέψει την παγίδευση της διαδικασίας μάθησης σε τοπικά βέλτιστα [73], [63]. Επίσης μία άλλη χρήσιμη ιδιότητα της on-line μάθησης είναι η δυνατότητα της να παρακολουθεί τις μικρές μεταβολές στα δεδομένα εκπαίδευσης, ιδιαίτερα όταν το περιβάλλον που είναι υπεύθυνο για την παραγωγή των δεδομένων είναι μη στατικό. Παρόλα αυτά, δεν πρέπει να παραγνωρίζουμε το γεγονός ότι η batch μάθηση έχει το πλεονέκτημα της επακριβής εκτίμησης του διανύσματος κλίσης (δηλαδή της μερικής παραγώγου της συνάρτησης ενέργειας ως τις παραμέτρους), πράγμα το οποίο εγγυάται, στις απλές περιπτώσεις, σύγκλιση της μεθόδου gradient descent σε ένα τοπικό βέλτιστο [74].

Το ζεύγος των on-line αλγορίθμων εκπαίδευσης του SOHMMM που περιγράφηκε και μελετήθηκε προηγούμενα είναι απλά αντιπροσωπευτικό των εναλλακτικών μορφών που μπορεί να πάρει ένας αλγόριθμος προσαρμογής των παραμέτρων του SOHMMM. Αυτές οι δύο εκδοχές προφανώς μπορούν να αποτελέσουν τη βάση για την ανάπτυξη ενός αριθμού παραλλαγών και διαφορετικών υλοποιήσεων. Για παράδειγμα είναι σχετικά απλό να κατασκευαστεί μία διαδικασία εκπαίδευσης του SOHMMM που να συμπεριλαμβάνει έναν όρο ορμής θ ούτως ώστε να ρυθμίζει την επίδραση της προηγούμενης κατεύθυνσης της μεθόδου gradient descent. Το πλεονέκτημα του όρου ορμής είναι ότι επιταχύνει την κατάβαση όταν η μερική παράγωγος της ενέργειας έχει το ίδιο πρόσημο σε διαδοχικές επαναλήψεις, ενώ παράλληλα, έχει σταθεροποιητική επίδραση σε κατευθύνσεις των οποίων το πρόσημο ταλαντώνει. Επιπλέον, η συμπερίληψη του όρου ορμής ενδέχεται να εμποδίζει τη διαδικασία μάθησης να τερματίσει σε ένα τοπικό ελάχιστο στην επιφάνεια σφάλματος. Τέλος, αξίζει να τονιστεί ότι η κατασκευή batch εκδοχών του αλγορίθμου SOHMMM αποδεικνύεται μία ευθεία διαδικασία. Αυτό που απαιτείται ως πρώτο βήμα είναι η εφαρμογή της μεθόδου gradient descent στη συνάρτηση ενέργειας (4.8) αντί στη δειγματική συνάρτηση. Λαμβάνοντας υπόψη ότι οι ακολουθίες παρατηρήσεων είναι ανεξάρτητες μεταξύ τους, τα επόμενα βήματα είναι παρόμοια με αυτά που απαιτήθηκαν για την εξαγωγή των on-line εκδοχών. Το αποτέλεσμα είναι batch κανόνες μάθησης που ενσωματώνουν πληροφορία προερχόμενη από το σύνολο των διαθέσιμων δεδομένων ακολουθιών, και που ουσιαστικά συνιστούν αθροίσματα των on-line σχέσεων προσαρμογής (οι οποίες αντιστοιχούν σε κάθε ξεχωριστή ακολουθία παρακολούθησεων).

4.5.15. Υπολογιστική Πολυπλοκότητα

Η ανάλυση του on-line στοχαστικού αλγορίθμου μάθησης του SOHMMM δείχνει ότι η υπολογιστική πολυπλοκότητα κάθε επανάληψης είναι της τάξης των $E(N^2T + NMT)$ πράξεων (συνυπολογίζοντας πολλαπλασιασμούς και προσθέσεις), υπενθυμίζουμε ότι E είναι το πλήθος HMMs στη συστοιχία SOHMMM, N είναι το μέτρο του χώρου καταστάσεων, M είναι ο συνολικός αριθμός παρατηρήσιμων συμβόλων και με T ορίζουμε το πλήθος των παρατηρήσεων στην ακολουθία που χρησιμοποιείται ως πρότυπο εκπαίδευσης στην υπό εξέταση επανάληψη. Το γεγονός ότι το υπολογιστικό κόστος κάθε επανάληψης είναι σταθερό και ανεξάρτητο του συνολικού πλήθους προτύπων εκπαίδευσης (ακολουθιών παρατηρήσεων εν προκειμένω) έχει ως αποτέλεσμα ο συνολικός χρόνος εκτέλεσης της διαδικασίας της εκπαίδευσης να είναι γραμμικά ανάλογος του πλήθους των επαναλήψεων του αλγορίθμου SOHMMM.

4.5.16. Παραδείγματα Αλγορίθμων

Στις ενότητες που προηγήθηκαν μελετήθηκε και αναλύθηκε μία πλειάδα ζητημάτων και προβλημάτων που άπτονται της αλγοριθμικής πραγμάτωσης του SOHMMM. Σε ορισμένα εξ αυτών δόθηκαν ακριβείς και μαθηματικά στιβαρές λύσεις, ενώ σε άλλα επιστρατεύθηκε η πειραματική μαρτυρία, η πρακτική εμπειρία και η συσσωρευμένη τεχνογνωσία (συχνά δανειζόμενες από συγγενή ερευνητικά πεδία), τέλος δεν πρέπει να παραγνωρίζουμε ότι κάποια εκ των θεμάτων παραμένουν σε μεγάλο βαθμό ανοικτά. Στη συνέχεια, προσπαθώντας να κινηθούμε σε λιγότερο θεωρητικό και περισσότερο πρακτικό επίπεδο παραθέτουμε δύο ενδεικτικά παραδείγματα στοχαστικών αλγορίθμων εκπαίδευσης του SOHMMM. Συγκεκριμένα παρουσιάζονται υποδείγματα σε ψευδοκώδικα ενός κανονικού και ενός τροποποιημένου αλγορίθμου μη επιβλεπόμενης μάθησης. Ο μεν πρώτος on-line αλγόριθμος μάθησης ακολουθεί την προσέγγιση της μετατροπής κλίμακας και συμπεριλαμβάνει έναν όρο ορμής, ο δε δεύτερος batch αλγόριθμος μάθησης στηρίζεται στη μεγιστοποίηση μίας συνάρτησης ενέργειας ενώ χρησιμοποιεί μία παραλλαγή της μεθόδου καθορισμού του νικητή νευρώνα.

Σε αμφότερα τα χαρακτηριστικά παραδείγματα θεωρούμε ένα σύνολο D ακολουθιών παρατηρήσεων το οποίο συμβολίζουμε ως $OS = \{O^{(1)}, O^{(2)}, \dots, O^{(D)}\}$,

$O^{(d)} = o_1 o_2 \dots o_{T_d}$ είναι η d -οστή ακολουθία παρατηρήσεων, κάθε παρατήρηση αυτής λαμβάνει μία συγκεκριμένη τιμή από το σύνολο συμβόλων $F = \{f_1, f_2, \dots, f_G\}$, ενώ με T_d ορίζουμε το πλήθος των παρατηρήσεων στην ακολουθία.

/* on-line στοχαστικός αλγόριθμος SOHMMM με μετατροπή κλίμακας και χρήση όρου ορμής */

for $e = 1$ to E

$$w_{ij}^{(e)} \leftarrow \text{uniformDistribution}(), 1 \leq i \leq N, 1 \leq j \leq N;$$

$$r_{jt}^{(e)} \leftarrow \text{uniformDistribution}(), 1 \leq j \leq N, 1 \leq t \leq M;$$

$$u_j^{(e)} \leftarrow \text{uniformDistribution}(), 1 \leq j \leq N;$$

$$a_{ij}^{(e)} \leftarrow e^{w_{ij}^{(e)}} / \sum_{l=1}^N e^{w_{il}^{(e)}}, 1 \leq i \leq N, 1 \leq j \leq N;$$

$$b_j^{(e)}(t) \leftarrow e^{r_{jt}^{(e)}} / \sum_{l=1}^M e^{r_{jl}^{(e)}}, 1 \leq j \leq N, 1 \leq t \leq M;$$

$$\pi_j^{(e)} \leftarrow e^{u_j^{(e)}} / \sum_{l=1}^N e^{u_l^{(e)}}, 1 \leq j \leq N;$$

end for

for $y = 1$ to $trainingSteps$

$$O \leftarrow \text{randomCyclicSelection}(OS);$$

for $e = 1$ to E

$$\tilde{\alpha}_t^{(e)}(i) \leftarrow \left[\sum_{j=1}^N \hat{\alpha}_{t-1}^{(e)}(j) a_{ji}^{(e)} \right] b_i^{(e)}(o_t), 2 \leq t \leq T, 1 \leq i \leq N;$$

$$sc_t^{(e)} \leftarrow \left[\sum_{i=1}^N \tilde{\alpha}_t^{(e)}(i) \right]^{-1}, 1 \leq t \leq T;$$

$$\hat{\alpha}_t^{(e)}(i) \leftarrow sc_t^{(e)} \tilde{\alpha}_t^{(e)}(i), 2 \leq t \leq T, 1 \leq i \leq N;$$

$$\tilde{\beta}_t^{(e)}(i) \leftarrow \sum_{j=1}^N a_{ij}^{(e)} b_j^{(e)}(o_{t+1}) \hat{\beta}_{t+1}^{(e)}(j), 1 \leq t \leq T-1, 1 \leq i \leq N;$$

$$\hat{\beta}_t^{(e)}(i) \leftarrow sc_t^{(e)} \tilde{\beta}_t^{(e)}(i), 1 \leq t \leq T-1, 1 \leq i \leq N;$$

end for

$$c \leftarrow \arg \max_e \left\{ - \sum_{t=1}^T \log(sc_t^{(e)}) \right\};$$

for $e = 1$ to E

$$w_{ij}^{(e)} \leftarrow w_{ij}^{(e)} + \eta(y) h_{ce}(y) \left\{ \mathbf{a}_{ij} \sum_{l=1}^{T-1} \left[\hat{\alpha}_l(i) \left(b_j(o_{l+1}) \hat{\beta}_{l+1}(j) - s c_l^{-1} \hat{\beta}_l(i) \right) \right] \right\}_{\lambda_e} \left. \right\} +$$

$$\theta(y) \Delta w_{ij}^{(e)}, 1 \leq i \leq N, 1 \leq j \leq N;$$

$$r_{jt}^{(e)} \leftarrow r_{jt}^{(e)} + \eta(y) h_{ce}(y) \left\{ \sum_{l=1}^T \left[s c_l^{-1} \hat{\alpha}_l(j) \hat{\beta}_l(j) (I\{o_l = t \mid \lambda\} - b_j(t)) \right] \right\}_{\lambda_e} \left. \right\} +,$$

$$\theta(y) \Delta r_{jt}^{(e)}, 1 \leq j \leq N, 1 \leq t \leq M;$$

$$u_j^{(e)} \leftarrow u_j^{(e)} + \eta(y) h_{ce}(y) \left\{ \left[\pi_j b_j(o_1) \hat{\beta}_1(j) - \pi_j \right] \right\}_{\lambda_e} \left. \right\} +$$

$$\theta(y) \Delta u_j^{(e)}, 1 \leq j \leq N;$$

$$\mathbf{a}_{ij}^{(e)} \leftarrow e^{w_{ij}^{(e)}} / \sum_{l=1}^N e^{w_{il}^{(e)}}, 1 \leq i \leq N, 1 \leq j \leq N;$$

$$b_j^{(e)}(t) \leftarrow e^{r_{jt}^{(e)}} / \sum_{l=1}^M e^{r_{jl}^{(e)}}, 1 \leq j \leq N, 1 \leq t \leq M;$$

$$\pi_j^{(e)} \leftarrow e^{u_j^{(e)}} / \sum_{l=1}^N e^{u_l^{(e)}}, 1 \leq j \leq N;$$

end for

end for

Με Δ συμβολίζεται η διαφορά μεταξύ της παρούσας τιμής και της αμέσως προηγούμενης τιμής καθεμίας παραμέτρου εκάστοτε HMM νευρώνα (δηλαδή η μεταβολή καθεμίας παραμέτρου μεταξύ δύο διαδοχικών βημάτων του αλγορίθμου SOHMMM).

/* batch εκδοχή του τροποποιημένου αλγορίθμου SOHMMM */

for $e = 1$ to E

$$w_{ij}^{(e)} \leftarrow \text{uniformDistribution}(), 1 \leq i \leq N, 1 \leq j \leq N;$$

$$r_{jt}^{(e)} \leftarrow \text{uniformDistribution}(), 1 \leq j \leq N, 1 \leq t \leq M;$$

$$u_j^{(e)} \leftarrow \text{uniformDistribution}(), 1 \leq j \leq N;$$

$$\mathbf{a}_{ij}^{(e)} \leftarrow e^{w_{ij}^{(e)}} / \sum_{l=1}^N e^{w_{il}^{(e)}}, 1 \leq i \leq N, 1 \leq j \leq N;$$

$$b_j^{(e)}(t) \leftarrow e^{r_{jt}^{(e)}} / \sum_{l=1}^M e^{r_{jl}^{(e)}}, 1 \leq j \leq N, 1 \leq t \leq M;$$

$$\pi_j^{(e)} \leftarrow e^{u_j^{(e)}} / \sum_{l=1}^N e^{u_l^{(e)}}, 1 \leq j \leq N;$$

end for

for $y = 1$ to *trainingEpochs*

for $d = 1$ to D

$O \leftarrow \text{randomCyclicSelection}(OS)$;

for $e = 1$ to E

$$\alpha_1^{(e)}(j) \leftarrow \pi_j^{(e)} b_j^{(e)}(o_1), 1 \leq j \leq N;$$

$$\alpha_{t+1}^{(e)}(j) \leftarrow \left[\sum_{i=1}^N \alpha_t^{(e)}(i) \mathbf{a}_{ij}^{(e)} \right] b_j^{(e)}(o_{t+1}), 1 \leq t \leq T-1, 1 \leq j \leq N;$$

$$\beta_T^{(e)}(j) \leftarrow 1, 1 \leq j \leq N;$$

$$\beta_t^{(e)}(j) \leftarrow \sum_{i=1}^N \mathbf{a}_{ji}^{(e)} b_i^{(e)}(o_{t+1}) \beta_{t+1}^{(e)}(i), 1 \leq t \leq T-1, 1 \leq j \leq N;$$

end for

$$c \leftarrow \arg \max_e \sum_{\varepsilon=1}^E \exp\left(-\|\delta_e - \delta_\varepsilon\|^2 / 2\sigma^2(y)\right) P(O | \lambda_\varepsilon);$$

for $e = 1$ to E

$$w_{ij}^{(e)} \leftarrow w_{ij}^{(e)} + \eta(y) \exp\left(-\|\delta_c - \delta_e\|^2 / 2\sigma^2(y)\right).$$

$$\left\{ \mathbf{a}_{ij} \sum_{l=1}^{T-1} [\alpha_l(i) (b_j(o_{l+1}) \beta_{l+1}(j) - \beta_l(i))] \right\}_{\lambda_e}, 1 \leq i \leq N, 1 \leq j \leq N;$$

$$r_{jt}^{(e)} \leftarrow r_{jt}^{(e)} + \eta(y) \exp\left(-\|\delta_c - \delta_e\|^2 / 2\sigma^2(y)\right).$$

$$\left\{ \sum_{l=1}^T [\alpha_l(j) \beta_l(j) (I_{\{o_l=t\}} - b_j(t))] \right\}_{\lambda_e}, 1 \leq j \leq N, 1 \leq t \leq M;$$

$$u_j^{(e)} \leftarrow u_j^{(e)} + \eta(y) \exp\left(-\|\delta_c - \delta_e\|^2 / 2\sigma^2(y)\right).$$

$$\left\{ \pi_j [b_j(o_1) \beta_1(j) - P(O | \lambda)] \right\}_{\lambda_e}, 1 \leq j \leq N;$$

end for

end for

for $e = 1$ to E

$$a_{ij}^{(e)} \leftarrow e^{w_{ij}^{(e)}} / \sum_{l=1}^N e^{w_{il}^{(e)}}, 1 \leq i \leq N, 1 \leq j \leq N;$$

$$b_j^{(e)}(t) \leftarrow e^{r_j^{(e)}} / \sum_{l=1}^M e^{r_{jl}^{(e)}}, 1 \leq j \leq N, 1 \leq t \leq M;$$

$$\pi_j^{(e)} \leftarrow e^{u_j^{(e)}} / \sum_{l=1}^N e^{u_l^{(e)}}, 1 \leq j \leq N;$$

end for

end for

4.6. Τεχνικές Οπτικοποίησης και Γραφικής Απεικόνισης

Η ομαδοποίηση (που συχνά καλείται μη επιβλεπόμενη κατηγοριοποίηση) συνιστά ένα κλάδο της διερευνητικής ανάλυσης δεδομένων όπου η εκ των προτέρων πληροφορία και η γνώση πεδίου περί του δοθέντος προβλήματος είναι (ή θεωρούνται) ανύπαρκτες και άγνωστες. Με αυστηρή θεώρηση, μία μεθοδολογία κατηγοριοποιείται ως μη επιβλεπόμενη εάν παράγει αποτελέσματα χρησιμοποιώντας αποκλειστικά ανεπεξέργαστα δεδομένα ή/και αξιοποιώντας πληροφορίες που δύναται να εξαχθούν από τέτοια δεδομένα. Η υπόθεση αυτή είναι επιβεβλημένο να ισχύει σε καθένα στάδιο μίας διαδικασίας ομαδοποίησης (καθώς και στις πιθανές ακόλουθες διαδικασίες οπτικοποίησης): συνεπώς, εκ των προτέρων γνώση και πληροφορία πεδίου πρέπει να θεωρούνται ανύπαρκτες εντός των ορίων ενός αυστηρά ορισμένου μη επιβλεπόμενου πλαισίου εργασίας. Είναι αναμενόμενο ότι τα αποτελέσματα που εξάγουν τεχνικές ομαδοποίησης είναι προσανατολισμένα περισσότερο προς μία ποιοτική περιγραφή του υπό εξέταση προβλήματος και λιγότερο προς μία επακριβή ποσοτική ανάλυση του. Προφανώς θα ήταν παράδοξο εάν η έλλειψη εκ των προτέρων πληροφορίας δεν είχε κάποιο τίμημα. Ωστόσο, σε πραγματικά προβλήματα όπου οι αλληλουχίες είναι αταξινόμητες, ορφανές ή αγνώστου προελεύσεως τα μόνα μέσα που διατίθενται για μία πρώτη αδρή έστω επεξεργασία και ανάλυση είναι οι μη επιβλεπόμενες μεθοδολογίες.

Ειδικότερα, στις περιπτώσεις που αταξινόμητες ακολουθίες και βιολογικά μόρια αλληλουχιών ορίζουν το χώρο του προβλήματος, η μόνη διαθέσιμη πληροφορία είναι οι σχετικές θέσεις των μονομερών στις ακολουθίες (και προφανώς τα μήκη αυτών των αλληλουχιών). Οιαδήποτε επιπρόσθετη εκ των προτέρων γνώση πεδίου συνιστά σαφή παρέκκλιση σε ένα αυστηρό μη επιβλεπόμενο πλαίσιο. Το SOHMMM είναι σε θέση να επιτελεί μη επιβλεπόμενες διεργασίες (όπως για παράδειγμα ομαδοποιήσεις)

λειτουργώντας σε χώρους προβλημάτων που ορίζονται από τεχνητές ακολουθίες και βιολογικές αλληλουχίες. Οι προτεινόμενες τεχνικές, που θα αναλυθούν λεπτομερώς στη συνέχεια, συνιστούν επέκταση και διεύρυνση του φάσματος λειτουργίας του SOHMMM εξοπλίζοντας το με τη δυνατότητα να υλοποιεί οπτικοποιήσεις και να παράγει μη γραμμικές αναπαραστάσεις (δηλαδή καθιστώντας το ικανό να προβάλει δεδομένα υψηλής διάστασης ως σημεία μίας απεικόνισης χαμηλής διάστασης). Αυτό επιτυγχάνεται επινοώντας την απεικόνιση πυκνότητας αλληλουχιών - Sequence Data Density Display (SDDD) και την προβολή πιθανοφάνειας αλληλουχίας - Sequence Likelihood Projection (SLP). Η πρώτη αξιοποιεί και εκμεταλλεύεται, εν παραλλήλω, το σύνολο των διαθέσιμων δεδομένων εν αντιθέσει με τη δεύτερη που χρησιμοποιεί μία δεδομένη ακολουθία για να επιτύχει μία συγκεκριμένη προβολή. Αμφότερες οι δύο τεχνικές οπτικοποίησης μοιράζονται τη δυνατότητα να μπορούν να λειτουργούν σε μία αυστηρά μη επιβλεπόμενη βάση. Προφανώς χαμηλής διάστασης απεικονίσεις είναι προτιμητέες καθώς επιτρέπουν την οπτική παρατήρηση των αποτελεσμάτων/γραφημάτων.

4.6.1 Απεικόνιση Πυκνότητας Αλληλουχιών

Κατ' ουσίαν, η SDDD αναπαριστά το πλήθος/αριθμό των αλληλουχιών που ανατίθενται σε κάθε ξεχωριστό νευρώνα του πλέγματος SOHMMM. Ο νευρώνας νικητής (δηλαδή ο νευρώνας με τη βέλτιστη ταύτιση) καθορίζεται βάσει του κριτηρίου της μέγιστης πιθανοφάνειας. Επιπροσθέτως, η παραγόμενη γραφική απεικόνιση λαμβάνει υπόψη, εκτός της πυκνότητας των αλληλουχιών, τον τύπο και τις διαστάσεις του πλέγματος SOHMMM όπως επίσης και την τοπολογία και τη θέση των νευρώνων. Το σχετικό μέγεθος κάθε νευρώνα στην εν λόγω αναπαράσταση είναι ανάλογο του πλήθους των ακολουθιών που αντιστοιχίζονται σε αυτόν. Απόρροια αυτής της διαδικασίας είναι ότι, σε συγκεκριμένες περιπτώσεις, συστάδες αλληλουχιών δύναται να εντοπιστούν οπτικά, αναζητώντας ομάδες τοπολογικά γειτνιαζόντων κόμβων μεγάλου μεγέθους που να διαχωρίζονται από περιοχές κόμβων μικρού (ή ακόμη και μηδενικού) μεγέθους. Σε αυτό το σημείο χρειάζεται να επισημανθεί ότι ένας κόμβος δεν απεικονίζεται καθόλου (και συχνά αναφέρεται ως μηδενικού μεγέθους) εάν δεν είναι νευρώνας νικητής για καμία από τις δοθείσες αλληλουχίες. Ακολούθως παρατίθεται η διατύπωση του SDDD αλγορίθμου.

/* απεικόνιση πυκνότητας αλληλουχιών */

for $d = 1$ to D

for $e = 1$ to E

$$\tilde{\alpha}_t^{(e)}(i) \leftarrow \left[\sum_{j=1}^N \hat{\alpha}_{t-1}^{(e)}(j) \mathbf{a}_{ji}^{(e)} \right] \mathbf{b}_i^{(e)}(o_t), \quad 2 \leq t \leq T, \quad 1 \leq i \leq N;$$

$$sc_t^{(e)} \leftarrow \left[\sum_{i=1}^N \tilde{\alpha}_t^{(e)}(i) \right]^{-1}, \quad 1 \leq t \leq T;$$

$$\hat{\alpha}_t^{(e)}(i) \leftarrow sc_t^{(e)} \tilde{\alpha}_t^{(e)}(i), \quad 2 \leq t \leq T, \quad 1 \leq i \leq N;$$

$$\log P(O^{(d)} | \lambda_e) \leftarrow - \sum_{t=1}^T \log(sc_t^{(e)});$$

end for

$$c \leftarrow \arg \max_e \{ \log P(O^{(d)} | \lambda_e) \};$$

$$numOfSeqs[c] \leftarrow numOfSeqs[c] + 1;$$

end for

$$maxNumOfSeqs \leftarrow \max_e \{ numOfSeqs[e] \};$$

$$minNumOfSeqs \leftarrow \min_e \{ numOfSeqs[e] \};$$

for $e = 1$ to E

$$SDDD_Paint(numOfSeqs[e], maxNumOfSeqs, minNumOfSeqs);$$

end for

Δύο βασικές επισημάνσεις δύνανται να γίνουν εξετάζοντας τον SDDD αλγόριθμο. Πρώτον, η προκύπτουσα απεικόνιση βασίζεται αλλά και λαμβάνει υπόψη τόσο τις δοθείσες αλληλουχίες όσο και το ήδη εκπαιδευμένο (βάσει αυτών) SOHMMM. Δεύτερον, βάσιμα μπορεί να υποστηριχθεί ότι η SDDD συνιστά μία αυστηρά μη επιβλεπόμενη τεχνική οπτικοποίησης μιας και η μόνη απαιτούμενη/απαραίτητη πληροφορία είναι η παρεχόμενη συλλογή ακολουθιών από μονομερή.

Ένα εγγενές μειονέκτημα της SDDD πηγάζει από την εξάρτησή της από τα δεδομένα εισόδου. Σε συγκεκριμένες περιπτώσεις στις οποίες το πλήθος των αλληλουχιών που ανήκουν σε κάποιες συστάδες είναι τάξεις μεγέθους υψηλότερο εν συγκρίσει με το πλήθος των αλληλουχιών που ανατίθενται στις υπόλοιπες συστάδες, ο

οπτικός εντοπισμός των δευτέρων δυσχεραίνεται (καθώς αυτές συρρικνώνονται στην γραφική απεικόνιση). Για την αντιμετώπιση αυτού του φαινομένου θα μπορούσαν να χρησιμοποιηθούν κατάλληλα επιλεγμένα κατώφλια. Ωστόσο, ο καθορισμός αυτών των κατωφλίων θα παρεξέκλινε εμφανώς της μη επιβλεπόμενης θεώρησης, γεγονός που καθιστά αυτήν την προσέγγιση ανεδαφική.

4.6.2 Προβολή Πιθανοφάνειας Αλληλουχίας

Κατ' ουσίαν, η SLP συνιστά μία γραφική αναπαράσταση της τιμής της πιθανοφάνειας μίας δεδομένης αλληλουχίας όπως αυτή υπολογίζεται για κάθε ξεχωριστό νευρώνα του πλέγματος SOHMMM. Όπως αναμένεται, η αντίστοιχη γραφική απεικόνιση λαμβάνει υπόψη τον τύπο και τις διαστάσεις της συστοιχίας SOHMMM όπως επίσης και την τοπολογία και τη θέση των νευρώνων. Ειδικότερα, η απόχρωση κάθε νευρώνα είναι ανάλογη της τιμής της πιθανοφάνειας που παράγει· η απόχρωση αυτή καθαυτή καθορίζεται από τη χρησιμοποιούμενη χρωματική κλίμακα. Απόρροια αυτής της διαδικασίας είναι το ότι μία συστάδα δύναται να εντοπιστεί οπτικά (θεωρώντας βέβαια ότι η δοθείσα αλληλουχία ανήκει σε κάποια εκ των συστάδων που αναπαρίστανται) αναζητώντας ομάδες τοπολογικά γειτνιαζόντων κόμβων υψηλών πιθανοφανειών που να διαχωρίζονται από περιοχές κόμβων χαμηλών πιθανοφανειών. Στη συνέχεια παρατίθεται η διατύπωση του SLP αλγορίθμου.

/ προβολή πιθανοφάνειας αλληλουχίας */*

$O \leftarrow \text{sequenceSelection}(OS);$

for $e = 1$ to E

$$\tilde{\alpha}_t^{(e)}(i) = \left[\sum_{j=1}^N \hat{\alpha}_{t-1}^{(e)}(j) \mathbf{a}_{ji}^{(e)} \right] \mathbf{b}_i^{(e)}(o_t), \quad 1 \leq i \leq N, 2 \leq t \leq T;$$

$$sc_t^{(e)} = \left(\sum_{i=1}^N \tilde{\alpha}_t^{(e)}(i) \right)^{-1}, \quad 1 \leq t \leq T;$$

$$\hat{\alpha}_t^{(e)}(i) = sc_t^{(e)} \tilde{\alpha}_t^{(e)}(i), \quad 1 \leq i \leq N, 2 \leq t \leq T;$$

$$\log P(O | \lambda_e) \leftarrow - \sum_{t=1}^T \log(sc_t^{(e)});$$

$$likelihood[e] \leftarrow \log P(O | \lambda_e);$$

end for

```
maxLikelihood ←  $\max_e \{likelihood[e]\};$   
minLikelihood ←  $\min_e \{likelihood[e]\};$   
for  $e = 1$  to  $E$   
  SLP_Paint(likelihood[ $e$ ], maxLikelihood, minLikelihood );  
end for
```

Ομοίως και σε αυτήν την περίπτωση δύο βασικές επισημάνσεις δύνανται να γίνουν εξετάζοντας τον SLP αλγόριθμο. Πρώτον, η προκύπτουσα απεικόνιση βασίζεται αλλά και λαμβάνει υπόψη τόσο την επιλεγείσα αλληλουχία όσο και το ήδη εκπαιδευμένο SOHMMM. Δεύτερον, βάσιμα μπορεί να υποστηριχθεί ότι η SLP συνιστά μία αυστηρά μη επιβλεπόμενη τεχνική οπτικοποίησης μιας και η μόνη απαιτούμενη/απαραίτητη πληροφορία είναι η παρεχόμενη ακολουθία μονομερών.

5. ΠΕΙΡΑΜΑΤΑ ΚΑΙ ΕΦΑΡΜΟΓΕΣ

Το πειραματικό πρωτόκολλο που έχει ακολουθηθεί αποσκοπεί στην παρουσίαση των βασικών ιδιοτήτων και των κύριων χαρακτηριστικών του προτεινόμενου αλγορίθμου σε μία προσπάθεια να δοθούν κατευθυντήριες γραμμές για (πιθανές) μελλοντικές υλοποιήσεις και εφαρμογές. Ο αντικειμενικός στόχος των πειραμάτων με τις τεχνητές ακολουθίες παρατηρήσεων είναι να μελετηθούν και να αναλυθούν οι δυνατότητες του SOHMMM ως ενός ενοποιημένου μοντέλου μη επιβλεπόμενης μηχανικής μάθησης για ομαδοποίηση, μη γραμμική απεικόνιση, και οπτικοποίηση δεδομένων ακολουθιών και συστάδων. Αμφότερες οι πειραματικές διατάξεις, τόσο αυτές που βασίζονται σε τεχνητές ακολουθίες παρατηρήσεων όσο και αυτές που βασίζονται στην πρωτεϊνική οικογένεια των σφαιρινών, διαδραματίζουν το ρόλο ενδεικτικών παραδειγμάτων για την επίλυση διαφόρων πραγματικών προβλημάτων. Το πρόβλημα προσδιορισμού των splice sites χρησιμοποιείται για έναν διαφορετικό λόγο. Ο σκοπός αυτού του τελευταίου πειραματικού πλαισίου είναι η συγκριτική μελέτη του SOHMMM έναντι μίας πλειάδας αλγορίθμων μηχανικής μάθησης.

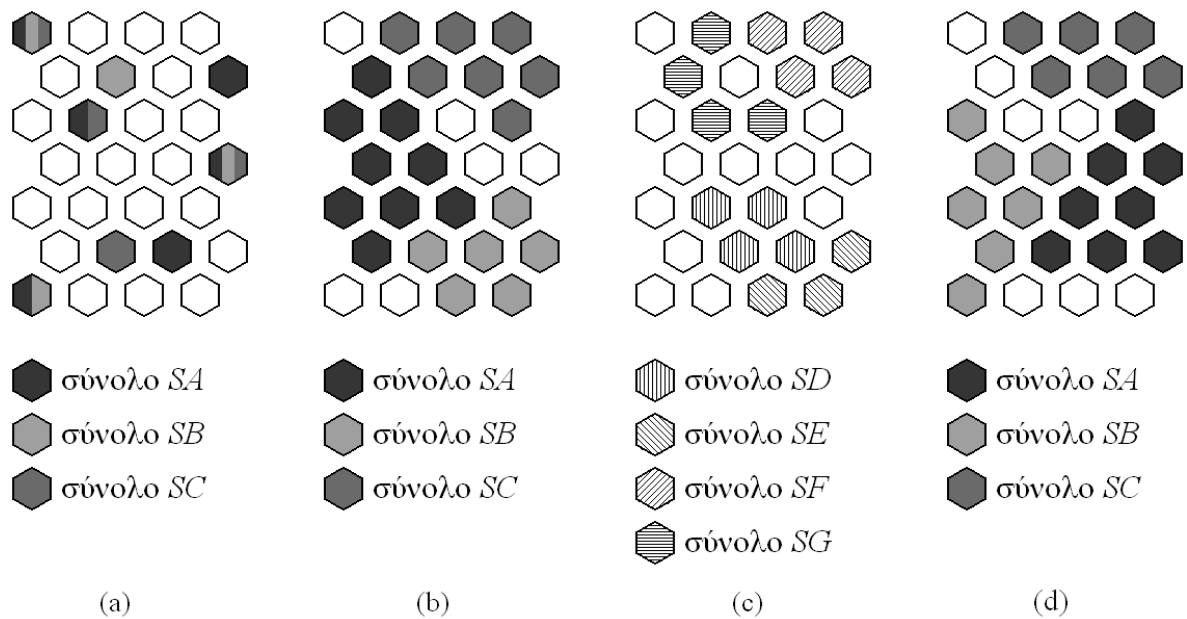
5.1. Τεχνητές Ακολουθίες Παρατηρήσεων

Αρχικά, θεωρούμε το αλφάβητο τριάντα γραμμάτων $F = \{f_1, f_2, \dots, f_{30}\}$. Έστω ότι SA είναι ένα σύνολο δεδομένων ακολουθιών όπου τα σύμβολα κάθε ακολουθίας παρατηρήσεων επιλέγονται ισοπίθانا από το διακριτό βασικό χώρο $\{f_1, f_2, \dots, f_{30}\}$. Ομοίως, έστω ότι SB είναι ένα σύνολο δεδομένων ακολουθιών όπου τα σύμβολα κάθε ακολουθίας παρατηρήσεων επιλέγονται ισοπίθانا από το διακριτό βασικό χώρο $\{f_{16}, f_{17}, \dots, f_{30}\}$. Ακόμη, έστω ότι SC είναι ένα σύνολο δεδομένων ακολουθιών όπου η διαδικασία επιλογής συμβόλων για κάθε ακολουθία παρατηρήσεων είναι αυτή που περιγράφεται στη συνέχεια. Θεωρούμε πέντε, ξένους μεταξύ τους, βασικούς χώρους των έξι γραμμάτων, τους $DN_1 = \{f_1, f_2, \dots, f_6\}$, $DN_2 = \{f_7, f_8, \dots, f_{12}\}$, $DN_3 = \{f_{13}, f_{14}, \dots, f_{18}\}$, $DN_4 = \{f_{19}, f_{20}, \dots, f_{24}\}$ και τέλος τον $DN_5 = \{f_{25}, f_{26}, \dots, f_{30}\}$. Επιπλέον, επιβάλλουμε τον περιορισμό ότι σε κάθε βασικό χώρο ισχύει η ομοιόμορφη κατανομή για την επιλογή ενός συμβόλου, δηλαδή κάθε σύμβολο (το οποίο ανήκει σε ένα βασικό χώρο) επιλέγεται ισοπίθانا ως προς τα υπόλοιπα σύμβολα που απαρτίζουν τον υπό εξέταση βασικό χώρο. Επίσης, θεωρούμε ότι μία χρονική κυκλική μετάθεση

είναι ορισμένη επί αυτών των πέντε βασικών χώρων, συγκεκριμένα αυτή είναι η $DN_1 \rightarrow DN_3 \rightarrow DN_5 \rightarrow DN_2 \rightarrow DN_4 \rightarrow DN_1$. Ένας βασικός χώρος έναρξης επιλέγεται με τυχαίο τρόπο μεταξύ των πέντε. Από τον εν λόγω βασικό χώρο επιλέγεται ισοπίθανα ένα σύμβολο. Ο επόμενος βασικός χώρος καθορίζεται βάσει της χρονικής κυκλικής μετάθεσης και η διαδικασία επιλογής συμβόλου επαναλαμβάνεται. Η συνέχιση της συνολικής διαδικασίας που μόλις περιγράφηκε παράγει ακολουθίες παρατηρήσεων με το επιθυμητό μήκος (δηλαδή με το επιθυμητό πλήθος συμβόλων παρατήρησης). Επίσης, πρέπει να τονιστεί ότι όλα τα πειράματα έχουν διεξαχθεί με ακολουθίες παρατηρήσεων διαφόρων μηκών τα οποία ανήκουν στο διάστημα τιμών $[1000, 10000]$. Κατά συνέπεια, τα μήκη των προκυπτουσών ακολουθιών δύναται να διαφέρουν ακόμη και κατά μία τάξη μεγέθους (κατά μέγιστο).

Το σύνολο εκπαίδευσης για την πρώτη πειραματική διάταξη αποτελείται από 1000 ακολουθίες του συνόλου δεδομένων SA , από 2000 ακολουθίες του συνόλου δεδομένων SB και από 3000 ακολουθίες προερχόμενες από το σύνολο δεδομένων SC . Συνολικά, τα πρότυπα εκπαίδευσης είναι 6000. Η συστοιχία SOHMMM απαρτίζεται από 28 HMMs διατεταγμένα σε ένα εξαγωνικό πλέγμα 4×7 . Ο ρυθμός μάθησης μειώνεται εκθετικά μεταξύ των τιμών 1.0 και 0.1, το εύρος της τοπολογικής γειτονιάς αρχικά είναι ίσο με το ήμισυ της μεγαλύτερης διάστασης του δικτύου και στη συνέχεια μειώνεται γραμμικά μέχρι να συμπεριλαμβάνει μόνο τους άμεσα γειτονικούς νευρώνες του νικητή νευρώνα. Το συνολικό πλήθος εποχών που επιστρατεύθηκαν για την εκπαίδευση του SOHMMM είναι πέντε. Αυτή η συνήθης και κοινή επιλογή παραμέτρων εκπαίδευσης (η οποία αποτελεί κοινό τόπο για πλείστα SOM δίκτυα) αποδεικνύεται περισσότερο από επαρκής. Αντιπροσωπευτικά αποτελέσματα που προκύπτουν μετά την ολοκλήρωση του στοχαστικού αλγορίθμου εκπαίδευσης του SOHMMM παρουσιάζονται στην εικόνα 5.1. Προτείνεται μία διαμέριση του συνόλου εκπαίδευσης σε τρεις συστάδες, καθεμία εκ των οποίων συσχετίζεται με ακριβώς ένα σύνολο ακολουθιών από αυτά που απαρτίζουν το σύνολο προτύπων εισόδου. Οι συστάδες αυτές είναι ξεχωριστές χωρίς επικαλυπτόμενες περιοχές (δηλαδή παρατηρείται υψηλός βαθμός εξωτερικού διαχωρισμού), και επιπλέον, το προσαρμοσμένο SOHMMM αντιστοιχίζει καθεμία εκ των 6000 ακολουθιών παρατηρήσεων στη συστάδα που της αναλογεί (πράγμα που συνεπάγεται αυξημένη εσωτερική ομοιογένεια). Εκτός της ακριβούς αντιστοίχισης των τριών συνόλων ακολουθιών, κάθε συστάδα αντιπροσωπεύεται από μία συλλογή τοπολογικά

γεινιαζόντων HMMs τα οποία σχηματίζουν ένα πιθανοτικό περιγραφέα υψηλότερου επιπέδου.



Εικόνα 5.1. Χαρακτηριστικά πλέγματα SOHMMM όπου κάθε HMM νευρώνας αναπαρίσταται από ένα εξάγωνο. Ένα τυχαία αρχικοποιημένο (ανεκπαιδευτο) SOHMMM εμφανίζεται στην (a). Το ίδιο SOHMMM μετά το πέρας του στοχαστικού αλγορίθμου μάθησης παρουσιάζεται στην (b). Διαφορετικές αποχρώσεις του γκρι χρησιμοποιούνται για τη διάκριση των HMMs που περιγράφουν ακολουθίες παρατηρήσεων προερχόμενες από τα τρία κύρια σύνολα δεδομένων. Στην εικόνα (c) νευρώνες που αναπαριστούν ακολουθίες παρατηρήσεων οι οποίες ανήκουν σε καθένα από τα τέσσερα διαφορετικά παράγωγα σύνολα δεδομένων διαχωρίζονται με τη χρήση κάθετων, διαγώνιων και οριζόντιων διαγραμμίσεων. Μία ενδεικτική συστοιχία SOHMMM μετά την ολοκλήρωση της αντίστοιχης διαδικασίας μη επιβλεπόμενης μάθησης που βασίστηκε στο περιορισμένο σύνολο εκπαίδευσης απεικονίζεται στην (d).

Η δεύτερη πειραματική διάταξη αξιοποιεί το εκπαιδευμένο SOHMMM από την πρώτη σειρά πειραμάτων και απαιτεί επιπλέον τον ορισμό τεσσάρων νέων συνόλων δεδομένων ακολουθιών. Το SD είναι ένα σύνολο δεδομένων ακολουθιών όπου τα σύμβολα κάθε ακολουθίας παρατηρήσεων επιλέγονται ισοπίθانا από το βασικό χώρο $\{f_9, f_{10}, \dots, f_{30}\}$. Ομοίως, το SE είναι ένα σύνολο δεδομένων ακολουθιών όπου τα σύμβολα κάθε ακολουθίας παρατηρήσεων επιλέγονται ισοπίθانا από το βασικό χώρο $\{f_{24}, f_{25}, \dots, f_{30}\}$. Θεωρούμε πέντε, ξένους μεταξύ τους, βασικούς χώρους των τριών γραμμμάτων, τους $\{f_1, f_2, f_3\}$, $\{f_8, f_{10}, f_{12}\}$, $\{f_{13}, f_{15}, f_{18}\}$, $\{f_{22}, f_{23}, f_{24}\}$, $\{f_{25}, f_{27}, f_{29}\}$. Επιπλέον, επιβάλλουμε τον περιορισμό ότι σε κάθε βασικό χώρο ισχύει η ομοιόμορφη

κατανομή για την επιλογή ενός συμβόλου. Επίσης, θεωρούμε ότι μία χρονική κυκλική μετάθεση είναι ορισμένη επί αυτών των πέντε βασικών χώρων, αυτή είναι η $\{f_1, f_2, f_3\} \rightarrow \{f_{13}, f_{15}, f_{18}\} \rightarrow \{f_{25}, f_{27}, f_{29}\} \rightarrow \{f_8, f_{10}, f_{12}\} \rightarrow \{f_{22}, f_{23}, f_{24}\} \rightarrow \{f_1, f_2, f_3\}$.

Τότε το SF είναι ένα σύνολο δεδομένων ακολουθιών που η διαδικασία δημιουργίας του είναι παρόμοια με αυτή που παράγει ακολουθίες παρατηρήσεων για το σύνολο SC . Τέλος, θεωρούμε ότι μία χρονική αμφίδρομη μετάθεση είναι ορισμένη επί των βασικών χώρων έξι γραμμάτων που ορίστηκαν προηγούμενα, συγκεκριμένα αυτή είναι η $DN_1 \leftrightarrow DN_3 \leftrightarrow DN_5 \leftrightarrow DN_2 \leftrightarrow DN_4 \leftrightarrow DN_1$. Σε κάθε βήμα το ποια κατεύθυνση θα ακολουθηθεί καθορίζεται με τυχαίο τρόπο θεωρώντας ίσες πιθανότητες για αμφότερες τις κατευθύνσεις. Το SG είναι ένα σύνολο δεδομένων ακολουθιών που η διαδικασία δημιουργίας του είναι ταυτόσημη με αυτή που παράγει ακολουθίες παρατηρήσεων για το προηγούμενο σύνολο SC , θεωρώντας την αμφίδρομη μετάθεση. Τα HMMs βέλτιστης ταύτισης, δηλαδή αυτά που επιστρέφουν τη μέγιστη τιμή πιθανοφάνειας (ή αθροίσματος σταθμισμένων πιθανοφανειών), για τις ακολουθίες παρατηρήσεων κάθε συνόλου εμφανίζονται στην εικόνα 5.1(c). Είναι σημαντικό να σημειωθεί ότι οι ακολουθίες παρατηρήσεων που περιέχονται στο σύνολο SE (SF) ανατίθενται σε νευρώνες που εντοπίζονται εντός της συστάδας που αφορά στο σύνολο SB (SC). Αυτό είναι κάτι αναμενόμενο από τη στιγμή που βάσει του ορισμού του το SE (SF) είναι ειδική περίπτωση, ή ισοδύναμα υποσύνολο, του συνόλου SB (SC). Από την άλλη πλευρά, ακολουθίες παρατηρήσεων που ανήκουν στο σύνολο SD (SG) συσχετίζονται με HMMs που βρίσκονται περίξ και επί των συνόρων των συστάδων που περιγράφουν τα σύνολα SA και SB (SA και SC). Απεικονίσεις τέτοιου τύπου μπορούν να αιτιολογηθούν εάν αναλογιστούμε ότι οι ακολουθίες παρατηρήσεων του SD (SG) κατασκευάζονται ως ενδιάμεσες περιπτώσεις των ακολουθιών παρατηρήσεων που περιέχονται στα σύνολα SA και SB (SA και SC). Είναι εύκολο να διαπιστώσει κανείς ότι όλες οι προηγούμενες επισημάνσεις αντικατοπτρίζονται απευθείας στην απεικόνιση του SOHMMM στην εικόνα 5.1(c).

Η τρίτη πειραματική διάταξη περιλαμβάνει ένα SOHMMM πλήρως ταυτόσημο με αυτό που χρησιμοποιήθηκε στην πρώτη σειρά πειραμάτων, επίσης χρησιμοποιεί ένα σύνολο εκπαίδευσης που αποτελείται από 30 ακολουθίες παρατηρήσεων προερχόμενες από καθένα εκ των τριών αρχικών συνόλων δεδομένων ακολουθιών, δηλαδή των SA , SB και SC . Έτσι στη προκειμένη περίπτωση, το συνολικό πλήθος των προτύπων εκπαίδευσης είναι δύο τάξεις μεγέθους μικρότερο αυτού που χρησιμοποιήθηκε στα

πρώτα πειράματα. Τα αποτελέσματα κατόπιν της ολοκλήρωσης του στοχαστικού αλγορίθμου μάθησης του SOHMMM παρουσιάζονται στην εικόνα 5.1(d). Τα ευρήματα και οι παρατηρήσεις αναφορικά με τις σχηματισθείσες συστάδες είναι πανομοιότυπα με αυτά της πρώτης πειραματικής διάταξης. Τα ποιοτικά χαρακτηριστικά και οι βασικές ιδιότητες των δύο SOHMMs συμπίπτουν, επουσιώδεις διαφορές που αφορούν στη θέση και τη μορφή των αντίστοιχων συστάδων οφείλονται κατά κύριο λόγο στις τυχαίες αρχικοποιήσεις και στη διαφορετική σειρά παρουσίασης των δεδομένων ακολουθιών στις εισόδους των SOHMMs. Επίσης παρεμφερείς επισημάνσεις μπορούν να γίνουν εάν ληφθούν υπόψη τα τέσσερα σύνολα δεδομένων ακολουθιών της δεύτερης πειραματικής διάταξης. Επιπλέον, όταν οι 6000 ακολουθίες παρατηρήσεων (της πρώτης πειραματικής διάταξης) εφαρμοστούν σε αυτό το δεύτερο SOHMMM, όλες αντιστοιχίζονται στη συστάδα που αναπαριστά το σύνολο δεδομένων ακολουθιών στο οποίο ανήκουν.

Αυτή η αρχική σειρά πειραμάτων, που βασίζεται σε τεχνητές ακολουθίες παρατηρήσεων, μας επιτρέπει να εξάγουμε ορισμένα συμπεράσματα αναφορικά με τις ιδιότητες και τις δυνατότητες του SOHMMM. Το SOHMMM ενσωματώνει ένα μηχανισμό για την επεξεργασία ακολουθιών απαρτιζόμενων από διακριτά σύμβολα προερχόμενα από αλφάβητα αυθαίρετου μεγέθους, όπως το χρησιμοποιηθέν αλφάβητο τριάντα γραμμάτων. Επίσης, είναι σε θέση να χειρίζεται αποδοτικά ακολουθίες παρατηρήσεων υψηλής διάστασης και διαφορετικών μηκών, που ενδέχεται να διαφέρουν ακόμη και κατά μία τάξη μεγέθους. Αξιοποιώντας συγκεκριμένα χαρακτηριστικά των HMMs πετυχαίνει να χειρίζεται τα μη αριθμητικά σύμβολα κάθε ακολουθίας παρατηρήσεων απευθείας, χωρίς την παρεμβολή καμίας διαδικασίας προεπεξεργασίας ή μετασχηματισμού των συμβολικών ποσοτήτων σε αριθμητικές. Το SOHMMM μπορεί να συλλαμβάνει και να αξιοποιεί τη λανθάνουσα πληροφορία που βρίσκεται κρυμμένη στις χρονικές/χωρικές συσχετίσεις και στην γραμμική ακολουθιακή φύση μορίων αλληλουχιών, όπως στις περιπτώσεις των συνόλων SC , SF και SG . Επιπρόσθετα, το SOHMMM εμφανίζει μία σχετική ανοχή απέναντι σε φαινόμενα περιορισμένου πλήθους προτύπων εισόδου (δηλαδή σε καταστάσεις μικρού αριθμού διαθέσιμων δεδομένων εκπαίδευσης), κάτι το οποίο γίνεται εμφανές από τα αποτελέσματα της τρίτης σειράς πειραμάτων στην οποία λίγα μόνο πρότυπα εκπαίδευσης αποδείχθηκαν αρκετά για την προσαρμογή του SOHMMM. Ωστόσο, πρέπει να σημειωθεί ότι ο ελάχιστος απαιτούμενος αριθμός δεδομένων εκπαίδευσης εξαρτάται σε μεγάλο βαθμό από το υπό εξέταση πρόβλημα, μιας και καθορίζεται από

την πολυπλοκότητα και τη φύση του ίδιου του προβλήματος όπως επίσης και από το πόσο περιγραφικά είναι τα πρότυπα εισόδου του συνολικού χώρου δεδομένων ακολουθιών.

Μία από τις βασικές ιδιότητες του SOHMMM είναι η δυνατότητα του να προσαρμόζεται και να μαθαίνει, χρησιμοποιώντας αποκλειστικά και μόνο τα παραδείγματα εισόδου, ακολουθώντας μία ευθεία διαδικασία μη επιβλεπόμενης εκπαίδευσης. Είναι αξιοσημείωτο ότι αυτό το επιτυγχάνει χωρίς να απαιτεί ή να καταφεύγει σε κανενός είδους εκ των προτέρων ή εκ των υστέρων γνώσης (όπως επί παραδείγματι πληροφορία κατηγοριοποίησης). Έτσι, στο πρώτο και στο τρίτο πείραμα, το SOHMMM κατάφερε να ανιχνεύσει την ύπαρξη των τριών συστάδων και να ομαδοποιήσει κατάλληλα τις ακολουθίες παρατηρήσεων του συνόλου εκπαίδευσης. Συγκεκριμένα, με χρήση μηδενικής εκ των προτέρων γνώσης και αξιοποιώντας τη σημαντική πληροφορία που ενυπάρχει στις χωρικές/χρονικές εξαρτήσεις των συμβόλων κάθε ακολουθίας παρατηρήσεων, το SOHMMM επιτυγχάνει τη δημιουργία ανώτερων αφαιρέσεων (οι οποίες περιγράφουν κάθε συστάδα) γραμμένων στη στατιστική γλώσσα των HMMs. Από την άλλη πλευρά, στις περιπτώσεις που παρέχεται η δυνατότητα χαρακτηρισμού περιοχών του χάρτη (δηλαδή εφόσον δίνονται ακολουθίες παρατηρήσεων στις οποίες έχει ήδη γίνει ανάθεση χαρακτηρισμών κλάσεων) το SOHMMM δύναται να χρησιμοποιηθεί για αναγνώριση, αναζήτηση και κατηγοριοποίηση άλλων, εν γένει αγνώστων, ακολουθιών παρατηρήσεων. Μία τέτοια χαρακτηριστική εφαρμογή είναι αυτή που περιγράφεται στα τρίτα πειράματα όπου ομαδοποιήθηκαν 6000 ακολουθίες παρατηρήσεων οι οποίες δεν είχαν χρησιμοποιηθεί προηγουμένως κατά την εκπαίδευση του δικτύου.

Όπως έχει δειχθεί, το SOHMMM απεικονίζει συστάδες, καθώς και τις σχετικές θέσεις τους, επάνω σε πλέγματα μικρής διάστασης. Καθεμία εκ των τριών συστάδων συνιστά μία συνοπτική περιγραφή των ιδιοτήτων και των χαρακτηριστικών του αντίστοιχου συνόλου δεδομένων ακολουθιών. Τα HMMs κάθε συστάδας εξελίσσονται σε αποκωδικοποιητές επιμέρους περιοχών του συνολικού χώρου ακολουθιών, με αυτόν τον τρόπο υποσύνολα δεδομένων ακολουθιών αναπαριστώνται από πιθανοτικά μοντέλα με μία διαδικασία που παραπέμπει σε συμπίεση πληροφορίας. Αποδείχθηκε επίσης πειραματικά ότι στις περιπτώσεις των συνόλων SD και SG , το SOHMMM έφτασε ένα επίπεδο γενίκευσης που ξεπερνά την απλή περιγραφή χαρακτηριστικών εντός συστάδας και επεκτείνεται στην πιο σύνθετη περιγραφή ιδιοτήτων μεταξύ συστάδων. Εν μέρη, αυτό οφείλεται στο γεγονός ότι το SOHMMM καλύπτει ένα

διευρυμένο σύνολο κατανομών πιθανοτήτων πράγμα που σημαίνει ότι είναι σε θέση να αναπαραστήσει συσχετίσεις που είναι αδύνατο να μοντελοποιηθούν από ασύνδετα, μη διατεταγμένα, ανεξάρτητα HMMs.

5.2. Πρωτεϊνική Οικογένεια Σφαιρινών

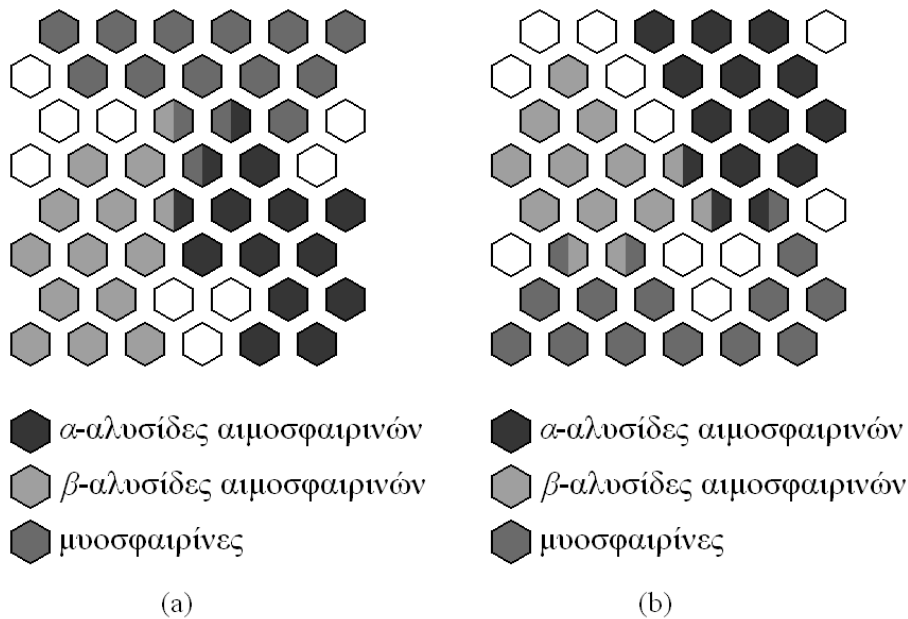
Στο παρόν σύνολο πειραμάτων η προσοχή μας θα είναι εστιασμένη στις σφαιρίνες. Οι σφαιρίνες συνιστούν μία ευρέως γνωστή οικογένεια πρωτεϊνών που αποτελούνται από προσθετικές ομάδες οι οποίες ονομάζονται αίμη και από πρωτεϊνικές αλυσίδες. Καθοριστικό ρόλο παίζει η παρουσία σιδήρου σε αυτές τις προσθετικές ομάδες, καθώς αυτός οποίος έχει υψηλότερη τάση σύνδεσης με το οξυγόνο και χαμηλότερη με το διοξείδιο του άνθρακα. Η λειτουργία των (αιμο)σφαιρινών έχει να κάνει με τη μεταφορά οξυγόνου στους ιστούς και την απαγωγή του διοξειδίου του άνθρακα από αυτούς. Ειδικότερα η μυοσφαιρίνη, που εντοπίζεται στους μύες, χρησιμεύει στην παραλαβή οξυγόνου από το αίμα, στην αποθήκευση του και στη μετακίνηση του μέσα στα μυϊκά κύτταρα. Η οικογένεια (κλάση) των σφαιρινών είναι μία εκτενής οικογένεια που απαρτίζεται από υποοικογένειες (υποκλάσεις). Από κρυσταλλογραφικές μελέτες προκύπτει ότι όλες οι σφαιρίνες έχουν παρόμοιες τρισδιάστατες δομές, αλλά απαρτίζονται από αρκετά διαφορετικές αλληλουχίες αμινοξέων. Οι αλληλουχίες σφαιρινών που χρησιμοποιήθηκαν σε όλα τα πειράματα ανακτήθηκαν από την βάση γνώσης iProClass [75], η οποία είναι μία εκτενής βάση δεδομένων που παρέχει ενσωμάτωση πληροφοριών προερχόμενων από την ενοποίηση περισσότερων από ενενήντα επιμέρους βάσεων πρωτεϊνών. Συνολικά ανακτήθηκαν 560 πρωτεΐνες οι οποίες ανήκουν σε τρεις εκ των βασικών υποκατηγοριών σφαιρινών (α – αλυσίδες και β – αλυσίδες αιμοσφαιρινών καθώς και μυοσφαιρίνες). Το προκύπτον σύνολο δεδομένων απαρτίζεται από 194 α – σφαιρίνες, 216 β – σφαιρίνες και 150 μυοσφαιρίνες. Συνεπώς, για το σύνολο των πειραμάτων που ακολουθούν θεωρούμε το αλφάβητο είκοσι γραμμμάτων των αμινοξέων.

5.2.1. Βασική Διερεύνηση

Η πρώτη σειρά πειραμάτων χρησιμοποιεί ολόκληρο το σύνολο διαθέσιμων πρωτεϊνών. Η συστοιχία SOHMMM απαρτίζεται από 48 HMMs διατεταγμένα σε ένα εξαγωνικό πλέγμα 6x8. Σε πρώτη φάση για να ελεγχθούν (σε γενικές γραμμές) οι δυνατότητες και οι επιδόσεις του SOHMMM, διενεργήθηκαν πολλαπλά πειράματα με

διαφόρους ρυθμούς μάθησης και διαφορετικές συναρτήσεις τοπολογικής γειτονιάς. Στο δια ταύτα οι συνήθειες και ευρέως χρησιμοποιούμενες επιλογές παραμέτρων και συναρτήσεων εκπαίδευσης (οι οποίες αποτελούν κοινό τόπο για πλείστα SOM δίκτυα) αποδεικνύονται περισσότερο από επαρκείς. Συνεπώς, αποφασίστηκε ο ρυθμός μάθησης να μειώνεται εκθετικά μεταξύ των τιμών 1.0 και 0.1. Ακόμη αποφασίστηκε το εύρος της τοπολογικής γειτονιάς αρχικά να είναι ίσο με το ήμισυ της μεγαλύτερης διάστασης του δικτύου και στη συνέχεια να μειώνεται γραμμικά μέχρι να συμπεριλαμβάνει μόνο τους άμεσα γειτονικούς νευρώνες του νικητή νευρώνα. Επίσης, το συνολικό πλήθος εποχών που επιστρατεύθηκαν για την εκπαίδευση του SOHMMM είναι πέντε. Στην παρούσα μελέτη η μέση τιμή των ορθά ομαδοποιημένων πρωτεϊνικών ακολουθιών επιστρατεύτηκε ως στατιστικό μέτρο αξιολόγησης της επίδοσης του δικτύου. Με άλλα λόγια, πληροφορία ανάθεσης κατηγοριών η οποία έχει εξαιρεθεί από τη διαδικασία προσαρμογής χρησιμοποιείται για την εκ των υστέρων εκτίμηση των δυνατοτήτων αναπαράστασης και γενίκευσης του εκπαιδευμένου πλέον μοντέλου. Η μέση τιμή αυτού του μέτρου, υπολογιζόμενη επί 100 επαναλήψεων του αλγορίθμου μη επιβλεπόμενης μάθησης, ισούται με $95.31 \pm 0.57\%$. Το μέτρο αυτό όπως υπολογίζεται με τη μεθοδολογία stratified ten-fold cross-validation για όλες τις 560 σφαιρίνες, λαμβάνει την τιμή $95.23 \pm 0.61\%$. Αντιπροσωπευτικά αποτελέσματα που προκύπτουν μετά την ολοκλήρωση του στοχαστικού αλγορίθμου εκπαίδευσης του SOHMMM παρουσιάζονται στην εικόνα 5.2. Το SOHMMM καταφέρνει να συλλαμβάνει και να επεξεργάζεται τις στατιστικές ιδιότητες των σφαιρινών. Επίσης, επιτυγχάνει να διαιρέσει το σύνολο δεδομένων σε τρεις συστάδες πρωτεϊνικά συγγενών και παρεμφερών σφαιρινών. Καθεμία εξ αυτών των συστάδων συσχετίζεται με μία υποοικογένεια σφαιρινών από αυτές που απαρτίζουν το σύνολο δεδομένων. Η εξέταση των αποτελεσμάτων μπορεί να επαληθεύσει ότι οι συστάδες, που σχηματίζονται κατά τη διάρκεια μίας πλήρως αυτοματοποιημένης διαδικασίας μη επιβλεπόμενης μάθησης, είναι διακριτές και συνεκτικές με σαφώς καθορισμένα σύνορα. 185 α -σφαιρίνες, 205 β -σφαιρίνες και 143 μυοσφαιρίνες ανατίθενται στους HMM νευρώνες που σχηματίζουν τις συστάδες των αντίστοιχων πρωτεϊνικών υποκλάσεων. Συνεπώς, η μεγάλη πλειοψηφία αλληλουχιών πρωτεϊνών ομαδοποιείται σωστά μιας και συσχετίζεται με HMMs που αναπαριστούν τις επιμέρους περιοχές των χώρων που ορίζονται από τις εκάστοτε υποοικογένειες σφαιρινών. Μόνο ένας μικρός αριθμός ακολουθιών (27 πρωτεΐνες για την ακρίβεια) ανατίθενται σε HMMs που εντοπίζονται

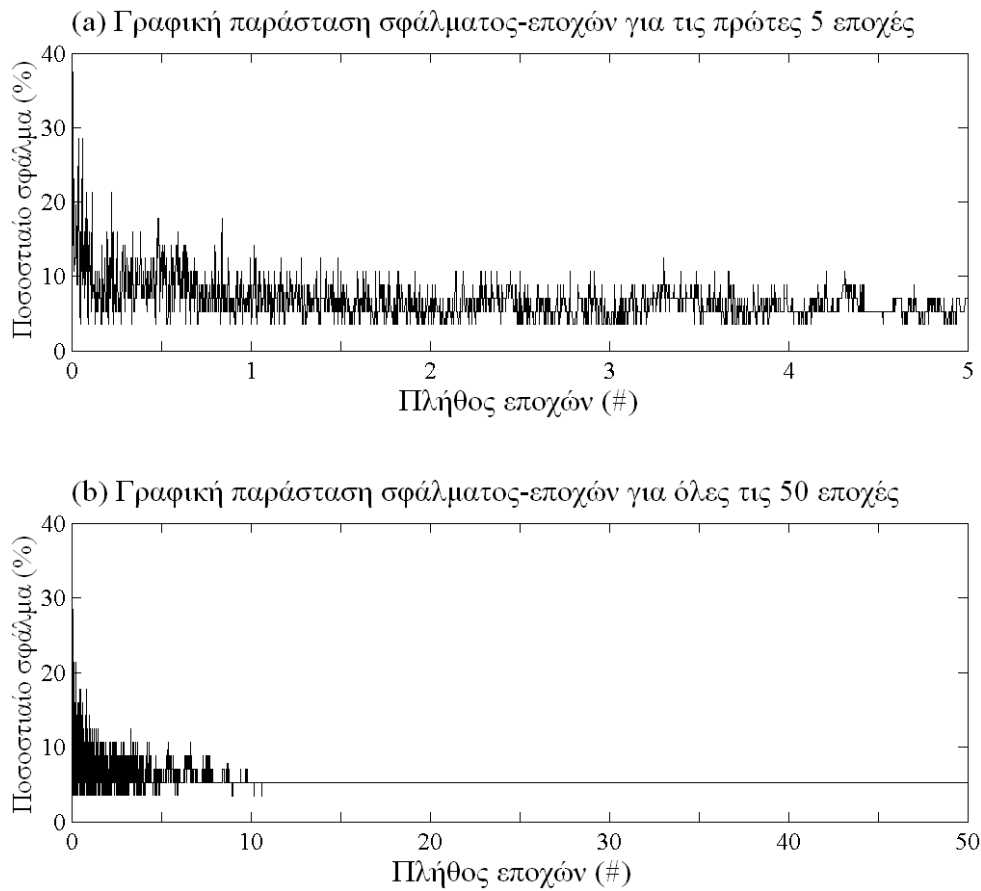
στα όρια των συστάδων. Τα εν λόγω HMMs έχουν την τάση να αντιπροσωπεύουν σφαιρίνες που ανήκουν σε δύο διαφορετικές υποκατηγορίες πρωτεϊνών, εμποδίζοντας έτσι την ακριβή ομαδοποίηση τους σε μία συγκεκριμένη πρωτεϊνική υποκλάση. Πάραυτα, τέτοια φαινόμενα πρέπει να θεωρούνται δικαιολογημένα και αναμενόμενα από τη στιγμή που το SOHMMM παράγει μία ομαλή μη γραμμική απεικόνιση των υποοικογενειών των σφαιρινών σε ένα επίπεδο χαμηλών διαστάσεων (υπό τη μορφή εφαπτόμενων/συνορευόντων συστάδων) και τα υπό εξέταση HMMs συγκεντρώνονται στις περιοχές διαχωρισμού των τριών συστάδων.



Εικόνα 5.2. (a) Ενδεικτικό παράδειγμα μίας συστοιχίας SOHMMM μετά το πέρας της διαδικασίας προσαρμογής που χρησιμοποιεί το πλήρες σύνολο των 560 σφαιρινών. (b) Εναλλακτικό παράδειγμα ενός SOHMMM, μετά την ολοκλήρωση του αλγορίθμου μη επιβλεπόμενης μάθησης, για το περιορισμένο σύνολο εκπαίδευσης των 225 σφαιρινών.

Πειράματα διεξήχθησαν επίσης για να μελετηθεί το με ποιο τρόπο και το σε ποιο βαθμό η συμπεριφορά και οι επιδόσεις του SOHMMM επηρεάζονται από τη διάρκεια της φάσης εκπαίδευσης, από το μέγεθος της συστοιχίας και από την αρχιτεκτονική των HMM νευρώνων (δηλαδή των αριθμό των καταστάσεων τους). Οι συναρτήσεις του ρυθμού μάθησης και της τοπολογικής γειτονιάς ακολούθησαν τα σχήματα ρύθμισης που εφαρμόστηκαν σε όλα τα προηγούμενα πειράματα (υιοθετήθηκε μία τέτοια

προσέγγιση καθώς η απόδοση και η λειτουργικότητα της είναι αρκετά καλή στην πράξη).



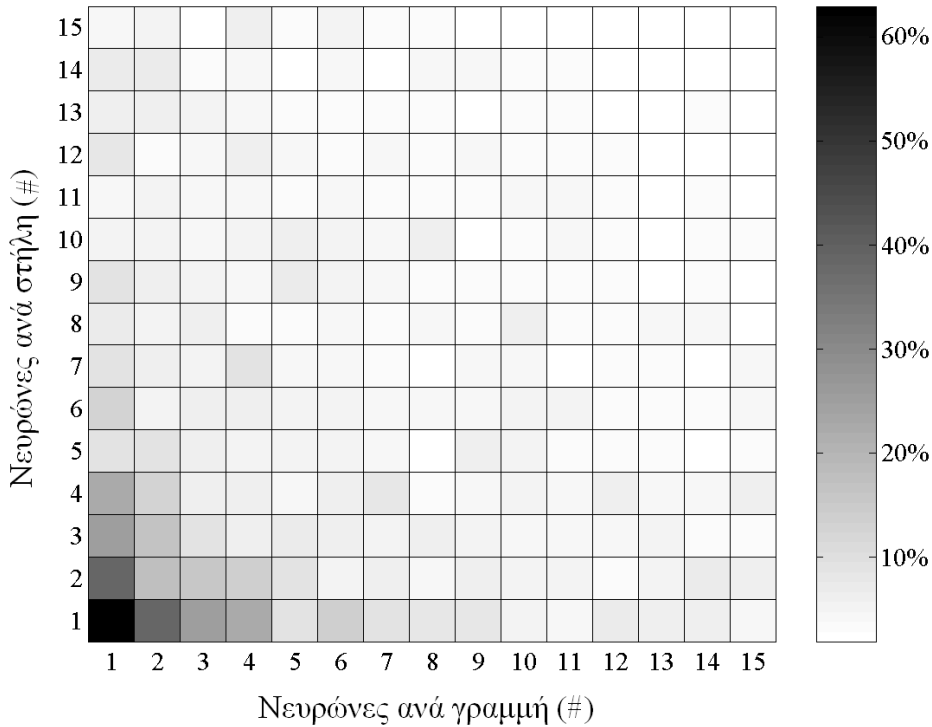
Εικόνα 5.3. Γραφικές παραστάσεις ποσοστιαίου σφάλματος-πλήθους εποχών του SOHMMM, όπως αυτές προκύπτουν για το σύνολο δεδομένων των σφαιρινών. Η γραφική παράσταση (a) αφορά τις πρώτες 5 εποχές εκπαίδευσης από τις συνολικά 50 που χρησιμοποιήθηκαν.

Ένα ενδεικτικό σενάριο μίας διαδικασίας προσαρμογής (ενός 6x8 πλέγματος SOHMMM) που διαρκεί δεκαπλάσιο αριθμό εποχών απ' ότι προηγούμενα φαίνεται στην εικόνα 5.3. Βασιζόμενοι σε αυτό μπορούμε να κάνουμε ορισμένες επισημάνσεις, ξεκινώντας από την εμφανή παρατήρηση ότι το μεγαλύτερο κομμάτι της εκπαίδευσης ολοκληρώνεται στις λίγες πρώτες εποχές, κατόπιν των οποίων το SOHMMM περνάει σε μία φάση σύγκλισης. Αυτά τα ευρήματα είναι σε συμφωνία με τις φάσεις αδρούς και λεπτομερούς προσαρμογής που έχουν ενσωματωθεί στον αλγόριθμο μη επιβλεπόμενης μάθησης του μοντέλου. Επιπλέον, ένας μπορεί να παρατηρήσει ότι η μέση τιμή του σφάλματος μειώνεται πράγματι, γρήγορα στην αρχή και βαθμιαία προς

το τέλος, παρά το γεγονός ότι η προσαρμογή βασίζεται στη βελτιστοποίηση μίας συνάρτησης διάκρισης η οποία δεν απαιτεί, και κατά συνέπεια δεν χρησιμοποιεί, κανενός είδους εκ των προτέρων γνώσης και καμίας μορφής πληροφορίας ανάθεσης κατηγοριών. Αυτή η παρατήρηση μας δίνει τη δυνατότητα να υποστηρίξουμε βάσιμα ότι όσο οι κατηγοριοποιήσεις και οι αναθέσεις κλάσεων βρίσκονται σε αντιστοιχία με τα υποκείμενα χαρακτηριστικά, τις ιδιότητες και τη δομή των δεδομένων, το SOHMMM παράγει σχετικά ακριβή αποτελέσματα. Επιπρόσθετα, το γεγονός ότι με το πέρας των πρώτων πέντε εποχών εκπαίδευσης το SOHMMM έχει ήδη επιτύχει ένα καλό επίπεδο επίδοσης (το οποίο βελτιώνεται ελάχιστα κατά την διάρκεια των εποχών που απομένουν), είναι μία περαιτέρω ένδειξη προς υποστήριξη της πρότασης ότι ακόμη και σύντομες περίοδοι προσαρμογής συχνά αρκούν. Εν μέρει, αυτό οφείλεται στον αποδεδειγμένο αλγόριθμο μη επιβλεπόμενης μάθησης του SOHMMM, ο οποίος παράγει πιθανοτικές μονάδες (τουτέστιν τους HMM νευρώνες) που είναι σε θέση να εκμεταλλεύονται το σύνολο της διαθέσιμης πληροφορίας κατά τρόπο σαφώς πιο αποδοτικό από άλλες (συνήθως ευρετικές) ντετερμινιστικές διανυσματικές μεθοδολογίες. Πέραν τούτου, κατόπιν ορισμένων ταλαντώσεων στην αρχή της διαδικασίας προσαρμογής, το SOHMMM συγκλίνει σε ένα βέλτιστο, επιδεικνύοντας ευστάθεια καθ' όλη τη διαδρομή, με αποτέλεσμα στο τέλος οι διακυμάνσεις πρακτικά να εξαλείφονται, τουλάχιστον για το υπό εξέταση πρόβλημα.

Η γραφική παράσταση που εμφανίζεται στην εικόνα 5.4 παρουσιάζει τα αποτελέσματα μίας πολλαπλής διαδικασίας stratified ten-fold cross-validation όταν αυτή εφαρμόζεται σε μία πλειάδα διαστάσεων της συστοιχίας SOHMMM. Το πιο βασικό συμπέρασμα που απορρέει από τα παρουσιαζόμενα αποτελέσματα έχει να κάνει με το γεγονός ότι υπάρχει ένα ευρύ φάσμα μεγεθών πλεγμάτων το οποίο παράγει ικανοποιητικά χαμηλές τιμές σφάλματος. Αυτό σημαίνει ότι εύλογες αυξομειώσεις στο μέγεθος του πλέγματος έχουν μικρό αντίκτυπο στην επίδοση του SOHMMM. Αυτή η σχετική ανοχή ως προς τη μεταβλητότητα ορισμένων ελεύθερων παραμέτρων (στη συγκεκριμένη περίπτωση του πλήθους των νευρώνων ανά γραμμή και στήλη) είναι μία επιθυμητή ιδιότητα για μία μέθοδο μη επιβλεπόμενης μάθησης μιας και ακριβείς τιμές είναι δύσκολο να καθοριστούν σε συνθήκες απύσας ή άγνωστης πληροφορίας κατηγοριοποίησης. Η μόνη παγίδα που πρέπει να αποφευχθεί είναι η χρησιμοποίηση εκφυλισμένων πλεγμάτων SOHMMM (για παράδειγμα ένα πλέγμα 1x1). Προφανώς, ένας περιορισμένος αριθμός HMMs δεν είναι σε θέση να μοντελοποιήσει τα δεδομένα και να τα ομαδοποιήσει σε συστάδες, με αποτέλεσμα ορισμένες ιδιότητες και

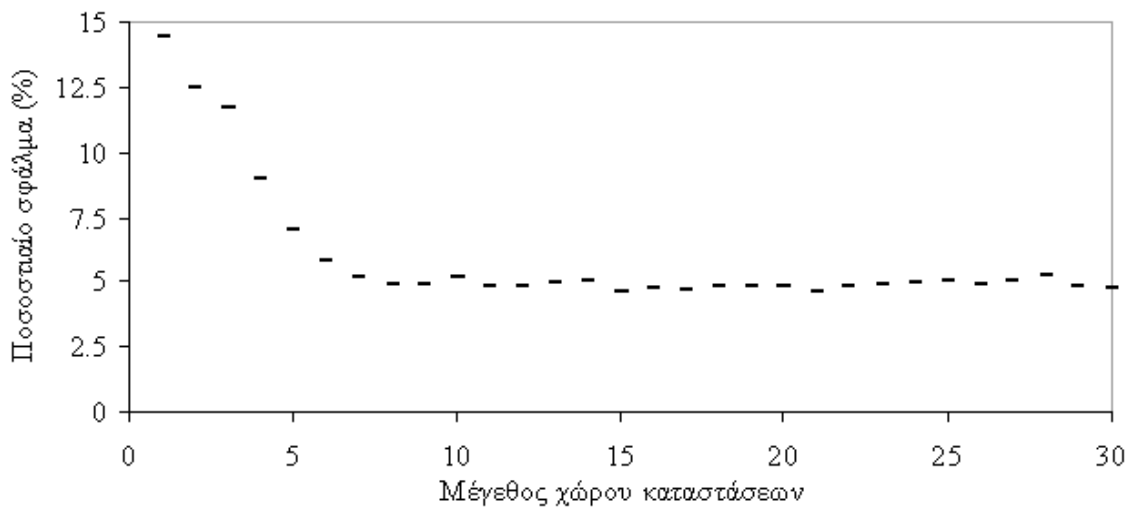
συσχετίσεις να αναπαριστώνται λαθεμένα. Σε τέτοιες περιπτώσεις υπερισχύουν εσφαλμένες καταστάσεις πράγμα που καθιστά την υποβάθμιση των επιδόσεων αναπόφευκτη. Αντίθετα, πρέπει να είναι εμφανές ότι η αύξηση του μεγέθους ενός πλέγματος SOHMMM που ήδη δουλεύει ικανοποιητικά δεν είναι μία έξυπνη κίνηση· οριακά μπορεί να διαπιστωθεί κάποιο κέρδος στην επίδοση, ενώ παράλληλα το υπολογιστικό κόστος αυξάνεται.



Εικόνα 5.4. Μέσες τιμές ποσοστιαίων σφαλμάτων που παράγονται από πλέγματα SOHMMM διαφόρων διαστάσεων.

Η γραφική παράσταση που εμφανίζεται στην εικόνα 5.5 παρουσιάζει τα αποτελέσματα μίας σειράς διαδικασιών stratified ten-fold cross-validation για διαφορετικά μεγέθη του χώρου καταστάσεων των HMM νευρώνων (μίας 6x8 συστοιχίας SOHMMM). Θεωρούμε ότι μπορούν να γίνουν ορισμένες παρατηρήσεις και να εξαχθούν κάποια συμπεράσματα που πηγάζουν από τα παρουσιαζόμενα αποτελέσματα. Πρώτον, για κρυφές καταστάσεις περισσότερες από ένα (σχετικά) μικρό αριθμό, περίπου 10, ο αλγόριθμος επιτυγχάνει χαμηλές τιμές σφάλματος. Πέραν αυτού του κατωφλίου οι επιδόσεις του SOHMMM δεν παρουσιάζουν κάποια εμφανή βελτίωση και παραμένουν σχεδόν σταθερές για όλα τα επόμενα μεγέθη των χώρων καταστάσεων. Ισοδύναμα μπορούμε να υποστηρίξουμε ότι τουλάχιστον για τη

συγκεκριμένη πειραματική διάταξη (και ειδικότερα για το συγκεκριμένο σύνολο δεδομένων) ένα μεγάλο, κάτω φραγμένο, εύρος καταστάσεων αποδίδει ικανοποιητικά. Οι χώροι καταστάσεων με μικρά μεγέθη εμφανίζουν προβληματική συμπεριφορά κάτι που είναι εύλογο εάν λάβει κανείς υπόψη ότι ένα περιορισμένο πλήθος κατανομών δεν επαρκεί για την ακριβή και πλήρη περιγραφή σύνθετων και πολύπλοκων χώρων ακολουθιών, όπως αυτούς των σφαιρινών. Η ομαδοποίηση και η προκύπτουσα χαρτογράφηση σε αυτές τις περιπτώσεις είναι μάλλον αδρές και προσεγγιστικές.



Εικόνα 5.5. Γραφική παράσταση ποσοστιαίου σφάλματος-μέγους χώρου κρυφών καταστάσεων των HMM νευρώνων ενός 6x8 SOHMMM, η οποία αντιστοιχεί στο σύνολο δεδομένων των σφαιρινών.

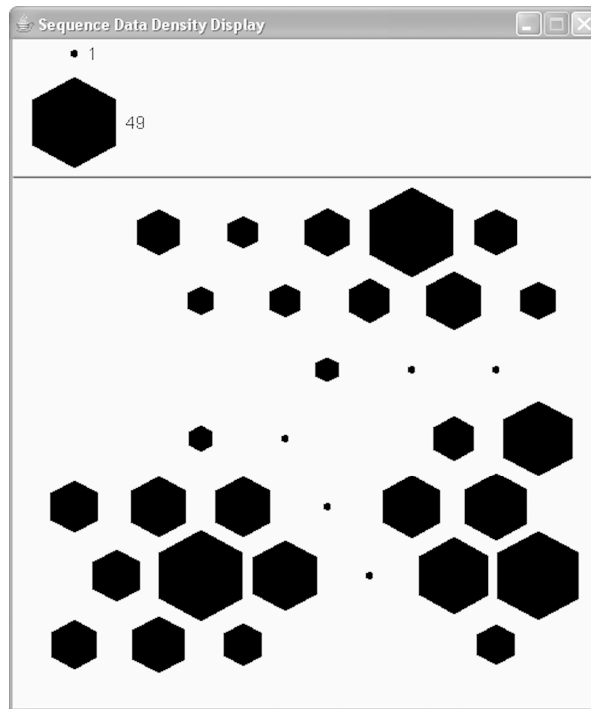
Συνολικότερα, ένα συμπέρασμα βαρύνουσας σημασίας έχει να κάνει με το γεγονός ότι ακόμη και στην περίπτωση που η πληροφορία ανάθεσης κατηγοριών είναι διαθέσιμη και αξιοποιήσιμη, έστω για εκ των υστέρων αξιολογήσεις και μετρήσεις, δεν φαίνεται να επαρκεί και να μπορεί να καθοδηγήσει μία διαδικασία εξαγωγής και ρύθμισης της αρχιτεκτονικής των HMM νευρώνων. Καθώς αυξάνει το πλήθος των καταστάσεων N ο αριθμός των ελεύθερων παραμέτρων αυξάνει με αριθμητική πρόοδο ενώ η πολυπλοκότητα παρουσιάζει τετραγωνική εξάρτηση ως προς το N , έτσι εάν και τριπλασιάζεται το μέγεθος του χώρου καταστάσεων, από 10 σε 30, πάραυτα δεν εντοπίζεται κάποια πρακτική βελτίωση στην αποτελεσματικότητα και τις επιδόσεις του SOHMMM. Ουσιαστικά, παρότι διερευνούμε λεπτομερώς μία μεγάλη γκάμα πιθανών διαστάσεων του χώρου καταστάσεων μία, έστω τοπικά, βέλτιστη επιλογή δεν φαίνεται να υπάρχει. Επομένως, γίνεται κατανοητό ότι η αναζήτηση και η εύρεση (με την επιστράτευση οποιασδήποτε προσέγγισης [70], [71], [62]) κάτι εν τέλει

ανύπαρκτου είναι ανεδαφικές, τουλάχιστον στα στενά πλαίσια του SOHMMM. Αυτό αποσαφηνίζεται εάν λάβουμε απλά υπόψη ότι τόσο ο αριθμός των κατηγοριών όσο και η αντιστοίχιση αλληλουχιών σε κατηγορίες είναι δύο πράγματα που θεωρούνται μη δεδομένα και σε πολλές πρακτικές εφαρμογές είναι κυριολεκτικά άγνωστα. Συνεπώς ακόμη και η κατασκευή αναλόγων γραφικών παραστάσεων είναι ανέφικτη, πολύ περισσότερο δε η αξιοποίηση τους για την κατάρτιση της αρχιτεκτονικής των HMM νευρώνων.

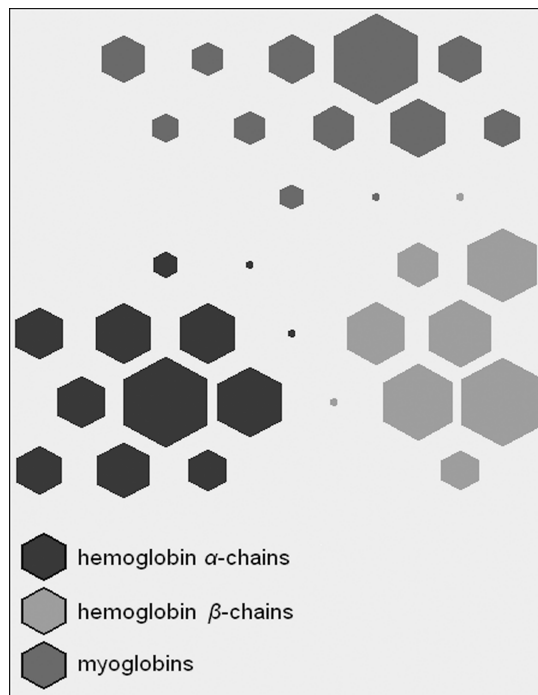
Θεωρούμε ένα εξαγωνικό πλέγμα 6x7 αποτελούμενο από 42 νευρώνες SOHMMM το οποίο έχει εκπαιδευτεί με βάση το πλήρες σύνολο δεδομένων σφαιρινών. Η SDDD που αντιστοιχεί στο εν λόγω SOHMMM φαίνεται στην εικόνα 5.6. Μία εξέταση αυτής της γραφικής αναπαράστασης αποκαλύπτει ότι οι δύο βασικοί στόχοι της SDDD επιτυγχάνονται σε ικανοποιητικό βαθμό. Πρώτον, το πλήθος αλληλουχιών που ανατίθεται σε κάθε νευρώνα, δηλαδή η πυκνότητα τους, εντοπίζεται οπτικά. Δεύτερον, τρεις συστάδες σχηματίζονται στην SDDD, αυτές αποτελούνται από περιοχές κόμβων μεγάλου μεγέθους οι οποίες διαχωρίζονται από διάκενα αποτελούμενα από κόμβους μικρού ή και μηδενικού μεγέθους. Επιπροσθέτως, σχετικά απλά μπορεί να επαληθευθεί ότι οι αυτές οι τρεις συστάδες βρίσκονται σε πλήρη αντιστοιχία με τις τρεις υποοικογένειες σφαιρινών. Πληροφορία ανάθεσης κατηγοριών η οποία έχει εξαιρεθεί από τις διαδικασίες προσαρμογής και οπτικοποίησης χρησιμοποιείται για την εκ των υστέρων αναγνώριση/κατηγοριοποίηση καθενός κόμβου SOHMMM. Το αποτέλεσμα αυτής της διαδικασίας παρουσιάζεται στην εικόνα 5.7, η ένα-προς-ένα αντιστοιχία μεταξύ κάθε απεικονιζόμενης συστάδας και της υποοικογένειας σφαιρινών που περιγράφει είναι εμφανής.

Η γλώσσα Java έχει χρησιμοποιηθεί για τον προγραμματισμό αμφοτέρων των SDDD και SLP γραφικών απεικονίσεων. Ο βασικός λόγος για αυτήν την επιλογή ήταν η υποστήριξη γραφικών η οποία είναι ανεξάρτητη της πλατφόρμας υλοποίησης. Η μεγάλη πλειοψηφία των παρουσιαζόμενων εικόνων αποτελείται από στιγμιότυπα οθόνης της γραφικής διεπαφής που υλοποιήθηκε. Η SDDD διεπαφή, εκτός της κλίμακας που χρησιμοποιείται σε κάθε περίπτωση και η οποία εμφανίζεται πάνω αριστερά, έχει επιπλέον χαρακτηριστικά και λειτουργίες. Εμφανίζεται το αριθμητικό πλήθος των αλληλουχιών που ανατίθενται σε κάθε νευρώνα ενώ παράλληλα παρέχεται η δυνατότητα ανάκτησης ή/και αποθήκευσης: (1) των αντίστοιχων ακολουθιών από αμινοξέα ή νουκλεοτίδια· (2) των αναγνωριστικών αυτών των αλληλουχιών· (3) των παραμέτρων του αντίστοιχου HMM αναφοράς. Όπως έχει ήδη αναφερθεί αυτή η

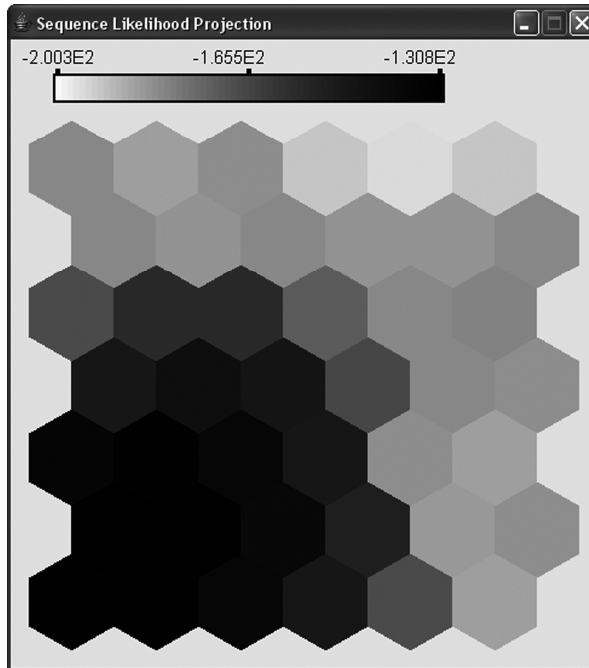
πληροφορία δύναται να χρησιμοποιηθεί για τη μελέτη των χαρακτηριστικών και των στατιστικών ιδιοτήτων των εκάστοτε περιγραφόμενων αλληλουχιών.



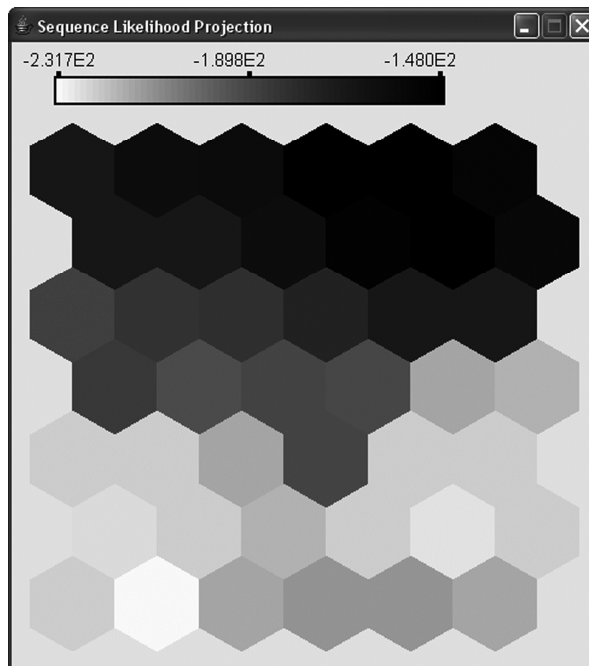
Εικόνα 5.6. SDDD βασιζόμενο στην πλήρη πρωτεϊνική οικογένεια των σφαιρινών. Κάθε εξάγωνο, ανεξαρτήτως μεγέθους, αναπαριστά έναν ξεχωριστό νευρώνα SOHMMM.



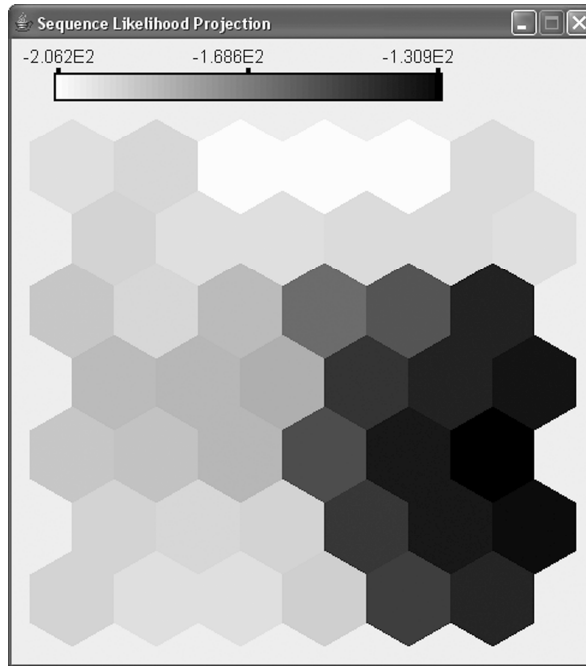
Εικόνα 5.7. Η υλοποιηθείσα SDDD απεικόνιση εμπλουτισμένη με πληροφορία ανάθεσης κατηγοριών προερχόμενης από εκ των υστέρων χαρακτηρισμό των συστάδων.



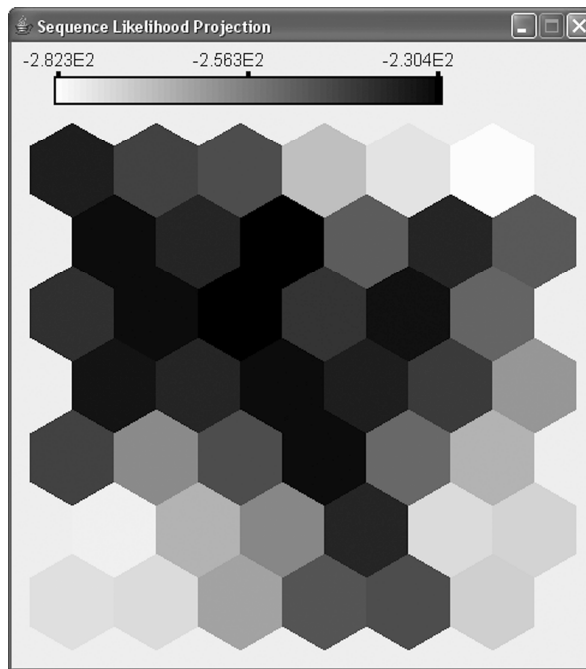
Εικόνα 5.8. Χαρακτηριστική SLP απεικόνιση που αντιστοιχεί σε μία α -σφαιρίνη. Κάθε εξάγωνο, ανεξαρτήτως απόχρωσης, αντιστοιχεί σε έναν ξεχωριστό κόμβο SOHMMM.



Εικόνα 5.9. Ενδεικτική SLP βασισμένη σε μυσφαιρίνη.



Εικόνα 5.10. Παράδειγμα μίας SLP γραφικής αναπαράστασης που αντιστοιχεί σε μία β -σφαιρίνη.



Εικόνα 5.11. Χαρακτηριστικό αποτέλεσμα μίας SLP που αφορά σε μία πρωτεϊνική αλυσίδα κυτοσφαιρίνης.

Η παρούσα πειραματική μελέτη της SLP κάνει χρήση του προηγούμενου SOHMMM το οποίο έχει εκπαιδευτεί με βάση το πλήρες σύνολο δεδομένων σφαιρινών. Οι προκύπτουσες SLP γραφικές αναπαραστάσεις που αφορούν σε τρεις αλληλουχίες, οι οποίες ανήκουν στις τρεις διαφορετικές υποκατηγορίες σφαιρινών, φαίνονται στις εικόνες 5.8-5.10. Οι αρχικά καθορισμένοι στόχοι επιτυγχάνονται σε

ικανοποιητικό βαθμό. Η κατανομή των τιμών πιθανοφάνειας (ή αλλιώς το τοπίο πιθανοφάνειας) κάθε αλληλουχίας αναπαριστάται οπτικά. Σημαντικότερο είναι ίσως το γεγονός ότι ακριβώς μία συστάδα σχηματίζεται σε καθεμία SLP. Η συστάδα αυτή αποτελείται από συνεχείς/συνεκτικές περιοχές νευρώνων υψηλής πιθανοφάνειας περιτριγυρισμένες από περιοχές νευρώνων με χαμηλές πιθανοφάνειες. Επιπροσθέτως, λαμβάνοντας υπόψη ότι η πρώτη SLP (εικόνα 5.8) περιγράφει μία α -σφαιρίνη, η δεύτερη SLP (εικόνα 5.9) σχετίζεται με μία μυοσφαιρίνη και η τρίτη SLP (εικόνα 5.10) αφορά μία β -αλυσίδα σφαιρίνης, υπάρχει μία προφανής αναλογία/αντιστοιχία με τις συστάδες που εντοπίζονται στην SDDD γραφική απεικόνιση (εικόνα 5.6).

Αντιθέτως, μία κυτοσφαιρίνη (*Xenopus Laevis*) η οποία δεν χρησιμοποιήθηκε κατά τη διαδικασία εκπαίδευσης του SOHMMM, τουτέστιν δεν αναπαριστάται ρητά στο επίπεδο SOHMMM, φαίνεται στην εικόνα 5.11. Σε αυτήν την περίπτωση αντί να παρατηρούνται συνεχείς περιοχές με καλώς ορισμένα σύνορα εμφανίζονται κατατμήσεις που οπτικά παραπέμπουν σε κάποια μορφή ακανόνιστο ψηφιδωτό. Η αιτιολόγηση αυτού του φαινομένου βασίζεται στο ότι το συγκεκριμένο SOHMMM δεν έχει προσαρμοστεί/ρυθμιστεί κατάλληλα έτσι ώστε να μοντελοποιεί την υποοικογένεια των κυτοσφαιρινών ή έτσι ώστε να ομαδοποιεί τις αντίστοιχες βιολογικές αλυσίδες. Ωστόσο, ακόμη και στην προκειμένη περίπτωση, οι υψηλότερες τιμές πιθανοφάνειας εντοπίζονται σε περιοχές που δεν περιγράφουν καμία εκ των τριών συστάδων από πρωτεΐνες (εικόνα 5.6). Το γεγονός αυτό αποτελεί μία έμμεση απόδειξη του ότι το SOHMMM παράγει μία μη γραμμική, διατεταγμένη χαρτογράφηση ακολουθιών. Παρότι, η κυτοσφαιρίνη εισάγεται στο σύστημα για πρώτη φορά δεν ανατίθεται σε κάποιο νευρώνα που περιγράφει μία εκ των τριών συστάδων που έχουν εντοπιστεί, πράγμα που θα ήταν εσφαλμένο, αλλά αντιθέτως, αντιστοιχίζεται σε ένα κόμβο SOHMMM που δεν αποτελεί τμήμα καμίας συστάδας.

Όπως αναφέρθηκε και προηγούμενα η γλώσσα Java έχει χρησιμοποιηθεί για τον προγραμματισμό της SLP γραφικής απεικόνισης καθώς και της αντίστοιχης διεπαφής. Ομοίως, το σύνολο των παρουσιαζόμενων εικόνων αποτελείται από στιγμιότυπα οθόνης της γραφικής διεπαφής που υλοποιήθηκε. Η SLP διεπαφή περιλαμβάνει χρωματική κλίμακα, η οποία βρίσκεται σε αναλογία με τη μέγιστη και την ελάχιστη λογαριθμική τιμή των πιθανοφανειών που εμφανίζονται. Επίσης παρέχει τη δυνατότητα ανάκτησης του αναγνωριστικού αλλά και της ακολουθίας αμινοξέων/νουκλεοτιδίων της υπό εξέταση αλληλουχίας. Επιπλέον, εμφανίζεται η ακριβής τιμή της πιθανοφάνειας που εξάγει καθένας κόμβος SOHMMM. Τέλος,

παρέχεται η δυνατότητα επισκόπησης ή/και αποθήκευσης των παραμέτρων του HMM αναφοράς που επιλέγεται.

5.2.2. Περιορισμένος Αριθμός Δεδομένων Εκπαίδευσης

Η δεύτερη πειραματική διάταξη περιλαμβάνει ένα SOHMMM πλήρως ταυτόσημο με αυτό που χρησιμοποιήθηκε στην προηγούμενη σειρά πειραμάτων (τουτέστιν μία 6x8 συστοιχία SOHMMM). Το σύνολο εκπαίδευσης απαρτίζεται από 75 ακολουθίες πρωτεϊνών τυχαία επιλεγμένες από καθεμία εκ των τριών αρχικών υποοικογενειών σφαιρινών, συνολικά επομένως το μέγεθος του συνόλου εκπαίδευσης ισούται με 225 αλληλουχίες σφαιρινών. Οι 335 πρωτεΐνες που περισσεύουν αποθηκεύονται έτσι ώστε να μελετηθεί και να ελεγχθεί το SOHMMM με δεδομένα που δεν έχουν χρησιμοποιηθεί κατά τη διάρκεια της εκπαίδευσης. Τα αποτελέσματα κατόπιν της ολοκλήρωσης του στοχαστικού αλγορίθμου μάθησης του SOHMMM παρουσιάζονται στην εικόνα 5.2(b). Τα ευρήματα και οι παρατηρήσεις αναφορικά με τις σχηματισθείσες συστάδες είναι πανομοιότυπα με αυτά της πρώτης πειραματικής διάταξης. Τα ποιοτικά χαρακτηριστικά και οι βασικές ιδιότητες των δύο SOHMMMs συμπίπτουν, επουσιώδεις διαφορές που αφορούν στη θέση και τη μορφή των αντίστοιχων συστάδων οφείλονται κατά κύριο λόγο στις τυχαίες αρχικοποιήσεις και στη διαφορετική σειρά παρουσίασης των δεδομένων αλληλουχιών στις εισόδους των SOHMMMs. Μία βάσιμη ποιοτική σύγκριση της πρώτης και της δεύτερης σειράς πειραμάτων πρέπει να λάβει υπόψη την απόκριση του μοντέλου στο σύνολο των 560 πρωτεϊνικών αλληλουχιών (αμφότερων αυτών που χρησιμοποιήθηκαν κατά την εκπαίδευση και αυτών που δεν αξιοποιήθηκαν). Σε αυτήν την αντιπροσωπευτική περίπτωση 183 α -σφαιρίνες, 204 β -σφαιρίνες και 141 μυοσφαιρίνες ανατίθενται στις συστάδες των αντίστοιχων πρωτεϊνικών υποκλάσεων. Η μέση τιμή σωστά ομαδοποιημένων σφαιρινών ισούται με $94.47 \pm 0.59\%$ (υπολογιζόμενη επί 100 επαναλήψεων του αλγορίθμου μη επιβλεπόμενης μάθησης). Αυτά τα αποτελέσματα είναι υποδεέστερα, αφού μειώνεται οριακά το ποσοστό σωστά ομαδοποιημένων πρωτεϊνών, ενώ παράλληλα, αυξάνεται το ποσοστό σφαιρινών που ανατίθενται σε HMMs που περιγράφουν πρωτεΐνες που ανήκουν σε δύο διαφορετικές υποκλάσεις. Το αποτέλεσμα αυτό είναι δικαιολογημένο από τη στιγμή που μικρότερος αριθμός αλληλουχιών (λιγότερες από τις μισές διαθέσιμες) προφανώς περιέχει λιγότερη

πληροφορία, συνεπώς, το προκύπτον SOHMMM αντικατοπτρίζει τη συρρικνωμένη και τμηματική φύση τις παρεχόμενης πληροφορίας.

Αυτό το σύνολο πειραμάτων, που βασίζεται στην πρωτεϊνική οικογένεια των σφαιρινών, μας επιτρέπει να εξάγουμε ορισμένα συμπεράσματα αναφορικά με τις ιδιότητες και τις δυνατότητες του SOHMMM καθώς και να επιβεβαιώσουμε ορισμένα από τα προτερήματα του. Εξαιτίας του γεγονότος ότι οι ψηφιακές ακολουθίες παρατηρήσιμων συμβόλων είναι πιστές περιγραφές πρωτεϊνικών αλληλουχιών, επιδεικνύουν το ίδιο εύρος μηκών όπως τα βιολογικά ανάλογα τους. Είναι φανερό ότι το SOHMMM ενσωματώνει ένα μηχανισμό για την επεξεργασία ακολουθιών απαρτιζόμενων από διακριτά σύμβολα προερχόμενα από αλφάβητα αυθαίρετου μεγέθους, όπως το χρησιμοποιηθέν αλφάβητο των είκοσι αμινοξέων. Επίσης, είναι σε θέση να χειρίζεται αποδοτικά τέτοιες αλληλουχίες υψηλής διάστασης και διαφορετικών μηκών. Αξιοποιώντας συγκεκριμένα χαρακτηριστικά των HMMs πετυχαίνει να χειρίζεται τα μη αριθμητικά σύμβολα (αμινοξέα) κάθε βιολογικής αλληλουχίας απευθείας, χωρίς την παρεμβολή καμίας διαδικασίας προεπεξεργασίας ή μετασχηματισμού των συμβολικών ποσοτήτων σε αριθμητικές. Το SOHMMM μπορεί να συλλαμβάνει και να αξιοποιεί τη λανθάνουσα πληροφορία που βρίσκεται κρυμμένη στις χρονικές/χωρικές συσχετίσεις και στην γραμμική ακολουθιακή φύση βιολογικών μορίων αλληλουχιών. Οι ακριβείς τιμές των παραμέτρων εκπαίδευσης, παρότι προσεγγιστικές και πρόχειρες, δεν φαίνεται να επηρεάζουν την αποδοτικότητα του SOHMMM. Όπως φάνηκε προηγούμενα, χονδρικές εκτιμήσεις αυτών των παραμέτρων (π.χ. του μεγέθους της συστοιχίας SOHMMM, του ρυθμού μάθησης και της συνάρτησης τοπολογικής γειτονιάς) αποδεικνύονται περισσότερο από επαρκείς. Επίσης, ακόμη και ένα μικρό πλήθος κύκλων προσαρμογής (πέντε εποχές κατά μέγιστο) είναι συχνά αρκετό για την εξασφάλιση ευστάθειας και σύγκλισης.

Ακολουθώντας πιστά τις στρατηγικές της αυτο-οργάνωσης και του ανταγωνισμού, το SOHMMM επιτυγχάνει την επινόηση ανώτερων αφαιρέσεων (συμβολισμών) οι οποίες εδράζονται στις στατιστικές ιδιότητες τοπολογικά γειτνιαζόντων HMMs. Η SOHMMM προσέγγιση είναι σε θέση να εντοπίζει και να επεξεργάζεται τις σημαντικές στατιστικές ιδιότητες της κλάσης των σφαιρινών, αυτό το επιτυγχάνει καλύπτοντας ένα διευρυμένο σύνολο κατανομών και αρά εκφράζοντας σχέσεις που δεν είναι προσβάσιμες από μεμονωμένα ασυσχέτιστα HMMs. Έτσι, το SOHMMM κατάφερε να ανιχνεύσει την ύπαρξη των τριών πρωτεϊνικών υποκλάσεων και να ομαδοποιήσει όλες τις βιολογικές αλληλουχίες στις τρεις συστάδες που αντιπροσωπεύουν αυτές τις

υποκατηγορίες. Είναι σημαντικό να επισημανθεί ότι αυτό επετεύχθη μέσω μίας διαδικασίας προσαρμογής απευθείας από πρωτεϊνικές αλληλουχίες, με χρήση μηδενικής εκ των προτέρων γνώσης και αξιοποιώντας τη σημαντική πληροφορία που ενυπάρχει στις χωρικές/χρονικές εξαρτήσεις των αμινοξέων κάθε σφαιρίνης. Ακόμη και στην περίπτωση του συρρικνωμένου συνόλου εκπαίδευσης (της δεύτερης πειραματικής διάταξης) τα αποτελέσματα ήταν μόνο οριακά υποδεέστερα. Το SOHMMM πέτυχε ένα επίπεδο γενίκευσης τέτοιο που εξάλειψε φαινόμενα υπερεκπαίδευσης, και ταυτόχρονα, ομαδοποίησε με ακρίβεια τη μεγάλη πλειοψηφία των σφαιρινών, τόσο αυτών που επιστρατεύθηκαν κατά την εκπαίδευση όσο και αυτών που χρησιμοποιήθηκαν για τον έλεγχο. Επιπρόσθετα, το SOHMMM εμφανίζει μία σχετική ανοχή απέναντι σε φαινόμενα περιορισμένου πλήθους προτύπων εισόδου (δηλαδή σε καταστάσεις μικρού αριθμού διαθέσιμων δεδομένων εκπαίδευσης), κάτι το οποίο γίνεται εμφανές από το γεγονός ότι λιγότερες από τις μισές διαθέσιμες σφαιρίνες αποδείχθηκαν αρκετές για την προσαρμογή του SOHMMM. Ωστόσο, πρέπει να σημειωθεί ότι ο ελάχιστος απαιτούμενος αριθμός δεδομένων εκπαίδευσης εξαρτάται σε μεγάλο βαθμό από το υπό εξέταση πρόβλημα μιας και καθορίζεται από την πολυπλοκότητα και τη φύση του ίδιου του προβλήματος όπως επίσης και από το πόσο περιγραφικά είναι τα πρότυπα εισόδου του συνολικού χώρου δεδομένων αλληλουχιών.

Ένας από τους βασικούς σκοπούς του SOHMMM είναι η εύρεση ή η δημιουργία περιεκτικών περιγραφών και αναπαραστάσεων για σύνολα ακολουθιών παρατηρήσεων. Όπως έχει ήδη δείχθει απεικονίζει υποκλάσεις σφαιρινών (οι οποίες είναι πολύπλοκα μόρια υψηλή διάστασης) ως σημεία ενός δισδιάστατου επιπέδου. Αυτές οι απεικονίσεις αναπαριστούν τις δεδομένες αλληλουχίες σε ένα χώρο χαμηλής διάστασης με τέτοιο τρόπο ώστε να διατηρούνται οι συστάδες και οι συσχετίσεις των πρωτεϊνών κατά τον καλύτερο δυνατό τρόπο. Επιπροσθέτως, τα HMMs που συνιστούν τις εκάστοτε συστάδες, εξελίσσονται σε αποκωδικοποιητές των αντίστοιχων χώρων πρωτεϊνών, και επίσης, αυτοί οι αποκωδικοποιητές είναι οργανωμένοι στο πλέγμα SOHMMM με συστηματικό τρόπο. Οι σφαιρίνες που ανατίθενται σε ένα νευρώνα SOHMMM ουσιαστικά περιγράφονται από ένα πιθανοτικό μοντέλο σύμφωνα με μία διαδικασία που παραπέμπει σε μείωση της διαστατικότητας. Τέλος, όταν έχει παρουσιαστεί ένας ικανοποιητικός αριθμός πρωτεϊνικών ακολουθιών και τα HMMs του πλέγματος, με την εφαρμογή του αλγορίθμου μάθησης SOHMMM, έχουν συγκλίνει σε πρακτικά στάσιμες τιμές, όλοι οι νευρώνες δύναται να χαρακτηριστούν αναλόγως των υποκλάσεων ή αναλόγως της πληροφορίας ανάθεσης κατηγοριοποίησης

των αντίστοιχων σφαιρινών, ακολουθώντας μία διαδικασία που θυμίζει χαρακτηρισμό περιοχών του χάρτη όπως αυτή του σχήματος 4.2(b). Από τη στιγμή που μπορεί να υπολογιστεί ο νικητής νευρώνας μίας άγνωστης αλληλουχίας από αμινοξέα, η εν λόγω πρωτεΐνη μπορεί να χαρακτηριστεί από τη συστάδα (άρα και την υποοικογένεια) που αντιπροσωπεύεται από αυτόν τον HMM νευρώνα. Διεργασίες βασιζόμενες σε αυτή τη στρατηγική δύναται να χρησιμοποιηθούν στην αναζήτηση, στο διαχωρισμό και στην κατηγοριοποίηση ορφανών ή άγνωστων πρωτεϊνών. Εφόσον υπάρχουν οικογένειες πρωτεϊνών με ανάθεση χαρακτηρισμού κατηγορίας μπορούμε να κατασκευάσουμε SOHMMs που να τις μοντελοποιούν, και στη συνέχεια, να τα χρησιμοποιούμε για αναζητήσεις (βάσει του κριτηρίου της ομοιότητας) σε βάσεις πρωτεϊνών. Μεταξύ άλλων, η δεύτερη σειρά πειραμάτων επιδεικνύει μία χαρακτηριστική διαδικασία κατηγοριοποίησης βάσει της οποίας καθεμία εκ των 335 σφαιρινών που εξαιρέθηκαν από τη φάση της εκπαίδευσης χαρακτηρίστηκαν από τη συστάδα (ή ισοδύναμα την υποοικογένεια) που αναφέρεται στον HMM νευρώνα με τη μεγαλύτερη τιμή πιθανοφάνειας.

5.2.3. Συγκριτικά Αποτελέσματα

Στην [62] περιγράφεται μία προσπάθεια για την αυτόματη ανακάλυψη υποοικογενειών από σφαιρίνες με τη χρήση μίας προσέγγισης ανταγωνιστικής μάθησης. Η προτεινόμενη μεθοδολογία παράγει ένα σύνθετο HMM το οποίο απαρτίζεται από επιμέρους HMMs. Εν τέλει, κάθε ξεχωριστή συστάδα πρωτεϊνών αναπαριστάται από ένα συγκεκριμένο επιμέρους HMM. Υλοποιούνται επτά μη κενές συστάδες που περιλαμβάνουν πρωτεϊνικές ακολουθίες προερχόμενες από γνωστές υποκλάσεις σφαιρινών. Τέσσερις εξ αυτών των συστάδων περιέχουν διάφορες αλληλουχίες που ανήκουν σε διαφορετικές υποοικογένειες σφαιρινών ή σε συγκεκριμένους οργανισμούς. Όλες ανεξαιρέτως οι α -σφαιρίνες, β -σφαιρίνες και μυοσφαιρίνες κατανεμήθηκαν στις τρεις (μεγαλύτερες) συστάδες που απέμειναν. Η πρώτη συστάδα περιέχει σχεδόν αποκλειστικά α , α -type και α -like σφαιρίνες. Σχεδόν όλες οι β , β -type και β -like σφαιρίνες περιέχονται στη δεύτερη συστάδα. Τέλος, η υποκλάση των μυοσφαιρινών είχε ανατεθεί στην τρίτη συστάδα. Κατ' ουσίαν, αυτά τα αποτελέσματα είναι σε συμφωνία με τα ευρήματα των δύο σειρών πειραμάτων που διεξήχθησαν προηγούμενα. Τέτοιες υβριδικές προσεγγίσεις μεταξύ μοντέλων μη

επιβλεπόμενης μάθησης και HMMs είναι σε θέση να διακρίνουν, και ακολούθως, να ομαδοποιούν σωστά τις κύριες υποκατηγορίες σφαιρινών.

Η τρίτη σειρά πειραμάτων περιλαμβάνει ένα 6x4 πλέγμα SOHMMM, όλες οι άλλες παράμετροι (τόσο του δικτύου όσο και της διαδικασίας εκπαίδευσης) είναι ταυτόσημες με αυτές που χρησιμοποιήθηκαν στις προηγούμενες δύο σειρές πειραμάτων που βασίστηκαν πάνω στις σφαιρίνες. Στις [76], [77], [50] 226 πρωτεϊνικές αλληλουχίες (οι οποίες ανήκουν σε πέντε διαφορετικές υποοικογένειες σφαιρινών) χρησιμοποιούνται ως σημείο αναφοράς μετρήσεων για την αξιολόγηση των αποδόσεων διαφόρων NG και SOM αλγορίθμων. Τα αποτελέσματα που εμφανίζονται στον πίνακα 5.1 είναι μέσες τιμές, υπολογιζόμενες επί 100 επαναλήψεων, που έχουν ληφθεί με χρήση της μεθοδολογίας stratified ten-fold cross-validation. Όλα τα μοντέλα που αναφέρονται (εκτός του SOHMMM) είναι ειδικά σχεδιασμένα για να χειρίζονται σχεσιακά δεδομένα, όπως για παράδειγμα είναι οι (αν)ομοιότητες ζευγών που εξάγονται από ευθυγραμμίσεις και ταιριάσματα βιολογικών ακολουθιών.

Μία εξέταση των αποτελεσμάτων αποκαλύπτει ότι το SOHMMM υπερσχύει με όρους απόδοσης, όχι τόσο επειδή πετυχαίνει την υψηλότερη ακρίβεια (ομολογουμένως οριακά), αλλά κυρίως επειδή το επιτυγχάνει αυτό μέσω μίας ευθείας μεθοδολογίας μη επιβλεπόμενης μάθησης. Προς υποστήριξη των λεγόμενων μας αρκεί να αναφέρουμε ότι οι υπόλοιποι αλγόριθμοι που πετυχαίνουν τιμές μεγαλύτερες από 91% κάνουν χρήση πληροφορίας κατηγοριών καθώς επίσης και προκαθορισμένων εξελικτικών αποστάσεων, γεγονός που τις καθιστά ισχυρά εξαρτώμενες από εκ των προτέρων γνώση. Την ίδια στιγμή, το SOHMMM υπερτερεί όλων των προσεγγίσεων χωρίς να μετέρχεται κανένα είδος πληροφορίας κατηγοριών ή πολύπλοκων αποστάσεων (οι οποίες πρέπει να προκαθοριστούν από εμπειρογνώμονες). Η αποδοτικότητα του SOHMMM στηρίζεται κυρίως στην ικανότητα που έχει να χειρίζεται ακολουθίες με συστηματικό τρόπο βάσει ενός αποδεδειγμένου στατιστικού πλαισίου. Το SOHMMM καταφέρνει να συλλαμβάνει και να αποκρυπτογραφεί την πληροφορία που ενυπάρχει στη σύσταση των πρωτεϊνών, καθώς επίσης, και την πληροφορία που είναι κατανομημένη κατά μήκος των χωρικών συσχετίσεων των πρωτεϊνικών μονομερών. Αντιθέτως, όλες οι άλλες προσεγγίσεις που αναφέρονται για το χειρισμό μη αριθμητικών ακολουθιών, καταφεύγουν στη λύση των προϋπολογισμένων αποστάσεων ζευγών και των γενικευμένων ενδιάμεσων, πράγμα που αναπόφευκτα οδηγεί σε απώλεια πληροφορίας.

ΠΙΝΑΚΑΣ 5.1
ΜΕΣΕΣ ΤΙΜΕΣ ΠΟΣΟΣΤΙΑΙΩΝ ΑΚΡΙΒΕΙΩΝ

Model	Overall
SOHMMM	94.93±0.62
Supervised Relational Batch NG	94.4±1.3
Relational Batch NG	91.9±1.6
Supervised Relational SOM	91.5
Supervised Relational Hyperbolic SOM	91.5
Deterministic Annealing	90.7±0.8
Supervised Median Batch NG	89.4
Supervised Relational k -Means	88.2
Relational k -Means	88.0
Median Batch NG	76.3
Median k -Means	76.1

Αποτελέσματα κατηγοριοποίησης της οικογένειας των σφαιρινών που επιτυγχάνονται από διαφορετικές προσεγγίσεις όπως SOHMMM, σχεσιακά SOMs επιβλεπόμενης μάθησης, (σχεσιακές) NG παραλλαγές και τεχνικές k – μέσων .

Σε πολλές εφαρμογές δεν αρκεί ένας αλγόριθμος απλά να ανιχνεύει ένα βιολογικό σήμα, αλλά θα όφειλε να παρέχει τα μέσα για τη μετάφραση της προτεινόμενης λύσης με σκοπό τη βιολογική ενδοσκόπηση καθώς και την εξαγωγή γνώσης αναφορικά με το υπό εξέταση πρόβλημα ανάλυσης αλληλουχιών. Κάθε νευρώνας SOHMMM, που συσχετίζεται με μία εκάστοτε συστάδα, είναι μία πηγή επιπρόσθετης πληροφορίας υπό την έννοια ότι μπορεί να χρησιμοποιηθεί για τη μελέτη των προφίλ και των στατιστικών ιδιοτήτων των πρωτεϊνών, που περιγράφονται από το αντίστοιχο HMM. Με αυτή τη λογική, τα αποτελέσματα των υπόλοιπων αλγορίθμων δεν συνιστούν τίποτα άλλο παρά ομαδοποιήσεις προκαθορισμένων (αν)ομοιοτήτων (οι οποίες δεν βασίζονται σε κάποιο πρότυπο ή στατιστικό μοντέλο), και συνεπώς, δεν αφήνουν πολλά περιθώρια για περαιτέρω διερεύνηση και ανάλυση.

Κλείνοντας, ένα επιπλέον πλεονέκτημα του SOHMMM, έναντι των μεθοδολογιών που βασίζονται σε σχεσιακά δεδομένα, έχει να κάνει με το υπεισερχόμενο υπολογιστικό κόστος. Η υπολογιστική πολυπλοκότητα της εκπαίδευσης σχεσιακών

μοντέλων είναι ανάλογη του κύβου του πλήθους των δεδομένων ακολουθιών, αντιθέτως, το SOHMM αποδεικνύεται υπολογιστικά εύρωστο καθώς η πολυπλοκότητα του στοχαστικού αλγορίθμου προσαρμογής που το συνοδεύει είναι γραμμικώς ανάλογη του πλήθους των δεδομένων ακολουθιών. Όπως έγινε και λόγος προηγούμενα, παρότι έχουν προταθεί ορισμένες ταχύτερες υλοποιήσεις [49], η ανά βήμα εκπαίδευσης τετραγωνική φύση σχεσιακών αλγορίθμων δεν δύναται να αποφευχθεί, γιατί ουσιαστικά όλα τα πρότυπα εισόδου τα οποία αναπαριστώνται από αριθμητικές τιμές (αν)ομοιοτήτων περιγράφονται εγγενώς από ένα τετραγωνικό πλήθος αποστάσεων ζευγών.

5.3. *Splice Junctions Ακολουθιών Γονιδίων*

Τα splice junctions σηματοδοτούν μεταβάσεις μεταξύ περιοχών ενός γονιδίου που εκφράζονται και περιοχών που δεν εκφράζονται, και αντιστρόφως. Συνεπώς είναι σημαντικά για τη σύνθεση δομικών και ρυθμιστικών πρωτεϊνών που απαρτίζουν και ελέγχουν πλείστους βιοχημικούς μεταβολισμούς. Ειδικότερα, τα splice sites είναι θέσεις στο DNA στα σύνορα μεταξύ εξωνίων - exons (περιοχές γονιδίων που μεταγράφονται και μεταφράζονται, δηλαδή φέρουν την αλληλουχία των κωδικονίων που κωδικοποιούν αμινοξέα, άρα κωδικοποιούνται σε πρωτεΐνες) και εσωνίων - introns (οι περιοχές του γονιδίου που μεταγράφονται αλλά τελικά αποκόπτονται από το ετερογενές RNA και δεν μεταφράζονται, δηλαδή οι αλληλουχίες τους δεν κωδικοποιούν αμινοξέα, άρα δεν κωδικοποιούνται σε πρωτεΐνες). Ουσιαστικά το μάτισμα - splicing αναφέρεται στη διαδικασία επεξεργασίας του RNA κατά την οποία γίνεται αποκοπή των εσωνίων και συρραφή των εξωνίων σε ένα μόριο RNA, με τη βοήθεια συμπλόκων. Επομένως, γίνεται σαφές ότι η αναγνώριση των splice sites είναι ένα βασικό πεδίο ενδιαφέροντος για την κατανόηση σχέσεων γονότυπων/φαινοτύπων.

Η αποθήκη μηχανικής μάθησης UCI [78] περιέχει ένα σύνολο δεδομένων για τον καθορισμό splice junctions σε γονίδια από πρωτεύοντες (ευκαρυωτικούς) οργανισμούς. Το πρόβλημα που τίθεται σε αυτό το σύνολο δεδομένων είναι η αναγνώριση ορίων εξωνίων-εσωνίων (που συχνά καλούνται donor splice sites), ορίων εσωνίων-εξωνίων (acceptor splice sites) και θέσεων που δεν γνωρίζουμε να περιέχουν splice sites. Το σύνολο δεδομένων απαρτίζεται από 3190 αλληλουχίες DNA από τις οποίες 767 κατηγοριοποιούνται ως donors, 768 κατηγοριοποιούνται ως acceptors και οι υπόλοιπες 1655 ως κανένα από τα δύο (neither). Προφανώς για όλες τις αλληλουχίες θεωρούμε το

αλφάβητο τεσσάρων γραμμάτων των νουκλεοτιδίων $\{(A)denine, (T)hymine, (G)uanine, (C)ytosine\}$. Ένας θα πρέπει να λαμβάνει υπόψη ότι αυτό το πρόβλημα αναφοράς είναι ειδικά σχεδιασμένο για αλγορίθμους επιβλεπόμενης μάθησης που μπορούν να χειριστούν αποκλειστικά και μόνο αλληλουχίες σταθερού μήκους με το splice junction τοποθετημένο στο κέντρο ακριβώς κάθε ακολουθίας (δηλαδή συνθήκες που ευνοούν την πλειοψηφία των μοντέλων επιβλεπόμενης μάθησης με σταθερό αριθμό εισόδων). Προφανώς, η απόδοση αρκετών αλγορίθμων επιβλεπόμενης μάθησης θα μειωνόταν εάν τα splice junctions ήταν αυθαίρετα τοποθετημένα σε διαφορετικές θέσεις των ακολουθιών. Πολύ περισσότερο, εάν οι αλληλουχίες ήταν υψηλών διαστάσεων και μεταβλητών μηκών πολλές εκ των αναφερόμενων τεχνικών επιβλεπόμενης μάθησης θα ήταν πρακτικά ανεφάρμοστες. Πάραυτα, εάν και δεν είναι σύνηθες (για ευνόητους λόγους) να συγκρίνεται μία μεθοδολογία μη επιβλεπόμενης μάθησης με προσεγγίσεις επιβλεπόμενης μάθησης (που κάνουν χρήση πληροφορίας κατηγοριοποίησης, εκ των προτέρων γνώσης και θεωρίας πεδίου), πιστεύουμε ότι μία τέτοια πειραματική πρακτική παρέχει μία επιπλέον απόδειξη περί της απόδοσης και των δυνατοτήτων του SOHMMM. Επίσης, η παρούσα πειραματική διάταξη μας δίνει την ευκαιρία να συγκρίνουμε το SOHMMM έναντι αρκετών εναλλακτικών προσεγγίσεων, για την ακρίβεια 20, σε ένα γνωστό και εδραιωμένο πρόβλημα κατηγοριοποίησης ακολουθιών (παρότι όπως προείπαμε και οι 20 είναι επιβλεπόμενης μάθησης).

Για αυτή τη σειρά πειραμάτων χρησιμοποιήθηκε μία συστοιχία SOHMMM αποτελούμενη από 88 HMMs διατεταγμένα σε ένα εξαγωνικό πλέγμα 11x8. Ο ρυθμός μάθησης μειώνεται εκθετικά μεταξύ των τιμών 1.0 και 0.1. Το δε εύρος της τοπολογικής γειτονιάς αρχικά είναι ίσο με το ήμισυ της μεγαλύτερης διάστασης του δικτύου και στη συνέχεια μειώνεται γραμμικά μέχρι να συμπεριλαμβάνει μόνο τους άμεσα γειτονικούς νευρώνες του νικητή νευρώνα. Επίσης, το συνολικό πλήθος εποχών που επιστρατεύθηκαν για την εκπαίδευση του SOHMMM είναι δέκα. Η απόδοση του SOHMMM υπολογίστηκε και πάλι βάσει της μεθοδολογίας stratified ten-fold cross-validation. Ο πίνακας 5.2 παρουσιάζει τη μέση τιμή σφάλματος που παράγεται από το SOHMMM, και επίσης, συνοψίζει τις μέσες τιμές των σφαλμάτων μίας πλειάδας αλγορίθμων μηχανικής μάθησης: τα αποτελέσματα αυτά, εκτός αυτών που αντιστοιχούν στο Supervised HMM (SHMM), έχουν ανακτηθεί από τις [79]-[84]. Η πλειοψηφία των αποτελεσμάτων που αναφέρονται έχουν προκύψει από την εφαρμογή

ενός σχήματος stratified ten-fold cross-validation επί 1000 τυχαία επιλεγμένων ακολουθιών. Οι επιδόσεις των πρώτων έξι μοντέλων όπως επίσης και οι τιμές σφαλμάτων των δύο Constrained Confidence-Rated ILP-Boosting (C²RIB) παραλλαγών είναι εκτιμήσεις που έχουν προκύψει εφαρμόζοντας την τεχνική stratified ten-fold cross-validation σε όλες τις διαθέσιμες 3190 αλληλουχίες. Σε αυτή την πειραματική διάταξη, που περιλαμβάνει όλες τις 3190 ακολουθίες, οι επιδόσεις του SOHMMM βελτιώνονται ελαφρά, έτσι οι μέσες τιμές των σφαλμάτων ελαττώνονται στο $9.31\pm 0.55\%$ (overall), $2.73\pm 0.31\%$ (donor), $3.11\pm 0.36\%$ (acceptor), και $3.47\pm 0.38\%$ (neither). Ένα τέτοιο αποτέλεσμα θα έπρεπε να αναμένεται από τη στιγμή που το SOHMMM αξιοποιεί την επιπλέον πληροφορία που περιέχεται στο πλήρες σύνολο δεδομένων, βελτιώνοντας έτσι την ακρίβεια του. Οι τιμές των σφαλμάτων που παράγονται από τα τρία τελευταία μοντέλα αντιστοιχούν σε ένα σχήμα stratified ten-fold cross-validation επί των donor και acceptor ακολουθιών (δηλαδή λαμβάνοντας υπόψη μόνο 1535 αλληλουχίες). Συγκεκριμένα, τα αποτελέσματα που αφορούν σε συστήματα κατηγοριοποίησης που βασίζονται σε Support Vector Machines (SVMs) που λειτουργούν σε χώρους χαρακτηριστικών Selective Markov Models (SMMs), αλυσίδων Markov υψηλής τάξης και Interpolated Markov Models (IMMs), αντιστοιχούν σε ένα πρόβλημα κατηγοριοποίησης ακολουθιών που ανήκουν σε μία εκ των donor ή acceptor κλάσεων.

ΠΙΝΑΚΑΣ 5.2

ΜΕΣΕΣ ΤΙΜΕΣ ΠΟΣΟΣΤΙΑΙΩΝ ΣΦΑΛΜΑΤΩΝ

Model	Donor	Acceptor	Neither	Overall
LI SVM	0.8±0.2	0.9±0.3	2.0±0.3	3.7
FK SVM	1.6±0.5	1.6±0.4	2.1±0.4	5.3
TOP SVM	1.5±0.4	1.7±0.3	2.2±0.4	5.4
PCL	—	—	—	5.6
C4.5	—	—	—	5.7
SHMM-DDA	1.0±0.4	2.4±0.7	2.6±0.5	6.0
SHMM	2.02±0.25	3.08±0.29	2.83±0.32	7.93±0.43
SOHMMM	2.90±0.35	3.25±0.38	4.05±0.46	10.20±0.57
C ² RIB ^D	—	—	—	12.0

BRAIN	5.0	4.0	4.0	13.0
KBANN	7.56	8.47	4.62	20.65
Back-Propagation	5.74	10.75	5.29	21.78
PEBLS	8.18	7.55	6.86	22.59
ID3	10.58	13.99	8.84	33.41
COBWEB	15.04	9.46	11.80	36.30
Perceptron	16.32	17.41	3.99	37.72
C ² RIB	—	—	—	46.9
Nearest Neighbor	11.65	9.09	31.11	51.85
SVM SMM	12.31	—	—	—
SVM MC	12.32	—	—	—
SVM IMM	17.23	—	—	—

Οι επιδόσεις του SOHMMM και διαφόρων άλλων αλγορίθμων μηχανικής μάθησης στο πρόβλημα καθορισμού των splice junctions.

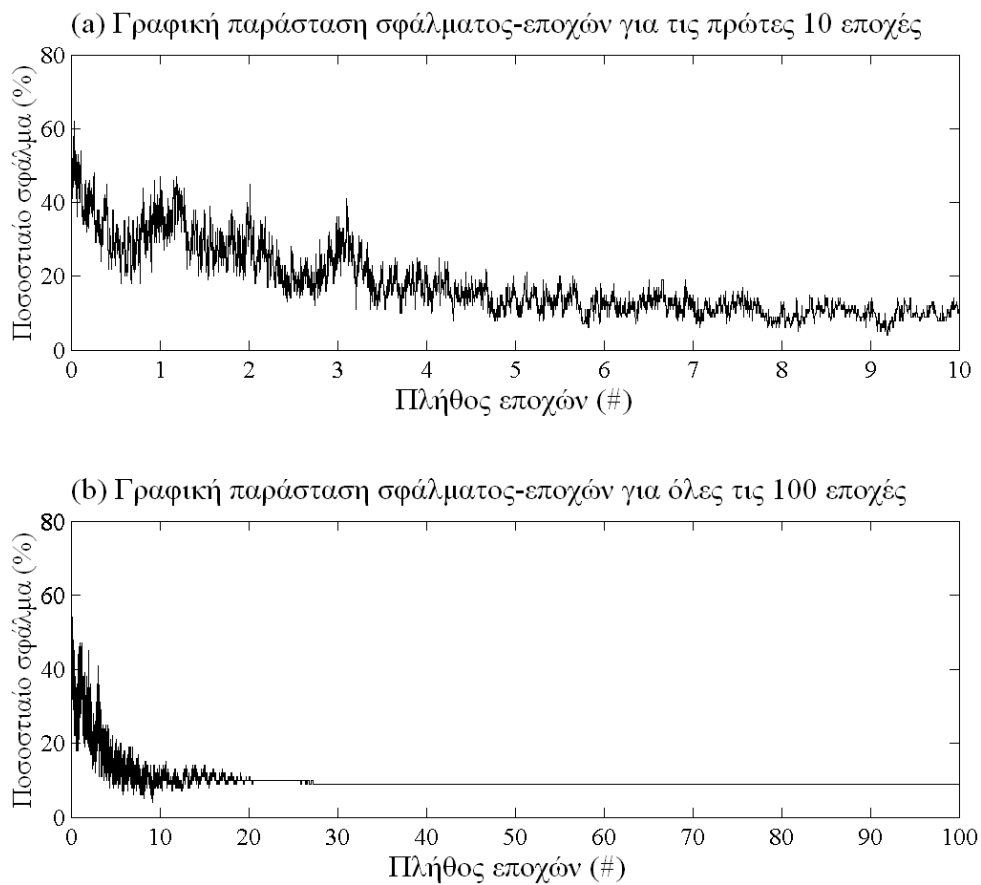
Δύο βασικές επισημάνσεις μπορούν να γίνουν εξετάζοντας τα αποτελέσματα που περιέχονται στον πίνακα 5.2. Πρώτον, τα πιθανοτικά μοντέλα (και ειδικότερα αυτά που ενσωματώνουν HMMs όπως τα Fisher Kernel (FK) SVM και Tangent vectors Of Posterior log-odds (TOP) SVM) ξεπερνούν κατά πολύ τις υπόλοιπες μεθόδους που παρουσιάζονται. Αυτό το εύρημα στηρίζει περαιτέρω την πρόταση ότι η ανάλυση βιολογικών ακολουθιών θα αναβαθμιστεί ακολουθώντας απευθείας μία συστηματική πιθανοτική κατεύθυνση [2]. Δεύτερον, όλα τα αποτελέσματα που αναφέρονται προκύπτουν από μοντέλα επιβλεπόμενης μάθησης που κάνουν χρήση πληροφορίας κατηγοριοποίησης και εκ των προτέρων γνώσης. Τα μοντέλα αυτά καλύπτουν ένα μεγάλο εύρος εδραιωμένων τεχνικών μηχανικής μάθησης όπως (σταθμισμένους) πλησιέστερους γείτονες - (weighted) nearest neighbors, (πολυεπίπεδα) perceptrons - (multi-layer) perceptrons, υβριδικά (υπο)συμβολικά συστήματα κατηγοριοποίησης - hybrid (sub)symbolic classifiers, συστήματα κατηγοριοποίησης με βάσεις κανόνων συνδυαζόμενων με διακριτική ανάλυση - rule-based classifiers combined with discriminant analysis, προώθησης με περιορισμούς - constrained boosting, SVMs που λειτουργούν σε χώρους χαρακτηριστικών από μοντέλα Markov υψηλής τάξης, συστήματα κατηγοριοποίησης SHMMs, SVMs με Locally Improved (LI) ή πιθανοτικούς πυρήνες. Στον αντίποδα, η απόδοση του SOHMMM είναι ανώτερη εν

συγκρίσει με αρκετές τεχνικές επιβλεπόμενης μάθησης, παρά το γεγονός ότι είναι ένας αλγόριθμος μη επιβλεπόμενης μάθησης και ως τέτοιος δεν μετέρχεται κανενός είδους επιπρόσθετης πληροφορίας και εκ των προτέρων γνώσης. Οι μόνες μέθοδοι που ξεπερνούν το SOHMMM με όρους ακρίβειας είναι τα συστήματα κατηγοριοποίησης με βάσεις κανόνων, τα SHMMs και τα SVMs. Αυτό είναι κάτι αναμενόμενο, όχι μόνο επειδή αυτές οι προσεγγίσεις χρησιμοποιούν πληροφορία κατηγοριοποίησης και γνώση πεδίου για τη ρύθμιση των παραμέτρων τους, αλλά επειδή είναι εξοπλισμένες με HMM υπομονάδες ειδικά σχεδιασμένες για τη μοντελοποίηση του προβλήματος αναγνώρισης των splice junctions.

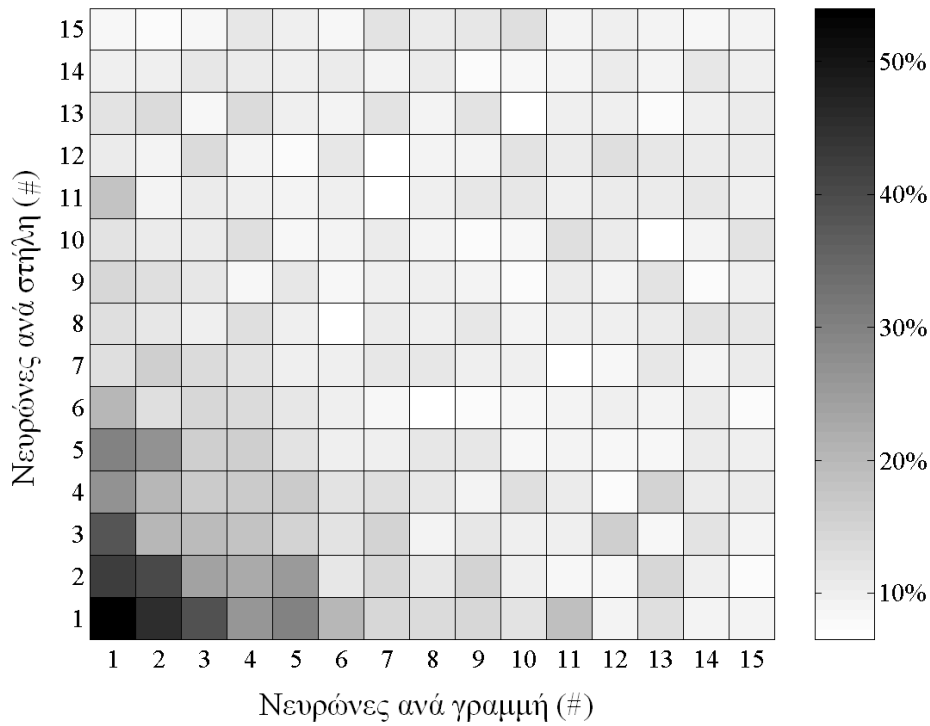
Το σύστημα κατηγοριοποίησης SHMM, το οποίο υπερτερεί του SOHMMM ως προς την απόδοση, απαρτίζεται από τρεις μονάδες, κάθε μία εκ των οποίων αναπαριστά αποκλειστικά και μόνο ένα συγκεκριμένο τύπο splice junction. Για την προσαρμογή των παραμέτρων του SHMM ακολουθήθηκε μία μεθοδολογία πλήρως επιβλεπόμενης μάθησης, καθώς κάθε μονάδα του (που στην πραγματικότητα είναι ένα σύνολο από HMMs) εκπαιδεύτηκε για να περιγράψει μία συγκεκριμένη κατηγορία splice site. Πέραν αυτού, το σύστημα κατηγοριοποίησης Supervised HMM Domain Dependent Architecture (SHMM-DDA), που επίσης παράγει καλύτερα από το SOHMMM, διαθέτει μία προκαθορισμένη αρχιτεκτονική βιολογικά εμπνευσμένη, η επιλογή των υπερ-παραμέτρων του έχει λάβει χώρα ανεξάρτητα και προηγείται της διαδικασίας εκπαίδευσης, και επίσης, διαφορετικά HMM οδηγούμενα με επιβλεπόμενη μάθηση έχουν ανατεθεί για την περιγραφή κάθε ξεχωριστής κατηγορίας splice junction. Το γεγονός ότι το SOHMMM υπερτερεί όλων των υπόλοιπων προσεγγίσεων επιβλεπόμενης μάθησης στο πρόβλημα καθορισμού των splice sites, οφείλεται στα χαρακτηριστικά και τις ιδιότητες του όπως αυτές έχουν αναλυθεί προηγουμένα.

Πειράματα διεξήχθησαν επίσης για μελετηθεί το με ποιο τρόπο και το σε ποιο βαθμό η συμπεριφορά και οι επιδόσεις του SOHMMM επηρεάζονται από τη διάρκεια της φάσης εκπαίδευσης, από το μέγεθος της συστοιχίας και από την αρχιτεκτονική των HMM νευρώνων (δηλαδή των αριθμό των καταστάσεων τους). Οι συναρτήσεις του ρυθμού μάθησης και της τοπολογικής γειτονιάς ακολούθησαν τα σχήματα ρύθμισης που εφαρμόστηκαν σε όλα τα προηγούμενα πειράματα. Ένα ενδεικτικό σενάριο μίας διαδικασίας προσαρμογής (ενός 11x8 πλέγματος SOHMMM) που διαρκεί δεκαπλάσιο αριθμό εποχών απ' ότι προηγούμενα φαίνεται στην εικόνα 5.12. Η δε γραφική παράσταση που εμφανίζεται στην εικόνα 5.13 παρουσιάζει τα αποτελέσματα μίας πολλαπλής διαδικασίας stratified ten-fold cross-validation όταν αυτή εφαρμόζεται σε

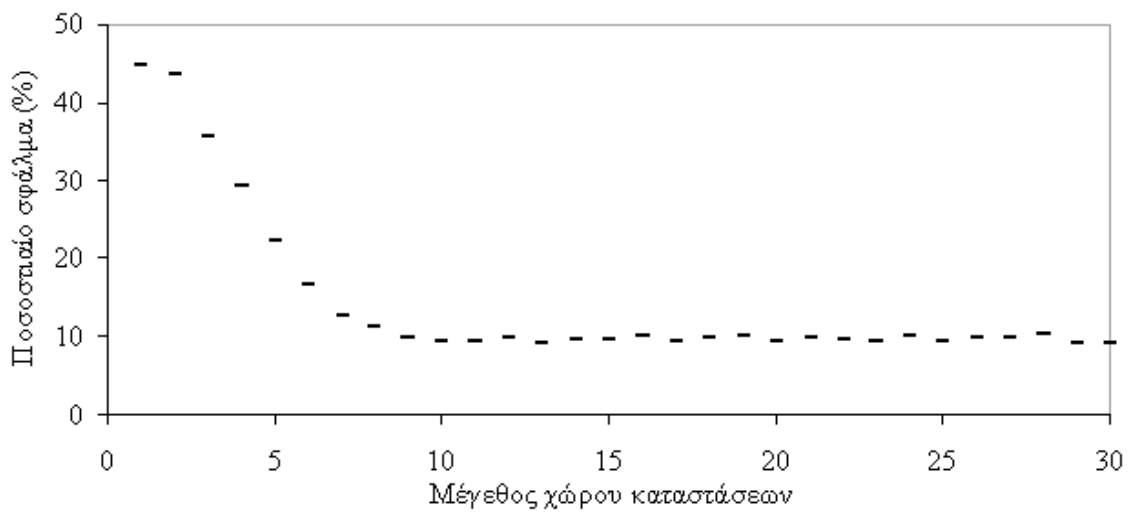
μία πλειάδα διαστάσεων της συστοιχίας SOHMMM. Τέλος, η γραφική παράσταση στην εικόνα 5.14 αποτυπώνει τα αποτελέσματα μίας σειράς διαδικασιών stratified ten-fold cross-validation για διαφορετικά μεγέθη του χώρου καταστάσεων των HMM νευρώνων (μίας 11x8 συστοιχίας SOHMMM). Κατόπιν εξέτασης και ανάλυσης των αποτελεσμάτων που αποτυπώνονται στις εικόνες 5.12-5.14 είναι έκδηλο ότι οι παρατηρήσεις, τα ευρήματα και τα συμπεράσματα είναι παρεμφερή, εάν όχι ταυτόσημα, με αυτά που αφορούν στην πρωτεϊνική οικογένεια των σφαιρινών.



Εικόνα 5.12. Γραφικές παραστάσεις ποσοστιαίου σφάλματος-πλήθους εποχών του SOHMMM, όπως αυτές προκύπτουν για το σύνολο δεδομένων των splice junctions. Η γραφική παράσταση (a) αφορά τις πρώτες 10 εποχές εκπαίδευσης από τις συνολικά 100 που χρησιμοποιήθηκαν.



Εικόνα 5.13. Μέσες τιμές ποσοστιαίων σφαλμάτων που παράγονται από πλέγματα SOHMMM διαφόρων διαστάσεων.



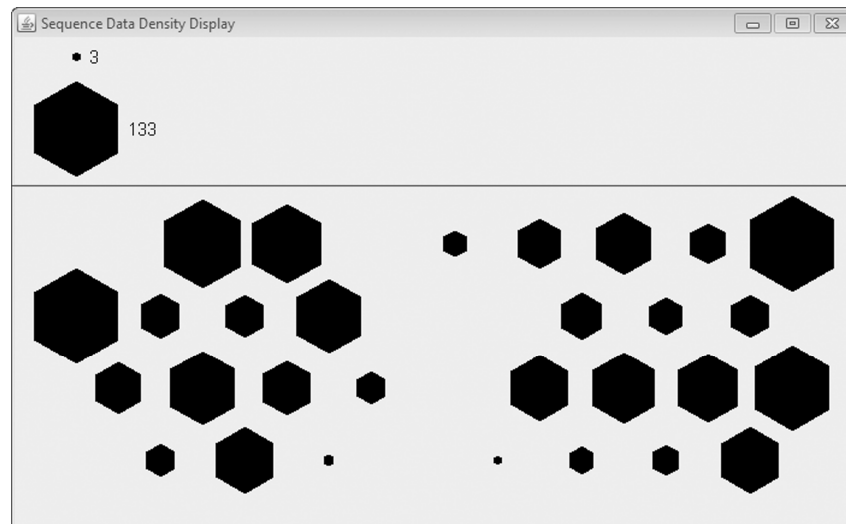
Εικόνα 5.14. Γραφική παράσταση ποσοστιαίου σφάλματος-μεγέθους χώρου κρυφών καταστάσεων των HMM νευρώνων ενός 11x8 SOHMMM, η οποία αντιστοιχεί στο σύνολο δεδομένων των splice sites.

Επιχειρώντας μία δεύτερη πειραματική διερεύνηση και μελέτη των προτεινόμενων οπτικοποιήσεων και γραφικών απεικονίσεων θεωρούμε ένα εξαγωνικού πλέγματος SOHMMM διαστάσεων 9x4 το οποίο έχει εκπαιδευτεί βάσει του παρόντος συνόλου δεδομένων αλληλουχιών (με πανομοιότυπο τρόπο με αυτόν που περιγράφηκε και

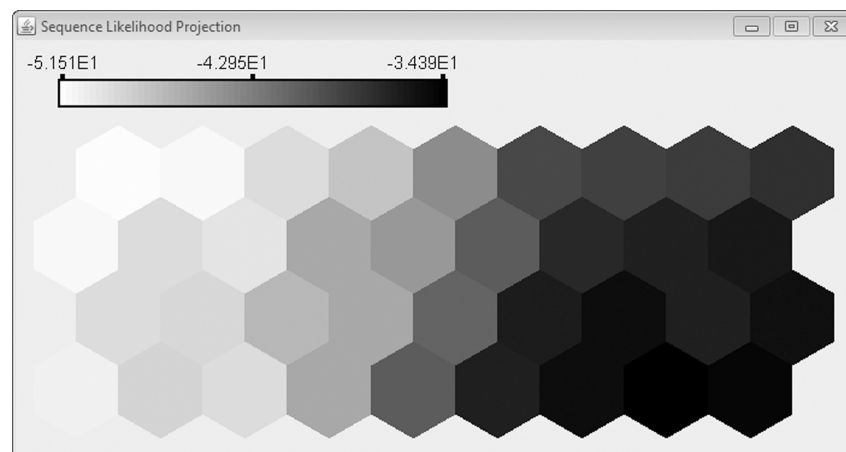
αναλύθηκε προηγούμενα). Το επακριβές πρόβλημα που τίθεται είναι η αναγνώριση των ορίων εξωνίων-εσωνίων (donor splice sites) καθώς και των ορίων εσωνίων-εξωνίων (acceptor splice sites). Στην προκειμένη περίπτωση οι προκύπτουσες γραφικές αναπαραστάσεις φαίνονται στα σχήματα 5.15-5.17.

Αρχικά, είναι άμεσα επαληθεύσιμο το γεγονός ότι η SDDD, σε συνέργια με το υποκείμενο SOHMMM, εντοπίζει την ύπαρξη των δύο συστάδων από splice junctions, δηλαδή της donor συστάδας και της acceptor συστάδας. Επιπλέον, οι εντοπισμένες συστάδες αναγνωρίζονται και αποτυπώνονται οπτικά με σαφή τρόπο (εικόνα 5.15) μετά το πέρας της διαδικασίας μη επιβλεπόμενης μάθησης (η οποία σε κανένα στάδιο της δεν απαιτεί ούτε χρησιμοποιεί εκ των προτέρων πληροφορία ανάθεσης κατηγοριών). Η ερμηνεία των SLP αποτελεσμάτων (εικόνες 5.16, 5.17) είναι διττή. Αφενός μεν μπορούν να θεωρηθούν αναθέσεις αγνώστων/ορφανών αλληλουχιών στις συστάδες που ήδη έχουν εντοπιστεί· τουτέστιν οι νουκλεοτιδικές ακολουθίες που ανατίθενται στην donor συστάδα θεωρείται ότι περιέχουν ένα όριο εξωνίου-εσωνίου ενώ αντίθετα οι νουκλεοτιδικές ακολουθίες που ανατίθενται στην acceptor συστάδα θεωρείται ότι περιέχουν ένα όριο εσωνίου-εξωνίου. Αφετέρου δε μπορούν να θεωρηθούν ως μία επιπρόσθετη επιβεβαίωση/ένδειξη της ύπαρξης (σε πρώτο χρόνο) και της ορθότητας (σε δεύτερο χρόνο) των δύο συστάδων που εντοπίζονται, καθώς καθεμία εκ των 1535 αλληλουχιών παράγει υψηλές τιμές πιθανοφάνειας ακριβώς σε μία εκ των δύο διακριτών περιοχών στο επίπεδο SOHMMM και αυτές οι περιοχές συμπίπτουν με τις εν λόγω συστάδες.

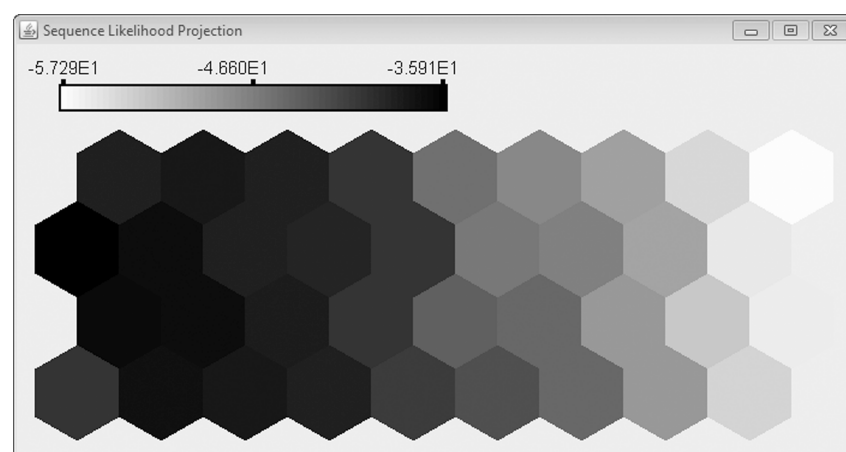
Επιπροσθέτως, θεωρώντας ότι διατίθεται πληροφορία ανάθεσης κατηγοριών αυτή δύναται να αξιοποιηθεί για έναν εκ των υστέρων χαρακτηρισμό των εντοπιζόμενων συστάδων στην SDDD απεικόνιση (όπως έχει γίνει στην εικόνα 5.7). Σε αυτήν την περίπτωση μία άγνωστη ή μη ταξινομημένη splice junction μπορεί να κατηγοριοποιηθεί βάσει της κατηγορίας του νευρώνα SOHMMM που παράγει την υψηλότερη πιθανοφάνεια. Είναι αξιοσημείωτο ότι η γραφική απεικόνιση που εξάγεται με την SLP συνιστά ένα τοπίο πιθανοφάνειας το οποίο δεν περιορίζεται μόνο στην αποτύπωση του νευρώνα νικητή και της τιμής της μέγιστης πιθανοφάνειας αλλά αντιθέτως εκτείνεται πέραν αυτών παρέχοντας μία συνολικότερη/περιεκτικότερη περιγραφή.



Εικόνα 5.15. SDDD που αντιστοιχεί στο πρόβλημα προσδιορισμού των splice junctions.



Εικόνα 5.16. Ενδεικτική SLP απεικόνιση μίας αλληλουχίας που περιέχει ένα όριο εξωνίου-εσωνίου.



Εικόνα 5.17. Χαρακτηριστικό παράδειγμα μίας SLP γραφικής αναπαράστασης που αφορά σε ένα acceptor splice site.

6. ΕΠΙΛΟΓΟΣ

Το SOHMMM συνιστά ένα υβριδικό μοντέλο μηχανικής μάθησης που εντοπίζεται στα σύνορα μεταξύ των SOM και HMM προσεγγίσεων. Κατά συνέπεια, βασιζόμενοι στη θεωρητική θεμελίωση και την πειραματική επαλήθευση, που έχουν προηγηθεί, μπορούμε να υποστηρίξουμε ότι σε πρακτικό επίπεδο, λιγότερο ή περισσότερο επιτυχώς, το SOHMMM είναι σε θέση να:

- 1) Αναλύει απευθείας (χωρίς την παρεμβολή κανενός σταδίου μετασχηματισμού ή προεπεξεργασίας) ακολουθίες παρατηρήσεων και βιολογικές αλληλουχίες (τουτέστιν γραμμικές περιγραφές πρωτεϊνών και γονιδίων).
- 2) Συλλαμβάνει, να επεξεργάζεται και να αξιοποιεί την ενυπάρχουσα αλλά αφανή πληροφορία που υποκρύπτεται στις χρονικές/χωρικές συσχετίσεις και στα γραμμικά ακολουθιακά χαρακτηριστικά τέτοιων μορίων.
- 3) Χειρίζεται και να μοντελοποιεί συμβολικές/μη αριθμητικές ακολουθίες υψηλών διαστάσεων και διαφόρων μηκών.
- 4) Καλύπτει ένα διευρυμένο σύνολο στατιστικών κατανομών οι οποίες αναπαριστούν με ποιο πιστό τρόπο τον εκάστοτε χώρο ακολουθιών.
- 5) Ολοκληρώνει λειτουργίες ομαδοποίησης, μείωσης διαστατικότητας, οπτικοποίησης συστάδων σε ένα ενοποιημένο παράδειγμα.
- 6) Παρέχει, υπό συγκεκριμένες προϋποθέσεις, διαδικασίες διαχωρισμού, αναζήτησης και κατηγοριοποίησης ακολουθιών παρατηρήσεων.

Παρόλα αυτά, το SOHMMM πάσχει από δύο σημαντικές αδυναμίες, τις οποίες έχει κληρονομήσει από τα HMMs. Πρώτον, τα HMMs για βιολογικές ακολουθίες συχνά συνεπάγονται έναν υψηλό αριθμό ελεύθερων παραμέτρων, πράγμα που έχει ως αποτέλεσμα να απαιτείται ένα μεγάλο πλήθος περιεκτικών (σε πληροφορία) και συνάμα περιγραφικών (του υπό εξέταση χώρου προβλήματος) μορίων αλληλουχιών. Το πρόβλημα αυτό γίνεται εντονότερο στις περιπτώσεις εκείνες στις οποίες μόνο ένας μικρός αριθμός αλληλουχιών είναι διαθέσιμος, κάτι το οποίο είναι σύνηθες σε πρώιμα στάδια προσδιορισμού γονιδιωμάτων ή σε ορισμένες οικογένειες πρωτεϊνών και γονιδίων. Δεύτερον, τα HMMs πρώτης τάξης περιορίζονται εγγενώς από την ιδιότητα Markov· δεν μπορούν να εκφράσουν εξαρτήσεις μεταξύ τριών ή περισσότερων καταστάσεων, με άλλα λόγια, βασίζονται στην (αυθαίρετη) υπόθεση ότι δοθέντος ενός παρόντος γεγονότος η δεσμευμένη πιθανότητα πραγματοποίησης αυτού με την προϋπόθεση του παρελθόντος ιστορικού συμβάντων εξαρτάται αποκλειστικά και μόνο

από το άμεσο παρελθόν συμβάν και όχι από τα οποιοδήποτε απώτερα παρελθόντα συμβάντα. Οι πρωτεΐνες αποκτούν πολύπλοκες τρισδιάστατες δομές οι οποίες και καθορίζουν σε μεγάλο βαθμό τη λειτουργία τους. Είναι πιθανό να υπάρχουν περίπλοκες απομακρυσμένες συσχετίσεις μεταξύ των στοιχείων μίας μοριακής αλληλουχίας, αυτές είναι προσβάσιμες από HMMs υψηλών τάξεων αλλά στην πλειονότητα των περιπτώσεων δεν εντοπίζονται καν από HMMs πρώτης τάξης. Επομένως γίνεται σαφές ότι και το SOHMMM αποτυγχάνει στην ανίχνευση τέτοιων εξαρτήσεων. Ωστόσο, δεν πρέπει να παραβλέπουμε το γεγονός ότι ένας ανυπέρβλητος περιορισμός των HMMs υψηλής τάξης αφορά στην εκθετική αύξηση των ελεύθερων μεταβλητών τους και του υπολογιστικού κόστους που αυτή συνεπάγεται.

Το SOHMMM θα μπορούσε να εμπλουτιστεί ενσωματώνοντας ποικίλα πιθανοτικά παραδείγματα, οριακά ή εντελώς διαφορετικά από το HMM. Για παράδειγμα, υψηλής τάξης HMMs, κρυφά μοντέλα semi-Markov, μοντέλα Markov μέγιστης εντροπίας, πολυεπίπεδα HMMs ή ακόμη και conditional random fields θα μπορούσαν να αντικαταστήσουν τους πρώτης τάξης HMM νευρώνες. Στη γενική περίπτωση, το SOHMMM περιλαμβάνει HMMs με τη μέγιστη δυνατή πολυπλοκότητα, δηλαδή πλήρως συνεκτικά HMMs όπου οι πιθανότητες μετάβασης μεταξύ όλων των καταστάσεων είναι μη μηδενικές (γεγονός που συνεπάγεται πυκνούς στοχαστικούς πίνακες μετάβασης). Σε μία πλειάδα πραγματικών εφαρμογών (όπως για παράδειγμα αυτών της αναγνώρισης φωνής και της ανακάλυψης μοτίβων) χρειάζεται να επιστρατεύσουμε κατά περίπτωση αρχιτεκτονικές που να ενσωματώνουν HMMs με μεγαλύτερους χώρους καταστάσεων και αραιότερες συνδέσεις. Παραδείγματος χάριν, τα γραμμικά ακολουθιακά χαρακτηριστικά βιολογικών αλληλουχιών αποκρυπτογραφούνται από τις left-to-right αρχιτεκτονικές, με την κανονική γραμμική αρχιτεκτονική να αποτελεί την πιο βασική και πλέον διαδεδομένη εξ αυτών. Οι εν λόγω αρχιτεκτονικές επιβάλλουν ορισμένους περιορισμούς στον στοχαστικό πίνακα της υποκείμενης αλυσίδας Markov. Μία άλλη επέκταση θα μπορούσε να είναι η ανάπτυξη ενός ανάλογου παραδείγματος για εφαρμογές που απαιτούν συνεχείς συναρτήσεις πιθανοτήτων εμφάνισης. Κατά συνέπεια, η κατασκευή παραλλαγών του SOHMMM για την αντιμετώπιση ζητημάτων αναγνώρισης φωνής και παρεμφερών προβλημάτων συνιστά μία ευθεία διαδικασία. Επιπλέον, η υλοποίηση τους εξοικονομεί υπολογιστικούς πόρους αφού τα μήκη των αντίστοιχων αλυσίδων Markov, που πρέπει να χρησιμοποιηθούν, είναι σαφώς μικρότερα, από αυτά που απαιτούνται για τις βιολογικές αλληλουχίες.

Εξαιτίας της απλότητας και της αποδοτικότητας του, το SOM έχει διαδραματίσει ένα σημαντικό ρόλο στο προσκήνιο της αυτο-οργανούμενης μάθησης. Το γεγονός αυτό έχει αποτελέσει το εφαλτήριο για τη δημιουργία μιας σειράς παραλλαγών και επεκτάσεων. Συνεπώς, μία εύλογη ερευνητική κατεύθυνση θα μπορούσε να είναι η αναβάθμιση αυτών των εναλλακτικών τεχνικών με τη δυνατότητα να επεξεργάζονται και να αναλύουν απευθείας ακολουθίες παρατηρήσεων και βιολογικές αλληλουχίες. Μαρκοβιανές αλλά και άλλες πιθανοτικές υπομονάδες δύναται να ενσωματωθούν σε ιεραρχικά/αυξητικά/ασαφή SOMs, GTMs, NGs ή LArge Memory SStorage And Retrieval (LAMSTAR) νευρωνικά δίκτυα.

Σε ένα ευρύτερο πιο αφαιρετικό επίπεδο, υβριδικά συστήματα μηχανικής μάθησης δύναται να σχηματιστούν συνδυάζοντας νευρωνικά δίκτυα μη επιβλεπόμενης μάθησης και πιθανοτικά μοντέλα. Έχει διαπιστωθεί ότι το πιο κρίσιμο στάδιο σε τέτοιες διαδικασίες συνένωσης και συγκερασμού είναι η ανάπτυξη ενός αξιόπιστου και ολοκληρωμένου σχήματος προσαρμογής/εκπαίδευσης. Ιδανικά, ο προκύπτων αλγόριθμος μάθησης θα πρέπει να:

- 1) είναι κατ' ουσίαν αυτόνομος, πράγμα που σημαίνει ότι κατά περίπτωση ευρετικές εξωγενείς ρυθμίσεις και ανθρώπινες παρεμβάσεις πρέπει να περιορίζονται στο ελάχιστο,
- 2) εκμεταλλεύεται (στο έπακρο) κάθε δυνατή ή πιθανή πηγή πληροφορίας, συμπεριλαμβανόμενης της λανθάνουσας αφανούς πληροφορίας που κρύβεται στις χωροχρονικές συσχετίσεις των στοιχείων του εκάστοτε χώρου δεδομένων,
- 3) απαιτεί ελάχιστη, ή ακόμη και μηδενική, εκ των προτέρων γνώση αναφορικά με το υπό εξέταση πρόβλημα που ανατίθεται κάθε φορά, πλην αυτής βεβαίως που μπορεί να εξαχθεί από τα δεδομένα με διαδικασίες αυτόνομης μάθησης,
- 4) είναι υπολογιστικά εύρωστος.

Κλείνοντας, αξίζει να αναφερθεί ότι η συλλογικά συσσωρευμένη ερευνητική εμπειρία έχει δείξει ότι πολλές κατευθύνσεις μελλοντικής έρευνας, συχνά είναι παραπλανητικές ή οδηγούν σε αδιέξοδα. Θα ήταν παράδοξο να θεωρήσουμε ότι οι κατευθύνσεις που παρουσιάστηκαν ακροθιγώς προηγούμενα διεκδικούν το αλάθητο.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Mount, D. W. (2004). *Bioinformatics: Sequence and Genome Analysis* (2nd ed.). New York: Cold Spring Harbor Laboratory Press.
- [2] Baldi, P., & Brunak, S. (2001). *Bioinformatics: The Machine Learning Approach* (2nd ed.). Cambridge, Massachusetts: The MIT Press.
- [3] Douglas, R., & Sejnowski, T. (2007). *Future Challenges for the Science and Engineering of Learning* (4th version). USA: National Science Foundation Workshop Report.
- [4] Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16, 645-678.
- [5] Du, K.-L. (2010). Clustering: a neural network approach. *Neural Networks*, 23, 89-107.
- [6] Kohonen, T. (2001). *Self-Organizing Maps* (3rd ed.). Berlin: Springer-Verlag.
- [7] Seiffert, U., & Jain, L. C. (2002). *Self-Organizing Neural Networks: Recent Advances and Applications*. Heidelberg, New York: Physica-Verlag.
- [8] Kraaijveld, M. A., Mao, J., & Jain, A. K. (1995). A nonlinear projection method based on Kohonen's topology preserving maps. *IEEE Transactions on Neural Networks*, 6, 645-678.
- [9] Ultsch, A. (2003). Maps for the visualization of high-dimensional data spaces. In *Proceedings of the Workshop on Self-Organizing Maps* (pp. 225-230).
- [10] Brugger, D., Bogdan, M., & Rosenstiel, W. (2008). Automatic cluster detection in Kohonen's SOM. *IEEE Transactions on Neural Networks*, 19, 442-459.
- [11] Tasdemir, K., & Merenyi, E. (2009). Exploiting data topology in visualization and clustering of self-organizing maps. *IEEE Transactions on Neural Networks*, 20, 549-562.
- [12] Tasdemir, K. (2010). Graph based representations of density distribution and distances for self-organizing maps. *IEEE Transactions on Neural Networks*, 21, 520-526.
- [13] Koski, T. (2001). *Hidden Markov Models for Bioinformatics*. Dordrecht, The Netherlands: Kluwer Academics Publishers.
- [14] Durbin, R., Eddy, S. R., Krogh, A., & Mitchison, G. (1998). *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge: Cambridge University Press.

- [15] Ferles, C., & Stafylopatis, A. (2008a). Sequence clustering with the self-organizing hidden Markov model map. In *8th IEEE International Conference on Bioinformatics and Bioengineering* (pp. 1-7).
- [16] Ferles, C., & Stafylopatis, A. (2008b). A hybrid self-organizing model for sequence analysis. In *20th IEEE International Conference on Tools with Artificial Intelligence* (pp. 105-112).
- [17] Ferles, C., Siolas, G., & Stafylopatis, A. (2011). Scaled on-line unsupervised learning algorithm for a SOM-HMM hybrid. In *26th International Symposium on Computer and Information Sciences* (pp. 533-537).
- [18] Ferles, C., & Stafylopatis, A. (2013). Self-Organizing Hidden Markov Model Map (SOHMMM). *Neural Networks*, 48, 133-147.
- [19] Ferles, C., Siolas, G., & Stafylopatis, A. (2013) Scaled self-organizing map - hidden Markov model architecture for biological sequence clustering. *Applied Artificial Intelligence*, 27, 461-495.
- [20] Ferles, C., & Stafylopatis, A. Cluster visualization and nonlinear projection techniques for biological sequences. Submitted for publication.
- [21] Hoekstra, A., & Drossaers, M. F. J. (1993). An extended Kohonen feature map for sentence recognition. In *International Conference on Artificial Neural Networks* (pp. 404-407).
- [22] Chappell, G., & Taylor, J. (1993). The temporal Kohonen map. *Neural Networks*, 6, 441-445.
- [23] Varsta, M., Heikkonen, J., & Millan, J. R. (1997). Context learning with the self-organizing map. In *Proceedings of the Workshop on Self-Organizing Maps* (pp. 197-202).
- [24] Koskela, T., Varsta, M., Heikkonen, J., & Kaski, K. (1998). Time series prediction using recurrent SOM with local linear models. *International Journal of Knowledge-Based and Intelligent Engineering Systems*, 2, 60-68.
- [25] Voegtlin, T. (2002). Recursive self-organizing maps. *Neural Networks*, 15, 979-992.
- [26] Strickert, M., & Hammer, B. (2004). Self-organizing context learning. In *Proceedings of the European Symposium on Artificial Neural Networks* (pp. 39-44).

- [27] Hagenbuchner, M., Sperduti, A., & Tsoi, A. C. (2003). A self-organizing map for adaptive processing of structured data. *IEEE Transactions on Neural Networks*, *14*, 491-505.
- [28] Sperduti, A. (2001). Neural networks for adaptive processing of structured data. In *International Conference on Artificial Neural Networks* (pp. 5-12).
- [29] Hagenbuchner, M., Sperduti, A., & Tsoi, A. C. (2005). Contextual processing of graphs using self-organizing maps. In *Proceedings of the European Symposium on Artificial Neural Networks* (pp. 399-404).
- [30] Vanco, P., & Farkas, I. (2010). Experimental comparison of recursive self-organizing maps for processing tree-structured data. *Neurocomputing*, *73*, 1362-1375.
- [31] Hammer, B., Micheli, A., Strickert, M., & Sperduti, A. (2004). A general framework for unsupervised processing of structured data. *Neurocomputing*, *57*, 3-35.
- [32] Euliano, N. R., & Principe, J. C. (1999). A spatiotemporal memory based on SOMs with activity diffusion. In E. Oja, & S. Kaski (Eds.), *Kohonen Maps* (pp. 253-265). Amsterdam, The Netherlands: Elsevier.
- [33] Horio, K., & Yamakawa, T. (2004). Feedback adaptive subspace self-organizing map for robust spatio-temporal pattern classification. In *International Congress Series* (pp. 125-128).
- [34] Euliano, N. R., & Principe, J. C. (1996). Spatio-temporal self-organizing feature maps. In *Proceedings of the International Conference on Neural Networks* (pp. 1900-1905).
- [35] Mozayyani, N., Alanou, V., Dreyfus, J., & Vaucher, G. (1995). A spatio-temporal data coding applied to Kohonen maps. In *Proceedings of the International Conference on Artificial Neural Networks* (pp. 75-79).
- [36] Verbeek, J. J., Vlassis, N., & Krose, B. J. (2005). Self-organizing mixture models. *Neurocomputing*, *63*, 99-123.
- [37] Lebbah, M., Rogovschi, N., & Bennani, Y. (2007). BeSOM: Bernoulli on self-organizing map. In *International Joint Conference on Neural Networks* (pp. 631-636).
- [38] Graepel, T., Burger, M., & Obermayer, K. (1998). Self-organizing maps: generalizations and new optimization techniques. *Neurocomputing*, *21*, 173-190.

- [39] Bishop, C. M., Svensen, M., & Williams, C. K. I. (1998). GTM: the generative topographic mapping. *Neural Computation*, *10*, 215-234.
- [40] Girolami, M. (2001). The topographic organisation and visualisation of binary data using multivariate-Bernoulli latent variable models. *IEEE Transactions on Neural Networks*, *12*, 1367-1374.
- [41] Kaban, A., & Girolami, M. (2001). A combined latent class and trait model for the analysis and visualization of discrete data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*, 859-872.
- [42] Tino, P., Kaban, A., & Sun, Y. (2004). A generative probabilistic approach to visualizing sets of symbolic sequences. In *Proceedings of the 10th International Conference on Knowledge Discovery and Data Mining* (pp. 701-706).
- [43] Gianniotis, N., & Tino, P. (2008). Visualization of tree-structured data through generative topographic mapping. *IEEE Transactions on Neural Networks*, *19*, 1468-1493.
- [44] Heskes, T. (2001). Self-organizing maps, vector quantization, and mixture modelling. *IEEE Transactions on Neural Networks*, *12*, 1299-1305.
- [45] Barreto, G. de A., Araujo, A., & Kremer, S. (2003). A taxonomy of spatiotemporal connectionist networks revisited: the unsupervised case. *Neural Computation*, *15*, 1255-1320.
- [46] Somervuo, P. (2004). Online algorithm for the self-organizing map of symbol strings. *Neural Networks*, *17*, 1231-1239.
- [47] Kohonen, T., & Somervuo, P. (2002). How to make large self-organizing maps for nonvectorial data. *Neural Networks*, *15*, 945-952.
- [48] Kohonen, T., & Somervuo, P. (1998). Self-organizing maps of symbol strings. *Neurocomputing*, *21*, 19-30.
- [49] Conan-Guez, B., Rossi, F., & Gollob, A. El (2006). Fast algorithm and implementation of dissimilarity self-organizing maps. *Neural Networks*, *19*, 855-863.
- [50] Hammer, B., Hasenfuss, A., Rossi, F., & Strickert, M. (2007). Topographic processing of relational data. In *Proceedings of the 6th International Workshop on Self-Organizing Maps*.
- [51] Baldi, P., & Chauvin, Y. (1996). Hybrid modeling, HMM/NN architectures, and protein applications. *Neural Computation*, *8*, 1541-1565.

- [52] Kang, J., Feng, C.-J., Shao, Q., & Hu, H.-Y. (2007). Prediction of chatter in machining process based on hybrid SOM-DHMM architecture. In *Proceedings of the 3rd International Conference on Computing* (pp. 1004-1013).
- [53] Recanati, C., Rogovschi, N., & Bennani, Y., (2007). The structure of verbal sequences analyzed with unsupervised learning techniques. In *Proceedings of the 3rd Language & Technology Conference*.
- [54] Tsuruta, N., Iuchi, H., Sagheer, A., & Tobely, T. (2003). Self-organizing feature maps for HMM based lip-reading. In *Proceedings of the 7th International Conference on Knowledge-Based Intelligence Information and Engineering Systems* (pp. 162-168).
- [55] Kurimo, M., & Somervuo, P. (1996). Using the self-organizing map to speed up the probability density estimation for speech recognition with mixture density HMMs. In *4th International Conference on Spoken Language* (pp. 358-361).
- [56] Kurimo, M. (1997). Training mixture density HMMs with SOM and LVQ. *Computer Speech and Language*, 11, 321-343.
- [57] Kurimo, M., & Torkkola, K. (1992). Training continuous density hidden Markov models in association with self-organizing maps and LVQ. In *Proceedings of the IEEE Workshop on Neural Networks Signal Processing*.
- [58] Somervuo, P. (2000). Competing hidden Markov models on the self-organizing map. In *Proceedings of the International Joint Conference on Neural Networks* (pp. 169-174).
- [59] Minamino, K., Aoyama, K., & Shimomura, H. (2005). Voice imitation based on self-organizing maps with HMMs. In *Proceedings of the International Workshop on Intelligent Dynamics Humanoids*.
- [60] Rogovschi, N., Lebbah, M., & Bennani, Y. (2010). Learning self-organizing mixture markov models. *Journal of Nonlinear Systems and Applications*, 1, 63-71.
- [61] Baum, L. E. (1972). An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. *Inequalities*, 3, 1-8.
- [62] Krogh, A., Brown, M., Mian, I. S., Sjolander, K., & Haussler, D. (1994). Hidden Markov models in computational biology: applications to protein modeling. *Journal of Molecular Biology*, 235, 1501-1531.

- [63] Jang, J.-S. R., Sun, C.-T., & Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Upper Saddle River, New Jersey: Prentice-Hall.
- [64] Baldi, P. (1995). Gradient descent learning algorithms overview: a general dynamical systems perspective. *IEEE Transactions on Neural Networks*, 6, 182-195.
- [65] Kushner, H. J., & Clark, D. S. (1978). *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Berlin: Springer.
- [66] Kuan, C.-M., & Hornik, K. (1991). Convergence of learning algorithms with constant learning rates. *IEEE Transactions on Neural Networks*, 2, 484-489.
- [67] Baldi, P., & Chauvin, Y. (1994). Smooth on-line learning algorithms for hidden Markov models. *Neural Computation*, 6, 305-316.
- [68] Heskes, T. (1996). Transition times in self-organizing maps. *Biological Cybernetics*, 75, 49-57.
- [69] Heskes, T. (1999). Energy functions for self-organizing maps. In E. Oja, & S. Kaski (Eds.), *Kohonen Maps* (pp. 303-315). Amsterdam, The Netherlands: Elsevier.
- [70] Stolcke, A., & Omohundro, S. (1993). Hidden Markov model induction by Bayesian model merging. In *Advances on Neural Information Processing Systems 5* (pp. 11-18).
- [71] Fujiwara, Y., Asogawa, M., & Konagaya, A. (1994). Stochastic motif extraction using hidden Markov models. In *Proceedings of the 2nd International Conference on Intelligent Systems and Molecular Biology* (pp. 121-129).
- [72] Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77, 257-286.
- [73] Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press.
- [74] Nakama, T. (2009). Theoretical analysis of batch and on-line training for gradient descent learning in neural networks. *Neurocomputing*, 73, 151-159.
- [75] The iProClass Protein Knowledgebase (release 3.79). [Online]. Available: <http://pir.georgetown.edu/>
- [76] Hammer, B., & Hasenfuss, A. (2010). Topographic mapping of large dissimilarity data sets. *Neural Computation*, 22, 2229-2284.

-
- [77] Hammer, B., & Hasenfuss, A. (2007). Relational neural gas. In *Proceedings of the 30th Conference on Artificial Intelligence* (pp. 190-204).
- [78] UCI Machine Learning Repository. [Online]. Available: <http://www.ics.uci.edu/~mllearn/MLRepository.html>
- [79] Sonneburg, S., Ratsch, G., Jagota, A. K., & Muller, K.-R. (2002). New methods for splice site recognition. In *Proceedings of the International Conference on Artificial Neural Networks* (pp. 329-336).
- [80] Li, J., & Wong, L. (2003). Using rules to analyse bio-medical data: a comparison between C4.5 and PCL. In *Proceedings of the 4th International Conference on Web-Age Information and Management* (pp. 254-265).
- [81] Hoche, S., & Wrobel, S. (2002). Scaling boosting by margin-based inclusion of features and relations. In *Proceedings of the 13th European Conference on Machine Learning* (pp. 148-160).
- [82] Rampone, S. (1998). Recognition of splice junctions on DNA sequences by BRAIN learning algorithm. *Bioinformatics*, 14, 676-684.
- [83] Towell, G. G., & Shavlik, J. W. (1994). Knowledge-based artificial neural networks. *Artificial Intelligence*, 70, 119-165.
- [84] Deshpande, M., & Karypis, G. (2002). Evaluation of techniques for classifying biological sequences. In *Proceedings of the 6th Pacific-Asia Conference on Advances on Knowledge Discovery and Data Mining* (pp. 417-431).