



Διπλωματική Εργασία

Στατιστικός Έλεγχος Ποιότητας σε Αυτοσυσχετισμένες
Πολυμεταβλητές Χρονοσειρές μέσω Αναδρομικών Νευρωνικών
Δικτύων

Θεμιστοκλής Νικόλης

ΑΜ: 09116003

Επιβλέπων Καθηγητής: Χρήστος Κουκουβίνος

Τριμελής Επιτροπή:

Καθηγητής ΕΜΠ, Χρήστος Κουκουβίνος

Καθηγήτρια ΕΜΠ, Χρυσή Καρώνη

Επίκ. Καθηγητής ΕΜΠ, Πέτρος Στεφανέας

Ιούλιος, 2021

©(2021) Εθνικό Μετσόβιο Πολυτεχνείο. *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 το συγγραφέα και δεν πρέπει να ερμηνευτεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.



NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF APPLIED MATHEMATICS AND PHYSICAL SCIENCES

Thesis

Statistical Quality Control on Autocorrelated Multivariate Time Series with Recurrent Neural Networks

Themistoklis Nikolis

AM: 09116003

Supervising Professor: Christos Koukouvinos

Τριμελής Επιτροπή:

NTUA Professor, Christos Koukouvinos

NTUA Professor, Chrysiis Karoni

NTUA Assis. Professor, Petros Stefaneas

July, 2021

©(2021) National Technical University of Athens. All rights Reserved. It is prohibited to copy, store and distribute this work, in whole or in part, for commercial purposes. Reprinting, storing and distributing for non-profit, educational or research purposes is permitted, provided the source is acknowledged and this message retained. Questions regarding the use of the work for profit should be addressed to the author. The views and conclusions contained in this document are those of the author and should not be construed as representing the official positions of the National Technical University of Athens.

Περίληψη

Υπό το πλαίσιο των εφαρμόσιμων στατιστικών κλάδων, ο Στατιστικός Έλεγχος Ποιότητας αφορά στην παρακολούθηση διεργασιών με σκοπό την αξιολόγηση και τον έλεγχο τους μέσω εργαλείων βασισμένα σε στατιστικά, μεθόδους λανθάνουσας δομής και μεθόδους ανάλυσης χρονοσειρών. Μια καινή προοπτική, ανεξάντλητης ισχύος, τα νευρωνικά δίκτυα, συνεχώς διευρύνουν το φάσμα εφαρμογής τους και την αποτελεσματικότητα μοντελοποίησης τους, καθιστώντας τα ένα σημαντικό υποψήφιο για την αξιολόγηση διεργασιών.

Η παρούσα διπλωματική εργασία δοκιμάζει την εφαρμογή της δομής ενός αναδρομικού νευρωνικού δικτύου προς μοντελοποίηση πραγματικών και προσομοιωμένων δεδομένων για την παραγωγή υπολοίπων ως χρονοσειρά δεδομένων. Η πρακτική αυτή, εν συνεχεία, συγκρίνεται με τις συνήθεις πρακτικές του Στατιστικού Ελέγχου Ποιότητας καταλήγοντας σε πλέον αξιολογικά αποτελέσματα εφαρμογής αναδρομικών δικτύων.

Η εργασία δομείται ως ακολούθως. Στο πρώτο εισαγωγικό κεφάλαιο θεμελιώνονται οι απαραίτητες αρχές και έννοιες για την ολότητα του περιεχόμενου της εργασίας. Στο δεύτερο κεφάλαιο παρουσιάζονται οι συμβατικές πρακτικές του Στατιστικού Ελέγχου Ποιότητας για την αξιολόγηση διεργασιών. Στο τρίτο κεφάλαιο αναπτύσσονται οι ιδιότητες και τα δημοφιλέστερα μοντέλα χρονοσειρών. Στο τέταρτο κεφάλαιο, αναλύεται η δομή και η λειτουργία των αναδρομικών νευρωνικών δικτύων. Τέλος, στο πέμπτο κεφάλαιο, πραγματοποιείται η σύγκριση και σχολιάζονται τα αποτελέσματα εφαρμογής των μεθόδων πάνω σε πραγματικά και προσομοιωμένα δεδομένα.

Abstract

Under the applicable statistical disciplines, Statistical Quality Control refers to the monitoring of processes for the purpose of their evaluation and control through tools based on statistics, latent structure methods and time series analysis methods. A new perspective, of inexhaustible power, neural networks are constantly expanding their range of application and modeling efficiency, making them an important candidate for process evaluation.

This thesis tests the application of the structure of a recurrent neural network to model real and simulated data for the production of residuals as a time series of data. This practice is then compared with the usual practices of Quality Statistical Control resulting in remarkable results for the application of the recurrent neural network.

The thesis is structured as follows. The first introductory chapter establishes the necessary principles and concepts for the completeness of the content of the work. The second chapter presents the conventional practices of Statistical Quality Control for process evaluation. The third chapter develops the properties and the most popular time series models. In the fourth chapter, the structure and function of recurrent neural networks is analyzed. Finally, in the fifth chapter, the comparison is made and the results of the application of the methods on real and simulated data are commented.

Ευχαριστίες

Ευχαριστώ μεγίστως και εγκαρδίως τον Κύριο Καθηγητή Χρήστο Κουκουβίνο, την Κυρία Αγγελική Λάππα, τον Κύριο Σπύρο Παρασκευά, την Κυρία Φανή Πετροπούλου και την οικογένεια μου.

Vae, puto deus fio.

Περιεχόμενα

1	Εισαγωγή	1
1.1	Στατιστικός Έλεγχος Ποιότητας	1
1.2	Θεμελιώδεις Αρχές	1
2	Διαγράμματα Ελέγχου	5
2.1	Θέσπιση Πολυμεταβλητών Βασικών Μεγεθών	5
2.2	Hotelling T^2	6
2.3	T^2 Διάγραμμα Ελέγχου	7
2.4	T^2 Διάγραμμα Ελέγχου για Μεμονωμένες Παρατηρήσεις	8
2.5	MCUSUM	9
2.6	MEWMA	11
2.7	Συμβατικές Μοντελοποιήσεις	13
2.7.1	PCA	13
2.7.2	Partial Least Squares	14
2.8	Ανασκόπηση	15
3	Χρονοσειρές	17
3.1	Εισαγωγή	17
3.2	Διανυσματικά Μοντέλα Χρονοσειρών	19
3.2.1	VMA	19
3.2.2	VAR	20
3.2.3	VARMA	22
3.2.4	MESSV	23
3.3	Διαγράμματα Υπολοίπων	25
4	Αναδρομικά Νευρωνικά Δίκτυα	30
4.1	Εισαγωγή	30
4.2	Νευρωνικά Δίκτυα	32
4.3	Αναδρομικά Νευρωνικά Δίκτυα - RNN	48
5	Συγκρίσεις Απόδοσης - Αποτελέσματα	62
5.1	Περιγραφή Δεδομένων	62
5.2	mRMR	63
5.3	Σύσταση πειράματος	66
5.3.1	Γενική Βάση - Ειδικό κριτήριο	66
5.3.2	Το πρόβλημα της μετατόπισης	67
5.3.3	Τρόπος αξιολόγησης διαγραμμάτων σε πραγματικά δεδομένα	70

5.3.4	Εφαρμογή της μεθόδου mRMR στα πραγματικά δεδομένα	71
5.3.5	Περιγραφή δομής Νευρωνικού Δικτύου	71
5.3.6	Επιλογή παραμέτρων	74
5.3.7	Επιλογή Παραμέτρων για το RNN	75
5.4	Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές	77
5.5	Αξιολόγηση Διαγραμμάτων σε Πραγματικά Δεδομένα	111
5.6	Επίλογος - Συμπεράσματα	112

List of Figures

1	Ένας τυπικός νευρώνας	32
2	Συνάρτηση Ενεργοποίησης ReLU	33
3	Συνάρτηση Ενεργοποίησης PReLU	34
4	Συνάρτηση Ενεργοποίησης Logistic Sigmoid	35
5	Παράγωγος Logistic Sigmoid	35
6	Βαθμωτή κάθοδος	46
7	Κωδικοποίηση και συμβολισμός εισόδου - εξόδου	49
8	Συνολική δομή ενός δικτύου RNN	51
9	Κόμβος RNN	51
10	Λειτουργία κόμβου RNN	53
11	Οπισθοδιάδοση σε RNN	54
12	Gated Recurrent Unit	57
13	Long Short Term Memory Node	59
14	Δομή δικατευθυντικού RNN	60
15	Δομή ενός απλού Deep RNN	61
16	Δομή της επιλεγμένης σύνθεσης νευρωνικού δικτύου	73
17	Σύγκριση απόδοσης Κλασσικών Διαγραμμάτων με Διαγράμματα Υπολοίπων σε VARMA(1,0,0) διαδικασία	107
18	Σύγκριση απόδοσης Κλασσικών Διαγραμμάτων με Διαγράμματα Υπολοίπων σε VARMA(1,0,1) διαδικασία	108
19	Σύγκριση απόδοσης Κλασσικών Διαγραμμάτων με Διαγράμματα Υπολοίπων σε VARMA(2,0,1) διαδικασία	109

1 Εισαγωγή

1.1 Στατιστικός Έλεγχος Ποιότητας

Σε αυτήν την ενότητα παρουσιάζονται οι βασικές αρχές και το αντικείμενο του στατιστικού ελέγχου ποιότητας. Ο στατιστικός έλεγχος ποιότητας εισηγήθηκε στην γενική βιομηχανική δραστηριότητα περί το 1920 με τον Walter A. Stewart, στην επιχείρηση Bell Laboratories σε μια πρώτη προσπάθεια για δειγματοληπτική εξέταση με λογική στατιστική βάση. Βέβαια, ο Δρ. Vilfredo Pareto πρώτος έθεσε έναν πυλώνα συστηματικής πραγματικότητας με την διατύπωση πως πολλές αποτυχίες έχουν λίγες και κοινές αιτίες. Προηγήθηκε δηλαδή της ανάλυσης αρχικών αιτιών (Root Cause Analysis, RCA). Η εμφώλευση του RCA στον PCA ήταν σχεδόν άμεση και φυσική, με αποτέλεσμα να μην γίνεται λόγος συγκεκριμένα για ανάλυση αρχικών αιτιών αλλά αυτόματα να υπάγεται τούτη στον στατιστικό έλεγχο ποιότητας. Με το πέρας του Β' Παγκοσμίου Πολέμου και την τεχνολογική πρόοδο, ο στατιστικός έλεγχος ποιότητας είναι το ακρολίθι της γενικότερης και ευρύτερης έννοιας "Συνολική Διαχείριση Ποιότητας" (Total Quality Management, TQM), που περιβάλλει ολοκληρωτικά την έννοια της παραγωγής, ξεκινώντας από την αλυσίδα προμηθειών και λήγωντας με τον κύκλο ζωής του προϊόντος.

Ο στατιστικός έλεγχος ποιότητας, όπως έχει διαμορφωθεί, πραγματεύεται την παρακολούθηση διεργασιών (process monitoring) και ποιότητας προϊόντων καθώς και την επιτήρηση της γενικότερης διακύμανσης της διεργασίας μέσω μερικών καίριων διεργασιών μεταβλητών. Περισσότερο στοιχειοθετημένος γύρω από το προϊόν και την βελτίωση της διεργασίας με κίνητρο την μείωση κόστους, τον περιορισμό ελαττωματικών μονάδων τανύζοντας ανέκαθεν στην ανάπτυξη, διόρθωση και βελτίωση της διεργασίας μέσω αύξησης της αξιοπιστίας και ελάττωση της διακύμανσης. Σε έλεγχο διεργασιών πάσας φύσης δηλώνονται οι ελεγχουσυναρτήσεις, στον στατιστικό έλεγχο ποιότητας ωστόσο απουσιάζουν, μα το "έλεγχος" δικαιώνεται μερικώς κατά την σύμβαση ότι ο επιβλέπων της διεργασίας υλοποιεί την ελεγχουσυνάρτηση.

1.2 Θεμελιώδεις Αρχές

Το δείγμα που λαμβάνεται από την παραγωγική διαδικασία, έχει την μορφή μετρήσεων, είτε ενός χαρακτηριστικού του προϊόντος, για παράδειγμα η ποσότητα γάλακτος μέσα στο μπουκάλι, είτε πολλών, για παράδειγμα η διάμετρος και το μήκος μιας ράβδου αλουμινίου. Απαίτηση αποτελεί οι μετρήσεις αυτές να κυμαίνονται γύρω από την τιμή στόχο που θέτει ο κατασκευαστής, με την αποτυχία της διεργασίας να κρίνεται όταν παρεκκλίνουν παραπάνω από τα αντίστοιχα άνω και κάτω φράγματα ανοχής. Οι διακριτές αυτές μετρήσεις σπάνια είναι σταθερές και ίσες με την προκαθορισμένη τιμή στόχο και συνήθως εμφανίζουν ελαφρές τυχαίες διακυμάνσεις μεταξύ τους, εξού και η στοχαστική φύση τους. Σημαντικά μεγέθη της διεργασίας, αναπόσπαστα από οποιαδήποτε στατιστική ανάλυση είναι ο μέσος και η διασπορά της. Ιδανικά, όσο μεγαλύτερο το δείγμα, τόσο κοντύτερα στην επιθυμητή τιμή στόχο διαγράφεται ο μέσος και

τόσο αποσβαίνει η διασπορά. Κατά την αξιολόγηση και παρακολούθηση διεργασιών έχει διασχευασθεί η υπόθεση ότι και ο μέσος και η διασπορά είναι σταθερά. Ο συνήθης συμβολισμός για την αναμενόμενη τιμή - μέσο διεργασίας είναι ο ακόλουθος:

$$E[x] = \bar{x}$$

Ο δειγματικός μέσος \hat{x} παρέχει την εκτίμηση, με τρόπο συνεπή, του πραγματικού μέσου της διεργασίας αν αυτός δεν είναι a priori γνωστός. (Θα συμβολίζεται απλά με \hat{x} .)

$$\hat{x} = \frac{1}{n} \sum_{x \in X} x, \quad \hat{x} \xrightarrow{n \rightarrow \infty} \bar{x}$$

Παραπάνω με X συμβολίζεται το δείγμα, με x οι δειγματικές μονάδες και με n το μέγεθος δείγματος. Οι δειγματικές μονάδες δύναται επομένως να εκφραστούν σαν το άθροισμα του μέσου (ή της εκτίμησης του) με μια στοχαστική διακύμανση γύρω από το μέσο (s_0). Δηλαδή:

$$s = \bar{x} + s_0$$

Αυτή η διακύμανση, ουσιαστικά αποτυπώνεται στα συνήθη διαγράμματα ελέγχου και συγκρίνεται με τα εκάστοτε όρια ελέγχου που απορρέουν από το σύνολο αναφοράς της υπό εξέταση μεταβλητής. Η διασπορά, σε περιπτώσεις που δεν γινώσκεται, εκτιμάται από το δείγμα, τυπικά, από τον ακόλουθο συνεπή εκτιμητή.

$$\hat{\sigma}^2 = s^2, \quad s^2 = \frac{1}{n-1} \sum_{x \in X} (x - \hat{x})^2, \quad s^2 \xrightarrow{n \rightarrow \infty} \sigma^2$$

Ακόμα μία υπόθεση που πραγματοποιείται είναι ότι σε περίπτωση που δεν γινώσκεται η κατανομή των μετρήσεων, τότε οι μετρήσεις ακολουθούν την κανονική κατανομή με μέσο \bar{x} και διασπορά σ^2 , ή σε περίπτωση που οι παράμετροι δεν είναι γνωστοί, από τις αντίστοιχες εκτιμήσεις τους. Δηλαδή:

$$x \sim N(\bar{x}, \sigma^2) \longrightarrow f(x) = \frac{1}{\sqrt{2\sigma^2\pi}} \exp\left(-\frac{(x - \bar{x})^2}{2\sigma^2}\right)$$

Επόμενος είναι και ο τρόπος με τον οποίο κρίνεται αν μια διεργασία είναι εντός ή εκτός ελέγχου. Για παράδειγμα, έστω ότι μια διεργασία ελέγχεται με άνω και κάτω όρια ανοχής a και b αντίστοιχα. Τότε θα ισχύει ότι:

$$F(a, b) = P[a \leq x \leq b] = \frac{1}{\sqrt{2\sigma^2\pi}} \int_a^b \exp\left(-\frac{(x - \bar{x})^2}{2\sigma^2}\right)$$

Τα a και b είναι τα κάτω (LCL) και αντίστοιχα άνω όρια ελέγχου (UCL) της διεργασίας με την κεντρική γραμμή (CL) να αντιπροσωπεύει την μέση τιμή των μετρήσεων του ποιοτικού χαρακτηριστικού υπό παρακολούθηση (μ ή \hat{x}), έχοντας επικρατήσει η τιμή $a = -3\sigma$ και $b = 3\sigma$ στα Shewhart διαγράμματα, δηλαδή το επίπεδο εμπιστοσύνης στην κανονική κατανομή να είναι $\alpha = 0.0027$. Στα Shewhart διαγράμματα έχουν θεσμοθετηθεί 6 προϋποθέσεις που οφείλει να μην παραβιάζει το δείγμα για να θεωρείται η διεργασία εντός ελέγχου (in-control):

1. Ένα σημείο εκτός ορίων ελέγχου
2. Δύο ή τρία διαδοχικά σημεία εκτός $\pm 2\sigma$ ορίων
3. Τέσσερα από πέντε διαδοχικά σημεία εκτός $\pm 1\sigma$ ορίων
4. Εφτά διαδοχικά σημεία ομοεπίπεδα σε σχέση με την κεντρική γραμμή του διαγράμματος
5. Δέκα από έντεκα διαδοχικά σημεία ομοεπίπεδα σε σχέση με την κεντρική γραμμή του διαγράμματος
6. Εφτά διαδοχικά γνησίως μονότονα σημεία

Μια διεργασία καλείται εκτός ελέγχου (out-of-control) αν δεν ικανοποιείται τουλάχιστον ένα από τα παραπάνω μέτρα. Κατά γενική αποδοχή, μια διεργασία εντός ελέγχου κατατάσσεται σε 4 κατηγορίες ανάλογα με την επίδοση της και το διάγραμμα ελέγχου της: *ιδανική, θετική, αναξιόπιστη, παραχώδης*. Μια ιδανική διεργασία παρουσιάζει σταθερότητα, με την μέση τιμή και την διασπορά της να μην διαφοροποιούνται από δείγμα σε δείγμα και από χρονικό πλαίσιο σε χρονικό πλαίσιο. Μια θετική διεργασία παρουσιάζει επίσης σταθερότητα, ωστόσο τα όρια ελέγχου είναι πλατύτερα και δεν ταυτίζονται με τα όρια ανοχής του κατασκευαστή. Μια αναξιόπιστη διεργασία είναι ασταθής με τα όρια ελέγχου να είναι μικρότερα ή ίσα αυτών της ανοχής. Τέλος, μια παραχώδης διεργασία είναι ασταθής με πλατύτερα όρια ελέγχου από αυτά της ανοχής.

Μέσο Μήκος Ροής

Προκειμένου να επιλεγεί επαρκές και αποδοτικό δείγμα για οικονομικούς ή διαχειριστικούς λόγους καθότι το προϊόν δύναται να είναι μιας χρήσης και να καταστρέφεται με την διαδικασία ελέγχου, χρησιμοποιείται το μέσο μήκος ροής (ARL) για τον καθορισμό της συχνότητας και της ποσότητας της δειγματοληψίας. Το ARL δηλώνει τον αναμενόμενο αριθμό δειγμάτων για την εμφάνιση ενός σημείου εκτός ελέγχου (συνήθως αυτό αφορά μονάχα την πρώτη από τις συνθήκες καθορισμού μιας εντός ελέγχου διεργασίας) και ορίζεται ως:

$$ARL = \frac{1}{p}$$

, όπου p η πιθανότητα να βρεθεί ένα σημείο εκτός ελέγχου. Στην πράξη πραγματοποιείται ο εξής έλεγχος υποθέσεων:

$$H_0: \text{Η διεργασία είναι εντός ελέγχου} \quad H_1: \text{η διεργασία είναι εκτός ελέγχου}$$

Στον παραπάνω έλεγχο υποθέσεων το **Σφάλμα Τύπου I**, α , συνδέεται με το εντός ελέγχου μήκος ροής, ARL_0 :

$$ARL_0 = \frac{1}{p} = \frac{1}{\alpha}$$

$$\alpha = \int_{-\infty}^{LCL} f_0(x)dx + \int_{UCL}^{\infty} f_0(x)dx$$

, όπου η $f_0(\bullet)$ η συνάρτηση πυκνότητας πιθανότητας του υπό εξέταση ποιοτικού χαρακτηριστικού. Αντιστοίχως για μια διεργασία εκτός ελέγχου λόγω μετατοπισμένου μέσου ορίζεται το εκτός ελέγχου μήκος ροής, ARL_1 , το οποίο συνδέεται με το **Σφάλμα Τύπου II**, β , ως εξής:

$$ARL_1 = \frac{1}{p} = \frac{1}{1 - \beta}$$

$$\beta = \int_{LCL}^{UCL} f_1(x)dx$$

, όπου η $f_1(\bullet)$ είναι η συνάρτηση πυκνότητας πιθανότητας που αντιστοιχεί στην διεργασία του μετατοπισμένου μέσου. Στόχος της παραγωγής είναι η ελαχιστοποίηση των σφαλμάτων α και β , δηλαδή η μεγιστοποίηση του ARL_0 και η ελαχιστοποίηση του ARL_1 .

Πολυμεταβλητότητα

Η εφαρμογή μονομεταβλητών μεθόδων αξιολόγησης της διεργασίας για κάθε μία από τις συνιστώσες της αποβαίνει επισφαλής, αναξιόπιστη και μάταια στην πλειονότητα των περιπτώσεων, καθώς σε μια διεργασία παραβάλλονται πλήθος εγγενών χαρακτηριστικών που την περιγράφουν, καθιστώντας χρεία την επίταξη κατάλληλων μεθόδων μεταχείρισης του πλήθους παραγόντων. Η εύρεση και η εφαρμογή τέτοιων μεθόδων αναλαμβάνεται από το πεδίο του πολυμεταβλητού στατιστικού ελέγχου ποιότητας.

2 Διαγράμματα Ελέγχου

2.1 Θέσπιση Πολυμεταβλητών Βασικών Μεγεθών

Κατ' αναλογία με την μονομεταβλητή περίπτωση, η προσέγγιση της κατανομής των δεδομένων μέσω της κανονικής κατανομής επεκτείνεται μέσω της πολυμεταβλητής κανονικής κατανομής. Για την παρουσίαση της απαιτείται ο καθορισμός ενός σαφούς συστήματος συμβολισμών. Έστω ότι μια παραγωγική διαδικασία εμφανίζει p ποιοτικά χαρακτηριστικά, δηλαδή p μεταβλητές. Τότε $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ το διάνυσμά τους, $\boldsymbol{\mu}' = [\mu_1, \mu_2, \dots, \mu_p]$ το διάνυσμα των μέσων των X και $\boldsymbol{\Sigma}$ ο πίνακας συνδιακύμανσής τους.

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

Ως εκ τούτου η πολυμεταβλητή κανονική συνάρτηση πυκνότητας πιθανότητας δίνεται από την σχέση:

$$f(\mathbf{X}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})\right)$$

Παρόμοια ορίζονται το διάνυσμα του δειγματικού μέσου $\bar{\mathbf{X}}$ και ο δειγματικός πίνακας συνδιακύμανσης, που είναι συνεπείς και αμερόληπτες εκτιμήτριες του μέσου και της συνδιακύμανσης, για ένα τυχαίο δείγμα, $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ από την πολυμεταβλητή κατανομή:

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i, \quad \mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n n(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$$

Σημειώνεται ότι κατά την αξιολόγηση της διεργασίας ο μέσος και οι συνδιακυμάνσεις θεωρούνται σταθερές.

2.2 Hotelling T^2

Το Hotelling T^2 αποτελεί ανάλογο των Shewhart διαγραμμάτων και εισηγήθηκε περί το 1947 από τον Harold Hotelling με την κατασκευή ενός στατιστικού (T^2) δυνητικά ικανού για την απεικόνιση πολυμεταβλητών παρατηρήσεων στο επίπεδο. Χρησιμεύει στην ανίχνευση μετατοπίσεων από το μέσο και δύναται να χρησιμοποιηθεί είτε σε δεδομένα υπό μορφή υπο-ομάδων είτε σε μεμονωμένες παρατηρήσεις.

Υπο-ομάδες

Η στατιστική συνάρτηση που χρησιμοποιείται για τον έλεγχο των μέσων, μ_1 και μ_2 , δύο ποιοτικών χαρακτηριστικών X_1, X_2 από κοινού κατανεμημένων σύμφωνα με την διμεταβλητή κανονική κατανομή είναι η ακόλουθη:

$$\chi_0^2 = \frac{n}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} [\sigma_2^2 (\bar{X}_1 - \mu_1)^2 + \sigma_1^2 (\bar{X}_2 - \mu_2)^2 - 2\sigma_{12} (\bar{X}_1 - \mu_1)(\bar{X}_2 - \mu_2)]$$

και ακολουθεί την χ_2^2 κατανομή. Βάσει αυτής κατασκευάζεται το χ^2 διάγραμμα ελέγχου με κάτω όριο ελέγχου ίσο με 0, ενώ το άνω όριο ισούται με $\chi_{\alpha,2}^2$.

$$LCL = 0 \quad UCL = \chi_{\alpha,2}^2$$

Σε γενική βάση το χ_0^2 δίνεται από τον τύπο:

$$\chi_0^2 = n(\bar{\mathbf{X}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu})$$

Η διαδικασία παρακολούθησης της διεργασίας δύναται να παρασταθεί γραφικά μέσω της έλλειψης ελέγχου (control ellipse), της οποίας το σχήμα απορρέει από την εξάρτηση των χαρακτηριστικών, αν είναι μηδενική τότε πρόκειται για κύκλο, ενώ αν είναι απολύτως εξαρτημένα, τετριμμένα σχηματίζεται γραμμή (προφανές, αφού η τιμή του ενός χαρακτηριστικού καθορίζει πλήρως την τιμή του άλλου). Αν όλα τα σημεία βρίσκονται εντός της έλλειψης τότε η διαδικασία κρίνεται εντός ελέγχου. Καθώς μια τέτοια παράσταση δεν διατηρεί την χρονική ακολουθία των μετρήσεων, προτείνεται η χρήση του χ^2 διαγράμματος ελέγχου για τον εντοπισμό συστηματικών αποκλίσεων ή άλλων μη τυχαίων μοτίβων.

Πρακτικά, ο μέσος και οι συνδιακυμάνσεις μιας διεργασίας αμφιγνοούνται και εκτιμούνται από την ανάλυση προκαταρκτικών δεδομένων. Για m δείγματα μεγέθους n με p ποιοτικά χαρακτηριστικά υπολογίζονται οι δειγματικοί μέσοι και διακυμάνσεις ως ακολούθως:

$$\bar{X}_{jk} = \frac{1}{n} \sum_{i=1}^n X_{ijk} \quad S_{jk}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_{ijk} - \bar{X}_{jk})^2 \quad \begin{cases} j=1,2,\dots,p \\ k=1,2,\dots,m \end{cases}$$

, όπου η X_{ijk} είναι η i -οστή παρατήρηση για το j -οστό ποιοτικό χαρακτηριστικό στο k -οστό δείγμα. Με προφανή τρόπο προκύπτει και η συνδιακύμανση μεταξύ δύο χαρακτηριστικών στο k -οστό δείγμα:

$$S_{jhk}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_{ijk} - \bar{X}_{jk})(X_{ihk} - \bar{X}_{hk}) \quad \begin{cases} j=1,2,\dots,p \\ j \neq h \end{cases}$$

Πλέον λαμβάνοντας τους μέσους όρους ανά τα m δείγματα προσδιορίζονται οι εκτιμήσεις του μέσου και της συνδιακύμανσης ως:

$$\bar{X}_j = \frac{1}{m} \sum_{i=1}^m \bar{X}_{jk} \quad j = 1, 2, \dots, p \Rightarrow$$

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}} = [\bar{X}_1, \bar{X}_2, \dots, \bar{X}_p]$$

$$\bar{S}_j = \frac{1}{m} \sum_{i=1}^m \bar{S}_{jk} \quad j = 1, 2, \dots, p, \quad \bar{S}_{jh} = \frac{1}{m} \sum_{i=1}^m \bar{S}_{jhk} \quad j \neq h \Rightarrow$$

$$\hat{\boldsymbol{\Sigma}} = \mathbf{S} = \begin{bmatrix} \bar{S}_1^2 & \bar{S}_{12} & \bar{S}_{13} & \cdots & \bar{S}_{1p} \\ & \bar{S}_2^2 & \bar{S}_{23} & \cdots & \bar{S}_{2p} \\ & & \bar{S}_3^2 & & \bar{S}_{3p} \\ & & & \ddots & \\ & & & & \bar{S}_p^2 \end{bmatrix}$$

2.3 T^2 Διάγραμμα Ελέγχου

Από τις παραπάνω σχέσεις, αντικαθιστώντας τις εκτιμώμενες τιμές για την μέση τιμή και τις συνδιακυμάνσεις προκύπτει το στατιστικό T^2 :

$$T^2 = n(\bar{\mathbf{X}} - \bar{\mathbf{X}})' \mathbf{S}^{-1} (\bar{\mathbf{X}} - \bar{\mathbf{X}})$$

Το διάγραμμα που προκύπτει από το στατιστικό T^2 είναι αμετάβλητο από πλευράς κατεύθυνσης, δηλαδή η μετατόπιση στο μέσο εξαρτάται μόνο από το μέγεθος και όχι από την κατεύθυνση. Εδώ τα όρια ελέγχου διαφοροποιούνται ανάλογα με την φάση χρήσης διαγραμμάτων ελέγχου. Η Φάση I αφορά

2.4 T^2 Διάγραμμα Ελέγχου για Μεμονωμένες Παρατηρήσεις

στην λήψη ενός εντός ελέγχου συνόλου παρατηρήσεων, ενώ η Φάση II στην θέσπιση ορίων για μελλοντική παρακολούθηση της παραγωγής. Έτσι τα αντίστοιχα όρια διαμορφώνονται ως εξής:

$$\begin{cases} \text{Φάση I:} & \begin{cases} UCL = \frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1} \\ LCL = 0 \end{cases} \\ \text{Φάση II:} & \begin{cases} UCL = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1} \\ LCL = 0 \end{cases} \end{cases}$$

Για επαρκώς μεγάλα δείγματα χρησιμοποιείται $UCL = \chi_{\alpha,p}^2$ αμφότερα φάσεων. Οι Jensen et al. (2006) και οι Lowry και Montgomery (1995) επισημαίνουν την σημαντικότητα επιλογής ορθού και συνήθως αυξημένου μεγέθους δείγματος ανάλογα τον αριθμό των χαρακτηριστικών και των μεγεθών των ομάδων του δείγματος αποσκοπώντας αφενός στην αναμενόμενη απόδοση ARL της Φάσης II και αφετέρου για την ακριβή προσέγγιση ορίων από την χ^2 κατανομή.

2.4 T^2 Διάγραμμα Ελέγχου για Μεμονωμένες Παρατηρήσεις

Όταν τα m δείγματα έχουν μέγεθος $n = 1$ και αφορούν p ποιοτικά χαρακτηριστικά, το στατιστικό T^2 μορφοποιείται ως εξής:

$$T^2 = n(\mathbf{X} - \bar{\mathbf{X}})' \mathbf{S}^{-1} (\mathbf{X} - \bar{\mathbf{X}})$$

Παρόμοια μεταβάλλονται και τα όρια για τις Φάσεις I και II:

$$\begin{cases} \text{Φάση I:} & \begin{cases} UCL = \frac{(m-1)^2}{m-p} \beta_{\alpha,p/2,(m-p+1)/2} \\ LCL = 0 \end{cases} \\ \text{Φάση II:} & \begin{cases} UCL = \frac{p(m+1)(\mu-1)}{m^2-mp} F_{\alpha,p,m-p} \\ LCL = 0 \end{cases} \end{cases}$$

, όπου $\beta_{\alpha,p/2,(m-p+1)/2}$ το άνω α ποσοστημόριο της $\text{Beta}(p/2, (m-p+1)/2)$ κατανομής. Κατά την Φάση II για μεγάλα δείγματα ($m > 100$) το άνω όριο ελέγχου δύναται να αντικατασταθεί από τις προσεγγίσεις:

$$UCL = \frac{p(m-1)}{m-p} F_{\alpha,p,m-p}$$

ή

$$UCL = \chi_{\alpha,p}^2$$

Μείζονα μέριμνα συνιστά η εύστοχη εκτίμηση του πίνακα συνδιασποράς από το δείγμα, ζήτημα που απασχόλησε ιδίως τους Williams, Woodall, Birch, and Sullivan (2006), στην προσπάθεια προσδιορισμού ενός ανεκτικού αμερόληπτου και συνεπή εκτιμητή στις ακραίες τιμές. Τελικώς οι προτεινόμενες και πιο συνήθεις εκτιμήτριες οι:

$$\mathbf{S}_1^2 = \frac{1}{m-1} \sum_{i=1}^m (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$$

$$\mathbf{S}_2^2 = \frac{1}{2} \frac{\mathbf{V}'\mathbf{V}}{m-1} \text{ για } \mathbf{V} = \begin{bmatrix} (\mathbf{X}_2 - \mathbf{X}_1)' \\ (\mathbf{X}_3 - \mathbf{X}_2)' \\ \vdots \\ (\mathbf{X}_m - \mathbf{X}_{m-1})' \end{bmatrix}$$

Σημειώνεται το αυταπόδεικτο, ότι οι εκτιμήτριες \mathbf{S}_1^2 και \mathbf{S}_2^2 προσδίδουν διαφορετικές τιμές στο στατιστικό T^2 με την \mathbf{S}_2^2 να έχει σε γενικές γραμμές στενότερα όρια αλλά παρεμφερή ευαισθησία με την \mathbf{S}_1^2 .

2.5 MCUSUM

Το Multivariate Cumulative Sum (MCUSUM) διάγραμμα ελέγχου είναι η πολυμεταβλητή επέκταση των CUSUM διαγραμμάτων. Αφορούν σε Φάση II παρακολούθηση διεργασίας για τον μέσο και τη συνδιακύμανση, κουβαλώντας την ορμή της αδράνειας των προηγηθέντων παρατηρήσεων εμφανίζουν ανώτερη απόδοση στην ανίχνευση εκτός ελέγχου τιμών και ευαισθησία σε μικρές μετατοπίσεις. Για αυτούς τους λόγους χρήζουν σημαντικής εφαρμογής στον παραγωγικό τομέα και προτιμούνται έναντι των T^2 διαγραμμάτων. Εισηγήθηκαν από τον Page (1961), σαν μια άμεση γενίκευση της μονομεταβλητής περίπτωσης, ωστόσο εν συνεχεία διατυπώθηκαν πιο ακριβείς και συνεπείς τροποποιήσεις: Woodall και Ncube (1985), Healy (1987), Crosier (1988), Pignatiello and Runger (1990). Παρακάτω παρουσιάζονται τα δύο διαγράμματα με την καλύτερη *ARL* απόδοση (Crosier και Pignatiello, Runger). Για την παρακολούθηση

διεργασιών σχηματίζονται τα στατιστικά T_i^2 και συγκρίνονται με το άνω όριο του ελέγχου, $UCL = h$. Η διαδικασία υπολογισμού των χαρακτηριστικών εξαρτάται από τις υπερπαραμέτρους k , h και n_0 , οι οποίες καθορίζονται από τον ερευνητή είτε άμεσα ή μέσω βελτιστοποίησης για την εκάστοτε παραγωγική διαδικασία. Κρίνεται απαραίτητο να σημειωθεί ότι αποτελεί σημαντικό και άκρως καίριο καθήκον του ερευνητή η ορθή υπόθεση κατανομής των δεδομένων, με προβληματική και μειωμένη απόδοση των MCUSUM σε εσφαλμένη υπόθεση πολυμεταβλητής κανονικής κατανομής, αν και η τελευταία εκτιμά το σύνολο των δεδομένων σε περιπτώσεις που παραμένει άγνωστη η πραγματική κατανομή.

Crosier

$$T_i^2 = \left[S_i' \left(\frac{\Sigma}{n} \right)^{-1} S_i \right]^{\frac{1}{2}}$$

$$S_i = \begin{cases} 0, & C_i \leq k \\ (S_{i-1} + \bar{\mathbf{X}}_i - \boldsymbol{\mu}_0) \left(1 - \frac{k}{C_i} \right), & C_i > k \end{cases}$$

, όπου $S_0 = 0$, $k > 0$ και

$$C_i = \left[(S_{i-1} + \bar{\mathbf{X}}_i - \boldsymbol{\mu}_0)' \left(\frac{\Sigma}{n} \right)^{-1} (S_{i-1} + \bar{\mathbf{X}}_i - \boldsymbol{\mu}_0) \right]^{\frac{1}{2}}$$

Τα όρια ελέγχου διαμορφώνονται ως:

$$UCL = h$$

$$LCL = 0$$

Pignatiello and Runger

$$T_i^2 = \max \left\{ 0, \left[S_i' \left(\frac{\Sigma}{n} \right)^{-1} S_i \right]^{\frac{1}{2}} - kn_i \right\}$$

$$, \text{ όπου } S_i = \sum_{i-n_i+1}^i (\bar{\mathbf{X}}_i - \boldsymbol{\mu}_0)$$

$$\text{ και } n_i = \begin{cases} n_{i-1} + 1, & T_{i-1}^2 > 0 \\ 1, & T_{i-1}^2 \leq 0 \end{cases}$$

Τα όρια ελέγχου διαμορφώνονται ως:

$$UCL = h$$

$$LCL = 0$$

Σε περίπτωση που ένα σήμα παρατηρηθεί εκτός των ορίων ελέγχου, η διεργασία κρίνεται εκτός ελέγχου.

2.6 MEWMA

Το Multivariate Exponentially Weighted Moving Average (MEWMA) διάγραμμα εισηγήθηκαν οι Lowry et. al το 1992, επεκτείνοντας το Exponentially Weighted Moving Average (EWMA) σε μονομεταβλητές περιπτώσεις. Πρόκειται για διαγράμματα δίχως κατεύθυνση, δηλαδή η απόδοση της διεργασίας εξαρτάται μονάχα από την οποιαδήποτε μετατόπιση στο μέσο. Τείνουν να είναι ευαίσθητα στην ανίχνευση μικρών και μεσαίων μετατοπίσεων και χρησιμοποιούνται ευρέως, ιδίως στην μορφή των Fuzzy EWMA. Όπως και τα MCUSUM διαγράμματα αφορούν την Φάση II της διεργασίας, δηλαδή για μελλοντική παρακολούθηση αφού έχει τεθεί εντός ελέγχου. Καθορίζεται από το:

$$\mathbf{Z}_i = \lambda \mathbf{X}_i + (1 - \lambda) \mathbf{Z}_{i-1}$$

, όπου $0 \leq \lambda \leq 1$ και $\mathbf{Z}_0 = 0$. Ο προσδιορισμός πίνακα συνδιασποράς είναι ανάλογος με την μονομεταβλητή περίπτωση και δίνεται από τον τύπο:

$$\boldsymbol{\Sigma}_{\mathbf{Z}_i} = \frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2i}] \boldsymbol{\Sigma}$$

Στο διάγραμμα σχεδιάζεται η ποσότητα:

$$T_i^2 = \mathbf{Z}_i' \boldsymbol{\Sigma}_{\mathbf{Z}_i}^{-1} \mathbf{Z}_i$$

Το κάτω όριο ελέγχου είναι $LCL = 0$, ενώ το άνω όριο $UCL = h$, όπου h υπερπαράμετρος που επιλέγεται ανάλογα με την επίτευξη ζητούμενου ARL. Σημειώνεται ότι για $\lambda = 1$ το MEWMA ταυτίζεται με το Hotelling T^2 διάγραμμα ελέγχου. Δεν σπανίζει η χρήση των χ^2 ορίων του Hotelling διαγράμματος σαν προστατευτικός μηχανισμός απέναντι στην αδράνεια που παρουσιάζει το MEWMA διάγραμμα ελέγχου με αντιστάθμισμα την μείωση του ARL_0 .

Στην εργασία των Alkahatani και Schaffer (2012) αναπτύσσεται θεωρητικά και πρακτικά το Double MEWMA (dMEWMA) διάγραμμα ελέγχου στα πλαίσια της παρακολούθησης διεργασιών. Το dMEWMA παρουσιάζει βελτιωμένη απόδοση σε σχέση με το απλό MEWMA, καθώς επιτυγχάνει μεγαλύτερα μήκη ροής ARL_0 και μικρότερα ARL_1 για μετατοπίσεις στον μέσο, $\delta = (\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})$. Ουσιαστικά οι τιμές των χαρακτηριστικών λειάνονται εις διπλούν με τα \mathbf{Z}_i να προκύπτουν ως εξής:

$$\begin{aligned}\mathbf{Y}_i &= \lambda\mathbf{X}_i + (1 - \lambda)\mathbf{Y}_{i-1} \\ \mathbf{Z}_i &= \lambda\mathbf{Y}_i + (1 - \lambda)\mathbf{Z}_{i-1}\end{aligned}$$

Το στατιστικό T_i^2 ορίζεται όπως προηγουμένως.

Ιδιαίτερου ενδιαφέροντος χαίρει το FEWMA διάγραμμα ελέγχου όπως προτάθηκε από τους Hawkins et al. (2007), το οποίο γενικεύει το λ , αντικαθιστώντας το με τον πίνακα \mathbf{R} επιτρέποντας τον εντοπισμό μετατοπίσεων πιο έντονα σε συγκεκριμένες κατευθύνσεις. Κατά αυτόν τον τρόπο το διάγραμμα περιγράφεται ως:

$$\mathbf{Z}_i = \mathbf{R}\mathbf{X}_i + (\mathbf{I} - \mathbf{R})\mathbf{Z}_{i-1}$$

Με τις διαγωνίους του πίνακα \mathbf{R} να καθορίζονται από τις παραμέτρους λ και c όπως:

$$r_{main} = \frac{\lambda}{1 + (p + 1)c} \quad r_{off} = \frac{c\lambda}{1 + (p - 1)c}$$

Τα FEWMA δύνανται να βελτιώσουν την απόδοση των MEWMA. Αναφέρεται ότι σε μερικές περιπτώσεις, όπου τα ποιοτικά χαρακτηριστικά είναι πολυπληθή, η εκλογή και προσαρμογή ενός MEWMA διαγράμματος στα πιο κύρια χαρακτηριστικά για την παρακολούθηση της διαδικασίας αποβαίνει πιο ακριβής και βελτιωμένη σε απόδοση από την περίπτωση που έχουν ληφθεί όλες οι μεταβλητές που εκφράζουν τα ποιοτικά χαρακτηριστικά.

Ολοκληρώνοντας την παρουσίαση των διαγραμμάτων ελέγχου για πολυμεταβλητές διεργασίες, επισημαίνεται ότι έχουν συνταχθεί διαγράμματα για παρακολούθηση της μεταβλητότητας της διεργασίας όπως

του Alt (1985) που αποτελεί ανάλογο του S^2 , των Shen et al. (2014) μέσω MEWMA και των Hidayatul et al. (2019) μέσω του Max-MCUSUM.

2.7 Συμβατικές Μοντελοποιήσεις

Τα διαγράμματα ελέγχου αποβαίνουν ισχυρά εργαλεία με έντονα πρακτικό χαρακτήρα λόγω της απεικόνισης και της γρήγορης και απλής εφαρμογής τους. Ωστόσο, καθώς αυξάνεται ο αριθμός των χαρακτηριστικών η ικανότητα ανίχνευσης ανωμαλιών κατά την παρακολούθηση της διαδικασίας μειώνεται, δηλαδή αυξάνεται το ARL_1 . Αυτό συμβαίνει επειδή η μετατόπιση είναι αραιωμένη στον πολυδιάστατο χώρο, καθιστώντας δύσκολο τον εντοπισμό της. Μέθοδοι λανθάνουσας δομής (latent structure methods) χρησιμοποιούνται πολλές φορές συμπληρωματικά ή αυτούσια για την εξαγωγή ενός υπόχωρου του αρχικού που εγκιβωτίζει σημαντική πληροφορία. Αναφέρεται ότι τέτοιες μέθοδοι υποθέτουν γραμμική σχέση και ανεξαρτησία μεταξύ των μεταβλητών των μετρήσεων, με το σύνολο της διεργασίας να ακολουθεί την πολυμεταβλητή κανονική κατανομή.

2.7.1 PCA

Έστω παραγωγική διαδικασία με τον πίνακα δεδομένων \mathbf{X} [$n \times p$], n μετρήσεις p χαρακτηριστικών, όπου κάθε χαρακτηριστικό έχει κανονικοποιηθεί έτσι ώστε να έχει μηδενική μέση τιμή και διασπορά ίση με την μονάδα. Τότε ο πίνακας συνδιακύμανσης του \mathbf{X} ορίζεται να είναι:

$$\Sigma = \frac{\mathbf{X}'\mathbf{X}}{n - 1}$$

Τότε συμβολίζοντας με $\lambda_i, i = 1 \dots p$ τις ιδιοτιμές του πίνακα σε φθίνουσα σειρά, υπολογίζονται οι κύριες συνιστώσες (PCs) και τα βάρη που τους αντιστοιχούν, \mathbf{p}_i . Έπειτα επιλέγονται τα k μεγαλύτερα βάρη για την κατασκευή του PCA μοντέλου και ο πίνακας δεδομένων εκφράζεται ως:

$$\mathbf{X} = \sum_{i=1}^k \mathbf{t}_i \mathbf{p}_i' + \mathbf{E}$$

, όπου τα \mathbf{t}_i είναι τα scores, δηλαδή οι συντεταγμένες του κάθε σημείου (μέτρησης) στον χώρο των αξόνων όπως προκύπτουν από την μέθοδο, και \mathbf{E} είναι ένας πίνακας υπολοίπων (Qin 2003). Για κάθε καινούρια παρατήρηση, η πρόβλεψη του μοντέλου είναι:

$$\hat{\mathbf{x}}_{new} = \mathbf{t}_{new}\mathbf{P}' = \mathbf{x}_{new}\mathbf{P}\mathbf{P}'$$

Η διαφορά της πρόβλεψης από την πραγματική τιμή αποτελεί το υπόλοιπο για την καινούρια παρατήρηση: $\mathbf{e}_{new} = \mathbf{x}_{new} - \hat{\mathbf{x}}_{new}$. Αποσκοπώντας στην παρακολούθηση της διεργασίας ορίζονται δύο στατιστικά T^2 και SPE βάσει των \mathbf{t}_{new} και \mathbf{e}_{new} :

$$T_{new}^2 = \mathbf{t}_{new}\mathbf{\Lambda}\mathbf{t}_{new}'$$

$$SPE = \mathbf{e}_{new}\mathbf{e}_{new}'$$

, όπου $\mathbf{\Lambda}$ είναι ο πίνακας των ιδιοτιμών του μοντέλου PCA. Το όριο ελέγχου των παραπάνω στατιστικών καθορίζεται ως εξής:

$$T^2 \leq T_{lim}^2 = \frac{k(n-1)}{n-k} F_{k,n-k,\alpha}$$

$$SPE \leq SPE_{lim} = g\chi_{h,\alpha}^2, \quad g = \frac{v}{2m}, \quad h = 2\frac{m^2}{v}$$

, όπου α το επίπεδο σημαντικότητας και m και v η μέση τιμή και η διασπορά των παρατηρήσεων στο σύνολο εκπαίδευσης. Όταν παρατηρηθεί τιμή που η τιμή της υπερβαίνει την τιμή του στατιστικού, η διαδικασία θεωρείται εκτός ελέγχου.

2.7.2 Partial Least Squares

Η αρχή που διέπει την μέθοδο PCA ομοιάζει με αυτή που εφαρμόζεται στα Μερικά Ελάχιστα Τετράγωνα (Partial Least Squares PLS), με τα τελευταία να περιλαμβάνουν την πληροφορία των ποιοτικών μεταβλητών της διεργασίας. Δοσμένων δύο συνόλων δεδομένων διεργασίας και ποιότητας αντιστοιχώς $\{\mathbf{X}, \mathbf{Y}\}$, η μέθοδος PLS αποσκοπεί στην αποσύνθεση του \mathbf{X} και \mathbf{Y} σε έναν συνδυασμό πινάκων score \mathbf{T} , φόρτου \mathbf{P} και \mathbf{Q} , και βαρών \mathbf{W} . Η σχέση μεταξύ των \mathbf{X} και \mathbf{Y} περιγράφεται από τις εξής εξισώσεις με \mathbf{E} και \mathbf{F} τους πίνακες υπολοίπων:

$$\mathbf{X} = \mathbf{T}'\mathbf{P} + \mathbf{E}$$

$$\mathbf{Y} = \mathbf{T}\mathbf{Q}' + \mathbf{F}$$

Ο πίνακας παλινδρόμησης του μοντέλου μερικών ελάχιστων τετραγώνων ορίζεται ως:

$$\mathbf{R} = \mathbf{W}(\mathbf{P}'\mathbf{W})^{-1}\mathbf{Q}'$$

Δοθέντος μιας νέας μέτρησης από την διεργασία, $(\mathbf{x}_{new}, \mathbf{y}_{new})$, οι κύριες συνιστώσες υπολογίζονται ως:

$$\mathbf{t}_{new} = \mathbf{x}_{new}\mathbf{W}(\mathbf{P}'\mathbf{W})^{-1}$$

Όμοια με πριν υπολογίζονται τα στατιστικά με τα οποία ελέγχεται η διεργασία:

$$T^2 = \mathbf{t}'_{new}\mathbf{S}_{tt}^{-1}\mathbf{t}_{new} = \mathbf{x}'_{new}\mathbf{R}\mathbf{S}_{tt}^{-1}\mathbf{R}'\mathbf{x}_{new}$$

, όπου ο \mathbf{S}_{tt} είναι ο διαγώνιος πίνακας που στεγάζει τα στοιχεία των διασπορών των t , που αν δεν είναι a priori γνωστά εκτιμούνται από το δείγμα μετρήσεων. Η τιμή του στατιστικού T^2 ακολουθεί την χ^2 κατανομή με $n_x - n$ βαθμούς ελευθερίας αν ο πίνακας διασποράς είναι a priori γνωστός, ειδάλως ακολουθεί *scaled - F* κατανομή. Κατ' αναλογία με το SPE υπολογίζεται το στατιστικό Q :

$$Q_\alpha = g\chi_{h,\alpha}^2$$

Και ακριβώς ορίζονται όπως και προηγουμένως οι παράμετροι g και h .

2.8 Ανασκόπηση

Τα μοντέλα του πολυμεταβλητού στατιστικού ελέγχου ποιότητας υποθέτουν, όπως αναφέρθηκε, ότι η κατανομή των ποιοτικών χαρακτηριστικών είναι γκαουσιανή, κάτι που δεν ανταποκρίνεται στην πράξη. Όχι μόνον αυτό, αλλά και σε πλήθος περιπτώσεων τα χαρακτηριστικά δεν είναι ανεξάρτητα μεταξύ τους, μάλιστα στην ανάλυση χρονοσειρών που ακολουθεί και είναι και το κύριο ενδιαφέρον της εργασίας αυτής, η υψηλή αυτοσυσχέτιση καθιστά μειωμένη την απόδοση των συμβατικών μοντέλων και διαγραμμάτων που έχουν αναπτυχθεί. Οι μέθοδοι λανθάνουσας δομής παρουσιάζουν βέβαια δυνατότητα τροποποίησης, έτσι ώστε να μετράζονται τα αρνητικά στοιχεία της μη συμμόρφωσης των δεδομένων στην κανονική κατανομή. Η Ανάλυση Ανεξάρτητων Συνιστωσών (Independent Component Analysis, ICA) διακρίνει ομάδες χαρακτηριστικών που δεν είναι κανονικά κατανομημένες ούτε εξαρτημένες μεταξύ τους, χρησιμοποιούμενη ευρέως σε συνδυασμό με PCA για την εξαγωγή σημαντικών χαρακτηριστικών. Τέλος, μοντέλα λανθάνουσας δύνανται να τροποποιηθούν ώστε να ανταποκρίνονται καλύτερα σε περιπτώσεις που παραβιάζεται η γραμμικότητα μεταξύ των μεταβλητών, για παράδειγμα η μέθοδος πηρούνων PCA (KPCA) (Lee et al. 2004; Choi et al. 2005; Cho et al. 2005), με κύρια ιδέα την απεικόνιση του

χώρου των χαρακτηριστικών μέσω μιας μη γραμμικής απεικόνισης σε χώρο όπου η γραμμικότητα της διεργασίας είναι πιο ισχυρή.

3 Χρονοσειρές

3.1 Εισαγωγή

Στην σύγχρονη εποχή παρατηρείται ένας κατακλυσμός από πληροφορία και δεδομένα που προηγουμένως δεν ήταν διαθέσιμα, και αν ήταν, τότε η υπολογιστική ισχύς των μέσων παρακολούθησης ήταν χαμηλότατη και ανεπαρκής. Από μετοχές μέχρι την παραγωγή πρώτων υλών, οι χρονοσειρές εμφανίζονται παντού και η αποδοτική, ταυτόχρονη και αναλυτική αξιοποίηση τους κρίνεται επιτακτική. Οι μετρήσεις που λαμβάνονται από μια παραγωγική διαδικασία δειγματοληπτούνται με χρονική σειρά, και εκτός αυτού η τιμή τους συνήθως εξαρτάται από τις προηγούμενες τιμές. Για παράδειγμα τους μήνες του καλοκαιριού σε ένα εργοστάσιο κατασκευής μεταλλικών ράβδων, λόγω θερμοκρασίας τα μηχανήματα υπερθερμαίνονται και προκύπτουν ράβδοι με πιο έντονες διακυμάνσεις αν δεν έχει ψυχθεί επαρκώς το μηχανήμα κοπής. Λογικό θα ήταν να πραγματοποιηθεί η υπόθεση ότι και το επόμενο καλοκαίρι αν δεν υπάρξει κάποια αναβάθμιση του ψυκτικού μηχανισμού πως τα μορφολογικά χαρακτηριστικά των ράβδων θα παρουσιάζουν και πάλι αυξημένες διακυμάνσεις. Γίνεται αντιληπτό, με λίγα λόγια πως είναι φυσικό μια προηγούμενη κατάσταση (μέτρηση) να επηρεάζει την επόμενη. Αυτό το φαινόμενο καλείται αυτοσυσχέτιση και τα συνήθη μέσα και διαγράμματα του στατιστικού ελέγχου ποιότητας αδυνατούν να την συλλάβουν επαρκώς και να ξεχωρίσουν την σημαντική πληροφορία που παρέχεται από τις προηγούμενες χρονικές στιγμές. Για το σκοπό αυτό, στο παρόν κεφάλαιο περιγράφονται τα βασικά στοιχεία των πολυμεταβλητών χρονοσειρών, καθώς και τα μέσα τα οποία επιτρέπουν την παρακολούθηση της διεργασίας όταν αυτή είναι αυτοσυσχετισμένη.

Κατά την παρακολούθηση ενός φαινομένου ή μιας διεργασίας, συναντώνται πολλές μεταβλητές $Z_{i,t}$, όπου $i = 1, 2, \dots, m$ και οι παρατηρήσεις λαμβάνονται με χρονική σειρά t . Γίνεται χρήση του διανύσματος $\mathbf{Z}_t = [Z_{1,t}, Z_{2,t}, \dots, Z_{m,t}]'$ προς υπόδειξη του συνόλου των μεταβλητών, με τον χρόνο t να είναι είτε συνεχής (λ.χ. ηλεκτρικά σήματα) είτε διακριτός (λ.χ. ημερήσια τιμή κλεισίματος μιας μετοχής). Στην πράξη, ακόμα και για συνεχή χρόνο λαμβάνονται διακριτές τιμές προς ανάλυση. Στις χρονοσειρές ένα θεμελιώδες φαινόμενο αποτελεί η ύπαρξη συσχέτισης όχι μονάχα μεταξύ των όρων i , αλλά και μεταξύ του t (χρονική συσχέτιση).

Στασιμότητα

Η στασιμότητα χαρακτηρίζει τις χρονοσειρές, στις οποίες οι διακυμάνσεις των τιμών της δεν διαφοροποιούνται με το χρόνο. Γενικεύοντας στην πολυμεταβλητή περίπτωση, στάσιμες καλούνται οι χρονοσειρές αν κάθε συνιστώσα χρονοσειρά είναι στάσιμη και οι δύο πρώτες ροπές της παρουσιάζονται ανεξάρτητες από τον χρόνο. Η στασιμότητα στο διάγραμμα ιστορίας (history diagram), στο οποίο

αναπαρίστανται γραφικά οι μετρήσεις με την χρονική σειρά τους, μεταφράζεται με την απουσία αργών και σταθερών μεταβολών στην τιμή/επίπεδο της χρονοσειράς.

Πίνακες Συσχέτισης και μερικής συσχέτισης

Έστω $\mathbf{Z}_t = [Z_{1,t}, Z_{2,t}, \dots, Z_{m,t}]'$, μια m -διάστατη στάσιμη πραγματική διαδικασία έτσι ώστε $E(Z_{i,t}) = \mu_i \forall i = 1, 2, \dots, m$ και η συνδιακύμανση μεταξύ των $Z_{i,t}$ και $Z_{j,s} \forall i = 1, 2, \dots, m, j = 1, 2, \dots, m$ να είναι συναρτήσεις μονάχα της χρονικής διαφοράς ($s - t$). Επομένως, το διάνυσμα του μέσου δίνεται από:

$$E(\mathbf{Z}_t) = \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{bmatrix},$$

και ο πίνακας συνδιασποράς σε υστέρηση (lag) k από:

$$\begin{aligned} \boldsymbol{\Gamma}(k) &= Cov\{\mathbf{Z}_t, \mathbf{Z}_{t+k}\} = E[(\mathbf{Z}_t - \boldsymbol{\mu})(\mathbf{Z}_{t+k} - \boldsymbol{\mu})'] \\ &= E \begin{bmatrix} Z_{1,t} - \mu_1 \\ Z_{2,t} - \mu_2 \\ \vdots \\ Z_{m,t} - \mu_m \end{bmatrix} [Z_{1,t+k} - \mu_1, Z_{2,t+k} - \mu_2, \dots, Z_{m,t+k} - \mu_m] \\ &= \begin{bmatrix} \gamma_{1,1}(k) & \gamma_{1,2}(k) & \dots & \gamma_{1,m}(k) \\ \gamma_{1,1}(k) & \gamma_{2,2}(k) & \dots & \gamma_{2,m}(k) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m,1}(k) & \gamma_{m,2}(k) & \dots & \gamma_{m,m}(k) \end{bmatrix} \end{aligned}$$

Βάσει αυτού ορίζεται ο πίνακας αυτοσυσχέτισης σε lag k , με $\mathbf{D} = diag[\gamma_{1,1}(0), \gamma_{2,2}(0), \dots, \gamma_{m,m}(0)]$:

$$\boldsymbol{\rho}(k) = \mathbf{D}^{-1/2} \boldsymbol{\Gamma}(k) \mathbf{D}^{-1/2}$$

Αποσύνθεση Χρονοσειρών

Σε γενικές γραμμές μια χρονοσειρά δύναται να αναλυθεί σε τρεις συνιστώσες ως εξής:

$$\mathbf{x}_t = \mathbf{b}_t + \mathbf{s}_t + \mathbf{e}_t$$

, όπου \mathbf{b}_t είναι η συνιστώσα της τάσης, \mathbf{s}_t η συνιστώσα της εποχικότητας/περιοδικότητας για κάποια περίοδο d ($\mathbf{s}_{t-d} = \mathbf{s}_t$) και \mathbf{e}_t η χρονοσειρά των υπολοίπων αν αφαιρεθεί από την παρατηρούμενη σειρά η τάση και η εποχικότητα. Σαν μεγέθη, η τάση και η περιοδικότητα είναι συναρτήσεις του χρόνου και δεν περιέχουν πληροφορία για την δυναμική του συστήματος, δηλαδή δεν μεταφράζουν την σχέση εξάρτησης της παρατήρησης από προηγούμενες. Ανάλογα με τον διερευνητικό σκοπό, άλλες φορές η ανάλυση συγκεντρώνεται πάνω στην εκτίμηση και εξήγηση της τάσης και περιοδικότητας, και άλλες φορές αυτό απαλείφεται και μελετάται η δυναμική του συστήματος.

White Noise Processes

Οι διαδικασίες λευκού θορύβου (White Noise Processes) αποτελούν την πιο απλή μορφή χρονοσειρών και ορίζονται να είναι οι χρονικά ασυσχέτιστες στάσιμες διαδικασίες με μηδενικό μέσο και πεπερασμένη διασπορά. Δηλαδή, δεδομένου του πίνακα Σ των συνδιασπορών ικανοποιούν την ιδιότητα:

$$E[\mathbf{a}_t \mathbf{a}'_{t+k}] = \begin{cases} \Sigma, & k = 0, \\ \mathbf{0}, & k \neq 0 \end{cases}$$

Αν επιπλέον το \mathbf{a}_t ακολουθεί την γκαουσιανή κατανομή τότε η διαδικασία καλείται Gaussian white noise process.

3.2 Διανυσματικά Μοντέλα Χρονοσειρών

3.2.1 VMA

Το διανυσματικό μοντέλο κινητού μέσου (Vector Moving Average) για m χαρακτηριστικά τάξης q περιγράφεται ως:

$$\begin{aligned} \mathbf{Z}_t &= \boldsymbol{\mu} + \mathbf{a}_t - \Theta_1 \mathbf{a}_{t-1} - \dots - \Theta_q \mathbf{a}_{t-q} \\ &= \boldsymbol{\mu} + \Theta_q(B) \mathbf{a}_t \end{aligned}$$

, με το B να είναι ο τελεστής οπισθοδρόμησης (Backshift Operator): $\Theta_q(B) = \mathbf{I} - \Theta_1 B - \dots - \Theta_q B^q$. Πρόκειται δηλαδή για ένα στάσιμο μοντέλο με μέσο: $E(\mathbf{Z}_t) = \boldsymbol{\mu}$ και πίνακα συνδιακύμανσης:

$$\begin{aligned} \Gamma(k) &= E[(\mathbf{Z}_t - \boldsymbol{\mu})(\mathbf{Z}_{t+k} - \boldsymbol{\mu})'] \\ &= \begin{cases} \sum_{j=0}^{q-k} \boldsymbol{\Theta}_j \boldsymbol{\Sigma} \boldsymbol{\Theta}'_{j+k}, & k = 0, 1, \dots, q \\ 0, & k > q \end{cases} \end{aligned}$$

, όπου $\boldsymbol{\Theta}_0 = \mathbf{I}$. Μια προφανής ιδιότητα του μοντέλου αυτού είναι ότι ο πίνακας συνδιασπορών μηδενίζει μετά από την q -οστή υστέρηση.

3.2.2 VAR

Το διανυσματικό μοντέλο αυτοπαλινδρόμησης (Vector AutoRegressive) για m χαρακτηριστικά τάξης p περιγράφεται ως:

$$\mathbf{Z}_t = \boldsymbol{\vartheta}_0 + \boldsymbol{\Phi}_1 \mathbf{Z}_{t-1} + \dots + \boldsymbol{\Phi}_p \mathbf{Z}_{t-p}$$

ή διαφορετικά

$$\boldsymbol{\Phi}_p(B) \mathbf{Z}_t = \boldsymbol{\vartheta}_0 + \mathbf{a}_t$$

Το VAR είναι εμφανώς αντιστρέψιμο μοντέλο και μάλιστα είναι στάσιμο αν οι ρίζες του

$$|\lambda^p \mathbf{I} - \lambda^{p-1} \boldsymbol{\Phi}_1 - \dots - \boldsymbol{\Phi}_p| = 0$$

είναι εντός της μοναδιαίας μπάλας. Μάλιστα τότε:

$$\begin{aligned} \boldsymbol{\mu} &= E(\mathbf{Z}_t) = E(\boldsymbol{\vartheta}_0 + \boldsymbol{\Phi}_1 \mathbf{Z}_{t-1} + \dots + \boldsymbol{\Phi}_p \mathbf{Z}_{t-p}) \\ &= \boldsymbol{\vartheta}_0 + \boldsymbol{\Phi}_1 \boldsymbol{\mu} + \dots + \boldsymbol{\Phi}_p \boldsymbol{\mu} \\ \Rightarrow \boldsymbol{\mu} &= (\mathbf{I} - \boldsymbol{\Phi}_1 - \dots - \boldsymbol{\Phi}_p)^{-1} \boldsymbol{\vartheta}_0 \end{aligned}$$

Επομένως, η στάσιμη διεργασία μπορεί να εκφραστεί ως:

$$\boldsymbol{\Phi}_p(B) \dot{\mathbf{Z}}_t = \mathbf{a}_t$$

, όπου $\dot{\mathbf{Z}}_t = \mathbf{Z}_t - \boldsymbol{\mu}$. Πλέον, ο πίνακας συνδιακύμανσης της διεργασίας μπορεί να εκφραστεί ως εξής:

$$\begin{aligned} \Gamma(k) &= E(\dot{\mathbf{Z}}_t \dot{\mathbf{Z}}_{t+k}') \\ &= E[\dot{\mathbf{Z}}_t (\boldsymbol{\Phi}_1 \dot{\mathbf{Z}}_{t+k-1}' + \boldsymbol{\Phi}_2 \dot{\mathbf{Z}}_{t+k-2}' + \dots + \boldsymbol{\Phi}_p \dot{\mathbf{Z}}_{t+k-p}')] \\ &= \begin{cases} \Gamma(-1)\boldsymbol{\Phi}'_1 + \Gamma(-2)\boldsymbol{\Phi}'_2 + \dots + \Gamma(-p)\boldsymbol{\Phi}'_p + \boldsymbol{\Sigma}, & k = 0 \\ \Gamma(k-1)\boldsymbol{\Phi}'_1 + \Gamma(k-2)\boldsymbol{\Phi}'_2 + \dots + \Gamma(k-p)\boldsymbol{\Phi}'_p, & k > 0 \end{cases} \end{aligned}$$

, το τελευταίο είναι ένα σύστημα γενικευμένων εξισώσεων πινάκων Yule-Walker, από το οποίο οι πίνακες $\boldsymbol{\Phi}'_j$ υπολογίζονται από τα $\Gamma(-j)$ ως εξής:

$$\begin{bmatrix} \boldsymbol{\Phi}'_1 \\ \boldsymbol{\Phi}'_2 \\ \vdots \\ \boldsymbol{\Phi}'_p \end{bmatrix} = \begin{bmatrix} \Gamma(0) & \Gamma(-1) & \Gamma(-2) & \dots & \Gamma(1-p) \\ \Gamma(-1) & \Gamma(0) & \Gamma(-1) & \dots & \Gamma(2-p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Gamma(p-1) & \Gamma(p-2) & \Gamma(p-3) & \dots & \Gamma(0) \end{bmatrix}^{-1} \begin{bmatrix} \Gamma(1) \\ \Gamma(2) \\ \vdots \\ \Gamma(p) \end{bmatrix}$$

Αφού υπολογιστούν τα $\boldsymbol{\Phi}'_j$ κατά τα παραπάνω, τότε ο πίνακας συνδιασποράς προκύπτει:

$$\boldsymbol{\Sigma} = \Gamma(0) - \Gamma(-1)\boldsymbol{\Phi}'_1 + \Gamma(-2)\boldsymbol{\Phi}'_2 + \Gamma(-p)\boldsymbol{\Phi}'_p$$

Αιτιότητα Granger

Η αιτιότητα Granger αποτελεί μια πολύ σημαντική ιδιότητα των πολυμεταβλητών χρονοσειρών και εκφράζει τον επηρεασμό των τιμών μιας μεταβλητής όχι μονάχα από το δικό της παρελθόν αλλά και από το παρελθόν των υπόλοιπων μεταβλητών. Δηλαδή ολοκληρώνει την αιτιότητα που ενυπάρχει σε μια πολυμεταβλητή χρονοσειρά (μεταξύ των μεταβλητών, διάφορες χρονικές υστερήσεις μεταξύ των μεταβλητών). Για παράδειγμα σε ένα VAR μοντέλο:

$$\boldsymbol{\Phi}_p(B)\mathbf{Z}_t = \boldsymbol{\vartheta}_0 + \mathbf{a}_t$$

διαμερίζεται ο όρος \mathbf{Z}_t σε δύο συνιστώσες:

$$\mathbf{Z}_t = [\mathbf{Z}'_{1,t}, \mathbf{Z}'_{2,t}]'$$

, έτσι ώστε:

$$\begin{bmatrix} \Phi_{1,1}(B) & \Phi_{1,2}(B) \\ \Phi_{2,1}(B) & \Phi_{2,2}(B) \end{bmatrix} \begin{bmatrix} \mathbf{Z}_{1,t} \\ \mathbf{Z}_{2,t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\vartheta}_1 \\ \boldsymbol{\vartheta}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{a}_{1,t} \\ \mathbf{a}_{2,t} \end{bmatrix}$$

Αν το $\Phi_{1,2} = \mathbf{0}$, τότε λαμβάνονται:

$$\begin{cases} \Phi_{1,1}(B)\mathbf{Z}_{1,t} = \boldsymbol{\vartheta}_1 + \mathbf{a}_{1,t}, \\ \Phi_{2,2}(B)\mathbf{Z}_{2,t} = \boldsymbol{\vartheta}_2 + \Phi_{2,1}(B)\mathbf{Z}_{1,t} + \mathbf{a}_{2,t} \end{cases}$$

Επομένως, είναι φανερό πως οι τιμές των $\mathbf{Z}_{2,t}$ επηρεάζονται από το παρελθόν των τιμών του $\mathbf{Z}_{1,t}$, χωρίς να ισχύει το αντίστροφο (Eichler et al. (2017)).

3.2.3 VARMA

Το διανυσματικό μοντέλο αυτοπαλινδρόμησης κινητού μέσου, VARMA αποτελεί ένα από τα σημαντικότερα και πιο διαδεδομένα στοχαστικά μοντέλα παρακολούθησης και πρόβλεψης χρονοσειρών. Στην απλή του μορφή είναι το ακριβές παράγωγο των μοντέλων VAR και VMA με τις δύο εξισώσεις να συνδυάζονται κατά προφανή τρόπο. Αυτή η μοντελοποίηση είναι επιτεύξιμη επειδή η διεργασία θεωρείται στάσιμη, οπότε και τα δύο μοντέλα είναι αντιστρέψιμα. Εδώ παρουσιάζεται μια ελαφριά παραλλαγή του συνηθισμένου VARMA μοντέλου, αυτή που περιλαμβάνει τον συνυπολογισμό της εποχικότητας των δεδομένων, όπως παρακάτω:

$$\boldsymbol{\alpha}_P(B^s)\Phi_P(B)\mathbf{Z}_t = \boldsymbol{\vartheta}_0 + \boldsymbol{\beta}_Q(B^s)\Theta_Q(B)\mathbf{a}_t$$

, όπου

$$\boldsymbol{\alpha}_P(B^s) = \mathbf{I} - \boldsymbol{\alpha}_1 B^s - \dots - \boldsymbol{\alpha}_P B^{Ps} = \mathbf{I} - \sum_{k=1}^P \boldsymbol{\alpha}_k B^{ks}$$

$$\boldsymbol{\beta}_Q(B^s) = \mathbf{I} - \boldsymbol{\beta}_1 B^s - \dots - \boldsymbol{\beta}_Q B^{Qs} = \mathbf{I} - \sum_{k=1}^Q \boldsymbol{\beta}_k B^{ks}$$

Με s συμβολίζεται η περίοδος της εποχικότητας, $\boldsymbol{\alpha}_k$ και $\boldsymbol{\beta}_k$ είναι οι εποχικές παραμέτροι των AR και MA μοντέλων αντιστοίχως. P είναι η τάξη του εποχικού AR, ενώ Q του εποχικού MA. Τέλος, είναι

3.2 Διανυσματικά Μοντέλα Χρονοσειρών

εφικτό να προσαρμοστεί σε μη στάσιμα δεδομένα μια ακόμα επέκταση του εποχικού VARMA μοντέλου, το διανυσματικό εποχικό ολοκληρωμένο μοντέλο αυτοπαλινδρόμησης κινητού μέσου (VSARIMA):

$$\alpha_P(B^s)\Phi_P(B)\mathbf{D}(B^s)\mathbf{Z}_t = \boldsymbol{\vartheta}_0 + \beta_Q(B^s)\Theta_Q(B)\mathbf{a}_t$$

, όπου

$$\mathbf{D}(B^s) = \begin{bmatrix} (1 - B^s)^{d_1} & 0 & \dots & \dots & 0 & 0 \\ 0 & (1 - B^s)^{d_2} & 0 & \dots & \dots & 0 \\ 0 & 0 & (1 - B^s)^{d_3} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & \dots & \dots & 0 & (1 - B^s)^{d_m} \end{bmatrix}$$

Δηλαδή, για κάθε μεταβλητή που αποτελείται η πολυμεταβλητή χρονοσειρά εφαρμόζεται ένας παράγων που λαμβάνει τις διαφορές d τάξης μεταξύ των μετρήσεων έτσι ώστε να φαίνεται στάσιμη. Ακόμη, δεν είναι σπάνιος ο μετασχηματισμός της χρονοσειράς, πάντα αποσκοπώντας στην στασιμότητα, μέσω εκθετικών, πολυωνυμικών και λογαριθμικών συναρτησοειδών.

3.2.4 MESSV

Τα πολυμεταβλητά εποχικά μοντέλα εκθετικής λείανσης (MESSV) αποτελούν πρακτική που εφαρμόζεται από τον Β' Παγκόσμιο Πόλεμο για την πρόβλεψη τροχιών βομβών και ακόμα και στην σύγχρονη εποχή για το σκοπό της μοντελοποίησης χρονοσειρών με αρκετά μεγάλη ακρίβεια. Ειδικά στον χρηματοοικονομικό μετοχικό τομέα, πολλές φορές εφαρμόζονται για την αρωγή στην πάρση απόφασης σχετικά με αγορά ή πώληση μετοχών.

Κοινό σημείο στην οικογένεια αυτών των μοντέλων αποτελεί η διαίρεση της χρονοσειράς σε στοιχεία που περιγράφουν την τάση της (trend), την εποχικότητα (seasonality), το επίπεδο της (level) και ένα τυχαίο σφάλμα. Περιγράφεται από το παρακάτω σύστημα στοχαστικών εξισώσεων:

$$y_t = l_{t-1} + b_{t-1} + s_{t-p|t} + e_t,$$

$$l_t = l_{t-1} + b_{t-1} + \alpha_l e_t,$$

$$b_t = b_{t-1} + \alpha_b e_t,$$

3.2 Διανυσματικά Μοντέλα Χρονοσειρών

$$s_{t|t} = s_{t-p|t-1} + g_1 \alpha_s e_t,$$

$$s_{t-i|t} = s_{t-i|t-1} + g_2 \alpha_s e_t,$$

για $t = 1, 2, \dots, T$, όπου l_t και b_t συμβολίζουν το επίπεδο και την τάση της χρονοσειράς για κάθε χρονική στιγμή t και ο εποχικός όρος με $s_{t-p|t}$. Οι όροι α_l , α_b , α_s συμβολίζουν τις παραμέτρους λείανσης, συνήθως περιορισμένοι στο διάστημα $[0, 1]$. Οι όροι g_1 και g_2 αποτελούν όρους κανονικοποίησης και θέτονται ίσοι με $\frac{p-1}{p}$ και $-\frac{1}{p}$ αντιστοίχως, προκειμένου ο όρος της εποχικότητας να αθροίζει στο μηδέν σε κάθε χρονικό βήμα και δεν επηρεάζεται από το επίπεδο της χρονοσειράς.

Τέτοια μοντέλα που υπακούουν τις παραπάνω εξισώσεις δεν δύνανται να αφομοιώσουν την πληροφορία της αυτοσυσχέτισης εντός των μεταβλητών, καθότι ούτε που εκφράζουν τις αναμεταξύ τους σχέσεις με οποιοδήποτε τρόπο. Προς επίλυση αυτού εισάγεται το εποχικό μοντέλο διανυσματικής τοπικής τάσης (Vector Local Trend Seasonal Model VLTS) για την επαρκή μοντελοποίηση και ακρίβεια στην παρακολούθηση αυτοσυσχετιζόμενων χρονοσειρών.

$$\mathbf{y}_t = \mathbf{l}_{t-1} + \mathbf{b}_{t-1} + \mathbf{s}_{t-p|t} + \mathbf{e}_t,$$

$$\mathbf{l}_t = \mathbf{l}_{t-1} + \mathbf{b}_{t-1} + \mathbf{A}_l \mathbf{e}_t,$$

$$\mathbf{b}_t = \mathbf{b}_{t-1} + \mathbf{A}_b \mathbf{e}_t,$$

$$\mathbf{s}_{t|t} = \mathbf{s}_{t-p|t-1} + \mathbf{G}_1 \mathbf{A}_s \mathbf{e}_t,$$

$$\mathbf{s}_{t-i|t} = \mathbf{s}_{t-i|t-1} + \mathbf{G}_2 \mathbf{A}_s \mathbf{e}_t$$

, όπου τώρα οι όροι σε έντονη γραφή αντιπροσωπεύουν διανύσματα τιμών, και τα \mathbf{A}_l , \mathbf{A}_b , \mathbf{A}_s τους πίνακες ανθεκτικότητας της εκθετικής λείανσης. Οι πίνακες \mathbf{G}_1 και \mathbf{G}_2 είναι διαγώνιοι των στοιχείων $\frac{p-1}{p}$ και $-\frac{1}{p}$ αντιστοίχως.

Το παραπάνω μοντέλο τείνει να υπερεκτιμά (overestimate) τις μελλοντικές τιμές εξαιτίας του ότι η τάση προστίθεται ισοπόσως σε κάθε βήμα. Για αυτό προτείνεται μια ελαφριά παραλλαγή του παραπάνω μοντέλου, η λεγόμενη διανυσματική εποχική μοντελοποίηση αποσβαινουσας τάσης (Vector Damped Local Trend Seasonal Model VDLTS), με την εισαγωγή του πίνακα παραμέτρων Φ στην εξίσωση τάσης:

$$\mathbf{b}_t = \Phi \mathbf{b}_{t-1} + \mathbf{A}_b \mathbf{e}_t,$$

Κατά την υλοποίηση των μεθόδων MESSV για κάθε παράμετρο δύναται να διαλεχθεί ως προς την φύση/επιρροή της στο μοντέλο, δηλαδή δύναται να αποφασιστεί αν δρα αθροιστικά (όπως παρουσιάστηκαν

τα παραπάνω μοντέλα) ή πολλαπλασιαστικά (εφαρμογή όταν λ.χ. η τάση μεταβάλλεται εκθετικά).

3.3 Διαγράμματα Υπολοίπων

Τα διαγράμματα υπολοίπων είναι ο μεσολαβητής των προγνωστικών μοντέλων και των μοντέλων ελέγχου διεργασίας γενικότερα. Όπως έχει ήδη υπογραμμιστεί, αυτοσυσχετιζόμενες διεργασίες καταστούν ανεπαρκή την χρήση κλασικών διαγραμμάτων ελέγχου, αφού τα τελευταία δεν δύνανται να αποτυπώσουν τα χρονικά μοτίβα της διαδικασίας. Έχοντας αναφέρει τα κυριότερα μέσα μοντελοποίησης χρονοσειρών επισημαίνεται ότι συνήθως ακόμα και ένα πολύ απλό μα κατάλληλο μοντέλο όπως είναι το $AR(1)$, δηλαδή το αυτοπαλινδρομητικό μοντέλο με συμπερίληψη μονάχα του αμέσως προηγούμενου όρου της χρονοσειράς, παρουσιάζεται ανταγωνιστικό. Πρακτικά, η ουσία των μοντέλων υπολοίπων όπως καλούνται, δεν είναι παρά η μοντελοποίηση μέσω υπολοίπων, ενός κλασσικού διαγράμματος. Δηλαδή, τα μη παραμετρικά μοντέλα που μοντελοποιούν τις χρονοσειρές προφανώς παρουσιάζουν κάποια εφαρμογή (fit) πάνω στην χρονοσειρά και λαμβάνοντας τη διαφορά μεταξύ της "εκτιμώμενης" και της πραγματικής τιμής λαμβάνονται τα υπόλοιπα. Εν συνεχεία, κάποιο μοντέλο, με συνηθέστερα το T^2 και το $EWMA$, δέχεται ως τιμές της χρονοσειράς την τιμή των υπολοίπων και κατασκευάζει τα αντίστοιχα διαγράμματα. Τούτη η προσέγγιση παρουσιάζει δύο κύρια πλεονεκτήματα έναντι της άμεσης προσαρμογής μοντέλου. Το πρώτο συνίσταται στην α priori γνώστη κατανομή των υπολοίπων, κάτι που επάγει την καθυσηχαστικότερη εφαρμογή οποιουδήποτε μοντέλου αναγκαίεται κανονικότητας δεδομένων. Το δεύτερο είναι η πρόληψη της θέσης μιας διεργασίας εκτός ελέγχου. Αυτό είναι παράγωγο της ιδιότητας των μοντέλων χρονοσειρών να παράγουν προβλέψεις. Εξάλλου, μοντέλα όπως το $VARIMA$ χρησιμοποιούνται κυρίως λόγω της προβλεπτικής τους ικανότητας. Ο τρόπος εκμετάλλευσης αυτής της ιδιότητας είναι η προφανής κατασκευή ενός διαγράμματος με δεδομένα τα σφάλματα που προκύπτουν από την πρόβλεψη k βημάτων μπροστά σε σχέση με την μέση τιμή της διεργασίας και την προσθήκη ενός θορύβου. Αν και η διασπορά των εκτιμηθέντων αυξάνεται ανά βήμα που προβλέπεται, συνολικά πρόκειται για μια πάγια τακτική πρόληψης καταστροφής και εξασφάλισης συνολικής ποιότητας σε προϊόντα αλλά και αποφυγής μείωσης αξίας χαρτοφυλακίου στο χρηματιστήριο και πρόληψης διαδικτυακής εισροής δικτύων. Όπως καθίσταται ευνόητο, η προοπτική πρόβλεψης ξεκλειδώνει φαρέτρα με κατά πολύ ισχυρότερα βέλη.

Πρόσφορο στην διατύπωση της αξιολόγησης μοντέλων κρίνεται το περίγραμμα πρόβλεψης των μοντέλων των χρονοσειρών. Για τον σκοπό αυτό περιγράφεται ο τρόπος εκτίμησης μελλοντικών τιμών σε διαδικασίες $VARMA$, οι οποίες όπως έχει γίνει κατανοητό εγχολπώνουν τα απλούστερα αυτοπαλινδρομητικά και κινητού μέσου μοντέλα. Σημειώνεται ότι γενικά οι διανυσματικές αυτοπαλινδρομητικές διεργασίες κινητού μέσου κρίνονται κατάλληλες για την παραγωγή γραμμικών προβλέψεων συνόλων μεταβλητών χρονοσειρών. Ο Teräsvita (2006), θιασώτης της γραμμικής προσαρμογής, υποστηρίζει ότι τα γραμμικά μοντέλα σε πλήθος περιπτώσεων παρουσιάζονται στιβαρότερα των μη γραμμικών. Είναι

εξάλλου φειδωλές αναπαραστάσεις της παραγωγής γραμμικών δεδομένων διεργασιών. Προς ενίσχυση μιας τέτοιας προοπτικής μοντελοποίησης τάσσεται και η σπουδή του Harvey (2006) που διατείνεται ότι το πλαίσιο εφαρμογής *VARMA* μοντέλων είναι τελέσφορο στην πρόβλεψη χρονοσειρών.

Σε αυτό το σημείο παρουσιάζονται τα θεωρητικά αποτελέσματα εφαρμογής *VARMA* που ισχύουν υπό ιδανικές συνθήκες.

Κατά την πρόβλεψη ενός συνόλου μεταβλητών είναι χρήσιμη η σύσταση ενός κριτηρίου αξιολόγησης της προβλεπτικής ικανότητας. Δοσμένου ενός τέτοιου κριτηρίου, βέλτιστες προβλέψεις δύναται να παραχθούν. Από την φύση τους, οι μοντελοποιήσεις *VARMA* είναι κατάλληλες για την ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος (MSE). Το MSE θα χρησιμοποιηθεί στην παρούσα εξέταση, ωστόσο και άλλα κριτήρια έχουν διατυπωθεί και αξιολογηθεί από τους Granger και Newbold (1977). Οι προβλέψεις λαμβάνονται άμεσα από την απλή *VAR* μορφή:

$$y_t = \sum_{i=1}^{\infty} \Xi_i y_{t-i} + u_t$$

, όπου εδώ υιοθετείται ο συμβολισμός που εκφράζει την αντιστρεψιμότητα σε μια *MA* διαδικασία και το σφάλμα σαν u_t που θεωρείται μια White Noise IID ανεξάρτητη διεργασία. Τότε λαμβάνεται βέλτιστα μια ελάχιστου τετραγωνικού σφάλματος πρόβλεψη h βημάτων στο χρόνο τ ως η δεσμευμένη αναμενόμενη τιμή δοσμένου του y_t , με $t \leq \tau$:

$$y_{\tau+h|\tau} = E(y_{\tau+h} | y_{\tau}, y_{\tau-1}, \dots)$$

Επιπλέον, μπορεί να ορισθεί αναδρομικά για $h = 1, 2, \dots$, ως εξής:

$$y_{\tau+h|\tau} = \sum_{i=1}^{\infty} \Xi_i y_{\tau+h-i|\tau}$$

, όπου $y_{\tau+h|\tau} = y_{\tau+h}$ για $h \leq 0$. Ακόμη και αν οι u_t δεν είναι ανεξάρτητες, αλλά ασυσχέτιστες white noise διεργασίες, η πρόβλεψη που λαμβάνεται κατά αυτόν τον τρόπο είναι η καλύτερη γραμμική πρόβλεψη αν και μπορεί να υπολείπεται μιας μεγάλης κλάσης μη γραμμικών συναρτήσεων προσαρμοσμένων σε παρελθοντικές παρατηρήσεις. Για δοσμένες αρχικές τιμές, η u_t δύναται να οριστεί σαν την παροντική υπόθεση μιας γνωστής σειράς. Επομένως, οι h βήματα μπροστά προβλέψεις μπορούν να καθοριστούν εναλλακτικώς:

$$y_{\tau+h|\tau} = A_0^{-1}(A_1 y_{\tau+h-1|\tau} + \dots + A_p y_{\tau+h-p|\tau}) + A_0^{-1} \sum_{i=h}^q M_i u_{\tau+h-i}$$

, όπου με A συμβολίζονται οι πίνακες αυτοπαλινδρόμησης και με M του κινητού μέσου και ισχύει ότι το άθροισμα είναι μηδενικό αν ισχύει ότι $h > q$.

Με οποιονδήποτε από τους δύο τρόπους εκτιμηθούν οι προβλέψεις για το $VARMA$ μοντέλου απαιτούνται αρχικές τιμές. Ειδικότερα, στην πρώτη περίπτωση κατά αυστηρό κανόνα απαιτούνται άπειρες τιμές y_t αν η αυτοπαλινδρομητική αναπαράσταση είναι τάξης απείρου, ενώ στη δεύτερη περίπτωση απαιτείται πλήρης γνώση σχετικά με τις χρονοσειρές σφάλματος (white noise u_t), κάτι που δεν παρατηρείται στην πράξη αφού αυτό συνεπάγεται την γνώση όλων των αρχικών συνθηκών ή και τιμών της y_t . Μια εναλλακτική προσέγγιση έχει προταθεί από τους Brockwell και Davis (1987) γνωστή ως πολυμεταβλητός καινοτόμος αλγόριθμος (Multivariate Innovations Algorithm). Υπό υποθέσεων περί στάσιμων χρονοσειρών, οι ιδιότητες των σφαλμάτων πρόβλεψης εξάγονται εκφράζοντας την διεργασία σαν μεταβλητού μέσου:

$$y_t = u_t + \sum_{i=1}^{\infty} \Phi_i u_{t-i}$$

, με $A_0 = M_0$. Βάσει αυτής της αναπαράστασης η βέλτιστη πρόβλεψη λαμβάνεται ως:

$$y_{\tau+h|\tau} = \sum_{i=h}^{\infty} \Phi_i u_{\tau+h-i}$$

Πλέον τα σφάλματα πρόβλεψης εξάγονται ως:

$$y_{\tau+h} - y_{\tau+h|\tau} = u_{\tau+h} + \Phi_1 u_{\tau+h-1} + \dots + \Phi_{h-1} u_{\tau+1}$$

Έτσι, η πρόβλεψη είναι αμερόληπτη, αφού τα σφάλματα έχουν μηδενικό μέσο και το μέσο τετραγωνικό σφάλμα ή όπως είναι γνωστός ο πίνακας συνδιασποράς των σφαλμάτων πρόβλεψης είναι:

$$\Sigma_y(h) = E[(y_{\tau+h} - y_{\tau+h|\tau})(y_{\tau+h} - y_{\tau+h|\tau})'] = \sum_{j=0}^{h-1} \Phi_j \Sigma_u \Phi_j'$$

Ένα πόρισμα είναι το ακόλουθο:

Αν τα u_t είναι κανονικώς κατανομημένα τότε και τα σφάλματα πρόβλεψης είναι κανονικώς κατανομημένα:

$$y_{\tau+h} - y_{\tau+h|\tau} \sim N(0, \Sigma_y(h))$$

Διαστήματα εμπιστοσύνης δύναται να εξαχθούν από τα παραπάνω βάση αυτής της κανονικής κατανομής. Επιπλέον, ο πίνακας συνδιασποράς των σφαλμάτων πρόβλεψης φράσσεται από τον πίνακα συνδιασποράς των y_t :

$$\Sigma_y(h) \rightarrow_{h \rightarrow \infty} \Sigma_y = E(y_t, y_t') = \sum_{j=0}^{h-1} \Phi_j \Sigma_u \Phi_j'$$

Άρα τα διαστήματα εμπιστοσύνης της πρόβλεψης έχουν επίσης φραγμένο μήκος καθώς ο ορίζοντας πρόβλεψης μεγαλώνει.

Όσον αφορά την παρούσα εργασία, αυτά τα υπόλοιπα θα χρησιμοποιηθούν για την προσαρμογή μοντέλων διαγραμμάτων χρονοσειρών.

Χάριν πληρότητας, παρουσιάζονται και οι εξισώσεις για την οικογένεια μοντέλων εκθετικής λείανσης που αφορούν την πρόβλεψη και το σφάλμα της (D.Pfeffermann J.Allon). Οι μονοβηματικές προβλέψεις θα είναι:

$$\mathbf{l}_{t+1} = \mathbf{A}_l(\mathbf{y}_{t+1} - \mathbf{s}_{t+1|t}) + (\mathbf{I} - \mathbf{A}_l)(\mathbf{l}_t + \mathbf{b}_t)$$

$$\mathbf{b}_{t+1} = \mathbf{A}_b(\mathbf{l}_{t+1} - \mathbf{l}_t) + (\mathbf{I} - \mathbf{A}_b)\mathbf{b}_t$$

$$\mathbf{s}_{t+1|t+1} = \mathbf{A}_s(\mathbf{y}_{t+1} - \mathbf{l}_{t+1}) + (\mathbf{I} - \mathbf{A}_s)\mathbf{s}_{t+1|t}$$

$$\sum_{i=0}^p \mathbf{s}_{t+1+i|t+1} = \mathbf{0}$$

Γενικά η πρόβλεψη της τιμής της χρονοσειράς για h βήματα μπροστά θα είναι:

$$\hat{\mathbf{y}}_{t+h} = \mathbf{l}_t + h\mathbf{b}_t + \mathbf{s}_{t+h|t}$$

Το σφάλμα πρόβλεψης δίνεται από τον τύπο:

$$\mathbf{e}_{t+1} = [\mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1}]$$

Μέσω των σφαλμάτων δύναται να ληφθεί η διαφορετική διατύπωση των προβλέψεων της τάσης, της εποχικότητας και του επιπέδου της χρονοσειράς:

$$\mathbf{l}_{t+1} = \mathbf{l}_t + \mathbf{b}_t + \mathbf{A}_l \mathbf{e}_{t+1}$$

$$\mathbf{b}_{t+1} = \mathbf{b}_t + \mathbf{A}_b \mathbf{A}_l \mathbf{e}_{t+1}$$

$$\mathbf{s}_{t+1|t+1} = \mathbf{s}_{t+1|t} + \frac{p-1}{p} \mathbf{A}_s (\mathbf{I} - \mathbf{A}_l) \mathbf{e}_{t+1}$$

$$\mathbf{s}_{t+1+i|t+1} = \mathbf{s}_{t+1+i|t} - \frac{1}{p} \mathbf{A}_s (\mathbf{I} - \mathbf{A}_l) \mathbf{e}_{t+1}, i = 1, \dots, p-1$$

Το παραπάνω σύστημα καταδεικνύει τον τρόπο με τον οποίο κάθε πτυχή της χρονοσειράς ενημερώνεται σαν συνάρτηση των πιο πρόσφατων παρατηρούμενων σφαλμάτων. Σε αυτό το σημείο ολοκληρώνεται ο τρόπος λήψης των σφαλμάτων για τα γραμμικά μοντέλα εφαρμογής για τα δύο γενικότερα πλαίσιο των δύο οικογενειών μοντέλων (*VARMA* και *MESSV*).

4 Αναδρομικά Νευρωνικά Δίκτυα

4.1 Εισαγωγή

Τα νευρωνικά δίκτυα αφορμήθηκαν με στόχο την κατασκευή ενός συστήματος μοντελοποίησης και επεξεργασίας δεδομένων κατά τρόπο που ομοιάζει τον ανθρώπινο εγκέφαλο και τις λειτουργίες του, κυνηγώντας αυτή την περίπλοκη, μη-γραμμική, εμπειρική απόκριση που παρέχει ανάλογα το input. Καθώς ορίζει το ονόμα τους, σχηματικά δύναται να κατανοηθούν σαν διατάξεις ή δίκτυα τεχνητών νευρώνων. Τα νευρωνικά δίκτυα προτάθηκαν το 1944 από τους McCullough και Pitts, αλλά στην μορφή που χρησιμοποιούνται σήμερα πρωτοδιαμορφώθηκαν το 1965 από τους Ivachenko και Lara σαν μια μέθοδος διαχειρισμού και ομαδοποίησης δεδομένων.

Αναλυτικότερα, τα νευρωνικά δίκτυα αποτελούν το μέσο πραγματοποίησης μηχανικής μάθησης, στο οποίο ο υπολογιστής μαθαίνει την εκτέλεση ενός σκοπού αναλύοντας δεδομένα εκπαίδευσης, τα οποία είναι τις περισσότερες φορές επισημασμένα. Αποτελούνται, από χιλιάδες ή εκατομμύρια κόμβους επεξεργασίας που είναι πυκνώς συνδεδεμένοι. Συνήθως οι τελευταίοι είναι οργανωμένοι σε στρώματα κόμβων από τα οποία τα δεδομένα ακολουθούν την ίδια, μία κατεύθυνση (feed-forward). Κάθε κόμβος δύναται να είναι συνδεδεμένος με πολλούς κόμβους του προηγούμενου στρώματος, από τους οποίους λαμβάνει τα δεδομένα και να καταλήγει σε πολλούς κόμβους του επόμενου στρώματος, τους οποίους τροφοδοτεί με δεδομένα. Για κάθε σύνδεση από τους κόμβους του προηγούμενου στρώματος, ο εκάστοτε κόμβος αναθέτει ένα βάρος ("weight"). Τα βάρη αυτά πολλαπλασιάζονται με την εκάστοτε τιμή που φτάνει από τους αντίστοιχους κόμβους του προηγούμενου στρώματος και το άθροισμα τους συγκρίνεται με μια τιμή ανοχής (threshold value) ώστε να αποφασιστεί αν ο κόμβος χρησιμεύει σαν διαδοτής πληροφορίας στο επόμενο στρώμα. Κατά την εκπαίδευση ενός δικτύου, τα βάρη και οι ανοχές αρχικοποιούνται σε τυχαίες τιμές, και τα δεδομένα εισάγονται στο στρώμα εισαγωγής (input layer) ακολουθώντας όλα τα στρώματα του δικτύου, καταλήγοντας στο στρώμα εξαγωγής (output layer). Τα βάρη και οι τιμές ανοχής συνεχώς ενημερώνονται έως ώτου τα δεδομένα εκπαίδευσης επισημασμένα κατά το ίδιο τρόπο παράγουν συνεπώς όμοιο αποτέλεσμα.

Κατηγορίες Μοντέλων Μηχανικής Μάθησης

Ως προς την Παρουσία Ετικετών

Τα μοντέλα μηχανικής μάθησης μπορούν να ταξινομηθούν σε τρεις γενικές κατηγορίες με βάση τη διαθεσιμότητα των ετικετών. (1) Επιβλεπόμενη ανίχνευση βαθιάς ανωμαλίας (supervised). (2) Ημι-επιβλεπόμενη ανίχνευση βαθιάς ανωμαλίας (semi-supervised). (3) Χωρίς επίβλεψη ανίχνευση βαθιάς ανωμαλίας (unsupervised).

1. Supervised Μοντέλα Περιλαμβάνουν την εκμάθηση ενός επιβλεπόμενου δυαδικού ή πολλαπλού ταξινομητή, χρησιμοποιώντας ετικέτες. Παρά τη βελτιωμένη απόδοση των επιβλεπόμενων μεθόδων, αυτές οι μέθοδοι δεν είναι τόσο δημοφιλείς όσο ημι-επιβλεπόμενες ή μη επιβλεπόμενες μέθοδοι, λόγω της έλλειψης διαθεσιμότητας επισημασμένων εκπαιδευτικών δειγμάτων. Επιπλέον, αποδίδουν υπό-βέλτιστη σε σύνολα δεδομένων που υπάρχει ανισορροπία τάξης.

2. Semi-Supervised Μοντέλα Αυτού του είδους τα μοντέλα αξιοποιούν όσες ετικέτες υπάρχουν στο δείγμα, συνήθως μέσω χρήσης ενός συστήματος κωδικοποίησης και αποκωδικοποίησης..

3. Unsupervised Μοντέλα Οι μη επιβλεπόμενες τεχνικές αξιοποιούν αποκλειστικά τις εγγενείς ιδιότητες των δεδομένα προσπαθώντας να ανακαλύψουν πιθανές δομές που κρύβονται πίσω από τα δεδομένα. Οι μη επιβλεπόμενες τεχνικές χρησιμοποιούνται ευρέως στην αυτόματη επισήμανση δειγμάτων δεδομένων χωρίς ετικέτα, όπου αυτές απουσιάζουν.

Ως προς το πλήθος ενδιάμεσων στρωμάτων

Από την σύλληψη των νευρωνικών δικτύων έχει επέλθει τόση μεταλλαγή στο μέσον υλοποίησης που από ένα συμπίλημα σποραδικών νευρώνων, τα δίκτυα έχουν αποκτήσει σαφές και καθορισμένο στρωματοποιημένο πλαίσιο με μεταβαλλόμενο πλήθος στρωμάτων ανάλογα με το εκάστοτε αντιμετωπιζόμενο πρόβλημα, τον τύπο, το πλήθος και την ποιότητα των δεδομένων, την ζητούμενη πολυπλοκότητα (διάθεση χρόνου για μοντελοποίηση και εκπαίδευση) και την προοπτική μοντελοποίησης. Έτσι, χωρίζονται σε δύο κύριες κατηγορίες: Δίκτυα Ρηχής Μάθησης και Δίκτυα Βαθιάς Μάθησης.

1. Δίκτυα Ρηχής Μάθησης Το πλήθος των ενδιάμεσων στρωμάτων σε αυτήν την κατηγορία είναι μικρό. Τέτοιου είδους μοντέλα βασίζονται κυρίως σε ευρετικά (heuristics) που επιβάλλονται πάνω σε χαρακτηριστικά που έχουν επισημανθεί από τον επιστήμονα που εξετάζει ένα πρόβλημα. Η ροή που ακολουθείται σε ένα πρόβλημα είναι συνήθως ο διαχωρισμός σε δύο στάδια, αυτό της εξαγωγής χαρακτηριστικών και αυτό της εκπαίδευσης. Παράδειγμα τέτοιων δικτύων αποτελούν τα Support Vector Machines, Naive Bayes και Decision Trees.

2. Δίκτυα Βαθιάς Μάθησης Το πλήθος των ενδιάμεσων στρωμάτων είναι μεγάλο. Ενδεικτικά μεγάλα θεωρούνται μεγέθη από 10 και πάνω ενδιάμεσα στρώματα. Στα βαθιά δίκτυα η εκπαίδευση και η εξαγωγή χαρακτηριστικών ουσιαστικά ταυτίζονται. Συνήθως απαιτείται αρκετά παραπάνω χρόνος για εκπαίδευση, καθότι πρόκειται για μια υπολογιστικά ακριβή διαδικασία. Βέβαια, η εκμάθηση είναι ικανή στην συμπερίληψη εκφραστικών και περίπλοκων χαρακτηριστικών, δημιουργώντας γενικά καλύτερα μοντέλα. Παράδειγμα τέτοιων δικτύων αποτελούν τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional

Neural Networks, CNN), οι Autoencoders και τα Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks, RNN) που θα αναπτυχθούν εκτενώς εν συνεχεία.

4.2 Νευρωνικά Δίκτυα

Νευρώνας

Οι νευρώνες είναι το κύτταρο λειτουργίας των νευρωνικών δικτύων. Πρόκειται για στοιχεία ή σύνολα βαρών μεταξύ του εκάστοτε νευρώνα και άλλων νευρώνων. Όλες οι ενώσεις του νευρώνα έχουν αντίστοιχα βάρη προκειμένου να προσομοιωθεί ο τρόπος με τον οποίο η ενεργοποίηση ενός νευρώνα επηρεάζει τους υπόλοιπους. Η προσομοίωση διέγερσης επιτυγχάνεται αθροίζοντας το γινόμενο των εισαγόμενων νευρώνων και των αντίστοιχων συναπτικών βαρών τους. Οι βιολογικοί νευρώνες εμφανίζουν ανεξαιρέτως κάποιο όριο ενεργοποίησης - απόκρισης, κάτι που δεν εφαρμόζεται στους τεχνητούς νευρώνες συνήθως, παραμόνο στο στρώμα εξόδου (output layer), αλλά τα παραγόμενα στοιχεία της άθροισης χρησιμοποιούνται για την ενεργοποίηση επόμενων νευρώνων. Στο στρώμα εξόδου στο παράγωγο των νευρώνων δύναται να εφαρμοστεί μια δείκτρια συνάρτηση η οποία παρέχει μια τιμή ανοχής και ενεργοποιείται.

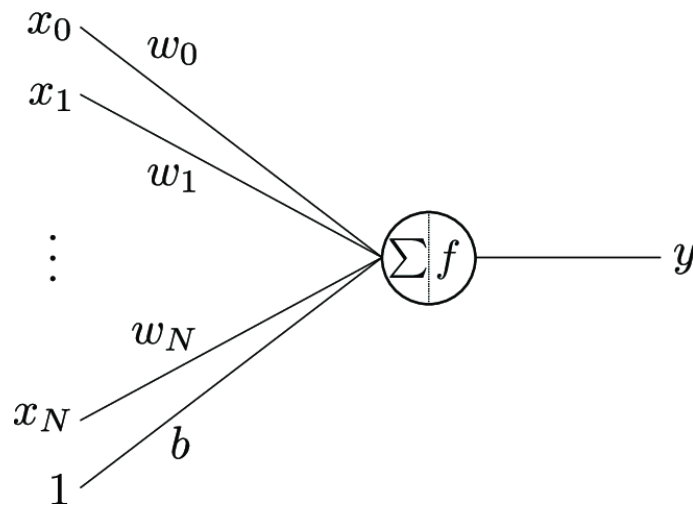


Figure 1: Ένας τυπικός νευρώνας

Όπως φαίνεται στο σχήμα 1 συμβολίζοντας με x το διάνυσμα των inputs $[x_0, x_1, \dots, x_N]$, με τα αντίστοιχα βάρη τους w , $[w_0, w_1, \dots, w_N]$ και b μια σταθερά διόρθωσης μεροληψίας (bias), τότε ακριβώς στο Σ γίνεται ο υπολογισμός:

$$z = w^T x + b$$

, δηλαδή με z συμβολίζεται το άθροισμα των γινομένων των τιμών x με τα αντίστοιχα βάρη και έπειτα σαν output y ορίζεται να είναι το z μέσω μιας συνάρτησης ενεργοποίησης f . Δηλαδή:

$$y = f(z)$$

Συναρτήσεις Ενεργοποίησης

Η συνάρτηση ενεργοποίησης αποτελεί συνάρτηση με χρήση τέτοια ώστε να λαμβάνει μια μικρή τιμή όταν το z είναι μικρό σε αντίθετη περίπτωση λαμβάνει μεγάλη τιμή, αν υπερβαίνεται το όριο ανοχής. Παρομοιάζεται με πύλη μέσω της οποίας δεν περνάνε τιμές μεγαλύτερες ορίου ανοχής. Η χρησιμότητα τους έγκειται στο γεγονός ότι εισάγουν μη γραμμικότητα στα νευρωνικά δίκτυα, επιτρέποντας την εκμάθηση εκφραστικών και σύνθετων διαδικασιών. Πράγματι, αν οι συναρτήσεις ενεργοποίησης απουσιάζουν από ένα πολυστρωματικό νευρωνικό δίκτυο, τότε αυτό θα ήταν ισοδύναμο με μια απλή γραμμική παλινδρόμηση ή μετασχηματισμό ενός πίνακα πάνω στο input και δεν θα κρινόταν κατάλληλο για εκμάθηση περίπλοκων χαρακτηριστικών. Οι πιο συνήθεις συναρτήσεις ενεργοποίησης είναι οι σιγμοειδείς (sigmoids) και ReLU (Rectified Linear Unit).

ReLU

Οι συναρτήσεις ReLU είναι από τις απλούστερες και πιο αποδοτικές συναρτήσεις ενεργοποίησης η οποία είναι μηδενική αν το input είναι αρνητικό, ειδάλλως σαν output παίρνει το input. Δηλαδή:

$$f(x) = \max(0, x)$$

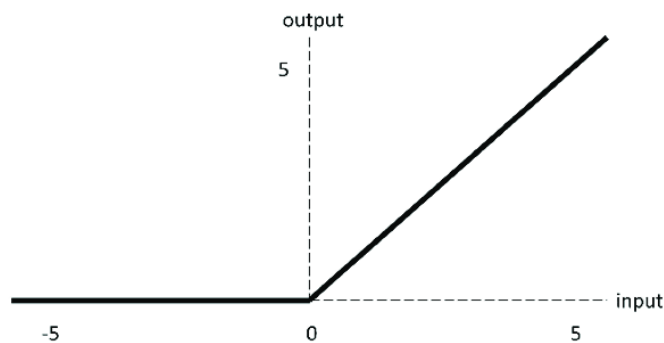


Figure 2: Συνάρτηση Ενεργοποίησης ReLU

Με παράγωγο:

$$\frac{df(x)}{dx} = \begin{cases} 1 & , x > 0 \\ 0 & , x < 0 \\ UNDEF & , x = 0 \end{cases}$$

Η παράγωγος εξετάζεται επειδή οι μηδενικές τιμές δύναται να είναι προβληματικές και παγιδευμένη στην μηδενική περιοχή η οπισθοδρόμηση, επομένως αναφέρεται σαν έναυσμα για την παρακάτω διατύπωση της παραμετρικής ReLU (PReLU).

PReLU

$$\frac{df(x)}{dx} = \begin{cases} x & , x \geq 0 \\ \alpha x & , x < 0 \end{cases}$$

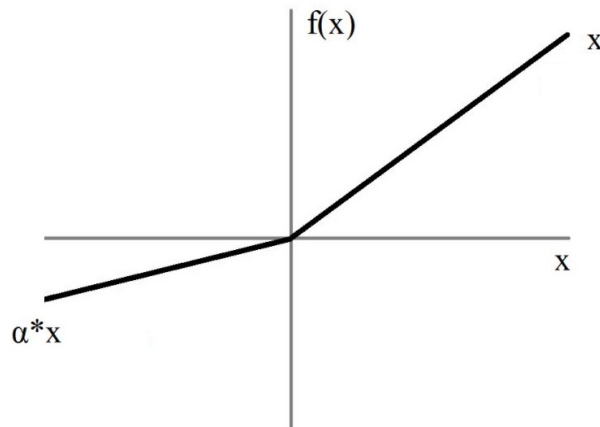


Figure 3: Συνάρτηση Ενεργοποίησης PReLU

Το πλεονέκτημα που παρουσιάζει η επιλογή της παραμετρικής ReLU είναι η μη μηδενικότητα της παραγώγου εκτός του σημείου 0 που δεν ορίζεται. Συνήθως, το α επιλέγεται να είναι μια μικρή τιμή, για παράδειγμα 0.01, 0.05 ή 0.1, ώστε πρακτικά σχεδόν να εξουδετερώνονται οι επιδράσεις των τιμών κάτω από το όριο ανοχής που εν παρόντι ισούται με 0.

Λογιστική Σιγμοειδής

$$S(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1}$$

Η λογιστική σιγμοειδής συνάρτηση έχει την χρήσιμη ιδιότητα ότι η παράγωγος της ορίζεται παντού και το output κυμαίνεται μεταξύ των τιμών 0 και 1. Το αρνητικό που παρουσιάζει είναι η πολυπλοκότητα υπολογισμού για την εκτίμηση των εκθετικών και για αυτό τον σκοπό δεν προτιμάται εκτενώς.

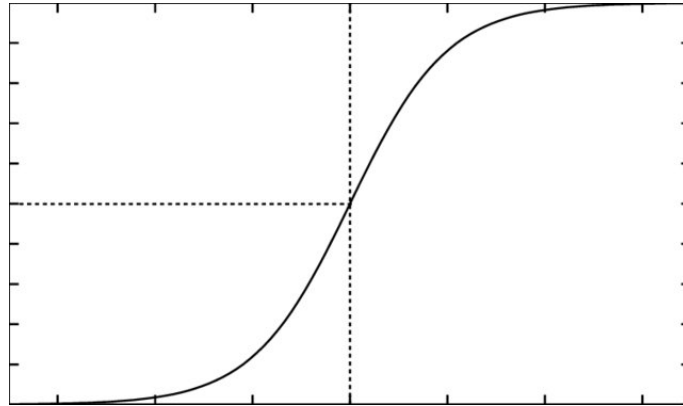


Figure 4: Συνάρτηση Ενεργοποίησης Logistic Sigmoid

Η παράγωγος της λογιστικής σιγμοειδούς συνάρτησης υπολογίζεται ως:

$$\frac{dS(x)}{ds} = S(x)(1 - S(x))$$

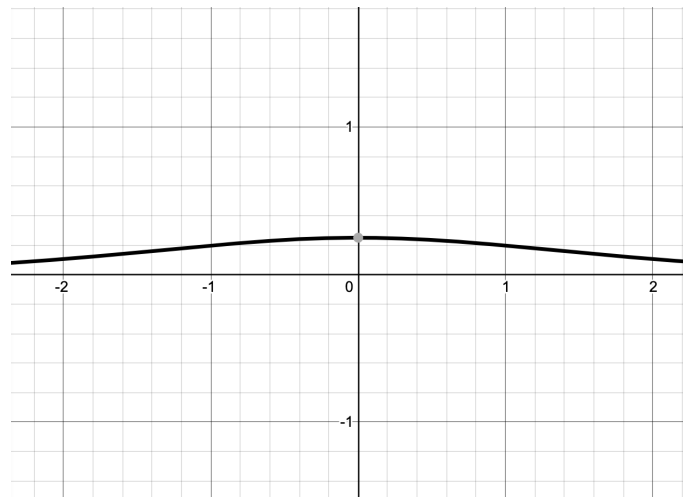


Figure 5: Παράγωγος Logistic Sigmoid

Όπως φαίνεται στο σχήμα 5 η παράγωγος της λογιστικής σιγμοειδούς συνάρτησης είναι μη μηδενική σε κάθε σημείο, το οποίο θεωρείται πλεονέκτημα κατά το βήμα εφαρμογής του αλγορίθμου της οπισθοδρόμησης. Βέβαια καθότι η παράγωγος λαμβάνει αρκετά μικρές τιμές για μεγάλα κατά απόλυτη τιμή ορίσματα, τοποθετείται το πρόβλημα των εξαφανιζομένων παραγώγων (το φαινόμενο κατά το οποίο οι

συνεχείς μικρές ή μηδενικές τιμές των παραγώγων οδηγούν σε πολύ αργούς ρυθμούς μάθησης. Επιλύεται μέσω της χρήσης PReLU συναρτήσεων ή υπολειπόμενων νευρωνικών δικτύων).

Softmax

Η softmax συνάρτηση ενεργοποίησης είναι ιδιαίτερα διαδεδομένη σε προβλήματα ταξινόμησης, ειδικά σε περιπτώσεις πολλαπλών κλάσεων. Χρησιμοποιείται κυρίως στο output layer ακριβώς για αυτό τον σκοπό. Επιστρέφει σκορ εμπιστοσύνης για κάθε κλάση και εκλέγεται το μεγαλύτερο εξ αυτών ως ανταπόκριση για την ταξινόμηση. Η μαθηματική έκφραση της softmax είναι η ακόλουθη:

$$f(x) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

, όπου το i συμβολίζει το κάθε μέρος του output των στοιχείων τα οποία είναι K στο σύνολο.

Αντικειμενική Συνάρτηση

Ουσιαστικά, ο σκοπός ενός νευρωνικού δικτύου είναι να αντιστοιχίσει τα δεδομένα εισόδου σε έξοδο μέσω εκμάθησης με το σύνολο εκπαίδευσης. Ο ακριβής υπολογισμός όλων των βαρών των νευρώνων είναι κάτι το αδύνατο, για αυτό το πρόβλημα τίθεται σαν ένα πρόβλημα βελτιστοποίησης, με τον αλγόριθμο βελτιστοποίησης κατά την οπισθοδρόμηση να χρησιμοποιείται για να "ψάξει" των χώρο τιμών των βαρών, ώστε το μοντέλο να είναι ικανοποιητικό σε αξιολόγηση διαδικασίας και πρόβλεψη. Στο πλαίσιο ενός αλγορίθμου βελτιστοποίησης, η συνάρτηση που χρησιμοποιείται για την εξέταση μιας πιθανής λύσης αναφέρεται σαν αντικειμενική συνάρτηση. Σκοπός είναι η μεγιστοποίηση ή ελαχιστοποίηση της αντικειμενικής συνάρτησης, όπου γίνεται λόγος για συνάρτηση κέρδους και απώλειας αντιστοίχως. Ο ρόλος της αντικειμενικής συνάρτησης είναι εξαιρετικά κρίσιμος, αφού καλείται να αποτυπώσει όλα τα εκφραστικά χαρακτηριστικά του συνόλου δεδομένων σε έναν αριθμό, έτσι ώστε οι βελτιώσεις του να σημαίνουν καλύτερη προσαρμογή. Επομένως, κατά τον υπολογισμό του σφάλματος ενός μοντέλου, επιλέγεται μια κατάλληλη αντικειμενική συνάρτηση, η οποία αντιπροσωπεύει τους σκοπούς της εκπαίδευσης και τα χαρακτηριστικά των δεδομένων. Οι πιο κοινώς διαδεδομένες παρουσιάζονται κάτωθι:

Maximum Likelihood

Η εκτίμηση Μέγιστης Πιθανοφάνειας (Maximum Likelihood Estimation) είναι το πλαίσιο στο οποίο αναπτύσσονται και τεκμηριώνονται οι αντικειμενικές συνάρτησεις για την βελτιστοποίηση μιας λύσης. Θεωρώντας ένα σύνολο δεδομένων με μία ή παραπάνω μεταβλητές (χαρακτηριστικά) απαιτείται εύρεση ενός μοντέλου βαρών παραμέτρων που κάλλιστα απεικονίζει τα δεδομένα εισόδου στην έξοδο. Μέσω της

εκτίμησης με την μέθοδο της Μέγιστης Πιθανοφάνειας, η αντικειμενική συνάρτηση εκτιμά πόσα κοντά είναι η κατανομή των παραγόμενων από το μοντέλο τιμών με την κατανομή των τιμών στόχων των δεδομένων εκπαίδευσης. Επιπλέον, με τη χρήση τούτου του πλαισίου επάγεται και το σημαντικό φαινόμενο της βελτίωσης αποτελεσμάτων του δικτύου με την αύξηση του πλήθους δεδομένων εκπαίδευσης.

Υπό αυτό το πλαίσιο, η ποσοτικοποίηση του σφάλματος μεταξύ δύο κατανομών γίνεται μέσω της μεταξύ τους εντροπίας (cross entropy). Για παράδειγμα, σε ένα πρόβλημα ταξινόμησης δύο κλάσεων, δύναται να επιλεγεί ως κλάση μιας τιμής αυτή που έχει την μεγαλύτερη πιθανότητα να ανήκει σε κάποια από τις δύο κλάσεις. Δοσμένων δεδομένων εκπαίδευσης αποσκοπάται ο προσδιορισμός των βαρών των νευρώνων έτσι ώστε να ελαχιστοποιείται η διαφορά μεταξύ της προβλεπόμενης από το δίκτυο κατανομής των στοιχείων και αυτής που εντοπίζεται στο σύνολο δεδομένων. Ωστόσο, υπάρχουν διάφορες άλλες αντικειμενικές συναρτήσεις που χρησιμοποιούνται, οι οποίες παρουσιάζονται παρακάτω.

Μέσο Τετραγωνικό Σφάλμα (MSE)

Το μέσο τετραγωνικό σφάλμα χρησιμοποιείται συνήθως για παλινδρομητικούς σκοπούς. Η αντικειμενική συνάρτηση, που εδώ είναι συνάρτηση απώλειας, υπολογίζεται ως ο μέσος των τετραγωνικών διαφορών της τιμής στόχου και της προβλεπόμενης τιμής.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Μέσο Απόλυτο Σφάλμα (MAE)

Το μέσο απόλυτο σφάλμα χρησιμοποιείται συνήθως για παλινδρομητικούς σκοπούς, όπως το MSE. Η αντικειμενική συνάρτηση, που εδώ είναι συνάρτηση απώλειας, υπολογίζεται ως ο μέσος των απόλυτων διαφορών της τιμής στόχου και της προβλεπόμενης τιμής.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

Binary Crossentropy (BCE)

Η Binary Crossentropy απώλεια χρησιμοποιείται για το πρόβλημα της ταξινόμησης δύο κλάσεων και είναι γενίκευση της λογιστικής απώλειας. Με την χρήση της απαιτείται μονάχα ένας νευρώνας εξόδου για την ταξινόμηση, του οποίου η τιμή "περνά" από μια σιγμοειδή συνάρτηση ώστε η απόκριση να είναι 0 ή 1.

$$BCE = -\frac{1}{N} \sum_{i=1}^N y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)$$

Categorical Crossentropy (CCE)

Η Categorical Crossentropy απώλεια χρησιμοποιείται για το πρόβλημα της ταξινόμησης πολλαπλών κλάσεων και είναι γενίκευση της BCE απώλειας. Με την χρήση της απαιτείται ένα στρώμα εξόδου με κόμβους όσους και οι κλάσεις προς ταξινόμηση, στο οποίο εφαρμόζεται μια softmax συνάρτηση ενεργοποίησης προκειμένου οι τιμές των κόμβων να κυμαίνονται μεταξύ του 0 και του 1. Σημειώνεται ότι κατά την εκπαίδευση του δικτύου χρησιμοποιείται one-hot encoding για τον ορθό προσδιορισμό των κλάσεων των αντικειμένων του συνόλου εκπαίδευσης.

$$BCE = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^J y_j \log(\hat{y}_j) + (1 - y_j) \log(1 - \hat{y}_j)$$

, όπου με J συμβολίζεται ο αριθμός των κλάσεων στο εκάστοτε πρόβλημα.

Sparse Categorical Cross Entropy

Η Sparse Categorical Crossentropy απώλεια χρησιμοποιείται για το πρόβλημα της ταξινόμησης πολλαπλών κλάσεων και διαφοροποιείται από την CCE καθώς δεν προϋποθέτει την one-hot κωδικοποίηση της απόκρισης του συνόλου εκπαίδευσης, αλλά απλά την επισήμανση σαν δείκτη της κλάσης που ανήκει κάθε τιμή.

$$SCCE = -\sum_{i=1}^C t_i \log(f(s_i))$$

, όπου με t_i συμβολίζεται ο δείκτης της κλάσης των δεδομένων, f είναι η softmax συνάρτηση ενεργοποίησης που εφαρμόζεται στις τιμές του τελευταίου στρώματος του δικτύου.

Οπισθοδιάδοση- Backpropagation

Το βήμα της διάδοσης σφάλματος προς τα πίσω, τελείται μέσω του αλγορίθμου Backpropagation (backward propagation of errors). Χρησιμοποιείται σε supervised neural networks και εφαρμόζει τεχνικές απότομης καθόδου (gradient descent). Δοσμένου ενός νευρωνικού δικτύου και μιας αντικειμενικής συνάρτησης, η μέθοδος υπολογίζει την κλίση (gradient) της αντικειμενικής συνάρτησης σε σχέση με τα βάρη του δικτύου. Είναι η γενίκευση του delta rule του perceptron σε πολυστρωματικά feed-forward νευρωνικά δίκτυα.

Το όνομα "οπισθοδιάδοση" προκύπτει με άμεσο τρόπο από το γεγονός ότι ο υπολογισμός των κλίσεων εξελίσσεται προς τα πίσω στο δίκτυο, με την κλίση του τελευταίου στρώματος να υπολογίζεται πρώτη και του πρώτου τελευταία. Υπολογισμοί κλίσης ενός στρώματος χρησιμοποιούνται μερικώς για τον υπολογισμό των προηγούμενων, κάτι που επιτρέπει στην οπίσθια διάδοση τον αποδοτικό υπολογισμό της κλίσης κάθε στρώματος αντί της αφελούς πολύπλοκης εύρεσης της κλίσης κάθε στρώματος χωριστά.

Για την εφαρμογή του αλγορίθμου της οπισθοδιάδοσης απαιτούνται τρία πράγματα:

1. Το σύνολο εκπαίδευσης που αποτελείται από ζευγάρια input-output $(\mathbf{x}_i, \mathbf{y}_i)$. Το σύνολο αυτών των ζευγαριών, μεγέθους N θα συμβολίζεται με $X = \{(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_N, \mathbf{y}_N)\}$.
2. Ένα feed-forward νευρωνικό δίκτυο με παραμέτρους θ . Κατά την οπισθοδιάδοση, οι παράμετροι του κύριου ενδιαφέροντος είναι τα βάρη, τα οποία μεταξύ του κόμβου j στο στρώμα l_k και του κόμβου i στο στρώμα l_{k-1} , συμβολίζονται με $w_{i,j}^k$ και το bias, το οποίο συμβολίζεται με b_i^k για τον κόμβο i στο l_k στρώμα. Υπόθεση αποτελεί ότι κόμβοι του ίδιου στρώματος δεν συνδέονται μεταξύ τους, καθώς και ότι τα στρώματα είναι πλήρως συνδεδεμένα (fully connected).
3. Μια αντικειμενική συνάρτηση, $E(X, \theta)$, που σε όλες τις περιπτώσεις θα θεωρείται συνάρτηση κόστους, εκτός αν διευκρινίζεται διαφορετικά, η οποία ορίζει το σφάλμα μεταξύ της επιθυμητής εξόδου \mathbf{y}_i και της υπολογισμένης από το νευρωνικό δίκτυο εξόδου $\hat{\mathbf{y}}_i$ για είσοδο \mathbf{x}_i , όπου το ζευγάρι $(\mathbf{x}_i, \mathbf{y}_i) \in X$ για μια συγκεκριμένη τιμή των παραμέτρων θ .

Για την προπόνηση του νευρωνικού δικτύου με την μέθοδο της απότομης καθόδου, απαιτείται ο υπολογισμός της κλίσης της αντικειμενικής συνάρτησης $E(X, \theta)$, ως προς τα βάρη $w_{i,j}^k$ και τα biases b_i^k . Έπειτα, σύμφωνα με την παράμετρο εκμάθησης α , κάθε επανάληψη της απότομης καθόδου ενημερώνει τα βάρη και τα biases (θ) σύμφωνα με την εξίσωση:

$$\theta^{t+1} = \theta^t - \alpha \frac{\partial E(X, \theta^t)}{\partial \theta}$$

, όπου με θ^t συμβολίζονται οι παράμετροι του δικτύου κατά την t -οστή επανάληψη.

Όπως έχει ήδη αναφερθεί, θεμελιώδης στόχος σε ένα πολυστρωματικό δίκτυο είναι η εκμάθηση καλών και εκφραστικών αναπαραστάσεων, δηλαδή κατάλληλα βάρη και biases. Σε αντίθεση με το perceptron, το οποίο με τη χρήση του delta rule, προσεγγίζει ικανοποιητικά μια τιμή στόχο, τα κρυφά στρώματα δεν διαθέτουν τιμές στόχους αφού ουσιαστικά είναι τα ενδιάμεσα βήματα του υπολογισμού. Δύο θέματα που προκύπτουν, ωστόσο επιλύει ο αλγόριθμος οπισθοδιάδοσης, είναι ότι: δεν μπορεί να

οριστεί μια αντικειμενική συνάρτηση ειδικά για κάθε κόμβο, καθώς η τιμή του εξαρτά όλους τους προηγούμενους και επόμενους κόμβους, και η περιπλοκότητα της ζεύξης παραμέτρων μεταξύ στρωμάτων, που ανορθόλογη επιλογή δύναται να καταστήσει το πρόβλημα πολύπλοκο.

Ο συμβολισμός που θα ακολουθηθεί είναι ο ακόλουθος:

- $w_{i,j}^k$: το βάρος του κόμβου j στο στρώμα l_k για τον εισερχόμενο κόμβο i
- b_i^k : το bias για τον κόμβο i στο στρώμα l_k , παρακάτω θα συμβολίζεται και ως $w_{0,j}^k$, δηλαδή σαν το εισερχόμενο βάρος από έναν υποθετικό κόμβο.
- a_i^k : η ενεργοποίηση (το άθροισμα του εσωτερικό γινόμενου των βαρών και του bias) του κόμβου i στο στρώμα l_k
- o_i^k : το output του κόμβου i στο στρώμα l_k . Επίσης, θα θεωρείται το output του bias $w_{0,j}^k$, ίσο με την μονάδα: $o_0^k = 1$
- r_k : πλήθος κόμβων στρώμα l_k
- g : η συνάρτηση ενεργοποίησης για τα ενδιάμεσα-κρυφά στρώματα
- g_0 : η συνάρτηση ενεργοποίησης για τους κόμβους του στρώματος εξόδου

Για την ανάλυση θα χρησιμοποιηθεί η συνήθης συνάρτηση κόστους που χρησιμοποιείται στα ενδιάμεσα στρώματα της οπισθοδρόμησης:

$$E(X, \theta) = \frac{1}{2N} \sum_{i=1}^N (\hat{y}_i - y_i)^2$$

Στόχος είναι η ελαχιστοποίηση της, θεωρώντας ως μεταβλητή τα βάρη. Επομένως λαμβάνεται η παράγωγος ως ακολούθως:

$$\frac{\partial E(X, \theta)}{\partial w_{i,j}^k} = \frac{1}{N} \sum_{d=1}^N \frac{\partial}{\partial w_{i,j}^k} \left(\frac{1}{2} (\hat{y}_d - y_d)^2 \right) = \frac{1}{N} \sum_{d=1}^N \frac{\partial E_d}{\partial w_{i,j}^k}$$

Ο αλγόριθμος της οπισθοδρόμησης εφαρμόζει τον κανόνα της αλυσίδας στην μερική παράγωγο της συνάρτησης κόστους E :

$$\frac{\partial E}{\partial w_{i,j}^k} = \frac{\partial E}{\partial a_j^k} \frac{\partial a_j^k}{\partial w_{i,j}^k}$$

Η αποσύνθεση αυτή του υπολογισμού αποσκοπεί στην διατύπωση πως η αλλαγή στην συνάρτηση κόστους λόγω βαρών είναι το γινόμενο της μεταβολής της συνάρτησης κόστους λόγω της ενεργοποίησης a_j^k και της μεταβολής της ενεργοποίησης εξαιτίας των βαρών.

Ο όρος $\frac{\partial E}{\partial a_j^k}$ καλείται συνήθως σφάλμα και συμβολίζεται:

$$\delta_j^k = \frac{\partial E}{\partial a_j^k}$$

Για τον δεύτερο όρο ισχύει:

$$\frac{\partial a_j^k}{\partial w_{i,j}^k} = \frac{\partial}{\partial w_{i,j}^k} \left(\sum_{l=0}^{r_k-1} w_{lj}^k o_l^{k-1} \right) = o_i^{k-1}$$

Τώρα, η μερική παράγωγος της συνάρτησης κόστους E ως προς το βάρος $w_{i,j}^k$ μπορεί να εκφραστεί ως:

$$\frac{\partial E}{\partial w_{i,j}^k} = \delta_j^k o_i^{k-1}$$

Επομένως, η μερική παράγωγος ως προς ένα βάρος είναι το γινόμενο του όρου του σφάλματος δ_j^k του κόμβου j στο στρώμα k , και η έξοδος o_i^{k-1} του κόμβου i στο στρώμα $k-1$. Μια τέτοια έκφραση γίνεται και νοητά αντιληπτή καθώς το βάρος $w_{i,j}^k$ είναι ουσιαστικά ο σύνδεσμος μεταξύ της εξόδου του κόμβου i στο στρώμα $k-1$ και της εισόδου j στο στρώμα k .

Σημειώνεται, ότι ο όρος του σφάλματος σχετίζεται με την συνάρτηση σφάλματος E , καθώς και με τις τιμές των σφαλμάτων σε επόμενα στρώματα. Εξάλλου, ακριβώς αυτό εκφράζει ο αλγόριθμος της οπισθοδιάδοσης, δηλαδή την "σειριακή" εκτίμηση του σφάλματος ξεκινώντας από το στρώμα εξόδου και πηγαίνοντας προς τα πίσω. Για την ευχρινότερη εικόνα της λειτουργίας του αλγορίθμου θα

χρησιμοποιηθούν η συνάρτηση μέσου τετραγωνικού σφάλματος και η σιγμοειδής για τα ενδιάμεσα και το τελευταίο στρώμα αντίστοιχα, όπου θα λάβει χώρα ο υπολογισμός του σφάλματος δ_j^k για την καθεμία.

Υπολογισμός για το στρώμα εξόδου

Ξεκινώντας από το τελευταίο στρώμα, η οπισθοδρόμηση προσπαθεί να ορίσει την τιμή δ_1^m , όπου m είναι το τελευταίο στρώμα. Ο δείκτης 1 εμφανίζεται καθότι στο τελευταίο στρώμα υποτίθεται ότι χρησιμοποιείται η σιγμοειδής συνάρτηση, η οποία απεικονίζει σε μια μονάχα τιμή, και για αυτό τον λόγο, δεν θα υπάρχουν παραπάνω από ένας κόμβοι στο τελευταίο στρώμα. Αφού το σφάλμα δ_1^m είναι η μερική παράγωγος ως προς a_1^m , εκφράζοντας την συνάρτηση κόστους ως προς την τιμή a_1^m λαμβάνεται ότι:

$$E = \frac{1}{2}(\hat{y} - y)^2 = \frac{1}{2}(g_0(a_1^m) - y)^2$$

, όπου $g_0(x)$ είναι η συνάρτηση ενεργοποίησης του στρώματος εξόδου.

Εφαρμόζοντας την μερική παράγωγο και τον κανόνα της αλυσίδας προκύπτει:

$$\delta_1^m = (g_0(a_1^m) - y)g_0'(a_1^m) = (\hat{y} - y)(g_0'(a_1^m))$$

Και τελικώς:

$$\frac{\partial E}{\partial w_{i,1}^m} = \delta_1^m o_i^{m-1} = (\hat{y} - y)(g_0'(a_1^m))o_i^{m-1}$$

Θεωρώντας $\sigma(x) = g(x)$ την σιγμοειδή συνάρτηση, τότε λαμβάνονται τα εξής αποτελέσματα:

$$g'(x) = \sigma(x)(1 - \sigma(x))$$

$$g_0'(x) = 1$$

$$\delta_1^m = \hat{y}_d - y_d$$

Υπολογισμός για τα ενδιάμεσα στρώματα

Για οποιοδήποτε από τα στρώματα $1 \leq k < m$ ισχύει:

$$\delta_j^k = \frac{\partial E}{\partial a_j^k} = \sum_{l=1}^{r^{k+1}} \frac{\partial E}{\partial a_l^{k+1}} \frac{\partial a_l^{k+1}}{\partial a_j^k}$$

, όπου το l λαμβάνει τιμές από το 1 μέχρι τον αριθμό κόμβων στο επόμενο στρώμα που ισούται με r^{k+1} . Σημειώνεται σε αυτό το σημείο ότι ο όρος του bias o_0^k που αναφέρεται στο βάρος w_{0j}^{k+1} είναι καθορισμένος και η τιμή του δεν εξαρτάται από τις εξόδους των προηγούμενων στρωμάτων, και άρα το l δεν λαμβάνει την τιμή 0.

Αν στην παραπάνω εξίσωση αντικατασταθεί ο πρώτος όρος του γινομένου με το σφάλμα δ_l^{k+1} , τότε προκύπτει το παρακάτω:

$$\delta_j^k = \sum_{l=1}^{r^{k+1}} \delta_l^{k+1} \frac{\partial a_l^{k+1}}{\partial a_j^k}$$

, όπου ο όρος a_l^{k+1} ισούται με:

$$a_l^{k+1} = \sum_{j=1}^{r^k} w_{jl}^{k+1} g(a_j^k)$$

Θεωρώντας την $g(x)$ την συνάρτηση ενεργοποίησης των ενδιάμεσων στρωμάτων προκύπτει:

$$\frac{\partial a_l^{k+1}}{\partial a_j^k} = w_{jl}^{k+1} g'(a_j^k)$$

Τέλος, ο όρος σφάλματος για τα ενδιάμεσα στρώματα θα είναι:

$$\delta_j^k = \sum_{l=1}^{r^{k+1}} \delta_l^{k+1} w_{jl}^{k+1} g'(a_j^k) = g'(a_j^k) \sum_{l=1}^{r^{k+1}} \delta_l^{k+1} w_{jl}^{k+1}$$

Τελικά συνολικά προκύπτει:

$$\frac{\partial E}{\partial w_{ij}^k} = \delta_j^k o_i^{k-1} \sum_{l=1}^{r^{k+1}} \delta_l^{k+1} w_{jl}^{k+1}$$

Αν η συνάρτηση ενεργοποίησης είναι η σιγμοειδής $\sigma(x)$, τότε το σφάλμα ισούται με:

$$\delta_j^k = o_j^k (1 - o_j^k) \sum_{l=1}^{r^{k+1}} \delta_l^{k+1} w_{jl}^{k+1}$$

Περιγραφή Αλγορίθμου

Έστω μια σταθερά εκμάθησης α και μια τυχαία αρχικοποίηση των βαρών w_{ij}^k . Τότε τα βήματα του αλγορίθμου είναι τα ακόλουθα:

1. Για κάθε ζευγάρι εισόδου-εξόδου (x_d, y_d) υλοποίησε την εμπρόσθια διάδοση και αποθήκευσε τα αποτελέσματα \hat{y}_d, a_j^k, o_j^k για κάθε κόμβο j του στρώματος k , ξεκινώντας από το πρώτο στρώμα (1), μέχρι το τελευταίο (m).
2. Υπολοποίησε την οπίσθια διάδοση για κάθε ζευγάρι εισόδου-εξόδου (x_d, y_d) και αποθήκευσε τα αποτελέσματα $\frac{\partial E_d}{\partial w_{ij}^k}$ για κάθε βάρος w_{ij}^k που συνδέει τον κόμβο i του στρώματος $k-1$ με τον κόμβο j στο στρώμα k , ξεκινώντας από το τελευταίο στρώμα (m), μέχρι το πρώτο στρώμα (1):

(a) Υπολόγισε τον όρο του σφάλματος για το τελευταίο στρώμα δ_1^m , κάνοντας χρήση της εξίσωσης: $\delta_1^m o_i^{m-1} = (\hat{y} - y)(g'_1(a_1^m))o_i^{m-1}$

(b) Οπισθοδρόμησε τους όρους σφάλματος δ_j^k για τα ενδιάμεσα στρώματα, ξεκινώντας από το τελευταίο στρώμα εφαρμόζοντας την εξίσωση: $\delta_j^k = g'(a_j^k) \sum_{l=1}^{r^{k+1}} \delta_j^{k+1} w_{jl}^{k+1}$

(c) Υπολόγισε τις μερικές παραγώγους του σφάλματος E_d ως προς τα βάρη χρησιμοποιώντας την εξίσωση: $\frac{\partial E}{\partial w_{i,j}^k} = \delta_j^k o_i^{k-1}$

3. Συνδύασε τις εκάστοτε παραγώγους για κάθε ζευγάρι εισόδου-εξόδου (x_d, y_d) , $\frac{\partial E_d}{\partial w_{ij}^k}$, ώστε να σχηματιστεί η ολική κλίση $\frac{\partial E(X, \theta)}{\partial w_{ij}^k}$ για το σύνολο των ζευγαριών εισόδου-εξόδου, $X = \{(x_1, y_1), \dots, (x_N, y_N)\}$,

χρησιμοποιώντας την εξίσωση: $\frac{\partial E_d}{\partial w_{ij}^k} = \frac{1}{N} \sum_{d=1}^N \frac{\partial}{\partial w_{ij}^k} \left(\frac{1}{2} (\hat{y}_d - y_d)^2 \right) = \frac{1}{2} \sum_{d=1}^N \frac{\partial E_d}{\partial w_{ij}^k}$

4. Ενημέρωσε τα βάρη σύμφωνα με τον ρυθμό εκμάθησης α και την ολική κλίση $\frac{\partial E(X, \theta)}{\partial w_{ij}^k}$ χρησιμοποιώντας την εξίσωση: $\Delta w_{ij}^k = -\alpha \frac{\partial E(X, \theta)}{\partial w_{ij}^k}$

Βαθμωτή Κάθοδος

Η βαθμωτή κάθοδος είναι μία πρώτης τάξεως επαναληπτική διαδικασία βελτιστοποίησης προς εύρεση ενός τοπικού ελαχίστου μιας παραγωγίσιμης συνάρτησης. Στα νευρωνικά δίκτυα, η χρήση έγκειται στην

γρήγορη και αποτελεσματική εύρεση των κατάλληλων βαρών κατά το στάδιο της ενημέρωσης με μικρές αποκλίσεις από την "κατάλληλη" τιμή. Η ιδέα στην οποία βασίζεται η μέθοδος είναι ότι, η καταλληλότερη κατεύθυνση για να κάνει κανείς βήματα προς το σημείο του ελαχίστου είναι λαμβάνοντας αυτή την κατεύθυνση να ταυτίζεται με την αντίθετη φορά της παραγώγου, καθότι αυτή είναι η κατεύθυνση της πιο απότομης καθόδου.

Θεωρώντας $E(X, \theta)$ την αντικειμενική συνάρτηση, όπου με θ συμβολίζονται οι παράμετροι (βάρος, bias, ...), η βαθμωτή κάθοδος λαμβάνει την παράγωγο της αντικειμενικής συνάρτησης ως προς τις παραμέτρους, σε μια συγκεκριμένη θέση και έπειτα ενημερώνει τις τελευταίες στην κατεύθυνση της παραγώγου της αντικειμενικής συνάρτησης. Η έκφραση αυτή αποτυπώνεται με την εξής σχέση:

$$\theta_{k+1} = \theta_k - \Delta\theta_k$$

, όπου με k συμβολίζεται η τιμή των παραμέτρων θ την k -οστή επανάληψη της βαθμωτής καθόδου.

Με τον όρο $\Delta\theta_k$ εννοείται η ποσότητα:

$$\Delta\theta_k = \mu \frac{\partial E}{\partial \theta}$$

, όπου μ η παράμετρος εκμάθησης, η οποία δηλώνει το μέγεθος του βήματος που θα ληφθεί κατά την διεύθυνση της παραγώγου και $\frac{\partial E}{\partial \theta}$ είναι προφανώς η παράγωγος της αντικειμενικής συνάρτησης ως προς τις παραμέτρους. Στην πραγματικότητα η διεργασία αυτή συντελείται για κάθε παράμετρο χωριστά, δηλαδή για κάθε βάρος που χρήζει ενημέρωσης χωριστά. Επομένως, η σχέση διατυπώνεται για το εκάστοτε βάρος ως:

$$w_{k+1} = w_k - \mu \frac{\partial E}{\partial w}$$

Για κάθε ζευγάρι εισόδου-εξόδου i , χρησιμοποιώντας τον κανόνα της αλυσίδας λαμβάνεται:

$$\frac{\partial E_i}{\partial w} = \frac{\partial E_i}{\partial y_i} \frac{\partial y_i}{\partial w}$$

Θεωρώντας ως αντικειμενική συνάρτηση το μέσο τετραγωνικό σφάλμα και το $y_i = x_i w$:

$$\Delta w = 2\mu x_i (y_i - \hat{y}_i)$$

Και άρα:

$$\Delta w = 2\mu \frac{1}{N} \sum_{i=1}^N x_i (y_i - \hat{y}_i)$$

Ενδεικτικώς παρουσιάζεται γραφικά η επαναληπτική λειτουργία του αλγορίθμου βαθμωτής καθόδου προς την εύρεση ελαχίστου:

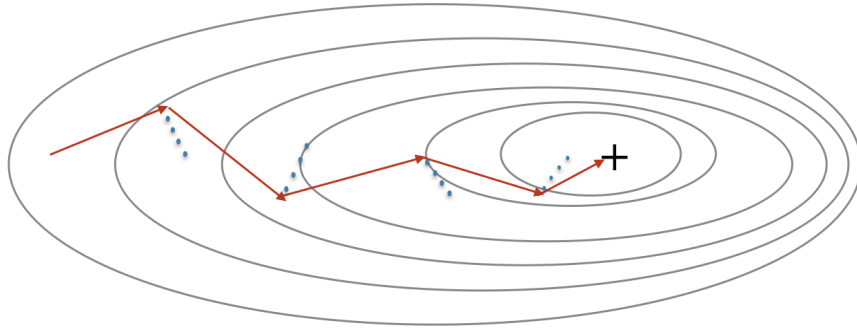


Figure 6: Βαθμωτή κάθοδος

Αλγόριθμος Βελτιστοποίησης Adam

Ο αλγόριθμος βελτιστοποίησης αποτελεί μια παραλλαγή-βελτίωση της μεθόδου βαθμωτής καθόδου, συνδυάζοντας δύο άλλες μεθόδους βελτιστοποίησης, την RMSprop και την δυναμική βαθμωτή κάθοδο. Πιθανότατα πρόκειται για αλγόριθμο που πραγματικά βελτιώνει την απόδοση της βαθμωτής καθόδου. Ο αλγόριθμος αποθηκεύει έναν εκθετικά φθίνων μέσο των προηγούμενων τετραγωνισμένων παραγώγων v_t και ακόμα έναν για τα m_t . Τα m_t και v_t είναι αντίστοιχα οι εκτιμήσεις των πρώτων και δεύτερων ροπών των παραγώγων αντιστοίχως. Περιγράφεται ως εξής:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

, όπου $g_t = \frac{\partial E}{\partial \theta}$, με το t να συμβολίζει το χρονικό βήμα.

Καθότι τα διανύσματα των m και v αρχικοποιούνται στα μηδενικά διανύσματα, μεροληπτούν προς το 0, ιδιαίτερως στα πρώτα χρονικά βήματα, και ότι οι παράμετροι β_1 και β_2 είναι κοντά στην μονάδα. Για

αυτόν τον λόγο εισάγεται ο διορθωτικός μηχανισμός μεροληψίας:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

Εν συνεχεία, η ενημέρωση των παραμέτρων υπακούει στην παρακάτω εξίσωση (όμοια με την μέθοδο βελτιστοποίησης RMSprop):

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t$$

, όπου το ϵ είναι μια παράμετρος ομαλοποίησης η οποία χρησιμοποιείται προς αποφυγή της διαίρεσης με το μηδέν και το η είναι ο ρυθμός εκμάθησης.

Οι προτεινόμενες τιμές αρχικοποίησης των παραμέτρων είναι οι εξής:

- $\eta = 0.001$
- $\beta_1 = 0.9$
- $\beta_2 = 0.999$
- $\epsilon = 10^{-8}$

Συνεχώς προτείνονται βελτιώσεις του αλγορίθμου Adam, όπως ο Nadam που συνδυάζει τον Adam με την επιταχυνόμενη παράγωγο Nesterov για την ενημέρωση των παραμέτρων, και η AMSGrad που επιλύει το πρόβλημα της βραχυπρόθεσμης μνήμης της μεθόδου Adam για προηγούμενες τιμές (που κρίνεται θεμέλιο σε εφαρμογές αναγνώρισης αντικειμένων). Ωστόσο, ο Adam βελτιώνει αισθητά την απόδοση της βαθμωτής καθόδου και χρησιμοποιείται ευρέως σε πλήθος εφαρμογών, όπου δεν κρίνεται απαραίτητη επιπλέον βελτιστοποίηση.

Mini-Batching

Οι περιπτώσεις που το σύνολο εκπαίδευσης είναι πολύ μεγάλο δεν είναι σπάνια και όλο και περισσότερες εφαρμογές των νευρωνικών δικτύων απαιτούν ή έχουν στην διάθεση τους τεράστια σύνολα δεδομένων. Με την βαθμωτή κάθοδο, προκειμένου να ολοκληρωθεί μια επανάληψη της, απαιτείται η προσπέλαση ολόκληρου του συνόλου εκπαίδευσης. Αυτή η διαδικασία είναι χρονοβόρα και ο επιστήμονας οφείλει να περιμένει την προσπέλαση εκατομμυρίων τιμών προκειμένου να παρατηρήσει μείωση της συνάρτησης κόστους. Επομένως, το mini-batching εφαρμόζεται σε τέτοιες περιπτώσεις. Το mini-batching αποτελεί πρακτική κατά την οποία το σύνολο εκπαίδευσης χωρίζεται σε μικρότερα σύνολα, με

τυχαίο ή σειριακό τρόπο, και η βαθμωτή κάθοδος εφαρμόζεται σε καθένα από αυτά τα σύνολα σειριακά. Η μέθοδος αυτή παρουσιάζει την ιδιομορφία της ασταθούς σύγκλισης κόστους. Δηλαδή κατά την εφαρμογή της, παρατηρούνται σακαμπανεύσματα στην συνάρτηση κόστους, η οποία εν τέλει θα συγκλίνει σε μια τιμή γύρω από την οποία θα ταλαντεύεται. Ακόμα και αν είναι πιο αργή στην προσπέλαση ολόκληρου του συνόλου εκπαίδευσης, πολλές φορές προτιμάται λόγω της δυνατότητας εξακρίβωσης κάποιου σφάλματος ή ατασθαλίας κατά την κατασκευή του ανευρωνικού δικτύου (ορισμός υπερπαραμέτρων και αρχιτεκτονικής του δικτύου ή λανθασμένο-ακατάλληλο σύνολο εκπαίδευσης). Μάλιστα η κατάλληλη επιλογή μεγέθους υποσυνόλων, δύναται να επιταχύνει την διαδικασία εκμάθησης.

Η τεχνική Dropout

Η τεχνική Dropout είναι μια μέθοδος κανονικοποίησης που προσιδιάζει στην παράλληλη εκπαίδευση ενός μεγάλου αριθμού νευρωνικών δικτύων με διαφορετική δομή.

Με την εφαρμογή της τεχνικής Dropout, κατά το στάδιο της εκπαίδευσης, ένας αριθμός από εξόδους στρωμάτων αγνοείται ("dropped out"), με τρόπο τυχαίο (απλή τυχαία δειγματοληψία) ή πλήρως προκαθορισμένο (γνώμη-αυθεντία ερευνητή). Αυτό έχει ως αποτέλεσμα το στρώμα στο οποίο γίνεται το dropout να μοιάζει διαφορετικό κάθε φορά με διαφορετικούς κόμβους να διαλέγονται σε κάθε εποχή. Κατά το πέρας αυτής της διαδικασίας η εκπαίδευση του νευρωνικού δικτύου καθίσταται θορυβώδης, αναγκάζοντας τους κόμβους πιθανοτικά να λαμβάνουν περισσότερη ή λιγότερη υπευθυνότητα για τα inputs. Μια τέτοια προοπτική σημαίνει ότι με την εφαρμογή της μεθόδου Dropout παύουν να ισχύουν καταστάσεις που είχαν εκμαυθεί στα προηγούμενα στρώματα, αναγκάζοντας το νευρωνικό δίκτυο να διορθώσει κάποιο λάθος εκμάθησης καθιστώντας το πιο στιβαρό με αντίτιμο τον κίνδυνο σε αυτή την περίπτωση του overfitting. Σαν παράπλευρο όφελος σημειώνεται η μείωση των κόμβων που απαιτούνται προς προπόνηση σε κάθε εποχή.

4.3 Αναδρομικά Νευρωνικά Δίκτυα - RNN

Τα αναδρομικά νευρωνικά δίκτυα (Recurrent Neural Networks, RNN) πρωτοδιατυπώθηκαν από τον David Rumelhart, αλλά έλαβαν τη γενική μορφή που εφαρμόζεται στο παρόν το 1997 με την εισαγωγή των LSTM (Long Short Term Memory) δικτύων. Πραγματικά θεωρούνται το έναυσμα για την διεύρυνση και καταξίωση των νευρωνικών δικτύων αφού επανάστησαν τον τρόπο μοντελοποίησης σημαντικών προβλημάτων, όπως η αναγνώριση φωνής, η παραγωγή μουσικής, η ανάλυση αλύσεων DNA και η παρακολούθηση βίντεο. Πολλές state-of-the-art βασίζονται στον πηρύνα τους στην εκμάθηση RNN. Τα RNN κρίνονται κατάλληλα για την παρακολούθηση διεργασιών καθότι τα δεδομένα είναι χρονικές ακολουθίες και όπως θα διαπιστωθεί, τα RNN, παρουσιάζουν εξαιρετικά αποτελέσματα στην μοντελοποίηση και ανάλυση δεδομένων υπό μορφή χρονικής ακολουθίας, ειδικά στις πολυμεταβλητές

περιπτώσεις (στην παρουσία πολλών χαρακτηριστικών). Αυτά τα χαρακτηριστικά θα συμβολίζονται για ευκολία με $x^{<t>}$ και οι αντίστοιχες αποκρίσεις ως $y^{<t>}$. Η πληθικότητα είναι αντίστοιχα T_x και T_y . Επιπρόσθετα, δεν είναι μία μέθοδος ανάλυσης που διαρκώς απαιτεί ενημέρωση παραμέτρων, μιας και ένα ήδη προπονημένο μοντέλο καλύπτει επαρκώς και επιτελεί τον ρόλο της ανάλυσης δεδομένων για αρκετά μεγάλο διάστημα, ειδικά αν δεν παρουσιαστεί concept drift (αλλαγή στην διεργασία που καθιστά το νευρωνικό δίκτυο πλέον ακατάλληλο).

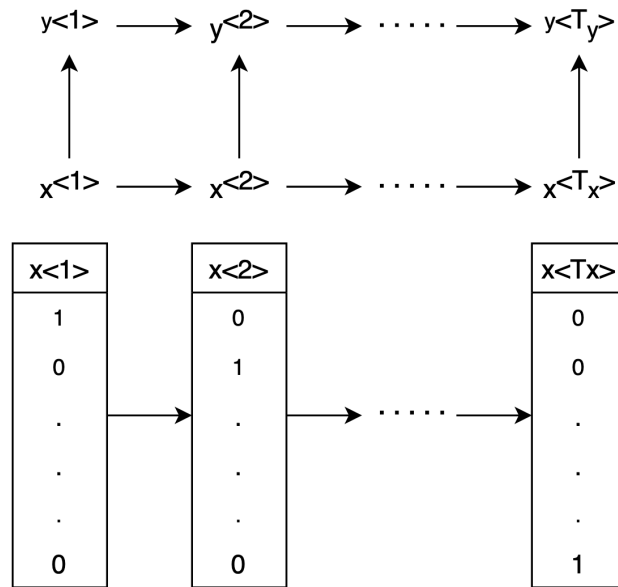


Figure 7: Κωδικοποίηση και συμβολισμός εισόδου - εξόδου

Με την κατασκευή ενός RNN το ζητούμενο είναι η εκμάθηση μιας απεικόνισης του input x στο output y . Για τον σκοπό αυτό κάλλιστα δόκιμη θα ήταν η χρήση ενός απλού νευρωνικού δικτύου, ίσως με την μοντελοποίηση των χαρακτηριστικών σε one-hot διανύσματα και την τροφοδότηση τους σε ένα δίκτυο με μερικά στρώματα. Μια τέτοια προσέγγιση, ωστόσο, δεν παρουσιάζει ικανοποιητικά αποτελέσματα επειδή το input ίσως είναι ανισομήχες με το output και δεύτερον και σημαντικότερο, η αφελής αναπαράσταση του δικτύου δεν επιτρέπει την από κοινού επεξεργασία χαρακτηριστικών σε διάφορα στρωματικά σημεία. Που σημαίνει ελάχιστη ικανοποίηση των ιδιοτήτων της αυτοσυσχέτισης, η οποία είναι επί το πλείστον συνυφασμένη με χρονικές ακολουθίες. Είναι θελκτική ιδιότητα το να γενικεύονται οι ιδιότητες ενός χαρακτηριστικού και να γίνονται κτήμα ολόκληρου του δικτύου.

Περιγραφή δομής

Έστω το πρώτο στοιχείο $x^{<1>}$ με το οποίο τροφοδοτείται το δίκτυο στο πρώτο κρυφό στρώμα και προσπαθεί να προβλέψει το output. Αφού το κάνει αυτό ελέγχει το δεύτερο στοιχείο, έστω $x^{<2>}$, αλλά αντί να προσπαθήσει να προβλέψει το output $y^{<2>}$ χρησιμοποιώντας μονάχα το $x^{<2>}$, εγκιβωτίζει και μερική πληροφορία από το πρώτο χρονικά στοιχείο $x^{<1>}$. Επομένως η τιμή απενεργοποίησης της πρώτης χρονική στιγμής μεταφέρεται σαν input για την δεύτερη. Εν συνεχεία, το νευρωνικό δίκτυο τροφοδοτείται με την επόμενη στιγμή $x^{<3>}$ και προσπαθεί να αποκριθεί το $y^{<3>}$, και ούτω καθεξής μέχρι να φτάσει στο στοιχείο εισόδου $x^{<T_x>}$, και αποκριθεί για το $y^{<T_y>}$. Αυτό βέβαια συμβαίνει αν $T_x = T_y$. Στην παρούσα εργασία καθότι απαιτείται απόκριση για κάθε input, τα T_x και T_y θα μπορούσαν να είναι ισοπληθικά, μα διατυπώνεται μια πιο εκλεπτυσμένη δομή. Ουσιαστικά σε κάθε χρονικό βήμα το RNN μεταφέρει την προηγούμενη απόκριση σαν input για το επόμενο. Για την ενεργοποίηση του πρώτου χρονικού βήματος συνήθως γίνεται χρήση μηδενικού διανύσματος ή ενός τυχαία παραγόμενου αλλά συναφή με τα δεδομένα διανύσματος. Το RNN σκανάρει τα δεδομένα από δεξιά προς τα αριστερά και οι παράμετροι που χρησιμοποιεί σε κάθε χρονικό βήμα είναι κοινόχρηστες. Για παράδειγμα, οι παράμετροι που ορίζουν την σύνδεση του $x^{<1>}$ με το πρώτο κρυφό στρώμα, θα είναι κάποιο σύνολο παραμέτρων που θα συμβολίζεται με w_{ax} και αυτές θα αποτελούν τις παραμέτρους που χρησιμοποιούνται σε κάθε χρονικό βήμα. Οι απενεργοποιήσεις, δηλαδή οι οριζόντιες συνδέσεις καθορίζονται από τις παραμέτρους w_{aa} και αυτές θα χρησιμοποιούνται σε κάθε χρονικό βήμα. Ομοίως και το άθροισμα w_{ya} που καθορίζει τις προβλέψεις για έξοδο. Για παράδειγμα, κατά την πρόβλεψη του $y_{<2>}$, λαμβάνεται πληροφορία όχι μονάχα από το $x_{<2>}$, αλλά και η πληροφορία από το $x_{<1>}$, επειδή η πληροφορία περνάει μέσω του w_{aa} . Η παρούσα υλοποίηση επεκτείνεται με απλό τρόπο και για την περίπτωση που είναι θελκτική η μετάβαση πληροφορίας με παρόμοιο τρόπο προς τα πίσω. Δηλαδή στην περίπτωση που ένα χρονικό σημείο ακολουθείται από πλήθος τιμών οι οποίες φαίνονται να έχουν κάποια συστηματική ή ανώμαλη συμπεριφορά, τότε θα ήταν θεμιτή η παροχή αυτής της πληροφορίας σε προηγούμενες χρονικές στιγμές, ώστε να αναλογισθούν τυχόν συμπεριφορές ή περιπτώσεις. Κάτι τέτοιο θα παρουσιαστεί εφόσον καταδειχθούν η οπισθοδιάδοση και μερικά κύρια συστατικά των RNN.

Ωστόσο, αξίζει να αναπτυχθεί η εμπρόσθια διάδοση του δικτύου που μόλις περιγράφηκε. Η περιγραφή θα γίνει σύμφωνα με το παρακάτω σχήμα:

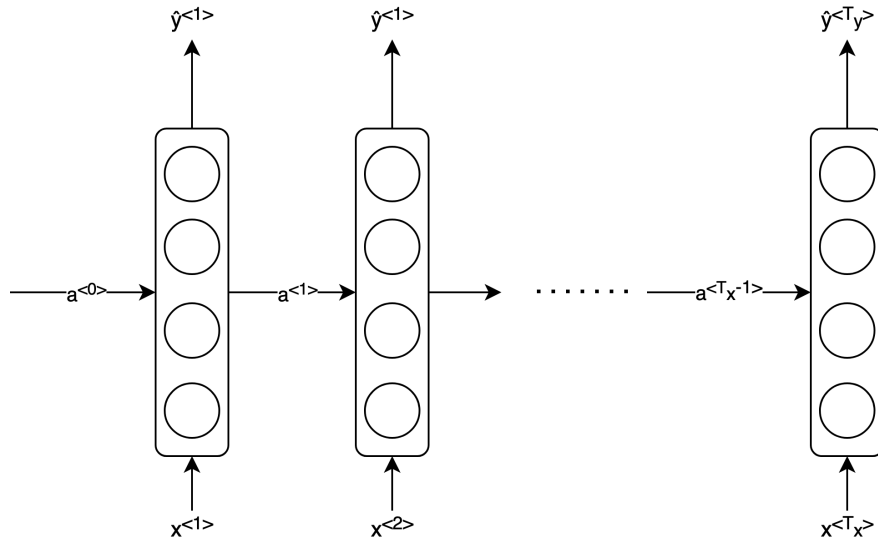


Figure 8: Συνολική δομή ενός δικτύου RNN

Για κάθε κόμβο δύναται να σχηματιστεί η παρακάτω εικόνα:

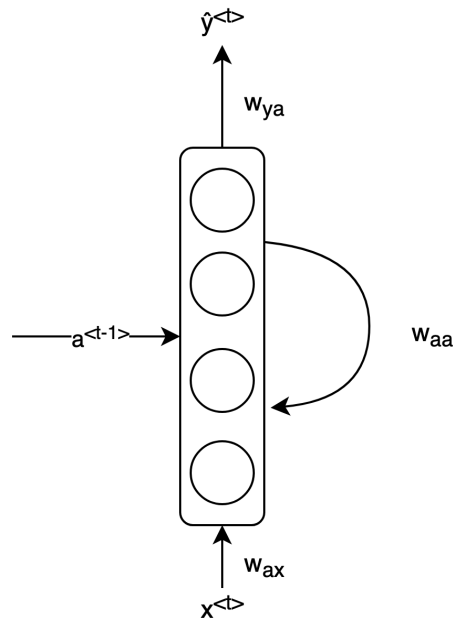


Figure 9: Κόμβος RNN

Όπως αναφέρθηκε προηγουμένως το $a^{<0>}$ αρχικοποιείται να είναι το μηδενικό ή ένα κατάλληλο

4.3 Αναδρομικά Νευρωνικά Δίκτυα - RNN

τυχαίο διάνυσμα. Για τον υπολογισμό του $a^{<1>}$, θεωρώντας g μια συνάρτηση ενεργοποίησης, η g_1 λαμβάνει σαν όρισμα το γινόμενο του w_{aa} με το $a^{<0>}$ αυξημένο κατά το γινόμενο του w_{ax} με το $x^{<1>}$ συν έναν όρο που δηλώνει το bias, b_a . Δηλαδή θα είναι:

$$a^{<1>} = g_1(w_{aa}a^{<0>} + w_{ax}x^{<1>} + b_a)$$

Για τον υπολογισμό της απόκρισης \hat{y}_1 εφαρμόζεται μια πιθανώς διαφορετική από την προηγούμενη συνάρτηση ενεργοποίησης g_2 στο γινόμενο του w_{ya} με το $a^{<1>}$ αυξημένο κατά το bias, b_y :

$$\hat{y}_1 = g_2(w_{ya}a^{<1>} + b_y)$$

Ο δεύτερος δείκτης των βαρών w_{ij} δείχνει τις διαστάσεις της αναμενόμενης ποσότητας πολλαπλασιασμού του για τον υπολογισμό μιας ποσότητας που αναφέρεται στον πρώτο δείκτη. Οι συναρτήσεις ενεργοποίησης δεν είναι αυστηρώς καθορισμένες ως προς την επιλογή τους, αλλά συνήθως επιλέγεται η υπερβολική εφαπτομένη \tanh . Επιπλέον, εύλογη κρίνεται η χρήση της σιγμοειδούς συνάρτησης όταν το output εκφράζει διμεταβλητή ταξινόμηση, της softmax για ταξινόμηση k κλάσεων.

Για ένα τυχαίο χρονικό βήμα t θα ισχύει:

$$a^{<t>} = g_1(w_{aa}a^{<t-1>} + w_{ax}x^{<t>} + b_a)$$

$$\hat{y}^{<t>} = g_2(w_{ya}a^{<t>} + b_y)$$

Η απλοποίηση της τελευταίας σχέσης γίνεται θεωρώντας w_a τον πίνακα που προκύπτει από ένωση των δύο πινάκων w_{aa} και w_{ax} κατά στήλες, αφού έχουν ίδια πρώτη διάσταση και αντίστοιχα θα σχηματίζεται και η ποσότητα που θα περιέχει το $a^{<t-1>}$ και το $x^{<t>}$.

$$w_a = [w_{aa}|w_{ax}]$$

$$[a^{<t-1>}, x^{<t>}] = \begin{bmatrix} a^{<t-1>} \\ x^{<t>} \end{bmatrix}$$

Και άρα:

$$a^{<t>} = g_1(w_a [a^{<t-1>}, x^{<t>}] + b_a)$$

Στο σχήμα παρουσιάζεται ενδεικτικά η λειτουργία ενός τυχαίου κόμβου:

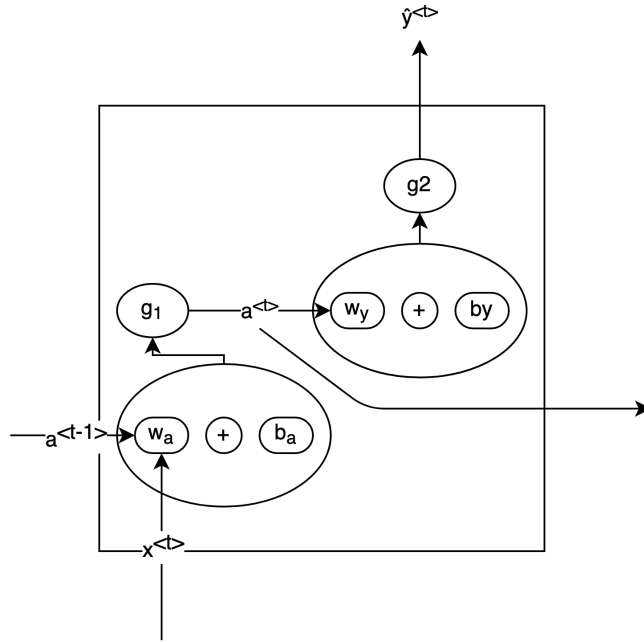


Figure 10: Λειτουργία κόμβου RNN

Οπισθοδιάδοση

Όπως και στα κλασικά νευρωνικά δίκτυα η οπισθοδιάδοση ακολουθεί την "αντίθετη" πορεία από αυτήν της εμπρόσθιας διάδοσης. Θεωρώντας την ακολουθία εισόδου $x^{<1>}, x^{<2>}$ έως $x^{<T_x>}$ υποτίθενται οι παράμετροι w_a, b_a και w_y, b_y που καθορίζουν την εμπρόσθια διάδοση. Έπειτα, για τον ορισμό της οπισθοδιάδοσης απαιτείται ο ορισμός μιας αντικειμενικής συνάρτησης, εν προκειμένω συνάρτησης απώλειας. Χάριν απλότητας, για ένα πρόβλημα διαδικής ταξινόμησης θεωρείται η logistic regression loss, η οποία καταδεικνύει και ποσοτικοποιεί το πλήθος "θέσεων" που διαφέρουν οι χρονικές ακολουθίες αν αυτές θεωρηθούν διαδικές ακολουθίες. Για κάθε χρονικό βήμα, η απώλεια είναι:

$$L^{<t>}(\hat{y}^{<t>}, y^{<t>}) = -y^{<t>} \log \hat{y}^{<t>} - (1 - y^{<t>}) \log(1 - \hat{y}^{<t>})$$

Συνεπώς, η συνολική απώλεια θα ισούται με το άθροισμα της παραπάνω ποσότητας για κάθε χρονικό βήμα είτε σε πλήθος T_x είτε σε T_y , αλλά στο παράδειγμα τα δύο αυτά μεγέθη είναι ίσα:

$$L(\hat{y}^{<t>}, y^{<t>}) = \sum_{t=1}^{T_y} L^{<t>}(\hat{y}^{<t>}, y^{<t>})$$

Ο τρόπος με τον οποίο πραγματοποιείται ένας πλήρης υπολογισμός έχει ως εξής: αφού υπολογιστεί η συνολική απώλεια L από τις επιμέρους, τότε πηγαίνοντας προς τα πίσω πρώτα αλλάζουν οι παράμετροι των y ξεκινώντας από πρώτο μέχρι το T_y και έπειτα οι παράμετροι των a ξεκινώντας από T_x μέχρι το πρώτο. Η ενημέρωση των παραμέτρων γίνεται όπως και στα κλασσικά νευρωνικά δίκτυα με τον αλγόριθμο της βαθμωτής καθόδου. Αυτός ο τρόπος λογαριασμού εμπίπτει του ονόματος οπισθοδρόμηση διά του χρόνου (backpropagation through time).

Γραφικώς, η οπισθοδιάδοση στα RNN παρουσιάζεται ως εξής:

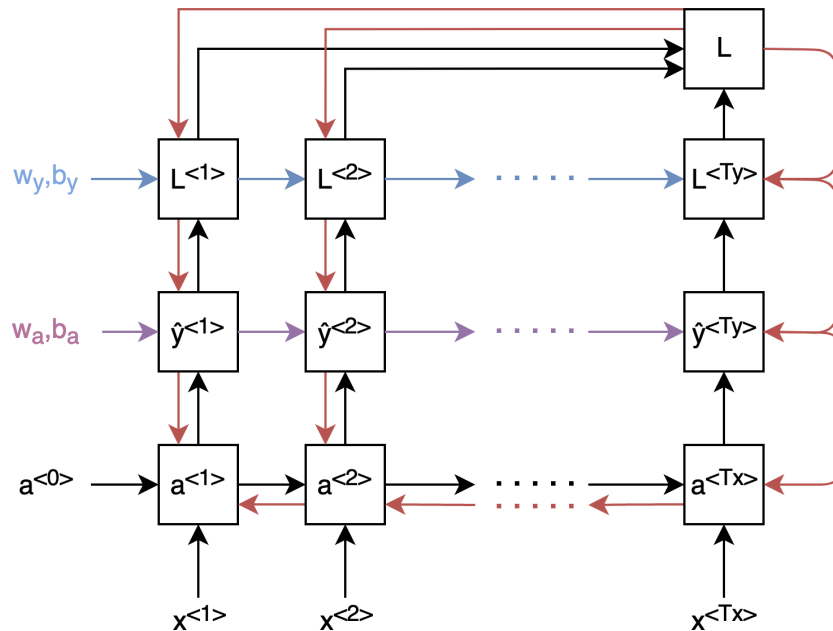


Figure 11: Οπισθοδιάδοση σε RNN

Εξαφανιζόμενες Παράγωγοι

Οι εξαφανιζόμενες παράγωγοι είναι ένα φαινόμενο το οποίο κυρίως ενδιαφέρει χρονοσειρές στις οποίες υπάρχει μακροπρόθεσμη συσχέτιση μεταξύ των χρονικών στιγμών. Τα RNN που έχουν παρουσιαστεί μέχρι στιγμής δεν συλλαμβάνουν επαρκώς ικανοποιητικά τις σχέσεις με συσχέτιση χρονικά μεγάλης διαφοράς. Αυτό αιτιολογείται από το γεγονός ότι καθώς ολοκληρωθεί η εμπρόσθια διάδοση

από τα αριστερά προς τα δεξιά σε ένα πολυστρωματικό δίκτυο, τότε η οπισθοδρόμηση από ένα μοναδικό output y είναι ένα εξαιρετικά δύσκολο εγχείρημα ώστε να επηρεαστούν τα βάρη των στρωμάτων, ειδικά όσων βρίσκονται στην "αρχή" του δικτύου. Ουσιαστικά, παρουσιάζει δυσχέρεια η "κατανόηση" από το νευρωνικό δίκτυο πως χρειάζεται να αποθηκεύσει πληροφορία των τελευταίων χρονικά στιγμών ώστε η κατάλληλη απόκριση να ενσωματωθεί και στα πρώτα χρονικά συμβάντα. Κατά συνέπεια αυτού είναι επίσης δύσκολο για το output να επηρεαστεί από κάποιο input που βρίσκεται στην αρχή της χρονοσειράς. Οποιοδήποτε είναι το ορθό output, είτε αν σε κάποια χρονική στιγμή η απόκριση του RNN είναι ορθή και σε κάποια άλλη είναι λανθασμένη, η οπισθοδρόμηση στην "αρχή" της χρονικής ακολουθίας δεν θα έχει ουσιαστικό αντίκτυπο στον τρόπο με τον οποίο λαμβάνεται η απόκριση από το δίκτυο.

Αξίζει να αναφερθεί, ότι παρόμοια με το πρόβλημα των εξαφανιζόμενων παραγώγων υπάρχει και το πρόβλημα των λεγόμενων εκρηγνυμένων παραγώγων (exploding gradients). Βέβαια, ο εντοπισμός αυτού του φαινομένου είναι εύκολος καθώς οι παράμετροι λαμβάνουν εξαιρετικά μεγάλες τιμές ή NaNs (Not a Number), όπως και η αντιμετώπιση του μέσω gradient clipping. Το gradient clipping είναι μια τεχνική κατά την οποία αν οι παράμετροι υπερβαίνουν ένα συγκεκριμένο άνω φράγμα, τότε αλλάζουν οι πολύ μεγάλες τιμές και τίθενται υπό κλίμακα.

Η επίλυση του προβλήματος των εξαφανιζόμενων παραγώγων απαιτεί μια πιο ειδική αντιμετώπιση μέσω πιο εξεζητημένων τρόπων οι οποίοι εγχολπώνουν συσχετίσεις μεγάλων χρονικών διαστημάτων και παρουσιάζονται κάτωθι.

GRU - Gated Recurrent Unit

Το GRU είναι μια τροποποίηση στο κρυφό στρώμα των RNN που τα καθιστά ικανότερα στην σύλληψη συσχετίσεων σε μεγάλη χρονική διαφορά και αρωγεί στο πρόβλημα των εξαφανιζόμενων παραγώγων. Η δομή του t -οστού νευρώνα στα RNN υπολογίζει το $a^{<t>}$, δεχόμενη σαν input το $a^{<t-1>}$ και το $x^{<t>}$, με συνήθη συνάρτηση ενεργοποίησης την υπερβολική εφαπτομένη:

$$a^{<t>} = \tanh(w_a [a^{<t-1>}, x^{<t>}] + b_a)$$

Επίσης, το $a^{<t>}$ δύναται να δωθεί σαν input ενός softmax unit για την απόκριση $y^{<t>}$. Η δομή GRU εισάγει μια καινούρια μεταβλητή, η οποία καλείται κελί μνήμης και συμβολίζεται με c (memory cell). Ο στόχος εισαγωγής τούτης της μεταβλητής είναι η παροχή μέρους μνήμης για την συγκράτηση πληροφορίας η οποία δεν είναι κοντινή στο παρόν. Η τιμή $c^{<t>}$ τίθεται ίση με την ποσότητα $a^{<t>}$. Σε κάθε χρονική στιγμή το GRU αναλογίζει μια τιμή, $\tilde{c}^{<t>}$, για την επεγγραφή του κελιού $c^{<t>}$. Ο υπολογισμός του $\tilde{c}^{<t>}$ έχει ως εξής:

$$\tilde{c}^{<t>} = \tanh(w_c [c^{<t-1>}, x^{<t>}] + b_c)$$

, με το w_c να είναι ένας πίνακας βαρών και b_c μια σταθερά που δηλώνει το bias. Επομένως, αυτή αποτελεί μια υποψήφια τιμή για την επεγγραφή του $c^{<t>}$. Η κυριότερη και σημαντικότερη ιδέα που υλοποιείται στα GRU είναι η ύπαρξη πύλης (gate). Ορίζεται η πύλη ενημέρωσης Γ_u , η οποία λαμβάνει τιμές μεταξύ του 0 και του 1. Ορίζεται ως εξής:

$$\Gamma_u = \sigma(w_u [c^{<t-1>}, x^{<t>}] + b_u)$$

, με την σ να είναι η σιγμοειδής συνάρτηση και w_u, b_u τα βάρη και η μεροληψία αντιστοίχως. Για την πλειονότητα των τιμών, η σιγμοειδής συνάρτηση λαμβάνει τιμές που είτε είναι πολύ κοντά στο 0 ή πολύ κοντά στο 1, και επομένως τέτοιες τιμές θα έχει και η συνάρτηση Γ_u . Η πύλη ουσιαστικά καταδεικνύει την αντικατάσταση ή όχι της τιμής του $c^{<t>}$ από το $\tilde{c}^{<t>}$, αν είναι 1 τότε γίνεται η αντικατάσταση, ενώ αν είναι 0 δεν γίνεται. Η τιμή της αντικατάστασης αναλυτικά είναι η εξής:

$$c^{<t>} = \Gamma_u \tilde{c}^{<t>} + (1 - \Gamma_u) c^{<t-1>}$$

Φαίνεται ότι αν η τιμή της πύλης είναι σχεδόν μονάδα τότε, η νέα τιμή του κελιού μνήμης θα ισούται με $\tilde{c}^{<t>}$, αν είναι σχεδόν μηδενική θα ισούται με την ήδη υπάρχουσα τιμή και στην περίπτωση που η πύλη έχει μια ενδιάμεση τιμή, τότε λαμβάνεται ένας κυρτός συνδυασμός των δύο προαναφερθέντων τιμών. Η τελευταία περίπτωση αν και εξαιρετικά σπάνια, αποφεύγεται και συνήθως τιμές που είναι ενδιάμεσες προτιμούνται να τίθενται ίσες με 0. Μια εξήγηση για αυτό αποτελεί το γεγονός ότι κατά αυτόν τον τρόπο το πρόβλημα των εξαφανιζόμενων παραγώγων βρίσκει λύσει αφού η τιμή του $c^{<t>}$ παραμένει απόφια σχεδόν και δεν μειώνεται εκθετικά και αλλάζει παραμόνο εάν κρίνεται τελείως απαραίτητο. Σημειώνεται ότι οι ποσότητες $c^{<t>}, \tilde{c}^{<t>}, \Gamma_u$ είναι διανύσματα ίσων μεταξύ τους διαστάσεων, οπότε οι πολλαπλασιαστικές πράξεις μεταξύ τους είναι στοιχείο προς στοιχείο. Βοηθητική κρίνεται η σύλληψη του κελιού μνήμης σαν ένα διάνυσμα που κάθε θέση του "θυμάται" και από ένα χαρακτηριστικό, η ύπαρξη ή μη του οποίου αποτυπώνεται δυαδικά. Αν μια χρονοσειρά το εμφανίζει τότε η τιμή τίθεται ίση με την μονάδα αλλιώς με το μηδέν. Ένα πλήρες κελί μνήμης GRU εμφανίζει ακόμα μία πύλη, Γ_r , η οποία έχει τον ρόλο να κρίνει την σχετικότητα μεταξύ της τιμής $c^{<t-1>}$ και της $c^{<t>}$. Αν δηλαδή η τιμή $c^{<t>}$ είναι αρκετά σχετική με τον πρόγονο της, ώστε να λάβει την τιμή. Με την εισαγωγή της πύλης Γ_r εισάγονται και τα αντίστοιχα βάρη και το bias. Στην ολότητα του, το σύνολο των εξισώσεων που περιγράφουν το GRU είναι το ακόλουθο:

$$\tilde{c}^{<t>} = \tanh(\Gamma_r w_c [c^{<t-1>}, x^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(w_u [c^{<t-1>}, x^{<t>}] + b_u)$$

4.3 Αναδρομικά Νευρωνικά Δίκτυα - RNN

$$\Gamma_r = \sigma(w_r [c^{<t-1>, x^{<t>}] + b_r)$$

$$c^{<t>} = \Gamma_u \tilde{c}^{<t>} + (1 - \Gamma_u) c^{<t-1>}$$

$$a^{<t>} = c^{<t>}$$

Παρακάτω, δίνεται σχηματικά η λειτουργία και δομή μιας μονάδας GRU:

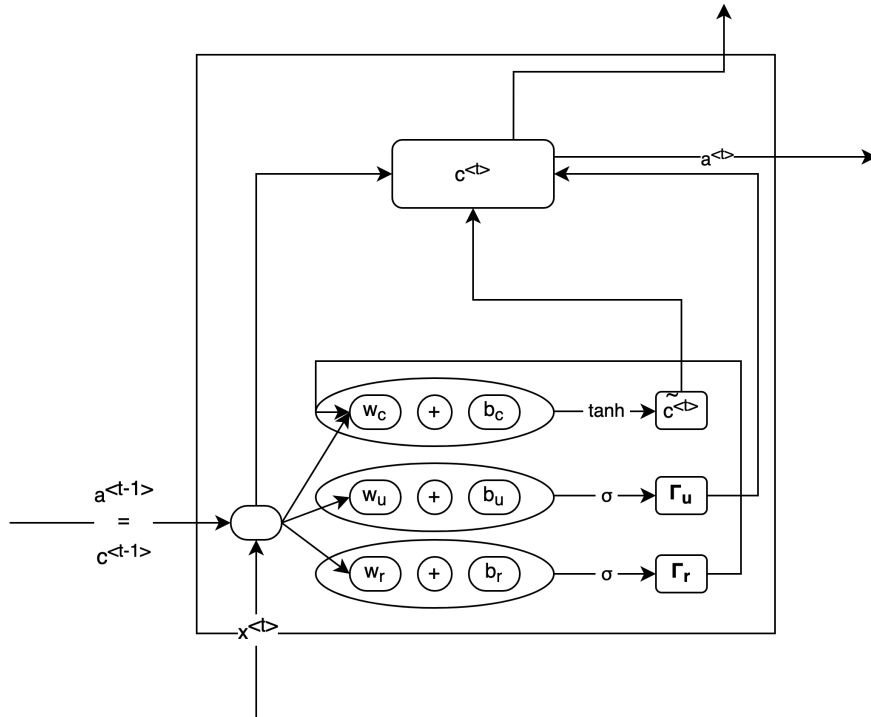


Figure 12: Gated Recurrent Unit

LSTM - Long Short Term Memory

Τα LSTMs είναι ακόμα μία δομή που χαίρει ευρείας χρήσης στα RNN για τον σκοπό της εύρεσης σχετιστικών συμπεριφορών μεγάλης χρονικής διαφοράς και αντιμετώπισης του φαινομένου της εξαφανιζόμενης παραγωγού. Μάλιστα, είναι ακόμη "ισχυρότερα" και γενικότερα από τα GRUs αν και τελευταία έχει μειωθεί η ακροαματικότητα τους εξαιτίας της πολυπλοκότερης δομής τους χωρίς αυτό βέβαια να υποτιμά την σημαντικότητά τους, εφαρμοζόμενα σε πληθώρα καίριων προβλημάτων. Οι εξισώσεις που περιγράφουν την δομή των LSTMs είναι οι ακόλουθες:

$$\tilde{c}^{<t>} = \tanh(w_c [a^{<t-1>, x^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(w_u [a^{<t-1>}, x^{<t>}] + b_u)$$

$$\Gamma_f = \sigma(w_f [a^{<t-1>}, x^{<t>}] + b_f)$$

$$\Gamma_o = \sigma(w_o [a^{<t-1>}, x^{<t>}] + b_o)$$

$$c^{<t>} = \Gamma_u \tilde{c}^{<t>} + \Gamma_f c^{<t-1>}$$

$$a^{<t>} = \Gamma_o \tanh(c^{<t>})$$

Παρατηρείται κάποια ελαφριά αλλά θεμελιώδη διαφοροποίηση με την δομή των GRUs, η οποία οφείλεται εν μέρει στο γεγονός ότι τα LSTMs αναπτύχθηκαν θέτοντας σαν θεμέλιο τη θεωρία των εξαφανιζόμενων παραγώγων προς τον σκοπό την αντιμετώπισης τους. Φαίνονται να είναι ελαφρώς περιπλοκότερες δομές, καθότι συνθέτονται από τρεις πύλες, τις Γ_u , Γ_f και Γ_o . Η πύλη της θύμησης, Γ_r , έχει απολυθεί, και έχουν οριστεί δύο διαφορετικές οι Γ_f και Γ_o που τα ονόματά τους είναι αντίστοιχα πύλη της λήθης (forget) και πύλη εξόδου (output). Εδώ, οι τιμές του κελιού μνήμης δεν ταυτίζονται με τις τιμές του $a^{<t>}$, αλλά το τελευταίο ορίζεται να είναι η τιμή της πύλης εξόδου πολλαπλασιασμένη με την υπερβολική εφαπτομένη της ποσότητας $\tilde{c}^{<t>}$. Με την σειρά της η $c^{<t>}$ επιδέχεται ενός συνδυασμού των αποκρίσεων των πυλών θύμησης και λύπης, οι οποίες ακριβώς αναλαμβάνουν να αποτυπώσουν την σημαντικότητα και σχετικότητα της προηγούμενης χρονικά πληροφορίας με το παρόν. Οι ποσότητες $a^{<t-1>}$ και $x^{<t>}$ χρησιμεύουν στον υπολογισμό των αποκρίσεων όλων των πυλών και της ποσότητας $\tilde{c}^{<t-1>}$. Ουσιαστικά το RNN θα απαρτίζεται πλέον από την χρονική σύζευξη τέτοιων μονάδων.

Μερικοί ερευνητές ανά διάφορες εφαρμογές προσαρμόζουν ελάχιστα την δομή των LSTMs εισάγοντας και την ποσότητα $c^{<t-1>}$ να λαμβάνει μέρος στον υπολογισμό των τιμών των πυλών αποβλέποντας στην ακόμα πιο ισχυρή συνένωση του $c^{<t-1>}$ και του $c^{<t>}$.

Παρακάτω, δίνεται σχηματικά η λειτουργία και δομή μιας μονάδας LSTM:

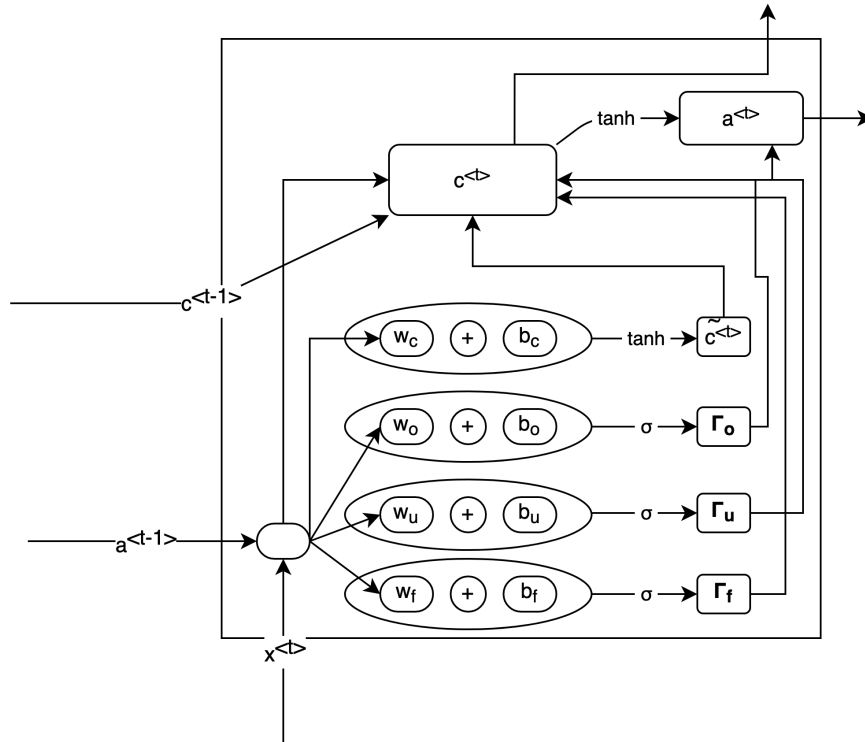


Figure 13: Long Short Term Memory Node

Ολοκληρώνοντας την περιγραφή των δομών GRU και LSTM αναφέρεται ότι αν και τελευταία είναι αποτελεσματικότερα και ισχυρότερα από τα πρώτα, τα νεοεμφανιζόμενα GRU χαίρουν χρήσης κυρίως λόγω της απλότητας τους σε σχέση με τα LSTM καθώς επεκτείνονται καλύτερα και ευκολότερα σε μεγάλα και βαθιά δίκτυα παρέχοντας σχεδόν το ίδιο ποιοτικά αποτελέσματα.

Δικατευθυντικά RNN - BRNN

Σε αυτήν την περίπτωση η γενική δομή είναι η ίδια όπως και προηγουμένως με τη διαφορά ότι αντιστοιχες τιμές όπως οι $a^{<t>}$ θα πηγαίνουν και από την αντίθετη κατεύθυνση. Δηλαδή, ορίζεται η ποσότητα $a'^{<t>}$ η οποία ξεκινάει τροφοδοτούμενη από το τελευταίο χρονικά βήμα και πηγαίνει προς τα πίσω. Αρχικοποιείται, αντιστοίχως με το $a^{<t>}$, δεχόμενη σαν input ένα μηδενικό ή τυχαίο διάνυσμα. Πλέον η απόκριση του κάθε χρονικού κόμβου χρησιμοποιεί στοιχεία από το παρελθόν, το παρόν και το μέλλον και θα ορίζεται από την εξίσωση:

$$\hat{y}^{<t>} = g_2(w_y [a^{<t>}, a'^{<t>}] + b_y)$$

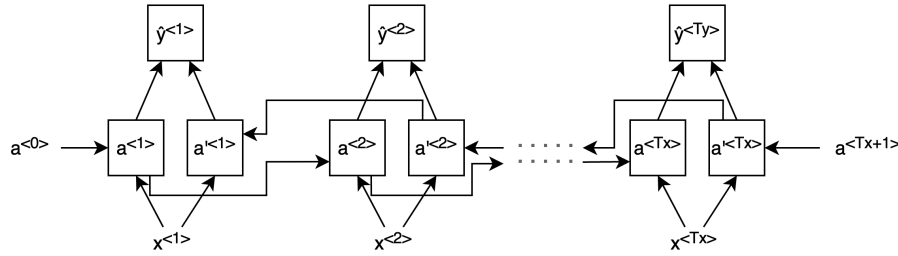


Figure 14: Δομή δικατευθυντικού RNN

Βέβαια, αυτή η απλή και αφελής επέκταση εμφανίζει το ελάττωμα της αναγκαιότητας επεξεργασίας ολόκληρης της χρονικής ακολουθίας προτού δωθεί η απόκριση. Κάποιες πιο εκλεπτυσμένες διαφοροποιήσεις της παρούσας διάταξης επιτρέπουν και την επεξεργασία δεδομένων με απόκριση σε πραγματικό χρόνο.

Βαθιά RNN - Deep RNNs

Τα Deep RNNs αποτελούν την χωρική επέκταση των κλασικών RNN. Μέχρι στιγμής τα RNN που παρουσιάστηκαν ουσιαστικά εναπόθεταν μια μονάδα υπολογισμού για κάθε χρονικό βήμα. Στην χωρική επέκταση ουσιαστικά στοιβάζονται αυτά τα χρονικά στρώματα το ένα πάνω από το άλλο εννοιολογώντας το βάθος. Τώρα, στις τιμές $a^{[l]<t>}$ προστίθεται ο δείκτης l για να δηλώσει το χωρικό στρώμα που εντοπίζεται ο κόμβος ξεκινώντας από την τιμή 1. Ο υπολογισμός που λαμβάνει χώρα σε ένα τυχαίο κόμβο είναι ο εξής:

$$a^{[l]<t>} = g_1(w_a^{[2]} [a^{[l]<t-1>}, a^{[l-1]<t>}] + b_a^{[l]})$$

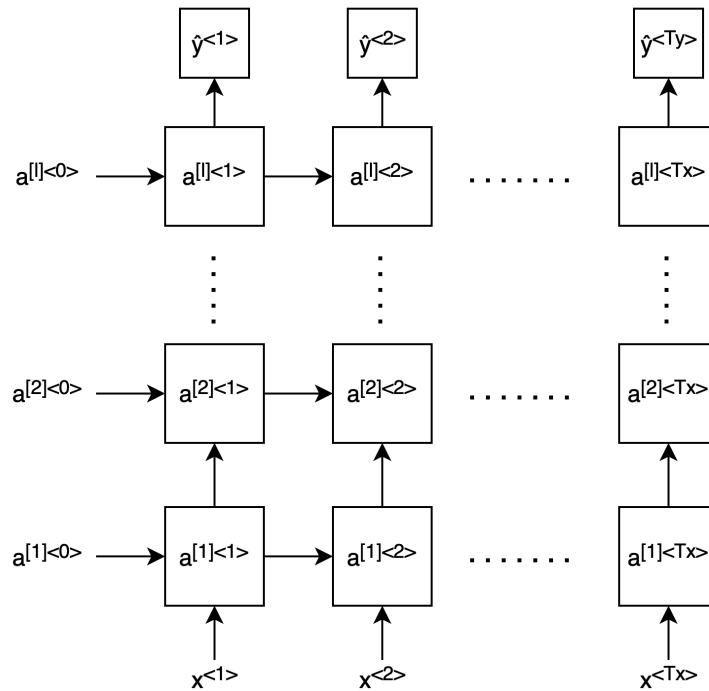


Figure 15: Δομή ενός απλού Deep RNN

Όπως είναι φυσικό οι μονάδες υπολογισμού δύναται να είναι GRUs ή LSTMs και η αρχιτεκτονική να είναι αυτή του δικατευθυντικού RNN. Αναφέρεται ότι το πλήθος των χωρικών στρωμάτων στα RNN είναι συνήθως αρκετά μικρό, καθότι πρόκειται ήδη για ένα πολύπλοκο υπολογιστικά δίκτυα και εκτός αυτού ακόμα και με πολύ λίγα χωρικά στρώματα η αποτελεσματικότητα και η σύλληψη εκφραστικών αναπαραστάσεων των RNN είναι περισσότερο από ικανοποιητική.

5 Συγκρίσεις Απόδοσης - Αποτελέσματα

5.1 Περιγραφή Δεδομένων

Ένα σύνολο δεδομένων από τον πραγματικό βιομηχανικό κόσμο αποτελεί το παρακάτω αρχείο:

Αρχείο Δεδομένων.

, που περιγράφει την παραγωγική διαδικασία ενός εργοστασίου κατασκευής χαρτοπολτού και χαρτιού. Το εργοστάσιο διαθέτει μηχανές επεξεργασίας των υλών, μεγάλου μεγέθους, οι οποίες παράγουν καρούλια χαρτιού. Διάφοροι δέκτες και αισθητήρες είναι τοποθετημένοι κατά μήκος και πλάτος των μηχανών, ώστε να παρακολουθήσουν την λειτουργία τους, μετρώντας τα ακατέργαστα υλικά και τις πρώτες ύλες που εισάγονται στις μηχανές (χημικά στοιχεία, ποσότητα πολτού, κ.ά.) καθώς επίσης και τις μεταβλητές παραγωγής (τύπος λεπίδας, απόσταση ρουλεμάν, ταχύτητα περιστροφής μηχανών, κ.ά.). Η παραγωγή χαρτιού είναι μια συνεχής και αδιάκοπη διαδικασία, που σημαίνει ότι, αν ένα κομμάτι χαρτιού κοπεί ή σπάσει στο καρούλι, η διαδικασία παύεται ολοκληρωτικά, το εν λόγω καρούλι αφαιρείται και διορθώνεται το τυχόν πρόβλημα, έπειτα επανατοποθετείται και η διαδικασία επανεκκινά. Η παύση της διαδικασίας δύναται να διαρκέσει περισσότερο από μισή ώρα. Επόμενο είναι και το τεράστιο κόστος που επιφέρει μια τέτοια δυσλειτουργία σε ένα μύλο χαρτιού. Επομένως, τα μοντέλα που έχουν γρήγορο initial response, θα προτιμούνται. Γενικά, τούτο το σύνολο δεδομένων κρίνεται κατάλληλο εφόσον είναι δεδομένα πολυμεταβλητής αυτοσυσχετισμένης χρονοσειράς. Ειδικότερα, τα δεδομένα εμφανίζουν τα εξής χαρακτηριστικά:

- 18398 καταγραφές - γραμμές
- 63 στήλες στις οποίες:
 - Στην πρώτη εμφανίζεται η χρονική στιγμή μέτρησης
 - Στην δεύτερη αποθηκεύεται η μεταβλητή απόκρισης y , η οποία λαμβάνει την τιμή 1 αν η διεργασία είναι ανεπιτυχής και την τιμή 0 αν λειτουργεί κανονικά. Στις 18398 καταγραφές η απόκριση $y = 1$ εντοπίζεται 124 φορές.
 - Οι υπόλοιπες στήλες, στο σύνολο 61, αποθηκεύονται οι επεξηγηματικές μεταβλητές - χαρακτηριστικά, οι οποίες εμφανίζονται κεντραρισμένες στο μέσο τους. Η αναλυτική τους περιγραφή ωστόσο παραλείπεται προκειμένου να διατηρηθεί η ανωνυμία σχετικά με τα δεδομένα.

5.2 mRMR

Από την προσαρμογή ενός μοντέλου προηγείται η συλλογή και αποθήκευση του συνόλου δεδομένων, αποτελούμενο από πλήθος χαρακτηριστικών, το οποίο κατά το παρόν παρουσιάζεται εύκολο και φθηνό. Από την μια πλευρά, η χρήση ολόκληρου αυτού του συνόλου, αν χρησιμοποιηθεί με σωφροσύνη, αποβαίνει σε πλεονεκτική απόδοση του μοντέλου. Ωστόσο, από την άλλη, η άμεση, δίχως προεπεξεργασία χρήση όλων των χαρακτηριστικών μιας διεργασίας και εμφλώλευσή τους στο μοντέλο αποδεικνύεται ότι οδηγεί σε υπολογιστική αναποτελεσματικότητα, υπερ-προσαρμογή, υψηλές απαιτήσεις συντήρησης και δύσκολη ερμηνευσιμότητα του μοντέλου. Η επιλογή χαρακτηριστικών είναι ένα ουσιώδες και αναγκαίο βήμα σε μεγάλης κλίμακας εφαρμογές προκειμένου να αναδειχθούν τα προτερήματα των πλούσια εκφραστικών χαρακτηριστικών υπερπηδώντας τις συναφείς προκλήσεις και κόστη. Ειδικότερα, τα οφέλη που αποκωμίζονται με την χρήση μεθόδων επιλογής χαρακτηριστικών είναι: (1) ταχύτερη υπολογιστική διεκπεραίωση, αφού με ένα μικρότερο σύνολο χαρακτηριστικών επιταχύνεται η προσαρμογή του μοντέλου καθώς και η διαδικασία της πρόβλεψης, (2) ακριβέστερη πρόβλεψη εξαιτίας της απόλειψης ανάρμοστων χαρακτηριστικών, της αποφυγής υπερ-προσαρμογής και της δυνατότητας εισαγωγής περισσότερων παρατηρήσεων στην μνήμη του υπολογιστή σε χώρο που εξασφαλίζει η απουσία πολλών πλεονάζουσων χαρακτηριστικών, (3) χαμηλότερο κόστος συντήρησης με την μείωση των απαιτούμενων χαρακτηριστικών, (4) αυξημένη ερμηνευτική ικανότητα και διαγνωστικότητα του μοντέλου, μιας και, με λιγότερα αλλά σημαντικότερα χαρακτηριστικά διευκολύνεται η πληροφοριακή και εκφραστική διάσταση της πρόβλεψης.

Σε τρεις κύριες κατηγορίες κατατάσσονται οι τεχνικές επιλογής χαρακτηριστικών: μέθοδοι φιλτραρίσματος (filter methods), μέθοδοι περιτυλίγματος (wrapper methods) και ένθετες μέθοδοι (embedded methods). Οι μέθοδοι φιλτραρίσματος χρησιμοποιούν τεχνικές κατάταξης μεταβλητών, όπως τα πρωταρχικά κριτήρια (principle criteria) για την διαλογή μεταβλητών μέσω διάταξης. Είναι απλές στην εφαρμογή και πρακτικές. Βασίζονται στην ιδέα ότι πρέπει να αποτελεί βασική ιδιότητα ενός χαρακτηριστικού η συμπερίληψη πλούσιας πληροφορίας. Τα κριτήρια που χρησιμοποιούνται κατά την υλοποίηση μεθόδων φιλτραρίσματος είναι συνήθως τα κριτήρια συσχέτισης (correlation criteria) (π.χ. Pearson) ή αμοιβαίας πληροφορίας (mutual information - MI) (π.χ. απόκλιση Kullback-Leibler). Οι μέθοδοι περιτυλίγματος χρησιμοποιούν σαν εκτιμητή ένα σύνολο το οποίο αξιολογείται μέσω μιας αντικειμενικής συνάρτησης (ο πληθάρθμος των υποσυνόλων ενός συνόλου είναι 2^N). Συνήθως γίνεται χρήση ευρετικών αλγορίθμων (γεννητικοί αλγόριθμοι όπως CHCGA) ή ακολουθιακής επιλογής (Sequential Forward/Backward Selection). Οι ένθετες μέθοδοι στοχεύουν στην μείωση του υπολογιστικού χρόνου που απαιτείται για την επαν-ταξινόμηση διαφορετικών υποσυνόλων που πραγματοποιείται στις μεθόδους περιτυλίγματος, ενσωματώνοντας την επιλογή χαρακτηριστικών κατά το στάδιο εκπαίδευσης του αλγορίθμου. Τέτοιες γνωστές μέθοδοι είναι τα Support Vector Machines (SVMs) και οι μέθοδοι που βασίζονται στους κοντινότερους γείτονες (k-NN methods). Η *MRMR* (Minimum Relevance Maxi-

mum Redudancy) μέθοδος ανήκει στην κατηγορία των μεθόδων φιλτραρίσματος και πλεονεκτεί έναντι άλλων εξαιτίας της υπολογιστικής αποδοτικότητας που παρουσιάζει καθώς και της ευρείας γενίκευσης που χαίρει σε πολλά είδη προβλημάτων.

Το κίνητρο σύλληψης και διαμόρφωσης της *mRMR* είναι η ουσιαστική μείωση περιττών χαρακτηριστικών ενώ διατηρούνται τα συναφή. Η αρχή που διακατέχει γενικά τις μεθόδους μείωσης χαρακτηριστικών ότι "τα k καλύτερα χαρακτηριστικά δεν είναι τα καλύτερα k " αναπτρώνεται και στην *mRMR*, εντοπίζοντας χαρακτηριστικά λαμβάνοντας υπ'όψιν την συνάφεια για το προβλεπόμενο αποτέλεσμα και τον πλεονασμό στις διαλεγμένες μεταβλητές. Κατά την χρήση της μεθόδου, ο ερευνητής καλείται ουσιαστικά να κάνει μονάχα μία επιλογή: τον αριθμό των χαρακτηριστικών που θα επιλέξει για να βάλει στο μοντέλο. Αυτή η επιλογή γίνεται συνήθως με εμπειρική γνώση ή λόγω περιορισμών μοντέλου, μνήμης ή χρόνου. Σε κάθε επανάληψη της μεθόδου *mRMR* εντοπίζεται το "καλύτερο" χαρακτηριστικό και προστίθεται στο σύνολο, χωρίς μετέπειτα αντικατάστασή του. Η μέθοδος πρωτοπαρουσιάστηκε από τους Pang et al. για συστήματα κατάταξης μοτίβων όπως διατυπώνεται ακριβώς παρακάτω.

Έστω ότι υπάρχουν στο σύνολο m χαρακτηριστικά και για ένα δοσμένο χαρακτηριστικό X_i ($i \in 1, 2, \dots, m$) η σημαντικότητα του βασίζεται στο *MRMR* που εκφράζεται ως:

$$f^{MRMR}(X_i) = I(Y, X_i) - \frac{1}{|S|} \sum_{X_s \in S} I(X_s, X_i)$$

,όπου Y είναι η μεταβλητή απόκρισης (σηματοδότηση κλάσης), S είναι το σύνολο των επιλεγμένων χαρακτηριστικών, $|S|$ είναι το μέγεθος του συνόλου χαρακτηριστικών, δηλαδή ο αριθμός χαρακτηριστικών, το $X_s \in S$ συμβολίζει ένα χαρακτηριστικό που ανήκει στο σύνολο χαρακτηριστικών S και το X_i συμβολίζει ένα χαρακτηριστικό που μέχρι στιγμής δεν είναι επιλεγμένο: $X_i \notin S$. Η συνάρτηση $I(\cdot)$ είναι η αμοιβαία πληροφορία:

$$I(Y, X) = \int_{\Omega_Y} \int_{\Omega_X} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) dx dy$$

, όπου Ω_Y και Ω_X είναι οι δειγματικοί χώροι που αφορούν στο Y και το X , $p(x, y)$ είναι η από κοινού συνάρτηση πυκνότητας πιθανότητας και $p(\cdot)$ είναι η περιθώρια συνάρτηση πυκνότητας πιθανότητας. Για διακριτές μεταβλητές Y και X , η φόρμουλα της από κοινού πληροφορίας λαμβάνει την εξής μορφή:

$$I(Y, X) = \sum_{y \in \Omega_Y} \sum_{x \in \Omega_X} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right)$$

Κατά τη διαδικασία διαλογής χαρακτηριστικών της *MRMR*, σε κάθε βήμα, το χαρακτηριστικό με το μεγαλύτερο σκορ σημαντικότητας:

$$\max_{X_i \notin S} f^{MRMR}(X_i)$$

θα προστεθεί στο σύνολο επιλεγμένων χαρακτηριστικών.

Για τα διακριτά χαρακτηριστικά υπάρχουν οι εξής επιλογές:

- Το MID (mutual information difference) αποτελεί την πρώτη σύλληψη της συνάρτησης f και είναι το ακόλουθο:

$$f^{MID}(X_i) = I(Y, X_i) - \frac{1}{|S|} \sum_{X_s \in S} I(X_s, X_i)$$

- Θα ήταν εφικτό αντί για την λήψη της διαφοράς να λαμβανόταν ο λόγος. Το MIQ (mutual information quotient) κάνει ακριβώς αυτό:

$$f^{MIQ}(X_i) = \frac{I(Y, X_i)}{\frac{1}{|S|} \sum_{X_s \in S} I(X_s, X_i)}$$

Για συνεχή χαρακτηριστικά, καθότι η εκτίμηση της συνάρτησης πυκνότητας πιθανότητας αποβαίνει συνήθως υπολογιστικά ακριβή, δύο εναλλακτικές προοπτικές είναι οι ακόλουθες:

- Η FCD (F-test correlation difference) χρησιμοποιεί το στατιστικό F για να αξιολογήσει την συνάφεια, και την συσχέτιση για αξιολογήσει τον πλεονασμό:

$$f^{FCD}(X_i) = F(Y, X_i) - \frac{1}{|S|} \sum_{X_s \in S} \rho(X_s, X_i)$$

, όπου $\rho(X_s, X_i)$ είναι η συσχέτιση Pearson και $F(Y, X_i)$ είναι το στατιστικό F .

- Όμοια με τη διακριτή περίπτωση το FCQ (F-test correlation quotient) λαμβάνει τον λόγο:

$$f^{FCQ}(X_i) = \frac{F(Y, X_i)}{\frac{1}{|S|} \sum_{X_s \in S} \rho(X_s, X_i)}$$

5.3 Σύσταση πειράματος

Στο παρόν και τελευταίο κεφάλαιο εξετάζεται η απόδοση των κλασικών διαγραμμάτων και μεθόδων επίβλεψης μιας διεργασίας σε σχέση με τα μοντέλα υπολοίπων, όπως το VARMA, το MESSV και τα αναδρομικά νευρωνικά δίκτυα. Η αξιολόγηση απόδοσης θα γίνει με δύο τρόπους. Πρώτα, αξιολογείται η απόδοση σε προσομοιωμένα δεδομένα, τα οποία εμφανίζουν διάφορες μορφές αυτοσυσχέτισης σε ένα πλήθος μεταβλητών με διαφορετικές εποχικότητες. Ο τρόπος παραγωγής των προσομοιωμένων δεδομένων συνίσταται στην ακόλουθη διαδικασία:

1. Παραγωγή δεδομένων από μία πολυμεταβλητή κανονική κατανομή
2. Ορισμός των απαραίτητων μητρών της VARMA διαδικασίας, ίδιας διάστασης με την διάσταση της πολυμεταβλητής κανονικής κατανομής, για την περιγραφή των αυτοπαλινδρομητικών και των στοιχείων των κινητών μέσων ως μεγέθη που περιγράφουν την διεργασία ως προς την τάση και την εποχικότητά της, τηρουμένων των περιορισμών που αφορούν τους συντελεστές.
3. Εφαρμογή των μητρών στα στοιχεία της πολυμεταβλητής κατανομής
4. Προσθήκη θορύβου σε κάθε συνιστώσα της χρονοσειράς που αφορά στον όρο σφάλματος και ακολουθεί $N(0, \epsilon)$.

Το γινόμενο των παραπάνω βημάτων είναι μια χρονοσειρά που περιγράφεται μεν από ένα μοντέλο VARMA, ωστόσο αποτελεί το κατάλληλο μέσο για την αξιολόγηση χρονοσυσχετισμένων δεδομένων που πολλές φορές συναντάται σε μια παραγωγική διαδικασία. Ειδικότερα, όσον αφορά τα προσομοιωμένα δεδομένα, για κάθε τύπο διαγράμματος εφαρμόστηκαν διάφορες τιμές των παραμέτρων και καταγράφηκαν τα μέσα μήκη ροής για τις περιπτώσεις που η διαδικασία είναι εντός ή εκτός. Σε κάθε περίπτωση που εξετάζεται στην συνέχεια προσομοιώθηκαν 1000 χρονοσειρές από δύο χιλιάδες στοιχεία για κάθε χαρακτηριστικό.

Σημειώνεται ότι ένα τέτοιο εγχείρημα επάγει υπολογιστική πολυπλοκότητα μεγάλης κλίμακας και τα μέσα προγραμματισμού και υπολογισμού της ολότητας των παραμέτρων δεν επέτρεψαν μια ενδελεχή και αναλυτική μελέτη του διακριτού χώρου τιμών ορισμού τους αν και διατηρήθηκε η βασικότερη και πραγματικότερη διακριτοποίηση των τιμών των παραμέτρων που εμφανίζεται σε πλήθος μελετών για τα παρόντα διαγράμματα.

5.3.1 Γενική Βάση - Ειδικό κριτήριο

Στην πλειονότητα των δημοσιεύσεων που αφορούν μελέτες και συγκρίσεις των πολυμεταβλητών μοντελοποιήσεων διεργασιών θεσπίζεται συνήθως το κριτήριο με βάση τα μέσα μήκη ροής των διαγραμ-

μάτων. Αρχικά, ορίζεται η τιμή στόχος για το μέσο μήκος ροής εντός ελέγχου, με συνήθεις τιμές τις 200, 370, 500, 1000. Έπειτα για κάθε διάγραμμα ελέγχου εκτιμώνται οι παράμετροι για τις οποίες ικανοποιείται το παραπάνω ARL_0 , οι οποίες προφανώς δύναται να είναι περισσότερες από μία. Καθότι, η ακριβής εύρεση τιμών που ικανοποιούν ένα συγκεκριμένο μέσο μήκος ροής είναι στριφνή διαδικασία, τις περισσότερες φορές αρκεί μια τιμή κοντά στο δοθέν ARL_0 . Τελικώς, το κριτήριο απόφασης για τον τύπο διαγράμματος που υπερτερεί έναντι άλλων κρίνεται από το μέσο μήκος ροής εκτός ελέγχου, δηλαδή πόσο γρήγορα εντοπίζει το εκάστοτε διάγραμμα την όποια μετατόπιση του μέσου, για τις τιμές παραμέτρων που ικανοποιούν το δοθέν ARL_0 , προτιμώντας πάντοτε το μικρότερο εξ αυτών, δηλαδή την γρηγορότερη ανίχνευση. Η παρούσα διπλωματική στοχεύει στην εκτέλεση παρόμοιας μεθοδολογίας με αυτήν που μόλις αναπτύχθηκε. Εξαιτίας του γεγονότος μικρού βήματος πλέγματος διακριτών τιμών αλλά ευρείας έκτασης, οι τιμές του μέσου μήκους ροής εντός ελέγχου φαίνεται να λαμβάνουν τιμές σε ένα μεγάλο διάστημα. Αποφασίστηκε η εξέταση των εξής περιπτώσεων:

1. $ARL_0 \approx 300$,δηλαδή ένα σχετικά μικρό ARL_0 για μια παραγωγική διαδικασία
2. $ARL_0 \approx 550$,δηλαδή ένα κανονικό ARL_0 για μια παραγωγική διαδικασία
3. $ARL_0 \approx 1000$,δηλαδή ένα σχετικά μεγάλο ARL_0 για μια παραγωγική διαδικασία

Μάλιστα, στην εκλογή κατάλληλου μοντέλου, προς αποφυγή έλλειψης ακρίβειας, θα διαλέγεται κάθε φορά το μοντέλο, του οποίου ο λόγος:

$$\frac{ARL_0}{ARL_1}$$

είναι μεγαλύτερος.

5.3.2 Το πρόβλημα της μετατόπισης

Στην μονομεταβλητή περίπτωση, ο τρόπος με τον οποίο εξετάζεται η απόδοση ενός διαγράμματος όταν η διεργασία τίθεται εκτός ελέγχου είναι εξαιρετικά απλός. Το μόνο που οφείλει ο ερευνητής να πράξει είναι η μετατόπιση των τιμών της χρονοσειράς κατά το μέγεθος της μετατόπισης ως προς τις μονάδες διασποράς που εμφανίζει η χρονοσειρά και να καταγράψει την στιγμή που το διάγραμμα διατείνεται πως η διεργασία είναι εκτός ελέγχου. Επαναλαμβάνοντας την παραπάνω διαδικασία μερικές χιλιάδες φορές καταλήγει στα μεγέθη ARL_1 της διαδικασίας για διάφορες τιμές των μετατοπίσεων. Μια άμεση επέκταση του παραπάνω στην πολυμεταβλητή περίπτωση δεν θα κάλυπτε το ερευνητικό ενδιαφέρον, εφόσον η μετατόπιση μιας συνιστώσας (χαρακτηριστικού) της διεργασίας πολλές φορές δεν αρκεί για να περιγράψει το σύνολο των πιθανών μονάδων με διαφορετικές αποκλίσεις που μπορεί ταυτόχρονα να βρίσκονται εκτός ελέγχου. Για τον λόγο αυτό κατασκευάζεται ένας πίνακας μετατοπίσεων της χρονοσειράς, διαφορετικός

για τις πέντε και τις δέκα διαστάσεις που θεωρείται ότι είναι κατά πολύ πιο αξιόπιστος και κατάλληλος για την εξέταση των ARL_1 . Ενδεικτικά, εξετάζονται οι εξής περιπτώσεις: Ποιο το μέσο μήκος ροής όταν μια συνιστώσα έχει μεταβληθεί κατά ενάμιση φορές την διασπορά της, μια άλλη κατά μία και ακόμα μία κατά μισή?, ή Ποιο το μέσο μήκος ροής όταν όλες οι συνιστώσες έχουν μετατοπιστεί κατά τρεις φορές την εκάστοτε διασπορά τους? (Δηλαδή η παραγωγική διαδικασία έχει αποτύχει εξ ολοκλήρου). Ακόμη ένα στοιχείο που συνηγορεί υπέρ μιας τέτοιας προσέγγισης αποτελεί το διότι κατά μία πολυμεταβλητή διεργασία δύναται να υπάρχει η λεγόμενη διασυσχέτιση (cross correlation ή inter correlation), δηλαδή η συσχέτιση μεταξύ των μεταβλητών χωροχρονικά (οι παρελθοντικές τιμές ενός χαρακτηριστικού να επηρεάζουν τις μελλοντικές ενός άλλου), παρούσης της οποίας φαίνεται αφελής η εξέταση μονάχα της μετατόπισης μιας συνιστώσας.

Ο προαναφερθέν τρόπος χρησιμοποιήθηκε για την μελέτη τεσσάρων γενικών περιπτώσεων και κάποιων υποπεριπτώσεων θεσπισμένες βάσει της αρχής της συχνότητας απάντησης τους σε πραγματικές εφαρμογές και ερευνητικού ενδιαφέροντος. Οι τέσσερις γενικές περιπτώσεις που εξετάζονται είναι οι ακόλουθες:

1. Μια διεργασία με διάσταση πέντε, η οποία περιγράφεται από το μοντέλο VARMA(1,0,0) δηλαδή μονάχα το αυτοπαλινδρομητικό κομμάτι πρώτης τάξης. Αυτό το μοντέλο μελετάται στις περισσότερες δημοσιεύσεις και χαιρεί ευρείας περιγραφής δραστηριοτήτων. Θεωρείται το απλούστερο μοντέλο από την οικογένεια των VARMA μοντέλων. Ειδικότερα, προσομοιώθηκαν δύο υποπεριπτώσεις για αυτή την διεργασία. Κατά την πρώτη, μελετήθηκαν θετικές συσχετίσεις, τυχαίες για κάθε χρονοσειρά με μέση τιμή του συντελεστή ϕ ίση με 0.5 από την κανονική κατανομή. Κατά την δεύτερη, σε αντιστοιχία με την πρώτη μελετήθηκαν αρνητικές συσχετίσεις με μέση τιμή του συντελεστή ϕ ίση με -0.5 από την κανονική κατανομή.
2. Ξεπερνώντας το δεύτερο απλούστερο μοντέλου μιας παραγωγής από διαδικασία κινητού μέσου πρώτης τάξης, η επόμενη διεργασία η οποία εξετάζεται είναι το μοντέλο VARMA(1,0,1), δηλαδή το μοντέλο που περιέχει ϕ και θ με όρους πρώτης τάξης. Όπως και στην παραπάνω περίπτωση αυτό το μοντέλο παρουσιάζεται αρκετά συχνά σε πρακτικές εφαρμογές και θεωρείται εξίσου απλό. Επιπροσθέτως, τα δεδομένα που προσομοιώνονται είναι και αυτά διάστασης πέντε με συντελεστές, για το αυτοπαλινδρομητικό κομμάτι του μοντέλου θετικής συσχέτισης με μέση τιμή ϕ ίση με 0.5 από την κανονική κατανομή, και για τους κινητούς μέσους με αρνητικές συσχετίσεις με μέση τιμή του θ ίση με -0.4 από την κανονική κατανομή.
3. Επόμενο μοντέλο που εξετάζεται είναι αυτό που περιγράφεται από το VARMA(2,0,1), ένα λίγο περιπλοκότερο μοντέλο με συντελεστές δεύτερης τάξης για το αυτοπαλινδρομητικό μέρος. Σημειώνεται ότι στο πλήθος των περιπτώσεων οι διαδικασίες εμφανίζουν μεγαλύτερης τάξης αυτοπαλινδρομητικό μέρος παρά κινητού μέσου. Εν πάση περιπτώσει, προτιμούνται απλά μοντέλα, συνήθως

μέχρι τρίτης τάξης για κάθε μέγεθος για την μοντελοποίηση χρονοσειρών, τουλάχιστον εκ πρώτης προσπάθειας μορφοποίησης του εκάστοτε προβλήματος. Σε αυτό το σημείο αναφέρεται και ένα θεώρημα περί αντιστρεψιμότητας χρονοσειρών μεταξύ της AR και της MA μοντελοποίησης που εμφωλεύεται σαν βασικό στοιχείο στα μοντέλα VARMA. Το θεώρημα συνίσταται σε τρεις αναγκαιότητες για τον ορισμό του VARMA(2, x, y), με x και y τυχαία μεγέθη. Πρώτο είναι ότι η νόρμα των συντελεστών οφείλει να είναι μικρότερη της μονάδας: $\|\phi_2\| < 1$. Δεύτερη αναγκαιότητα αποτελεί ότι το άθροισμα των ϕ οφείλει να είναι μικρότερο της μονάδας: $\phi_1 + \phi_2 < 1$. Και τρίτο, η διαφορά του όρου πρώτης τάξης από τον δεύτερο πρέπει να είναι μικρότερη της μονάδας: $\phi_2 - \phi_1 < 1$. Η διεργασία που μελετήθηκε επιλέχθηκε και πάλι να είναι διάστασης πέντε. Για τους συντελεστές του αυτοπαλινδρομητικού κομματιού θεωρήθηκε ότι θα είναι και οι δύο θετικοί, με τον πρώτο μεγαλύτερο του δεύτερου με μέση τιμή $\phi_1 = 0.6$ και για τον δεύτερο με μέση τιμή $\phi_1 = 0.3$, πάντοτε τηρουμένων των απαιτήσεων για τον ορισμό του μοντέλου δεύτερης τάξης. Για τους κινητούς επιλέχθηκαν μέσοι με αρνητικές συσχετίσεις με μέση τιμή του θ ίση με -0.4 από την κανονική κατανομή, όπως προηγουμένως.

4. Τελευταία περίπτωση που εξετάστηκε αποτέλεσε η χαοτική προσομοίωση μιας δεκαμεταβλητής διεργασίας από το μοντέλο VARMA(x, 0, z)(X, 0, Z). Το x και το z είναι αριθμοί μεταξύ του 0 και του 3, το X και το Z είναι αριθμοί μεταξύ του 0 και του 2 και το y. Συνολικά δηλαδή προκύπτει ένα μοντέλο δέκα διαστάσεων, όπου η κάθε διάσταση, δηλαδή ένα χαρακτηριστικό σε μια παραγωγική διαδικασία ακολουθεί το VARMA μοντέλο με αυτοπαλινδρομητικό μέρος μέχρι τρίτης τάξης με τυχαίους συντελεστές συσχέτισης, όπως και για το κομμάτι των κινητών μέσων, πάντα τηρουμένων των περιορισμών περί ορθότητας ορισμού της διαδικασίας. Όμοια, συμπεριλαμβάνεται και η εποχική συνιστώσα του μοντέλου με όρους μέχρι δεύτερης τάξης. Όπως γίνεται κατανοητό πρόκειται για μια υπερβολικά περίπλοκη διαδικασία, η οποία πιθανώς να μην ανταποκρίνεται σε πραγματικές καταστάσεις, αλλά επιλέχθηκε εξαιτίας του ερευνητικού ενδιαφέροντος για την περίπτωση εξαιρετικά καλύτερης προσαρμογής του μοντέλου RNN πάνω στην εξαιρετικά ποικιλόμορφη και χαρακτηριστικά δυσεπίλυτη διεργασία.

Σαν σημείωση παρουσιάζεται το γεγονός ότι όπως έγινε καταληπτό απουσιάζει το μέρος των διαφορών για τα γνωστά μοντέλα VARMA τα οποία δοκιμάστηκαν, και αυτό δεν συνέβη δίχως αιτία. Μια προσέγγιση στην οποία θα θεωρούνταν οι διαφορές k τάξης θα αύξανε κατά πολύ την περιπλοκότητα νοήματος η οποία θα απαιτούνταν να ανακαλυφθεί και να απομνημονευθεί από τα νευρωνικά δίκτυα. Ωστόσο αυτό δεν σημαίνει ότι λόγω αβλεψίας παραλείφθηκε η εξέταση αυτών των περιπτώσεων, διότι υπάρχουν πλήθος διαγνωστικών ελέγχων όπως αυτοί των Box-Pierce ή των Ljung-Box για την επιλογή ενός κατάλληλου λάμδα για την εφαρμογή του μετασχηματισμού Box-Cox ή κάποιου άλλου, προκειμένου η χρονοσειρά να γίνει στάσιμη. Επιπροσθέτως, είτε λαμβάνοντας τις διαφορές ή τις λογαριθμοποιημένες διαφορές οποιασδήποτε τάξης για τον στόχο στασιμότητας δεν στερεί πληροφορία από την διαδικασία αλλά μονάχα αλλάζει την ερμηνεία που θα δοθεί στις παραμέτρους και τις μεταβλητές του μοντέλου.

5.3.3 Τρόπος αξιολόγησης διαγραμμάτων σε πραγματικά δεδομένα

Βέβαια, είναι γνωστό ότι οι θεωρητικές ποσότητες θα πρέπει να μεταφράζονται και στην πράξη, αλλά μερικές φορές αυτός ο κανόνας δεν ισχύει. Καθότι κάθε παραγωγική διεργασία είναι διαφορετική, σε άλλες είναι πιο εύλογο να γίνει χρήση απλούστερων διαγραμμάτων όπως είναι το T^2 και άλλες φορές πιο εξειδικευμένα διαγράμματα, όπως είναι η εφαρμογή των *MCUSUM* και *MEWMA* στα υπόλοιπα της προσαρμογής ενός αναδρομικού νευρωνικού δικτύου. Είναι εμφανής η διαφορά στην τάξη των παραμέτρων που μένουν να καθοριστούν για κάθε μοντέλο, αφού στην πρώτη περίπτωση απαιτείται μονάχα η εύρεση μιας παραμέτρου, ενώ στην δεύτερη περίπτωση μένουν να εκτιμηθούν μερικές δεκάδες χιλιάδες παράμετροι και δεκάδες υπερ-παράμετροι για την εξακρίβωση της κάλλιστης δομής σε μια συγκεκριμένη διαδικασία. Επομένως, ακόμη και αν τα μέσα μήκη ροής είναι ένα ασφαλές, δοκιμασμένο και αξιόπιστο κριτήριο για την σύγκριση διαγραμμάτων, θα προτιμηθεί μια άλλη κατεύθυνση, η οποία είναι η κάτωθι.

Ευρισκόμεθα σε πλεονεκτική θέση καθότι το σύνολο δεδομένων από το εργοστάσιο παραγωγής χαρτιού έχει σημειώσει τις τιμές στις οποίες η διεργασία έχει κριθεί εκτός ελέγχου και αυτές στις οποίες είναι εντός, επιτρέποντας την εφαρμογή ενός κριτηρίου λίγο διαφορετικού από τα μέσα μήκη ροής. Ουσιαστικά, θα θεωρείται σωστή η απόκριση ενός διαγράμματος αν καταφέρνει να εντοπίσει την διεργασία εκτός λίγο πριν αυτή κριθεί εκτός ελέγχου. Αυτή η πρότερη διάγνωση δεν είναι κάποια παρόρμηση αλλά απαίτηση που δίνεται κιόλας από τους επιστήμονες που έθεσαν το ερώτημα για το σύνολο αυτών των δεδομένων, εφόσον απαίτησαν ουσιαστικά πρόγνωση της μετατόπισης της διεργασίας. Με άλλα λόγια, χωρίς να διευκρινίζεται επακριβώς, η υπόθεση ότι η διεργασία είναι εκτός ελέγχου προτού η ίδια κριθεί εκτός ελέγχου από τους μηχανικούς του εργοστασίου φαίνεται απόλυτα λογική. Επομένως, αν κάποιο διάγραμμα ανιχνεύσει μετατόπιση του μέσου της διεργασίας μέχρι δέκα στιγμές προτού αυτός κριθεί εκτός, θα θεωρείται επιτυχής ανίχνευση μετατόπισης, δηλαδή η πρακτική αυτή θα αντικαταστήσει κατά κάποιο τρόπο την πληροφορία και την σημαντικότητα του μέσου μήκους ροής εκτός ελέγχου. Χάριν πληρότητας φυσικά θα καταμετρούνται και οι φορές που το διάγραμμα εσφαλμένα εμφάνισε εκτός ορίων σημεία παρόλο που η διαδικασία είναι εντός ελέγχου. Το τελευταίο, ο καταμετρητής των false positive θα αντικαταστήσει τον ρόλο του μέσου μήκους ροής εντός ελέγχου. Συμπληρώνοντας τα παραπάνω, αξίζει να σημειωθεί, ότι η πρώτη καταμέτρηση είναι ισοδύναμη με την καταμέτρηση των false negative περιστάσεων. Στο παρόν πρόβλημα διασαφηνίζεται ότι οι false negative ποσότητες βλάπτουν κατά πολύ περισσότερο την παραγωγική διαδικασία, σε σχέση με τις false positive. Το επακριβές μέτρο κόστους αυτής της σχέσης μένει να καθοριστεί σαν πρόβλημα επιχειρησιακής έρευνας ή να αναληφθεί η διατύπωση της ακριβούς σχέσης από τους υπεύθυνους της επιχείρησης και για αυτό τον λόγο δεν θα μας απασχολήσει ιδιαίτερα, παρότι σε μια τέτοια περίπτωση μια συστάδα νευρωνικών δικτύων με κατάλληλη συνάρτηση κόστους για το πρόβλημα θα αναμένονταν ότι θα υπερτερούσε ενάντια στα κλασικά δια-

γράμματα ελέγχου. Επομένως, ακόμα και αν υπολογιστούν οι δύο καταμετρήσεις, ο λόγος αυτών στην παρούσα περίπτωση θα ήταν ανούσιος και περιττός. Συμπερασματικά και λακωνικά, κατά κάποιον τρόπο ποινικοποιείται η εσφαλμένη ένδειξη ενός διαγράμματος ότι είναι εκτός ελέγχου και επιβραβεύεται στην ορθή κρίση μετατόπισης.

5.3.4 Εφαρμογή της μεθόδου mRMR στα πραγματικά δεδομένα

Τα πραγματικά δεδομένα από το εργοστάσιο παραγωγής χαρτιού αποτελούνται από 61 μεταβλητές-χαρακτηριστικά που παρακολουθούνται, ένας αριθμός αρκετά μεγάλος που αξίζει να γίνει προσπάθεια μείωσης του. Η mRMR μέθοδος που παρουσιάστηκε προηγουμένως κρίνεται η πλέον κατάλληλη καθότι δεν χάνεται κανένα είδος πληροφορίας που η κάθε μεταβλητή κατέχει. Για την δοκιμή που η διπλωματική αυτή έφερε εις πέρας, με την εφαρμογή της μεθόδου επιλέχθηκαν οι πέντε και οι δέκα μεταβλητές μεγαλύτερης σημαντικότητας και μικρότερης συσχέτισης που φανέρωσε η μέθοδος. Πραγματικά σωτήρια φανερώθηκε η εφαρμογή της μεθόδου mRMR αφού ειδάλλως η πολυπλοκότητα των υπολογισμών θα αυξανόταν κατά ένα πολύ μεγάλο ποσοστό.

5.3.5 Περιγραφή δομής Νευρωνικού Δικτύου

Η δομή που επιλέχθηκε για το νευρωνικό δίκτυο φαίνεται να παράγει αρκετά καλά αποτελέσματα για την πρόβλεψη των χρονοσειρών, αποτελούμενο ουσιαστικά από 4 στρώματα-συστάδες. Τεχνικά, η ίδια δομή τοποθετείται μια φορά πάνω από την ίδια, αλλά προτού αναλυθεί αυτή κρίνεται απαραίτητη η ανάλυση της μοντελοποίησης των δεδομένων πριν εισαχθούν στο νευρωνικό δίκτυο.

Τα βήματα που ακολουθούνται είναι τα κάτωθι:

1. Για κάθε συνιστώσα της διεργασίας υπολογίζονται οι παράμετροι τυποποίησής της, η μέση τιμή και η διασπορά. Αν η διεργασία έχει οριστεί για συγκεκριμένο μέσο ή διασπορά τότε χρησιμοποιούνται αυτά τα μεγέθη έναντι των υπολογισμένων από τις τιμές της χρονοσειράς.
2. Τυποποιούνται τα δεδομένα βάση των παραπάνω παραμέτρων και ξεχωρίζονται σε ένα σύνολο πρόβλεψης οι πρώτες lag παρατηρήσεις. Οι lag παρατηρήσεις είναι ουσιαστικά πόσες παρατηρήσεις θα χρησιμοποιηθούν για την πρόβλεψη της χρονοσειράς. Για παράδειγμα άμα οριστούν τα lags να είναι δώδεκα, τότε αυτό σημαίνει ότι οι 12 προηγούμενες κάθε φορά παρατηρήσεις χρησιμοποιούνται για την πρόβλεψη των επόμενων, prediction το πλήθος, τιμών. Δηλαδή υπάρχουν δύο σταθερές, τα lags και τα predictions, με το πρώτο να καθορίζει το πλήθος των παρατηρήσεων για πρόβλεψη και το δεύτερο το πλήθος των παρατηρήσεων που προβλέπονται. Αυτή η πρακτική καλείται, μέθοδος μετακινούμενων παραθύρων (moving window) καθότι δεν αναλογίζεται όλες τις προηγούμενες - επόμενες παρατηρήσεις για την πρόβλεψη της επόμενης τιμής.

3. Διαχωρίζονται τα δεδομένα σε σύνολο εκπαίδευσης και σε σύνολο αξιολόγησης. Η επιλογή αυτή είναι εξαιρετικά σημαντική για την ορθή αξιολόγηση της εκπαίδευσης του νευρωνικού δικτύου. Σε παρόμοια προβλήματα, το σύνολο εκπαίδευσης συνήθως είναι το αριθμός των παρατηρήσεων μειωμένος κατά τα lags και τα predictions και αυξημένος κατά ένα.
4. Κατασκευάζεται η μεταβλητή απόκρισης y του νευρωνικού δικτύου ως οι προβλεπόμενες κάθε φορά τιμές από το νευρωνικό δίκτυο.

Μέχρι στιγμής, έχουν οριστεί κατά κάποιο τρόπο ήδη τρεις υπερ-παράμετροι του νευρωνικού δικτύου, οι σταθερές lags και predictions, και το ποσοστό διαχωρισμού του συνόλου δεδομένων σε σύνολο εκπαίδευσης και σύνολο αξιολόγησης. Τώρα περιγράφεται και αναλύεται η αρχιτεκτονική του δικτύου.

- Το πρώτο στρώμα του νευρωνικού δικτύου είναι ένα στρώμα διχατευθεντικού αναδρομικού νευρωνικού δικτύου, με έναν ενδεδειγμένο αριθμό μονάδων, που σαν έξοδο έχει την ίδια διάσταση με το input.
- Το δεύτερο στρώμα δέχεται το output του αναδρομικού στρώματος και αποτελεί ένα στρώμα στο οποίο εφαρμόζεται η τεχνική Dropout με ένα συγκεκριμένο ποσοστό.
- Το output του δεύτερου στρώματος τροφοδοτείται σε ένα τρίτο στρώμα ίδιο μορφολογικά με το πρώτο.
- Το output του τρίτου στρώματος τροφοδοτείται στο τέταρτο στρώμα που είναι ίδιο μορφολογικά με το δεύτερο.
- Τέλος, το output του τέταρτου στρώματος τροφοδοτείται στο στρώμα εξόδου, στο οποίο εφαρμόζεται μια συνάρτηση απώλειας.

Ακριβώς παρακάτω, δίνεται μια εποπτική αναπαράσταση της δομής του νευρωνικού δικτύου που θα υλοποιηθεί για την εύρεση των υπολοίπων πρόβλεψης:

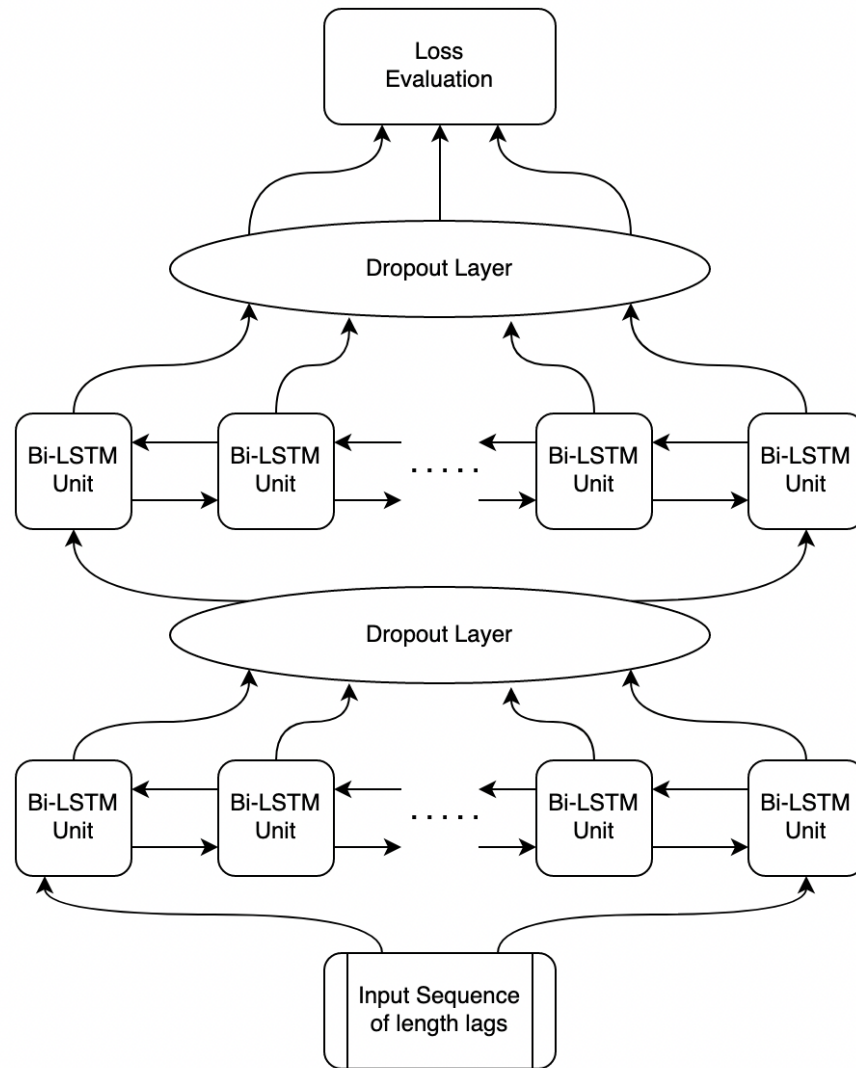


Figure 16: Δομή της επιλεγμένης σύνθεσης νευρωνικού δικτύου

Παρατήρηση

Η επιλογή του δικατευθυντικού νευρωνικού δικτύου υπό άλλες συνθήκες δεν θα ήταν η πλέον σοφή, καθότι απαιτεί την γνώση ολόκληρης της ακολουθίας προτού κάνει οποιονδήποτε υπολογισμό, δηλαδή θα εμφάνιζε πιο αργή απόδοση σε μια διαδικασία που ελέγχεται άμεσα. Βέβαια, ακόμα και αν η παραγωγική διαδικασία του εργοστασίου είναι σύγχρονη με τον έλεγχο, εξαιτίας της ανάγκης πρόγνωσης και όχι άμεσης υπογράμμισης της μετατόπισης, τα δικατευθυντικά νευρωνικά δίκτυα κρίνονται κατάλληλα ακόμα και αν απαιτούν γνώση ακόμα και του τελευταίου μέχρι στιγμής στοιχείου.

Ανακεφαλαιώνοντας, για τον αριθμό των υπερ-παραμέτρων ισχύει τώρα ότι σε αυτόν προστίθενται ο αριθμός των κρυφών κόμβων κάθε στρώματος και η σταθερά ποσοστού Dropout που θα λαμβάνει

χώρα μετά τα αναδρομικά στρώματα. Σαν υπερ-παράμετρος δύναται να οριστεί και η επιλογή της αντικειμενικής συνάρτησης, αν και αυτή που γνωρίζει ευρεία χρήση σε τέτοια προβλήματα είναι το μέσο τετραγωνικό σφάλμα καθώς επίσης και ο τύπος των units που θα χρησιμοποιηθούν (απλό RNN, GRU ή LSTM), με την τελεσίδικη απόφαση υπέρ της χρήσης των LSTMs.

Αφετέρου φυσικά διεισδώντας βαθύτερα στην μορφοποίηση των νευρωνικών δικτύων, οφείλει να καθοριστεί ο αλγόριθμος βελτιστοποίησης της βαθμωτής καθόδου, που στην παρούσα εργασία αποφασίστηκε να είναι ο αλγόριθμος Adam. Αυτός με την σειρά του φέρνει άλλες δύο υπερ-παραμέτρους που μένουν να καθοριστούν για την εκπαίδευση του μοντέλου.

Τρόπος παραγωγής σφαλμάτων

Το πλαίσιο που έχει οριστεί μέχρι στιγμής απλά κατασκευάζει και προπονεί ένα νευρωνικό δίκτυο που μαθαίνει τα χαρακτηριστικά μιας χρονοσειράς και προβλέπει την επόμενη τιμή της με όσο το δυνατόν μικρότερο σφάλμα (αποτέλεσμα της συνάρτησης απώλειας). Βέβαια σημειώνεται ότι αν αυτή δομή νευρωνικού δικτύου προπονηθεί επαρκώς, τότε συλλαμβάνει τις σχέσεις που διέπουν έναν τύπο μοντέλων VARMA. Δηλαδή, αν η εφαρμογή του νευρωνικού δικτύου γίνει πάνω σε μια VARMA(1,0,0), τότε το νευρωνικό εγκολπώνει τα χαρακτηριστικά που ορίζουν αυτού του είδους διεργασιών, δηλαδή για θετικά συσχετισμένες χρονοσειρές αυτοπαλινδρομητικού χαρακτήρα, μία εκπαίδευση νευρωνικού δικτύου καλύπτει εξ ολοκλήρου αυτήν την οικογένεια. Ειδικά αν θά ήταν σχεδόν ανώφελη η προσαρμογή ενός νευρωνικού δικτύου για κάθε διεργασία χωριστά, καθώς η πολυπλοκότητα εκπαίδευσης θα υπερνίκαιε τις μικρές βελτιώσεις των διαγραμμάτων. Οπότε, για μια προπόνηση ενός νευρωνικού δικτύου με δεδομένα μιας οικογένειας καλύπτονται όλες οι προσομοιωμένες χρονοσειρές αυτού του τύπου. Θεωρώντας ότι έχει τελεσθεί η προπόνηση ενός μοντέλου, τότε ο τρόπος με τον οποίο λαμβάνονται τα υπόλοιπα είναι μέσω της διαφοράς των προσαρμοσμένων τιμών πρόβλεψης (fitted values) από τις πραγματικές (real values). Ενθυμώμενοι το βήμα τυποποίησης χρονοσειρών οφείλει κανείς να πράξει το αντίστροφο στα fitted values, δηλαδή να πολλαπλασιάσει με την διασπορά και να προσθέσει την μέση τιμή.

5.3.6 Επιλογή παραμέτρων

Εν συνεχεία παρουσιάζονται αναλυτικότερα τα πεδία στα οποία διερευνήθηκαν οι παράμετροι για κάθε μοντέλο.

- T^2 : Για το T^2 διάγραμμα ελέγχου έχει να διευθετηθεί η τιμή μονάχα της παραμέτρου εμπιστοσύνης από την οποία εξαρτώνται τα άνω όρια ελέγχου. Η συνήθης τιμή που αυτή λαμβάνει είναι (1 –

$0.0027)^p$, όπου p ο αριθμός μεταβλητών-χαρακτηριστικών που υπάρχουν στη διεργασία. Σε αυτή την περίπτωση τα αυτοσυσχετισμένα δεδομένα απαιτούν χαλαρότερα άνω φράγματα και επομένως μικρότερα επίπεδα εμπιστοσύνης για την κατασκευή. Για την παρούσα διπλωματική, $\text{conf.level} = 0.99, 9901, \dots, 0.9998$.

- **MCUSUM**: Για το MCUSUM διάγραμμα δοκιμάστηκαν οι παράμετροι που συνήθως εφαρμόζονται για την ανίχνευση μετατοπίσεων και την εύρεση των μέσων μηκών ροής σε περιπτώσεις που απουσιάζει η αυτοσυσχέτιση διευρυμένες κατά ένα ποσοστό της τάξης του 50% προς τα πάνω. Δηλαδή, για το h δοκιμάστηκαν οι τιμές από το 2 μέχρι το 10 με βήμα 0.5 και για το k από το 0.375 μέχρι το 2.25 με βήμα 0.125. Σαν γενική παρατήρηση αναφέρεται ότι τα διάγραμμα ελέγχου για αυτοσυσχετισμένες διεργασίες τείνουν στην ολότητα τους να εμφανίζουν αρκετά μεγαλύτερα h προκειμένου να πετύχουν τις τιμές ARL_0 στόχους, με τις θετικά αυτοσυσχετισμένες να εμφανίζουν και γενικά μεγαλύτερες τιμές στα k , ενώ οι αρνητικά συσχετισμένες μικρότερα k .
- **MEWMA**: Για το διάγραμμα MEWMA οφείλουν να αποφασιστούν δύο παράμετροι: το λ που είναι η σταθερά εξομάλυνσης. Τα λ εμφανίζονται να λαμβάνουν αρκετά μεγάλες τιμές επίσης, για την επίτευξη του ζητούμενου μέσου μήκους ροής. Οι τιμές οι οποίες δοκιμάζονται είναι οι ακόλουθες: $\lambda = (0.05, 0.075, \dots, 0.925, 0.95$. Αντίστοιχα για το άνω όριο h , εφαρμόστηκαν οι τιμές που δοκιμάστηκαν στο T^2 διάγραμμα.
- **PCA**: Η μέθοδος PCA από μόνη της δεν εμφανίζει κάποια επιπλέον παράμετρο για τον προσδιορισμό των μέσων μηκών ροής, εκτός βέβαια από τον υπολογισμό του άνω φράγματος, που η συνήθης μέθοδος το έχει υπολογισμένο a priori. Ωστόσο, κάτι τέτοιο θεωρήθηκε αρκετά γενικό και για την περίπτωση των δεδομένων που εξετάζει η διεργασία αποφασίστηκε ο ορισμός μιας υπερπαραμέτρου, a , η οποία πολλαπλασιάζει το άνω φράγμα, δηλαδή τελεί τον ίδιο σκοπό με ένα επίπεδο εμπιστοσύνης του στατιστικού αλλά με αρκετά απλούστερο τρόπο. Σημειώνεται, ότι μια τέτοια προοπτική θεωρήθηκε επιβλητική καθότι γυρεύονται τα μέσα μήκη ροής εντός ελέγχου για τέσσερις διαφορετικές τιμές, κάτι που θα καθιστούσε αδύνατη την σύγκριση των PCA με τα υπόλοιπα διαγράμματα εφόσον η τιμή του μέσου μήκους ροής θα προσδιοριζόταν μονοσήμαντα. Αυτή η υπερπαραμέτρος χρησιμοποιήθηκε καθότι στις περιπτώσεις που διαλέγονταν από την μέθοδο PCA λιγότερες μεταβλητές τότε το στατιστικό δεν ήταν αντιπροσωπευτικό των ορίων. Ιδιαίτερα μεγάλες αποκλίσεις παρατηρήθηκαν στην δεκαμεταβλητή περίπτωση όταν διαλέγονται για παράδειγμα οι 5 πιο αντιπροσωπευτικές μεταβλητές, που η υπερπαραμέτρος καλούνταν να λαμβάνει τιμές στο διάστημα 5 – 8 προκειμένου τα μέσα μήκη ροής εντός ελέγχου να καταλήξουν να είναι τουλάχιστον άνω του 100.

5.3.7 Επιλογή Παραμέτρων για το RNN

Για την εκπαίδευση του νευρωνικού δικτύου, αποφασίστηκε η χρήση 350 εποχών για τα προσομοιωμένα

δεδομένα, επιλογή που κρίθηκε κατάλληλη αφενός λόγω της κατάληξης της διαδικασίας εκπαίδευσης σε plateau με αμελητέα βελτίωση στην μείωση της συνάρτησης απώλειας και αφετέρου εξαιτίας της βέλτιστης απόκρισης σε δεδομένα ελέγχου από παρόμοιες ακολουθίες που δεν ήταν πρότεινως εισαγμένα στο σύνολο εκπαίδευσης του νευρωνικού δικτύου. Δηλαδή, ικανοποιούνται οι δύο βασικές προσδοκίες της διαδικασίας εκμάθησης ενός νευρωνικού δικτύου, η καλή προσαρμογή και η απουσία overfitting (υπερ-εφαρμογής) του δικτύου πάνω στα δεδομένα εκπαίδευσης. Η σταθερά Dropout τέθηκε ίση με 25% για κάθε ένα από τα στρώματα, τιμή που συνήθως προτιμάται για την ευελξία μεγέθους της, αφού δεν είναι τόσο μικρή ώστε πρακτικά η απώλεια να λογαριάζεται από τα ίδια σχεδόν δεδομένα κάθε φορά, ούτε τόσο μεγάλη που να αντιστοιχεί σε συμπεριφορά παρόμοια με στοχαστική εκπαίδευση, αλλά περίπου το 56.25% των αρχικών δεδομένων καταλήγει κάθε φορά να συνεισφέρει στον υπολογισμό της απώλειας και να καθορίσει την διαδικασία της οπισθοδιάδοσης (backpropagation). Να σημειωθεί ότι δοκιμάστηκε η επιλογή μικρότερων συνόλων εκπαίδευσης από ότι το σύνολο των δεδομένων για κάθε εποχή, δηλαδή εφαρμογή mini-batching, ωστόσο σε ορισμένες περιπτώσεις παρατηρήθηκε αργή σύγκλιση και εκπαίδευση, πολύ μεγάλες ταλαντώσεις της συνάρτησης απώλειας και μη σύγκλιση της, για αυτούς τους λόγους και δεν προτιμήθηκε. Η επιλογή της συνάρτησης απώλειας επίσης διερευνήθηκε, με τις δύο κύριες επιλογές να είναι το μέσο τετραγωνικό σφάλμα και το μέσο απόλυτο σφάλμα. Το τελευταίο αποδείχθη πιο ακριβές, οπότε και προτιμήθηκε αν και ορθότερη οδός κρίνεται η ειδικότερη κατασκευή συνάρτησης απώλειας που λειτουργεί καλύτερα σε κάθε πρόβλημα. Ο αριθμός μονάδων LSTM σε κάθε στρώμα διαλέχθηκε να είναι ίσος με 65 - 90, ανάλογα με την περιπλοκότητα του εκάστοτε μοντέλου.

Για τα πραγματικά δεδομένα η ίδια δομή νευρωνικού δικτύου εκπαιδεύτηκε για 1000 εποχές στα πραγματικά δεδομένα, με σταθερά Dropout 40% και συνάρτηση απώλειας το μέσο απόλυτο σφάλμα. Ο αριθμός μονάδων LSTM σε κάθε στρώμα διαλέχθηκε να είναι ίσος με 120.

ΣΧΟΛΙΟ: Το overfitting είναι το φαινόμενο κατά το οποίο ένα νευρωνικό δίκτυο υπερ-εκπαιδεύεται πάνω στα δεδομένα του συνόλου εκπαίδευσης, κάτι που έχει ως αποτέλεσμα την ελαχιστοποίηση της συνάρτησης κόστους, όμως ως αντίκτυπο παρουσιάζει την αδυναμία γενίκευσης σε δεδομένα τα οποία απουσιάζουν του συνόλου εκπαίδευσης, εμφανίζοντας υπο-βέλτιστη και ίσως ακατάλληλη απόδοση σε αυτά τα δεδομένα. Το overfitting απασχολεί το σύνολο των προβλημάτων πρόβλεψης (μαζί με πλήθος άλλων προβλημάτων) και κάθε εφαρμογή οφείλει να μεριμνά για την αποφυγή του.

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

ARL1		MCUSUM							
Είδος Μετατόπισης	Αριθμός Χρονοσειρών που μετατοπίζονται	(h,k)							
		(6.5, 2.25) <i>ARL</i> ₀ = 278	(7, 2.25) <i>ARL</i> ₀ = 301	(8, 2.25) <i>ARL</i> ₀ = 472	(8.5, 2.25) <i>ARL</i> ₀ = 563	(3, 2.5) <i>ARL</i> ₀ = 322	(4, 2.5) <i>ARL</i> ₀ = 481	(4.5, 2.5) <i>ARL</i> ₀ = 537	(5, 2.5) <i>ARL</i> ₀ = 602
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(1, 0, 0, 0, 0, 0)	100	108	127	135	48	64	72	79
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(2, 0, 0, 0, 0, 0)	55	60	68	72	33	44	50	55
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(3, 0, 0, 0, 0, 0)	33	36	41	44	24	32	36	39
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(4, 0, 0, 0, 0, 0)	23	26	29	31	14	18	20	22
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(5, 0, 0, 0, 0, 0)	12	13	15	16	12	16	18	20
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(6, 0, 0, 0, 0, 0)	10	12	13	14	12	15	17	19
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(7, 0, 0, 0, 0, 0)	9	10	11	12	11	15	16	18
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(8, 0, 0, 0, 0, 0)	9	10	11	12	11	15	16	17
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(9, 0, 0, 0, 0, 0)	8	9	10	11	11	14	15	16
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(10, 0, 0, 0, 0, 0)	8	8	9	10	10	13	15	15
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 1, 0, 0, 0, 0)	56	61	70	74	34	35	39	42
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 2, 0, 0, 0, 0)	31	35	39	41	24	31	35	37
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 3, 0, 0, 0, 0)	22	25	28	30	15	20	22	23
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 4, 0, 0, 0, 0)	11	13	14	15	13	17	19	20
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 5, 0, 0, 0, 0)	10	11	12	13	12	16	18	19
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 6, 0, 0, 0, 0)	8	9	10	11	12	15	17	18
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 7, 0, 0, 0, 0)	8	8	10	10	10	14	15	16
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 8, 0, 0, 0, 0)	6	7	8	8	9	12	13	14
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 9, 0, 0, 0, 0)	5	5	6	7	8	11	12	13
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 10, 0, 0, 0, 0)	4	4	5	5	7	9	10	11
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 9, 0, 1, 0, 0)	7	8	9	10	8	11	12	13
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 8, 0, 2, 0, 0)	6	6	7	8	8	11	12	12
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 7, 0, 3, 0, 0)	3	3	4	4	5	7	8	9
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 6, 0, 4, 0, 0)	2	2	3	3	4	6	7	8
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 5, 0, 5, 0, 0)	2	2	3	3	4	6	7	7
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 4, 0, 6, 0, 0)	2	2	2	2	4	5	6	6
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 3, 0, 7, 0, 0)	2	2	2	2	3	4	4	4
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 2, 0, 8, 0, 0)	1	2	2	2	2	3	3	3
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 1, 0, 9, 0, 0)	1	1	2	2	2	2	2	2
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 10, 0, 0)	1	1	1	2	1	1	1	1
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 1, 0, 0, 0)	13	14	16	17	8	11	12	13
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 2, 0, 0, 0)	8	9	10	11	7	10	11	11
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 3, 0, 0, 0)	4	5	5	6	6	8	9	9
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 4, 0, 0, 0)	3	3	4	4	5	7	7	8
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 5, 0, 0, 0)	2	3	3	3	4	6	7	7
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 6, 0, 0, 0)	2	2	3	3	4	5	5	6
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 7, 0, 0, 0)	2	2	3	3	3	4	4	5
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 8, 0, 0, 0)	2	2	3	3	3	4	4	4
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 9, 0, 0, 0)	2	2	2	2	2	3	3	3
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 10, 0, 0, 0)	2	2	2	2	2	2	2	3
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 9, 1, 0, 0)	2	2	2	2	2	2	2	2
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(1, 0, 8, 1, 0, 0)	2	2	2	2	2	2	2	2
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(2, 0, 7, 1, 0, 0)	2	2	3	3	2	3	3	3
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(3, 0, 6, 1, 0, 0)	2	2	3	3	2	3	3	3
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(4, 0, 5, 1, 0, 0)	2	2	3	3	3	4	4	4

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

ARL1		MCUSUM							
Είδος Μετατόπισης	Αριθμός Χρονοσειρών που μετατοπίζονται	(h,k)							
		(6.5, 2.25) <i>ARL</i> ₀ = 278	(7, 2.25) <i>ARL</i> ₀ = 301	(8, 2.25) <i>ARL</i> ₀ = 472	(8.5, 2.25) <i>ARL</i> ₀ = 563	(3, 2.5) <i>ARL</i> ₀ = 322	(4, 2.5) <i>ARL</i> ₀ = 481	(4.5, 2.5) <i>ARL</i> ₀ = 537	(5, 2.5) <i>ARL</i> ₀ = 602
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(5, 0, 4, 1, 0, 0)	2	2	3	3	3	4	5	5
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(6, 0, 3, 1, 0, 0)	3	3	4	4	4	5	5	6
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(7, 0, 2, 1, 0, 0)	5	5	6	7	5	6	6	6
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(8, 0, 1, 1, 0, 0)	7	8	9	10	5	7	7	7
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(9, 0, 0, 1, 0, 0)	8	9	10	11	9	12	13	14
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 9, 0, 1, 0)	2	2	3	3	2	3	3	3
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 1, 8, 0, 1, 0)	2	2	3	3	2	3	3	3
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 2, 7, 0, 1, 0)	2	2	3	3	2	3	3	3
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 3, 6, 0, 1, 0)	2	2	3	3	2	3	3	3
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 4, 5, 0, 1, 0)	2	2	3	3	2	3	3	4
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 5, 4, 0, 1, 0)	2	2	3	3	2	3	4	4
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 6, 3, 0, 1, 0)	2	2	3	3	2	3	4	4
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 7, 2, 0, 1, 0)	2	2	3	3	2	3	4	4
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 8, 1, 0, 1, 0)	2	2	3	3	2	3	4	5
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 9, 1, 0, 1, 0)	2	3	3	3	2	3	4	5
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 1, 0, 0)	3	3	4	4	3	4	5	5
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 2, 0, 0)	2	2	3	3	4	4	5	5
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 3, 0, 0)	2	2	3	3	3	4	4	4
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 4, 0, 0)	1	2	2	2	2	3	3	3
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 5, 0, 0)	1	2	2	2	2	3	3	3
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 6, 0, 0)	1	1	2	2	1	2	2	2
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 7, 0, 0)	1	1	2	2	1	1	1	2
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 8, 0, 0)	1	1	1	1	1	1	1	2
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 9, 0, 0)	1	1	1	1	1	1	1	1
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 10, 0, 0)	1	1	1	1	1	1	1	1
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 9, 0, 1)	1	1	1	1	1	1	1	1
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 1, 0, 8, 0, 1)	1	1	1	1	1	1	1	1
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 2, 0, 7, 0, 1)	1	1	2	2	1	1	1	1
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 3, 0, 6, 0, 1)	1	1	2	2	1	1	1	2
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 4, 0, 5, 0, 1)	1	1	2	2	1	1	1	2
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 5, 0, 4, 0, 1)	1	1	2	2	1	1	2	2
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 6, 0, 3, 0, 1)	1	1	2	2	1	2	2	2
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 7, 0, 2, 0, 1)	2	2	2	2	1	2	2	2
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 8, 0, 1, 0, 1)	2	2	3	3	1	2	2	2
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 9, 0, 0, 0, 1)	2	2	3	3	1	2	2	2
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 0, 0, 1)	2	2	3	3	1	2	3	3
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 0, 0, 2)	2	2	2	2	1	2	2	2
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 0, 0, 3)	1	1	2	2	1	1	2	2
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 0, 0, 4)	1	1	2	2	1	1	1	1
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 0, 0, 5)	1	1	1	1	1	1	1	1
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 0, 0, 6)	1	1	1	1	1	1	1	1
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 0, 0, 7)	1	1	1	1	1	1	1	1
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 0, 0, 8)	1	1	1	1	1	1	1	1
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 0, 0, 9)	1	1	1	1	1	1	1	1
(0.5, 1, 1.5 ,2 ,2.5 ,3)	(0, 0, 0, 0, 0, 10)	1	1	1	1	1	1	1	1

Table 1: Τιμές Παραμέτρων και Μέσα Μήκη Ροής Εντός και Εκτός Ελέγχου για Διαδικασία VARMA(x,0,z)(X,0,Z) για τα διαγράμματα MCUSUM

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

Τα αποτελέσματα όπως παρουσιάζονται στον παραπάνω πίνακα για το διάγραμμα MCUSUM εφαρμόστηκαν και στα υπόλοιπα είδη διαγραμμάτων. Αρχικά για τα δεδομένα έχουμε τα εξής συμπεράσματα:

1. VARMA(1,0,0): Για αυτού του είδους διεργασίες, όλα τα διαγράμματα ήταν σε θέση να αντεπεξέλθουν ικανοποιητικά στον εντοπισμό τιμών εκτός ελέγχου. Ειδικότερα, η υποπερίπτωση των αρνητικών αυτοσυσχετισμών στα δεδομένα, για τις ίδιες τιμές ARL_0 , έδωσε κατά πολύ μικρότερες τιμές ARL_1 , όπως παρουσιάζεται παρακάτω:

(a) Θετική Αυτοσυσχέτιση:

- T^2

Στον παρακάτω πίνακα παρουσιάζονται οι τιμές των παραμέτρων εμπιστοσύνης που εμφανίζουν τα ζητούμενα μέσα μήκη ροής εντός ελέγχου:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
0.9968	0.9983	0.9993

Table 2: Τιμές Παραμέτρων για τα Ζητούμενα Μέσα Μήκη Ροής Εντός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) για το T^2 διάγραμμα

Για τις ως άνω τιμές υπολογίστηκαν τα μέσα μήκη ροής εκτός ελέγχου της διεργασίας, με τα αποτελέσματα να είναι τα ακόλουθα:

Για μικρές μετατοπίσεις τα διαγράμματα T^2 φαίνονται να είναι τα λιγότερο αποδοτικά δίχως βέβαια να είναι τόσο πίσω σε σχέση με τα πιο αυστηρά διαγράμματα ελέγχου. Ενδεικτικά, για μετατόπιση 0.5σ σε δύο χαρακτηριστικά της διεργασίας προκύπτουν ARL_1 ίσα με 39.48, 47.23 και 56.24 αντιστοίχως. Για μεσαίες μετατοπίσεις φαίνεται πως συμβαδίζουν σχεδόν σε απόδοση με τα άλλα διαγράμματα. Τα ARL_1 για μετατόπιση δύο τυπικών αποκλίσεων σε ένα χαρακτηριστικό και μιας σε ένα άλλο χαρακτηριστικό φαίνονται στον παρακάτω πίνακα.

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
9.74	12.94	15.43

Table 3: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) με μετατόπιση του μέσου κατά $(1 * 2\sigma, 1 * 1\sigma)$ για το T^2 διάγραμμα

Για μετατόπιση τριών τυπικών αποκλίσεων από τον μέσο δύο χρονοσειρών, τα μέσα μήκη ροής εκτός ελέγχου είναι τα ακόλουθα:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
3.1	4.13	4.89

Table 4: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) με μετατόπιση του μέσου κατά $(2 * 3\sigma)$ για το T^2 διάγραμμα

- MCUSUM

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
(9.5, 1.25)	(8.5, 1.5)	(10, 1.5)
(7.5, 1.5)	(5.5, 2)	(8.25, 1.75)
(4.5, 2)	(4.5, 2.25)	(5.5, 2.25)

Table 5: Τιμές Παραμέτρων για τα ζητούμενα Μέσα Μήκη Ροής Εντός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) για το MCUSUM διάγραμμα

Για τις ως άνω τιμές υπολογίστηκαν τα μέσα μήκη ροής εκτός ελέγχου της διεργασίας, με τα αποτελέσματα να είναι τα ακόλουθα:

Για μικρές μετατοπίσεις του μέσου της διεργασίας σε οποιονδήποτε αριθμό μεταβλητών τα καλύτερα ARL_1 λαμβάνονται από τα μικρότερα h σε κάθε περίπτωση. Ενδεικτικά, για

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

μετατόπιση 0.5σ σε δύο χαρακτηριστικά της διεργασίας προκύπτουν ARL_1 ίσα με 34, 37 και 42 αντιστοίχως. Για μεσαίες μετατοπίσεις φαίνεται οι ενδιαμέσες τιμές για τα ζευγάρια παραμέτρων να δίνουν πλέον ικανοποιητικά ARL_1 όπως φαίνεται στον παρακάτω πίνακα.

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
$(h,k) = (7.5, 1.5)$	$(h,k) = (5.5, 2)$	$(h,k) = (8.25, 1.75)$
8.82	10.65	12.78

Table 6: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) με μετατόπιση του μέσου κατά $(1 * 2\sigma, 1 * 1\sigma)$ για το MCUSUM διάγραμμα

Για μεγαλύτερες μετατοπίσεις στο μέσο, δηλαδή όταν ένα ή περισσότερα στοιχεία της διαδικασίας έχουν καταστεί ακατάλληλα, τα μέσα μήκη ροής εκτός ελέγχου ανιχνεύουν αρκετά γρήγορα τις μετατοπίσεις. Ενδεικτικά, για μετατόπιση τριών τυπικών αποκλίσεων από τον μέσο δύο χρονοσειρών, τα μέσα μήκη ροής εκτός ελέγχου είναι τα ακόλουθα:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
$(h,k) = (4.5, 2)$	$(h,k) = (4.5, 2.25)$	$(h,k) = (5.5, 2.25)$
2.78	3.76	4.41

Table 7: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) με μετατόπιση του μέσου κατά $(2 * 3\sigma)$ για το MCUSUM διάγραμμα

- MEWMA Για τα διαγράμματα MEWMA ακολουθήθηκε παρόμοια προσέγγιση όπως παραπάνω. Οι παράμετροι που ικανοποιούν τα μέσα μήκη ροής εντός ελέγχου που θέτονται ως στόχοι παρουσιάζονται στον παρακάτω πίνακα:

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
0.7	0.85	0.925

Table 8: Τιμές Παραμέτρων για τα ζητούμενα Μέσα Μήκη Ροής Εντός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) για το MEWMA διάγραμμα

Για τις ως άνω τιμές υπολογίστηκαν τα μέσα μήκη ροής εκτός ελέγχου της διεργασίας, με τα αποτελέσματα να είναι τα ακόλουθα:

Για μικρές μετατοπίσεις τα διαγράμματα MEWMA φαίνονται να είναι περισσότερο αποδοτικά από τα MCUSUM. Ενδεικτικά, για μετατόπιση 0.5σ σε δύο χαρακτηριστικά της διεργασίας προκύπτουν ARL_1 ίσα με 24.54, 25.2 και 27.12 αντιστοίχως. Για μεσαίες μετατοπίσεις φαίνεται πως συμβαδίζουν σχεδόν σε απόδοση τα δύο διαγράμματα, με λίγο καλύτερα αποτελέσματα για το MCUSUM. Τα ARL_1 για μετατόπιση δύο τυπικών αποκλίσεων σε ένα χαρακτηριστικό και μιας σε ένα άλλο χαρακτηριστικό φαίνονται στον παρακάτω πίνακα.

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
8.31	11.8	13.91

Table 9: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) με μετατόπιση του μέσου κατά $(1 * 2\sigma, 1 * 1\sigma)$ για το MEWMA διάγραμμα

Για μεγαλύτερες μετατοπίσεις στο μέσο, δηλαδή όταν ένα ή περισσότερα στοιχεία της διαδικασίας έχουν καταστεί ακατάλληλα, τα μέσα μήκη ροής εκτός ελέγχου ανιχνεύουν αρκετά γρήγορα τις μετατοπίσεις, ωστόσο αργότερα από τα MCUSUM διαγράμματα. Ενδεικτικά, για μετατόπιση τριών τυπικών αποκλίσεων από τον μέσο δύο χρονοσειρών, τα μέσα μήκη ροής εκτός ελέγχου είναι τα ακόλουθα:

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
2.9	4.13	4.89

Table 10: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) με μετατόπιση του μέσου κατά $(2 * 3\sigma)$ για το MEWMA διάγραμμα

- PCA Στον παρακάτω πίνακα παρουσιάζονται οι τιμές της πολλαπλασιαστικής υπερ-παραμέτρου που εμφανίζουν τα ζητούμενα μέσα μήκη ροής εντός ελέγχου:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
1.62	1.91	2.14

Table 11: Τιμές Παραμέτρων για τα Ζητούμενα Μέσα Μήκη Ροής Εντός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) με την εφαρμογή της μεθόδου PCA

Για τις ως άνω τιμές υπολογίστηκαν τα μέσα μήκη ροής εκτός ελέγχου της διεργασίας, με τα αποτελέσματα να είναι τα ακόλουθα:

Για όλες τις μετατοπίσεις, τα διαγράμματα PCA ανταποκρίθηκαν συνολικά καλύτερα από όλα τα άλλα διαγράμματα ελέγχου για θετικές συσχετίσεις της χρονοσειράς, γεγονός που εξέπληξε κατά κάποιο τρόπο λόγω της συνηθισμένες υποβέλτιστης απόδοσης της μεθόδου σε κανονικά δεδομένα. Η υπερπαραμέτρος φαίνεται επίσης ότι είχε πολύ θετικό αντίκτυπο αφού τα μέσα μήκη ροής εκτός ελέγχου παρουσιάζονται αρκετά κοντινά για όλων των ειδών τις μετατοπίσεις. Ενδεικτικά, για μετατόπιση 0.5σ σε δύο χαρακτηριστικά της διεργασίας προκύπτουν ARL_1 ίσα με 17.83, 20.34 και 22.39 αντιστοίχως. Τα ARL_1 για μετατόπιση δύο τυπικών αποκλίσεων σε ένα χαρακτηριστικό και μιας σε ένα άλλο χαρακτηριστικό φαίνονται στον παρακάτω πίνακα.

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
4.87	5.59	6.29

Table 12: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) με μετατόπιση του μέσου κατά $(1 * 2\sigma, 1 * 1\sigma)$ με την εφαρμογή της μεθόδου PCA

Για μετατόπιση τριών τυπικών αποκλίσεων από τον μέσο δύο χρονοσειρών, τα μέσα μήκη ροής εκτός ελέγχου είναι τα ακόλουθα:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
1.8	2.06	2.39

Table 13: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) με μετατόπιση του μέσου κατά $(2 * 3\sigma)$ με την εφαρμογή της μεθόδου PCA

(b) Αρνητική Αυτοσυσχέτιση:

- T^2

Στον παρακάτω πίνακα παρουσιάζονται οι τιμές των παραμέτρων εμπιστοσύνης που εμφανίζουν τα ζητούμενα μέσα μήκη ροής εντός ελέγχου:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
0.9945	0.9969	0.9988

Table 14: Τιμές Παραμέτρων για τα Ζητούμενα Μέσα Μήκη Ροής Εντός Ελέγχου για Αρνητικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) για το T^2 διάγραμμα

Για τις ως άνω τιμές υπολογίστηκαν τα μέσα μήκη ροής εκτός ελέγχου της διεργασίας, με τα αποτελέσματα να είναι τα ακόλουθα:

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

Και για αρνητικές συσχετίσεις για μικρές μετατοπίσεις τα διαγράμματα T^2 φαίνονται να είναι τα λιγότερο αποδοτικά αλλά αρκετά κοντύτερα με τα πιο αυστηρά διαγράμματα ελέγχου. Ενδεικτικά, για μετατόπιση 0.5σ σε δύο χαρακτηριστικά της διεργασίας προκύπτουν ARL_1 ίσα με 11.31, 12.98 και 15.17 αντιστοίχως. Για μεσαίες μετατοπίσεις φαίνεται πως συμβαδίζουν σχεδόν σε απόδοση με τα άλλα διαγράμματα. Τα ARL_1 για μετατόπιση δύο τυπικών αποκλίσεων σε ένα χαρακτηριστικό και μιας σε ένα άλλο χαρακτηριστικό φαίνονται στον παρακάτω πίνακα.

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
3.11	4.0	4.78

Table 15: Μέσα Μήκη Ροής Εκτός Ελέγχου για Αρνητικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) με μετατόπιση του μέσου κατά $(1 * 2\sigma, 1 * 1\sigma)$ για το T^2 διάγραμμα

Για μετατόπιση τριών τυπικών αποκλίσεων από τον μέσο δύο χρονοσειρών τα διαγράμματα T^2 είχαν εξαιρετική απόκριση. Τα μέσα μήκη ροής εκτός ελέγχου είναι τα ακόλουθα:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
1.65	1.89	2.2

Table 16: Μέσα Μήκη Ροής Εκτός Ελέγχου για Αρνητικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) με μετατόπιση του μέσου κατά $(2 * 3\sigma)$ για το T^2 διάγραμμα

- MCUSUM

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
(5.5, 0.75)	(5.25, 1)	(4.5, 1.5)
(3.5, 1.5)	(4, 1.75)	(3.75, 2.25)
(2.25, 2)	(2.75, 2.25)	(3.25, 2.5)

Table 17: Τιμές Παραμέτρων για τα ζητούμενα Μέσα Μήκη Ροής Εντός Ελέγχου για Αρνητικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) για το MCUSUM διάγραμμα

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

Για τις ως άνω τιμές υπολογίστηκαν τα μέσα μήκη ροής εκτός ελέγχου της διεργασίας, με τα αποτελέσματα να είναι τα ακόλουθα:

Για μικρές μετατοπίσεις του μέσου της διεργασίας σε οποιονδήποτε αριθμό μεταβλητών τα καλύτερα ARL_1 λαμβάνονται και πάλι από τα μικρότερα h σε κάθε περίπτωση. Ενδεικτικά, για μετατόπιση 0.5σ σε δύο χαρακτηριστικά της διεργασίας προκύπτουν ARL_1 ίσα με 5.76, 6.99 και 7.95 αντιστοίχως. Για μεσαίες μετατοπίσεις φαίνεται οι ενδιάμεσες τιμές για τα ζευγάρια παραμέτρων να δίνουν πλέον ικανοποιητικά ARL_1 όπως φαίνεται στον παρακάτω πίνακα.

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
$(h,k) = (3.5, 1.5)$	$(h,k) = (5.5, 2)$	$(h,k) = (3.75, 2.25)$
2.42	3.22	4.48

Table 18: Μέσα Μήκη Ροής Εκτός Ελέγχου για Αρνητικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) με μετατόπιση του μέσου κατά $(1 * 2\sigma, 1 * 1\sigma)$ για το MCUSUM διάγραμμα

Για μεγαλύτερες μετατοπίσεις στο μέσο, δηλαδή όταν ένα ή περισσότερα στοιχεία της διαδικασίας έχουν καταστεί ακατάλληλα, τα μέσα μήκη ροής εκτός ελέγχου ανιχνεύουν άμεσα τις μετατοπίσεις. Ενδεικτικά, για μετατόπιση τριών τυπικών αποκλίσεων από τον μέσο δύο χρονοσειρών, τα μέσα μήκη ροής εκτός ελέγχου είναι τα ακόλουθα:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
$(h,k) = (2.25, 2)$	$(h,k) = (2.75, 2.25)$	$(h,k) = (3.25, 2.5)$
1.38	1.91	2.29

Table 19: Μέσα Μήκη Ροής Εκτός Ελέγχου για Αρνητικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) με μετατόπιση του μέσου κατά $(2 * 3\sigma)$ για το MCUSUM διάγραμμα

- MEWMA

Με παρουσία αρνητικής συσχέτισης, παρουσιάζεται εξαιρετικά βελτιωμένη απόδοση και στα διαγράμματα MEWMA. Οι παράμετροι που ικανοποιούν τα μέσα μήκη ροής εντός

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

ελέγχου που θέτονται ως στόχοι παρουσιάζονται στον παρακάτω πίνακα:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
0.2	0.55	0.725

Table 20: Τιμές Παραμέτρων για τα ζητούμενα Μέσα Μήκη Ροής Εντός Ελέγχου για Αρνητικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) για το MEWMA διάγραμμα

Σε αυτό το σημείο, φαίνονται κατά πολύ μικρότερες τιμές της παραμέτρου λ για την επίτευξη των ορίων μέσω των μηκών ροής εντός ελέγχου, καταδεικνύοντας ότι οι αρνητικώς συσχετισμένες χρονοσειρές εμφανίζουν μεγαλύτερη εξάρτηση από τις προηγούμενες παρατηρήσεις, παρότι οι θετικά συσχετισμένες. Μια λογική ερμηνεία για αυτό το φαινόμενο είναι ότι η μοντελοποίηση των αρνητικά συσχετισμένων χρονοσειρών απασχολείτε από τις παρελθοντικές τιμές καθότι αυτές παρέχουν γνήσια πληροφορία, ενώ στην περίπτωση των θετικά συσχετισμένων χρονοσειρών, οι προηγούμενες τιμές παρέχουν την ίδια περίπου πληροφορία και για αυτό δίνεται μεγαλύτερο βάρος στην παρούσα τιμή της χρονοσειράς. Για τις ως άνω τιμές ARL_0 υπολογίστηκαν τα μέσα μήκη ροής εκτός ελέγχου της διεργασίας, με τα αποτελέσματα να είναι τα ακόλουθα:

Για μικρές και μεσαίες μετατοπίσεις τα διαγράμματα MEWMA φαίνονται να είναι περισσότερο αποδοτικά από τα MCUSUM. Ενδεικτικά, για μετατόπιση 0.5σ σε δύο χαρακτηριστικά της διεργασίας προκύπτουν ARL_1 ίσα με 5.3, 5.8 και 6.35 αντιστοίχως. Για μετατόπιση δύο τυπικών αποκλίσεων σε ένα χαρακτηριστικό και μιας σε ένα άλλο χαρακτηριστικό, τα ARL_1 φαίνονται στον παρακάτω πίνακα.

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
2.17	2.86	4.62

Table 21: Μέσα Μήκη Ροής Εκτός Ελέγχου για Αρνητικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) με μετατόπιση του μέσου κατά $(1 * 2\sigma, 1 * 1\sigma)$ για το MEWMA διάγραμμα

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

Για μεγαλύτερες μετατοπίσεις στο μέσο, δηλαδή όταν ένα ή περισσότερα στοιχεία της διαδικασίας έχουν καταστεί ακατάλληλα, τα μέσα μήκη ροής εκτός ελέγχου ανιχνεύουν αρκετά γρήγορα τις μετατοπίσεις, ωστόσο αργότερα από τα MCUSUM διαγράμματα. Ενδεικτικά, για μετατόπιση τριών τυπικών αποκλίσεων από τον μέσο δύο χρονοσειρών, τα μέσα μήκη ροής εκτός ελέγχου είναι τα ακόλουθα:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
1.79	2.02	2.63

Table 22: Μέσα Μήκη Ροής Εκτός Ελέγχου για Αρνητικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) με μετατόπιση του μέσου κατά $(2 * 3\sigma)$ για το MEWMA διάγραμμα

- PCA Ο παρακάτω πίνακα παρουσιάζει τις τιμές της πολλαπλασιαστικής υπερπαραμέτρου που εμφανίζουν τα ζητούμενα μέσα μήκη ροής εντός ελέγχου. Παρατηρείται ότι και στα αρνητικά συσχετισμένα δεδομένα οι τιμές της υπερπαραμέτρου είναι αρκετά κοντινές με την περίπτωση των θετικώς συσχετισμένων δεδομένων, αλλά είναι λίγο μικρότερες.

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
1.59	1.86	2.06

Table 23: Τιμές Παραμέτρων για τα Ζητούμενα Μέσα Μήκη Ροής Εντός Ελέγχου για Αρνητικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) με την μέθοδο PCA

Για τις ως άνω τιμές υπολογίστηκαν τα μέσα μήκη ροής εκτός ελέγχου της διεργασίας, με τα αποτελέσματα να είναι τα ακόλουθα:

Οι τιμές των μέσων μηκών ροής εκτός ελέγχου είναι παρόμοιες με την περίπτωση των θετικών συσχετισμένων δεδομένων, αλλά εμφανώς μικρότερα. Ενδεικτικά, για μετατόπιση 0.5σ σε δύο χαρακτηριστικά της διεργασίας προκύπτουν ARL_1 ίσα με 14.83, 16.0 και 17.63 αντιστοίχως. Για μεσαίες και μεγάλες μετατοπίσεις φαίνεται πλησιάζει σε απόδοση τα άλλα διαγράμματα. Τα ARL_1 για μετατόπιση δύο τυπικών αποκλίσεων

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

σε ένα χαρακτηριστικό και μιας σε ένα άλλο χαρακτηριστικό φαίνονται στον παρακάτω πίνακα.

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
4.47	5.12	5.64

Table 24: Μέσα Μήκη Ροής Εκτός Ελέγχου για Αρνητικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) με μετατόπιση του μέσου κατά $(1 * 2\sigma, 1 * 1\sigma)$ με την μέθοδο PCA

Για μετατόπιση τριών τυπικών αποκλίσεων από τον μέσο δύο χρονοσειρών, τα μέσα μήκη ροής εκτός ελέγχου είναι τα ακόλουθα:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
1.78	2.03	2.29

Table 25: Μέσα Μήκη Ροής Εκτός Ελέγχου για Αρνητικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,0) με μετατόπιση του μέσου κατά $(2 * 3\sigma)$ με την μέθοδο PCA

2. VARMA(1,0,1): Για αυτού του είδους τις διαδικασίες τα διαγράμματα φαίνεται να είχαν πιο ικανοποιητική απόδοση από ότι στις VARMA(1,0,0) διαδικασίες τουλάχιστον στην περίπτωση που εξετάστηκε, δηλαδή για θετικά συσχετισμένα αυτοπαλινδρομητικό κομμάτι. Γενικότερα, παρατηρήθηκαν φυσιολογικότερες τιμές παραμέτρων, κοντά στις θεωρητικές για κανονικώς κατανεμημένα δεδομένα, για τα ζητούμενα μέσα μήκη ροής εντός ελέγχου.

- T^2

Στον παρακάτω πίνακα παρουσιάζονται οι τιμές των παραμέτρων εμπιστοσύνης που εμφανίζουν τα ζητούμενα μέσα μήκη ροής εντός ελέγχου:

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
0.9956	0.9974	0.9995

Table 26: Τιμές Παραμέτρων για τα ζητούμενα Μέσα Μήκη Ροής Εντός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,1) για το T^2 διάγραμμα

Για τις ως άνω τιμές υπολογίστηκαν τα μέσα μήκη ροής εκτός ελέγχου της διεργασίας, με τα αποτελέσματα να είναι τα ακόλουθα:

Για μετατόπιση 0.5σ σε δύο χαρακτηριστικά της διεργασίας προκύπτουν ARL_1 ίσα με 63.4, 68.78 και 80.5 αντιστοίχως. Για μεσαίες και μεγάλες μετατοπίσεις φαίνεται πως το διάγραμμα παρουσιάζει εξαιρετική απόδοση και κρίνεται πλέον κατάλληλο για την παρακολούθηση μιας τέτοιας διεργασίας. Τα ARL_1 για μετατόπιση δύο τυπικών αποκλίσεων σε ένα χαρακτηριστικό και μιας σε ένα άλλο χαρακτηριστικό φαίνονται στον παρακάτω πίνακα.

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
3.82	4.94	7.32

Table 27: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,1) με μετατόπιση του μέσου κατά $(1 * 2\sigma, 1 * 1\sigma)$ για το T^2 διάγραμμα

Για μετατόπιση τριών τυπικών αποκλίσεων από τον μέσο δύο χρονοσειρών τα διαγράμματα T^2 είχαν εξαιρετική απόκριση. Τα μέσα μήκη ροής εκτός ελέγχου είναι τα ακόλουθα:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
1.11	1.21	1.43

Table 28: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,1) με μετατόπιση του μέσου κατά $(2 * 3\sigma)$ για το T^2 διάγραμμα

• MCUSUM

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
(9.5, 0.75)	(8.5, 1)	(5.5, 1.5)
(7.5, 1)	(4, 1.75)	(3.5, 2)
(4, 1.5)	(3, 2.25)	(3.25, 2.5)

Table 29: Τιμές Παραμέτρων για τα ζητούμενα Μέσα Μήκη Ροής Εντός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,1) για το MCUSUM διάγραμμα

Για τις ως άνω τιμές υπολογίστηκαν τα μέσα μήκη ροής εκτός ελέγχου της διεργασίας, με τα αποτελέσματα να είναι τα ακόλουθα:

Για μικρές μετατοπίσεις του μέσου της διεργασίας σε οποιονδήποτε αριθμό μεταβλητών τα καλύτερα ARL_1 λαμβάνονται από τα μικρότερα k σε κάθε περίπτωση. Ενδεικτικά, για μετατόπιση 0.5σ σε δύο χαρακτηριστικά της διεργασίας προκύπτουν ARL_1 ίσα με 19.48, 23.26 και 34.68 αντιστοίχως. Για μεσαίες μετατοπίσεις φαίνεται οι ενδιάμεσες τιμές για τα ζευγάρια παραμέτρων να δίνουν πλέον ικανοποιητικά ARL_1 όπως φαίνεται στον παρακάτω πίνακα.

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
(h,k) = (7.5, 1)	(h,k) = (4, 1.75)	(h,k) = (3.5, 2)
4.64	4.08	5.52

Table 30: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,1) με μετατόπιση του μέσου κατά $(1 * 2\sigma, 1 * 1\sigma)$ για το MCUSUM διάγραμμα

Για μεγαλύτερες μετατοπίσεις στο μέσο, δηλαδή όταν ένα ή περισσότερα στοιχεία της διαδικασίας έχουν καταστεί ακατάλληλα, τα μέσα μήκη ροής εκτός ελέγχου ανιχνεύουν σχετικά ικανοποιητικά τις μετατοπίσεις. Παρατηρείτε ότι για μικρότερες μετατοπίσεις καλύτερη απόκριση αισθητά έχουν τα διαγράμματα με τις μικρότερες σταθερές k , αλλά για μεσαίες και μεγάλες

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

μετατοπίσεις, καθώς αυξάνεται το k προκειμένου να συγκριθούν τα ίδια ARL_0 , τα διαγράμματα με μικρότερο k και μεγαλύτερο h αργούν περισσότερο να εντοπίσουν μεγάλες μετατοπίσεις. Γίνεται αισθητή η λειτουργία και η σημαντικότητα της επιλογής ορθών παραμέτρων τους διαγράμματος αναλόγως με το εκάστοτε ειδικό πρόβλημα. Ενδεικτικά, για μετατόπιση τριών τυπικών αποκλίσεων από τον μέσο δύο χρονοσειρών, τα μέσα μήκη ροής εκτός ελέγχου είναι τα ακόλουθα:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
$(h,k) = (4, 1.5)$	$(h,k) = (3, 2.25)$	$(h,k) = (3.25, 2.5)$
2.16	1.16	1.62

Table 31: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,1) με μετατόπιση του μέσου κατά $(2 * 3\sigma)$ για το MCUSUM διάγραμμα

- MEWMA

Τα διαγράμματα MEWMA φαίνονται τα πλέον ταιριαστά για τα παρόντα δεδομένα του μοντέλου VARMA(1,0,1) για την ανίχνευση μικρών μετατοπίσεων. Οι τιμές των λ που ικανοποιούν τα ζητούμενα μέσα μήκη ροής εντός ελέγχου είναι οι ακόλουθες:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
0.42	0.11	0.04

Table 32: Τιμές Παραμέτρων για τα Ζητούμενα Μέσα Μήκη Ροής Εντός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,1) για το MEWMA διάγραμμα

Σε αυτό το σημείο, φαίνονται κατά πολύ μικρότερες τιμές της παραμέτρου λ για την επίτευξη των ορίων μέσων μηκών ροής εντός ελέγχου, καταδεικνύοντας ότι οι αρνητικώς συσχετισμένες χρονοσειρές εμφανίζουν μεγαλύτερη εξάρτηση από τις προηγούμενες παρατηρήσεις, παρότι οι θετικά συσχετισμένες. Μια λογική ερμηνεία για αυτό το φαινόμενο είναι ότι η μοντελοποίηση των αρνητικά συσχετισμένων χρονοσειρών απασχολείται από τις παρελθοντικές τιμές καθότι αυτές παρέχουν γνήσια πληροφορία, ενώ στην περίπτωση των θετικά

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

συσχετισμένων χρονοσειρών, οι προηγούμενες τιμές παρέχουν την ίδια περίπου πληροφορία και για αυτό δίνεται μεγαλύτερο βάρος στην παρούσα τιμή της χρονοσειράς.

Για τις ως άνω τιμές ARL_0 υπολογίστηκαν τα μέσα μήκη ροής εκτός ελέγχου της διεργασίας, με τα αποτελέσματα να είναι τα ακόλουθα:

Για μικρές και μεσαίες μετατοπίσεις τα διαγράμματα MEWMA φαίνονται να είναι περισσότερο αποδοτικά από τα MCUSUM. Ενδεικτικά, για μετατόπιση 0.5σ σε δύο χαρακτηριστικά της διεργασίας προκύπτουν ARL_1 ίσα με 18.45, 18.78 και 69.65 αντιστοίχως. Για μετατόπιση δύο τυπικών αποκλίσεων σε ένα χαρακτηριστικό και μιας σε ένα άλλο χαρακτηριστικό, τα ARL_1 φαίνονται στον παρακάτω πίνακα.

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
5.32	4.6	6.01

Table 33: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,1) με μετατόπιση του μέσου κατά $(1 * 2\sigma, 1 * 1\sigma)$ για το MEWMA διάγραμμα

Για μεγαλύτερες μετατοπίσεις στο μέσο, δηλαδή όταν ένα ή περισσότερα στοιχεία της διαδικασίας έχουν καταστεί ακατάλληλα, τα μέσα μήκη ροής εκτός ελέγχου ανιχνεύουν αρκετά γρήγορα τις μετατοπίσεις, ωστόσο αργότερα από τα MCUSUM διαγράμματα. Ενδεικτικά, για μετατόπιση τριών τυπικών αποκλίσεων από τον μέσο δύο χρονοσειρών, τα μέσα μήκη ροής εκτός ελέγχου είναι τα ακόλουθα:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
1.41	1.46	1.89

Table 34: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,1) με μετατόπιση του μέσου κατά $(2 * 3\sigma)$ για το MEWMA διάγραμμα

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

- PCA Ο παρακάτω πίνακα παρουσιάζει τις τιμές της πολλαπλασιαστικής υπερπαραμέτρου που εμφανίζουν τα ζητούμενα μέσα μήκη ροής εντός ελέγχου. Παρατηρείται ότι και στα αρνητικών συσχετισμένα δεδομένα οι τιμές της υπερπαραμέτρου είναι αρκετά κοντινές με την περίπτωση των θετικώς συσχετισμένων δεδομένων, αλλά είναι λίγο μικρότερες.

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
0.92	1.17	1.56

Table 35: Τιμές Παραμέτρων για τα Ζητούμενα Μέσα Μήκη Ροής Εντός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,1) με εφαρμογή της μεθόδου PCA

Παρατηρείται ότι σε αυτή την περίπτωση η τιμή της υπερπαραμέτρου λαμβάνει τιμές αρκετά κοντά στο ένα, που σημαίνει ότι η μέθοδος PCA εφαρμόζεται καλύτερα σε δεδομένα από το μοντέλο VARMA(1,0,1). Για τις ως άνω τιμές υπολογίστηκαν τα μέσα μήκη ροής εκτός ελέγχου της διεργασίας, με τα αποτελέσματα να είναι τα ακόλουθα:

Ενδεικτικά, για μετατόπιση 0.5σ σε δύο χαρακτηριστικά της διεργασίας προκύπτουν ARL_1 ίσα με 70.5, 124.46 και 165.24 αντιστοίχως. Για μεσαίες και μεγάλες μετατοπίσεις φαίνεται πλησιάζει σε απόδοση τα άλλα διαγράμματα. Τα ARL_1 για μετατόπιση δύο τυπικών αποκλίσεων σε ένα χαρακτηριστικό και μιας σε ένα άλλο χαρακτηριστικό φαίνονται στον παρακάτω πίνακα.

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
7.88	10.6	14.7

Table 36: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,1) με μετατόπιση του μέσου κατά $(1 * 2\sigma, 1 * 1\sigma)$ με εφαρμογή της μεθόδου PCA

Για μετατόπιση τριών τυπικών αποκλίσεων από τον μέσο δύο χρονοσειρών, τα μέσα μήκη ροής εκτός ελέγχου είναι τα ακόλουθα:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
1.92	2.22	4.72

Table 37: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(1,0,1) με μετατόπιση του μέσου κατά $(2 * 3\sigma)$ με εφαρμογή της μεθόδου PCA

3. VARMA(2,0,1): Για αυτού του είδους τις διαδικασίες τα διαγράμματα φαίνεται να είχαν λιγότερο ικανοποιητική απόδοση από ότι στις VARMA(1,0,0) και τις VARMA(1,0,1) διαδικασίες τουλάχιστον στην περίπτωση που εξετάστηκε, δηλαδή για θετικά συσχετισμένα αυτοπαλινδρομητικό κομμάτι. Διαφαίνεται ότι γενικά όσο αυξάνεται το αυτοπαλινδρομητικό κομμάτι του VARMA τόσο δυσκρινέστερη είναι η καλή απόδοση του για το μέσο μήκος ροής εντός ελέγχου. Γενικότερα, παρατηρήθηκαν εξαιρετικά μεγάλες τιμές παραμέτρων, μακριά από τις θεωρητικές για κανονικώς κατανομημένα δεδομένα, για τα ζητούμενα μέσα μήκη ροής εντός ελέγχου.

- T^2

Στον παρακάτω πίνακα παρουσιάζονται οι τιμές των παραμέτρων εμπιστοσύνης που εμφανίζουν τα ζητούμενα μέσα μήκη ροής εντός ελέγχου:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
0.9947	0.9969	0.9989

Table 38: Τιμές Παραμέτρων για τα Ζητούμενα Μέσα Μήκη Ροής Εντός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(2,0,1) για τα T^2 διαγράμματα

Για τις ως άνω τιμές υπολογίστηκαν τα μέσα μήκη ροής εκτός ελέγχου της διεργασίας, με τα αποτελέσματα να είναι τα ακόλουθα:

Για μετατόπιση 0.5σ σε δύο χαρακτηριστικά της διεργασίας προκύπτουν ARL_1 ίσα με 71.59, 85.79 και 90.34 αντιστοίχως. Για μεσαίες και μεγάλες μετατοπίσεις φαίνεται πως το διάγραμμα παρουσιάζει εξαιρετική απόδοση και κρίνεται πλέον κατάλληλο για την παρακολούθηση μιας

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

τέτοιας διεργασίας. Τα ARL_1 για μετατόπιση δύο τυπικών αποκλίσεων σε ένα χαρακτηριστικό και μιας σε ένα άλλο χαρακτηριστικό φαίνονται στον παρακάτω πίνακα.

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
7.83	9.02	10.59

Table 39: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(2,0,1) με μετατόπιση του μέσου κατά $(1 * 2\sigma, 1 * 1\sigma)$ για τα T^2 διαγράμματα

Για μετατόπιση τριών τυπικών αποκλίσεων από τον μέσο δύο χρονοσειρών τα διαγράμματα T^2 είχαν και πάλι εξαιρετική απόκριση. Τα μέσα μήκη ροής εκτός ελέγχου είναι τα ακόλουθα:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
1.51	1.79	1.93

Table 40: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(2,0,1) με μετατόπιση του μέσου κατά $(2 * 3\sigma)$ για τα T^2 διαγράμματα

- MCUSUM

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
(6.5, 2.25)	(7.5, 2.25)	(10, 2.25)
(7, 2.25)	(8, 2.25)	(5.5, 2.5)
(3.5, 2.5)	(4, 2.5)	(6.0, 2.5)

Table 41: Τιμές Παραμέτρων για τα ζητούμενα Μέσα Μήκη Ροής Εντός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(2,0,1) για τα MCUSUM διαγράμματα

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

Για τις ως άνω τιμές υπολογίστηκαν τα μέσα μήκη ροής εκτός ελέγχου της διεργασίας, με τα αποτελέσματα να είναι τα ακόλουθα:

Για μικρές μετατοπίσεις του μέσου της διεργασίας σε οποιονδήποτε αριθμό μεταβλητών τα καλύτερα ARL_1 λαμβάνονται από τα μικρότερα k σε κάθε περίπτωση. Ενδεικτικά, για μετατόπιση 0.5σ σε δύο χαρακτηριστικά της διεργασίας προκύπτουν ARL_1 ίσα με 65.57, 76.73 και 78.26 αντιστοίχως. Για μεσαίες μετατοπίσεις φαίνεται οι ενδιάμεσες τιμές για τα ζευγάρια παραμέτρων να δίνουν πλέον ικανοποιητικά ARL_1 όπως φαίνεται στον παρακάτω πίνακα.

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
$(h,k) = (7, 2.25)$	$(h,k) = (8, 2.25)$	$(h,k) = (5.5, 2.5)$
7.92	8.91	12.42

Table 42: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(2,0,1) με μετατόπιση του μέσου κατά $(1 * 2\sigma, 1 * 1\sigma)$ για τα MCUSUM διαγράμματα

Για μεγαλύτερες μετατοπίσεις στο μέσο, δηλαδή όταν ένα ή περισσότερα στοιχεία της διαδικασίας έχουν καταστεί ακατάλληλα, τα μέσα μήκη ροής εκτός ελέγχου ανιχνεύουν σχετικά ικανοποιητικά τις μετατοπίσεις. Ενδεικτικά, για μετατόπιση τριών τυπικών αποκλίσεων από τον μέσο δύο χρονοσειρών, τα μέσα μήκη ροής εκτός ελέγχου είναι τα ακόλουθα:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
$(h,k) = (3.5, 2.5)$	$(h,k) = (4, 2.5)$	$(h,k) = (6.5, 2.5)$
1.92	2.09	2.6

Table 43: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(2,0,1) με μετατόπιση του μέσου κατά $(2 * 3\sigma)$ για τα MCUSUM διαγράμματα

- MEWMA

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

Τα διαγράμματα MEWMA φαίνονται επίσης τα πλέον ταιριαστά για τα παρόντα δεδομένα του μοντέλου VARMA(2,0,1) για την ανίχνευση μικρών μετατοπίσεων. Οι τιμές των λ που ικανοποιούν τα ζητούμενα μέσα μήκη ροής εντός ελέγχου είναι οι ακόλουθες:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
0.86	0.78	0.5

Table 44: Τιμές Παραμέτρων για τα Ζητούμενα Μέσα Μήκη Ροής Εντός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(2,0,1) για τα MEWMA διαγράμματα

Σε αυτό το σημείο, φαίνονται και πάλι όπως στην περίπτωση του μοντέλου VARMA(1,0,0) μεγάλες τιμές της παραμέτρου λ για την επίτευξη των ορίων μέσων μηκών ροής εντός ελέγχου.

Για τις ως άνω τιμές ARL_0 υπολογίστηκαν τα μέσα μήκη ροής εκτός ελέγχου της διεργασίας, με τα αποτελέσματα να είναι τα ακόλουθα:

Για μετατόπιση 0.5σ σε δύο χαρακτηριστικά της διεργασίας προκύπτουν ARL_1 ίσα με 24.2, 28.78 και 69.65 αντιστοίχως. Για μετατόπιση δύο τυπικών αποκλίσεων σε ένα χαρακτηριστικό και μιας σε ένα άλλο χαρακτηριστικό, τα ARL_1 φαίνονται στον παρακάτω πίνακα.

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
15.32	14.43	16.03

Table 45: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(2,0,1) με μετατόπιση του μέσου κατά ($1 * 2\sigma$, $1 * 1\sigma$) για τα MEWMA διαγράμματα

Για μεγαλύτερες μετατοπίσεις στο μέσο, δηλαδή όταν ένα ή περισσότερα στοιχεία της διαδικασίας έχουν καταστεί ακατάλληλα, τα μέσα μήκη ροής εκτός ελέγχου ανιχνεύουν αρκετά γρήγορα τις μετατοπίσεις, ωστόσο αργότερα από τα MCUSUM διαγράμματα. Ενδεικτικά,

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

για μετατόπιση τριών τυπικών αποκλίσεων από τον μέσο δύο χρονοσειρών, τα μέσα μήκη ροής εκτός ελέγχου είναι τα ακόλουθα:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
3.72	3.66	3.67

Table 46: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(2,0,1) με μετατόπιση του μέσου κατά $(2 * 3\sigma)$ για τα MEWMA διαγράμματα

- PCA

Ο παρακάτω πίνακα παρουσιάζει τις τιμές της πολλαπλασιαστικής υπερπαραμέτρου που εμφανίζουν τα ζητούμενα μέσα μήκη ροής εντός ελέγχου.

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
2.17	2.66	3.46

Table 47: Τιμές Παραμέτρων για τα Ζητούμενα Μέσα Μήκη Ροής Εντός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(2,0,1) με εφαρμογή της μεθόδου PCA

Παρατηρείται ότι σε αυτή την περίπτωση η τιμή της υπερπαραμέτρου λαμβάνει τιμές αρκετά μεγαλύτερες του ένα, που σημαίνει ότι η μέθοδος PCA εφαρμόζεται με υποβέλτιστα αποτελέσματα σε δεδομένα από το μοντέλο VARMA(2,0,1). Για τις ως άνω τιμές υπολογίστηκαν τα μέσα μήκη ροής εκτός ελέγχου της διεργασίας, με τα αποτελέσματα να είναι τα ακόλουθα:

Ενδεικτικά, για μετατόπιση 0.5σ σε δύο χαρακτηριστικά της διεργασίας προκύπτουν ARL_1 ίσα με 75.58, 121.0 και 169.76 αντιστοίχως. Για μεσαίες και μεγάλες μετατοπίσεις φαίνεται πλησιάζει σε απόδοση τα άλλα διαγράμματα. Τα ARL_1 για μετατόπιση δύο τυπικών αποκλίσεων σε ένα χαρακτηριστικό και μιας σε ένα άλλο χαρακτηριστικό φαίνονται στον παρακάτω πίνακα.

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
7.22	9.64	25.67

Table 48: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(2,0,1) με μετατόπιση του μέσου κατά ($1 * 2\sigma, 1 * 1\sigma$) με εφαρμογή της μεθόδου PCA

Για μετατόπιση τριών τυπικών αποκλίσεων από τον μέσο δύο χρονοσειρών, τα μέσα μήκη ροής εκτός ελέγχου είναι τα ακόλουθα:

$ARL_0 \approx 300$	$ARL_0 \approx 550$	$ARL_0 \approx 1000$
2.52	3.44	9.08

Table 49: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(2,0,1) με μετατόπιση του μέσου κατά ($2 * 3\sigma$) με εφαρμογή της μεθόδου PCA

4. VARMA(x, 0, z)(X, 0, Z): Για αυτού του είδους τις χαστικές διαδικασίες τα διαγράμματα φαίνεται να είχαν παραδόξως ικανοποιητική απόδοση. Παρουσιάστηκαν δυσκολίες σε πολλά διαγράμματα για την επίτευξη μέσου μήκους ροής εντός ελέγχου για τιμές κοντά στο 1000, ενώ τα ARL_0 να είναι αρκετά μικρά ώστε να έχει νόημα η εφαρμογή διαγράμματος πάνω στα δεδομένα. Για τον λόγο αυτόν δοκιμάζονται τιμές παραμέτρων που δίνουν μέσο μήκος ροής εντός ελέγχου μέχρι την τιμή 550. Βέβαια, σημειώνεται ότι πιθανώς το κομμάτι της εποχικότητας ως ένα βαθμό αντιτασσόταν στο έντονο αυτοπαλινδρομητικό στοιχείο και επομένως οι τιμές των μέσων μηκών ροής εκτός ελέγχου κυμαίνονταν σε σχετικά φυσιολογικά πλαίσια.

- T^2

Στον παρακάτω πίνακα παρουσιάζονται οι τιμές των παραμέτρων εμπιστοσύνης που εμφανίζουν τα ζητούμενα μέσα μήκη ροής εντός ελέγχου:

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

$ARL_0 \approx 300$	$ARL_0 \approx 550$
0.9952	0.9971

Table 50: Τιμές Παραμέτρων για τα ζητούμενα Μέσα Μήκη Ροής Εντός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(x, 0, z)(X, 0, Z) για τα T^2 διαγράμματα

Για τις ως άνω τιμές υπολογίστηκαν τα μέσα μήκη ροής εκτός ελέγχου της διεργασίας, με τα αποτελέσματα να είναι τα ακόλουθα:

Για μετατόπιση 0.5σ σε δύο χαρακτηριστικά της διεργασίας προκύπτουν ARL_1 ίσα με 95.49 και 111.87 αντιστοίχως. Τα ARL_1 για μετατόπιση δύο τυπικών αποκλίσεων σε ένα χαρακτηριστικό και μιας σε ένα άλλο χαρακτηριστικό φαίνονται στον παρακάτω πίνακα.

$ARL_0 \approx 300$	$ARL_0 \approx 550$
25.12	30.83

Table 51: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(x, 0, z)(X, 0, Z) με μετατόπιση του μέσου κατά $(1 * 2\sigma, 1 * 1\sigma)$ για τα T^2 διαγράμματα

Για μετατόπιση τριών τυπικών αποκλίσεων από τον μέσο δύο χρονοσειρών τα διαγράμματα T^2 είχαν βελτιωμένη απόκριση. Τα μέσα μήκη ροής εκτός ελέγχου είναι τα ακόλουθα:

$ARL_0 \approx 300$	$ARL_0 \approx 550$
2.45	3.11

Table 52: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(x, 0, z)(X, 0, Z) με μετατόπιση του μέσου κατά $(2 * 3\sigma)$ για τα T^2 διαγράμματα

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

- MCUSUM

Τα μέσα μήκη ροής εκτός ελέγχου, καθώς και τα μέσα ροής εντός ελέγχου μαζί με τις παραμέτρους που τα ικανοποιούν, έχουν ήδη παρουσιαστεί στον πίνακα 1. Τα MCUSUM διαγράμματα εν παρόντι φάνηκαν ικανότερα για την αξιολόγηση χαοτικών διεργασιών με εξαιρετική απόκριση ακόμα και στην περίπτωση των πολλών μεταβλητών. Αναλυτικότερα συμπεράσματα θα δωθούν με το πέρας της παράθεσης των αποτελεσμάτων για τα προσομοιωμένα δεδομένα.

- MEWMA

Τα διαγράμματα MEWMA ανταπεξήλθαν βέλτιστα για τις μεσαίες μετατοπίσεις του μέσου, με πάρα πολύ μικρές επιλογές της παραμέτρου εμφάνισαν γενικότερα περίεργη και απρόβλεπτη συμπεριφορά. Οι τιμές των λ που ικανοποιούν τα ζητούμενα μέσα μήκη ροής εντός ελέγχου είναι οι ακόλουθες:

$ARL_0 \approx 300$	$ARL_0 \approx 550$
0.024	0.021

Table 53: Τιμές Παραμέτρων για τα Ζητούμενα Μέσα Μήκη Ροής Εντός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(x, 0, z)(X, 0, Z) για τα MEWMA διαγράμματα

Για τις ως άνω τιμές ARL_0 υπολογίστηκαν τα μέσα μήκη ροής εκτός ελέγχου της διεργασίας, με τα αποτελέσματα να είναι τα ακόλουθα:

Για μετατόπιση 0.5σ σε δύο χαρακτηριστικά της διεργασίας προκύπτουν ARL_1 ίσα με 142.35 και 165.2 αντιστοίχως. Για μετατόπιση δύο τυπικών αποκλίσεων σε ένα χαρακτηριστικό και μιας σε ένα άλλο χαρακτηριστικό, τα ARL_1 φαίνονται στον παρακάτω πίνακα.

$ARL_0 \approx 300$	$ARL_0 \approx 550$
9.41	8.02

Table 54: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(x, 0, z)(X, 0, Z) με μετατόπιση του μέσου κατά $(1 * 2\sigma, 1 * 1\sigma)$ για τα MEWMA διαγράμματα

Για μεγαλύτερες μετατοπίσεις στο μέσο, δηλαδή όταν ένα ή περισσότερα στοιχεία της διαδικασίας έχουν καταστεί ακατάλληλα, τα μέσα μήκη ροής εκτός ελέγχου ανιχνεύουν αρκετά γρήγορα τις μετατοπίσεις, ωστόσο αργότερα από τα MCUSUM διαγράμματα. Ενδεικτικά, για μετατόπιση τριών τυπικών αποκλίσεων από τον μέσο δύο χρονοσειρών, τα μέσα μήκη ροής εκτός ελέγχου είναι τα ακόλουθα:

$ARL_0 \approx 300$	$ARL_0 \approx 550$
8.13	7.19

Table 55: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(x, 0, z)(X, 0, Z) με μετατόπιση του μέσου κατά $(2 * 3\sigma)$ για τα MEWMA διαγράμματα

- **PCA** Η μέθοδος PCA έδωσε αρκετά καλά αποτελέσματα για μεγαλύτερες μετατοπίσεις αν και για την επίτευξη μέσου μήκους ροής εντός ελέγχου ίσου με 550, η υπερπαραμέτρος έλαβε την εξαιρετικά μεγάλη τιμή του 8.12, και για τον λόγο αυτό, η ανίχνευση μικρών μετατοπίσεων με αυτή την μέθοδο δεν λειτουργεί καθόλου αξιολογικά. Ο παρακάτω πίνακα παρουσιάζει τις τιμές της πολλαπλασιαστικής υπερπαραμέτρου που εμφανίζουν τα ζητούμενα μέσα μήκη ροής εντός ελέγχου.

$ARL_0 \approx 300$	$ARL_0 \approx 550$
5.77	8.12

Table 56: Τιμές Παραμέτρων για τα Ζητούμενα Μέσα Μήκη Ροής Εντός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(x, 0, z)(X, 0, Z) με εφαρμογή της μεθόδου PCA

Παρατηρείται ότι σε αυτή την περίπτωση η τιμή της υπερπαραμέτρου λαμβάνει τιμές αρκετά μεγαλύτερες του ένα, που σημαίνει ότι η μέθοδος PCA εφαρμόζεται με υποβέλτιστα αποτελέσματα σε δεδομένα από το μοντέλο VARMA(2,0,1). Για τις ως άνω τιμές υπολογίστηκαν τα μέσα μήκη ροής εκτός ελέγχου της διεργασίας, με τα αποτελέσματα να

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

είναι τα ακόλουθα:

Ενδεικτικά, για μετατόπιση 0.5σ σε δύο χαρακτηριστικά της διεργασίας προκύπτουν ARL_1 ίσα με 199 και 264 αντιστοίχως. Για μεσαίες και μεγάλες μετατοπίσεις ωστόσο, πάλι φαίνεται να πλησιάζει σε απόδοση τα άλλα διαγράμματα, δίχως να τα προσπερνάει. Τα ARL_1 για μετατόπιση δύο τυπικών αποκλίσεων σε ένα χαρακτηριστικό και μιας σε ένα άλλο χαρακτηριστικό φαίνονται στον παρακάτω πίνακα.

$ARL_0 \approx 300$	$ARL_0 \approx 550$
15.02	18.38

Table 57: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(x, 0, z)(X, 0, Z) με μετατόπιση του μέσου κατά $(1 * 2\sigma, 1 * 1\sigma)$ με εφαρμογή της μεθόδου PCA

Για μετατόπιση τριών τυπικών αποκλίσεων, από τον μέσο δύο χρονοσειρών, τα μέσα μήκη ροής εκτός ελέγχου είναι τα ακόλουθα:

$ARL_0 \approx 300$	$ARL_0 \approx 550$
2.73	4.25

Table 58: Μέσα Μήκη Ροής Εκτός Ελέγχου για Θετικά Αυτοσυσχετισμένη Διαδικασία VARMA(x, 0, z)(X, 0, Z) με μετατόπιση του μέσου κατά $(2 * 3\sigma)$ με εφαρμογή της μεθόδου PCA

Σχολιασμοί και αξιολόγηση των διαγραμμάτων για τα προσομοιωμένα δεδομένα

Αρχικά, αξίζει να σημειωθεί ότι τα δεδομένα, στο πλήθος των περιπτώσεων, δεν ακολουθούν την κανονική κατανομή και επομένως αυτομάτως κρίνουν ακατάλληλα τα συνήθη διαγράμματα ελέγχου που αναγκάζουν την ύπαρξη κανονικότητας στα δεδομένα. Εκτός αυτού, εμφανίζουν έντονη αυτοσυσχέτιση, γεγονός που επίσης δρα σαν τροχοπέδη για την εφαρμογή των συνήθων διαγραμμάτων. Η εποχικότητα

φαίνεται να έχει αμελητέα επίδραση σε όλα τα διαγράμματα εκτός του MEWMA που προβαίνει σε υπερ-εξομάλυνση. Στη μονομεταβλητή περίπτωση έχουν αναπτυχθεί πλήθη τροποποιημένων διαγραμμάτων, τα οποία σε περιπτώσεις αρνητικής συσχέτισης εμφανίζουν εξαιρετικά αποτελέσματα, αλλά δυσκολεύονται επίσης στην αποδοτική αξιολόγηση διεργασιών με υψηλή θετική αυτοσυσχέτιση. Κάτι παρόμοιο παρατηρήθηκε και στα πολυμεταβλητά απλά διαγράμματα. Αναλυτικά, οι σταθερές k και h στα MCUSUM διαγράμματα ήταν ιδιαίτερος υψηλές προκειμένου να επιτευχθούν αξιοπρεπή μέσα μήκη ροής για εντός ελέγχου διαδικασίες, ειδικά στα μοντέλα VARMA(1, 0, 0) και VARMA(1, 0, 1) όπου ο υπερκερασμός της θετικής αυτοσυσχέτισης ανάγκασε στην επιλογή πολύ μεγάλων τιμών παραμέτρων. Το κύριο μειονέκτημα αυτής της επιλογής είναι ότι πλέον τα MCUSUM διαγράμματα που κρίνονται πλέον κατάλληλα για την εύρεση μικρών μετατοπίσεων δεν είχαν την ζητούμενη απόδοση. Παρατηρήθηκε ότι για μικρότερες τιμές της παραμέτρου h και για μεγάλες αποκλίσεις, ήταν αρκετά γρήγορα συνολικά στην εύρεση της μετατόπισης, με μέτρια απόδοση στις μεσαίες αποκλίσεις και ασθενή στις μικρές. Ωστόσο, η απόδοση τους κατά την εφαρμογή του χασοτικού μοντέλου ήταν η πλέον επωφελής, αφού έδωσαν τις πιο ικανοποιητικές τιμές των μέσων μηκών ροής. Τα MEWMA διαγράμματα εμφάνισαν αλλόκοτη συμπεριφορά. Για τα μοντέλα με έντονο αυτοπαλινδρομητικό χαρακτήρα και θετική συσχέτιση επιλέχθηκαν εξαιρετικά μεγάλες τιμές της παραμέτρου λ , ενώ στην περίπτωση του χασοτικού μοντέλου και της διαδικασίας VARMA(1, 0, 1) η επιλογή της παραμέτρου ήταν αρκετά μικρή. Αυτό σημαίνει ότι στις δύο τελευταίες περιπτώσεις το διάγραμμα εξομαλύνει σε μεγάλο βαθμό τα δεδομένα, καθιστώντας την εύρεση μετατοπίσεων μεγάλης κλίμακας εξαιρετικά δύσκολη. Φάνηκε σε περιπτώσεις μεγάλων μετατοπίσεων να φτάνει σε ένα plateau για τα μέσα μήκη ροής εκτός ελέγχου και να μην δύναται να βελτιώσει περαιτέρω την ανίχνευση της μετατόπισης. Ωστόσο, στις μικρές μετατοπίσεις, ίσως είναι κατάλληλος ο χαρακτηρισμός του, ως το πιο αποδοτικό διάγραμμα. Τα T^2 διαγράμματα φάνηκαν να είναι τα πιο σταθερά σε απόδοση για όλων των ειδών τα μοντέλα, με εξαιρετική κάθε φορά απόκριση στις μεγάλες αποκλίσεις, όπου σε όλες τις περιπτώσεις πλησίαζαν την μονάδα, δηλαδή την άμεση επισήμανση. Αυτό το χαρακτηριστικό τους τα καθιστά ασφαλή και αξιόπιστη επιλογή σε όλες τις διαδικασίες στις οποίες στις οποίες ζητείται η ανίχνευση μεγάλων αποκλίσεων από το μέσο. Τα PCA διαγράμματα, με την εκλογή της υπερπαραμέτρου ήταν πολύ κοντά σε απόδοση με τα υπόλοιπα διαγράμματα για τα δύο απλούστερα μοντέλα, αλλά μόλις αυξήθηκε η περιπλοκότητα των χρονοσειρών κρίθηκαν ακατάλληλα για τον εντοπισμό μικρών και μεσαίων αποκλίσεων. Συνολικά, το συμπέρασμα που προκύπτει είναι το ακόλουθο. Για έντονα αυτοσυσχετισμένες διεργασίες, κατά τις οποίες η συνθήκη κανονικότητας παραβιάζεται τακτικά, η εκλογή των MEWMA για την ανίχνευση μικρών μετατοπίσεων, των MCUSUM ή PCA για την ανίχνευση μεσαίων μετατοπίσεων και των T^2 για την ανίχνευση μεγάλων μετατοπίσεων φαίνεται να είναι η κατά κανόνα πιο αποδοτική και πιο αξιόπιστη.

Σύγκριση με τα διαγράμματα υπολοίπων

Προκειμένου να συγκριθούν οι μέθοδοι που προσαρμόζουν στα υπόλοιπα μοντέλων κάποιο διάγραμμα,

5.4 Αποτελέσματα και Σχολιασμός για Προσομοιωμένες Τιμές

διαλέχθηκε η παρακάτω προσέγγιση. Για κάθε προσομοιωμένη χρονοσειρών ευρέθησαν τα υπόλοιπα αυτής με το εκάστοτε μοντέλα και εφαρμόστηκαν τα διαγράμματα MEWMA και MCUSUM (η επιλογή των δύο αυτών μοντέλων κρίνεται κατάλληλη, καθότι τα υπόλοιπα δεν παραβιάζουν, παρά σπανίως, την συνθήκη κανονικότητας, και εμφανίζουν αυτοσυσχέτιση σε πολύ μικρότερο βαθμό). Έπειτα, λήφθηκε η διαφορά της απόκρισης για κάθε προσομοιωμένη χρονοσειρά για τα εντός και τα εκτός μήκη ροής. Βάσει των διαφορών αυτών, κατασκευάστηκαν διαγράμματα που μελετούσαν την απόκριση σε κάθε περίπτωση των διαγραμμάτων υπολοίπων. Για τα δύο απλούστερα μοντέλα, το MEWMA διάγραμμα έδωσε τις καλύτερες αποκρίσεις, ενώ για τα δύο πιο σύνθετα το MCUSUM. Αυτή η διαφορά αιτιολογείται μερικώς από την ύπαρξη αυτοσυσχέτισης ως ένα βαθμό, αρκετά χαμηλότερο βέβαια, στις χρονοσειρές υπολοίπων. Για την υλοποίηση των διαγραμμάτων τα δεδομένα χωρίστηκαν σε τρεις κατηγορίες βάσει του βαθμού μετατόπισης που συμβαίνει στην διεργασία. Την πρώτη κατηγορία, σημάνουν οι μικρές μετατοπίσεις, δηλαδή αυτές που είναι μικρότερες της μιας τυπικής απόκλισης σε κάθε στοιχείο της χρονοσειράς. Τη δεύτερη κατηγορία σημάνουν οι μεσαίες παρατηρήσεις, αυτές που είναι έως δύο μονάδες τυπικών αποκλίσεων από την μέσα τιμή της κάθε χρονοσειράς πλην τις περιπτώσεις που καλύπτει η πρώτη κατηγορία. Την τρίτη και τελευταία κατηγορία κοσμούν οι μεγάλες μετατοπίσεις, δηλαδή αυτές που εμφανίζουν μέχρι τρεις μονάδες τυπικής απόκλισης μετατοπίσεις στον μέσο και δεν καλύπτονται από τις δύο πρώτες κατηγορίες. Βάσει αυτών προκύπτουν τα παρακάτω διαγράμματα που συγκρίνουν τις κατά μέσο όρο αποκρίσεις σε κάθε περίπτωση. Με την τιμή 100 εννοείται η καλύτερη απόδοση που τα συνήθη διαγράμματα ελέγχου είχαν σε κάθε περίπτωση, με τις άλλες παράλληλες μπάρες να εκφράζουν υψομετρικά την ποσοστιαία βελτίωση ή επιδείνωση της συμπεριφοράς των διαγραμμάτων υπολοίπων για κάθε μοντέλο. Γενικά αναμένεται το μοντέλο *ETS* να εμφανίσει τα χειρότερα αποτελέσματα αφού οι τιμές που έχουν προσομοιωθεί έχουν παραχθεί από την οικογένεια μοντέλων VARMA.

Σε κάθε σελίδα που ακολουθεί παρουσιάζονται τα διαγράμματα για τις διαδικασίες VARMA(1, 0, 0) για θετική αυτοσυσχέτιση, VARMA(1, 0, 1) και VARMA(2, 0, 1):

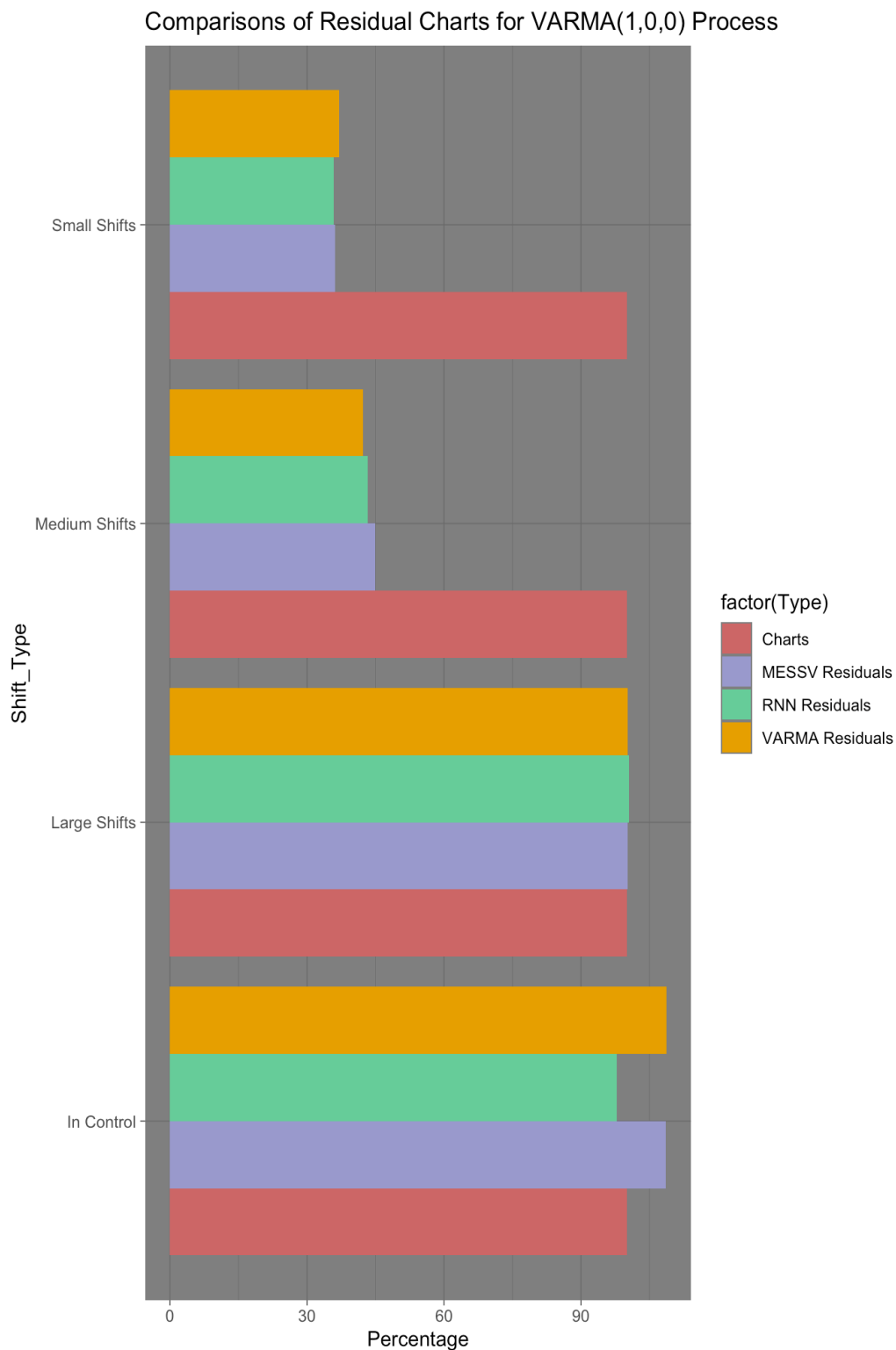


Figure 17: Σύγκριση απόδοσης Κλασικών Διαγραμμάτων με Διαγράμματα Υπολοίπων σε VARMA(1,0,0) διαδικασία

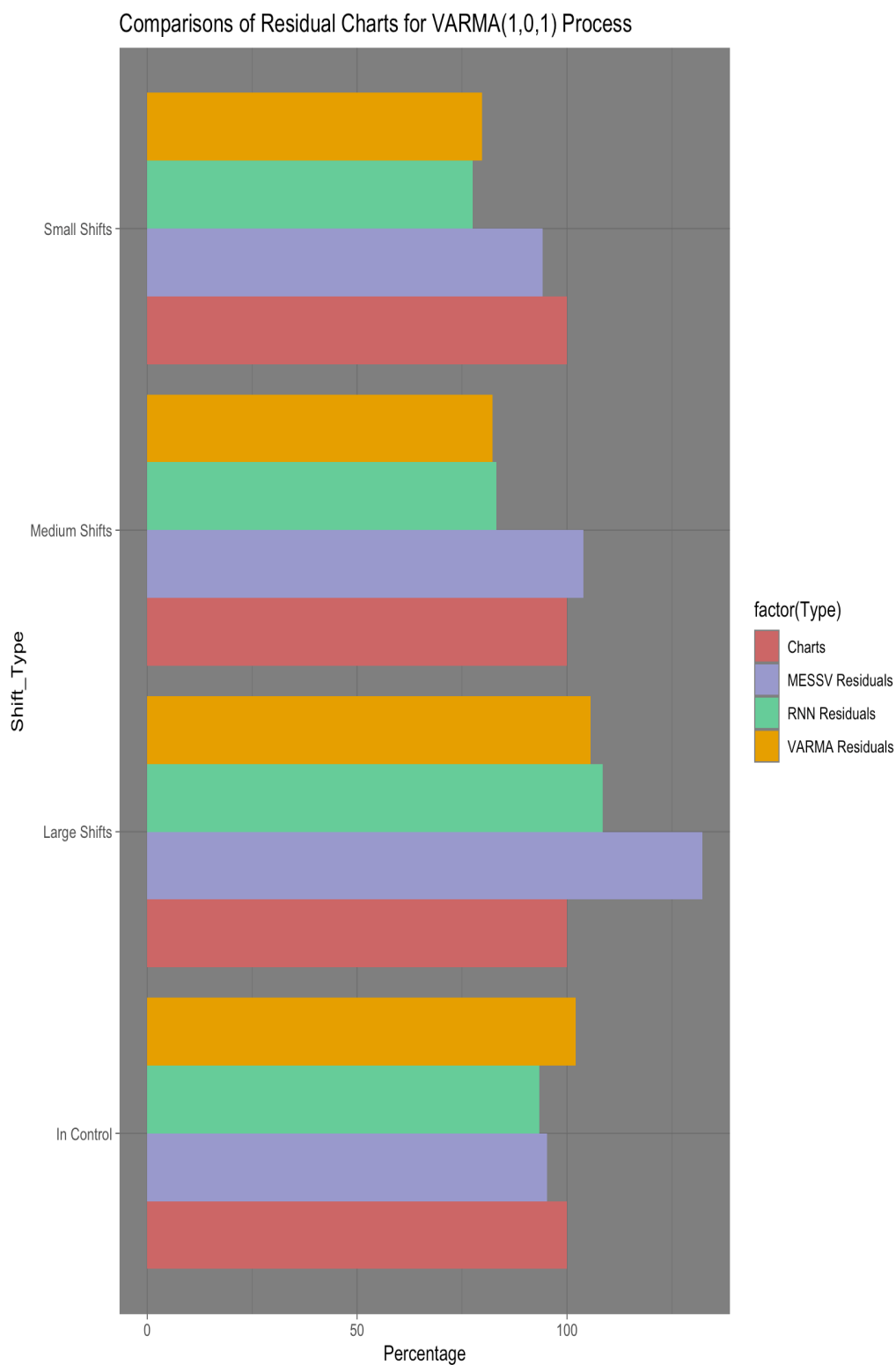


Figure 18: Σύγκριση απόδοσης Κλασικών Διαγραμμάτων με Διαγράμματα Υπολοίπων σε VARMA(1,0,1) διαδικασία

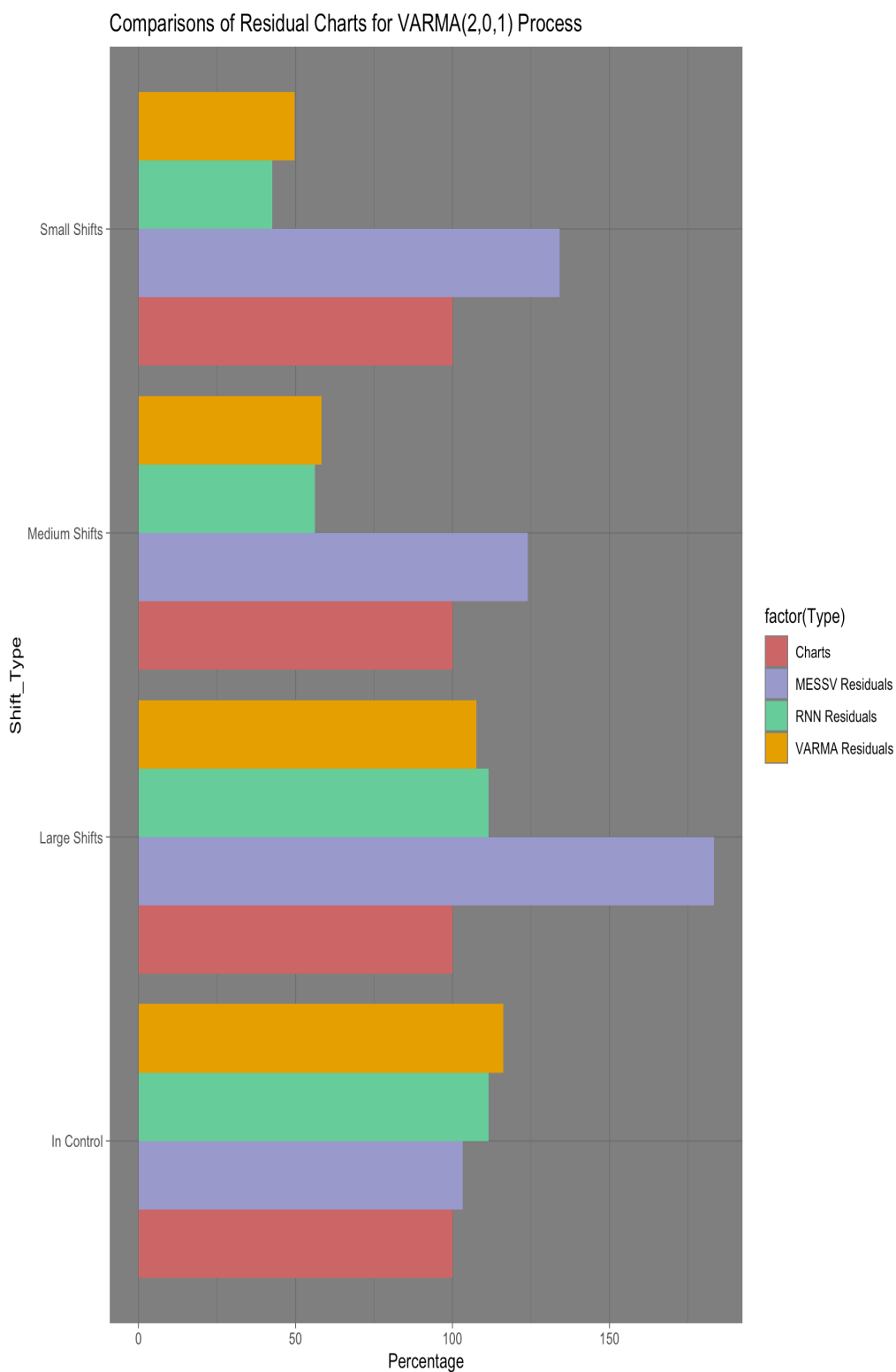


Figure 19: Σύγκριση απόδοσης Κλασικών Διαγραμμάτων με Διαγράμματα Υπολοίπων σε VARMA(2,0,1) διαδικασία

Σχόλια για τα Διαγράμματα Υπολοίπων

Αρχικά, αξίζει να σημειωθεί ότι στην περίπτωση των διεργασιών VARMA(1, 0, 0) με αρνητική συσχέτιση τα αποτελέσματα χρήσης διαγραμμάτων υπολοίπων ήταν εξίσου ικανοποιητικά με την εφαρμογή των κλασικών διαγραμμάτων ελέγχου ή χειρότερα σε μερικές περιπτώσεις. Επίσης, στην χαοτική προσομοιωμένη διαδικασία VARMA τα διαγράμματα υπολοίπων απέδωσαν εξαιρετικά καλά και πλέον ικανοποιητικά και οφείλει κανείς αναμφίβολα να τα προτιμήσει έναντι των κλασικών αναλόγων τους. Από τα διαγράμματα που παρουσιάστηκαν στις προηγούμενες σελίδες προκύπτουν τα ακόλουθα συμπεράσματα:

- VARMA(1, 0, 0): Και τα τρία μοντέλα υπολοίπων φάνηκαν να έχουν πολύ καλύτερη συμπεριφορά από τα κλασικά διαγράμματα ελέγχου. Για εντός ελέγχου διαδικασία το αναδρομικό νευρωνικό δίκτυο συνήθιζε να έχει λίγο μικρότερα μέσα μήκη ροής εντός ελέγχου από ότι τα υπόλοιπα διαγράμματα. Για μικρές και μεσαίες μετατοπίσεις, τα υπόλοιπα όλων των μοντέλων είχαν εξαιρετική απόδοση και βρισκόνταν αρκετά κοντά. Το ίδιο ισχύει και για μεγάλες μετατοπίσεις, αν και το T^2 παρουσιάζει πλέον αξιόλογα αποτελέσματα.
- VARMA(1, 0, 1): Σε αυτό το μοντέλο ξεκινά να φαίνεται η αδυναμία του *MESSV* να ανταπεξέλθει σε δεδομένα από μοντέλα VARMA υψηλότερης τάξης, ωστόσο είναι καλύτερο από τα κλασικά διαγράμματα ελέγχου στην ανίχνευση μικρών μετατοπίσεων. Και πάλι το αναδρομικό νευρωνικό δίκτυο και το μοντέλο VARMA είχαν με διαφορά τα καλύτερα αποτελέσματα στην ανίχνευση μικρών μετατοπίσεων και εξαιρετική απόκριση στις μεγάλες μετατοπίσεις.
- VARMA(2, 0, 1): Στο ακόμα περιπλοκότερο μοντέλο, τα διαγράμματα υπολοίπων του αναδρομικού νευρωνικού δικτύου, εμφανώς έχουν την καλύτερη απόδοση στην ανίχνευση μικρών μετατοπίσεων και συμπορεύονται με τα VARMA για τις περιπτώσεις των μεσαίων μετατοπίσεων. Εδώ, είναι η πρώτη φορά που τα δύο αυτά διαγράμματα είχαν αξιοσημείωτα καλύτερα μέσα μήκη ροής εντός ελέγχου από τα κλασικά διαγράμματα ελέγχου.

Συνολικά, τα διαγράμματα υπολοίπων είναι η προτεινόμενη επιλογή για την μοντελοποίηση πολυμεταβλητών δεδομένων από διεργασίες με παρουσία αυτοσυσχέτισης, ειδικά αν αυτή είναι θετική. Προσφέρουν ικανοποιητικά μέσα μήκη ροής εντός ελέγχου και ζηλευτά μέσα μήκη ροής εντός ελέγχου ακόμα και από τροποποιημένα διαγράμματα ελέγχου. Για μικρές μετατοπίσεις, το αναδρομικό νευρωνικό δίκτυο επανειλημμένα εμφάνιζε τις καλύτερες αποκρίσεις και για τις μεσαίες συμπορευόταν με τα υπόλοιπα που προέκυπταν από το μοντέλο VARMA. Για μεγάλες μετατοπίσεις τα διαγράμματα υπολοίπων δυσκολεύτηκαν να ξεπεράσουν την απόδοση των διαγραμμάτων T^2 .

5.5 Αξιολόγηση Διαγραμμάτων σε Πραγματικά Δεδομένα

Απόδοση Διαγραμμάτων σε Πραγματικά Δεδομένα		
	Επιλογή 5 μεταβλητών με την μέθοδο mRMR	
	Σωστές Επισημάνσεις Εκτός Ελέγχου	Λανθασμένες Επισημάνσεις Εκτός Ελέγχου
T^2	76.71 %	75
MCUSUM	69.35%	61
MEWMA	89.52%	140
MEWMA on RNN Residuals	91.93%	89
MEWMA on VARMA Residuals	90.32%	150
MEWMA on MESSV Residuals	77.41%	59
	Επιλογή 10 μεταβλητών με την μέθοδο mRMR	
	Σωστές Επισημάνσεις Εκτός Ελέγχου	Λανθασμένες Επισημάνσεις Εκτός Ελέγχου
T^2	78.23%	74
MCUSUM	73.39%	63
MEWMA	89.52%	106
MEWMA on RNN Residuals	92.74%	88
MEWMA on VARMA Residuals	89.52%	121
MEWMA on MESSV Residuals	79.84%	60

Table 59: Απόδοση Διαγραμμάτων σε Πραγματικά Δεδομένα με Εφαρμογή της μεθόδου mRMR για την Μείωση Διάστασης

Σχολιασμός Διαγραμμάτων σε Πραγματικά Δεδομένα

Αρχικά αξίζει να σημειωθεί ότι με την εφαρμογή της μεθόδου mRMR προς τον σκοπό της μείωσης της διάστασης των δεδομένων από 61 σε 5 και 10, απεδείχθη η ικανότητα της για την εκφραστική και ταυτόχρονα περιληπτική επιλογή μεταβλητών. Και στις 2 περιπτώσεις τα διαγράμματα, κλασικά και υπολοίπων, είχαν παρεμφερή απόδοση, με την μείωση κυρίων των λανθασμένων επισημάνσεων στη περίπτωση που διαλέγονται δέκα μεταβλητές. Όσον αφορά την ορθή αξιολόγηση τιμών ως εκτός ελέγχου, η χρήση δέκα μεταβλητών φαίνεται να έχει μικρό αντίκτυπο, με τα καλύτερα διαγράμματα να μένουν το ίδιο ικανά ή να εντοπίζουν μια παραπάνω θέση εκτός ελέγχου της διεργασίας. Παρατηρούμε ότι η άμεση εφαρμογή του διαγράμματος MCUSUM στη διεργασία προσέφερε απογοητευτικά αποτελέσματα. Το διάγραμμα T^2 καθώς και το διάγραμμα υπολοίπων του μοντέλου MESSV είχαν αρκετά ικανοποιητική απόκριση στον εντοπισμό εκτός ελέγχου σημάτων με αρκετά λιγότερες false positive ενδείξεις σε σχέση με τα υπόλοιπα δεδομένα. Ωστόσο, ο στόχος είναι η μέγιστη δυνατότητα ανίχνευσης της διεργασίας εκτός ελέγχου και για αυτό το ζητούμενο ξεχωρίζουν το διάγραμμα MEWMA καθώς και τα διαγράμματα υπολοίπων του αναδρομικού νευρωνικού δικτύου και του μοντέλου VARMA. Το αναδρομικό νευρωνικό δίκτυο παρουσίασε την μέγιστη απόδοση για την ορθή ανίχνευση εκτός ελέγχου της διεργασίας.

γασίας, ενώ παράλληλα είχε τις χαμηλότερες τιμές εσφαλμένων επισημάνσεων, σε σχέση με τα άλλα δύο διαγράμματα που ήταν κοντά σε απόδοση μεταξύ τους. Καταληκτικά, ακόμη και με την πολύωρη εκπαίδευση που απαιτείται για την ικανοποιητική εκμάθηση του νευρωνικού δικτύου, θεωρείται ότι εμφανώς είχε την καλύτερη απόδοση και θα πρέπει να προτιμηθεί έναντι των άλλων διαγραμμάτων για την παρακολούθηση της αυτοσυσχετισμένης πολυμεταβλητής διεργασίας, με δεύτερα καταλληλότερα το διάγραμμα MEWMA και το διάγραμμα υπολοίπων από το μοντέλο VARMA.

5.6 Επίλογος - Συμπεράσματα

Στην παρούσα εργασία παρουσιάστηκαν και τέθηκαν υπό έρευνα τα κλασσικά διαγράμματα ελέγχου διεργασιών, οι κύριες μοντελοποιήσεις χρονοσειρών και τα αναδρομικά νευρωνικά δίκτυα πάνω σε πολυμεταβλητές και αυτοσυσχετισμένες διεργασίες οι οποίες προέρχονται είτε από προσομοιώσεις είτε από πραγματικά δεδομένα. Παρατηρήθηκε ότι για θετικές αυτοσυσχετίσεις οι συνήθεις πρακτικές μοντελοποίησης δεν έφεραν ικανοποιητικά αποτελέσματα και κρίθηκαν ακατάλληλες για εφαρμογή. Εν αντιθέσει, η μοντελοποίηση του προβλήματος μέσω ενός αναδρομικού δικτύου και η εφαρμογή πάνω στα υπόλοιπα της εφαρμογής του διαγράμματος MEWMA φάνηκε να ξεχωρίζει, ιδιαίτερα κατά την εφαρμογή σε πραγματικά δεδομένα.

Η δυσκολία σύλληψης και ορθής αξιολόγησης της διεργασίας παρουσία θετικής συσχέτισης φαίνεται να ξεπερνάται στο σημείο που η εφαρμογή του αναδρομικού νευρωνικού δικτύου ή του μοντέλου VARMA γίνεται πάνω σε υπόλοιπα που ακολουθούν την κανονική κατανομή εμφανίζουν μικρό βαθμό αυτοσυσχέτισης. Αυτό το αποτέλεσμα επιτεύχθη εξαιτίας της αρχιτεκτονικής του νευρωνικού δικτύου που υπερβαίνει την απλή προσέγγιση της ένα προς ένα αντιστοιχίας συμβάντων και μονάδων στο αναδρομικό νευρωνικό δίκτυο, αλλά μέσα ουσιαστικά της διστρωματικής μοντελοποίησης η διάταξη είναι σε θέση να μάθει τον τρόπο λειτουργίας της διεργασίας καταλληλότερα. Επίσης, η τεχνική του μετακινούμενου παραθύρου φαίνεται να λειτούργησε πλήρως ικανοποιητικά στην πρόβλεψη της χρονοσειράς για την απόκτηση των υπολοίπων.

Βάσει αυτών των αποτελεσμάτων, σαν μελλοντική έρευνα θεωρείται ότι αξίζει η δοκιμή περιπλοκότερων διατάξεων νευρωνικών δικτύων προς εκμάθηση των στοιχείων χρονοσειρών των διεργασιών, με στοχευμένες συναρτήσεις απώλειας και ακόμα μεγαλύτερη δυνατότητα γενίκευσης του μοντέλου.

Βιβλιογραφία

- [1] Afifi Abdelmonem, May Susanne, Clark Virginia A.. Practical Multivariate Analysis, Fifth Edition. *CRC Press* 2012
- [2] Ahsan Muhammad, Mashuri Muhammad, Wibawati, Khusna Hidayatul and Lee Muhammad Hisyam Multivariate Control Chart Based on Kernel PCA for Monitoring Mixed Variable and Attribute Quality Characteristics *MDPI Symmetry* 2020
- [3] Anderson T. W.. Introduction to Multivariate Statistical Analysis, Third Edition. *Wiley-Interscience* 2003
- [4] Aslam Muhammad, Saghir Aamir, Ahmad Liaquat. Introduction to Statistical Process Control. *Wiley* 2021
- [5] Athanasopoulos George and Ashton de Silva Multivariate exponential smoothing for forecasting tourist arrivals to Australia and New Zealand. *Monash Uni.* 2010
- [6] Bersimis, Sotiris and Psarakis, Stelios and Panaretos, John. Multivariate Statistical Process Control Charts: An Overview. *MPRA* 7 November 2006
- [7] Bersimis S., Panaretos J. and Psarakis S.. Multivariate Statistical Process Control Charts and the Problem of Interpretation: A Short Overview and Some Applications in Industry. (2005)
- [8] Bermúdez José D., Corberán-Vallet Ana, Vercher Enriqueta. Multivariate exponential smoothing: A Bayesian forecast approach based on simulation. *Elsevier, Mathematics and Computers in Simulation* Volume 79, Issue 5, January 2009, 1761-1769
- [9] Boone Jeffrey Michael. Contributions to Multivariate Control Charting: Studies of the Z Chart and four non-parametric Charts. (2010)
- [10] Champ, Charles W. . Multivariate Statistical Process Control with Industrial Applications. *Journal of Quality Technology* Vol. 36, No.1, January 2004 126-136
- [11] Dataset: Rare Event Classification in Multivariate Time Series. Chitta Ranjan, Reddy Mahendranath, Mustonen Markku, Paynabar Kamran, Pourak Karim. *arXiv:1809.10717 [stat.ML]*
- [12] Chandrashekar Girish and Sahin Ferat. A survey on feature selection methods. *Elsevier, Computers and Electrical Engineering* Volume 40, Issue 1, January 2014: 16-28
- [13] Guarneri Jean Paulo, Souza Adriano Mendonça, Jacobi Luciane Flores, Reichert Bianca and da Veiga Claudimar Pereira. Control chart based on residues: Is a good methodology to detect outliers?. *Journal of Industrial Engineering International* volume 15 (2019) 119–130

- [14] Harvey Andrew. Handbook of Economic Forecasting. *North-Holland* 2006
- [15] Hyndman Rob J and Athanasopoulos George. Forecasting: Principles and Practice (3rd ed). *O Texts* 2021
- [16] Karaoglan Aslan Deniz and Bayhan Mirac. Performance comparison of residual control charts for trend stationary first order autoregressive processes. *Gazi University Journal of Science* 24 2011 (2):329-339
- [17] Koukouvinos Christos Statistical Process Control. *NTUA Publ.* 2020
- [18] Lowry, C. A, Montgomery, D. C. A review of Multivariate Control Charts. *IIE Transactions*, 27. 1995 :800-810.
- [19] X.S.Lu, M.Xie, T.N.Goh and C.D.Lai Control chart for multivariate attribute processes *INT. J. PROD. RES.* 1998:vol. 36, no. 12, 3477-3489
- [20] Mason, R. L, Young, J. C. Multivariate Statistical Process Control with Industrial Applications. *SIAM, Philadelphia, P.A.* 2002.
- [21] The performance of MCUSUM control charts when the multivariate normality assumption is violated. Nidsunkid S.,Borkowski John J and Budsaba K.. *Thai Statistical Association* 2018
- [22] Mohajerin Nima. Identification and Predictive Control Using RecurrentNeural Networks. (2012)
- [23] Mulder WD, Bethard S, Moens MF. A survey on the application of recurrent neural networks to statistical language modeling. *Computer Speech and Language.* 2015;30(1):61-98
- [24] Nasrullah Khan, Muhammad Aslam, Mansour Sattam Aldosari, and Chi-Hyuck Jun. A Multivariate Control Chart for Monitoring Several Exponential Quality Characteristics Using EWMA. *IEEE Access* 2018: 70349-70358
- [25] Pacella Massimo and Semeraro Quirico. Using recurrent neural networks to detect changes in autocorrelated processes for quality monitoring. *Computers and Industrial Engineering* 52 2007 (4):502-520
- [26] Pareto Vilfredo. Amoroso. *Econometrica* 6. 1938: 1–21.
- [27] Pudjihardjo Ivan, Mulyana Ig. Jaka, Asrini Luh Juni. Autocorrelated Multivariate Control Chart in Cooking Oil Industry. *Proceedings of the International Conference on Industrial Engineering and Operations Management IEOM Society International* 2019
- [28] Ruder Sebastian. An overview of gradient descent optimization algorithms. (2016)

- [29] Sourabh Shastri, Amardeep Sharma, Vibhakar Mansotra, Anand Sharma, Arun Singh Bhadwal, Monika Kumari. A Study on Exponential Smoothing Method for Forecasting. *International Journal of Computer Sciences and Engineering* Volume-6, Issue-4 2018 482-485
- [30] Taleb Hassen. Control charts applications for multivariate attribute processes. *Elsevier Computers and Industrial Engineering* 56. (2009) 399–410
- [31] Topalidou Elisabeth and Stelios Psarakis Review of Multinomial and Multiattribute Quality Control Charts *Qual. Reliab. Engng. Int.* 2009; 25:773–804
- [32] Mary Waterhouse, Ian Smith, Hassan Assareh and Kerrie Mengersen Implementation of multivariate control charts in a clinical setting. *International Journal for Quality in Health Care* 2010; Volume 22, Number 5: 408–414
- [33] Woodall William H. , Lowry Cynthia A., Champ Charles W. and Rigdon Steven E.. A Multivariate Exponentially Weighted Moving Average Control Chart. *American Statistical Association and the American Society for Quality Control, TECHNOMETRICS* February 1992: Vol. 34, no. 1, 46-53
- [34] Zhao Zhenyu, Anand Radhika and Wang Mallory. Maximum Relevance and Minimum Redundancy Feature Selection Methods for a Marketing Machine Learning Platform. (2019)