



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ
ΠΛΗΡΟΦΟΡΙΚΗΣ

Εφαρμογή Τεχνικών Μηχανικής Μάθησης για την Πρόβλεψη Κίνησης στα Οπτικά Δίκτυα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Διονύσιος Π. Αποστολόπουλος

Επιβλέπων : Εμμανουήλ Βαρβαρίγος
Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος, 2020



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ
ΠΛΗΡΟΦΟΡΙΚΗΣ

Εφαρμογή Τεχνικών Μηχανικής Μάθησης για την Πρόβλεψη Κίνησης στα Οπτικά Δίκτυα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Διονύσιος Π. Αποστολόπουλος

Επιβλέπων : Εμμανουήλ Βαρβαρίγος
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 31^η Οκτωβρίου 2020.

.....
Εμμανουήλ Βαρβαρίγος
Καθηγητής Ε.Μ.Π

.....
Ηρακλής Αβραμόπουλος
Καθηγητής Ε.Μ.Π

.....
Θεοδώρα Βαρβαρίγου
Επ. Καθηγήτρια Ε.Μ.Π.

Αθήνα, Οκτώβριος, 2020

.....
Διονύσιος Π. Αποστολόπουλος

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Διονύσιος Αποστολόπουλος, 2020

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Πρόλογος

Η παρούσα διπλωματική εργασία εντάσσεται στο ερευνητικό πεδίο των οπτικών δικτύων και πιο συγκεκριμένα της εφαρμογής τεχνικών μηχανικής μάθησης στα οπτικά δίκτυα. Η ολοένα αυξανόμενη κίνηση στο διαδίκτυο, καθώς και η παράλληλα αυξανόμενη πολυπλοκότητα των τηλεπικοινωνιακών συστημάτων έχουν ως αποτέλεσμα την μετατροπή των τηλεπικοινωνιακών δικτύων σε μέσα κίνησης τεράστιου όγκου ετερογενών δεδομένων. Για την εξαγωγή χρήσιμων πληροφοριών από τα δεδομένα αυτά και τη λήψη αποφάσεων σχετικά με την ορθή λειτουργία του δικτύου απαιτούνται προηγμένα μαθηματικά εργαλεία. Μεταξύ αυτών των μαθηματικών εργαλείων, η μηχανική μάθηση θεωρείται ως μια από τις πιο ελπιδοφόρες μεθοδολογικές προσεγγίσεις για την ανάλυση των δεδομένων του δικτύου και τη λήψη αποφάσεων. Στα πλαίσια της παρούσας εργασίας, μελετάται το πρόβλημα της πρόβλεψης κίνησης στα οπτικά δίκτυα. Στο πρώτο κεφάλαιο παρουσιάζονται βασικοί ορισμοί, έννοιες και αρχιτεκτονικές των οπτικών δικτύων. Στο δεύτερο κεφάλαιο γίνεται αναφορά στην μηχανική μάθηση. Παρουσιάζονται οι κατηγορίες μάθησης, θεμελιώδεις αλγόριθμοι καθώς επίσης πραγματοποιείται εκτενή περιγραφή της μάθησης με επίβλεψη και των νευρωνικών δικτύων. Εν συνεχεία, στο τρίτο κεφάλαιο παρουσιάζονται οι εφαρμογές και οι προκλήσεις της εφαρμογής τεχνικών μηχανικής μάθησης στα οπτικά δίκτυα. Στο τέταρτο κεφάλαιο περιγράφεται πως το πρόβλημα της πρόβλεψης κίνησης ανάγεται στο πρόβλημα πρόβλεψης χρονοσειρών και παρουσιάζονται τα βασικά χαρακτηριστικά των χρονοσειρών καθώς και μέθοδοι διαχείρισής τους. Τέλος, στο πέμπτο κεφάλαιο παρουσιάζονται αναλυτικά οι προσομοιώσεις που πραγματοποιήθηκαν για τις ανάγκες της εργασίας, με στόχο την πρόβλεψη κίνησης μεταξύ κόμβων σε επίπεδο δικτύου, ενός οπτικού δικτύου.

Λέξεις-κλειδιά: οπτικά δίκτυα, μηχανική μάθηση, νευρωνικά δίκτυα, μάθηση με επίβλεψη, διαχείριση κίνησης, πρόβλεψη κίνησης, πρόβλεψη χρονοσειρών

Abstract

This diploma thesis is part of the research field of optical networks and more specifically of the application of machine learning techniques in optical networks. The ever-increasing internet traffic, as well as the increasing complexity of telecommunication systems have resulted in the transformation of telecommunication networks into means of traffic of a huge volume of heterogeneous data. Advanced mathematical tools are required to extract useful information from this data and to make decisions about the proper operation of the network. Among these mathematical tools, machine learning is considered as one of the most promising methodological approaches for network data analysis and decision making. In the context of this research, the problem of traffic prediction in optical networks is studied. The first chapter presents main definitions, concepts and architectures of optical networks. The second chapter refers to machine learning. The categories of learning, fundamental algorithms are presented as well as an extensive description of supervised learning and neural networks. Then, in the third chapter, the applications and challenges of the application of machine learning techniques in optical networks are presented. The fourth chapter describes how the problem of traffic prediction is reduced to the problem of time series forecasting and presents the main characteristics of time series as well as their management methods. Finally, the fifth chapter presents in detail the simulations performed for the needs of the research, with the aim of predicting traffic between nodes at the network level of an optical network.

Key-words: optical networks, machine learning, neural networks, supervised learning, traffic engineering, traffic prediction, time series forecasting

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

Πρόλογος	7
Abstract	9
Κατάλογος Ακρωνυμίων	13
Εισαγωγή	15
Κεφάλαιο 1 - Οπτικά Δίκτυα	17
1.1 Πολυπλεξία με διαίρεση μήκους κύματος – WDM	17
1.2 Συνοπτική περιγραφή σύγχρονων WDM συστημάτων	18
1.3 Εξέλιξη Οπτικών Δικτύων	21
1.3.1 Οπτικά Δίκτυα 1 ^{ης} γενιάς.....	22
1.3.2 Οπτικά Δίκτυα 2 ^{ης} γενιάς.....	23
1.3.3 Οπτικά Δίκτυα 3 ^{ης} γενιάς.....	25
1.4 Αρχιτεκτονική WDM Οπτικών Δικτύων	27
1.5 Ιεραρχική δομή Οπτικών Δικτύων	28
1.6 Κατηγοριοποίηση Οπτικών Δικτύων	30
1.6.1 Κατηγοριοποίηση ως προς τον προορισμό της οπτικής σύνδεσης.....	30
1.6.2 Κατηγοριοποίηση ως προς την ζεύξη.....	32
1.6.3 Κατηγοριοποίηση ως προς τις τεχνολογίες πρόσβασης	33
1.7 Στοιχεία Οπτικών Δικτύων	35
1.8 Ιδιότητες WDM δικτύου	39
Κεφάλαιο 2 - Μηχανική Μάθηση	43
2.1 Εισαγωγή.....	43
2.2 Κατηγορίες αλγορίθμων μάθησης.....	43
2.3 Επιλογή αλγόριθμου Μηχανικής Μάθησης.....	45
2.4 Μέθοδος διαχωρισμού ενός dataset	46
2.5 Τύποι δεδομένων και ορισμοί	47
2.6 Μέθοδοι επεξεργασίας των δεδομένων.....	47
2.7 Εξαγωγή χαρακτηριστικών προβλήματος – Feature Engineering	49
2.8 Διαχείριση χαρακτηριστικών που λείπουν από ένα dataset.....	50
2.9 Ρύθμιση υπερπαραμέτρων	50
2.10 Supervised Learning.....	51
2.10.1 Λειτουργία μάθησης με επίβλεψη	51
2.10.2 Θεμελιώδεις αλγόριθμοι μάθησης με επίβλεψη	51
2.10.2.1 Linear Regression	51
2.10.2.2 Logistic Regression.....	52
2.10.2.3 Δέντρο Λήψης Απόφασης (Decision Tree Learning).....	54
2.10.2.4 Support Vector Machine – SVM	54
2.10.2.5 k-Nearest Neighbors – k-NN	55
2.10.3 Δομικά στοιχεία ενός αλγόριθμου μάθησης.....	56
2.10.4 Λειτουργία αλγόριθμου Gradient Descent	57
2.10.5 Εκτίμησης απόδοσης μοντέλου	58
2.10.6 Underfitting και Overfitting	62

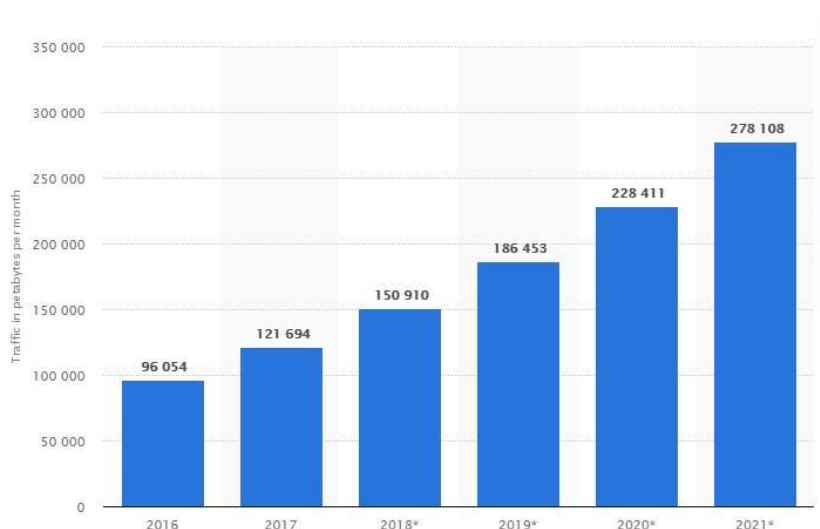
2.11 Νευρωνικά Δίκτυα	64
2.11.1 Νευρωνικά Δίκτυα πολλών επιπέδων (Multilayer Perceptron).....	65
2.11.2 Αρχιτεκτονική Feed-Forward (πρόσθιας τροφοδότησης)	66
2.11.3 Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks – RNN) ...	67
2.11.4 Δίκτυα LSTM (Long Short Term Memory)	69
2.12 Deep Learning	72
2.12.1 Συνελκτικά Νευρωνικά Δίκτυα	73
2.13 Unsupervised Learning	75
2.14 Semi-Supervised Learning	76
2.15 Reinforcement Learning.....	78
2.16 Διευκρινίσεις	79
Κεφάλαιο 3 - Εφαρμογές Μηχανικής Μάθησης στα Οπτικά Δίκτυα	81
3.1 Εισαγωγή.....	81
3.2 Προκλήσεις για την εφαρμογή τεχνικών Μηχανικής Μάθησης στα Οπτικά Δίκτυα	82
3.3 Εφαρμογές της Μηχανικής Μάθησης στο φυσικό επίπεδο	84
3.4 Εφαρμογές της Μηχανικής Μάθησης στο επίπεδο δικτύου	86
Κεφάλαιο 4 - Πρόβλεψη Χρονοσειρών	91
4.1 Εισαγωγή.....	91
4.2 Πρόβλεψη χρονοσειρών.....	91
4.3 ARIMA	92
4.4 SARIMA	96
Κεφάλαιο 5 - Προσομοιώσεις.....	99
5.1 Εισαγωγή.....	99
5.2 Κριτήρια επιλογής μοντέλων	99
5.3 Software	99
5.4 Συλλογή δεδομένων	100
5.4.1 Δημιουργία κίνησης και παραγωγή συνθετικών δεδομένων.....	100
5.4.2 Πραγματικά δεδομένα κίνησης	103
5.5 Εκτίμηση απόδοσης μοντέλων.....	104
5.6 Πραγματοποίηση Προσομοιώσεων.....	106
5.6.1 Πρόβλεψη κίνησης με χρήση CNN Νευρωνικού Δικτύου.....	107
5.6.2 Πρόβλεψη κίνησης με χρήση SARIMA μοντέλου.....	113
5.6.3 Πρόβλεψη κίνησης με χρήση LSTM Νευρωνικού Δικτύου	121
5.7 Αποτελέσματα.....	127
Επίλογος	129
Βιβλιογραφία	131

Κατάλογος Ακρωνυμίων

IoT	Internet of Things	SR-TSP	Short-Reach Transponder
EON	Elastic Optical Network	EDFA	Erbium Doped Fiber Amplifier
WDM	Wavelength Division Multiplexing	SOA	Semiconductor Optical Amplifier
ML	Machine Learning	WSS	Wavelength Selective Switch
AI	Artificial Intelligence	OADM	Optical add/drop Multiplexer
DWDM	Dense Wavelength Division Multiplexing	SL	Supervised Learning
TSP	Transponder	UL	Unsupervised Learning
OLT	Optical Line Terminal	SSL	Semi Supervised Learning
WWDM	Wide Wavelength Division Multiplexing	RL	Reinforcement Learning
CWDM	Coarse Wavelength Division Multiplexing	SVM	Support Vector Machine
OCH	Optical CHannel	k-NN	k-Nearest Neighbors
OMS	Optical Multiplex Section	MSE	Mean Squared Error
OTS	Optical Transmission Section	NN	Neural Network
P2P	Point-to-Point	MLP	Multilayer Perceptron
DXC	Digital Cross-Connector	FFNN	Feed-Forward Neural Network
OAMP	Optical AMPlifier	RNN	Recurrent Neural Network
WADM	Wavelength Add-Drop Multiplexers	LSTM	Long Short Term Memory
OXC	Optical Cross-Connector	CNN	Convolutional Neural Network
WCS	Wavelength Convertible Switches	GMM	Gaussian Mixture Model
RWA	Routing and Wavelength Assignment	MDP	Markov decision processes
PON	Passive Optical Network	BER	Bit Error Rate
OPS	Optical Packet Switching	QoS	Quality of Service
OBS	Optical Burst Switching	QoT	Quality of Transmission
OLS	Optical Label Switching	OSNR	Optical Signal-to-Noise Ratio
SONET	Synchronous Optical Networking	MFR	Modulation format recognition
SDH	Synchronous Digital Hierarchy	CD	Chromatic Dispersion
WAN	Wide Area Network	OPM	Optical performance monitoring
MAN	Metropolitan Area Network	VTD	Virtual Topology Design
AN	Access Networks	SLA	Service Level Agreement
FTTx	Fiber To The x	MTTR	Mean Time To Repair
PMP	Point-to-Multipoint	WITS	Waikato Internet Traffic Storage
AON	Active Optical Network	RMSE	Root Mean Squared Error
TSP	Transponder	RRSE	Root Relative Squared Error
LR-TSP	Long-Reach Transponder	ADF	Augmented Dickey-Fuller

Εισαγωγή

Η εισαγωγή του διαδικτύου στις ζωές μας τις τελευταίες δεκαετίες έχει μετασχηματίσει σε μεγάλο βαθμό τις συνήθειές μας, τον τρόπο που επικοινωνούμε μεταξύ μας, την ροή των πραγμάτων καθώς και το διαμοιρασμό της πληροφορίας. Θεμελιώδεις συνήθειες όπως η εργασία και η εκπαίδευση τείνουν να αποκτήσουν απομακρυσμένο χαρακτήρα (remote). Με την αυξανόμενη επιρροή του σε μεμονωμένους καταναλωτές και σε μεγάλες οικονομίες, το Διαδίκτυο έχει γίνει όλο και πιο ζωτικής σημασίας μέρος της καθημερινής μας ζωής. Το 2019 ο αριθμός των χρηστών του διαδικτύου παγκοσμίως ανήλθε σε 4,13 δισεκατομμύρια, πράγμα που σημαίνει ότι πάνω από το μισό του παγκόσμιου πληθυσμού είναι συνδεδεμένο με τον παγκόσμιο ιστό [1]. Επιπρόσθετα, νέες εφαρμογές και τεχνολογίες έχουν εισέλθει στις ζωές μας, όπως το υπολογιστικό «νέφος» (cloud), το διαδίκτυο των αντικειμένων (Internet of Things - IoT) καθώς και η έλευση των δικτύων κινητής τηλεφωνίας πέμπτης γενιάς (5G). Υπολογίζεται πως η κίνηση στο διαδίκτυο αναμένεται να αυξάνεται συνεχώς με ρυθμό περίπου 25% ανά χρόνο ως το 2021, ως αποτέλεσμα όλων αυτών των συνηθειών και των τεχνολογιών [2].



Εικόνα 1: Πρόβλεψη παγκόσμιας κίνησης δεδομένων μέχρι το 2021

Το μεγαλύτερο μέρος της κίνησης αυτής θα είναι τοπική, επιβαρύνοντας σημαντικά το δίκτυο πρόσβασης και το μητροπολιτικό και σε μικρότερο βαθμό το δίκτυο κορμού. Τα μητροπολιτικά δίκτυα και τα δίκτυα κορμού τα τελευταία χρόνια αναβαθμίζονται σε ελαστικά οπτικά δίκτυα (Elastic Optical Network - EON) που αποτελούν εξέλιξη της τεχνολογίας WDM (Wavelength Division Multiplexing) παρέχοντας μεγαλύτερη ευελιξία και αποδοτικότητα.

Πέραν όμως της αύξησης της κίνησης στο διαδίκτυο, πολλές από τις νέες εφαρμογές έχουν «εκρηκτική» συμπεριφορά, δηλαδή σε κάποιες χρονικές περιόδους οι απαιτήσεις

της δικτυακής χωρητικότητάς τους είναι πολύ μικρές και σε κάποιες άλλες αυξάνονται απότομα, ενισχύοντας την ανάγκη ελαστικής και ευέλικτης λειτουργίας των οπτικών δικτύων. Ένα τέτοιο παράδειγμα είναι η απευθείας μετάδοση ενός σημαντικού αθλητικού γεγονότος. Σε αυτό το περιβάλλον υπάρχει η ανάγκη δυναμικής και ευέλικτης λειτουργίας και συνεργασίας των διαφορετικών τμημάτων του δικτύου. Σημαντικό ρόλο παίζει η πρόβλεψη της μεταβολής της κίνησης σε συγκεκριμένα σημεία του δικτύου, ώστε να είναι δυνατή η επαναρύθμιση του δικτύου για να ανταποκριθεί στις καινούργιες απαιτήσεις.

Η Μηχανική Μάθηση (Machine Learning - ML), η οποία αποτελεί έναν υποκλάδο της Τεχνητής Νοημοσύνης (Artificial Intelligence - AI), θεωρείται η πιο δημοφιλής μεθοδολογική προσέγγιση για την επίλυση αυτών των προβλημάτων. Οι αλγόριθμοι μηχανικής μάθησης, χαρακτηρίζονται από μία μοναδική ικανότητα να μαθαίνουν τη συμπεριφορά του συστήματος από παρελθοντικά δεδομένα και να πραγματοποιούν εκτιμήσεις για τις μελλοντικές αποκρίσεις με βάση το μοντέλο του συστήματος. Έτσι, οι αλγόριθμοι αυτοί, εκμεταλλευόμενοι την πληθώρα διαθέσιμων δεδομένων, τόσο για τη συμπεριφορά του δικτύου, όσο και για τη συμπεριφορά των χρηστών, παρέχουν τη δυνατότητα δημιουργίας προβλέψεων, αναγνώρισης και κατηγοριοποίησης σφαλμάτων, ταξινόμησης ροών κυκλοφορίας, με σκοπό την λήψη αποφάσεων για τη βέλτιστη ευελιξία και αποδοτικότητα των οπτικών δικτύων.

Στόχος της παρούσας εργασίας είναι να μελετηθούν οι κλάδοι των οπτικών δικτύων και της μηχανικής μάθησης, να αναδειχθούν οι ανάγκες εφαρμογής τεχνικών μηχανικής μάθησης στον τομέα των οπτικών δικτύων και να μελετηθεί το πρόβλημα της πρόβλεψης κίνησης στο επίπεδο IP των οπτικών δικτύων.

Κεφάλαιο 1 - Οπτικά Δίκτυα

1.1 Πολυπλεξία με διαίρεση μήκους κύματος – WDM

Η τεχνική πολυπλεξίας στην οποία βασίζονται τα οπτικά δίκτυα είναι η πολυπλεξία με διαίρεση μήκους κύματος (Wavelength Division Multiplexing – WDM) [3]. Με τον όρο πολυπλεξία με διαίρεση μήκους κύματος, ουσιαστικά αναφερόμαστε σε οπτικό φέρον σήμα το οποίο προσδιορίζεται από το μήκος κύματος του. Κατά την τεχνική αυτή, ένας αριθμός από οπτικά φέροντα σήματα πολυπλέκονται σε ένα οπτικό μέσο (οπτική ίνα) χρησιμοποιώντας διαφορετικά μήκη κύματος. Το εύρος ζώνης δηλαδή, χωρίζεται σε έναν αριθμό από διαφορετικά διαστήματα, καθένα εκ των οποίων εκτείνεται γύρο από ένα βασικό μήκος κύματος (wavelength). Αυτά τα μήκη κύματος ονομάζονται και οπτικά κανάλια, όπως επίσης οπτικά μονοπάτια ή ακόμα και lightpaths. Χαρακτηριστικό τους είναι πως το κάθε κανάλι μπορεί να διαμορφώνεται ανεξάρτητα από τα υπόλοιπα.

Με αυτόν τον τρόπο δύναται η ταυτόχρονη χρησιμοποίηση πολλαπλών πηγών εκπομπής, καθώς η κάθε πηγή εκπέμπει ανεξάρτητα και στη συχνότητα που της αντιστοιχεί, δηλαδή στο μήκος κύματος που της έχει ανατεθεί. Αποτέλεσμα αυτού, η ταυτόχρονη και ανεξάρτητη μετάδοση οπτικού σήματος πάνω στον ίδιο οπτικό σύνδεσμο.

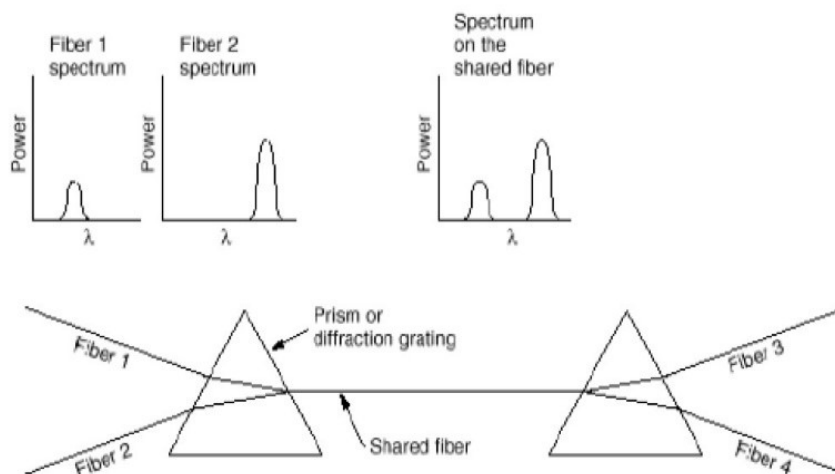
Αποτέλεσμα της εφαρμογής αυτής της τεχνικής είναι η αμφίδρομη επικοινωνία μέσω της ίνας και ο πολλαπλασιασμός της χωρητικότητά της [4]. Ο συνδυάζοντας της τεχνικής WDM με τη χρήση οπτικών ενισχυτών κατέστησε την οπτική ίνα ικανή να μπορεί να μεταφέρει τεράστιο όγκο δεδομένων σε υψηλές ταχύτητες.

Η αμερικάνικη εταιρεία AT&T πρωτοπόρησε στην ανάπτυξη της τεχνολογίας αυτής μέσω την φημισμένων εργαστηρίων της (AT&T Bell Laboratories) [5]. Εκεί επετεύχθη η ταυτόχρονη διάδοση διαφορετικών μηκών κύματος μέσω μίας οπτικής ίνας, καθώς τα μήκη κύματος αυτά διαχωρίζονταν από μία αρχική δέσμη φωτός με τη χρήση ενός πρίσματος. Έτσι, κατά μήκος μίας οπτικής ίνας δημιουργούνται πολλά οπτικά κανάλια διότι οι οπτικές δέσμες είναι διακριτές καθότι το μέγιστο μήκος κύματος έχει τοποθετηθεί σε απόσταση επαρκή από τα υπόλοιπα. Τα πρώτα WDM συστήματα διέθεταν 4 κανάλια [6]. Παρατηρήθηκε όμως η εισαγωγή ενός μη ανεκτού επιπέδου θορύβου στο μήκος κύματος του σήματος κατά την μετάδοση του, γεγονός που οφειλόταν σε ερβιομιγείς ενισχυτές, οι οποίοι δεν ενίσχυαν το σήμα ομοιόμορφα στην περιοχή του φασματικού κέρδους τους. Άμεσο αποτέλεσμα αυτού του θορύβου κατέστησε την τεχνολογία πολυπλεξίας μήκους κύματος μη πρακτική για επικοινωνίες ινών μεγάλων αποστάσεων.

Με σκοπό την αντιμετώπιση του παραπάνω θορύβου, αναπτύχθηκε ο πρώτος οπτικός ενισχυτής (optical amplifier) 2 επιπέδων [7]. Το πλεονέκτημα αυτού του ενισχυτή ήταν η δυνατότητα μετάδοσης πληροφορίας σε διαφορετικά και πολλαπλά μήκη κύματος αλλά με ομοιόμορφο κέρδος. Στον ενισχυτή αυτόν βασίστηκε και το πρώτο εμπορικό πυκνό σύστημα WDM (dense WDM - DWDM), το οποίο παρείχε 16 κανάλια και

επιτύγχανε ταχύτητες μετάδοσης έως και 40 Gbit/s. Χαρακτηριστικά, τα σύγχρονα συστήματα DWDM προσφέρουν ταχύτητες μετάδοσης 100 Gbit/s σε κάθε μήκος κύματος και συνολική ταχύτητα 1 Tbit/s [8].

Στο παρακάτω σχήμα παρουσιάζεται η λειτουργία της πολυπλεξίας μήκους κύματος με χρήση πολλαπλών καναλιών και μηκών κύματος:



Εικόνα 2: Λειτουργία πολυπλεξίας μήκους κύματος

Τα WDM συστήματα λειτουργούν κατά κόρον σε μονότροπες οπτικές ίνες. Ορισμένοι τύποι WDM συστημάτων μπορούν να λειτουργήσουν και σε πολύτροπες ίνες.

Η δημοτικότητα των WDM συστημάτων και η καθολική χρήση τους στα σύγχρονα τηλεπικοινωνιακά δίκτυα έγκειται στην δυνατότητα που παρέχουν για την διαχείριση πολλαπλών σημάτων και κατ' επέκταση την επέκταση ενός βασικού σήματος μέσω μίας ίνας με πολύ υψηλές ταχύτητες. Έτσι το δίκτυο επεκτείνεται χρησιμοποιώντας λιγότερους πόρους σε οπτική ίνα.

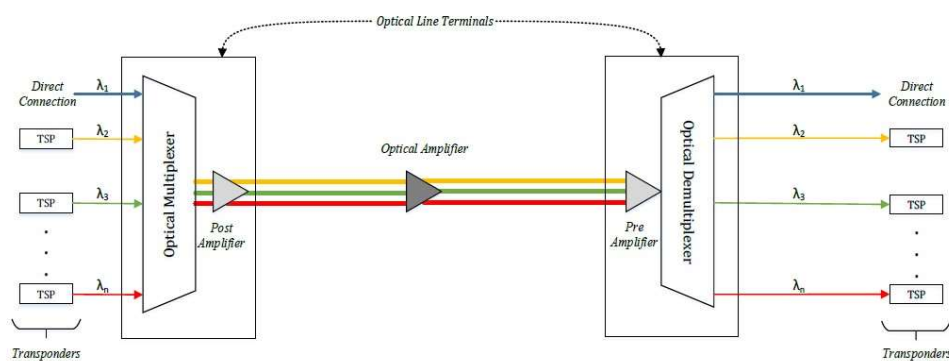
1.2 Συνοπτική περιγραφή σύγχρονων WDM συστημάτων

Όπως ήδη αναφέρθηκε, το πλεονέκτημα της WDM πολυπλεξίας είναι η ταυτόχρονη, ταχύτατη και ανεξάρτητη μετάδοση οπτικού σήματος πάνω από τον ίδιο οπτικό σύνδεσμο [9].

Η διασύνδεση ενός ζεύγους κόμβων μπορεί να επιτευχθεί όταν ένα οπτικό μονοπάτι που διατρέχει ένα πλήθος συνεχόμενων ινών στο δίκτυο πολυπλέκεται με άλλα οπτικά μονοπάτια, με μόνη προϋπόθεση αυτά να έχουν διαφορετικά μήκη κύματος. Η πολυπλεξία αυτή πραγματοποιείται στην κόμβο πρόσβασης (access node) και έτσι αυτά προωθούνται στο δίκτυο [10]. Στους κόμβους πρόσβασης είναι εγκατεστημένοι οι αποπολυπλέκτες.

Οι αποπολυπλέκτες λειτουργούν ως οπτικά φίλτρα, τα οποία είναι συντονισμένα στα αντίστοιχα μήκη κύματος, εκεί διαχωρίζονται και προωθούνται στους αντίστοιχους δέκτες όπου και αποδιαμορφώνονται.

Στο παρακάτω σχήμα παρουσιάζεται αναλυτικά ένα point-to-point WDM σύστημα μετάδοσης:



Εικόνα 3: WDM σύστημα μετάδοσης

Ένα WDM σύστημα όπως αυτό παρουσιάζεται λειτουργεί με τον εξής τρόπο:

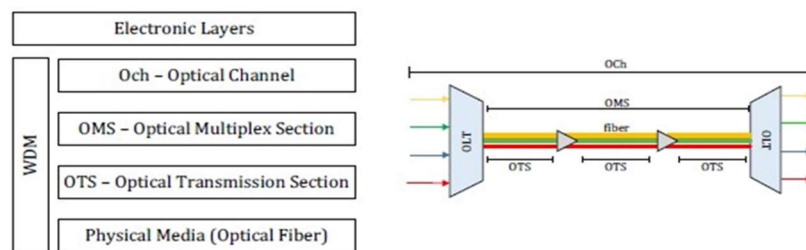
Ένας πομποδέκτης (Transponder - TSP) δέχεται ως είσοδο ένα οπτικό σήμα. Εν συνεχεία, αυτή το σήμα εισόδου αντιστοιχίζεται σε ένα μήκος κύματος (λ). Το μήκος κύματος αυτό αποτελεί ένα κανάλι το οποίο λειτουργεί ανάλογα με το ρυθμό μετάδοσης του αναμεταδότη. Τα μη επικαλυπτόμενα μήκη κύματος από τους πομποδέκτες πολυπλέκονται σε ένα μοναδικό οπτικό σήμα και περνάνε στην οπτική ίνα. Όταν τα σήματα φτάσουν στον αποπολυπλέκτη, αποπολυπλέκονται ώστε να αξιοποιηθεί το κάθε μήκος κύματος ξεχωριστά, το οποίο στη συνέχεια μπορεί να αντιστοιχηθεί σε στην κατάλληλη μορφή εξόδου προς τα τοπικά δίκτυα, χρησιμοποιώντας πάλι πομποδέκτες. Επίσης, για την ενίσχυση του σήματος, όπως αυτό εγκαταλείπει τη διάταξη με προορισμό τα τοπικά δίκτυα, ενδέχεται να συναντήσουμε έναν ενισχυτή (post amplifier) στο τερματικό οπτικής γραμμής (Optical Line Terminal - OLT) του πομπού. Όσο αυξάνεται ο αριθμός των μηκών κύματος που πολυπλέκονται, κατ' αναλογία αυξάνεται το χρησιμοποιούμενο εύρος ζώνης. Αυτό είναι και το πλεονέκτημα έναντι των ηλεκτρονικών δικτύων.

Τα WDM συστήματα μπορούν να ταξινομηθούν σε τρία διαφορετικά πρότυπα μήκους, στο ευρύ WDM (Wide WDM - WWDM), στο αδρό WDM (Coarse WDM - CWDM) και στο πυκνό WDM (Dense WDM - DWDM). Ανεξαρτήτως βέβαια της κατηγορίας στην οποία ανήκει, τα αναφέρουμε πάντα ως WDM [11].

- Το ευρύ WDM (WWDM) συναντάται κατά κόρον στα παθητικά οπτικά δίκτυα υποστηρίζει διαπόσταση καναλιών (channel spacing) μεγαλύτερη ή ίση με 50 nm στο φάσμα των 1550 nm. Τα χαρακτηριστικά των πομπών, των δεκτών και των ενισχυτών που χρησιμοποιούνται καθορίζονται από την διαπόσταση των καναλιών, από την οποία εξαρτάται και ο αριθμός των καναλιών.

- Το αδρό WDM (CWDM) συναντάται στα μητροπολιτικά δίκτυα. Υποστηρίζει διαπόσταση καναλιών 20 nm, μικρό αριθμό μηκών κύματος στο φάσμα των 1300-1600 nm. Επίσης πετυχαίνει υλοποιήσεις εφαρμογών χαμηλότερου κόστους.
- Το πυκνό WDM (DWDM) αποτελεί ιδανική λύση για δίκτυα μεγάλων αποστάσεων (long haul). Υποστηρίζει στενή διαπόσταση καναλιών στα 8 nm [12].

Ένα WDM οπτικό δίκτυο μπορεί να οριστεί ως μία ιεραρχική δομή [13], με σταθερή κατανομή και λειτουργίες κατά μήκος του δικτύου και των σημείων τερματισμού. Η ιεραρχική αυτή δομή περιλαμβάνει τέσσερα υπό-επίπεδα: Το Οπτικό Κανάλι (Optical Channel - OCh), το επίπεδο Οπτικής Πολυπλεξίας (Optical Multiplex Section - OMS), το επίπεδο Οπτικής Μετάδοσης (Optical Transmission Section - OTS) και το Φυσικό Επίπεδο. Το παραπάνω μοντέλο καθώς και οι λειτουργίες του κατανέμονται κατά μήκος του δικτύου. Μία τέτοια ιεραρχική δομή ενός WDM οπτικού δικτύου παρουσιάζεται στο παρακάτω σχήμα:



Εικόνα 4: Ιεραρχική δομή WDM δικτύου

- Το χαμηλότερο υπό-επίπεδο, το Φυσικό Επίπεδο δηλαδή, είναι μια οπτική ίνα η οποία εξυπηρετεί τη μετάδοση του σήματος στο επίπεδο οπτικής μετάδοσης.
- Το υψηλότερο υπό-επίπεδο είναι το Οπτικό Κανάλι (Och) (το λεγόμενο και υπό-επίπεδο μονοπατιού - path sublayer) το οποίο εξυπηρετεί μεμονωμένες συνδέσεις ή οπτικά μονοπάτια κατά μήκος του δικτύου. Ένα παράδειγμα λειτουργίας του είναι η προσθήκη ή ο τερματισμός μηκών κύματος για την δρομολόγηση των οπτικών καναλιών. Όπως έχει ήδη αναφερθεί, μέσω ενός οπτικού καναλιού μπορεί να επιτευχθεί η διασύνδεση ενός ζεύγους κόμβων, με τους ενδιάμεσους κόμβους να υποστηρίζουν το κανάλι παρέχοντας οπτική παράκαμψη (optical bypass).
- Το υπό-επίπεδο της Οπτικής Πολυπλεξίας (OMS) παρέχει την δυνατότητα ενός οπτικού σήματος πολλαπλών μηκών κύματος καθώς ελέγχει την πολυπλεξία όλων των οπτικών καναλιών που διατρέχουν μια συγκεκριμένη οπτική ίνα, συγκεκριμένου μήκους κύματος το καθένα. Η απόσταση μεταξύ των οπτικών ενισχυτών ή η απόσταση ανάμεσα σε έναν οπτικό ενισχυτή και ένα τερματικό οπτικής γραμμής δημιουργούν διάφορα τμήματα συνδέσμων. Το συγκεκριμένο υπό-επίπεδο απαρτίζεται από διάφορα τμήματα συνδέσμων, όπως αυτά που περιγράψαμε.

- Κάθε ένα από αυτά τα τμήματα αποτελεί ένα υπό-επίπεδο Οπτικής Μετάδοσης (OTS). Οι οπτικοί ενισχυτές εισάγονται κατά το μήκος της μετάδοσης ώστε να αντισταθμιστούν οι απώλειες ισχύος καθώς το σήμα διατρέχει την ίνα.

1.3 Εξέλιξη Οπτικών Δικτύων

Οι απαιτήσεις των σύγχρονων επικοινωνιακών δικτύων για πληθώρα υπηρεσιών ολοένα αυξανόμενης χωρητικότητας έχουν καταστήσει την οπτική ίνα ως το ιδανικό τηλεπικοινωνιακό κανάλι για την υλοποίησή τους [13].

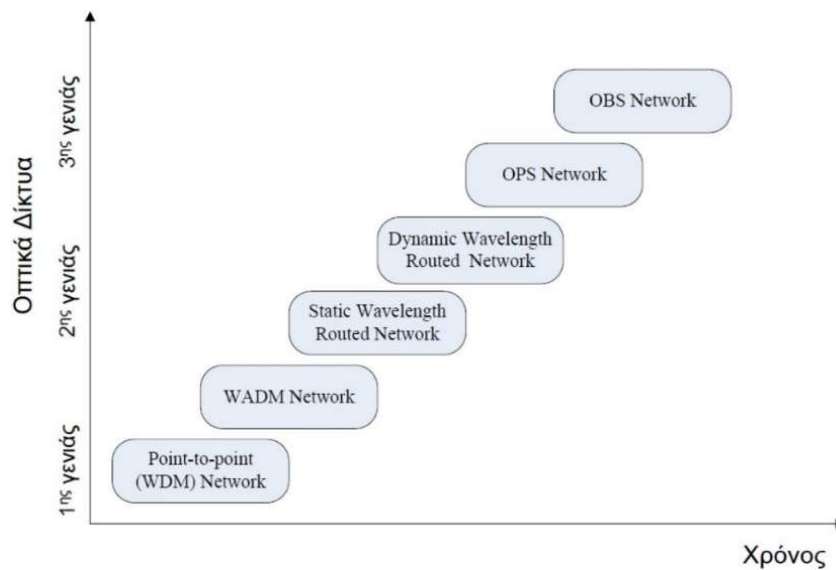
Παρά την ύπαρξη της οπτικής ίνας, δεν οδηγηθήκαμε άμεσα στην ανάπτυξη τηλεπικοινωνιακών συστημάτων που να βασίζονται σε αυτή. Για το λόγο αυτό πριν τη δεκαετία του 1970 η απώλεια ισχύος λάμβανε τεράστιες τιμές της τάξεως των 1000 dB/Km. Η πρόοδος όμως που ακολούθησε στον τομέα των οπτικής τεχνολογίας τις επόμενες δεκαετίες οδήγησε στην ανάπτυξη ινών χαμηλής εξασθένησης, προσεγγίζοντας τελικά την τιμή των 0.2 dB/Km για μήκος κύματος $\lambda=1.55 \mu\text{m}$, ενώ ταυτόχρονα οδηγηθήκαμε στην ανάπτυξη και εξέλιξη των απαραίτητων δομικών στοιχείων μιας οπτικής ζεύξης. Στοιχεία όπως ο οπτικός πομπός, ο οπτικός δέκτης που βασικό του τμήμα έχει το φωτοανιχνευτή για τη μετατροπή του οπτικού σήματος σε ηλεκτρικό, οι οπτικοί ενισχυτές για την αντιμετώπιση των απωλειών ισχύος και οι συνδετήρες για τις συνδέσεις του πομπού και του δέκτη με την ίνα ή για τις επιμέρους συνδέσεις τμημάτων ίνας. Τι ορίζουμε λοιπόν ως οπτικό δίκτυο;

Ένα οπτικό δίκτυο ορίζεται ως σύνολο κόμβων (εξοπλισμός μεταγωγής), που συνδέονται μεταξύ τους μέσω των ζευξέων οπτικών ινών (εξοπλισμός μετάδοσης).

Παρά το γεγονός ότι η μετάδοση πραγματοποιείται στην οπτική ίνα, δεν πρόκειται απαραίτητα για ένα αμιγώς οπτικό δίκτυο, αφού η μεταγωγή μπορεί να αφορά πακέτα, κύκλωμα ή ροές κλασμάτων μήκους κύματος οπότε και να έχουμε λειτουργία είτε στο ηλεκτρονικό φάσμα, είτε στο οπτικό είτε σε συνδυασμό των δύο με χρήση της κατάλληλης τεχνικής οπτικής πολυπλεξίας [6].

Η δυνατότητα που παρέχει η πολυπλεξία μήκους κύματος για την αξιοποίηση των πλεονεκτημάτων των οπτικών ινών, καθώς επιτρέπεται στο χρήστη να λειτουργεί με την ταχύτητα που υποστηρίζει ένα WDM κανάλι και η οποία μπορεί να επιλεγεί αυθαίρετα ώστε να συμβαδίζει με την ταχύτητα των ηλεκτρονικών διατάξεων, καθιστούν την τεχνολογία WDM ως την πιο διαδεδομένη τεχνική πολυπλεξίας στα υπάρχοντα οπτικά δίκτυα επικοινωνίας.

Παρόλα αυτά, με τον τρόπο που ταχύτατα εξελίσσεται ο τομέας των οπτικών δικτύων, οδηγούμαστε στην υιοθέτηση καινούριων τεχνικών οπτικής πολυπλεξίας καθότι η τεχνική WDM περιορίζει τις νέες γενιές συστημάτων.

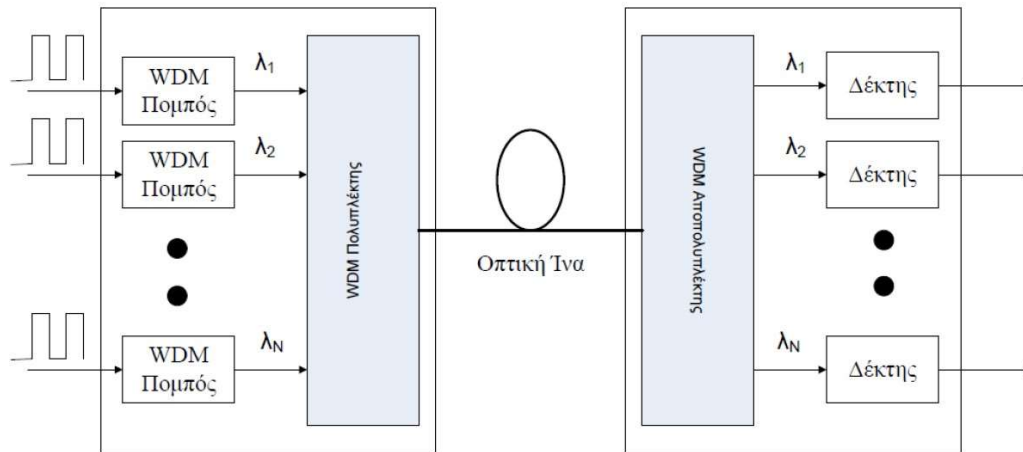


Εικόνα 5: Εξέλιξη οπτικών δικτύων

1.3.1 Οπτικά Δίκτυα 1^{ης} γενιάς

Στα οπτικά δίκτυα 1^{ης} γενιάς συναντάμε τη χρήση της οπτικής ίνας αποκλειστικά στη φυσική διασύνδεση δύο κόμβων. Ουσιαστικά αναφερόμαστε σε οπτικές ζεύξεις σημείο-προς-σημείο (point-to-point links – P2P) [6]. Για να πετύχουμε αύξηση του διαθέσιμου εύρους ζώνης στις σημείο-προς-σημείο οπτικές ζεύξεις, διατίθενται δύο εναλλακτικές σχεδιαστικές επιλογές:

- Η εγκατάσταση και χρήση επιπρόσθετου οπτικού εξοπλισμού (δηλαδή οπτικών ινών και τερματικού εξοπλισμού στη μεταξύ τους σύνδεση. Λύση που ενδείκνυται για σύνδεση δύο κόμβων σε απόσταση μικρότερη των 50 χλμ.
- Η χρήση τεχνολογίας WDM με σκοπό τη μετάδοση πολλαπλών καναλιών διαφορετικής συχνότητας μέσα από την ίδια οπτική ίνα. Η λύση αυτή ενδείκνυται για σύνδεση δύο κόμβων σε απόσταση μικρότερη των 50 χλμ.



Εικόνα 6: Οπτικό δίκτυο 1ης γενιάς με N κανάλια

Η ηλεκτρονική μετατροπή των οπτικών σημάτων καθίσταται αναγκαία, καθώς όλες οι λειτουργίες που σχετίζονται με τη μετάδοση της πληροφορίας, όπως η μεταγωγή (switching), η δρομολόγηση (routing), η μετατροπή του μήκους κύματος (wavelength conversion), η αναγέννηση των οπτικών παλμών (pulse regeneration) και η ανάκτηση ρολογιού (clock recovery) πραγματοποιούνται στην ηλεκτρονική περιοχή με χρήση των αντίστοιχων ηλεκτρονικών διατάξεων στους κόμβους του δικτύου. Οπτικών σημάτων που όχι μόνο ξεκινούν ή τερματίζουν σε έναν κόμβο, αλλά και οπτικών σημάτων που διαπερνούν έναν κόμβο χωρίς να προστίθενται ή να αφαιρούνται δεδομένα σε ένα μήκος κύματος (διερχόμενη κίνηση - bypass traffic).

Για την επεξεργασία των δεδομένων στους κόμβους χρησιμοποιήθηκαν ψηφιακοί καταναμητές (Digital Cross-Connectors-DXC), ενώ για το πρόβλημα της μακρινής σύνδεσης δύο κόμβων χρησιμοποιήθηκαν αναγεννητές (regenerators). Οι αναγεννητές αυτοί στην πορεία αντικαταστάθηκαν από οπτικούς ενισχυτές (Optical Amplifiers-OAMPs)

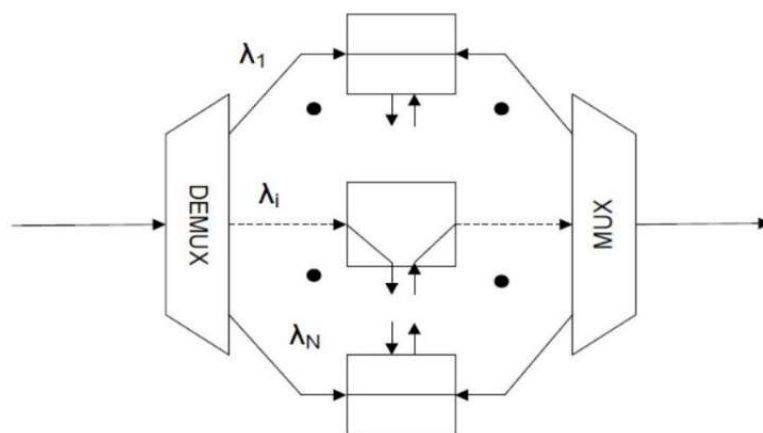
Όπως παρατηρούμε λοιπόν, τα οπτικά δίκτυα 1^{ης} χαρακτηρίζονται από την αδυναμία τους να εκμεταλλευτούν πλήρως τα πλεονεκτήματα που προσφέρει η χρήση της οπτικής ίνας. Η χαμηλή ταχύτητα λειτουργίας των ηλεκτρονικών στοιχείων των κόμβων περιορίζουν αισθητά την ταχύτητα μετάδοσης των δεδομένων (ηλεκτρονική συμφόρηση-electronic bottleneck), καθώς επίσης διαπιστώνουμε και αύξηση του ρυθμού εμφάνισης σφαλμάτων. Έτσι λοιπόν οδηγηθήκαμε στα οπτικά δίκτυα 2ης γενιάς, με στόχο να αντιμετωπιστούν τα παραπάνω προβλήματα.

1.3.2 Οπτικά Δίκτυα 2^{ης} γενιάς

Η 2^η γενιά οπτικών δικτύων σηματοδοτείται από τη χρήση οπτικών διατάξεων προσθήκης/αφαίρεσης μήκους κύματος (Wavelength Add-Drop Multiplexers - WADMs). Με τη χρήση αυτών των διατάξεων επιτυγχάνεται ο περιορισμός της οπτο-ηλεκτρονικής μετατροπής μόνο στα δεδομένα που ξεκινούν από έναν κόμβο καθώς προστίθενται σε ένα μήκος κύματος ή τερματίζουν σε έναν κόμβο και αφαιρούνται από αυτό το μήκος κύματος. Η τεχνολογία αυτή επιτρέπει την επιλεκτική επεξεργασία των

δεδομένων στους ενδιάμεσους κόμβους (που αποτελεί και το μεγαλύτερο μέρος της διαχειριζόμενης κίνησης, κυρίως σε τοπολογίες δακτυλίου) με αποτέλεσμα να αξιοποιείται καλύτερα η χωρητικότητα της οπτικής ίνας καθώς το overhead μειώνεται (το υψηλό overhead ήταν βασικό χαρακτηριστικό των οπτικών δικτύων πρώτης γενιάς).

Η χρήση οπτικών διατάξεων WADMs οδηγεί στην δημιουργία τοπολογιών δακτυλίου, οι οποίες τοπολογίες κατά κύριο λόγο στελεχώνουν τα μητροπολιτικά δίκτυα.



Εικόνα 7: Οπτικός πολυπλέκτης προσθήκης/αφαίρεσης (WADM)

Επιπρόσθετα, στη δεύτερη γενιά οπτικών δικτύων, ένα ακόμα χαρακτηριστικό των κόμβων είναι πως διαθέτουν και οπτικούς κατανεμητές (Optical Cross-Connectors - OXCs). Η λειτουργία των οπτικών κατανεμητών έγκειται στη μεταγωγή ενός μήκους κύματος από μία θύρα εισόδου τους σε άλλο μήκος κύματος σε μία από τις θύρες εξόδου τους. Επίσης διαθέτουν και διακόπτες μετατροπής μήκους κύματος (Wavelength Convertible Switches - WCSs), οι οποίοι δίνουν τη δυνατότητα μετατροπής ενός μήκους κύματος πριν από την είσοδό του στον πολυπλέκτη της ίνας εξόδου.

Η χρήση τεχνολογιών όπως οι οπτικοί κατανεμητές (OXCs) και οι διακόπτες μετατροπής μήκους κύματος (WADMs) συντέλεσαν σπουδαίο ρόλο στην εξέλιξη των οπτικών δικτύων δεύτερης γενιάς. Έτσι, οδηγηθήκαμε στην ανάπτυξη των οπτικών δικτύων δρομολόγησης μήκους κύματος ευρείας περιοχής (Wavelength Routed Wide Area Optical Networks). Στα δίκτυα αυτά για την σύνδεση δύο κόμβων του δικτύου χρησιμοποιούνται συνδεδιστρεφή (connection oriented) οπτικά μονοπάτια (lightpaths).

Αναλόγως τα χαρακτηριστικά της προσφερόμενης κίνησης, η ανάθεση των οπτικών μονοπατιών γίνεται με δύο μεθόδους, είτε με στατική είτε με δυναμική ανάθεση. Στη στατική ανάθεση το σύνολο των συνδέσεων είναι εξ αρχής γνωστό, ενώ στην δυναμική ανάθεση οι συνδέσεις διαμορφώνονται κατάλληλα κατά την άφιξη των κλήσεων. Το βασικό πρόβλημα που προκύπτει στην ανάθεση των οπτικών μονοπατιών, είτε αυτή γίνεται στατικά, είτε δυναμικά, είναι η εύρεση ενός μήκους κύματος κατά μήκος ενός

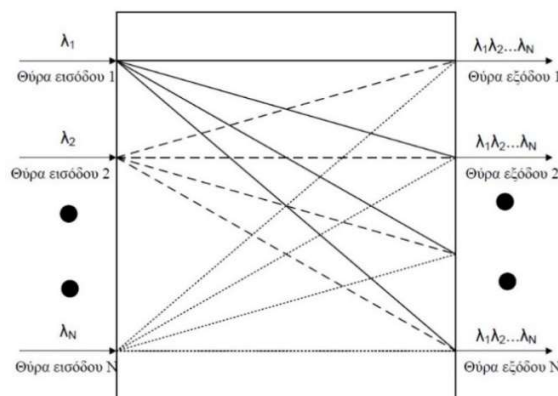
οπτικού μονοπατιού (Routing and Wavelength Assignment - RWA) λαμβάνοντας ως δεδομένη την κατάσταση του δικτύου.

1.3.3 Οπτικά Δίκτυα 3^{ης} γενιάς

Η 3^η γενιά οπτικών δικτύων αφορά ασυνδεδεστικά οπτικά δίκτυα (connectionless). Η πρόσβαση επιτυγχάνεται μέσω οπτικών δικτύων πρόσβασης, όπως τα παθητικά οπτικά δίκτυα για παράδειγμα (Passive Optical Networks - PON). Η οπτική μεταγωγή μπορεί να αφορά είτε πακέτα (Optical Packet Switching - OPS), είτε ριπές (Optical Burst Switching - OBS) είτε ετικέτες (Optical Label Switching - OLS).

Η εξέλιξη των οπτικών δικτύων σε αμιγώς οπτικά βασίζεται στη χρήση αποκλειστικά οπτικών διατάξεων/διακοπών (all optical devices). Στις διατάξεις αυτές, όλες οι λειτουργίες που σχετίζονται με την πληροφορία όπως η δρομολόγηση και η μεταγωγή, πραγματοποιούνται αποκλειστικά στο οπτικό επίπεδο, χωρίς παρουσία οπτο-ηλεκτρονικών μετατροπών πέρα από τους κόμβους παρυφής (edge nodes). Οι προαναφερθείσες διατάξεις κατατάσσονται σε τρεις βασικές κατηγορίες: τους παθητικούς δρομολογητές (passive routers), τους συζεύκτες παθητικού αστέρα (passive star couplers) και τους ενεργούς διακόπτες (active switchers).

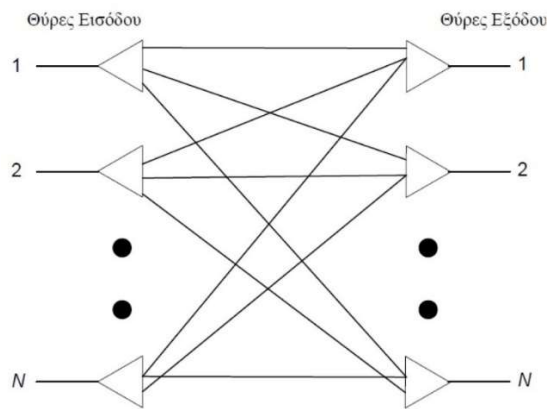
- Ο παθητικός αστέρας (Passive Star) είναι μία διάταξη ευρείας εκπομπής. Με την εγκατάστασή του, η ενέργεια ενός σήματος συγκεκριμένου μήκους κύματος λ το οποίο φτάνει σε μία θύρα εισόδου ισοκατανέμεται σε όλες τις θύρες εξόδου του χρησιμοποιώντας το ίδιο μήκος κύματος λ . Σε περίπτωση ταυτόχρονης μετάδοσης τουλάχιστον δύο σημάτων ίδιου μήκους κύματος στις θύρες εισόδου, δημιουργείται σύγκρουση (collision), εξέλιξη μη επιθυμητή.



Εικόνα 8: Παθητικός αστέρας NxN

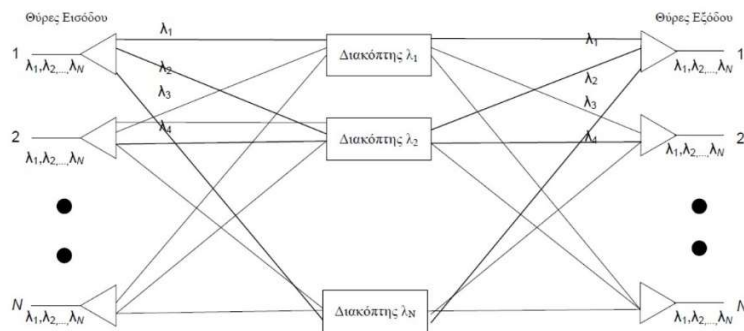
- Ο παθητικός δρομολογητής (Passive Router) είναι στην πραγματικότητα ένας στατικός δρομολογητής, καθώς δεν παρέχει την δυνατότητα επαναπρογραμματισμού του πίνακα δρομολόγησης για την μεταγωγή των σημάτων. Η λειτουργία του παθητικού δρομολογητή έγκειται στην

δρομολόγηση καθενός από τα μήκη κύματος που εμφανίζονται στις θύρες εισόδου του στα αντίστοιχα μήκη κύματος των θυρών εξόδου.



Εικόνα 9: Παθητικός δρομολογητής $N \times N$

- Ο ενεργός διακόπτης (Active Switcher) πραγματοποιεί την παραπάνω δρομολόγηση δυναμικά, καθώς προσαρμόζει το περιεχόμενο του πίνακα δρομολόγησης δυναμικά αναλόγως τις ανάγκες της κίνησης.



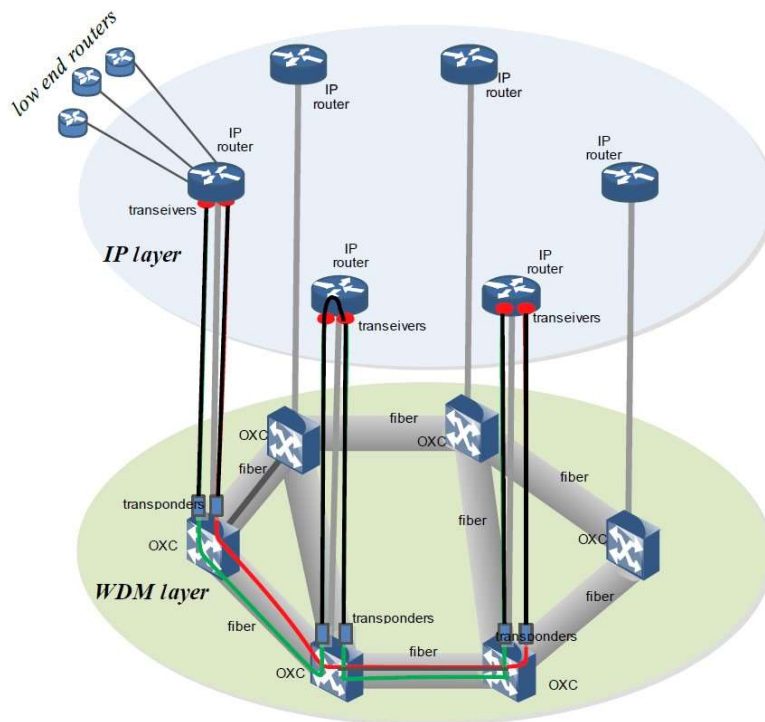
Εικόνα 10: Ενεργός διακόπτης $N \times N$

Αποτέλεσμα της χρήσης των παραπάνω διατάξεων που αναφέραμε είναι η πλήρης απουσία οπτο-ηλεκτρονικών μετατροπών της πληροφορίας που διακινείται στον πυρήνα του δικτύου. Άμεσο επακόλουθο της απουσίας των οπτο-ηλεκτρονικών μετατροπών είναι η πολύ υψηλή εκμετάλλευση του εύρους ζώνης της οπτικής ίνας καθώς η μείωση της κατανάλωσης ισχύος, παράγοντας ιδιαίτερα σημαντικός στα σύγχρονα επικοινωνιακά δίκτυα. Όπως γίνεται λοιπόν κατανοητό, τα αμιγώς οπτικά δίκτυα αποτελούν το μέλλον επικοινωνιακών δικτύων.

1.4 Αρχιτεκτονική WDM Οπτικών Δικτύων

Η μορφή των σύγχρονων αλλά και μελλοντικών WDM συστημάτων επικοινωνίας καθορίζεται σε μεγάλο βαθμό από τα διαθέσιμα οπτικά στοιχεία [6].

Ένα σύγχρονο οπτικό WDM δίκτυο, όπως αυτό που παρουσιάζεται στο παρακάτω σχήμα, αποτελείται από οπτικούς κόμβους μεταγωγής – OXC. Οι βασικές λειτουργίες των οπτικών κόμβων μεταγωγής είναι η προσθήκη/αφαίρεση κυκλωμάτων (add - drop) καθώς και η μεταγωγή των διερχόμενων οπτικών κυκλωμάτων. Όσον αφορά τα μικρότερα δίκτυα, όπως δίκτυα δακτυλίων τότε οι οπτικοί κόμβοι είναι τύπου ROADM. Το μέσο διασύνδεσης των οπτικών κόμβων είναι οι οπτικές ίνες. Για τη μετάδοση του σήματος στο οπτικό δίκτυο χρησιμοποιούνται οπτικοί . Στους οπτικούς κόμβους χρησιμοποιούνται οπτικοί ενισχυτές με στόχο να αντισταθμίσουν τις απώλειες καθώς επίσης και στις ίνες/συνδέσμους λειτουργώντας ως ενισχυτές γραμμής.

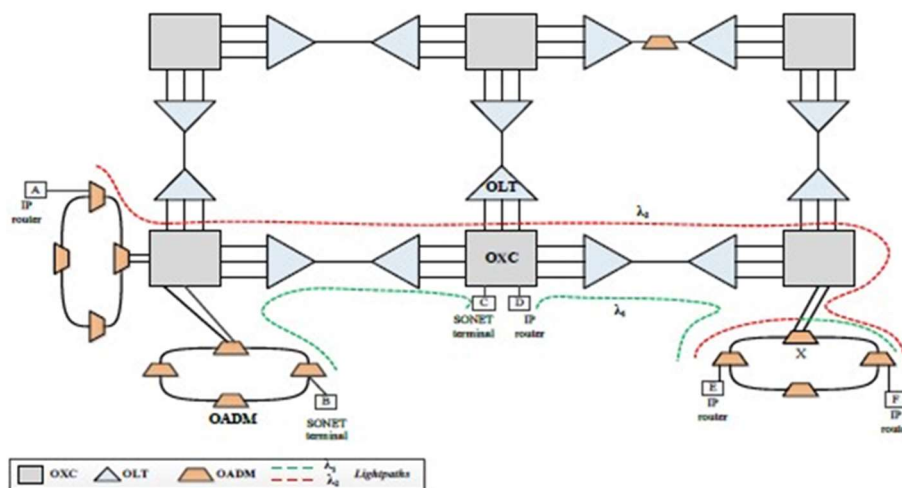


Εικόνα 11: Αρχιτεκτονική σύγχρονων οπτικών δικτύων

Όπως παρουσιάζεται στο παραπάνω σχήμα, το επίπεδο του WDM δικτύου ενώνεται με το επίπεδο του IP/TCP. Τα WDM δίκτυα συναντώνται στον κορμό (core/backbone) και σε μητροπολιτικά δίκτυα (metro) με στόχο να συνενώσουν τα IP υποδίκτυα που είναι κατανεμημένα γεωγραφικά και σε μεγάλες αποστάσεις μεταξύ τους. Στην πραγματικότητα οι οπτικοί κόμβοι δημιουργούν ένα WDM οπτικό δίκτυο στο οποίο ενώνονται οπτικοί κόμβοι (ROADM/OXC) όπου είναι συνδεδεμένοι ηλεκτρονικοί IP δρομολογητές (routers). Οι δρομολογητές αυτοί (IP routers) προωθούν τα πακέτα, τα οποία μετατρέπονται από οπτικούς πομποδέκτες σε οπτικό σήμα το οποίο μεταφέρεται μέσω οπτικών κυκλωμάτων πάνω από το WDM δίκτυο στο IP router του προορισμού.

Κατά το παρελθόν, η WDM τεχνολογία χρησιμοποιείτο μόνο για μεταδόσεις από σημείο σε σημείο, γεγονός το οποίο σήμαινε πως όλες οι συνδέσεις τερματίζονταν σε κάθε ενδιάμεσο κόμβο στο ηλεκτρονικό επίπεδο χωρίς να υπάρχει η δυνατότητα να περνούν τα πακέτα διαφανώς (transparent – bypass) από τους ενδιάμεσους κόμβους. Για το λόγο αυτό, στα σύγχρονα οπτικά δίκτυα, τα οπτικά πακέτα μπορεί είτε να μετατραπούν σε ηλεκτρικά σε κάποιους ενδιάμεσους κόμβους (πράσινη σύνδεση στο σχήμα) είτε να περάσουν διαφανώς (κόκκινη γραμμή στο σχήμα). Η επιλογή του τερματισμού ή μη, του σήματος ηλεκτρονικά στους ενδιάμεσους κόμβους, ανήκει στον κατασκευαστή του δικτύου, συνεισφέροντας άμεσα ή έμμεσα στο κόστος και την ενεργειακή κατανάλωση του δικτύου, το οποίο αποτελεί ιδιαίτερα σημαντικό παράγοντα στον σχεδιασμό των σύγχρονων οπτικών δικτύων.

Στο παρακάτω σχήμα παρουσιάζεται η σύνδεση του IP επιπέδου απευθείας με το WDM δίκτυο:



Εικόνα 12: Σύνδεση IP και WDM επιπέδου

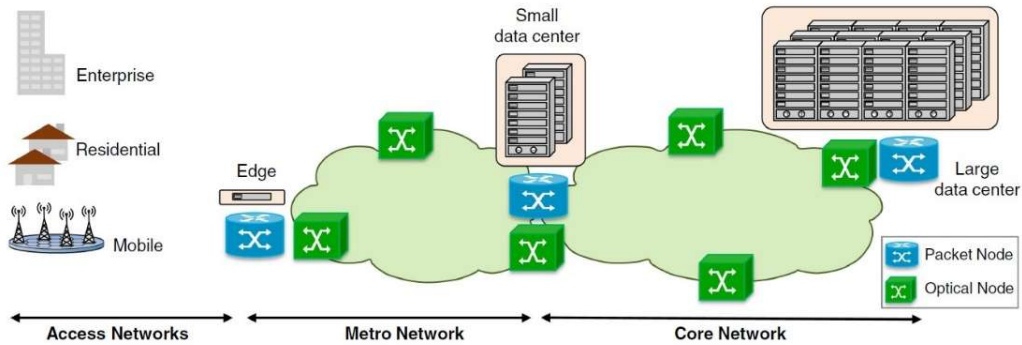
Μία ακόμη αρχιτεκτονική επιλογή στο σχεδιασμό των σύγχρονων οπτικών δικτύων, είναι η χρησιμοποίηση της τεχνολογίας SONET/SDH. Κατά την συγκεκριμένη επιλογή, θεωρείται ένα ενδιάμεσο επίπεδο μεταξύ των επιπέδων IP και WDM, στο οποίο χρησιμοποιούνται SONET/SDH μεταγωγείς οι οποίοι αντικαθιστούν τους IP δρομολογητές. Οι πομποδέκτες – αναμεταδότες δέχονται ως είσοδο το σήμα από τους IP routers, SONET/SDH switches ή όποιο άλλο στοιχείο χρησιμοποιούν τα τοπικά δίκτυα που ενώνονται στο WDM δίκτυο και μετατρέπουν αυτό το σήμα στην κατάλληλη μορφή για να περάσει μέσα στο WDM δίκτυο.

1.5 Ιεραρχική δομή Οπτικών Δικτύων

Τα οπτικά δίκτυα ιεραρχούνται ανάλογα με τη γεωγραφική επικράτεια στην οποία εκτείνονται. Οι κατηγορίες των δικτύων κατά την ιεραρχία αυτή είναι τα δίκτυα

μεγάλης απόστασης, τα δίκτυα ευρείας έκτασης, τα μητροπολιτικά δίκτυα και τα δίκτυα πρόσβασης [6].

- Τα οπτικά δίκτυα μεγάλης απόστασης (long haul networks) αποτελούν τα πρώτα δίκτυα που έγινε απαραίτητη η εφαρμογή τεχνολογίας οπτικών ινών. Στην ουσία πρόκειται για ηπειρωτικά και διηπειρωτικά δίκτυα και αφορούν ζεύξεις μεγάλων αποστάσεων (άνω των 10.000χλμ), δηλαδή ζεύξεις μεταξύ ωκεανών, ηπείρων και χωρών. Η αναγκαιότητα των δικτύων για χρήση οπτικών ινών έγκειται στις απαιτήσεις των ζεύξεων αυτών για μεγάλο εύρος ζώνης κατά την μετάδοση, η οποία πρέπει να πραγματοποιηθεί χωρίς πολλές ενισχύσεις κατά μήκος της. Η τεχνολογία που εξυπηρετεί τις απαιτήσεις αυτές είναι η πυκνή πολυπλεξία με διαίρεση μήκους κύματος (Dense Wavelength Division Multiplexing - DWDM). Τα δίκτυα αυτά αποτελούνται κατά κύριο λόγο από δακτυλίους (rings) οι οποίοι διαθέτουν τεχνολογία SONET/SDH (Synchronous Optical Networking/ Synchronous Digital Hierarchy). Επίσης, χρησιμοποιούν τεχνολογία πολυπλεξίας μήκους κύματος (WDM) και οπτικής μεταγωγής μήκους κύματος (Wavelength routed). Τέλος, διαθέτουν αμιγώς οπτικούς μεταγωγείς (optical cross connect).
- Τα δίκτυα ευρείας έκτασης (Wide Area Networks - WAN) που ονομάζονται επίσης και δίκτυα κορμού (backbone/core networks) αφορούν ζεύξεις στα όρια της εθνικής επικράτειας και εξυπηρετούν συνδέσεις μεταξύ πόλεων καθώς και βασικών κόμβων (παραδείγματος χάρη τα κέντρα δεδομένων). Οι αποστάσεις των ζεύξεων αυτών κυμαίνονται μεταξύ 200-10.000χλμ. Τα δίκτυα κορμού χρησιμοποιούν τεχνολογίες πολυπλεξίας μήκους κύματος (Wavelength Division Multiplexing - WDM) και οπτικής μεταγωγής μήκους κύματος (Wavelength routed). Επίσης, χρησιμοποιούν αποκλειστικά αμιγώς οπτικούς μεταγωγείς (optical cross connect).
- Τα μητροπολιτικά δίκτυα (Metropolitan Area Networks – MAN) αφορούν ζεύξεις μεταξύ πόλεων ή περιφερειών καθώς καλύπτουν αποστάσεις (αποστάσεις μεταξύ κόμβων) μικρότερες των 200χλμ. Τα μητροπολιτικά δίκτυα είναι συνήθως δακτύλιοι και βασίζονται στην πολυπλεξία μήκους κύματος. Χρησιμοποιούν τεχνολογίες μετάδοσης από-σημείο-σε-σημείο (point-to-point – P2P) ή οπτική μεταγωγή μήκους κύματος (Wavelength routed).
- Τα δίκτυα πρόσβασης (Access Networks) αποτελούν τα δίκτυα σύνδεσης με τους τελικούς χρήστες, δηλαδή επιχειρήσεις, οργανισμούς και απλούς χρήστες, καθώς καλύπτουν αποστάσεις μεταξύ 100 μέτρων έως 10χλμ. Αποτελούνται από συνδέσεις από σημείο σε πολλαπλά σημεία και θεωρούνται παθητικά οπτικά δίκτυα (Passive Optical Networks - PON) καθώς δεν χρησιμοποιούν ενεργά στοιχεία μεταγωγής ούτε ενισχυτές παρά μόνο διαχωριστές (splitters) και πομποδέκτες (transceivers). Επίσης, λόγω αυτής της αρχιτεκτονικής, έχουν χαμηλό κόστος και κατανάλωση ενέργειας.



Εικόνα 13: Ιεραρχική δομή οπτικών δικτύων

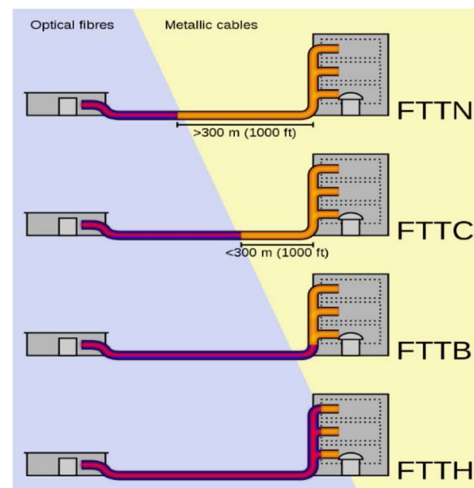
1.6 Κατηγοριοποίηση Οπτικών Δικτύων

Υπάρχουν τρεις βασικοί τρόποι κατηγοριοποίησης των οπτικών δικτύων [10] [13] [6]:

- 1) Ως προς τον προορισμό της οπτικής σύνδεσης
- 2) Ως προς την ζεύξη
- 3) Ως προς τις τεχνολογίες πρόσβασης

1.6.1 Κατηγοριοποίηση ως προς τον προορισμό της οπτικής σύνδεσης

Το Fiber to the x (FTTx) είναι ένας γενικός όρος για οποιαδήποτε αρχιτεκτονική ευρυζωνικού δικτύου που χρησιμοποιεί οπτικές ίνες για να παρέχει όλο ή μέρος του τοπικού βρόχου που χρησιμοποιείται στα οπτικά δίκτυα. Καθώς τα καλώδια οπτικών ινών είναι σε θέση να μεταφέρουν πολύ περισσότερα δεδομένα και σε υψηλότερες ταχύτητες από τα καλώδια χαλκού, ειδικά σε μεγάλες αποστάσεις, έχουμε κατηγοριοποίηση των αρχιτεκτονικών ανάλογα το που τερματίζει το οπτικό μέσο [14]:



Εικόνα 14: Αρχιτεκτονική FTTx

Επομένως έχουμε τις παρακάτω αρχιτεκτονικές:

Fiber To The Home – FTTH

Ως Fiber To The Home (οπτική ίνα ως το σπίτι) ορίζεται η τηλεπικοινωνιακή αρχιτεκτονική κατά την οποία μία επικοινωνιακή οδός εκτείνεται με χρήση αποκλειστικά οπτικού μέσου από τον εξοπλισμό μεταγωγής ενός δικτυακού παρόχου έως το χώρο διαμονής κάθε χρήστη. Η οπτική ίνα τερματίζεται εντός του χώρου διαμονής του χρήστη.

Fiber To The building – FTTB

Ως Fiber To The Building (οπτική ίνα ως το κτήριο) ορίζεται η τηλεπικοινωνιακή αρχιτεκτονική κατά την οποία μία επικοινωνιακή οδός εκτείνεται με χρήση αποκλειστικά οπτικού μέσου από τον εξοπλισμό μεταγωγής ενός δικτυακού παρόχου μέχρι το κτήριο (συνήθως μέχρι τον κατανομητή) στο οποίο βρίσκεται ο χώρος του τελικού χρήστη, αλλά δε φτάνει μέχρι το χώρο του. Για την υπολειπόμενη απόσταση μέχρι το σπίτι ή το γραφείο του χρήστη, χρησιμοποιούνται είτε οι κοινές τηλεφωνικές γραμμές (ομοαξονικό καλώδιο/συνεστραμμένα ζεύγη χαλκού) είτε ασύρματη ζεύξη.

Όμοια με το Fiber To The Home και το Fiber To The building ορίζονται το Fiber To The Office – FTTO και το Fiber To The Business – FTTB, με τη διαφορά ότι αναφερόμαστε σε γραφείο/εταιρικό πελάτη και σε εταιρία αντίστοιχα.

Fiber To The Curb/Cabinet – FTTC

Ως Fiber To The Curb/Cabinet (οπτική ίνα μέχρι το κουτί του παρόχου) ορίζεται η τηλεπικοινωνιακή αρχιτεκτονική κατά την οποία μία επικοινωνιακή οδός εκτείνεται με χρήση αποκλειστικά οπτικού μέσου από τον εξοπλισμό μεταγωγής ενός δικτυακού παρόχου μέχρι τον τοπικό διανομητή («καφάο»). Για την υπολειπόμενη απόσταση μέχρι το σπίτι του τελικού χρήστη, χρησιμοποιούνται συνήθως κοινά καλώδια χαλκού. Το FTTC μπορεί να εξυπηρετήσει μερικούς δεκάδες πελάτες σε ακτίνα μικρότερη από 300 μέτρα από το καφάο.

Fiber To The Node – FTTN

Ως Fiber To The Node (οπτική ίνα μέχρι τον κόμβο) ορίζεται η τηλεπικοινωνιακή αρχιτεκτονική κατά την οποία μία επικοινωνιακή οδός εκτείνεται με χρήση αποκλειστικά οπτικού μέσου από τον εξοπλισμό μεταγωγής ενός δικτυακού παρόχου μέχρι ένα κοινό network box, που λέγεται κόμβος (node). Για την υπολειπόμενη απόσταση μέχρι το σπίτι του τελικού χρήστη, χρησιμοποιούνται συνήθως οι κοινές τηλεφωνικές γραμμές. Το FTTN μπορεί να εξυπηρετήσει μερικούς εκατοντάδες πελάτες σε ακτίνα 300 μέτρων από τον κόμβο.

Fiber To The Premises – FTTP

Ο ορισμός Fiber To The Premises (οπτική ίνα μέχρι τις εγκαταστάσεις) μπορεί να χρησιμοποιηθεί είτε σαν γενικός όρος για τα FTTN και τα FTTB δίκτυα είτε οπτικά δίκτυα που περιλαμβάνουν και σπίτια καθώς και μικρές επιχειρήσεις.

1.6.2 Κατηγοριοποίηση ως προς την ζεύξη

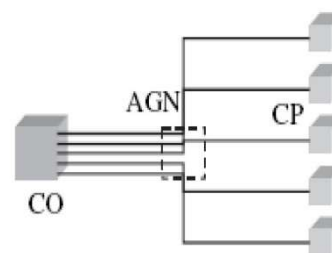
Η τοπολογία καθορίζει το αν η κάθε οπτική ίνα στη ζεύξη που ξεκινάει από το αστικό κέντρο αντιστοιχεί σε έναν συνδρομητή ή σε περισσότερους. Έτσι, διακρίνουμε δύο βασικές τοπολογίες οπτικών δικτύων:

- 1) Point-to-Point (P2P)
- 2) Point-to-Multipoint (PMP)

Point-to-Point (P2P) ζεύξη

Στην τοπολογία αυτή, κάθε οπτική ίνα αντιστοιχεί σε αποκλειστικά έναν χρήστη.

Η υποδομή παρέχει στο φυσικό επίπεδο, αστεροειδώς, διακριτές οπτικές οδούς από το απομακρυσμένο σημείο παρουσίας του παρόχου προς κάθε υποστηριζόμενο από το σημείο αυτό, χρήστη. Είναι σαφές ότι η «διακριτότητα» της κάθε σύνδεσης όσο απομακρυνόμαστε από τον παραπάνω κόμβο προς το εσωτερικό του δικτύου, μπορεί να εξασφαλίζεται όχι στο φυσικό επίπεδο αλλά σε ανώτερα επίπεδα.

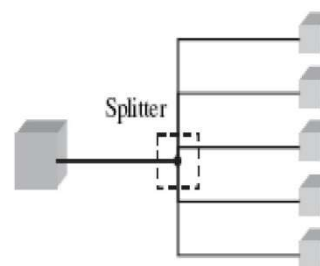


Εικόνα 15: Σύνδεση P2P

Γίνεται κατανοητό πως η P2P ζεύξη αποτελεί μία ακριβή λύση, καθώς απαιτείται μεγαλύτερη υποδομή για την υλοποίησή της. Τέλος, αξίζει να σημειωθεί πως χρησιμοποιείται και για παροχή μισθωμένων υπηρεσιών καθώς προσφέρει μεγαλύτερη ασφάλεια ως λύση [15].

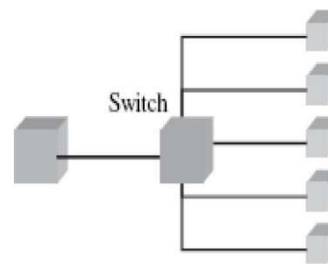
Point-to-Multipoint (PMP) ζεύξη

Στην τοπολογία αυτή, η οπτική ίνα συναντάει κάποιον αριθμό από splitters (συνήθως 4 ή 5), στους οποίους η οπτική ισχύς μειώνεται κατά το ήμισυ, έτσι ώστε μετά από 4 ή 5 στάδια splitting να εξυπηρετούνται 16 ή 32 συνδρομητές αντίστοιχα.



Εικόνα 16: Σύνδεση PMP

Στην περίπτωση αυτή, η οπτική υποδομή αναπτύσσεται με τρόπο ώστε να δίνει τη δυνατότητα δενδροειδούς διακλάδωσης οπτικών μονοπατιών, χωρίς ενδιάμεσους κόμβους, μέσω διακλαδωτήρων από το σημείο παρουσίας του παρόχου, προς διαδοχικές ομάδες γειτνιαζόντων χρηστών. Οι οπτικές οδοί από τον πάροχο προς τον διακλαδωτή, εξυπηρετούν την κίνηση από και προς πολλούς χρήστες, η οποία μεταφέρεται αυτούσια σε όλες τις ακτίνες του αστέρα. Ουσιαστικά οι χρήστες στα άκρα του αστέρα μοιράζονται το εύρος ζώνης μιας οπτικής οδού με επιπλέον απώλειες και αυξημένη πολυπλοκότητα για την απομόνωση των κινήσεων των χρηστών.



Εικόνα 17: Σύνδεση PMP

Η PMP ζεύξη αποτελεί μία πιο οικονομική λύση και για τον λόγο αυτό χρησιμοποιείται σε ευρεία κλίμακα.

1.6.3 Κατηγοριοποίηση ως προς τις τεχνολογίες πρόσβασης

Η τεχνολογία πρόσβασης εξαρτάται από την ύπαρξη ή μη ενεργών οπτικών στοιχείων έξω από το αστικό κέντρο.

Ένα οπτικό στοιχείο το οποίο καταναλώνει ρεύμα (άρα θέλει αδιάλειπτη παροχή ρεύματος και γεννήτρια ή μπαταρία για εφεδρεία) ονομάζεται ενεργό οπτικό στοιχείο. Αντιθέτως, παθητικό οπτικό στοιχείο ονομάζεται το οπτικό στοιχείο το οποίο δεν καταναλώνει ρεύμα [16].

Έτσι λοιπόν διακρίνουμε τρεις βασικές κατηγορίες οπτικών δικτύων:

- 1) Παθητικά Οπτικά Δίκτυα (Passive Optical Networks-PONs)
- 2) Ενεργά Οπτικά Δίκτυα (Active Optical Networks-AONs)
- 3) Υβριδικά Οπτικά Δίκτυα

Παθητικά Οπτικά Δίκτυα – PON

Τα Παθητικά Οπτικά Δίκτυα (Passive Optical Networks – PONs) είναι οπτικά δίκτυα βασισμένα στην αρχιτεκτονική σημείου προς πολλαπλά σημεία (PMP). Τα δίκτυα αυτά δεν περιέχουν ενεργά στοιχεία, δηλαδή δεν πραγματοποιείται οπτο-ηλεκτρονική μετατροπή σήματος από την πηγή μέχρι το προορισμό. Τα μόνα στοιχεία που χρησιμοποιούνται είναι παθητικοί ζεύκτες (couplers), διαιρέτες (splitters) και συνδυαστές (combiners).

Αποτελούν μια αξιόπιστη λύση για τα Δίκτυα Πρόσβασης (Access Networks – AN) καθώς δίνουν τη δυνατότητα χρήσης υπηρεσιών ευρείας ζώνης με οικονομικούς όρους. Αυτή η δυνατότητα καθιστά εφικτή την παροχή της συγκεκριμένης τεχνολογίας σε μεμονωμένους χρήστες ή μικρές επιχειρήσεις οι οποίοι δεν έχουν τη οικονομική δυνατότητα χρησιμοποίησης οπτικών ινών με απευθείας ζεύξη (P2P).

Ενεργά Οπτικά Δίκτυα – AON

Η κύρια διαφορά μεταξύ των Ενεργών και των Παθητικών Οπτικών Δικτύων είναι η αντικατάσταση του παθητικού διαχωριστή (splitter) από έναν ενεργό κόμβο. Συνέπεια αυτού είναι η απαραίτητη χρήση ενός ηλεκτροφόρου καλωδίου (power line) μεταξύ του κέντρου και του ενεργού κόμβου.

Εν αντιθέσει με τα παθητικά οπτικά δίκτυα που χρησιμοποιούν κατά βάση αρχιτεκτονική δένδρου, τα ενεργά οπτικά δίκτυα έχουν τη δυνατότητα να υλοποιηθούν και με αρχιτεκτονική δακτυλίου ή αστέρα. Η αρχιτεκτονική επιλογή για την υλοποίηση του δικτύου, εξαρτάται συνήθως από τον τύπο και το κόστος υλοποίησης, καθώς και από την τοπολογία και την διαθεσιμότητα της οπτικής ίνας. Οι προαναφερθείσες αρχιτεκτονικές, θα αναλυθούν σε επόμενη ενότητα στο κεφάλαιο αυτό.

Επιπρόσθετα, σε κάθε τελικό χρήστη προσφέρεται σύνδεση που του παρέχει το συνολικό αμφίδρομο εύρος ζώνης, αντί το εύρος ζώνης να διανέμεται μεταξύ πολλαπλών συνδρομητών. Η υπηρεσία αυτή παρέχεται εφαρμόζοντας τεχνικές SDM (Space Division Multiplexing) ή WDM. Όπως γίνεται κατανοητό, η αρχιτεκτονική αυτή εμπίπτει σε μία αρχιτεκτονική ζεύξης από-σημείο-σε-σημείο (P2P).

Τέλος, μία ακόμα σημαντική διαφορά μεταξύ των ενεργών και των παθητικών οπτικών δικτύων είναι ο περιορισμός της απόστασης. Ένα Ενεργό Οπτικό Δίκτυο έχει έναν περιορισμό απόστασης περίπου 80 χιλιομέτρων, ανεξάρτητα από τον αριθμό συνδρομητών που εξυπηρετεί. Εν αντιθέσει με ένα παθητικό οπτικό δίκτυο, στο οποίο ο πιο απομακρυσμένος συνδρομητής πρέπει να βρίσκεται σε μια ακτίνα 10-20 χιλιομέτρων από τον κεντρικό σταθμό, ανάλογα με το συνολικό αριθμό των διαμοιράσεων. Ο αριθμός συνδρομητών σε ένα ενεργό οπτικό δίκτυο περιορίζεται μόνο από τους κόμβους μεταγωγής (switches) που χρησιμοποιούνται και όχι από την ίδια την υποδομή, όπως στην περίπτωση παθητικών οπτικών δικτύων.

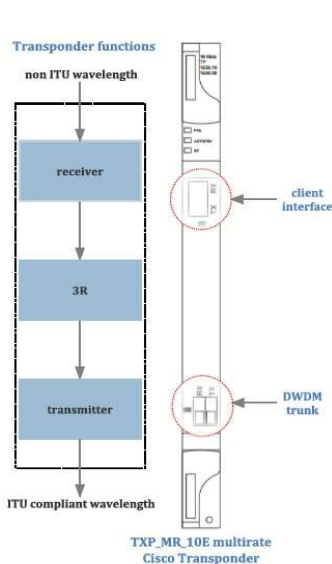
Υβριδικά Οπτικά Δίκτυα

Μια τρίτη κατηγορία οπτικών δικτύων που συναντάμε, είναι τα υβριδικά οπτικά δίκτυα. Τα υβριδικά οπτικά δίκτυα στην ουσία είναι ένας συνδυασμός ενός ενεργού κόμβου και μίας αρχιτεκτονικής παθητικού οπτικού δικτύου.

Σύμφωνα με αυτή την αρχιτεκτονική επιλογή, δύναται να επιτύχουμε μεγαλύτερη απόσταση μετάδοσης της υπηρεσίας απ' ότι σε μία αρχιτεκτονική παθητικών οπτικών δικτύων, χρησιμοποιώντας όμως λιγότερη και πιο απλή υποδομή σε σχέση με μία τοπολογία ενεργού οπτικού δικτύου, με ο,τι αυτό συνεπάγεται από άποψη κόστους και εξοπλισμού.

1.7 Στοιχεία Οπτικών Δικτύων

Στο σημείο αυτό θα αναλυθούν εκτενέστερα τα βασικά οπτικά στοιχεία του WDM επιπέδου [13] [6]:



Εικόνα 18: Transponder

Πομποδέκτες (Transpoders – TSP)

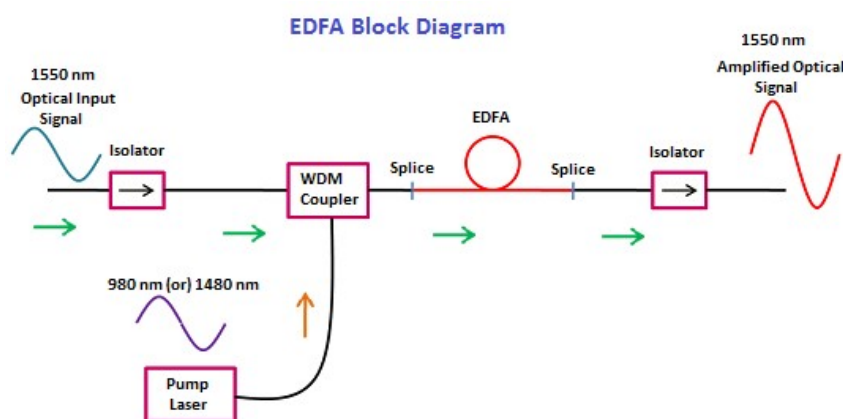
Η λειτουργία ενός πομποδέκτη έγκειται στην μετατροπή ενός σήματος εισόδου σε κατάλληλη για μετάδοση μορφή πάνω σε ένα WDM δίκτυο. Πάνω σε έναν πομποδέκτη είναι ενσωματωμένες τρεις λειτουργικές μονάδες. Από την πλευρά του χρήστη συναντάμε έναν πομπό, εν συνεχεία έναν αναγεννητή και έναν WDM πομποδέκτη. Έχει τη δυνατότητα υποστήριξης πρωτοκόλλων εισόδου όπως Ethernet, SONET/SDH, point-to-point Fiber channel αναλόγως τον τύπο του, ή και συνδυασμό αυτών. Στα WDM δίκτυα κορμού συναντάμε κατά κόρον Long-Reach Transponders (LR-TSPs) λειτουργώντας ως μετατροπείς του σήματος εισόδου σε σήμα WDM. Στα μητροπολιτικά δίκτυα συναντάμε συνήθως Short-Reach Transponders (SR-TSPs).

Οπτικοί Ενισχυτές – Optical Amplifiers

Τα κύρια χαρακτηριστικά ενός οπτικού δικτύου είναι η εξασθένηση του σήματος κατά την μετάδοσή του στην οπτική ίνα, η διασπορά των εκπεμπόμενων παλμών και φυσικά η απώλεια ισχύος. Για την αντιμετώπιση αυτού του φυσικού φαινομένου χρησιμοποιούμε οπτικούς ενισχυτές ανά διαστήματα στην οπτική ίνα, με στόχο τα οπτικά αυτά σήματα να ενισχυθούν και να διανύσουν μεγαλύτερη απόσταση. Το βασικό πλεονέκτημα τους έγκειται στην αντιμετώπιση της εξασθένησης σε μία ολόκληρη φασματική περιοχή, καθώς την ίδια στιγμή ενισχύονται αρκετά κανάλια. Επίσης, χαρακτηριστικό των οπτικών ενισχυτών είναι πως δεν επηρεάζονται από τη διαμόρφωση του σήματος καθώς και από το ρυθμό μετάδοσης.

Οι οπτικοί ενισχυτές ήρθαν να αντικαταστήσουν τους αναγεννητές. Οι αναγεννητές ήταν λειτουργικές μονάδες που ως βασική λειτουργία είχαν την μετατροπή του οπτικού σήματος σε ηλεκτρικό, να το φιλτράρουν, να το μετατρέψουν πάλι σε οπτικό σήμα και να το αναμεταδώσουν. Παρά το γεγονός πως με τη χρήση αναγεννητών επιτυγχανόταν ο περιορισμός των ατελειών μετάδοσης, όπως ο θόρυβος, η διασπορά και τα μη γραμμικά φαινόμενα, δεν είχαμε πλήρη εκμετάλλευση του εύρους ζώνης της οπτικής ίνας.

Στα σύγχρονα οπτικά δίκτυα, οι ενισχυτές που κατά κύριο λόγο χρησιμοποιούνται είναι οι ενισχυτές εμποτισμένοι με ιόντα ερβίου (Erbium Doped Fiber Amplifier - EDFA). Το έρβιο (Er) είναι χημικό στοιχείο της σειράς των λανθανίδων και προσφέρει κέρδος στη μήκη κύματος μεταξύ 1525nm και 1560nm και το οποίο μπορεί να ενισχύσει ένα οπτικό σήμα. Οι EDFA ενισχυτές είναι οπτικές ίνες εμποτισμένες με έρβιο. Στο παρακάτω διάγραμμα παρουσιάζεται η διάταξη ενός EDFA ενισχυτή:



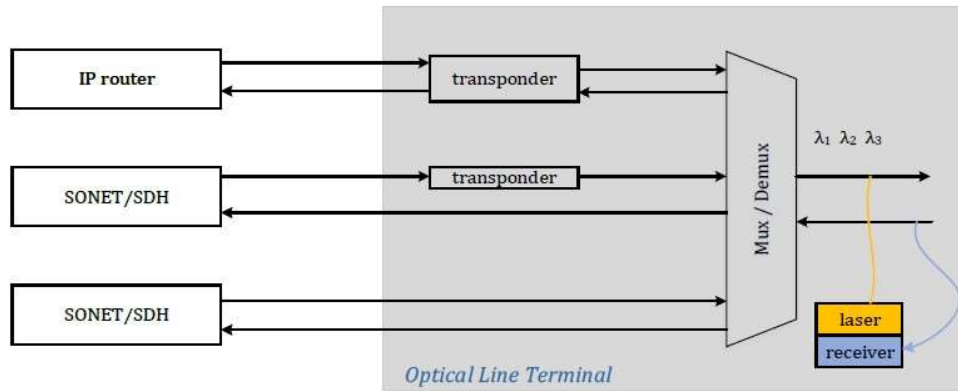
Εικόνα 19: Διάταξη EDFA

Κατά τη λειτουργία ενός EDFA ενισχυτή το λέιζερ άντλησης, το οποίο έχει μήκος κύματος 980nm ή 1480nm, πολυπλέκεται με το προς ενίσχυση οπτικό σήμα, το οποίο έχει μήκος κύματος 1550 nm στην εμποτισμένη ίνα. Τα εμποτισμένα άτομα διεγείρονται από το σήμα άντλησης σε υψηλότερο ενεργειακό επίπεδο, γεγονός που προκαλεί την απελευθέρωση φωτονίων.

Τέλος, αξίζει να αναφερθεί πως στα σύγχρονα οπτικά δίκτυα χρησιμοποιούνται κι άλλοι τύποι ενισχυτών όπως ο οπτικός ενισχυτής ημιαγωγού (Semiconductor Optical Amplifier - SOA) καθώς και ο ενισχυτής Raman, ο οποίος έχει πολύ καλά χαρακτηριστικά θορύβου και για το λόγο αυτό χρησιμοποιείται σε συστήματα μεγάλων αποστάσεων (long-haul systems).

Τερματικά Οπτικής Γραμμής (Optical Line Terminals – OLT)

Τα τερματικά οπτικής γραμμής είναι συσκευές που χρησιμοποιούνται στα άκρα μίας point to point σύνδεσης ενός παθητικού οπτικού δικτύου. Διαθέτουν δύο κύριες υπηρεσίες, τη μετατροπή σήματος μεταξύ του ηλεκτρικού που χρησιμοποιείται στον εξοπλισμό του παρόχου υπηρεσιών και του οπτικού που χρησιμοποιεί το οπτικό δίκτυο, καθώς και το συντονισμό της πολυπλεξίας μεταξύ των συσκευών μετατροπής στο άλλο άκρο του δικτύου (Optical Network Unit - ONU). Σε περίπτωση που το δίκτυο είναι αδιαφανές, συνήθως αποτελούνται από transponders, πολυπλέκτες, ενδεχομένως και από ενισχυτές. Στο παρακάτω διάγραμμα παρουσιάζεται μία διάταξη τερματικού οπτικής γραμμής:

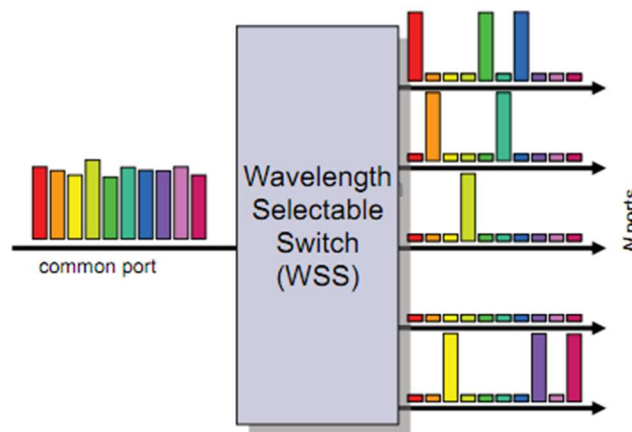


Εικόνα 20: διάταξη τερματικού οπτικής γραμμής

Επιλεκτικοί Μεταγωγείς Μήκους Κύματος (Wavelength Selective Switch - WSS)

Οι επιλεκτικοί μεταγωγείς μήκους κύματος (WSS) αποτελούν τη βάση των οπτικών πολυπλεκτών προσθήκης/αφαίρεσης κίνησης (OADM).

Ένας WSS μεταγωγέας μπορεί δυναμικά να δρομολογήσει, να μπλοκάρει και να εξασθενίσει όλα τα μήκη κύματος μέσα σε έναν κόμβο δικτύου. Στο παρακάτω σχήμα παρουσιάζεται η λειτουργία ενός WSS:



Εικόνα 21: Λειτουργία WSS

Αποτελείται από μία κοινή οπτική θύρα εισόδου και N θύρες εξόδου πολλαπλών μηκών κύματος, όπου κάθε μήκος κύματος μπορεί να δρομολογηθεί σε οποιαδήποτε από τις N θύρες εξόδου, ανεξαρτήτως του πως δρομολογούνται τα άλλα κανάλια μήκους κύματος. Αυτή η διαδικασία αλλαγής μήκους κύματος (δρομολόγησης) μπορεί να αλλάξει δυναμικά μέσω μιας διεπαφής ελέγχου ηλεκτρονικής επικοινωνίας στο WSS. Στην πραγματικότητα, το WSS αλλάζει κανάλια WDM ή μήκη κύματος. Επιπρόσθετα, στο WSS υπάρχουν μεταβλητοί μηχανισμοί εξασθένησης για κάθε μήκος κύματος.

Επομένως, κάθε μήκος κύματος μπορεί να διαμοιραστεί ανεξάρτητα με αντίκτυπο στον έλεγχο ισχύος και την εξισορρόπηση της.

Οπτικοί Πολυπλέκτες Προσθήκης/Αφαίρεσης Κίνησης (Optical Add/Drop Multiplexers - OADM)

Οι οπτικοί πολυπλέκτες προσθήκης/αφαίρεσης κίνησης αποτελούν την βάση μητροπολιτικών δικτύων καθώς και των δικτύων ευρείας κλίμακας. Η χρήση τους έγκειται στην δρομολόγηση και διαχείριση της κίνησης στα οπτικά δίκτυα. Αυτό επιτυγχάνεται καθώς μια OADM διάταξη παρέχει τη δυνατότητα σε ένα μήκος κύματος να ανέβει ή να κατέβει στη γραμμή ζεύξης χωρίς να επηρεάσει τα υπόλοιπα μήκη κύματος. Το βασικό πλεονέκτημα της χρήσης OADMs έγκειται στην δυνατότητα αντικατάστασης ενδιάμεσων τερματικών οπτικής γραμμής με οπτικούς πολυπλέκτες προσθήκης/αφαίρεσης κίνησης OADMs. Στον κόμβο αυτό αποφασίζεται ποια οπτικά σήματα θα γίνουν add/drop χωρίς όπως ήδη αναφέρθηκε να επηρεαστούν τα υπόλοιπα σήματα, καθώς επίσης να πραγματοποιείται κάποια μετατροπή σήματος. Αντίκτυπος αυτού είναι η μικρότερη χρησιμοποίηση πομποδεκτών στους ενδιάμεσους κόμβους, άρα κατ' επέκταση έχουμε και μία πιο οικονομική διάταξη.

Κύρια χαρακτηριστικά των OADMs είναι το πλήθος των μηκών κύματος που υποστηρίζονται και η επεκτασιμότητα τους, το ποσοστό των μηκών κύματος που μπορεί να γίνουν add/drop καθώς και οι περιορισμοί σε αυτούς, η επίδραση των λειτουργιών add/drop στο φυσικό επίπεδο όπως επίσης και η δυναμική αρχιτεκτονική που παρέχει τη δυνατότητα απομακρυσμένης διαχείρισης.

Οι οπτικοί πολυπλέκτες προσθήκης/αφαίρεσης κίνησης χωρίζονται σε δύο κατηγορίες, τους στατικούς και τους συντονισμένους.

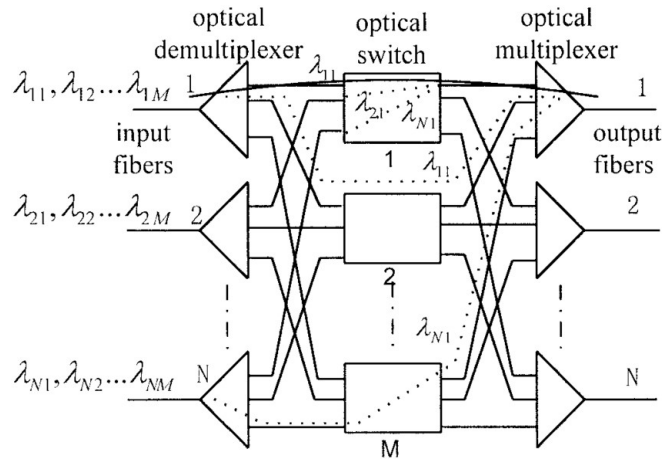
Η ιδιαιτερότητα των στατικών OADMs είναι η εκ των προτέρων ρύθμιση για την λειτουργία της προσθήκης/αφαίρεσης μηκών κύματος στο οπτικό δίκτυο εν αντιθέσει με τους συντονισμένους οπτικούς πολυπλέκτες προσθήκης/αφαίρεσης κίνησης (ROADMs), οι οποίοι παρέχουν την δυνατότητα δυναμικής διαχείρισης της κίνησης.

Συναντάμε τέσσερις βασικές αρχιτεκτονικές OADMs, είτε στατικών είτε ρυθμιζόμενων, την παράλληλη, την σπονδυλωτή (modular), την σειριακή και την τερματισμού ζώνης (band-drop).

Οπτικοί Μεταγωγείς (Optical Crossconnects – OXC)

Οι οπτικοί μεταγωγείς OXC είναι οπτικές μονάδες που διαθέτουν πολλές θύρες εισόδου οπτικών ινών καθώς επίσης και πολλές εξόδους. Κάθε οπτική ίνα πιθανόν να μεταφέρει περισσότερα από ένα μήκος κύματος. Θεωρώντας N θύρες εισόδου και M θύρες εξόδου, οι οπτικοί μεταγωγείς είναι διατάξεις $N \times M$. Βασικό πλεονέκτημα αυτής της διάταξης είναι η δυνατότητα προγραμματισμού και δρομολόγησης οποιουδήποτε μήκους κύματος εισόδου σε οποιαδήποτε θύρα εξόδου. Επίσης διαφοροποιούνται αναλόγως με το αν η μεταγωγή γίνεται οπτικά ή ηλεκτρονικά. Η οπτική μονάδα αυτή βρίσκει εφαρμογή σε πολύπλοκες δικτυακές τοπολογίες, όπως η τοπολογία δένδρου, καθώς επίσης και σε κόμβους που διαχειρίζονται μεγάλο όγκο πληροφορίας.

Παράδειγμα λειτουργίας ενός OXC είναι η διαχείριση κίνησης σε έναν κόμβο με παράλληλη συλλογή και τοπικό τερματισμό από και προς SDH τερματικά (terminals), IP δρομολογητές (routers) και μεταγωγείς ATM (switches).



Εικόνα 22: Διάγραμμα OXC

Οι σύγχρονοι οπτικοί μεταγωγείς OXC υλοποιούνται με διαφανή τρόπο (transparent), εν αντιθέσει με τις πρώτες γενιές οπτικών μεταγωγών που υλοποιούνταν με έναν αδιαφανή τρόπο (opaque).

Αμιγώς Οπτικές Διασυνδέσεις (All-optical OXCs)

Σε μία μονάδα αμιγούς οπτικής διασύνδεσης, το οπτικό σήμα φτάνει στην μονάδα μέσω διαφορετικών οπτικών ινών και αποπολυπλέκεται στα τερματικά οπτικής γραμμής. Στην διάταξη αυτή παρατηρούμε πως σε κάθε μεταγωγέα έχει ανατεθεί ένα συγκεκριμένο μήκος κύματος. Συνεπώς, το αποπολυπλεγμένο σήμα οδηγείται στον μεταγωγέα στον οποίο αντιστοιχεί το μήκος κύματος του και εν συνεχεία φτάνει στο τερματικό οπτικής γραμμής εισόδου όπου και πολυπλέκεται.

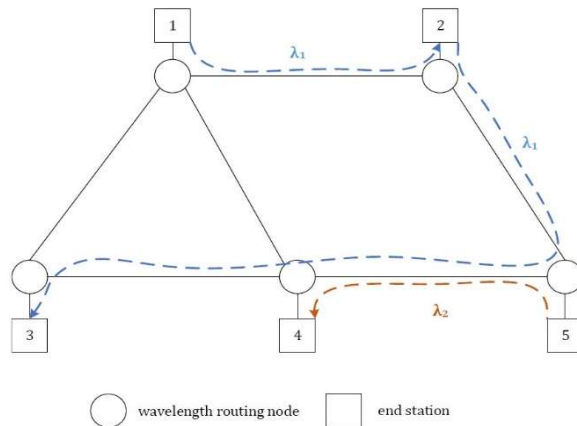
1.8 Ιδιότητες WDM δικτύου

Επαναχρησιμοποίηση μήκους κύματος (wavelength reuse)

Στα WDM οπτικά δίκτυα, το ίδιο μήκος κύματος (λ_1) μπορεί να χρησιμοποιηθεί από πολλαπλά οπτικά μονοπάτια, εφόσον αυτά δεν επικαλύπτονται σε κάποια σύνδεση. Συνεπώς, το δίκτυο έχει τη δυνατότητα να χρησιμοποιεί έναν περιορισμένο αριθμό μηκών κύματος για την υποστήριξη μεγάλου αριθμού οπτικών μονοπατιών. Η παραπάνω λειτουργία περιγράφεται και στο επόμενο σχήμα.

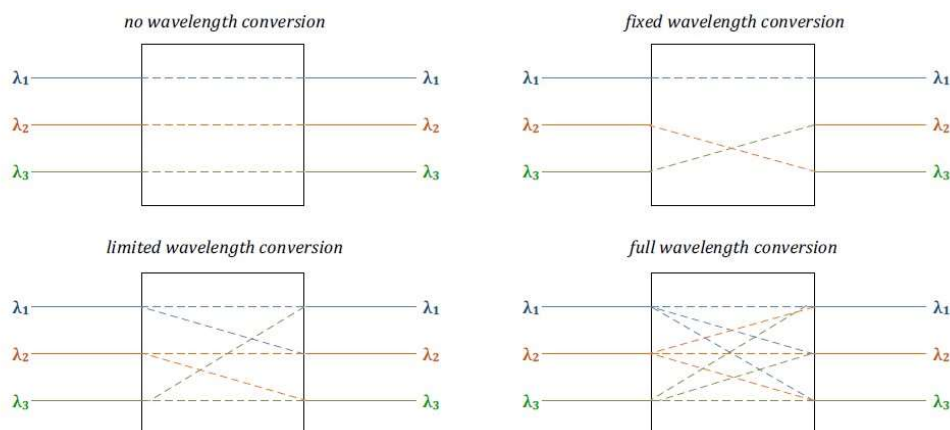
Μετατροπή μήκους κύματος (wavelength conversion)

Στα οπτικά δίκτυα WDM, η λειτουργία μετατροπής μήκους κύματος είναι ιδιαίτερα σημαντική. Το οπτικό σήμα το οποίο, από ένα σύνδεσμο εισόδου φτάνει σε έναν ενδιάμεσο κόμβο με μήκος κύματος λ_i , μπορεί να εξέλθει από τον ενδιάμεσο κόμβο προς το σύνδεσμο εξόδου με διαφορετικό μήκος κύματος λ_j , με $\lambda_i \neq \lambda_j$. Αυτή η μετατροπή από το ένα μήκος κύματος σε ένα άλλο πραγματοποιείται με τη χρήση ενός μετατροπέα μήκους κύματος που τοποθετείται στον ενδιάμεσο κόμβο.



Εικόνα 23: Λειτουργία μετατροπής μήκους κύματος

Υπάρχουν διάφοροι τύποι μετατροπής μήκους κύματος, οι οποίοι περιγράφονται στο παρακάτω σχήμα:



Εικόνα 24: Τύποι μετατροπής μήκους κύματος

- Κατά την πλήρη μετατροπή μήκους κύματος (full wavelength conversion), κάθε μήκος κύματος εισόδου μπορεί να μετατραπεί σε οποιοδήποτε μήκος κύματος εξόδου.
- Κατά την περιορισμένη μετατροπή μήκους κύματος (limited wavelength conversion), κάθε μήκος κύματος εισόδου μπορεί να μετατραπεί σε ένα μόνο υποσύνολο των μηκών κύματος εξόδου.
- Ειδική περίπτωση της περιορισμένης μετατροπής μήκους κύματος είναι η σταθερή μετατροπή μήκους κύματος (fixed wavelength conversion), κατά την οποία το μήκος κύματος εισόδου μπορεί να μετατραπεί σε ένα άλλο συγκεκριμένο μήκος κύματος εξόδου.
- Υπάρχει περίπτωση το μήκος κύματος εισόδου να μην μπορεί να μετατραπεί σε άλλο μήκος κύματος εξόδου. Στην περίπτωση αυτή (no wavelength conversion), το μήκος κύματος εξόδου θα είναι το ίδιο με το μήκος κύματος εισόδου.

Η ιδιότητα της μετατροπής του μήκους κύματος αυξάνει τις επιλογές δρομολόγησης και ως εκ τούτου μειώνεται η πιθανότητα απόρριψης (blocking probability).

Εφαρμογή της ιδιότητας της μετατροπής μήκους κύματος σε έναν κόμβο συναντάμε σε περιπτώσεις όπως, σε μια σύνδεση που φτάνει από έναν κόμβο μεταγωγής του δικτύου και θα πρέπει να κατευθυνθεί σε έναν άλλο (continuing connection), είτε σε μια σύνδεση που φτάνει από έναν κόμβο πρόσβασης (originating connection), είτε και σε μία σύνδεση που καταλήγει σε έναν κόμβο πρόσβασης (terminating connection).

Η ιδιότητα της μετατροπής του μήκους κύματος, με την κατάλληλη μετατροπή μήκους κύματος σε ένα WDM δίκτυο, αυξάνει τις επιλογές δρομολόγησης και ως εκ τούτου μειώνεται ο ρυθμός απόρριψης καθώς και η πιθανότητα απόρριψης (blocking probability) για μία σύνδεση. Επίσης, διαδραματίζει σπουδαίο ρόλο στην καλύτερη και αποδοτικότερη χρησιμοποίηση των διαθέσιμων μηκών κύματος του δικτύου.

Η μετατροπή μήκους κύματος αποτελεί μία πολύπλοκη αλλά και κοστοβόρα διαδικασία. Η πολυπλοκότητα έγκειται στο γεγονός πως όταν δεν υπάρχει δυνατότητα μετατροπής μήκους κύματος, προκύπτει μια αρκετά εύκολη υλοποίηση. Κατ' επέκταση, όσο αυξάνεται ο βαθμός μετατροπής τόσο αυξάνεται και η πολυπλοκότητα της υλοποίησης. Ως εκ τούτου, η πλήρης μετατροπή μήκους κύματος είναι πιο σύνθετη έναντι της περιορισμένης μετατροπής και η περιορισμένη μετατροπή πιο σύνθετη έναντι της σταθερής. Το κόστος της υλοποίησης, όπως γίνεται προφανές, είναι ανάλογο της πολυπλοκότητας της υλοποίησης, καθώς επίσης και σε περιπτώσεις που χρησιμοποιούνται αμιγώς οπτικά μέσα σε κάθε μετατροπή.

Στην πραγματικότητα, υπάρχει άμεση συσχέτιση της δυνατότητας για μετατροπή μήκους κύματος που διαθέτουν οι οπτικοί κόμβοι με το πλήθος των μεταγωγών που απαιτούνται για τη δρομολόγηση των σημάτων. Για κάθε βαθμό μετατροπής απαιτείται ο αντίστοιχος συνδυασμός από μετατροπείς μήκους κύματος και μεταγωγείς ώστε να επιτευχθεί η μετατροπή. Επίσης, μεγάλος βαθμός μετατροπής δεν συνεπάγεται αυτόματα και μεγαλύτερο κέρδος στους κόμβους του δικτύου. Επομένως, για τον περιορισμό στην χρήση υλικοτεχνικού εξοπλισμού όπως οι μεταγωγείς και οι μετατροπείς μήκους κύματος και κατ' επέκταση της πολυπλοκότητας και του κόστους υλοποίησης, συνίσταται η εφαρμογή της περιορισμένης μετατροπής μήκους κύματος έναντι της πλήρους μετατροπής.

Διαφάνεια (Transparency)

Όταν όλες οι λειτουργίες, όπως η μεταγωγή, η ενίσχυση και η μετατροπή μήκους κύματος γίνονται στο αμιγώς οπτικό επίπεδο, τότε αναφερόμαστε στα αμιγώς οπτικά δίκτυα WDM (all-optical WDM). Αποτέλεσμα αυτής της αρχιτεκτονικής επιλογής είναι η εξασφάλιση της διαφάνειας (Transparency) της λειτουργίας του δικτύου καθώς και η σημαντική αύξηση της ταχύτητας του.

Η διαφάνεια αναφέρεται στην ιδιότητα των οπτικών μονοπατιών να μεταδίδουν δεδομένα σε διαφορετικούς ρυθμούς μετάδοσης και διαφορετικά πρωτόκολλα, οπότε δεν εξαρτώνται από το πρωτόκολλο. Η ιδιότητα αυτή δίνει στο δίκτυο την δυνατότητα να υποστηρίζει την ίδια στιγμή διάφορα υψηλότερα επίπεδα.

Επιβίωση (Survivability)

Σύμφωνα με την ιδιότητα της επιβίωσης, ένα οπτικό δίκτυο μπορεί να ρυθμιστεί αυτόματα, έτσι ώστε σε περίπτωση απόρριψης τα οπτικά μονοπάτια να μπορούν να δρομολογηθούν ξανά μέσω εναλλακτικών μονοπατιών. Η ιδιότητα αυτή καθιστά το δίκτυο ιδιαίτερα προσαρμοστικό σε περίπτωση απόρριψης.

Τοπολογία οπτικών μονοπατιών (Lightpath Topology)

Ως τοπολογία οπτικών μονοπατιών ορίζεται ο γράφος που περιγράφει τους κόμβους του δικτύου, με τη χρήση των ακμών να αντιστοιχίζουν την ύπαρξη οπτικών μονοπατιών ανάμεσα στους κόμβους. Η τοπολογία οπτικών μονοπατιών αναφέρεται στην τοπολογία που αντιλαμβάνονται τα υψηλότερα επίπεδα, όπως το επίπεδο IP, που συνδέονται με το οπτικό επίπεδο. Στο IP επίπεδο, τα οπτικά μονοπάτια αντιστοιχούν στις συνδέσεις μεταξύ των IP routers. Η τοπολογία αυτή μπορεί να προσαρμοστεί στις ανάγκες της κίνησης των υψηλότερων επιπέδων.

Κεφάλαιο 2 - Μηχανική Μάθηση

2.1 Εισαγωγή

Η μηχανική μάθηση είναι ένας κλάδος της τεχνητής νοημοσύνης και κατ' επέκταση της επιστήμης των υπολογιστών που ασχολείται με τους αλγόριθμους δημιουργίας μοντέλων που βασίζονται σε μια συλλογή παραδειγμάτων κάποιου φαινομένου. Αυτά τα παραδείγματα μπορούν να προέρχονται είτε από τη φύση, είτε να είναι κατασκευασμένα από ανθρώπους είτε να δημιουργούνται από έναν άλλο αλγόριθμο [17].

Ο Tom M. Mitchell πρότεινε έναν πιο επίσημο ορισμό που χρησιμοποιείται ευρέως για την περιγραφή της μηχανικής μάθησης: «Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από εμπειρία E ως προς μια κλάση εργασιών T και ένα μέτρο επίδοσης P , αν η επίδοσή του σε εργασίες της κλάσης T , όπως αποτιμάται από το μέτρο P , βελτιώνεται με την εμπειρία E » [18].

Η μηχανική μάθηση μπορεί επίσης να οριστεί ως η διαδικασία επίλυσης ενός πρακτικού προβλήματος μέσω είτε της συλλογής ενός συνόλου δεδομένων είτε της αλγοριθμικής δημιουργίας ενός στατιστικού μοντέλου βάσει αυτού του συνόλου δεδομένων. Αυτό το στατιστικό μοντέλο θεωρείται ότι χρησιμοποιείται ποικιλοτρόπως για την επίλυση του πρακτικού προβλήματος.

2.2 Κατηγορίες αλγορίθμων μάθησης

Οι αλγόριθμοι μηχανικής μάθησης μπορούν να ταξινομηθούν σε τέσσερις βασικές κατηγορίες, αναλόγως το είδος της μάθησης:

- Supervised Learning (Μάθηση με Επίβλεψη)
- Unsupervised Learning (Μάθηση χωρίς Επίβλεψη)
- Semi-Supervised Learning (Μάθηση με Ημι-Επίβλεψη)
- Reinforcement Learning (Μάθηση με Ενίσχυση)

Supervised Learning - SL

Στην μάθηση με επίβλεψη, το σύνολο των δεδομένων (dataset) αποτελείται από παραδείγματα εισόδου $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$. Κάθε στοιχείο \mathbf{x}_i στο πλήθος των N στοιχείων ονομάζεται χαρακτηριστικό διάνυσμα. Το χαρακτηριστικό διάνυσμα είναι ένα διάνυσμα στο οποίο κάθε διάσταση $j = 1, \dots, D$ περιέχει μια τιμή που περιγράφει το παράδειγμα ως προς κάποιο χαρακτηριστικό του. Αυτή η τιμή ονομάζεται χαρακτηριστικό και δηλώνεται ως $x^{(j)}$. Για παράδειγμα, εάν κάθε στοιχείο x στη συλλογή αντιπροσωπεύει ένα άτομο, τότε το πρώτο χαρακτηριστικό, $x^{(1)}$, θα μπορούσε να περιέχει ύψος σε cm, το δεύτερο χαρακτηριστικό, $x^{(2)}$ θα μπορούσε να περιέχει βάρος σε kg, το $x^{(3)}$ θα μπορούσε περιέχουν φύλο και ούτω καθεξής. Για όλα

τα παραδείγματα στο σετ δεδομένων, το χαρακτηριστικό στη θέση j στο χαρακτηριστικό διάνυσμα περιέχει πάντα το ίδιο είδος πληροφοριών. Αυτό σημαίνει ότι εάν $x_i^{(2)}$ περιέχει βάρος σε kg σε κάποιο παράδειγμα \mathbf{x}_i , τότε το στοιχείο $x_k^{(2)}$ θα περιέχει επίσης βάρος σε kg σε κάθε παράδειγμα \mathbf{x}_k , $k = 1, \dots, N$. Η κλάση y_i μπορεί να είναι είτε ένα στοιχείο που ανήκει σε ένα πεπερασμένο σύνολο κλάσεων $\{1, \dots, C\}$, είτε ένας πραγματικός αριθμός, είτε μια πιο περίπλοκη δομή, όπως ένα διάνυσμα, μια μήτρα, ένα δέντρο ή ένας γράφος. Επιπρόσθετα, η κλάση y_i μπορεί να αποτελείται είτε από έναν πεπερασμένο αριθμό κλάσεων, είτε από έναν πραγματικό αριθμό. Θεωρούμε την κλάση ως μία κατηγορία στην οποία ανήκει το παράδειγμα εισόδου. Για παράδειγμα, εάν τα παραδείγματά είναι μηνύματα ηλεκτρονικού ταχυδρομείου και το πρόβλημά σας είναι η ανίχνευση ανεπιθύμητων μηνυμάτων, τότε έχετε δύο κατηγορίες $\{\text{spam}, \text{όχι spam}\}$.

Ο στόχος ενός SL αλγορίθμου είναι η χρήση του συνόλου δεδομένων με σκοπό την δημιουργία ενός μοντέλου που θα λαμβάνει ένα χαρακτηριστικό διάνυσμα x ως πληροφορίες εισόδου και εξόδου, που επιτρέπει την εξαγωγή της κλάσης για αυτό το χαρακτηριστικό διάνυσμα. Ένα παράδειγμα τέτοιου αλγορίθμου, είναι ένα μοντέλο που θα δέχεται ως είσοδο ένα χαρακτηριστικό διάνυσμα ενός ατόμου και ως έξοδο να δίνει την πιθανότητα το άτομο αυτό να έχει καρκίνο. Τέλος, οι Supervised Learning αλγόριθμοι αποτελούν τον τύπο μηχανικής μάθησης που βρίσκει πιο συχνή εφαρμογή.

Unsupervised Learning - UL

Στη μάθηση χωρίς επίβλεψη, το σύνολο δεδομένων είναι μια συλλογή παραδειγμάτων χωρίς κλάση $\{\mathbf{x}_i\}_{i=1}^N$. Και σε αυτή την περίπτωση, το \mathbf{x} είναι ένα χαρακτηριστικό διάνυσμα και ο στόχος ενός αλγορίθμου μάθησης χωρίς επίβλεψη είναι να δημιουργήσει ένα μοντέλο που παίρνει ένα χαρακτηριστικό διάνυσμα \mathbf{x} ως είσοδο και είτε το μετατρέπει σε άλλο διάνυσμα είτε σε μια τιμή που μπορεί να χρησιμοποιηθεί για την επίλυση ενός πρακτικού προβλήματος. Οι δύο βασικές υποκατηγορίες του SL είναι το Clustering και το Dimensionality Reduction. Στο Clustering (συσταδοποίηση), το μοντέλο επιστρέφει το αναγνωριστικό (id) της συστάδας (cluster) για κάθε χαρακτηριστικό διάνυσμα στο σύνολο δεδομένων. Στο Dimensionality Reduction (μείωση διαστάσεων), η έξοδος του μοντέλου είναι ένα χαρακτηριστικό διάνυσμα που έχει λιγότερα χαρακτηριστικά από την είσοδο x . Κατά την ανίχνευση των ακραίων τιμών (outliers), η έξοδος είναι ένας πραγματικός αριθμός που δείχνει πώς το x διαφέρει από ένα τυπικό παράδειγμα στο dataset.

Semi-Supervised Learning - SSL

Στη μάθηση με ημι-επίβλεψη, το σύνολο δεδομένων είναι ένας συνδυασμός ταξινομημένων και μη, σε κλάσεις, παραδειγμάτων. Συνήθως, ο αριθμός των παραδειγμάτων χωρίς κλάση είναι πολύ μεγαλύτερος από τον αριθμό των παραδειγμάτων με κλάση. Ο στόχος ενός αλγορίθμου μάθησης με ημι-επίβλεψη είναι ο ίδιος με τον στόχο του αλγορίθμου μάθησης με επίβλεψη. Στο σημείο αυτό, ιδιαίτερη σημασία έχει η χρήση πολλών παραδειγμάτων χωρίς κλάση καθώς μπορούν να βοηθήσουν τον αλγόριθμο μάθησης να δημιουργήσει ένα καλύτερο μοντέλο.

Reinforcement Learning - RL

Η μάθηση με ενίσχυση είναι μία κατηγορία της μηχανικής μάθησης όπου η μηχανή «ζει» σε ένα περιβάλλον και είναι ικανή να αντιληφθεί την κατάσταση αυτού του περιβάλλοντος ως ένα χαρακτηριστικό διάνυσμα. Το μηχάνημα μπορεί να εκτελέσει ενέργειες σε κάθε κατάσταση. Διαφορετικές ενέργειες φέρνουν διαφορετικές ανταμοιβές (rewards) και μπορούν να μετακινήσουν το μηχάνημα σε άλλη κατάσταση του περιβάλλοντος. Ο στόχος ενός αλγορίθμου RL είναι να μάθει μια πολιτική (policy). Μια πολιτική είναι μια συνάρτηση f (παρόμοια με το μοντέλο της μάθησης με επίβλεψη) που λαμβάνει το χαρακτηριστικό διάνυσμα μιας κατάστασης ως είσοδο και εξάγει μια βέλτιστη ενέργεια για εκτέλεση σε αυτήν την κατάσταση. Η ενέργεια είναι βέλτιστη αν μεγιστοποιήσει την αναμενόμενη μέση ανταμοιβή (expected average reward). Η μάθηση με ενίσχυση επιλύει ένα συγκεκριμένο είδος προβλημάτων όπου η λήψη αποφάσεων είναι διαδοχική και ο στόχος είναι μακροπρόθεσμος.

2.3 Επιλογή αλγορίθμου Μηχανικής Μάθησης

Η επιλογή του αλγορίθμου μηχανικής μάθησης που θα χρησιμοποιηθεί δεν είναι εύκολη διαδικασία. Αν και σε ερευνητικό επίπεδο θεωρητικά υπάρχει αρκετός διαθέσιμος χρόνος ακόμα και να εξεταστούν όλες οι επιλογές, στην πράξη ο χρόνος είναι περιορισμένος. Παρόλα αυτά, υπάρχουν κάποια κριτήρια τα οποία αν ληφθούν υπόψη, βοηθούν στη λήψη αυτής της απόφασης. Τέτοια κριτήρια είναι:

- Πλήθος χαρακτηριστικών και παραδειγμάτων. Ορισμένοι αλγόριθμοι όπως τα νευρωνικά δίκτυα ή αλγόριθμοι ενίσχυσης κλίσης (gradient boosting algorithm) μπορούν να διαχειριστούν τεράστια dataset με εκατομμύρια παραδειγμάτων και χαρακτηριστικών. Αντιθέτως, άλλοι αλγόριθμοι όπως οι SVM έχουν μικρή χωρητικότητα στη διαχείριση των δεδομένων.
- Τύπος χαρακτηριστικών, δηλαδή εάν τα χαρακτηριστικά του dataset είναι αριθμητικές τιμές ή κατηγορίες (κλάσεις) ή ανάμεικτα. Ορισμένοι αλγόριθμοι δεν έχουν την δυνατότητα να διαχειριστούν χαρακτηριστικά σε μορφή κλάσης, οπότε πρέπει να τα μετατραπούν σε αριθμητικές τιμές, χρησιμοποιώντας παραδείγματος χάρη την τεχνική one-hot encoding.
- Μη γραμμικότητα δεδομένων, δηλαδή εάν τα δεδομένα μας είναι γραμμικά διαχωρίσιμα ή αν μπορούν να μοντελοποιηθούν χρησιμοποιώντας ένα γραμμικό μοντέλο. Σε περίπτωση που τα δεδομένα μας είναι γραμμικά διαχωρίσιμα, μπορούν να χρησιμοποιηθούν αλγόριθμοι όπως linear regression, logistic regression ή SVM με γραμμικό πυρήνα. Σε αντίθετη περίπτωση δύναται η εφαρμογή deep neural networks ή συνδυαστικών μεθόδων (ensemble algorithms).
- Ταχύτητα εκπαίδευσης, δηλαδή πόσο χρόνο χρειάζεται ο αλγόριθμος να δημιουργήσει ένα μοντέλο. Για παράδειγμα τα νευρωνικά δίκτυα απαιτούν πολύ χρόνο για να εκπαιδευτούν. Πιο απλοί αλγόριθμοι όπως linear regression, logistic regression και decision tree learning είναι αρκετά πιο γρήγοροι στην εκπαίδευση του μοντέλου. Επίσης, άλλοι αλγόριθμοι έχουν την δυνατότητα

εκμετάλλευσης των πολλών πυρήνων (CPU cores) ενός υπολογιστικού συστήματος και της παράλληλης επεξεργασίας με αποτέλεσμα να μειώνουν αισθητά τον χρόνο εκπαίδευσης. Τέλος, υπάρχουν και πολλές βιβλιοθήκες διαθέσιμες για το σκοπό αυτό, διαθέτοντας συγκεκριμένους αλγόριθμους για την εκπαίδευση του μοντέλου.

- Ταχύτητα πρόβλεψης, δηλαδή το πόσο γρήγορα ένα μοντέλο πραγματοποιεί τις προβλέψεις. Ορισμένοι αλγόριθμοι όπως οι SVM, οι linear regression, οι logistic regression και ορισμένοι τύποι νευρωνικών δικτύων πραγματοποιούν πολύ γρήγορα τις προβλέψεις τους. Αντιθέτως, αλγόριθμοι όπως k-NN, ensemble, και βαθειάς μάθησης ή τα συνελκτικά νευρωνικά δίκτυα (RNN) είναι αρκετά πιο αργοί.
- Διαθέσιμη υπολογιστική μνήμη, δηλαδή εάν ένα dataset μπορεί να φορτωθεί πλήρως στη μνήμη RAM ενός υπολογιστικού συστήματος. Εάν το σύστημα διαθέτει μεγάλη μνήμη, τότε δύναται η επιλογή μέσα από μία πληθώρα αλγορίθμων. Σε αντίθετη περίπτωση, προτείνεται η χρήση αλγορίθμων σταδιακής μάθησης (incremental learning algorithms) οι οποίοι μπορούν να βελτιώσουν το μοντέλο προσθέτοντας σταδιακά περισσότερα δεδομένα.

Ένας καλός τρόπος επιλογής αλγορίθμου μηχανικής μάθησης, είναι να το τεστάρουμε στο validation set (σετ επικύρωσης) του dataset.

2.4 Μέθοδος διαχωρισμού ενός dataset

Ένα σετ δεδομένων (dataset) χωρίζεται σε τρία μέρη:

- Το Training set (σετ εκπαίδευσης)
- Το Validation set (σετ επικύρωσης)
- Το Test set

Ο κύκλος εργασίας δοθέντος ενός σετ δεδομένων είναι ο εξής:

Αρχικά τα δεδομένα ανακατεύονται (shuffle) και εν συνεχεία χωρίζονται (split) στα τρία σετ που αναφέρθηκαν πιο πάνω. Το Training set είναι συνήθως μεγαλύτερο από τα άλλα δύο και χρησιμοποιείται για να εκπαιδευτεί το μοντέλο. Το Validation set έχει περίπου το ίδιο μέγεθος με το Test set και είναι αρκετά μικρότερα από το training set. Ο αλγόριθμος μάθησης δεν μπορεί να χρησιμοποιήσει κάποιο από τα δύο αυτά σετ (τα οποία συχνά αποκαλούνται και hold-out sets) για να εκπαιδευτεί, παρά μόνο το training set.

Δεν υπάρχει βέλτιστο ποσοστό να χωριστούν αυτά τα τρία σετ, αν και στο παρελθόν είθισται να χωρίζονταν με τον ακόλουθο τρόπο: 70% καταλάμβανε το training set, 15% το validation και 15% το test set. Αν και στην εποχή των Big Data που διανύουμε, όπου τα dataset απαρτίζονται από εκατομμύρια παραδείγματα και ετικέτες, το training set καταλαμβάνει το 95% του dataset, το validation set καταλαμβάνει το 2.5% και το test set επίσης το 2.5%.

Η χρησιμότητα του validation set έγκειται στην επιλογή του κατάλληλου αλγορίθμου εκμάθησης, καθώς και στην εύρεση των βέλτιστων τιμών των υπερπαραμέτρων. Από την άλλη πλευρά, η χρησιμότητα του test set έγκειται στην εκτίμηση του μοντέλου ως προς τις προβλέψεις του.

Ο λόγος που το σετ δεδομένων χωρίζεται σε τρία μέρη και δεν αξιοποιείται ολόκληρο για την εκπαίδευση του μοντέλου οφείλεται στο γεγονός πως ο στόχος δεν είναι η ανάπτυξη απλά ενός μοντέλου που να εξάγει καλές προβλέψεις για τις κλάσεις των παραδειγμάτων που έχει ήδη δει. Ένας τέτοιος αλγόριθμος θα έχει απομνημονεύσει τα προς εκπαίδευση παραδείγματα και θα πραγματοποιεί προβλέψεις αλάνθαστα. Ο πραγματικός στόχος είναι να κάνει προβλέψεις για δεδομένα που δεν έχει ξαναεπεξεργαστεί.

2.5 Τύποι δεδομένων και ορισμοί

Βαθμωτά μεγέθη (Scalars)

Ένα βαθμωτό μέγεθος είναι τυπικά μια αριθμητική τιμή (όπως το 13 ή το -2.5).

Διανύσματα (Vectors)

Ένα διάνυσμα (παραδείγματος χάρη x ή w) είναι τυπικά μία ταξινομημένη λίστα βαθμωτών μεγεθών (attributes). Τα διανύσματα μπορούν να απεικονιστούν ως βέλη που δείχνουν προς ορισμένες κατευθύνσεις καθώς και σημεία σε έναν πολυδιάστατο χώρο.

Σετ (Sets)

Ένα σετ (παραδείγματος χάρη το S) αποτελείται από μία μη ταξινομημένη συλλογή μοναδικών στοιχείων. Ένα σετ από αριθμούς μπορεί να είναι είτε πεπερασμένο είτε άπειρο και να εμπεριέχει όλες τις τιμές μέσα σε ένα διάστημα.

2.6 Μέθοδοι επεξεργασίας των δεδομένων

One-Hot Encoding

Ορισμένοι αλγόριθμοι μάθησης λειτουργούν με χαρακτηριστικό διανύσματα που περιέχει αποκλειστικά αριθμητικές τιμές. Όταν κάποιο χαρακτηριστικό στο dataset μπορεί να θεωρηθεί ως κατηγορία, όπως «χρώματα» ή «ημέρες της εβδομάδας», μια τέτοια κατηγορία χαρακτηριστικών δύναται να μετατραπεί σε πολλές δυαδικές. Μία τέτοια τεχνική είναι η μετατροπή ενός χρώματος σε μορφή RGB.

Με τη διαδικασία αυτή αυξάνεται η διάσταση του χαρακτηριστικού διανύσματος. Βέβαια, στο σημείο αυτό να αναφερθεί πως χρειάζεται προσοχή σε αυτή την αύξηση της διάστασης, καθώς υπάρχει περίπτωση ο αλγόριθμος να «μπερδευτεί» και να

προκληθεί το λεγόμενο overfitting (υπερβολική προσαρμογή), καθώς ο αλγόριθμος θα ψάχνει να βρει μία κανονικότητα στις τιμές η οποία όμως δεν θα υπάρχει.

Binning

Μία αντίστροφη διαδικασία, η οποία βέβαια δεν εφαρμόζεται στην πράξη συχνά είναι το Binning (συναντάται όπως και ως bucketing). Στην περίπτωση αυτή το χαρακτηριστικό διάνυσμα αποτελείται από αριθμητικές τιμές και στόχος είναι οι τιμές αυτές να κατηγοριοποιηθούν. Ουσιαστικά είναι μία διαδικασία μετατροπής συνεχόμενων χαρακτηριστικών σε πολλαπλά δυαδικά χαρακτηριστικά, τα λεγόμενα buckets, συνήθως βάσει όπως εύρους τιμών. Σε ορισμένες περιπτώσεις, μια προσεκτικά σχεδιασμένη Binning διαδικασία μπορεί να βοηθήσει έναν αλγόριθμο μάθησης να «μάθει» χρησιμοποιώντας λιγότερα παραδείγματα. Αυτό συμβαίνει καθότι γίνεται «υπόδειξη» στον αλγόριθμο να μην βρει την ακριβή τιμή όπως χαρακτηριστικού αρκεί να βρει ότι η τιμή αυτή εμπίπτει σε ένα συγκεκριμένο εύρος τιμών.

Κανονικοποίηση (Normalization)

Κανονικοποίηση ονομάζεται η διαδικασία μετατροπής όπως εύρους τιμών που μπορεί να πάρει ένα χαρακτηριστικό σε ένα άλλο συγκεκριμένο εύρος τιμών, συνήθως στο διάστημα $[-1,1]$ ή $[0,1]$.

Η μαθηματική έκφραση όπως διαδικασίας όπως είναι:
$$\bar{x}^{(j)} = \frac{x^{(j)} - \min^{(j)}}{\max^{(j)} - \min^{(j)}} \quad (1),$$

όπου $\min^{(j)}$ και $\max^{(j)}$ είναι η ελάχιστη και μέγιστη τιμή του χαρακτηριστικού στο dataset.

Η κανονικοποίηση δεν είναι μία «αυστηρή» απαίτηση στην επεξεργασία των δεδομένων. Ωστόσο, στην πράξη μπορεί να οδηγήσει στην ταχύτερη μάθηση του αλγόριθμου. Επιπρόσθετα, είναι χρήσιμο να διασφαλιστεί πως τα δεδομένα εισόδου βρίσκονται περίπου στο ίδιο εύρος τιμών, ώστε να αποφευχθούν σφάλματα όπως όταν δουλεύουμε με πολύ μεγάλους ή πολύ μικρούς αριθμούς (σφάλμα γνωστό και ως αριθμητική υπερχείλιση – numerical overflow).

Τυποποίηση (Standardization)

Η τυποποίηση (ή αλλιώς κανονικοποίηση στον z-άξονα) είναι η διαδικασία κατά την οποία οι τιμές των χαρακτηριστικών επανακλιμακοποιούνται ώστε να έχουν όπως ιδιότητες κανονικής κατανομής με $\mu = 0$ και $\sigma = 1$, όπου μ είναι ο μέση τιμή όλων των χαρακτηριστικών στο πλήθος δειγμάτων του dataset και σ η τυπική απόκλιση από τη μέση τιμή. Η μαθηματική έκφραση όπως διαδικασίας όπως είναι:

$$\hat{x}^{(j)} = \frac{x^{(j)} - \mu^{(j)}}{\sigma^{(j)}} \quad (2).$$

Ένα ζήτημα που προκύπτει στην διαχείριση των δεδομένων αφορά το ποια από όπως παραπάνω δύο τεχνικές χρειάζεται να εφαρμοστεί. Η απάντηση δεν είναι ξεκάθαρη, αν και κατά βάση έγκειται στο μέγεθος του dataset, την ποιότητά του και τον διαθέσιμο χρόνο που έχουμε όπως επεξεργασία του.

Συνοπτικά όπως, έχουν προταθεί οι παρακάτω κανόνες:

- Οι SL αλγόριθμοι συνήθως επωφελούνται περισσότερο από την εφαρμογή όπως standardization παρά όπως normalization.
- Εάν οι τιμές όπως χαρακτηριστικού είναι κατανεμημένες κοντά σε μία κανονική κατανομή (καμπύλη Gauss), συνήθως εφαρμόζεται standardization.
- Σε περιπτώσεις που οι τιμές όπως χαρακτηριστικού παίρνουν πού υψηλές ή χαμηλές τιμές (ακραίες τιμές – outliers), προτιμάται η τυποποίηση, καθώς με κανονικοποίηση οι κανονικές τιμές θα συμπυκνωθούν σε ένα πολύ μικρό εύρος τιμών.
- Σε όπως όπως περιπτώσεις συνίσταται η κανονικοποίηση.

Τέλος να αναφερθεί πως αν και συνήθως η κλιμακοποίηση των δεδομένων των χαρακτηριστικών (feature rescaling) είναι ιδιαίτερα χρήσιμη για όπως περισσότερους αλγόριθμους μηχανικής μάθησης, σε ορισμένες περιπτώσεις ωστόσο, το μοντέλο είναι καλό να εκπαιδεύεται τα κανονικά αρχικά δεδομένα.

2.7 Εξαγωγή χαρακτηριστικών προβλήματος – Feature Engineering

Ένα από τα βασικά ζητήματα στην εφαρμογή τεχνικών μηχανικής μάθησης, είναι η ποιότητα των δεδομένων. Όπως αναφέραμε, το dataset πρέπει να είναι σε μορφή κατάλληλη ώστε να μπορεί να αναλυθεί από το σύστημα. Διαδικασίες που συνεισφέρουν ώστε να επιτευχθεί αυτό το επιθυμητό σημείο μπορεί να είναι ο καθαρισμός μη φυσιολογικών τιμών, η ομαλοποίηση των δεδομένων, η μορφοποίηση όπως, η μορφοποίηση του συνόλου των δεδομένων εισόδου, η μορφοποίηση του λεγόμενου κυλιόμενου παράθυρου (rolling window) που χρησιμοποιείται για την ανάλυση μοτίβων και δεδομένων ακολουθιών, ο ποιοτικός έλεγχος δεδομένων (data reconciliation), όταν δηλαδή τα δεδομένα εκφράζονται σε χρόνο/απόσταση συναρτήσει των υπολοίπων. Να τονιστεί στο σημείο αυτό πως οι διεργασίες προ-επεξεργασίας των δεδομένων διαφέρουν ανάλογα τον τομέα όπως μελέτης ή όπως εφαρμογής και την ποιότητα των διαθέσιμων δεδομένων.

Υπενθυμίζεται πως ένα dataset αποτελείται από μία συλλογή από παραδείγματα με κλάσεις, όπου $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$. Κάθε στοιχείο \mathbf{x}_i στο πλήθος των N στοιχείων ονομάζεται χαρακτηριστικό διάνυσμα. Ένα χαρακτηριστικό διάνυσμα με διαστάσεις $j = 1, \dots, D$ περιέχει μία τιμή που περιγράφει το παράδειγμα.

Η διαδικασία μετατροπής μιας σειράς δεδομένων σε dataset ικανό όπως επεξεργασία ονομάζεται «feature engineering». Ένα μοντέλο έχει χαμηλό bias όταν προβλέπει ορθά τα δεδομένα εκπαίδευσης. Αυτό σημαίνει πως πραγματοποιούνται λίγα λάθη κατά την διαδικασία πρόβλεψης όπως κλάσης του παραδείγματος που χρησιμοποιήθηκε για να δημιουργηθεί το μοντέλο.

2.8 Διαχείριση χαρακτηριστικών που λείπουν από ένα dataset

Όπως αλγόριθμος ενδέχεται να συναντήσει περιπτώσεις με dataset τα οποία περιέχουν χαρακτηριστικά στα οποία λείπουν κάποιες τιμές. Αυτό συνήθως συμβαίνει σε dataset που έχουν κατασκευαστεί χειρωνακτικά και κάποια τιμή είτε παραλείφθηκε από ανθρώπινο σφάλμα, είτε δεν προσμετρήθηκε ποτέ.

Η διαδικασία διαχείρισης δεδομένων στα οποία λείπουν κάποιες τιμές είναι:

- Η αφαίρεση από το dataset των παραδειγμάτων όπου λείπουν τα χαρακτηριστικά. Αυτό δύναται να συμβεί όταν το dataset είναι αρκετά μεγάλο και δεν θα έχει επίπτωση η αφαίρεση κάποιων όπως εκπαίδευση δεδομένων.
- Η χρήση αλγορίθμων μάθησης που μπορούν να διαχειριστούν dataset όπως αυτά που περιγράφονται (εξαρτάται από την βιβλιοθήκη και την συγκεκριμένη εφαρμογή του κάθε αλγορίθμου).
- Χρησιμοποιώντας τεχνικές εισαγωγής δεδομένων.

Με τη χρήση τεχνικών εισαγωγής δεδομένων (data imputation technique) ουσιαστικά εισάγονται τιμές όπως θέσεις των τιμών των χαρακτηριστικών που λείπουν από το dataset. Αναλόγως την τεχνική που χρησιμοποιείται, το κριτήριο επιλογής των τιμών αυτών μπορεί να διαφέρει. Για παράδειγμα, μια τεχνική συνιστά την αντικατάσταση όπως τιμής όπως χαρακτηριστικού που λείπει με το μέσο όρων των τιμών του χαρακτηριστικού. Άλλη τεχνική αντικατάστασης σε περίπτωση που οι τιμές ανήκουν σε ένα συγκεκριμένο διάστημα είναι να εισάγουμε μία τιμή εκτός του διαστήματος αυτού. Τέλος, μια πιο εξειδικευμένη τεχνική αντικατάστασης τιμής που λείπει, είναι ο ορισμός όπως όπως τιμής ως τον στόχο όπως regression προβλήματος.

2.9 Ρύθμιση υπερπαραμέτρων

Οι υπερπαραμέτροι (όπως η παράμετρος α στον αλγόριθμο gradient descent ή η παράμετρος k στον αλγόριθμο k-NN) δεν βελτιστοποιούνται από τον ίδιο τον αλγόριθμο. Ρυθμίζονται κατά την ανάλυση των δεδομένων από όπως αναλυτές, βρίσκοντας πειραματικά τον καλύτερο συνδυασμό των τιμών για κάθε παράμετρο.

Η ρύθμιση αυτή μπορεί να πραγματοποιηθεί μέσω όπως στρατηγικής όπως η grid search, αρκεί το dataset να είναι αρκετά μεγάλο ώστε να υπάρχει όπως ικανοποιητικά μεγάλος αριθμός παραδειγμάτων στο validation set και ο αριθμός των υπερπαραμέτρων καθώς και το εύρος όπως να μην είναι μεγάλο. Κατά την στρατηγική αυτή δημιουργούνται όλοι οι δυνατοί συνδυασμοί των τιμών των υπερπαραμέτρων, δημιουργούνται και γίνονται train τόσα μοντέλα όσα και το πλήθος των συνδυασμών και γίνεται εκτίμηση όπως επίδοσης των μοντέλων στο validation set. Τα αποτελέσματα συγκρίνονται μεταξύ όπως και έτσι γίνεται η επιλογή του καλύτερου συνδυασμού για όπως υπερπαραμέτρους αυτές. Για πιο σύνθετα dataset, εφαρμόζονται τεχνικές όπως η random search ή Μπαγεςιανή ρύθμιση υπερπαραμέτρων (Bayesian hyperparameter optimization).

2.10 Supervised Learning

2.10.1 Λειτουργία μάθησης με επίβλεψη

Μια Supervised Learning διαδικασία ξεκινά με τη συλλογή των δεδομένων. Τα δεδομένα για την μάθηση με επίβλεψη είναι μια συλλογή ζευγών δεδομένων (είσοδος, έξοδος). Η είσοδος μπορεί να είναι οτιδήποτε, για παράδειγμα, μηνύματα email, εικόνες ή μετρήσεις αισθητήρα. Οι έξοδοι είναι συνήθως πραγματικοί αριθμοί ή κλάσεις. Σε ορισμένες περιπτώσεις οι έξοδοι ενδέχεται να είναι διανύσματα, ακολουθίες είτε να έχουν κάποια άλλη δομή.

Μέσω ανάλυσης δεδομένων (data analytics) αποφασίζεται πώς μπορεί μια πραγματική οντότητα να μετατραπεί σε ένα χαρακτηριστικό διάνυσμα. Έτσι τα δεδομένα αποκτούν μορφή την οποία δύναται να επεξεργαστεί η μηχανή. Τα δεδομένα εξόδου παίρνουν μορφή κλάσης (συνήθως με τιμές 0 ή 1). Πρόκειται ουσιαστικά για μία επαναληπτική διαδικασία που πραγματοποιείται για κάθε ζεύγος δεδομένων. Σε πολλές περιπτώσεις τα δεδομένα μας ταξινομούνται σε διαφορετικές κλάσεις. Στη μηχανική μάθηση, το όριο που χωρίζει τα παραδείγματα διαφορετικών κλάσεων ονομάζεται όριο απόφασης (decision boundary).

Επομένως, με την παραπάνω διαδικασία έχουμε έτοιμο το dataset το οποίο μπορεί να χρησιμοποιηθεί από έναν αλγόριθμο μάθησης και να δημιουργήσει το μοντέλο μας.

2.10.2 Θεμελιώδεις αλγόριθμοι μάθησης με επίβλεψη

2.10.2.1 Linear Regression

Η γραμμική παλινδρόμηση (Linear Regression) αποτελεί έναν από τους πιο δημοφιλής regression learning αλγόριθμους. Ο αλγόριθμος αυτός «εκπαιδεύει» το μοντέλο, το οποίο είναι ένας γραμμικός συνδυασμός των χαρακτηριστικών των παραδειγμάτων εισόδου.

Το dataset αποτελείται από μία συλλογή παραδειγμάτων εισόδου $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, όπου N είναι το πλήθος των παραδειγμάτων, \mathbf{x}_i είναι το χαρακτηριστικό διάνυσμα D -διαστάσεων για κάθε παράδειγμα $i=1, \dots, N$, ο όρος y_i είναι ο στόχος και είναι μία πραγματική τιμή, και κάθε χαρακτηριστικό $x_i^{(j)}$ με $j = 1, \dots, D$ είναι επίσης πραγματικός αριθμός.

Ο σκοπός του αλγόριθμου είναι να δημιουργήσει ένα μοντέλο $f_{\mathbf{w},b}(\mathbf{x})$ ως έναν γραμμικό συνδυασμό των χαρακτηριστικών των παραδειγμάτων \mathbf{x} :

$$f_{\mathbf{w},b}(\mathbf{x}) = \mathbf{w}\mathbf{x} + b \quad (3),$$

όπου \mathbf{w} είναι διάνυσμα παραμέτρων D -διαστάσεων και b πραγματικός αριθμός. Ο όρος $f_{\mathbf{w},b}$ καταδεικνύει πως το μοντέλο f είναι συνάρτηση των τιμών \mathbf{w} και b .

Στη συνέχεια, δοθέντος του \mathbf{x} , το μοντέλο προβλέπει τον άγνωστο y : $y \leftarrow f_{\mathbf{w},b}(\mathbf{x})$. Ο στόχος είναι να βρεθούν οι βέλτιστες τιμές των παραμέτρων (\mathbf{w}^*, b^*) οι οποίες θα διαμορφώσουν το μοντέλο που θα κάνει τις πιο ακριβείς προβλέψεις.

Η διαδικασία βελτιστοποίησης που ακολουθείται για την εύρεση των βέλτιστων παραμέτρων \mathbf{w}^* και b^* προσπαθεί να ελαχιστοποιήσει την συνάρτηση:

$$\frac{1}{N} \sum_{i=1 \dots N} (f_{\mathbf{w},b}(\mathbf{x}_i) - y_i)^2 \quad (4).$$

Η μαθηματική έκφραση $(f(\mathbf{x}_i) - y_i)^2$ ονομάζεται συνάρτηση απώλειας (loss function) και είναι μέτρο ποινής (penalty) για την λάθος ταξινόμηση ενός παραδείγματος i . Η συγκεκριμένη συνάρτηση ονομάζεται επίσης και συνάρτηση απώλειας τετραγωνικού σφάλματος (squared error loss function) ή μέσο τετραγωνικό σφάλμα (mean squared error - MSE). Όλοι οι model-based learning αλγόριθμοι έχουν μία συνάρτηση απώλειας, γνωστή και ως συνάρτηση κόστους (cost function) την οποία προσπαθούμε να ελαχιστοποιήσουμε ώστε να καταλήξουμε στο καλύτερο μοντέλο. Σε ένα linear regression μοντέλο, η συνάρτηση κόστους δίνεται από την μέση απώλεια, η οποία ορίζεται ως ο μέσος όρος όλων των ποινών που επιβλήθηκαν εφαρμόζοντας το μοντέλο στα training data.

2.10.2.2 Logistic Regression

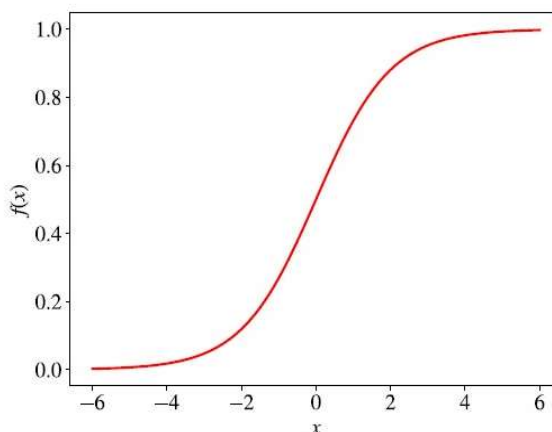
Οι logistic regression αλγόριθμοι στην ουσία είναι αλγόριθμοι μάθησης μέσω ταξινόμησης (classification learning). Ο όρος regression οφείλεται καθαρά στο γεγονός πως βασίζεται σε παρόμοια μαθηματική φόρμουλα με αυτή της linear regression.

Όμοια με τη γραμμική, έτσι και στη λογιστική παλινδρόμηση, στόχος είναι να μοντελοποιηθεί το y_i ως μια γραμμική συνάρτηση του \mathbf{x}_i . Ωστόσο, για ένα δυαδικό y_i , αυτό δεν είναι εύκολο. Ένας γραμμικός συνδυασμός των χαρακτηριστικών, όπως η συνάρτηση $\mathbf{w}\mathbf{x} + b$, είναι μία συνάρτηση που εκτείνεται από το μείον άπειρο έως το συν άπειρο, την ίδια στιγμή που το y_i μπορεί να πάρει μόνο δύο τιμές.

Για την ανάπτυξη ενός γραμμικού μοντέλου ταξινόμησης παρατηρήθηκε πως εάν ορίσουμε μία αρνητική κλάση ως 0 και μία θετική ως ένα, χρειαζόμαστε απλά μία συνεχή συνάρτηση της οποίας το πεδίο τιμών είναι το $(0,1)$. Η συνάρτηση που πληροί αυτή την ιδιότητα είναι η σιγμοειδής συνάρτηση (sigmoid function ή standard logistic function):

$$f(x) = \frac{1}{1+e^{-x}} \quad (5).$$

Παρατηρώντας τη γραφική παράσταση της σιγμοειδούς συνάρτησης, γίνεται αντιληπτό πως εξυπηρετείται το πρόβλημα της ταξινόμησης: εάν βελτιστοποιήσουμε κατάλληλα τις παραμέτρους \mathbf{w} και b , μπορούμε να ερμηνεύσουμε το αποτέλεσμα της συνάρτησης $f(x)$ ως την πιθανότητα να είναι θετικό το y_i . Επομένως, ορίζουμε ένα κατώφλι (threshold) αναλόγως το πρόβλημα και αν το αποτέλεσμα της συνάρτησης είναι μεγαλύτερο ή ίσο του κατωφλίου μας, θεωρούμε πως η κλάση του \mathbf{x} είναι θετική, αλλιώς είναι αρνητική.



Εικόνα 25: Γραφική παράσταση της σιγμοειδούς

Επομένως, η μαθητική φόρμουλα διαμορφώνεται ως εξής:

$$f_{\mathbf{w},b}(\mathbf{x}) \stackrel{\text{def}}{=} \frac{1}{1+e^{-(\mathbf{w}\mathbf{x}+b)}} \quad (6).$$

Το κριτήριο βελτιστοποίησης στους logistic regression αλγόριθμους ονομάζεται μέγιστη πιθανοφάνεια (maximum likelihood). Σε αντίθεση με τους linear regression όπου προσπαθούμε να ελαχιστοποιήσουμε τη συνάρτηση κόστους, εδώ προσπαθούμε να μεγιστοποιήσουμε την πιθανοφάνεια των training data, σύμφωνα με το μοντέλο:

$$L_{\mathbf{w},b} \stackrel{\text{def}}{=} \prod_{i=1 \dots N} f_{\mathbf{w},b}(\mathbf{x}_i)^{y_i} (1 - f_{\mathbf{w},b}(\mathbf{x}_i))^{(1-y_i)} \quad (7).$$

Εξαιτίας της χρήσης της συνάρτησης \exp , πρακτικά είναι βολικότερο να μεγιστοποιήσουμε την λογαριθμική πιθανοφάνεια έναντι της πιθανοφάνειας. Η λογαριθμική πιθανοφάνεια (log-likelihood) ορίζεται ως:

$$\log L_{\mathbf{w},b} \stackrel{\text{def}}{=} \ln(L(\mathbf{w}, b(\mathbf{x}))) = \sum_{i=1}^N y_i \ln f_{\mathbf{w},b}(\mathbf{x}) + (1 - y_i) \ln(1 - f_{\mathbf{w},b}(\mathbf{x})) \quad (8).$$

Καθώς η συνάρτηση \ln είναι μία αυστηρά αύξουσα συνάρτηση, η μεγιστοποίηση αυτής της συνάρτησης είναι το ίδιο με τη μεγιστοποίηση των όρων της. Επομένως η λύση σε αυτό το πρόβλημα βελτιστοποίησης είναι η ίδια με του αρχικού προβλήματος.

Εν αντιθέσει με την περίπτωση του linear regression, σε αυτό το πρόβλημα βελτιστοποίησης δεν υπάρχει κλειστή μορφή λύσης. Για τον λόγο αυτό, σε μία τυπική

αριθμητική διαδικασία βελτιστοποίησης όπως αυτή, επιστρατεύουμε τον αλγόριθμο Gradient Descent.

2.10.2.3 Δέντρο Λήψης Απόφασης (Decision Tree Learning)

Το δέντρο αποφάσεων είναι ένας μη-περιοδικός γράφος που χρησιμοποιείται για τη λήψη αποφάσεων. Σε κάθε κόμβο διακλάδωσης του γράφου εξετάζεται ένα συγκεκριμένο χαρακτηριστικό j του χαρακτηριστικού διανύσματος. Εάν η τιμή του χαρακτηριστικού είναι κάτω από ένα συγκεκριμένο κατώφλι που έχει οριστεί σύμφωνα με το πρόβλημα, τότε ακολουθείται η αριστερή διακλάδωση, ενώ σε αντίθετη περίπτωση η δεξιά. Φτάνοντας στον κόμβο των φύλλων (leaf node), έχει παρθεί και η απόφαση για την κλάση στην οποία ανήκει το χαρακτηριστικό.

Στον αλγόριθμο δέντρου λήψης αποφάσεων έχουμε ως δεδομένα εισόδου παραδείγματα με κλάσεις, τα οποία ανήκουν στο σετ $\{0,1\}$. Υπάρχουν διάφοροι αλγόριθμοι που επιλύουν τέτοιου είδους προβλήματα, τη δημιουργία δηλαδή δέντρων λήψης αποφάσεων. Αυτός που συναντάται με μεγαλύτερη συχνότητα είναι ο αλγόριθμος ID3 (Iterative Dichotomiser 3). Το κριτήριο βελτιστοποίησης στην περίπτωση αυτή είναι η μέση λογαριθμική πιθανοφάνεια (average log-likelihood):

$$\frac{1}{N} \sum_{i=1}^N y_i \ln f_{ID3}(\mathbf{x}_i) + (1 - y_i) \ln(1 - f_{ID3}(\mathbf{x}_i)) \quad (9),$$

όπου f_{ID3} είναι το δέντρο απόφασης.

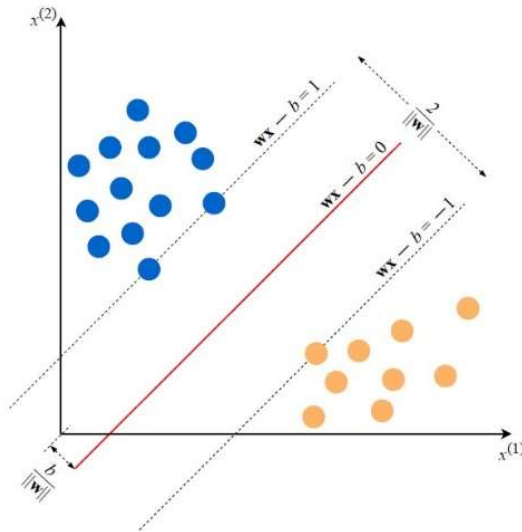
Ο αλγόριθμος ID3 βελτιστοποιεί το κριτήριο κατασκευάζοντας προσεγγιστικά ένα μη παραμετρικό μοντέλο $f_{ID}(\mathbf{x}) \stackrel{\text{def}}{=} \Pr(y = 1 | \mathbf{x})$.

2.10.2.4 Support Vector Machine – SVM

Ο αλγόριθμος SVM είναι επίσης ένας αλγόριθμος ταξινόμησης ο οποίος απαιτεί η θετική κλάση να παίρνει την αριθμητική τιμή +1 την αρνητική τιμή -1. Αντιμετωπίζει κάθε χαρακτηριστικό διάνυσμα ως ένα πολυδιάστατο χώρο D -διαστάσεων και τα τοποθετεί σε ένα φανταστικό επίπεδο D -διαστάσεων χαράσσοντας μία φανταστική γραμμή D -διαστάσεων (hyperplane), η οποία χωρίζει τα παραδείγματα με θετική κλάση από αυτά με την αρνητική. Στην μηχανική μάθηση, το όριο που χωρίζει τα παραδείγματα διαφορετικών κλάσεων, λέγεται όριο απόφασης (decision boundary).

Η εξίσωση της ευθείας αυτής (hyperplane) δίνεται από δύο παραμέτρους, ένα διάνυσμα \mathbf{w} ίδιων διαστάσεων με το χαρακτηριστικό διάνυσμα \mathbf{x} και έναν πραγματικό αριθμό b :

$$\mathbf{w}\mathbf{x} + b \quad (10).$$



Εικόνα 26: Γραφική παράσταση hyperplane

Ο αλγόριθμος προβλέπει την κλάση του παραδείγματος \mathbf{x} μέσω της συνάρτησης:

$$y = \text{sign}(\mathbf{w}\mathbf{x} - b) \quad (11),$$

όπου sign είναι μαθηματική συνάρτηση που δέχεται ως είσοδο οποιαδήποτε τιμή και επιστρέφει +1 σε περίπτωση που η είσοδος είναι θετικός αριθμός, ενώ στην αντίθετη περίπτωση επιστρέφει -1.

Στόχος του SVM αλγορίθμου είναι η νοητή αυτή γραμμή (hyperplane) που χωρίζει τα θετικά από τα αρνητικά παραδείγματα να έχει όσο το δυνατόν μεγαλύτερο περιθώριο (margin). Ως περιθώριο ορίζεται η απόσταση των δύο πιο κοντινών παραδειγμάτων που ανήκουν σε κάθε κλάση. Για να επιτευχθεί αυτό πρέπει να ελαχιστοποιηθεί η

Ευκλείδεια νόρμα του \mathbf{w} (συμβολίζεται $\|\mathbf{w}\|$) και δίνεται από τον τύπο: $\sqrt{\sum_{j=1}^D (w^{(j)})^2}$ (12).

Το κριτήριο βελτιστοποίησης που επιλύει το παραπάνω πρόβλημα συνοπτικά εκφράζεται ως: Minimize $\|\mathbf{w}\|$ subject to $y_i(\mathbf{w}\mathbf{x}_i - b) \geq 1$ for $i = 1, \dots, N$.

2.10.2.5 k-Nearest Neighbors – k-NN

Ο αλγόριθμος k-NN είναι ένας μη παραμετρικός αλγόριθμος μάθησης. Σε αντίθεση με τους άλλους αλγόριθμους μάθησης οι οποίοι παραμερίζουν τα training data αφότου δημιουργηθεί το μοντέλο, ο k-NN κρατάει όλα τα training data στην μνήμη. Έτσι, όταν δεχτεί ως είσοδο ένα καινούριο παράδειγμα \mathbf{x} , ο αλγόριθμος βρίσκει k παραδείγματα κοντινά του \mathbf{x} που έχουν γίνει ήδη train και επιστρέφει την κυρίαρχη κλάση (σε περίπτωση classification) ή την μέση κλάση (σε περίπτωση regression).

Η στενότητα των δύο σημείων δίνεται από μία συνάρτηση απόστασης. Μία συνάρτηση απόστασης που χρησιμοποιείται συχνά είναι η συνάρτηση αρνητικής συνημιτονικής ομοιότητας (negative cosine similarity). Η συνημιτονική ομοιότητα ορίζεται ως:

$$s(\mathbf{x}_i, \mathbf{x}_k) \stackrel{\text{def}}{=} \cos(\angle(\mathbf{x}_i, \mathbf{x}_k)) = \frac{\sum_{j=1}^D x_i^{(j)} x_k^{(j)}}{\sqrt{\sum_{j=1}^D (x_i^{(j)})^2} \sqrt{\sum_{j=1}^D (x_k^{(j)})^2}} \quad (13)$$

και είναι ένα μέτρο ομοιότητας της κατεύθυνσης δύο διανυσμάτων. Άλλες γνωστές συναρτήσεις απόστασης που χρησιμοποιούνται είναι η Ευκλείδεια απόσταση, η απόσταση Chebychev, η απόσταση Mahalanobis και η απόσταση Hamming. Η επιλογή της συνάρτησης απόστασης καθώς και της παραμέτρου k , γίνονται εκ των προτέρων κατά την ανάλυση των δεδομένων, συνεπώς αναφερόμαστε σε υπερπαραμέτρους.

Τέλος, αξίζει να σημειωθεί πως στον αλγόριθμο αυτό δεν χρησιμοποιείται συνάρτηση κόστους.

2.10.3 Δομικά στοιχεία ενός αλγόριθμου μάθησης

Τα δομικά στοιχεία ενός αλγόριθμου εκμάθησης είναι:

1. Η συνάρτηση απώλειας (loss function)
2. Ένα κριτήριο βελτιστοποίησης βασισμένο στη loss function (π.χ. συνάρτηση κόστους)
3. Ένα πρόβλημα βελτιστοποίησης που αξιοποιεί τα δεδομένα εκπαίδευσης (training data) για να βρει μια λύση στο κριτήριο βελτιστοποίησης

Ορισμένοι αλγόριθμοι (όπως Linear Regression, Logistic Regression, SVM) είναι σχεδιασμένοι να βελτιστοποιούν ρητά ένα συγκεκριμένο κριτήριο. Άλλοι αλγόριθμοι (όπως οι αλγόριθμοι Δέντρων Απόφασης, k-NN) βελτιστοποιούν το κριτήριο «σιωπηρά». Συγκεκριμένα, οι αλγόριθμοι decision tree learning και k-NN οι οποίοι συγκαταλέγονται μεταξύ των παλαιότερων αλγορίθμων μηχανικής μάθησης και αναπτύχθηκαν πειραματικά με βάση τη διαίσθηση, κατασκευάστηκαν χωρίς συγκεκριμένο κριτήριο βελτιστοποίησης. Τα κριτήρια βελτιστοποίησης αναπτύχθηκαν αργότερα για να εξηγήσουν το πως αυτοί οι αλγόριθμοι λειτουργούν.

Στη σύγχρονη βιβλιογραφία συχνά συναντάμε τον όρο Gradient Descent (αλγόριθμος απότομης καθόδου) ή Stochastic Gradient Descent (στοχαστικός αλγόριθμος απότομης καθόδου). Αυτοί οι δύο αλγόριθμοι αποτελούν δύο από τους πιο συχνά χρησιμοποιούμενους αλγορίθμους βελτιστοποίησης κριτηρίου.

Ο Gradient Descent είναι ένας επαναληπτικός αλγόριθμος βελτιστοποίησης για την εύρεση του ελαχίστου μίας συνάρτησης. Για να βρούμε ένα τοπικό ελάχιστο μίας συνάρτησης χρησιμοποιώντας τον Gradient Descent, ξεκινάμε από ένα τυχαίο σημείο και κάνουμε βήματα ανάλογα με το αρνητικό της κλίσης της συνάρτησης στο τρέχον σημείο.

Ο αλγόριθμος αυτός μπορεί να εφαρμοστεί για την εύρεση βέλτιστων παραμέτρων σε αλγόριθμους όπως linear και logistic regression, σε SVM καθώς επίσης και στα νευρωνικά δίκτυα. Για πολλά μοντέλα, όπως το logistic regression και το SVM, το κριτήριο βελτιστοποίησης εκφράζεται μέσω μίας κυρτής συνάρτησης (convex function). Οι κυρτές συναρτήσεις έχουν ένα μόνο ελάχιστο, το ολικό ελάχιστο. Στα νευρωνικά δίκτυα, αν και το κριτήριο βελτιστοποίησης δεν είναι κυρτή συνάρτηση, πολλές φορές αρκεί να βρεθεί ένα τοπικό ελάχιστο.

2.10.4 Λειτουργία αλγορίθμου Gradient Descent

Χρησιμοποιώντας ένα παράδειγμα αλγορίθμου linear regression θα περιγραφεί η λειτουργία του αλγορίθμου gradient descent.

Θυμίζουμε πως το μοντέλο linear regression είναι το: $f_{w,b}(\mathbf{x}) = \mathbf{w}\mathbf{x} + b$.

Συνεπώς, το κριτήριο βελτιστοποίησης περιέχει δύο παραμέτρους (w και b).

Στόχος είναι να βρούμε τις βέλτιστες παραμέτρους μέσω της συνάρτησης κόστους:

$$l = \frac{1}{N} \sum_{i=1}^N (y_i - (wx_i + b))^2 \quad (14).$$

Ο αλγόριθμος gradient descent ξεκινά να υπολογίζει τη μερική παράγωγο για κάθε παράμετρο:

$$\begin{aligned} \frac{\partial l}{\partial w} &= \frac{1}{N} \sum_{i=1}^N -2x_i(y_i - (wx_i + b)) \\ \frac{\partial l}{\partial b} &= \frac{1}{N} \sum_{i=1}^N -2(y_i - (wx_i + b)) \end{aligned} \quad (15).$$

Δίνονται αρχικές τιμές w_0 και b_0 και πραγματοποιείται επαναληπτικά η διαδικασία για κάθε παράδειγμα εισόδου ενημερώνοντας τις τιμές τους χρησιμοποιώντας τις τιμές που προκύπτουν από τις μερικές παραγώγους. Ο συντελεστής μάθησης α το μέγεθος της ενημέρωσης:

$$\begin{aligned} w_i &\leftarrow \alpha \frac{-2x_i(y_i - (w_{i-1}x_i + b_{i-1}))}{N} \\ b_i &\leftarrow \alpha \frac{-2(y_i - (w_{i-1}x_i + b_{i-1}))}{N} \end{aligned} \quad (16).$$

Ένας κύκλος εκπαίδευσης από όλα τα παραδείγματα ονομάζεται epoch. Τυπικά, χρειάζονται πολλές τέτοιες επαναλήψεις, δηλαδή epochs, μέχρις ότου οι τιμές w και b να αλλάζουν κατ' ελάχιστο, οπότε ο αλγόριθμος σταματάει.

Να σημειωθεί πως για την επιλογή της παραμέτρου α του αλγόριθμου gradient descent χρειάζεται ιδιαίτερη προσοχή. Επιπρόσθετα, ο αλγόριθμος αυτός καθίσταται αργός για μεγάλα dataset.

Μία εκδοχή του αλγόριθμου που περιγράφεται είναι ο Stochastic Gradient Descent (SGD), ο οποίος επιταχύνει τον υπολογισμό της κλίσης χρησιμοποιώντας μικρότερα υποδιαστήματα (batches) του training set. Διάφορες εκδοχές του στοχαστικού αλγόριθμου που συναντάμε συχνά είναι οι: Adagard, Momentum, RMSprop και Adam.

Στο σημείο αυτό να αναφερθεί πως ο αλγόριθμος gradient descent δεν αποτελεί αλγόριθμο μηχανικής μάθησης. Επί της ουσίας είναι αλγόριθμος επίλυσης προβλημάτων στα οποία η συνάρτηση ελαχιστοποίησης παρουσιάζει απότομη κλίση.

2.10.5 Εκτίμησης απόδοσης μοντέλου

Αφότου ένας αλγόριθμος εκμάθησης δημιουργήσει το μοντέλο, ιδιαίτερα σημαντική είναι η αξιολόγηση του μοντέλου. Για το λόγο αυτό χρησιμοποιούμε το test set.

Το test set περιέχει τα παραδείγματα που ο αλγόριθμος εκμάθησης δεν έχει ξαναδεί, οπότε αν το μοντέλο έχει καλή απόδοση στην πρόβλεψη των κλάσεων των παραδειγμάτων από το test set, λέμε ότι το μοντέλο γενικεύεται καλά.

Για μεγαλύτερη ακρίβεια, χρησιμοποιείται πληθώρα εργαλείων και μετρικών για την αξιολόγηση των αλγορίθμων μηχανικής μάθησης.

Για μοντέλα regression η αξιολόγηση του μοντέλου είναι αρκετά εύκολη, καθότι ένα well-fitting regression μοντέλο προβλέπει τιμές πολύ κοντά στις τιμές που έχουν παρατηρηθεί. Επομένως, η προσαρμογή ενός υπό αξιολόγηση regression μοντέλου θα πρέπει να είναι καλύτερη από την προσαρμογή του «μέσου» μοντέλου (mean model) το οποίο πάντα προβλέπει τον μέσο όρο των κλάσεων στο training set και θα χρησιμοποιούταν εάν δεν υπήρχαν πληροφοριακά χαρακτηριστικά. Εάν συμβεί αυτό, το επόμενο βήμα είναι η σύγκριση των επιδόσεων του μοντέλου στα training και test data.

Για να συμβεί αυτό, υπολογίζεται το μέσο τετραγωνικό σφάλμα (mean squared error - MSE) ξεχωριστά για τα δεδομένα του training set και ξεχωριστά του test set. Εάν το MSE που υπολογίστηκε για τα training data είναι σημαντικά υψηλότερο από το αντίστοιχο των test data, τότε υπάρχουν ενδείξεις overfitting.

Για μοντέλα classification, η αξιολόγηση είναι πιο περίπλοκη. Οι μετρικές που χρησιμοποιούνται πιο συχνά είναι:

- Ο πίνακας σύγχυσης (confusion matrix)
- Η ακρίβεια (accuracy)
- Η ακρίβεια ως προς το κόστος (cost-sensitive accuracy)
- Η ακρίβεια/ανάκληση (precision/recall)
- Η περιοχή κάτω από την καμπύλη ROC

Πίνακας σύγχυσης (confusion matrix)

Ο πίνακας σύγχυσης είναι ένας πίνακας στον οποίο συνοψίζεται πόσο επιτυχημένο είναι το classification μοντέλο στις προβλέψεις που έκανε. Ο ένας άξονας του πίνακα αποτελείται από τις κλάσεις που πρόβλεψε το μοντέλο και ο άλλος άξονας από τις πραγματικές ετικέτες. Σε ένα δυαδικό πρόβλημα classification, υπάρχουν δύο κλάσεις. Παρακάτω παρατίθεται ένα παράδειγμα πίνακα σύγχυσης, χρησιμοποιώντας το παράδειγμα ταξινόμησης των email.

	spam (predicted)	not_spam (predicted)
spam (actual)	23 (TP)	1 (FN)
not_spam (actual)	12 (FP)	556 (TN)

Το παραπάνω παράδειγμα περιλαμβάνει 24 emails τα οποία ήταν όντως spam, εκ των οποίων το μοντέλο έκανε ορθή πρόβλεψη για τα 23 και λανθασμένη για 1. Επομένως έχουμε 23 true positives (TP = 23) και 1 false negative (FN = 1). Αντίστοιχα, για τα not_spam email έχουμε 12 false positive (FP = 12) και 556 true negative (TN = 556).

Αντίστοιχα, ένας πίνακας σύγχυσης για multiclass classification προβλήματα αποτελείται από πολλές σειρές και στήλες, ανάλογα με τον αριθμό των κλάσεων. Η διαδικασία αυτή βοηθάει στον εντοπισμό μοτίβων λαθών.

Ο πίνακας σύγχυσης χρησιμοποιείται επίσης και για τον υπολογισμό των μετρικών precision και recall.

Precision/Recall

Οι μετρικές που χρησιμοποιούνται πιο συχνά για την αξιολόγηση της απόδοσης ενός μοντέλου είναι οι μετρικές precision και recall.

Η μετρική recall είναι ο λόγος των true positives, δηλαδή των σωστών θετικών προβλέψεων, προς το συνολικό αριθμό των θετικών προβλέψεων του test set.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (17).$$

Η μετρική precision είναι ο λόγος των true positives, δηλαδή των σωστών θετικών προβλέψεων, προς το συνολικό αριθμό των θετικών προβλέψεων.

$$\text{Precision} = \frac{TP}{TP+} \quad (18).$$

Για την κατανόηση της σημασίας των παραπάνω μετρικών για την αξιολόγηση του μοντέλου, είναι χρήσιμο να σκεφτούμε το πρόβλημα της έρευνας εγγράφων σε μία βάση δεδομένων χρησιμοποιώντας ένα query. Η μετρική precision είναι η αναλογία των σχετικών εγγράφων στη λίστα όλων των εγγράφων που επιστρέφει η αναζήτηση. Η μετρική recall είναι ο λόγος των σχετικών εγγράφων που επιστρέφονται από την αναζήτηση προς τον συνολικό αριθμό των σχετικών εγγράφων που θα μπορούσαν να είχαν επιστραφεί.

Σχεδόν πάντα τίθεται το δίλλημα της επιλογής μεταξύ μιας υψηλότερης τιμής precision και μιας υψηλότερης τιμής recall. Πρακτικά, είναι ανέφικτο να επιτευχθούν και τα δύο. Δύναται η επίτευξη κάποιας από τις δύο μετρικής με διάφορα μέσα, όπως είτε με ανάθεση μεγαλύτερων βαρών στις κλάσεις, είτε ρυθμίζοντας τις υπερπαραμέτρους ώστε να μεγιστοποιείται η τιμή είτε precision είτε η τιμή recall στο validation set, είτε μεταβάλλοντας το κατώφλι απόφασης για αλγόριθμους που επιστρέφουν πιθανότητες κλάσεων.

Παρά τον ορισμό των μετρικών precision και recall για δυαδικά προβλήματα classification, μπορούν να χρησιμοποιηθούν και για την αξιολόγηση μοντέλων με παραπάνω κλάσεις αφού πρώτα επιλεγεί η κλάση που θέλουμε να αξιολογηθεί. Εν συνεχεία ορίζονται όλα τα παραδείγματα της κλάσης που επιλέχθηκε ως θετικά και όλα τα υπόλοιπα των εναπομεινάντων κλάσεων ως αρνητικά.

Accuracy

Η μετρική accuracy (ακρίβεια) ορίζεται από το λόγο του πλήθους των ορθώς ταξινομημένων παραδειγμάτων ως προς τον συνολικό αριθμό των ταξινομημένων παραδειγμάτων.

$$\text{Accuracy} = \frac{TP+T}{TP+TN+FP+FN} \quad (19).$$

Η μετρική accuracy είναι μία πολύ χρήσιμη μετρική ειδικά όταν τα σφάλματα στην πρόβλεψη όλων των κλάσεων είναι εξίσου σημαντικά.

Cost-Sensitive Accuracy

Η cost-sensitive accuracy είναι ιδιαίτερα χρήσιμη σε περιπτώσεις που διαφορετικές κλάσεις έχουν διαφορετική σημαντικότητα. Για τον υπολογισμό της cost-sensitive accuracy αρχικά αναθέτεται ένα κόστος (η τιμή του κόστους είναι θετική) και στις δύο λάθος προβλέψεις (FP και FN). Εν συνεχεία γίνεται ο υπολογισμός των μετρικών TP, TN, FP και FN ως συνήθως και πολλαπλασιάζονται οι τιμές FP και FN με το αντίστοιχο κόστος προτού υπολογιστεί η accuracy.

Περιοχή κάτω από την καμπύλη ROC (Area under the ROC Curve - AUC)

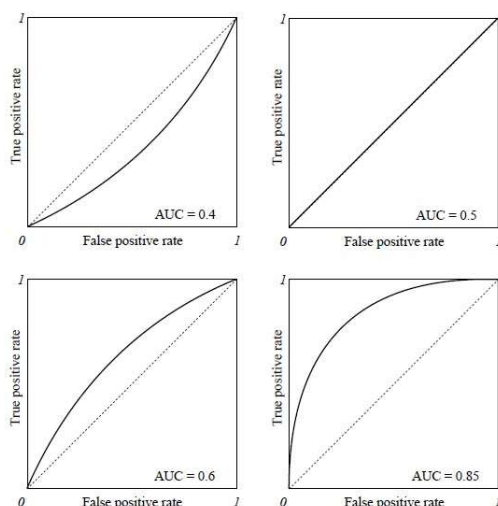
Η καμπύλη ROC (ROC σημαίνει «receiver operating characteristic» και ο όρος προέρχεται από τους μηχανικούς των ραντάρ) είναι μία τεχνική που χρησιμοποιείται συχνά για την αξιολόγηση της απόδοσης classification μοντέλων. Η καμπύλη ROC χρησιμοποιεί έναν συνδυασμό του ρυθμού σωστών θετικών προβλέψεων (όπως ορίστηκε στην recall) και ρυθμού λανθασμένων θετικών ώστε να δημιουργήσει μία συνολική εικόνα απόδοσης του classification.

Ο ρυθμός σωστών θετικών προβλέψεων (true positive rate – TPR) και ο ρυθμός λανθασμένων θετικών προβλέψεων (false positive rate - FPR) ορίζεται ως:

$$\text{TPR} = \frac{\text{TP}}{(\text{TP}+\text{FN})} \quad (20) \quad \text{και} \quad \text{FPR} = \frac{\text{FP}}{(\text{FP}+\text{TN})} \quad (21).$$

Η καμπύλη ROC μπορεί να χρησιμοποιηθεί μόνο για την αξιολόγηση ταξινομητών που επιστρέφουν κάποια διαστήματα εμπιστοσύνης (ή πιθανότητα) πρόβλεψης. Για παράδειγμα, τα logistic regression μοντέλα, τα νευρωνικά δίκτυα, τα δέντρα αποφάσεων (καθώς και τα ensemble μοντέλα που βασίζονται στα δέντρα αποφάσεων) μπορούν να αξιολογηθούν με χρήση καμπύλης ROC.

Για το σχεδιασμό τη καμπύλης, αρχικά διακριτοποιείται το εύρος του διαστήματος εμπιστοσύνης και στην συνέχεια χρησιμοποιείται κάθε διακριτή τιμή ως κατώφλι πρόβλεψης και γίνεται πρόβλεψη των κλάσεων των παραδειγμάτων στο σύνολο δεδομένων χρησιμοποιώντας το μοντέλο και αυτό το όριο. Ένα παράδειγμα αυτής της τεχνικής παρατίθεται στο παρακάτω γράφημα:



Εικόνα 27: Παράδειγμα καμπύλης ROC

Όπως φαίνεται και στο παραπάνω γράφημα, εάν το κατώφλι μας ισούται με μηδέν, όλες μας οι προβλέψεις θα είναι θετικές, συνεπώς οι τιμές TPR και FPR θα είναι ίσες

με 1. Αντιθέτως, εάν το κατώφλι μας ισούται με 1, δεν θα έχουμε θετικές προβλέψεις και τότε οι τιμές TPR και FPR θα παίρνουν μηδενικές τιμές. Όπως γίνεται κατανοητό, όσο πιο ψηλά είναι η περιοχή κάτω από την καμπύλη, τόσο καλύτερη είναι η ταξινόμηση.

2.10.6 Underfitting και Overfitting

Αναφέρθηκε νωρίτερα ο όρος bias. Σύμφωνα με τον όρο αυτό, ένα μοντέλο έχει χαμηλό bias εάν προβλέπει σωστά τις ετικέτες του training set. Όταν το μοντέλο κάνει αρκετά λάθη στις προβλέψεις του, τότε έχει υψηλό bias και το μοντέλο παρουσιάζει το φαινόμενο του Underfitting.

Ο όρος underfitting αναφέρεται στην αδυναμία του μοντέλου να προβλέψει σωστά τις κλάσεις των δεδομένων στα οποία έχει εκπαιδευτεί. Υπάρχουν πολλοί λόγοι που συμβαίνει αυτό με τους κυριότερους να είναι είτε διότι το μοντέλο είναι πολύ απλό για τη φύση των δεδομένων (για παράδειγμα ένα γραμμικό μοντέλο συχνά παρουσιάζει underfitting) είτε τα χαρακτηριστικά που επεξεργάστηκαν δεν εμπεριέχουν αρκετή πληροφορία για επεξεργασία.

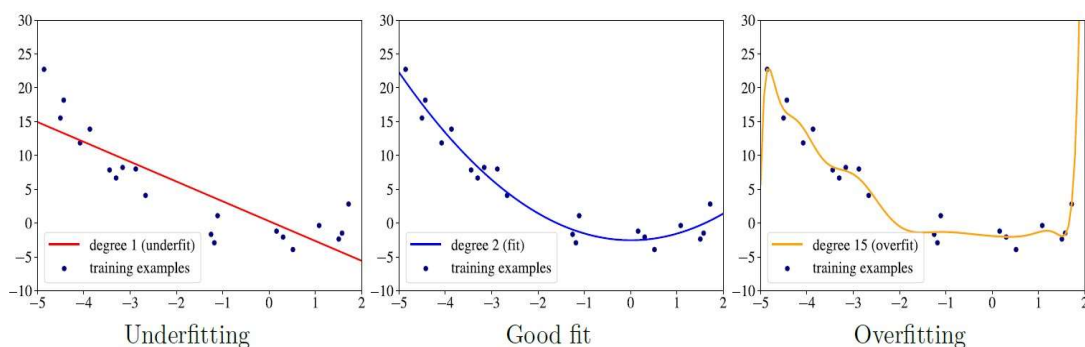
Για να αντιμετωπιστεί το πρόβλημα του underfitting συνίσταται είτε η χρήση ενός πιο πολύπλοκου μοντέλου, είτε η καλύτερη επεξεργασία των δεδομένων ώστε να αυξηθεί η προγνωστική δύναμή τους.

Ένα άλλο πρόβλημα που μπορεί να συναντήσει ένα μοντέλο είναι το λεγόμενο Overfitting. Κατά το φαινόμενο αυτό, το μοντέλο ενώ εξάγει πολύ σωστές προβλέψεις για τα δεδομένα του training σετ, δεν πραγματοποιεί τις ίδιες προβλέψεις σε τουλάχιστον ένα από τα δύο εναπομείναντα σετ. Πολλοί λόγοι μπορούν να οδηγήσουν σε overfitting, με τους κυριότερους να είναι είτε ένα ιδιαίτερα πολύπλοκο μοντέλο ως προς τη φύση των δεδομένων (όπως ένα πολύ «ψηλό» δέντρο αποφάσεων ή ένα πολύ «βαθύ» νευρωνικό δίκτυο) είτε ένα dataset με πολλά χαρακτηριστικά για μικρό πλήθος δεδομένων για εκπαίδευση.

Επιπρόσθετα, ένας ακόμα λόγος που οδηγεί σε overfitting είναι και η μεγάλη διακύμανση. Η διακύμανση είναι στατιστικός όρος και στην ουσία αφορά το σφάλμα του μοντέλου εξαιτίας των μικρών διακυμάνσεων που παρουσιάζουν οι τιμές του training set.

Υπάρχουν αρκετές τεχνικές για τη μείωση του overfitting, όπως η χρήση πιο απλών μοντέλων, η μείωση των διαστάσεων των παραδειγμάτων του dataset, η προσθήκη δεδομένων προς εκπαίδευση αν είναι εφικτό, καθώς και η κανονικοποίηση (Regularization) του μοντέλου.

Στο παρακάτω γράφημα παρουσιάζονται με γραφικό τρόπο τα φαινόμενα που περιγράφηκαν:



Εικόνα 28: Περιπτώσεις fitting

Regularization

Η κανονικοποίηση (Regularization) αποτελεί την πιο ευρέως χρησιμοποιούμενη προσέγγιση για την αποφυγή του προβλήματος του overfitting.

Ακόμα και τα πιο απλά μοντέλα, όπως τα γραμμικά ενδέχεται να παρουσιάσουν το φαινόμενο του data overfitting. Όπως ήδη αναφέρθηκε, αυτό συμβαίνει όταν τα δεδομένα είναι πολυδιάστατα και το μέγεθος του training σετ μικρό. Στην πραγματικότητα, όταν το χαρακτηριστικό διάνυσμα είναι πολυδιάστατο, ο γραμμικός αλγόριθμος μάθησης μπορεί να φτιάξει ένα μοντέλο που να εκχωρεί μη μηδενικές τιμές στις περισσότερες διαστάσεις $w^{(j)}$ στην παράμετρο διανύσματος w προσπαθώντας να βρει μία πολύ πολύπλοκη σχέση μεταξύ όλων των διαθέσιμων χαρακτηριστικών που είναι διαθέσιμα για να προβλέψει βέλτιστα την ετικέτα του παραδείγματος.

Ένα τόσο πολύπλοκο μοντέλο είναι πολύ πιθανό να κάνει μία «φτωχή» πρόβλεψη για τις ετικέτες των παραδειγμάτων των hold-out sets. Αυτό συμβαίνει διότι, προσπαθώντας να κάνει την βέλτιστη πρόβλεψη το μοντέλο θα έχει λάβει υπόψη και όλες τις ατέλειες του dataset, όπως είναι ο θόρυβος των δεδομένων, οι ατέλειες στη δειγματοληψία λόγω του μικρού μεγέθους του dataset καθώς και άλλα εξωγενή χαρακτηριστικά. Οι δύο πιο γνωστές τεχνικές regularization είναι οι λεγόμενες L1 regularization και L2 regularization.

Η τεχνική L1 στην πράξη δημιουργεί ένα «αραιό» μοντέλο (sparse model) του οποίου οι περισσότερες παράμετροι έχουν μηδενική τιμή. Έτσι, η L1 κάνει την επιλογή των χαρακτηριστικών αποφασίζοντας ποια χαρακτηριστικά είναι απαραίτητα για την πρόβλεψη και ποια όχι. Αυτό μπορεί να είναι χρήσιμο στην περίπτωση που χρειάζεται να αυξηθεί η επεξηγησιμότητα του μοντέλου. Ωστόσο, εάν ο στόχος είναι η μεγιστοποίηση της απόδοσης του μοντέλου στα δεδομένα των hold-out sets, τότε η τεχνική L2 συνήθως δίνει καλύτερα αποτελέσματα. Επίσης, η L2 έχει το πλεονέκτημα ότι είναι διαφοροποιήσιμο και έτσι μπορούμε να χρησιμοποιήσουμε gradient descent για τη βελτιστοποίηση της αντικειμενικής συνάρτησης (objective function).

Επιπρόσθετα, εκτός από τα γραμμικά μοντέλα, οι προσεγγίσεις L1 και L2 χρησιμοποιούνται πολύ συχνά και στα νευρωνικά δίκτυα με στόχο την ελαχιστοποίηση της objective function. Εκτός από τις τεχνικές L1 και L2, τα νευρωνικά δίκτυα επωφελούνται και από άλλες τεχνικές regularization, όπως η dropout και batch-

normalization, καθώς και από μη μαθηματικές μεθόδους όπως η data augmentation και early stopping.

2.11 Νευρωνικά Δίκτυα

Η βάση των Νευρωνικών Δικτύων (Neural Networks - NN) είναι ο αλγόριθμος logistic regression. Στην πραγματικότητα, το logistic regression μοντέλο και η γενίκευσή του για την ταξινόμηση πολλών κλάσεων (ονομάζεται και softmax regression model), αποτελεί μία τυπική μονάδα σε ένα νευρωνικό δίκτυο.

Ένα νευρωνικό δίκτυο, όπως και η παλινδρόμηση (regression) ή ένα SVM μοντέλο, είναι μία μαθηματική συνάρτηση:

$$y = f_{NN}(\mathbf{x}) \quad (22).$$

Η συνάρτηση f_{NN} έχει μία συγκεκριμένη μορφή, είναι μία εμφωλευμένη συνάρτηση (τα λεγόμενα επίπεδα ενός νευρωνικού δικτύου). Επομένως ένα νευρωνικό δίκτυο τριών επιπέδων προκύπτει από μία βαθμωτή συνάρτηση της μορφής:

$$y = f_{NN}(\mathbf{x}) = f_3(f_2(f_1(\mathbf{x}))) \quad (23).$$

Στην παραπάνω εξίσωση, οι συναρτήσεις f_2 και f_3 είναι διανυσματικές συναρτήσεις της μορφής:

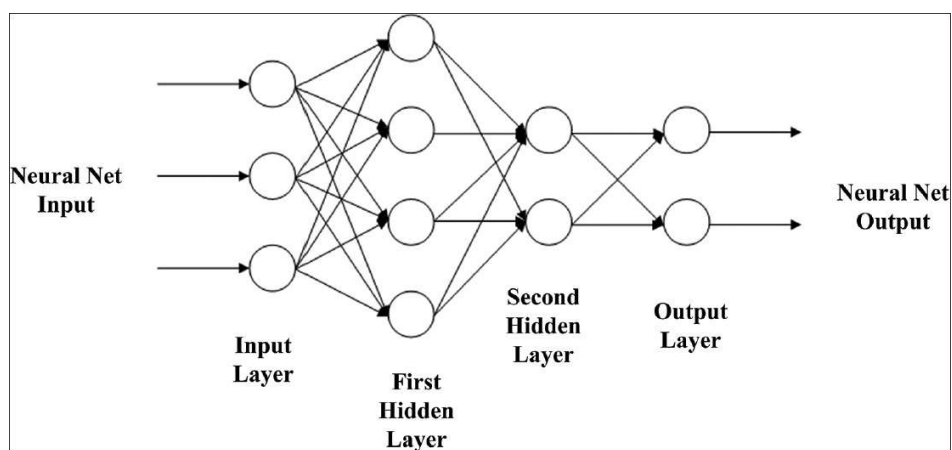
$$f_l(\mathbf{z}) \stackrel{\text{def}}{=} g_l(\mathbf{W}_l \mathbf{z} + \mathbf{b}_l) \quad (24).$$

όπου ο δείκτης l ονομάζεται δείκτης επιπέδου και μπορεί να πάρει τιμή από ένα μέχρι όσα είναι τα επίπεδα του νευρωνικού δικτύου. Η συνάρτηση g_l ονομάζεται συνάρτηση ενεργοποίησης και είναι μία σταθερή, συνήθως μη γραμμική συνάρτηση. Οι παράμετροι \mathbf{W}_l (πίνακας) και \mathbf{b}_l (διάνυσμα) προσδιορίζονται για κάθε επίπεδο χρησιμοποιώντας τον αλγόριθμο gradient descent μέσω βελτιστοποίησης μίας συνάρτησης κόστους (όπως το MSE). Η συνάρτηση f_l είναι μία βαθμωτή συνάρτηση για ένα regression πρόβλημα, αλλά μπορεί να είναι και μία διανυσματική συνάρτηση, αναλόγως το πρόβλημα. Ο λόγος που η παράμετρος \mathbf{W}_l είναι πίνακας και όχι διάνυσμα, έγκειται στο γεγονός πως η παράμετρος g_l είναι διανυσματική συνάρτηση. Για την καλύτερη κατανόηση της ιδιότητας αυτής θα περιγραφεί μια αρχιτεκτονική

νευρωνικών δικτύων που ονομάζεται multilayer perceptron (συχνά αναφέρεται και ως vanilla neural network).

2.11.1 Νευρωνικά Δίκτυα πολλών επιπέδων (Multilayer Perceptron)

Μία ευρέως διαδεδομένη διαμόρφωση νευρωνικών δικτύων η αρχιτεκτονική FFNN (Feed-Forward Neural Networks) και πιο συγκεκριμένα η αρχιτεκτονική Multilayer Perceptron (MLP). Κατά την αρχιτεκτονική αυτή, τα νευρωνικά δίκτυα διαθέτουν περισσότερα από ένα εσωτερικά επίπεδα, δηλαδή επίπεδα εκτός από το επίπεδο εισόδου και το εξόδο. Το νευρωνικό δίκτυο μπορεί να πάρει ως είσοδο ένα χαρακτηριστικό διάνυσμα πολλών διαστάσεων και ως έξοδο υπολογίζει έναν αριθμό. Τα FFNN δίκτυα μπορεί να είτε regression είτε classification μοντέλα, ανάλογα με τη συνάρτηση ενεργοποίησης που χρησιμοποιείται στο επίπεδο εξόδου. Μία τέτοια αρχιτεκτονική MLP περιγράφεται στο παρακάτω διάγραμμα:



Εικόνα 29: Αρχιτεκτονική MLP

Το νευρωνικό δίκτυο αναπαρίσταται γραφικά ως ένας συνδυασμός συνδεδεμένων μονάδων (units), λογικά οργανωμένων σε επίπεδα. Κάθε μονάδα μπορεί να αναπαρασταθεί είτε ως κύκλος είτε ως ορθογώνιο. Το εισερχόμενο βέλος αντιπροσωπεύει μια είσοδο μιας μονάδας και υποδεικνύει από πού προήλθε αυτή η είσοδος, ενώ το εξερχόμενο βέλος υποδεικνύει την έξοδο μιας μονάδας. Επίσης, η έξοδος κάθε μονάδας είναι το αποτέλεσμα της μαθηματικής λειτουργίας που εκτελείται στην εκάστοτε μονάδα. Οι μονάδες στο επίπεδο εισόδου δεν εκτελούν κάποια μαθηματική λειτουργία, παρά μόνο προωθούν τα δεδομένα από την είσοδό τους στην έξοδό τους.

Κατά την λειτουργία της αρχιτεκτονικής αυτής, αρχικά τα δεδομένα εισόδου κάθε μονάδας συνδέονται ώστε να δημιουργηθεί το διάνυσμα εισόδου. Στη συνέχεια, η μονάδα εφαρμόζει έναν γραμμικό μετασχηματισμό στο διάνυσμα εισόδου, όπως ακριβώς το μοντέλο της γραμμικής παλινδρόμησης με το χαρακτηριστικό διάνυσμα. Τέλος, η μονάδα εφαρμόζει μία συνάρτηση ενεργοποίησης g_i στο αποτέλεσμα του γραμμικού μετασχηματισμού και λαμβάνει την τιμή εξόδου, δηλαδή έναν πραγματικό

αριθμό. Σε ένα vanilla FFNN η τιμή εξόδου μιας μονάδας ενός επιπέδου γίνεται τιμή εισόδου καθεμιάς από τις μονάδες του επόμενου επιπέδου.

Η συνάρτηση ενεργοποίησης g_l εμπεριέχει τον δείκτη l , ο οποίος είναι ο δείκτης του επιπέδου στο οποίο ανήκει η μονάδα. Συνήθως, όλες οι μονάδες ενός επιπέδου χρησιμοποιούν την ίδια συνάρτηση ενεργοποίησης, κάτι το οποίο δεν είναι πάντα απαραίτητο. Κάθε επίπεδο μπορεί να έχει διαφορετικό πλήθος μονάδων. Κάθε μονάδα διαθέτει τις δικές της παραμέτρους $w_{l,u}$ και $b_{l,u}$, όπου u είναι ο δείκτης της μονάδας και l ο δείκτης του επιπέδου. Το διάνυσμα \mathbf{y}_{l-1} κάθε μονάδας ορίζεται ως $[\mathbf{y}_{l-1}^{(1)}, \mathbf{y}_{l-1}^{(2)}, \dots, \mathbf{y}_{l-1}^{(size_l)}]$, ενώ το διάνυσμα εισόδου x ως $[x^{(1)}, x^{(2)}, \dots, x^{(D)}]$.

Όπως φαίνεται και στην εικόνα, στην αρχιτεκτονική MLP όλες οι εξοδοί ενός επιπέδου συνδέονται σε κάθε μονάδα του επόμενου επιπέδου. Η αρχιτεκτονική αυτή ονομάζεται πλήρης συνδεσιμότητα (fully connected). Ένα νευρικό δίκτυο μπορεί να περιέχει πλήρως συνδεδεμένα επίπεδα. Αυτά είναι τα επίπεδα των οποίων οι μονάδες λαμβάνουν ως εισόδους τις εξόδους καθεμιάς από τις μονάδες του προηγούμενου επιπέδου.

2.11.2 Αρχιτεκτονική Feed-Forward (πρόσθιας τροφοδότησης)

Για την επίλυση regression ή classification προβλημάτων, το τελευταίο (το δεξιότερο) επίπεδο του νευρωνικού δικτύου αποτελείται μόνο από μία μονάδα. Εάν η συνάρτηση ενεργοποίησης g_{last} είναι γραμμική, τότε το νευρωνικό δίκτυο είναι ένα regression μοντέλο. Εάν η συνάρτηση g_{last} είναι λογιστική συνάρτηση, τότε το νευρωνικό δίκτυο είναι μοντέλο δυαδικής ταξινόμησης.

Οποιαδήποτε συνάρτηση $g_{l,u}$ μπορεί να επιλεγεί, αρκεί να είναι παραγωγίσιμη. Η τελευταία αυτή είναι απαραίτητη για τον αλγόριθμο gradient descent, ο οποίος χρησιμοποιείται για την εύρεση των τιμών των παραμέτρων $w_{l,u}$ και $b_{l,u}$ για κάθε l και u . Η ύπαρξη μη γραμμικών στοιχείων στη συνάρτηση f_{NN} επιτρέπει στο νευρωνικό δίκτυο να προσεγγίσει τις μη γραμμικές συναρτήσεις. Χωρίς την ύπαρξη μη γραμμικοτήτων, η συνάρτηση f_{NN} θα ήταν γραμμική, ανεξαρτήτως του πλήθους των επιπέδων που διαθέτει το νευρωνικό δίκτυο. Αυτό οφείλεται στο γεγονός ότι η συνάρτηση $\mathbf{W}_l \mathbf{z} + \mathbf{b}_l$ είναι γραμμική και μία γραμμική συνάρτηση μίας γραμμικής συνάρτησης, είναι επίσης γραμμική.

Για την επιλογή της συνάρτησης ενεργοποίησης προτείνεται η λογιστική συνάρτηση, όπως έχει ήδη περιγραφεί, καθώς και οι συναρτήσεις TanH και ReLU. Η συνάρτηση TanH είναι συνάρτηση υπερβολικής εφαιτομένης, παρόμοια με τη λογιστική συνάρτηση αλλά κυμαίνεται στο ανοιχτό διάστημα (-1,1). Η συνάρτηση ReLU (Rectified Linear Unit) είναι η συνάρτηση ανορθωμένης εισόδου, η οποία ισούται με μηδέν όταν το z είναι αρνητικό ενώ σε αντίθετη περίπτωση παίρνει την τιμή του z :

$$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (25).$$

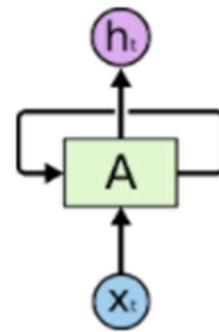
$$\text{relu}(z) = \begin{cases} 0 & \text{if } z < 0 \\ z & \text{otherwise} \end{cases} \quad (26).$$

Σε αντίθεση με τον αλγόριθμο linear regression ($\mathbf{wz} + b$), το στοιχείο \mathbf{W}_l στην έκφραση $\mathbf{W}_l \mathbf{z} + \mathbf{b}_l$ είναι πίνακας ενώ το στοιχείο \mathbf{b}_l είναι διάνυσμα. Στον πίνακα \mathbf{W}_l , κάθε σειρά u αντιστοιχεί σε ένα διάνυσμα της παραμέτρου $w_{l,u}$. Η διάσταση του διανύσματος $w_{l,u}$ ισούται με τον αριθμό των μονάδων στο επίπεδο $l - 1$. Η λειτουργία $\mathbf{W}_l \mathbf{z}$ οδηγεί σε ένα διάνυσμα $\mathbf{a}_l \stackrel{\text{def}}{=} [\mathbf{w}_{l,1} \mathbf{z}, \mathbf{w}_{l,2} \mathbf{z}, \dots, \mathbf{w}_{l, \text{size}_l} \mathbf{z}]$ και το άθροισμα $\mathbf{a}_l + \mathbf{b}_l$ δίνει το διάνυσμα \mathbf{c}_l διαστάσεων l . Τέλος, η συνάρτηση $g_l(\mathbf{c}_l)$ δημιουργεί το διάνυσμα εξόδου $\mathbf{y}_l \stackrel{\text{def}}{=} [y_l^{(1)}, y_l^{(2)}, \dots, y_l^{(\text{size}_l)}]$.

2.11.3 Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks – RNN)

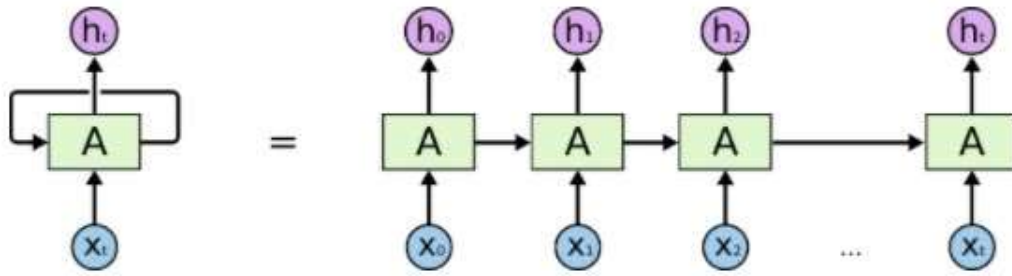
Τα δίκτυα RNN αποτελούν την πιο απλή μορφή αναδρομικών νευρωνικών δικτύων. Σε αντίθεση με τα υπόλοιπα νευρωνικά δίκτυα διαθέτουν βρόχους, γεγονός που τους επιτρέπει την διατήρηση των πληροφοριών καθώς ανατροφοδοτούν την έξοδο τους ως είσοδο [19].

Στο διάγραμμα παρουσιάζεται πως ένα κομμάτι του νευρωνικού δικτύου, A, εξετάζει κάποια είσοδο x_t και εξάγει μια τιμή h_t . Ο βρόχος επιτρέπει τη μετάδοση πληροφοριών από το ένα βήμα του δικτύου στο επόμενο.



Εικόνα 30: Μονάδα RNN

Αυτοί οι βρόχοι κάνουν τα RNN δίκτυα να φαίνονται αρκετά περίπλοκα. Ωστόσο, αποδεικνύεται ότι δεν διαφέρουν τόσο πολύ από ένα κανονικό νευρικό δίκτυο. Ένα RNN δίκτυο μπορεί να θεωρηθεί ως πολλαπλά αντίγραφα του ίδιου δικτύου, το καθένα από τα οποία μεταβιβάζει μία πληροφορία στον επόμενο βρόχο. Αναπτύσσοντας το βρόχο αυτό προκύπτει:

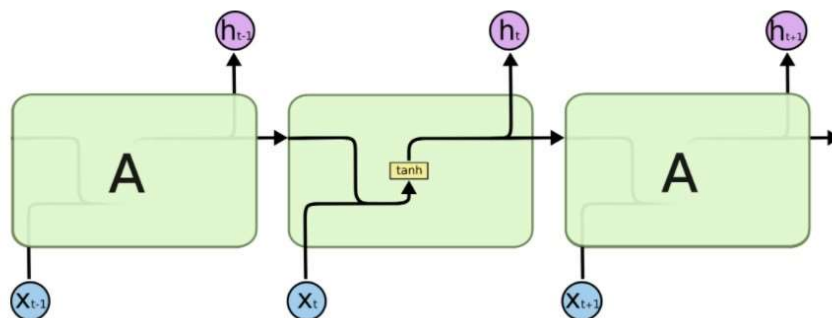


Εικόνα 31: Αλυσωτή διάταξη RNN

Η διάταξη αυτή που μοιάζει με αλυσίδα αποκαλύπτει ότι τα επαναλαμβανόμενα νευρικά δίκτυα σχετίζονται στενά με ακολουθίες και λίστες. Είναι η φυσική αρχιτεκτονική ενός νευρικού δικτύου που χρησιμοποιείται για τέτοιας φύσης δεδομένα. Τα τελευταία χρόνια η εφαρμογή των RNN δικτύων είναι πολύ δημοφιλής σε προβλήματα όπως η αναγνώριση ομιλίας, η μοντελοποίηση της γλώσσας, η μετάφραση, η αναγνώριση εικόνας [20].

Αρχιτεκτονική

Όλα τα αναδρομικά νευρωνικά δίκτυα έχουν τη μορφή μιας αλυσίδας επαναλαμβανόμενων μονάδων νευρικού δικτύου. Στα κλασικά RNN, αυτή η επαναλαμβανόμενη μονάδα διαθέτει μια πολύ απλή δομή, όπως ένα μονό στρώμα tanh. Σύμφωνα με τη δομή αυτή, τα δίκτυα RNN δέχονται την πληροφορία εισόδου η οποία ενώνεται με την πληροφορία εξόδου της προηγούμενης μονάδας του δικτύου και στη συνέχεια, η πληροφορία αυτή περνά από τη συνάρτηση ενεργοποίησης tanh. Έπειτα, η έξοδος της συνάρτησης ενεργοποίησης προωθείται στην έξοδο της μονάδας καθώς και στην είσοδο της επόμενης RNN μονάδας.



Εικόνα 32: Αρχιτεκτονική RNN

Τα αναδρομικά νευρωνικά δίκτυα αποτελούν ιδανική λύση στη διαχείριση προβλημάτων πρόβλεψης δεδομένων σε κοντινά χρονικά βήματα, δηλαδή δεδομένων

με μικρές χρονικές εξαρτήσεις, όπως για παράδειγμα η πρόβλεψη της επόμενης λέξης σε μία πρόταση. Το πρόβλημα αναδρομικών νευρωνικών δικτύων έγκειται στην διαχείριση των μακροπρόθεσμων (long-term) εξαρτήσεων των δεδομένων, δηλαδή εξαρτήσεις δεδομένων που απέχουν αρκετά χρονικά βήματα μεταξύ τους.

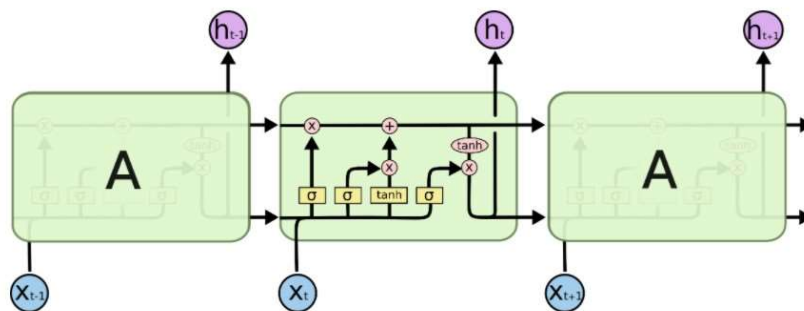
Για την επίλυση τέτοιων προβλημάτων αναπτύχθηκαν τα LSTM δίκτυα, τα οποία αποτελούν ειδική κατηγορία των δικτύων RNN.

2.11.4 Δίκτυα LSTM (Long Short Term Memory)

Τα δίκτυα LSTM παρουσιάστηκαν πρώτη φορά από τους Hochreiter και Schmidhuber [21] και βρίσκουν ευρεία εφαρμογή για την επίλυση προβλημάτων με μακροπρόθεσμες εξαρτήσεις δεδομένων, όπως δεδομένα εικόνας, αναγνώρισης ομιλίας καθώς και προβλήματα χρονοσειρών. Στην πραγματικότητα, τα LSTM είναι σχεδιασμένα για τη διαχείριση τέτοιων προβλημάτων [19].

Αρχιτεκτονική

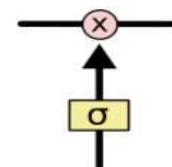
Τα LSTM έχουν και αυτά την ίδια αρχιτεκτονική αλυσίδας, με τη διαφορά ότι η μονάδα που επαναλαμβάνεται έχει διαφορετική εσωτερική αρχιτεκτονική.



Εικόνα 33: Αρχιτεκτονική LSTM

Το κλειδί στα LSTM είναι το κύτταρο κατάστασης (cell state) δηλαδή η άνω οριζόντια γραμμή του διαγράμματος) η οποία διατρέχει όλη την αλυσίδα των μονάδων έχοντας ελάχιστες μικρές αλληλεπιδράσεις. Στο κύτταρο αυτό ρέουν τα δεδομένα σχεδόν αμετάβλητα. Το LSTM έχει τη δυνατότητα να αφαιρεί ή να προσθέτει πληροφορίες στην κατάσταση του κελιού, μέσω προσεκτικής ρύθμισης από τις δομές που ονομάζονται πύλες.

Οι πύλες αποτελούν το μέσο να περάσει προαιρετικά η πληροφορία στο κύτταρο κατάστασης. Αποτελούνται από ένα σιγμοειδή νευρικό επίπεδο και μια λειτουργία πολλαπλασιασμού.



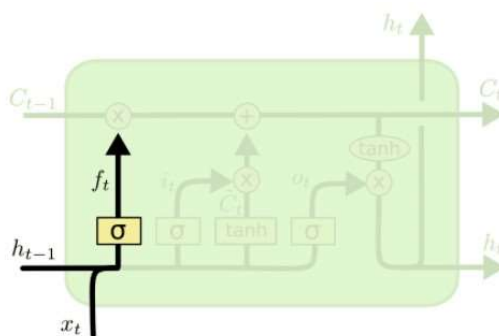
Εικόνα 34: Πύλη LSTM

Το σιγμοειδές επίπεδο εξάγει τιμές μεταξύ μηδέν και ένα, περιγράφοντας ποιο ποσοστό κάθε πληροφορίας πρέπει να περάσει. Όπως είναι προφανές, με την τιμή ένα η πληροφορία περνάει εξ' ολοκλήρου ενώ με την τιμή μηδέν αποκόπτεται. Ένα LSTM έχει τρεις τέτοιες πύλες, για την προστασία και τον έλεγχο της κατάστασης των κυττάρων.

Λειτουργία LSTM

Ο κύκλος λειτουργίας ενός LSTM νευρωνικού δικτύου χωρίζεται σε τέσσερις φάσεις:

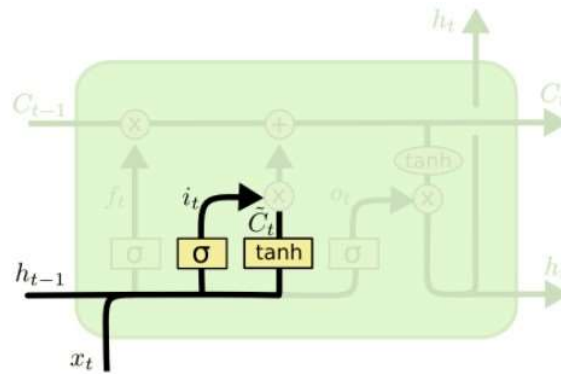
Στην πρώτη φάση το LSTM αποφασίζει ποια πληροφορία θα απορρίψει από το κύτταρο κατάστασης. Η λήψη αυτής της απόφασης πραγματοποιείται από ένα σιγμοειδές επίπεδο που ονομάζεται “forget gate layer”, λαμβάνοντας υπόψη τις τιμές h_{t-1} και x_t και εξάγοντας μία τιμή μεταξύ του μηδέν και του ένα για κάθε κύτταρο κατάστασης C_{t-1} .



Εικόνα 35: 1^ο στάδιο λειτουργίας LSTM

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (27)$$

Στην επόμενη φάση αποφασίζει ποια νέα πληροφορία θα αποθηκεύσει στο κύτταρο κατάστασης, το οποίο πραγματοποιείται σε δύο μέρη. Στο πρώτο μέρος, ένα σιγμοειδές επίπεδο το οποίο ονομάζεται “input gate layer” ποιες τιμές θα ενημερωθούν. Έπειτα, ένα επίπεδο tanh και δημιουργεί ένα διάνυσμα υποψήφιων τιμών C_t που θα μπορούσαν να προστεθούν στην κατάσταση. Στο δεύτερο μέρος, αυτά τα δύο συνδυάζονται και ενημερώνουν την κατάσταση.

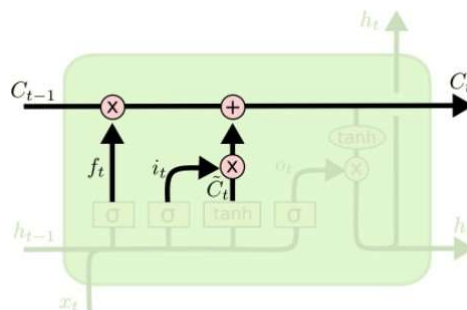


Εικόνα 36: 2^ο στάδιο λειτουργίας LSTM

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (28)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (29)$$

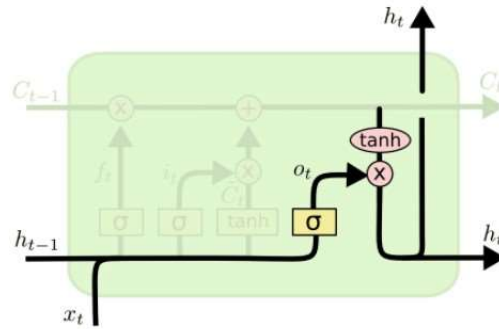
Στην τρίτη φάση ενημερώνεται το νέο κύτταρο κατάστασης C_t από το παλιό C_{t-1} βάσει των αποφάσεων που λήφθηκαν στις πρώτες δύο φάσεις. Έτσι, οι νέες υποψήφιας τιμές που προκύπτουν, κλιμακώνονται ανάλογα με το ποσοστό που αποφασίστηκε να ενημερωθεί η κάθε τιμή κατάστασης



Εικόνα 37: 3^ο στάδιο λειτουργίας LSTM

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (30)$$

Στο τελευταίο στάδιο πραγματοποιείται η λήψη απόφασης για την έξοδο του μοντέλου. Η έξοδος αυτή θα είναι μία φιλτραρισμένη εκδοχή του κυττάρου κατάστασης. Η διαδικασία του φιλτραρίσματος πραγματοποιείται σε δύο βήματα. Αρχικά ένα σιγμοειδές επίπεδο αποφασίζει ποια μέρη του κυττάρου κατάστασης θα εξαχθούν. Εν συνεχεία, την πληροφορία αυτή την επεξεργάζεται μία συνάρτηση tanh (ώστε να θέσει τις τιμές μεταξύ -1 και 1) και πολλαπλασιάζεται με την έξοδο της σιγμοειδούς πύλης.



Εικόνα 38: 4ο στάδιο λειτουργίας LSTM

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (31)$$

$$h_t = o_t * \tanh(C_t) \quad (32)$$

2.12 Deep Learning

Ο όρος Deep Learning (βαθιά μάθηση) αναφέρεται σε νευρωνικά δίκτυα με περισσότερα από δύο εσωτερικά (hidden) επίπεδα. Κατά το παρελθόν η δυσκολία εκπαίδευσης τέτοιων μοντέλων ήταν ανάλογη με τον αριθμό των επιπέδων όσο αυτά αυξάνονταν. Το πρόβλημα της απότομης αύξησης των gradient (exploding gradient) καθώς και της εξαφάνισης (vanishing gradient) αποτελούσαν τις μεγαλύτερες προκλήσεις, καθώς η μέθοδος gradient descent χρησιμοποιείται για τον υπολογισμό των παραμέτρων του νευρωνικού δικτύου. Ενώ το πρόβλημα της απότομης αύξησης μπορούσε εύκολα να αντιμετωπιστεί μέσω τεχνικών όπως η gradient clipping και L1 ή L2 regularization, το πρόβλημα της εξαφάνισης παρέμενε για δεκαετίες μεγάλη πρόκληση.

Για τον υπολογισμό των παραμέτρων, τα νευρωνικά δίκτυα συνήθως χρησιμοποιούν τον αλγόριθμο backpropagation. Ο αλγόριθμος αυτός αποτελεί έναν πολύ αποτελεσματικό αλγόριθμο για τον υπολογισμό των gradients (κλίσεις) χρησιμοποιώντας τον κανόνα αλυσίδα (δηλαδή τον κανόνα για τον υπολογισμό παραγώγων σύνθετων συναρτήσεων). Κατά τον αλγόριθμο gradient descent, οι παράμετροι του νευρικού δικτύου λαμβάνουν μια ενημέρωση ανάλογη με τη μερική παράγωγο της συνάρτησης κόστους σε σχέση με την τρέχουσα παράμετρο σε κάθε επανάληψη της εκπαίδευσης. Στο σημείο αυτό το πρόβλημα έγκειται στο γεγονός πως σε ορισμένες περιπτώσεις, η κλίση θα είναι μικρή (εξαφάνιση) εμποδίζοντας έτσι την αλλαγή της τιμής ορισμένων παραμέτρων. Η διαδικασία αυτή ενδεχομένως να οδηγήσει το νευρωνικό δίκτυο να σταματήσει την εκπαίδευση. Με τον τρόπο αυτό προκύπτει και το πρόβλημα της «εξαφάνισης της κλίσης».

Οι κλασικές συναρτήσεις ενεργοποίησης, όπως η συνάρτηση υπερβολικής εφαπτομένης που αναφέρθηκε νωρίτερα, έχουν κλίσεις στο διάστημα (0,1) και ο αλγόριθμος backpropagation υπολογίζει τις κλίσεις με τον κανόνα αλυσίδα.

Αποτέλεσμα αυτού είναι ο πολλαπλασιασμός n τέτοιων μικρών αριθμών για τον υπολογισμό των κλίσεων των προηγούμενων (αριστερότερων) επιπέδων σε ένα νευρωνικό δίκτυο n -επιπέδων, γεγονός που σημαίνει πως η κλίση μειώνεται εκθετικά ως προς n . Αυτό έχει ως αποτέλεσμα τα προηγούμενα στρώματα να εκπαιδεύονται πολύ αργά, αν όχι καθόλου.

Ωστόσο, οι σύγχρονες υλοποιήσεις των αλγορίθμων μάθησης των νευρωνικών δικτύων επιτρέπουν την εκπαίδευση πολύ «βαθιών» νευρωνικών δικτύων (με πάνω από εκατό επίπεδα). Η συνάρτηση ενεργοποίησης ReLU επηρεάζεται πολύ λιγότερο από το πρόβλημα του vanishing gradient. Επίσης, τα LSTM (long short-term memory) νευρωνικά δίκτυα καθώς και τεχνικές skip connections που χρησιμοποιούνται στα residual νευρωνικά δίκτυα επιτρέπουν την εκπαίδευση δικτύων χιλιάδων επιπέδων.

Επομένως, σήμερα, δεδομένου ότι τα προβλήματα εξαφάνισης και εκτόξευσης κλίσης ως επί το πλείστον έχουν επιλυθεί, ο όρος «deep learning» αναφέρεται στην εκπαίδευση νευρωνικών δικτύων χρησιμοποιώντας τη σύγχρονη αλγοριθμική και μαθηματική εργαλειοθήκη ανεξάρτητα από το πόσο βαθύ το νευρικό το δίκτυο είναι. Πρακτικά, πολλά προβλήματα μπορούν να επιλυθούν με νευρωνικά δίκτυα που έχουν 2-3 επίπεδα μεταξύ των επιπέδων εισόδου και εξόδου. Τα επίπεδα που δεν είναι ούτε είσοδος ούτε έξοδος καλούνται συχνά κρυμμένα επίπεδα (hidden layers).

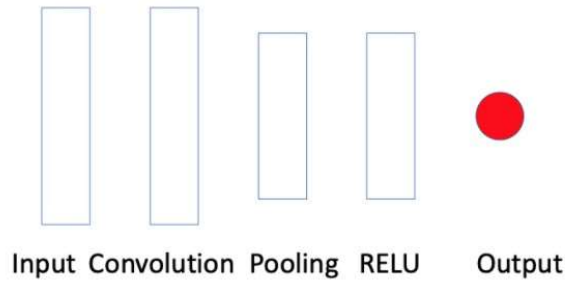
2.12.1 Συνελικτικά Νευρωνικά Δίκτυα

Τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNNs) αποτελούν μία κατηγορία των νευρωνικών δικτύων και συγκεκριμένα αρχιτεκτονικής FFNN που μειώνει σημαντικά τον αριθμό των παραμέτρων σε ένα βαθύ νευρωνικό δίκτυο με πολλές μονάδες χωρίς να μειώνεται ιδιαίτερα η ποιότητα του μοντέλου. Τα μοντέλα CNN αν και αρχικά αναπτύχθηκαν για τη διαχείριση προβλημάτων επεξεργασίας εικόνας, βρίσκουν ευρεία εφαρμογή σε προβλήματα γλωσσικής επεξεργασίας καθώς και σε προβλήματα χρονοσειρών, όπως η πρόβλεψη και η ταξινόμηση [22].

Κατά την ανάλυση των χρονοσειρών, εκτός του σήματος συναντάται και αρκετός θόρυβος πέραν του σήματος, γεγονός που καταδεικνύει την ανάγκη ύπαρξης μιας μεθόδου για την εξομάλυνση του σήματος πριν την ενδεχόμενη πρόβλεψη ή ταξινόμηση. Τέτοιες μέθοδοι μπορεί να είναι η εφαρμογή του κινούμενου μέσου (rolling mean) ή εφαρμογή ενός σταθμισμένου μέσου όρου. Το πρόβλημα που προκύπτει όμως στις μεθόδους αυτές είναι ο προσδιορισμός των παραμέτρων εξομάλυνσης. Στο σημείο αυτό, τα συνελικτικά νευρωνικά δίκτυα εκμεταλλευόμενα την αρχιτεκτονική τους έχουν την δυνατότητα να μάθουν τις παραμέτρους εξομάλυνσης.

Αρχιτεκτονική

Η αρχιτεκτονική ενός απλού συνελικτικού νευρωνικού δικτύου παρουσιάζεται στην παρακάτω εικόνα:



Εικόνα 39: Αρχιτεκτονική CNN

Σύμφωνα με την παραπάνω αρχιτεκτονική, το επίπεδο εισόδου δέχεται ως είσοδο μία υπο-ακολουθία σταθερού μήκους της αρχικής χρονοσειράς και την προωθεί στο συνελικτικό επίπεδο. Το συνελικτικό επίπεδο και το επίπεδο συγκέντρωσης (pooling layer) εξομαλύνουν την είσοδο. Στο επίπεδο ReLU, εφαρμόζεται ένας μη γραμμικός ReLU μετασχηματισμός στην εξομαλυμένη πλέον είσοδο και την προωθεί στο επίπεδο εξόδου. Το επίπεδο εξόδου δέχεται αυτή την τιμή διανύσματος και τη συνδέει σε μία άλλη συνάρτηση ενεργοποίησης με στόχο να δώσει την απόκριση εξόδου. Η απόκριση αυτή, ανάλογα με την συνάρτηση ενεργοποίησης ενδέχεται να είναι πιθανότητα κλάσης, απόκριση συνεχούς τιμής ή κάποιος αριθμός.

Συνελικτικό επίπεδο

Ως συνέλιξη μπορεί να θεωρηθεί ένα «σταθμισμένο άθροισμα αναμνήσεων ή ηχώ»[α]. Για καλύτερη κατανόηση και χρησιμοποιώντας το παράδειγμα του ήχου, υποθέτουμε πως ο όρος $f(t)$ είναι ο ήχος και $h(\tau)$ το ποσοστό του ήχου που ακούστηκε πριν από τ δευτερόλεπτα και αυτός που μπορεί να ακούσει κάποιος σε διακριτές χρονικές στιγμές. Η μαθηματική έκφραση αυτού που ακούγεται τη στιγμή t είναι:

$$h(0)f(t) + h(1)f(t - 1) + h(2)f(t - 2) + \dots + h(t)f(0) \quad (33)$$

Να σημειωθεί πως η παραπάνω έκφραση αποτελεί έναν σταθμισμένο κινούμενο μέσο, όπου τα βάρη προκύπτουν από τη συνάρτηση h . Επομένως, η συνέλιξη διακριτού χρόνου γενικεύει έναν κινούμενο μέσο όρο (moving average) έτσι ώστε τα βάρη να είναι μη μηδενικά και το άθροισμά τους να μην φτάνει το ένα. Με τη διαδικασία αυτή προκύπτει η εξομάλυνση της χρονοσειράς.

Επίπεδο Συγκέντρωσης (Pooling Layer)

Στο επίπεδο συγκέντρωσης πραγματοποιείται ο κατακερματισμός του διανύσματος σε ίσου μεγέθους, μη επικαλυπτόμενες ομάδες (ή δεξαμενές - pools) και έπειτα λαμβάνεται μία συνοπτική στατιστική για κάθε ομάδα. Η διαδικασία αυτή προκαλεί περεταίρω εξομάλυνση του θορύβου στο τοπικό δυναμικό. Οι τρεις πιο κοινοί τύποι

συγκέντρωσης είναι η μέγιστη συγκέντρωση (max pooling) η οποία χρησιμοποιείται συχνά στην επεξεργασία εικόνας, η μέση συγκέντρωση (average ή mean pooling) και η ελάχιστη συγκέντρωση (min pooling).

Επίπεδο ReLU

Το επίπεδο ReLU δέχεται το διάνυσμα το οποίο έχει ήδη εξομαλυνθεί από το συνελκτικό επίπεδο και το επίπεδο συγκέντρωσης και του εφαρμόζει έναν μη γραμμικό μετασχηματισμό για να το προετοιμάσει για το επίπεδο εξόδου.

Επίπεδο Εξόδου

Το επίπεδο εξόδου δέχεται μία αναπαράσταση των αρχικών δεδομένων τα οποία έχουν υποβληθεί σε δύο επίπεδα εξομάλυνσης και ένα επίπεδο μη γραμμικού μετασχηματισμού και εφαρμόζει μια συνάρτηση ενεργοποίησης σε ένα σταθμισμένο άθροισμα αυτής της αναπαράστασης ώστε να εξάγει δεδομένα που είναι αντίστοιχης μορφής με αυτή που επιλέχθηκε από τη συνάρτηση ενεργοποίησης.

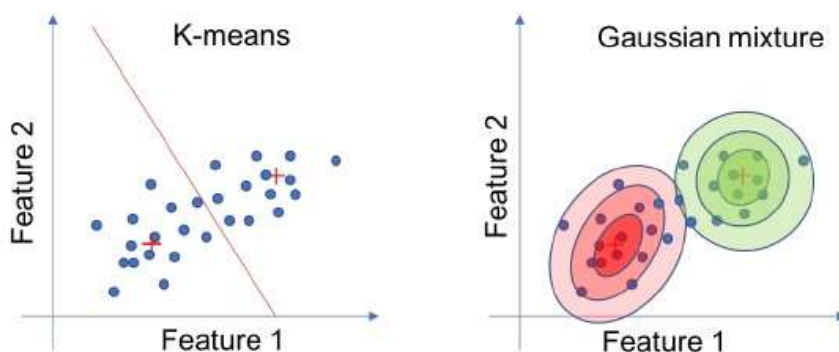
2.13 Unsupervised Learning

Η ανάλυση των κοινωνικών δικτύων, το clustering (συσταδοποίηση) γονιδίων και η μελέτη της αγοράς αποτελούν τις πιο συχνές εφαρμογές της μάθησης χωρίς επίβλεψη. Στην περίπτωση αυτή, το dataset αποτελείται από ένα σετ διανυσμάτων εισόδου x . Παρά το γεγονός πως η μάθηση χωρίς επίβλεψη μπορεί να επιλύσει πολλά προβλήματα, αυτά που κατά κύριο λόγο συναντάμε είναι το clustering και η ανάλυση cluster (cluster analysis).

Clustering ονομάζεται η διαδικασία ομαδοποίησης δεδομένων έτσι ώστε η ενδοσυσταδική ομοιότητα να είναι υψηλή και η διασυσταδική ομοιότητα να είναι χαμηλή. Η ομοιότητα αυτή εκφράζεται ως μία συνάρτηση απόστασης η οποία εξαρτάται από τον τύπο των δεδομένων. Οι δύο κυρίαρχοι model-based αλγόριθμοι που συναντώνται σε αυτή την κατηγορία μάθησης είναι ο αλγόριθμος k-means και το Γκαουσιανό μοντέλο ανάμιξης (Gaussian Mixture Model - GMM).

Ο k-means αποτελεί τον πιο γνωστό clustering αλγόριθμο [23]. Είναι ένας επαναληπτικός αλγόριθμος ο οποίος ξεκινά με το διαχωρισμό των δεδομένων σε k clusters. Στη συνέχεια, υπολογίζεται το κέντρο κάθε cluster και τα δεδομένα αντιστοιχίζονται στο cluster με το πλησιέστερο κέντρο. Η διαδικασία - υπολογισμός κέντρου και εκχώρηση δεδομένων - επαναλαμβάνεται έως ότου η εκχώρηση δεν αλλάξει ή ξεπεραστεί ο προκαθορισμένος μέγιστος αριθμός επαναλήψεων. Μέσω της διαδικασίας αυτής, ο αλγόριθμος ενδέχεται να τερματίσει σε ένα βέλτιστο τοπικό διαχωρισμό τμημάτων. Επιπρόσθετα, ο αλγόριθμος k-means είναι «ευαίσθητος» στις ακραίες τιμές (outliers).

Καθώς ο αλγόριθμος k-mean εκχωρεί κάθε παράδειγμα μοναδικά σε κάθε cluster, , οι πιθανοτικές προσεγγίσεις επιτρέπουν μια μαλακή ανάθεση και παρέχουν ένα μέτρο της αβεβαιότητας που σχετίζεται με την ανάθεση. Στο παρακάτω γράφημα παρουσιάζεται η διαφορά του αλγόριθμου k-means με το μοντέλο GMM:



Εικόνα 40: Διαφορά του αλγόριθμου k-means με μοντέλο GMM

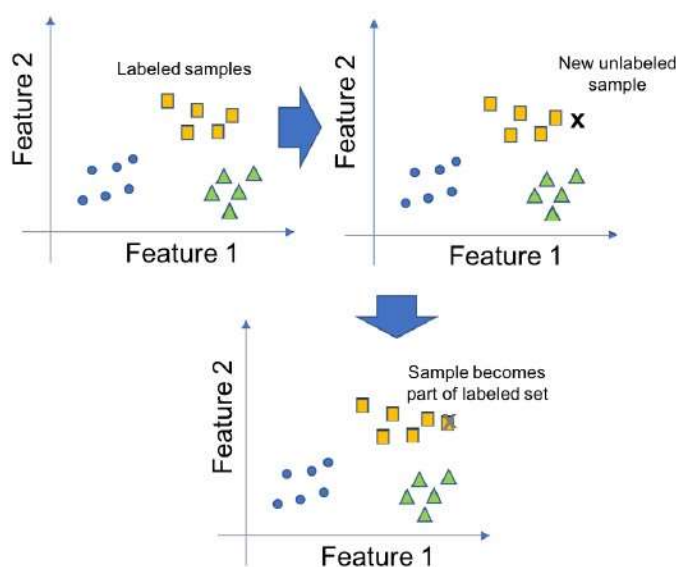
Το μοντέλο GMM αποτελεί μία γραμμική υπέρθεση Gaussian κατανομών και είναι μια από τις πιο ευρέως χρησιμοποιούμενες πιθανολογικές προσεγγίσεις για το clustering. Οι παράμετροι του μοντέλου είναι οι μεικτοί συντελεστές κάθε Gaussian συνιστώσας, δηλαδή ο μέσος όρος και η συνδιακύμανση κάθε κατανομής. Για τη μεγιστοποίηση της συνάρτησης λογαριθμικής πιθανοφάνειας, λαμβάνοντας υπόψη της παραμέτρους του εκάστοτε dataset, χρησιμοποιείται ο αλγόριθμος Expectation Maximization (EM). Η αρχικοποίηση των παραμέτρων μπορεί να γίνει χρησιμοποιώντας k-means. Μετά την αρχικοποίηση των παραμέτρων και την εκτίμηση της λογαριθμικής πιθανοφάνειας, ο αλγόριθμος εναλλάσσεται μεταξύ δύο βημάτων. Στο πρώτο βήμα (expectation step), οι τρέχουσες τιμές των παραμέτρων χρησιμοποιούνται για τον προσδιορισμό της «ευθύνης» κάθε στοιχείου για τα δεδομένα που παρατηρούνται. Το δεύτερο βήμα (maximization step), χρησιμοποιεί τις «ευθύνες» αυτές για να υπολογίσει τη μέγιστη πιθανοφάνεια εκτίμησης των παραμέτρων του μοντέλου. Η σύγκλιση ελέγχεται βάσει της συνάρτησης λογαριθμικής πιθανοφάνειας ή των παραμέτρων.

2.14 Semi-Supervised Learning

Οι αλγόριθμοι μάθησης με ημι-επίβλεψη (Semi-Supervised Learning) είναι υβριδικοί αλγόριθμοι μάθησης και επιλύουν προβλήματα στα οποία τα περισσότερα από τα παραδείγματα εισόδου δεν διαθέτουν κάποια κλάση, την στιγμή που ελάχιστα από τα δεδομένα εισόδου διαθέτουν κλάση. Το κύριο πλεονέκτημα είναι πως σε πολλούς τομείς είναι ήδη διαθέσιμα πληθώρα σετ δεδομένων χωρίς κλάση. Η μάθηση με ημι-επίβλεψη χρησιμοποιείται για τον ίδιο τύπο προβλημάτων με την μάθηση με επίβλεψη και είναι ιδιαίτερα χρήσιμη όταν τα η πρόσβαση στα δεδομένα με κλάση δεν είναι

εύκολη και η χρήση των παραδειγμάτων χωρίς κλάση μπορεί να βελτιώσει την απόδοση.

Η μέθοδος Self-training είναι η παλαιότερη μορφή μάθησης με ημι-επίβλεψη. Είναι μία επαναληπτική διαδικασία κατά την οποία, στο πρώτο στάδιο χρησιμοποιούνται από έναν SL αλγόριθμο τα δεδομένα με ετικέτα. Στη συνέχεια, σε κάθε βήμα, εκχωρείται μία κλάση σε παραδείγματα χωρίς κλάση σύμφωνα με την πρόβλεψη που προκύπτει από το μοντέλο εκπαίδευσης αυτά τα παραδείγματα χρησιμοποιούνται μαζί με τα προηγούμενα δεδομένα για επανεκπαίδευση, χρησιμοποιώντας τον ίδιο αλγόριθμο. Η διαδικασία αυτή παρουσιάζεται στο παρακάτω γράφημα:



Εικόνα 41: Διαδικασία SSL

Η ιδέα της χρησιμοποίησης παραδειγμάτων με και χωρίς κλάση, όπως στον self-training αλγόριθμο, έχει οδηγήσει στην δημιουργία νέων SSL αλγορίθμων. Οι αλγόριθμοι αυτοί μπορούν να ταξινομηθούν στις εξής κατηγορίες [24]:

- Σε μεθόδους βασισμένες σε γενετικά μοντέλα
- Σε μεθόδους που βασίζονται στην υπόθεση ότι το όριο απόφασης πρέπει να βρίσκεται σε μια περιοχή χαμηλής πυκνότητας
- Σε μεθόδους που βασίζονται σε γράφους
- Σε μεθόδους δύο βημάτων, όπου το πρώτο χωρίς επίβλεψη βήμα αλλάζει την αναπαράσταση των δεδομένων ή δημιουργεί έναν καινούριο πυρήνα και το δεύτερο, με επίβλεψη βήμα που βασίζεται στην καινούρια αναπαράσταση ή στον καινούριο πυρήνα.

2.15 Reinforcement Learning

Η μάθηση με ενίσχυση χρησιμοποιείται σε περιπτώσεις που ο στόχος είναι να βρεθεί μία πολιτική (policy), δηλαδή μια χαρτογράφηση μεταξύ των καταστάσεων του περιβάλλοντος σε ενέργειες που πρέπει να εκτελεστούν, ενώ αλληλοεπιδρούν άμεσα με το περιβάλλον.

Το πρότυπο RL επιτρέπει στον «πράκτορα» (agent) να μάθει, εξερευνώντας τις διαθέσιμες ενέργειες και βελτιώνοντας τη συμπεριφορά του χρησιμοποιώντας μόνο μια αξιολογήσιμη ανατροφοδότηση, που ονομάζεται ανταμοιβή (reward). Ο στόχος του πράκτορα είναι να μεγιστοποιήσει μακροπρόθεσμα την απόδοση. Ως εκ τούτου, ο πράκτορας δεν λαμβάνει απλώς υπόψη την άμεση ανταμοιβή, αλλά αξιολογεί τις συνέπειες των ενεργειών του στο μέλλον. Η καθυστερημένη ανταμοιβή και η «δοκιμή-και-σφάλμα» αποτελούν τα πιο σημαντικά χαρακτηριστικά της μάθησης με ενίσχυση.

Η μάθηση με ενίσχυση εκτελείται συνήθως στο πλαίσιο των διαδικασιών απόφασης Markov (Markov decision processes - MDP). Η αντίληψη του πράκτορα τη στιγμή k συμβολίζεται ως $\mathbf{s}_k \in \mathbf{S}$, όπου \mathbf{S} είναι το πεπερασμένο σύνολο των καταστάσεων του περιβάλλοντος. Ο πράκτορας αλληλοεπιδρά με το περιβάλλον εκτελώντας ενέργειες. Τη στιγμή k ο πράκτορας επιλέγει μία ενέργεια $\mathbf{a}_k \in \mathbf{A}$, όπου \mathbf{A} είναι το πεπερασμένο σύνολο των ενεργειών του πράκτορα, η οποία θα μπορούσε να προκαλέσει μία μετάβαση σε μία νέα κατάσταση. Ο πράκτορας θα λάβει μια ανταμοιβή ως αποτέλεσμα της μετάβασης, σύμφωνα με τη συνάρτηση ανταμοιβής: $\rho: \mathbf{S} \times \mathbf{A} \times \mathbf{S} \rightarrow R$. Ο στόχος του πράκτορα είναι να βρει την ακολουθία ζευγών κατάστασης-δράσης που μεγιστοποιεί την αναμενόμενη μειωμένη ανταμοιβή, δηλαδή τη βέλτιστη πολιτική. Στο πλαίσιο του MDP, έχει αποδειχθεί ότι υπάρχει μια βέλτιστη ντετερμινιστική και σταθερή πολιτική. Υπάρχει πλήθος αλγορίθμων που μαθαίνουν τη βέλτιστη πολιτική, τόσο σε περίπτωση που είναι γνωστές η μετάβαση κατάστασης και συνάρτηση ανταμοιβής (model-based learning) όσο και σε περιπτώσεις που είναι άγνωστες (model-free learning). Ο αλγόριθμος RL που χρησιμοποιείται πιο συχνά είναι ο Q-learning, ένας model-free αλγόριθμος ο οποίος εκτιμά τη βέλτιστη συνάρτηση δράσης-τιμής [25]. Η συνάρτηση δράσης-τιμής, η οποία ονομάζεται Q-function, είναι η αναμενόμενη απόδοση ενός ζεύγους κατάστασης-δράσης για μια δεδομένη πολιτική. Η βέλτιστη συνάρτηση δράσης-τιμής, (Q^*), αντιστοιχεί στη μέγιστη αναμενόμενη απόδοση για ένα ζεύγος κατάστασης. Μετά την εκμάθηση της συνάρτησης Q^* , ο πράκτορας επιλέγει την ενέργεια με την αντίστοιχη υψηλότερη τιμή Q σε αντιστοιχία με την τρέχουσα κατάσταση.

Μία λύση βάσει πίνακα (table-based solution) όπως αυτή που περιγράφεται παραπάνω είναι κατάλληλη μόνο σε περίπτωση προβλημάτων με περιορισμένο χώρο κατάστασης. Προκειμένου να γενικευθεί η πολιτική που μαθαίνεται σε αντιστοιχία με καταστάσεις που δεν είχαν προηγουμένως βιώσει ο πράκτορας, οι μέθοδοι RL μπορούν να συνδυαστούν με υπάρχουσες μεθόδους προσέγγισης λειτουργιών, όπως για παράδειγμα τα νευρωνικά δίκτυα.

2.16 Διευκρινίσεις

Classification vs. Regression

Η ταξινόμηση (Classification) αποτελεί ένα πρόβλημα αυτόματης εκχώρησης μιας κλάσης (label) σε ένα παράδειγμα χωρίς κλάση (unlabeled example). Ένα πολύ γνωστό παράδειγμα Classification προβλήματος αποτελεί η αναγνώριση ανεπιθύμητων μηνυμάτων (spam emails).

Στη μηχανική μάθηση, ένα πρόβλημα Classification επιλύεται από έναν αλγόριθμο μάθησης ταξινόμησης ο οποίος δέχεται μια συλλογή από παραδείγματα με κλάση ως εισόδο και παράγει ένα μοντέλο που μπορεί να πάρει ένα παράδειγμα χωρίς κλάση ως είσοδο και είτε εξάγει απευθείας μία κλάση είτε εξάγει έναν αριθμό που μπορεί να αναλυθεί (μέσω data analytics), ώστε να εξαχθεί εύκολα ένα συμπέρασμα για την κλάση. Συχνά ένας τέτοιος αριθμός είναι μία πιθανότητα.

Σε ένα πρόβλημα Classification, κάθε κλάση αποτελεί μέρος ενός πεπερασμένου σετ από κλάσεις. Εάν το πλήθος των κλάσεων είναι το δύο (όπως ‘spam’/’not spam’ στο προηγούμενο παράδειγμα), τότε αναφερόμαστε σε δυαδική ταξινόμηση (binary classification). Επίσης συναντάμε και ταξινόμηση πολλών κλάσεων με περισσότερες από δύο κλάσεις (Multiclass classification). Τέλος, ενώ πολλοί ML αλγόριθμοι επιτρέπουν περισσότερες από δύο κλάσεις, άλλοι είναι εκ φύσεως δυαδικοί.

Η παλινδρόμηση (Regression) αποτελεί ένα πρόβλημα πρόβλεψης μίας κλάσης με πραγματική τιμή (real-valued label), το οποίο συχνά αποκαλείται και στόχος (target), έχοντας ως είσοδο ένα παράδειγμα χωρίς κλάση. Ένα πολύ γνωστό παράδειγμα Regression προβλήματος είναι η εκτίμηση της αξίας ενός σπιτιού βάσει των χαρακτηριστικών του, δηλαδή την περιοχή στην οποία βρίσκεται, το μέγεθος του, το πλήθος των υπνοδωματίων, και ούτω καθεξής.

Στη μηχανική μάθηση ένα regression πρόβλημα επιλύεται από έναν regression learning αλγόριθμο ο οποίος δέχεται ως είσοδο μία συλλογή από παραδείγματα με κλάσεις (labeled examples) και παράγει ένα μοντέλο που μπορεί να δεχτεί ως είσοδο παραδείγματα χωρίς κλάση και να εξάγει μία πραγματική τιμή (στόχος).

Model-Based vs. Instance-Based Learning

Οι περισσότεροι SL αλγόριθμοι βασίζονται σε μοντέλα (model-based learning algorithms), όπως για παράδειγμα ο αλγόριθμος SVM. Οι model-based learning αλγόριθμοι χρησιμοποιούν τα δεδομένα εκπαίδευσης (training data) για να δημιουργήσουν ένα μοντέλο που έχει παραμέτρους τις οποίες «έμαθε» από τα training data. Αφότου κατασκευαστεί το μοντέλο, το training data set μπορεί να παραμεριστεί.

Οι Instance-Based Learning αλγόριθμοι χρησιμοποιούν ολόκληρο το dataset ως μοντέλο. Ένας Instance-Based Learning αλγόριθμος που χρησιμοποιείται πολύ συχνά είναι ο k-NN (k-Nearest Neighbors). Σε προβλήματα classification, για να προβλεφθεί η κλάση ενός παραδείγματος εισόδου, ο αλγόριθμος k-NN εξετάζει την «γειτονιά» του στοιχείου εισόδου στο χώρο του χαρακτηριστικού διανύσματος και δίνει ως έξοδο την κλάση που συνάντησε πιο συχνά.

Shallow vs. Deep Learning

Ένας shallow learning (ρηχής μάθησης) αλγόριθμος μαθαίνει τις παραμέτρους του μοντέλου κατευθείαν από το training set. Οι περισσότεροι SL αλγόριθμοι κατατάσσονται στους shallow. Η μόνη «κραυγαλέα» εξαίρεση είναι τα Νευρωνικά Δίκτυα (Neural Networks), ειδικά εκείνοι οι αλγόριθμοι που δημιουργούν νευρωνικά δίκτυα με περισσότερα από ένα επίπεδα μεταξύ της εισόδου και της εξόδου (deep neural networks). Στα deep neural networks, ή αλλιώς κατά το Deep Learning, εν αντιθέσει με το shallow learning, οι παράμετροι των μοντέλων δεν εκπαιδεύονται απευθείας από τα χαρακτηριστικά του training set, αλλά από τις εξόδους των προηγούμενων επιπέδων.

Κεφάλαιο 3 - Εφαρμογές Μηχανικής Μάθησης στα Οπτικά Δίκτυα

3.1 Εισαγωγή

Τα σημερινά τηλεπικοινωνιακά δίκτυα έχουν μετατραπεί σε μέσα κίνησης τεράστιου όγκου ετερογενών δεδομένων. Αυτά τα δεδομένα μπορεί να προέρχονται είτε από ίχνη κυκλοφορίας του δικτύου, είτε alarms του δικτύου, είτε δείκτες ποιότητας σήματος, δεδομένα συμπεριφοράς χρηστών κ.α. Για την εξαγωγή χρήσιμων πληροφοριών από αυτά τα δεδομένα (δεδομένα που έχουν παραχθεί από το ίδιο το δίκτυο) και τη λήψη αποφάσεων σχετικά με την ορθή λειτουργία του δικτύου απαιτούνται προηγμένα μαθηματικά εργαλεία. Μεταξύ αυτών των μαθηματικών εργαλείων, η Μηχανική Μάθηση θεωρείται ως μια από τις πιο ελπιδοφόρες μεθοδολογικές προσεγγίσεις για την ανάλυση των δεδομένων του δικτύου, την αυτοματοποίηση του και τη διαχείριση των βλαβών [26].

Η υιοθέτηση τεχνικών μηχανικής μάθησης στον τομέα των οπτικών δικτύων επικοινωνίας οφείλεται στην άνευ προηγουμένου αύξηση της πολυπλοκότητας του δικτύου που αντιμετωπίζουν τα οπτικά δίκτυα τα τελευταία χρόνια. Η αύξηση οφείλεται στην εισαγωγή τεράστιου αριθμού ρυθμιζόμενων και αλληλοεξαρτώμενων παραμέτρων συστήματος όπως για παράδειγμα οι ρυθμίσεις δρομολόγησης, η μορφή της διαμόρφωσης, ο ρυθμός συμβόλων, τα σχήματα κωδικοποίησης, οι οποίες παράμετροι ενεργοποιούνται με τη χρήση σύμφωνων τεχνολογιών μετάδοσης / λήψης, με την προηγμένη επεξεργασία του ψηφιακού σήματος και με την αντιστάθμιση μη γραμμικών επιδράσεων στη διάδοση μέσα στην οπτική ίνα.

Τα οπτικά δίκτυα αποτελούν τη βασική φυσική υποδομή όλων των δικτύων μεγάλων παρόχων παγκοσμίως, χάρη στην υψηλή χωρητικότητα, το χαμηλό κόστος και πολλές άλλες ελκυστικές ιδιότητες [10]. Τώρα διεισδύουν σε νέες σημαντικές αγορές τηλεπικοινωνιών όπως τα DataCom (Data Communications - δηλαδή η διαδικασία χρήσης τεχνολογιών υπολογιστών και επικοινωνιών για τη μεταφορά δεδομένων) [27] και τα δίκτυα πρόσβασης [28], και δεν υπάρχει ένδειξη ότι ενδέχεται να εμφανιστεί μια υποκατάστατη τεχνολογία στο άμεσο μέλλον. Έχουν μελετηθεί διάφορες προσεγγίσεις για τη βελτίωση της απόδοσης των οπτικών δικτύων, όπως η δρομολόγηση, η εκχώρηση μήκους κύματος, η διαχείριση της κυκλοφορίας και η επιβιωσιμότητα [29] [30].

Ένα εύλογο ερώτημα που προκύπτει σήμερα είναι γιατί η μηχανική μάθηση, ένας τομέας που έχει αναπτυχθεί και ερευνάται εδώ και τρεις δεκαετίες, ασχολείται τώρα με τα οπτικά δίκτυα. Η απάντηση είναι σύνθετη και δεν αφορά μόνο τεχνικούς λόγους. Από τεχνικής άποψης, η πρόσφατη τεχνική πρόοδος τόσο στα οπτικά συστήματα επικοινωνίας όσο και στο επίπεδο του δικτύου, στηρίζεται σε μεγάλο βαθμό στην ραγδαία αύξηση της πολυπλοκότητας των οπτικών δικτύων [31].

Από πλευράς συστήματος, ενώ η μοντελοποίηση των οπτικών καναλιών ήταν ανέκαθεν πολύπλοκη, με την πρόσφατη υιοθέτηση σύμφωνων τεχνικών [32] η μοντελοποίηση

αυτή έγινε ακόμα πολυπλοκότερη καθώς εισήχθησαν πληθώρα ρυθμιζόμενων παραμέτρων σχεδίασης, όπως οι μορφές διαμόρφωσης, οι ρυθμοί συμβόλων και οι προσαρμοστικοί ρυθμοί κωδικοποίησης με στόχο την βελτιστοποίηση του ρυθμού μετάδοσης. Επίσης, ένας παράγοντας που καθιστά αυτή τη βελτιστοποίηση αρκετά δύσκολη, είναι η μη γραμμικότητα των εξασθενήσεων των οπτικών μονοπατιών.

Από πλευράς δικτύου, η αύξηση της πολυπλοκότητας των συστημάτων μετάδοσης αντικατοπτρίζεται σε μια σειρά εξελίξεων τόσο στο επίπεδο δεδομένων όσο και στο επίπεδο ελέγχου. Στο επίπεδο δικτύου, η ιδέα των ελαστικών οπτικών δικτύων (Elastic Optical Network - EON) [33] [34] [35] έχει αναδειχθεί ως μια νέα αρχιτεκτονική οπτικών δικτύων ικανή να ανταποκριθεί στην αυξημένη ανάγκη ελαστικότητας στην κατανομή πόρων οπτικού δικτύου. Σε αντίθεση με τα παραδοσιακά σταθερά WDM δίκτυα, τα ελαστικά οπτικά δίκτυα προσφέρουν ευέλικτη (σχεδόν συνεχή) κατανομή εύρους ζώνης. Τα δίκτυα EON έχουν την δυνατότητα κατανομής πόρων ώστε να προσαρμόζονται στις διάφορες προαναφερθείσες μεταβλητές αποφάσεων που διατίθενται από νέα συστήματα μετάδοσης, συμπεριλαμβανομένων διαφορετικών τεχνικών μετάδοσης, όπως οι OFDM (Orthogonal Frequency Division Multiplexing), Nyquist WDM (NWDM), τύπους αναμεταδοτών (όπως BVT και S-BVT), μορφές διαμόρφωσης (όπως QPSK, QAM) και ρυθμών κωδικοποίησης. Αυτή η ευελιξία καθιστά τα προβλήματα κατανομής πόρων ως μία ιδιαίτερη πρόκληση για τους μηχανικούς δικτύων. Στο επίπεδο ελέγχου, ο δυναμικός έλεγχος, όπως στα SDN δίκτυα (Software-Defined Networks), δύναται να επιτρέψει την πολυαναμενόμενη κατά απαίτηση (on-demand) αναδιαμόρφωση (reconfiguration) και εικονοποίηση (virtualization). Επιπρόσθετα, η αναδιάρθρωση του οπτικού υποστρώματος θέτει πολλές προκλήσεις όσον αφορά την επανα-βελτιστοποίηση του δικτύου, τον κατακερματισμό του φάσματος, τις ρυθμίσεις ισχύος των ενισχυτών, τις μη αναμενόμενες βλάβες λόγω μη γραμμικότητας μετάδοσης, οι οποίες απαιτούν αυστηρή ενοποίηση μεταξύ των στοιχείων ελέγχου (SDN controllers, network orchestrators) και οπτικών συσκευών παρακολούθησης απόδοσης που λειτουργούν στο επίπεδο εξοπλισμού.

Όλοι αυτοί οι βαθμοί ελευθερίας καθώς και οι περιορισμοί θέτουν σοβαρές προκλήσεις για τη λήψη απόφασης για το ποια είναι η καλύτερη σχεδίαση συστήματος και/ή δικτύου. Η μηχανική μάθηση θεωρείται σήμερα ως το εργαλείο για την αλλαγή του τρόπου σχεδιασμού των μελλοντικών οπτικών δικτύων και συστημάτων.

3.2 Προκλήσεις για την εφαρμογή τεχνικών Μηχανικής Μάθησης στα Οπτικά Δίκτυα

Τα τελευταία χρόνια οι εφαρμογές τεχνικών μηχανικής μάθησης έχουν προσελκύσει το ενδιαφέρον των ερευνητών και των επιστημόνων στο χώρο των οπτικών επικοινωνιών και δικτύων. Τα βασικά κίνητρα του ενδιαφέροντος αυτού, είναι η αυξημένη πολυπλοκότητα των συστημάτων καθώς και αυξημένη διαθεσιμότητα των δεδομένων.

Η αυξημένη πολυπλοκότητα των συστημάτων οφείλεται στην υιοθέτηση προηγμένων τεχνικών μετάδοσης [32], καθώς και στην εισαγωγή αρκετά ευέλικτων αρχών δικτύωσης (όπως το παράδειγμα ελαστικών οπτικών δικτύων) και έχουν καταστήσει το σχεδιασμό και τη λειτουργία οπτικών δικτύων εξαιρετικά περίπλοκες διαδικασίες, λόγω του μεγάλου αριθμού των ρυθμιζόμενων παραμέτρων που πρέπει να ληφθούν υπόψη (μορφές διαμόρφωσης, ρυθμοί συμβόλων, προσαρμοστικοί ρυθμοί κωδικοποίησης, προσαρμοστικό εύρος ζώνης καναλιού). Στην περίπτωση αυτή, η ακριβής μοντελοποίηση του συστήματος είναι συχνά πολύ δύσκολη, αν όχι αδύνατη. Στην πραγματικότητα, τα αναλυτικά μοντέλα συνήθως υιοθετούν «περιθώρια» (margins) οδηγώντας το σύστημα σε υποχρησιμοποίηση πόρων και συνεπώς αυξημένο κόστος. Στον αντίποδα, με την χρήση τεχνικών μηχανικής μάθησης δύναται η καταγραφή σύνθετων, μη γραμμικών συμπεριφορών του συστήματος και με χρήση εκπαιδευμένων μοντέλων μάθησης με επίβλεψη, τα οποία έχουν εκπαιδευτεί από προηγούμενες καταγραφές δεδομένων του δικτύου, με στόχο την επίλυση σύνθετων προβλημάτων πολλαπλών επιπέδων, όπως αυτά του κλάδου των οπτικών δικτύων.

Η αυξημένη διαθεσιμότητα των δεδομένων οφείλεται στο γεγονός ότι τα σύγχρονα οπτικά δίκτυα είναι εξοπλισμένα με μεγάλο αριθμό συσκευών παρακολούθησης του δικτύου, ικανές να παρέχουν συνεχώς πληροφορίες για ολόκληρο το δίκτυο. Πληροφορίες όπως ίχνη κίνησης (traffic traces), δείκτες ποιότητας σήματος όπως το BER (bit error rate), alarms βλαβών του συστήματος, συμπεριφορά χρηστών. Στην περίπτωση αυτή, το πλεονέκτημα εφαρμογής τεχνικών μηχανικής μάθησης έγκειται στην αξιοποίηση της πληθώρας συλλεγμένων δεδομένων με την ταυτόχρονη μελέτη κρυφών εξαρτήσεων μεταξύ διαφορετικών τύπων δεδομένων.

Η εφαρμογή της μηχανικής μάθησης στο φυσικό επίπεδο οφείλεται κυρίως στην παρουσία μη γραμμικών φαινομένων στις οπτικές ίνες, τα οποία καθιστούν τα αναλυτικά τα μοντέλα ανακριβή ή ακόμη και πολύ περίπλοκα. Επιπτώσεις των φαινομένων αυτών συναντάμε σε περιπτώσεις όπως η πρόβλεψη απόδοσης των οπτικών δικτύων όσον αφορά το BER, τον παράγοντα ποιότητας (Q-factor) καθώς επίσης και την αποδιαμόρφωση του σήματος [36].

Προχωρώντας από το φυσικό επίπεδο στο επίπεδο δικτύου συναντάμε τις ίδιες προκλήσεις στην εφαρμογή μεθόδων μηχανικής μάθησης. Συγκεκριμένα, ο σχεδιασμός και η διαχείριση των οπτικών δικτύων εξελίσσονται συνεχώς εξαιτίας της τεράστιας αύξησης της μεταφερόμενης κίνησης και των δραστικών αλλαγών στις απαιτήσεις κυκλοφορίας όσον αφορά τη χωρητικότητα, την καθυστέρηση, την εμπειρία του χρήστη και την ποιότητα της υπηρεσίας (Quality of Service - QoS). Επομένως, τα σύγχρονα οπτικά δίκτυα αναμένεται να λειτουργούν με πολύ μεγαλύτερη αξιοποίηση σε σχέση με το παρελθόν, καθώς παρέχουν αυστηρότερες εγγυήσεις για την ποιότητα της παρεχόμενης υπηρεσίας. Για την επίτευξη του στόχου αυτού επιβάλλεται η εφαρμογή τεχνικών διαχείρισης κίνησης και βελτιστοποίησης, οι οποίες όμως ενδεχομένως να αντιμετωπίζουν ζητήματα επεκτασιμότητας και μη αποδεκτής υπολογιστικής πολυπλοκότητας. Στο πλαίσιο αυτό, η μηχανική μάθηση θεωρείται ως μία πολλά υποσχόμενη μεθοδολογική προσέγγιση των ζητημάτων αυτών, καθώς επιτρέπει την αυτόματη αυτό-διαμόρφωση του δικτύου και την άμεση λήψη αποφάσεων αξιοποιώντας την πληθώρα των διαθέσιμων δεδομένων του δικτύου και

επιτρέποντας έτσι τη δημιουργία μοντέλων για πιο ακριβή και βέλτιστη τροφοδοσία και διαχείριση του δικτύου.

3.3 Εφαρμογές της Μηχανικής Μάθησης στο φυσικό επίπεδο

Όπως έχει ήδη περιγραφεί, αρκετές προκλήσεις πρέπει να αντιμετωπιστούν στο φυσικό επίπεδο των οπτικών δικτύων για την εκτίμηση της απόδοσης του συστήματος μετάδοσης καθώς και τον έλεγχο επηρεασμού των οπτικών καναλιών από την υποβάθμιση του σήματος κατά την μετάδοση. Τα αποτελέσματα των διαδικασιών αυτών μπορούν να χρησιμοποιηθούν για άλλες προληπτικές διαδικασίες, όπως η ρύθμιση της παροχής ισχύος, ο έλεγχος του κέρδους των οπτικών ενισχυτών και οι διαφορετικές μορφές διαμόρφωσης, προτού υποβαθμιστεί το σήμα. Παρακάτω, παρουσιάζονται οι κυριότερες εφαρμογές της μηχανικής μάθησης στο επίπεδο του φυσικού δικτύου.

Εκτίμηση Ποιότητας Μετάδοσης (Quality of Transmission - QoT)

Για την ανάπτυξη ενός νέου οπτικού μονοπατιού απαιτείται η εκτίμηση της ποιότητας της μετάδοσης (QoT) του μονοπατιού αυτού, καθώς επίσης και των ήδη υπαρχόντων. Η έννοια της ποιότητας της μετάδοσης γενικά αναφέρεται σε έναν αριθμό παραμέτρων του φυσικού επιπέδου. Παράμετροι όπως ο λαμβάνον οπτικός σηματο-θορυβικός λόγος (Optical Signal-to-Noise Ratio – OSNR), ο ρυθμός λανθασμένων συμβόλων (bit error rate – BER) και ο παράγοντας Q (Q-factor), οι οποίες επηρεάζουν την «αναγνωσιμότητα» του οπτικού σήματος στον δέκτη και δίνουν ένα ποσοτικό μέτρο για τον έλεγχο εάν θα ήταν εγγυημένο ένα προκαθορισμένο επίπεδο QoT. Οι παράμετροι αυτοί επηρεάζονται από διάφορες άλλες σχεδιαστικές παραμέτρους, όπως η μορφή διαμόρφωσης, ο ρυθμός μετάδοσης συμβόλων (baud rate), ο ρυθμός κωδικοποίησης, τα φυσικά μονοπάτια του δικτύου. Όπως γίνεται κατανοητό, η βελτιστοποίηση της ποιότητας μετάδοσης δεν είναι μία απλή υπόθεση, καθώς απαιτείται η ρύθμιση όλων αυτών των παραμέτρων καθώς και η εύρεση του βέλτιστου συνδυασμού τους.

Οι υπάρχουσες τεχνικές εκτίμησης της ποιότητας μετάδοσης χωρίζονται σε δύο κατηγορίες, τα ακριβή αναλυτικά μοντέλα, τα οποία παρέχουν αρκετά ακριβή αποτελέσματα για τις αστοχίες στο φυσικό επίπεδο αλλά απαιτούν αρκετή υπολογιστική ισχύ, και τις διακριτές μεθόδους (marginated formulas) οι οποίες υπολογιστικά είναι πιο γρήγορες αλλά εισάγουν διαλείψεις στη μετάδοση και οδηγούν στην υποχρησιμοποίηση του δικτύου. Επίσης, αξίζει να σημειωθεί πως η δημιουργία ακριβών αναλυτικών μοντέλων είναι μία δύσκολη διαδικασία λόγω της πολύπλοκης αλληλεπίδρασης των πολλών παραμέτρων του συστήματος (όπως η ισχύς του σήματος εισόδου, το πλήθος των καναλιών, ο τύπος των συνδέσμων, ο ρυθμός συμβόλων) και κατά κύριο λόγο εξαιτίας της μη γραμμικής μετάδοσης του σήματος μέσω του οπτικού καναλιού. Εν αντιθέσει με τη μηχανική μάθηση που είναι μια πολλά υποσχόμενη προσέγγιση για την πρόβλεψη της ποιότητας της υπηρεσίας για κανάλια που δεν έχουν εγκατασταθεί ακόμα.

Μια τεχνική μηχανικής μάθησης που συναντάται στην περίπτωση αυτή είναι η χρήση ταξινομητών (classifiers) που μπορούν να εκπαιδευτούν μέσω μάθησης με επίβλεψη, με σκοπό τη δημιουργία άμεσης σχέσης εισόδου-εξόδου μεταξύ του δείκτη QoT που παρατηρήθηκε στο δέκτη με την αντίστοιχη διαμόρφωση του οπτικού μονοπατιού, όσον αφορά τη χρησιμοποιούμενη μορφή διαμόρφωσης, ρυθμό baud και τη φυσική διαδρομή στο δίκτυο.

Έλεγχος οπτικών ενισχυτών

Στα σύγχρονα οπτικά δίκτυα, λόγω της εμφάνισης νέων υπηρεσιών που απαιτούν τεράστιο εύρος ζώνης για περιορισμένες χρονικές περιόδους, η ανάθεση των οπτικών μονοπατιών γίνεται πλέον δυναμικά. Η δυναμική ρύθμιση όμως, καθώς και η κατάργηση των οπτικών καναλιών με διαφορετικά μήκη κύματος αναγκάζει τα δίκτυα να επαναρυθμιστούν κατά τη διάρκεια της διαδικασίας αυτής για να διατηρηθεί η σταθερότητα του δικτύου στο φυσικό επίπεδο. Ως απάντηση στην γρήγορη αλλαγή της εγκατάστασης των οπτικών μονοπατιών, οι EDFAs παρουσιάζουν κενά ισχύος εξαρτώμενες από το μήκος κύματος. Δηλαδή, όταν εγκαθίσταται ένα νέο οπτικό μονοπάτι ή καταργείται ένα ήδη εγκατεστημένο, η απόκλιση των επιπέδων ισχύος του σήματος μεταξύ διαφορετικών καναλιών (δηλαδή καναλιών που λειτουργούν σε διαφορετικό μήκος κύματος) εξαρτάται από το συγκεκριμένο μήκος κύματος που προστίθεται ή καταργείται στο σύστημα. Επομένως, απαιτείται ένας αυτόματος έλεγχος των επιπέδων ισχύος του σήματος προενίσχυσης για να αποφευχθεί αυτή η υπερβολική απόκλιση ισχύος μεταξύ των οπτικών μονοπατιών, η οποία απόκλιση ενδέχεται να προκαλέσει και παραμόρφωση του σήματος. Στο σημείο αυτό βρίσκουν εφαρμογή οι τεχνικές μηχανικής μάθησης. Χάρη στη διαθεσιμότητα των δεδομένων που έχουν καταγραφεί και ανακτηθεί από την παρακολούθηση του δικτύου, regression αλγόριθμοι μπορούν να εκπαιδευτούν για να προβλέψουν με ακρίβεια τα κενά ισχύος ως απόκριση στην προσθήκη/τερματισμό αυτού του συγκεκριμένου μήκους κύματος.

Αναγνώριση σχήματος διαμόρφωσης

Οι σύγχρονοι οπτικοί πομποί και δέκτες παρέχουν υψηλή ευελιξία στο χρησιμοποιούμενο εύρος ζώνης, τη συχνότητα του φέροντος σήματος και τη μορφή διαμόρφωσης, με σκοπό κυρίως την προσαρμογή της μετάδοσης στον απαιτούμενο ρυθμό μετάδοσης και της οπτικής εμβέλειας σε ένα ευέλικτο/ελαστικό περιβάλλον δικτύου. Δεδομένου ότι στην πλευρά της μετάδοσης υιοθετείται μια αυθαίρετη συνεκτική οπτική μορφή μετάδοσης, η εκ των προτέρων γνώση αυτής της απόφασης στην πλευρά του δέκτη δεν είναι πάντα εφικτή και αυτό μπορεί να επηρεάσει την μετάδοση του σήματος, και κατά συνέπεια την επεξεργασία του και την ανίχνευση του.

Η χρήση αλγορίθμων μάθησης με επίβλεψη ενισχύει την αναγνώριση της μορφής διαμόρφωσης (Modulation format recognition - MFR) στην πλευρά του δέκτη, εξαιτίας της δυνατότητας να μάθουν την τοπολογία μεταξύ της υιοθετημένης μορφής διαμόρφωσης και του εισερχόμενου οπτικού σήματος.

Εξομάλυνση μη γραμμικότητας

Η συμπεριφορά των παραμέτρων απόδοσης, παράμετροι όπως το BER, ο Q-factor, η χρωματική διασπορά (Chromatic Dispersion – CD) είναι αρκετά απρόβλεπτη εξαιτίας των μη γραμμικών φαινομένων που παρουσιάζουν οι οπτικές ίνες, όπως το φαινόμενο Kerr. Η απρόβλεπτη αυτή συμπεριφορά ενδεχομένως να προκαλέσει παραμόρφωση του σήματος στο δέκτη. Για το λόγο αυτό χρησιμοποιούνται συχνά πολύπλοκα αναλυτικά μοντέλα για να λειτουργήσουν αντίστροφα στην υποβάθμιση του σήματος ή να αντισταθμίσουν τα μη γραμμικά φαινόμενα.

Εν αντιθέσει με τα προσεγγιστικά αναλυτικά μοντέλα που χρησιμοποιούνται για την επίλυση τέτοιων σύνθετων μη γραμμικών προβλημάτων, δύναται η δημιουργία μοντέλων μηχανικής μάθησης με επίβλεψη, τέτοια ώστε να συλλαμβάνουν τα μη γραμμικά φαινόμενα αξιοποιώντας τη «γνώση» από δεδομένα που έχουν ανακτηθεί από το δίκτυο, και να δημιουργούν σχέσεις εισόδου-εξόδου μεταξύ των υπό εξέταση παραμέτρων και των επιθυμητών εξόδων.

Παρακολούθηση απόδοσης οπτικού επιπέδου (Optical performance monitoring - OPM)

Με τις απαιτήσεις για χωρητικότητα στα οπτικά δίκτυα επικοινωνίας ολοένα και να αυξάνονται, η παρακολούθηση του δικτύου στο οπτικό επίπεδο αποκτά ζωτική σημασία για την εξασφάλιση ισχυρών και αξιόπιστων δικτύων. Η παρακολούθηση της οπτικής απόδοσης στοχεύει στην εκτίμηση των παραμέτρων μετάδοσης του συστήματος (όπως BER, Q-factor, CD) κατά τη διάρκεια του κύκλου λειτουργίας των οπτικών μονοπατιών. Η γνώση αυτών των παραμέτρων μπορεί να αξιοποιηθεί για διάφορες διαδικασίες, όπως την ενεργοποίηση μονάδων αντισταθμιστή πόλωσης, την ρύθμιση παροχής ισχύος, τη μεταβολή της υιοθετούμενης μορφής διαμόρφωσης, την αλλαγή διαδρομής οπτικών καναλιών. Το γεγονός πως οι παράμετροι οπτικής απόδοσης πρέπει να συλλέγονται σε διάφορα σημεία «παρακολούθησης» κατά μήκος του οπτικού μονοπατιού απαιτεί μεγάλο αριθμό μονάδων καταγραφής αυξάνοντας το κόστος της υποδομής.

Για την μείωση του αριθμού των μονάδων καταγραφής κατά μήκος της υποδομής, ειδικά σε ενδιάμεσα σημεία του δικτύου, και κατ' επέκταση μείωση του κόστους υποδομής, μπορεί να χρησιμοποιηθούν αλγόριθμοι μηχανικής μάθησης με επίβλεψη ώστε να μάθουν τη χαρτογράφηση μεταξύ των παραμέτρων του οπτικού καναλιού και των ιδιοτήτων σήματος που λαμβάνει ο δέκτης, τα οποία μπορούν να ανακτηθούν από δεδομένα όπως στατιστικά διαγραμμάτων ισχύος, το πλάτος του σήματος, το OSNR.

3.4 Εφαρμογές της Μηχανικής Μάθησης στο επίπεδο δικτύου

Όπως στο φυσικό επίπεδο, έτσι και στο επίπεδο δικτύου συναντάμε πολλές εφαρμογές της μηχανικής μάθησης. Η παροχή καινούριων οπτικών μονοπατιών ή η αποκατάσταση των μονοπατιών που ήδη υπάρχουν μετά από μία βλάβη του δικτύου, απαιτούν πολύπλοκες και γρήγορες αποφάσεις οι οποίες εξαρτώνται από διάφορα,

ταχέως εξελισσόμενα δεδομένα καθώς για την διαχείριση της υπάρχουσας κίνησης απαιτείται η αναγνώριση του αντίκτυπου της κίνησης που εισήχθη πρόσφατα. Για την αποτελεσματική λειτουργία του δικτύου, δηλαδή την αποφυγή υπερβολικής παροχής πόρων στο δίκτυο, είναι ιδιαίτερα χρήσιμη η εκτίμηση των απαιτήσεων των χρηστών και των υπηρεσιών.

Παρακάτω παρουσιάζονται οι κύριες περιπτώσεις χρήσης εφαρμογών μηχανικής μάθησης στο επίπεδο δικτύου των οπτικών δικτύων.

Πρόβλεψη κίνησης

Η ακριβής πρόβλεψη της κίνησης στο πεδίο του χωροχρόνου επιτρέπει τον αποτελεσματικό σχεδιασμό και αποτελεσματική διαχείριση των δικτύων. Στην φάση του σχεδιασμού, η πρόβλεψη κίνησης επιτρέπει την μείωση της υπερτροφοδοσίας του δικτύου με πόρους που δεν θα αξιοποιηθούν. Η χρησιμοποίηση των πόρων μπορεί να βελτιστοποιηθεί μέσω εφαρμογής τεχνικών κίνησης, οι οποίες βασίζονται σε πραγματικά δεδομένα και τελικά να επαναδρομολογήσει την υπάρχουσα κίνηση εξοικονομώντας πόρους για μελλοντικές ανάγκες.

Η μηχανική μάθηση βρίσκει εφαρμογή χρησιμοποιώντας αλγόριθμους μάθησης με επίβλεψη, οι οποίοι εκμεταλλεύονται τα δεδομένα που έχουν καταγραφεί και αφορούν τα μοτίβα τις κίνησης καθώς και τη συμπεριφορά των χρηστών, με σκοπό τη δημιουργία μοντέλων που θα εκπαιδευτούν στην πρόβλεψη της μελλοντικής κίνησης καθώς και στις απαιτήσεις ως προς τη διαχείριση των πόρων. Με αυτόν τον τρόπο επιτρέπεται η προληπτική επαναδρομολόγηση της κίνησης καθώς και η τακτική επαναβελτιστοποίηση του δικτύου, έτσι ώστε να εξυπηρετείται όλη η κίνηση του δικτύου και την ίδια στιγμή να μειώνεται η χρήση των πόρων.

Επιπρόσθετα, μπορούν να χρησιμοποιηθούν αλγόριθμοι μάθησης χωρίς επίβλεψη για την εξαγωγή μοτίβων κίνησης σε διαφορετικά σημεία του δικτύου. Με τον τρόπο αυτό μπορούν να αναπτυχθούν παρόμοιες διαδικασίες σχεδιασμού και διαχείρισης (όπως η εγκατάσταση δικτύου και η δέσμευση χωρητικότητας) σε διαφορετικά σημεία του δικτύου ώστε να αναδεικνύονται ομοιότητες ως προς τις απαιτήσεις της κίνησης, δηλαδή να ανήκουν στο ίδιο cluster προφίλ κίνησης.

Επισημαίνεται στο σημείο αυτό πως η πρόβλεψη κίνησης καθώς και οι σχετικές τεχνικές μηχανικής μάθησης ποικίλουν σημαντικά ανάλογα με το τμήμα του δικτύου που εξετάζουμε κάθε φορά (διαφορετικές τεχνικές ακολουθούνται στα εσωτερικά δίκτυα των κέντρων δεδομένων και διαφορετικές στα δίκτυα πρόσβασης) εξαιτίας των χαρακτηριστικών της κίνησης που είναι άρρηκτα συνδεδεμένα με το αντίστοιχο τμήμα του δικτύου.

Σχεδιασμός και αναδιαμόρφωση εικονικής τοπολογίας

Η μοντελοποίηση των επικοινωνιακών δικτύων μέσω εικονικής τοπολογίας συναντά ευρεία εφαρμογή στην εποχή μας. Η μοντελοποίηση αποτελείται από την αναπαράσταση της σύνδεσης δύο τελικών σημείων (παραδείγματος χάρι δύο data center) μέσω μίας γειτνίασης στην εικονική τοπολογία, δηλαδή έναν εικονικό

σύνδεσμο, αν και τα δύο τελικά σημεία δεν είναι απαραίτητο να είναι συνδεδεμένα με φυσικό τρόπο. Μετά τον καθορισμό όλων των εικονικών συνδέσμων, δηλαδή μετά την ταυτοποίηση των αιτημάτων των οπτικών μονοπατιών, για το σχεδιασμό της εικονικής τοπολογίας (Virtual Topology Design - VTD) απαιτείται η επίλυση του προβλήματος της δρομολόγησης και της ανάθεσης μήκους κύματος (Routing and Wavelength Assignment - RWA) για κάθε οπτικό μονοπάτι πάνω από το φυσικό δίκτυο. Να σημειωθεί πως πολλές εικονικές τοπολογίες μπορεί να συνυπάρξουν με το φυσικό δίκτυο, αναπαριστώντας για παράδειγμα υπηρεσίες που απαιτούνται από διαφορετικούς χρήστες καθώς ακόμα και διαφορετικές υπηρεσίες, κάθε μία με ξεχωριστές απαιτήσεις όσον αφορά την ποιότητα της υπηρεσίας (Quality of Service - QoS), το εύρος ζώνης, την καθυστέρηση (latency) που παρέχονται στον ίδιο χρήστη.

Ο σχεδιασμός της εικονικής τοπολογίας δεν είναι απαραίτητος μόνο σε περιπτώσεις παροχής νέων υπηρεσιών και εκχώρησης νέων πόρων στο δίκτυο. Σε ορισμένες περιπτώσεις, όπως παραδείγματος χάρη περιπτώσεις που συναντάμε αστοχίες δικτύου ή όταν η χρησιμοποίηση των πόρων του δικτύου υπόκειται σε διαδικασίες επαναβελτιστοποίησης, οι υπάρχουσες (δηλαδή οι ήδη σχεδιασμένες) εικονικές τοπολογίες επαναρυθμίζονται. Η διαδικασία αυτή ονομάζεται αναδιαμόρφωση εικονικής τοπολογίας (VT reconfiguration).

Για το σχεδιασμό και την αναδιαμόρφωση των εικονικών τοπολογιών δεν απαιτείται μόνο η παροχή χωρητικότητας στο δίκτυο για τις απαιτούμενες υπηρεσίες, αλλά απαιτείται και η παροχή πρόσθετων πόρων σύμφωνα με τα ειδικά χαρακτηριστικά των υπηρεσιών. Αυτή η περίπτωση παροχής υπηρεσιών ονομάζεται τεμαχισμός δικτύου (network slicing), λόγω του γεγονότος ότι κάθε παρεχόμενη υπηρεσία, δηλαδή κάθε εικονική τοπολογία αναπαριστά ένα τμήμα (slice) του συνολικού δικτύου.

Η μηχανική μάθηση συμβάλλει στο σχεδιασμό των εικονικών τοπολογιών και την αναδιαμόρφωσή τους. Μέσω αλγορίθμων μάθησης με επίβλεψη και χρήση ταξινομητών (classification), το μοντέλο μπορεί να εκπαιδευτεί ώστε να αποφασίζει με βέλτιστο τρόπο την κατανομή των πόρων στο δίκτυο, λαμβάνοντας ταυτόχρονα υπόψιν έναν μεγάλο αριθμό διαφορετικών και ετερογενών απαιτήσεων υπηρεσιών για ένα είδος εικονικών τοπολογιών (network slice), επιτρέποντας έτσι τη γρήγορη λήψη αποφάσεων για τη βέλτιστη παροχή πόρων. Η δυνατότητα αυτή αποκτά ιδιαίτερη σημασία ειδικά σε περιπτώσεις που οι συνθήκες του δικτύου μεταβάλλονται δυναμικά.

Διαχείριση βλαβών

Κατά τη διαχείριση ενός δικτύου, η ικανότητα ανίχνευσης βλαβών και ο εντοπισμός τους, όπως επίσης και ο προσδιορισμός της αιτίας αυτής της βλάβης, είναι ζωτικής σημασίας καθώς επιτρέπει την έγκαιρη επαναδρομολόγηση της κίνησης ώστε να διατηρηθεί το καθορισμένο επίπεδο υπηρεσίας (Service Level Agreements - SLAs) και το δίκτυο να ανακάμψει γρήγορα από την βλάβη αυτή. Η διαχείριση των βλαβών μπορεί να επιτευχθεί σε διάφορα επίπεδα. Για παράδειγμα, η ανίχνευση της βλάβης, δηλαδή ο προσδιορισμός του συνόλου των οπτικών μονοπατιών που επηρεάστηκαν από την αποτυχία, είναι μία σχετικά απλή διαδικασία η οποία επιτρέπει την αναδιαμόρφωση των μονοπατιών αυτών μέσω επαναδρομολόγησης της αντίστοιχης κίνησης. Επίσης, η ικανότητα εντοπισμού των βλαβών επιτρέπει την ενεργοποίηση

διαδικασιών επαναφοράς του δικτύου. Με τον τρόπο αυτό μπορεί να ανακτηθεί η κατάσταση προ βλάβης του δικτύου, η οποία κατάσταση είναι γενικά μία βέλτιστη κατάσταση από άποψη διαχείρισης πόρων. Επιπρόσθετα, είναι χρήσιμος ο προσδιορισμός της αιτίας της βλάβης (όπως η προσωρινή συμφόρηση της κυκλοφορίας, η διακοπή λειτουργίας των συσκευών, ακόμα και οι αστοχίες στις μονάδες καταγραφής των δεδομένων) για την υιοθέτηση κατάλληλων διαδικασιών επαναφοράς και αναδιαμόρφωσης της κίνησης. Επιπλέον, η έγκαιρη αναγνώριση της αιτίας μιας βλάβης επιτρέπει τη γρήγορη επισκευή του εξοπλισμού και την επακόλουθη μείωση του μέσου χρόνου επισκευής (Mean Time To Repair - MTTR).

Η μηχανική μάθηση μπορεί να συνεισφέρει στη διαχείριση του τεράστιου όγκου δεδομένων από τη συνεχή καταγραφή της δραστηριότητας του δικτύου. Για παράδειγμα, αλγόριθμοι ταξινόμησης μπορούν να εκπαιδευτούν ώστε να διακρίνουν τη διαφορά μεταξύ κανονικής και ανώμαλης μετάδοσης. Αξίζει να σημειωθεί πως σε περιπτώσεις που τα δεδομένα με κλάση είναι λιγιστά και η πλειονότητα των δεδομένων είναι χωρίς κλάση, μπορούν να χρησιμοποιηθούν προσεγγίσεις μάθησης με ημι-επίβλεψη. Επιπλέον, μπορούν να χρησιμοποιηθούν αλγόριθμοι ταξινόμησης ώστε να διακρίνουν την αιτία της βλάβης, εκμεταλλευόμενοι τη γνώση από τις βλάβες που έχουν ήδη παρατηρηθεί.

Ταξινόμηση ροών κυκλοφορίας

Όταν συνυπάρχουν διαφορετικοί τύποι υπηρεσιών στην ίδια υποδομή του δικτύου, η ταξινόμηση των αντίστοιχων ροών κίνησης πριν από την παροχή τους επιτρέπει την αποτελεσματική κατανομή πόρων, ελαχιστοποιώντας τον κίνδυνο για ανεπαρκή ή υπερβολική παροχή πόρων. Επιπλέον, την ακριβή ταξινόμηση των ροών κυκλοφορίας την εκμεταλλεύονται και άλλες, ήδη παρεχόμενες υπηρεσίες για την εφαρμογή πολιτικών, όπως η διαχείριση της προτεραιότητας των πακέτων, ο έλεγχος της ροής και της συμφόρησης και η εγγύηση κατάλληλου QoS για κάθε ροή σύμφωνα με τα SLAs.

Με βάση τα διάφορα χαρακτηριστικά της κυκλοφορίας και εκμεταλλευόμενοι τον μεγάλο όγκο πληροφοριών που μεταφέρονται από πακέτα δεδομένων, αλγόριθμοι μάθησης με επίβλεψη μπορούν να εκπαιδευτούν για να εξαγάγουν κρυμμένα χαρακτηριστικά κυκλοφορίας και να εκτελούν γρήγορη ταξινόμηση πακέτων και διάκριση ροών.

Υπολογισμός διαδρομής

Κατά την κατανομή των πόρων στο δίκτυο για μια εισερχόμενη αίτηση για παροχή συγκεκριμένης υπηρεσίας, θα πρέπει να επιλεγεί μία κατάλληλη διαδρομή προκειμένου να αξιοποιηθούν αποτελεσματικά οι διαθέσιμοι πόροι του δικτύου για να εξυπηρετηθεί η ζητούμενη κίνηση με το επιθυμητό QoS και χωρίς να επηρεαστούν οι υπάρχουσες υπηρεσίες που ήδη παρέχει το δίκτυο.

Παραδοσιακά, ο υπολογισμός της διαδρομής πραγματοποιείται μέσω χρήσης αλγόριθμων δρομολόγησης βάσει κόστους (όπως οι αλγόριθμοι Dijkstra, Bellman-Ford και Yen), οι οποίοι βασίζονται σε μία προκαθορισμένη μετρική κόστους (όπως η

απόσταση μεταξύ αφετηρίας και τερματισμού, η από άκρο έως άκρο καθυστέρηση, η κατανάλωση ενέργειας, ακόμα και ο συνδυασμός άλλων μετρικών) για τη διάκριση μεταξύ εναλλακτικών διαδρομών.

Στο πλαίσιο αυτό, η χρήση μεθόδων μάθησης με επίβλεψη μπορεί να είναι χρήσιμη καθώς επιτρέπει την εξέταση πολλών παραμέτρων που περιλαμβάνει το εισερχόμενο αίτημα για παροχή υπηρεσίας, ταυτόχρονα με τις πληροφορίες για την τρέχουσα κατάσταση του δικτύου. Με αυτό τον τρόπο οι πληροφορίες αυτές χαρτογραφούνται σε μία βέλτιστη λύση δρομολόγησης χωρίς να χρειάζονται πολύπλοκες εκτιμήσεις κόστους δικτύου, επιτρέποντας έτσι την γρήγορη επιλογή διαδρομής και κατ' επέκταση γρήγορη παροχή υπηρεσίας.

Κεφάλαιο 4 - Πρόβλεψη Χρονοσειρών

4.1 Εισαγωγή

Η πραγματοποίηση προβλέψεων για μελλοντικές παρατηρήσεις ονομάζεται «κατά προσέγγιση υπολογισμός» (extrapolation) στην κλασική στατιστική διαχείριση δεδομένων χρονοσειρών [37]. Στα σύγχρονα επιστημονικά πεδία ενδιαφέροντος, ονομάζεται πρόβλεψη χρονοσειρών (time series forecasting).

Η πρόβλεψη χρονοσειρών αποτελεί έναν πολύ σημαντικό κλάδο της μηχανικής μάθησης. Η σημαντικότητα του έγκειται στο μεγάλο πλήθος προβλημάτων πρόβλεψης που περιλαμβάνουν ένα στοιχείο χρόνου. Αυτή η συνιστώσα του χρόνου καθιστά τα προβλήματα πρόβλεψης χρονοσειρών πολύ δύσκολα στην διαχείρισή της.

Η κίνηση, της οποίας η μελέτη και η πρόβλεψη αποτελεί αντικείμενο μελέτης της παρούσας εργασίας, αποτελεί ουσιαστικά πρόβλημα πρόβλεψης χρονοσειράς. Η κίνηση των δικτύων εκφράζεται μαθηματικά ως μία χρονοσειρά, καθώς η τιμή της οποιαδήποτε χρονικής στιγμής έχει άμεση εξάρτηση με της τιμές των προηγούμενων τιμών της.

Της γίνεται κατανοητό λοιπόν, το πρόβλημα της πρόβλεψης της κίνησης στα οπτικά δίκτυα εκπίπτει στο πρόβλημα πρόβλεψης χρονοσειρών και μελετάται αντίστοιχα.

4.2 Πρόβλεψη χρονοσειρών

Σε αντίθεση με τα υπόλοιπα προβλήματα πρόβλεψης, όπου κατά την πρόβλεψη των μελλοντικών τιμών όλα τα δεδομένα αντιμετωπίζονται σχεδόν πάντα ισότιμα, στην πρόβλεψη χρονοσειρών η ισότιμη αυτή αντιμετώπιση διαφέρει. Οι χρονοσειρές προσθέτουν μια ρητή εξάρτηση μεταξύ των παρατηρήσεων, δηλαδή προσθέτουν μια χρονική διάσταση. Αυτή η πρόσθετη διάσταση αποτελεί τόσο έναν περιορισμό, όσο και μία πηγή που παρέχει πρόσθετες πληροφορίες.

Η πρόβλεψη χρονοσειρών περιλαμβάνει τη χρήση μοντέλων κατάλληλων για να προσαρμοστούν σε ιστορικά δεδομένα και τη χρήση τους για την πρόβλεψη μελλοντικών παρατηρήσεων. Τα περιγραφικά μοντέλα που μπορούν να δανειστούν δεδομένα από το μέλλον για εξομάλυνση ή αφαίρεση θορύβου, επιδιώκουν απλά την καλύτερη περιγραφή των δεδομένων.

Μια ιδιαίτερα σημαντική λεπτομέρεια στην πρόβλεψη χρονοσειρών είναι ότι οι μελλοντικές τιμές δεν είναι διαθέσιμες και πρέπει να εκτιμηθούν μόνο από τα δεδομένα που έχουν ήδη καταγραφεί στο dataset.

Στο σημείο αυτό να διευκρινιστεί πως η περιγραφή των δεδομένων μίας χρονοσειράς αποτελεί διαφορετικό ερευνητικό πεδίο από την πρόβλεψη και ονομάζεται “time series analysis”. Η ανάλυση των χρονοσειρών παρέχει ένα σύνολο τεχνικών για την καλύτερη

κατανόηση ενός συνόλου δεδομένων. Μία ιδιαίτερα χρήσιμη τεχνική είναι η αποσύνθεση μίας χρονοσειράς σε τέσσερα βασικά μέρη:

- Το επίπεδο (Level), δηλαδή η βασική τιμή μίας χρονοσειράς, εάν ήταν ευθεία γραμμή.
- Την Τάση (Trend), δηλαδή η γραμμική αύξηση ή μείωση (συνήθως) της συμπεριφοράς της χρονοσειράς κατά την πάροδο του χρόνου.
- Την Εποχικότητα (Seasonality), δηλαδή τα επαναλαμβανόμενα μοτίβα ή οι κύκλοι συμπεριφοράς της χρονοσειράς κατά την πάροδο του χρόνου.
- Το Θόρυβο (Noise), δηλαδή η μεταβλητότητα στις παρατηρήσεις που δεν μπορεί να εξηγηθεί από το μοντέλο.

Όλες οι χρονοσειρές έχουν επίπεδο, οι περισσότερες έχουν θόρυβο, και η τάση καθώς και η εποχικότητα είναι προαιρετικές παράμετροι. Επίσης, μία ιδιαίτερη παράμετρος των χρονοσειρών είναι η συσχέτιση, δηλαδή η εξάρτηση μεταξύ δύο ή περισσότερων παρατηρήσεων της χρονοσειράς.

Η ικανότητα ενός μοντέλου πρόβλεψης χρονοσειρών καθορίζεται από την απόδοσή του στην πρόβλεψη του μέλλοντος.

Η θεμελιώδης διαίσθηση πίσω από την πρόβλεψη χρονοσειρών είναι ότι, το μέτρο μιας μεταβλητής σε μια χρονική περίοδο θα εξαρτηθεί από το μέτρο της ίδιας μεταβλητής σε μια προηγούμενη χρονική περίοδο, δύο χρονικές περιόδους πριν, τρεις... και ούτω καθεξής. Η πρόβλεψη χρονοσειρών βρίσκει εφαρμογή σε πολλά προβλήματα πρόβλεψης, και το στατιστικό μοντέλο ARIMA μπορεί να υλοποιήσει τέτοια προβλήματα.

4.3 ARIMA

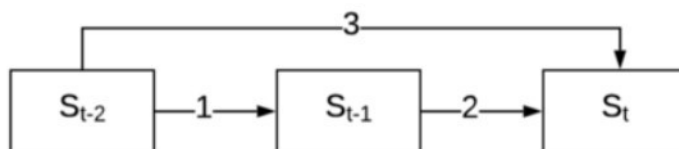
Ο όρος ARIMA προέρχεται από το **Auto Regressive Integrated Moving Average**. Το μοντέλο ARIMA είναι ένα από στοχαστικό μοντέλο χρονοσειρών το οποίο χρησιμοποιείται για να εκπαιδευτεί και να προβλέψει μελλοντικά χρονικά σημεία. Το μοντέλο αυτό μπορεί να συλλάβει πολύπλοκες σχέσεις καθώς παίρνει όρους σφάλματος και παρατηρήσεις καθυστερημένων (lagged) όρων. Μοντέλα σαν το ARIMA βασίζονται στην παλινδρόμηση μιας μεταβλητής σε προηγούμενες τιμές [38].

Συσχέτιση (Correlation) και παράγοντας καθυστέρησης (Lag Factor)

Ως συσχέτιση ορίζεται η εξάρτηση μεταξύ δύο ή περισσότερων παρατηρήσεων σε μία χρονοσειρά. Η εξάρτηση των μεταβλητών είναι γραμμική με αποτέλεσμα να αυξάνονται ή να μειώνονται ταυτόχρονα. Στην ανάλυση των χρονοσειρών, η συσχέτιση αφορά την εξάρτηση μεταξύ τιμών διαφορετικών χρονικών βημάτων.

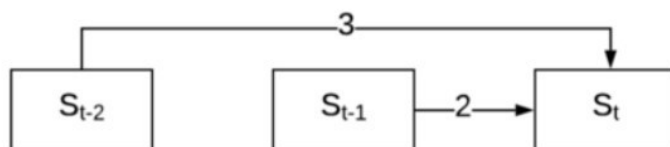
Η συσχέτιση είναι μια σημαντική έννοια των χρονοσειρών και παρουσιάζει δύο τύπους:

- Την Αυτοσυσχέτιση (Auto-correlation) η οποία λαμβάνει υπόψη τόσο άμεσα όσο και έμμεσα αποτελέσματα.



Εικόνα 42: Διάταξη αυτοσυσχέτισης

- Την Μερική αυτοσυσχέτιση (Partial Auto-correlation), η οποία λαμβάνει υπόψη μόνο τα άμεσα αποτελέσματα.



Εικόνα 43: Διάταξη μερικής αυτοσυσχέτισης

Ο παράγοντας καθυστέρησης αφορά το πλήθος των χρονικών βημάτων όπου μία παρατήρηση επηρεάζεται από μία άλλη προηγούμενη παρατήρηση.

Δομικά στοιχεία μοντέλου ARIMA

Ένα στατιστικό μοντέλο ARIMA αποτελείται από τα εξής δομικά στοιχεία:

- Auto Regressive (AR)

Το μοντέλο Auto Regressive (AR) είναι ένας συγκεκριμένος τύπος μοντέλου παλινδρόμησης όπου, η εξαρτημένη μεταβλητή εξαρτάται από τις προηγούμενες τιμές της. Αυτό σημαίνει απαραίτητα ότι οι τρέχουσες τιμές συσχετίζονται με τιμές στα προηγούμενα χρονικά βήματα. Και πιο συγκεκριμένα, ο τύπος συσχέτισης εδώ είναι η μερική αυτοσυσχέτιση.

Η εξίσωση του μοντέλου AR είναι:

$$Y_t = \beta_1 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} \quad (34)$$

Τα σχετικά βάρη ($\Phi_1, \Phi_2, \dots, \Phi_p$) των αντίστοιχων παρατηρήσεων με καθυστέρηση αποφασίζονται από τη συσχέτιση μεταξύ της καθυστερημένης παρατήρησης και της τρέχουσας παρατήρησης. Εάν ο συσχετισμός είναι μεγαλύτερος, το βάρος που αντιστοιχεί σε αυτήν την καθυστέρηση παρατήρησης είναι υψηλό (και αντίστροφα).

Στην εξίσωση παρατηρείται ο όρος p . Ο όρος αυτός ονομάζεται βαθμός καθυστέρησης (lag order) και αντιπροσωπεύει τον αριθμό των προηγούμενων παρατηρήσεων καθυστέρησης που συμπεριλαμβάνουμε στο μοντέλο, δηλαδή τον αριθμό των καθυστερήσεων που έχουν σημαντική συσχέτιση με την τρέχουσα παρατήρηση.

- Moving Average (MA)

Το μοντέλο Moving Average (MA) λειτουργεί αναλύοντας πόσο λάθος είχαν προβλεφθεί οι τιμές για τις προηγούμενες χρονικές περιόδους, με στόχο μια καλύτερη εκτίμηση για την τρέχουσα χρονική περίοδο.

Το μοντέλο λαμβάνει υπόψη τα λάθη από τις καθυστερημένες παρατηρήσεις. Τα αποτελέσματα αυτών των προηγούμενων (καθυστερημένων) σφαλμάτων παρατήρησης, στην τρέχουσα παρατήρηση, εξαρτάται από την αυτοσυσχέτιση μεταξύ τους. Η διαδικασία αυτή είναι παρόμοια με την έννοια του μοντέλου AR που εξετάζει τη μερική αυτοσυσχετισμούς.

$$Y_t = \beta_2 + \omega_1 \varepsilon_{t-1} + \omega_2 \varepsilon_{t-2} + \dots + \omega_q \varepsilon_{t-q} + \varepsilon_t \quad (35)$$

Οι όροι ε αντιπροσωπεύουν τα λάθη που παρατηρούνται στις αντίστοιχες καθυστερήσεις και τα βάρη $\omega_1, \omega_2, \dots, \omega_q$ υπολογίζονται στατιστικά ανάλογα με τους συσχετισμούς.

Στην εξίσωση παρατηρείται ο όρος q . Ο όρος αυτός αντιπροσωπεύει το μέγεθος του κινούμενου παραθύρου, δηλαδή τον αριθμό των λαθών των καθυστερημένων παρατηρήσεων που έχουν σημαντική επίδραση στην τρέχουσα παρατήρηση. Είναι παρόμοιο με τη βαθμό καθυστέρησης (p), αλλά εξετάζει λάθη αντί για τις ίδιες τις παρατηρήσεις.

Το μοντέλο MA συμπληρώνει το μοντέλο AR λαμβάνοντας υπόψη τα σφάλματα από τις προηγούμενες χρονικές περιόδους, βοηθώντας έτσι στην λήψη καλύτερης εκτίμησης.

Συνδυάζοντας τις εξισώσεις AR και MA προκύπτει:

$$Y_t = (\beta_1 + \beta_2) + (\Phi_1 Y_{t-1} + \dots + \Phi_p Y_{t-p}) + (\omega_1 \varepsilon_{t-1} + \dots + \omega_q \varepsilon_{t-q} + \varepsilon_t) \quad (36)$$

Στασιμότητα(Stationarity)

Τα μοντέλα που έχουν συζητηθεί μέχρι στιγμής (AR και MA) υποθέτουν ότι η χρονοσειρά είναι σταθερή, προϋπόθεση απαραίτητη για να μπορούν να διαχειριστούν αυτά τα μοντέλα οποιαδήποτε χρονοσειρά. Μια χρονοσειρά θεωρείται στάσιμη (stationary), όταν η μέση τιμή μ και η τυπική απόκλιση σ είναι σταθερές και η χρονοσειρά δεν εμφανίζει εποχικότητα (seasonality).

Στις περισσότερες περιπτώσεις, αυτές οι συνθήκες μπορούν να αναλυθούν γραφικά μελετώντας την γραφική παράσταση ως προς το χρόνο. Προφανώς δεν είναι όλες οι χρονοσειρές στάσιμες. Για τις περιπτώσεις αυτές λαμβάνει δράση ο όρος «I» του ARIMA.

- Integrated (I)

Ο όρος «I» σημαίνει ολοκληρωμένο, χωρίς όμως να σχετίζεται με την έννοια της μαθηματικής ολοκλήρωσης, και περιγράφει τη διαδικασία πρόβλεψης των διαφορών της σειράς από το ένα βήμα στο άλλο, αντί της πρόβλεψης της ίδιας της χρονοσειράς.

Η διαδικασία αυτή περιγράφεται από την εξίσωση:

$$Z_t = Y_{t+1} - Y_t \quad (37)$$

Στην περίπτωση αυτή έχει ληφθεί η διαφορά πρώτου βαθμού, δηλαδή μια μόνο φάση διαφοροποίησης διαδοχικών όρων. Η διαδικασία αυτή θα μπορούσε να επαναληφθεί πολλές φορές για να κάνει τη σειρά στάσιμη.

$$Z_t = Y_{t+1} - Y_t \quad \dots \quad d = 1$$

$$Q_t = Z_{t+1} - Z_t \quad \dots \quad d = 2 \quad (38)$$

Ο βαθμός διαφοροποίησης d αποτελεί μία πολύ σημαντική παράμετρο του ARIMA και καθορίζει την επιτυχία του μοντέλου.

Επομένως, το μοντέλο ARIMA παίρνει τη μορφή: ARIMA(p, d, q).

Το ARIMA είναι ένα ευρέως χρησιμοποιούμενο μοντέλο για πρόβλεψη χρονοσειρών εξαιτίας της απλότητάς του και της δυνατότητάς του να γενικεύει για μη στάσιμες χρονοσειρές. Παρόλα αυτά, ως στατιστική μέθοδος δεν ενδείκνυται για διαχείριση χρονοσειρών που παρουσιάζουν εποχικότητα.

4.4 SARIMA

Ένα SARIMA (Seasonal Auto Regressive Integrated Moving Average) μοντέλο διαφέρει ελάχιστα από ένα ARIMA μοντέλο. Η διαφοροποίηση αυτή έγκειται στον όρο S που αφορά την της εποχικότητας. Άλλωστε, πολλά dataset που προέρχονται από χρονοσειρές εμφανίζουν συχνά εποχικότητα [39].

Ένα μοντέλο είναι της μορφής: SARIMA(p, d, q)(P, D, Q) m , όπου οι όροι P, D και Q είναι ανάλογοι των p, d και q συμπεριλαμβάνοντας όμως τη συνιστώσα της εποχικότητας. Για την κατασκευή ενός τέτοιου μοντέλου ακολουθείται η εξής διαδικασία:

Ξεκινώντας από την εξίσωση πρώτου βαθμού της έκφρασης SARMA (όπου λείπει ο όρος I) έχουμε:

$$\begin{aligned} (1 - \Phi B^m)x_t &= (1 + \Theta B^m)w_t \\ x_t &= \Phi x_{t-m} + w_t + \Theta w_{t-m} \end{aligned} \quad (39)$$

Η έκφραση αυτή υποδηλώνει ότι το μοντέλο εμφανίζει αυτοσυσχέτιση σε προηγούμενες καθυστερήσεις πολλαπλασίων m (όπου m ορίζεται ως η περίοδος της εποχικότητας) τόσο για το μοντέλο AR όσο και για το MA. Έτσι, αυτό το μοντέλο θα ήταν SARMA (P, Q), όπου P και $Q = 1$.

Το πολλαπλασιαστικό SARMA μοντέλο, ορίζεται ως ο συνδυασμός ενός ARMA μοντέλου με εποχικότητα και ενός ARMA μοντέλου χωρίς εποχικότητα, δίνεται από την έκφραση:

$$\Phi_P(B^m)\phi(B)x_t = \Theta_Q(B^m)\theta(B)w_t \quad (40)$$

Σε ορισμένες χρονοσειρές ενδεχομένως η αυτοσυσχέτιση να φθίνει με τις καθυστερήσεις κατά τον πολλαπλασιασμό των περιόδων εποχικότητας.

Μία έκφραση μέσης εποχικότητας μπορεί να μοντελοποιηθεί ως:

$$\begin{aligned} x_t &= S_t + \omega_t \\ S_t &= S_{t-m} + v_t \end{aligned} \quad (41)$$

Ο όρος S_t αντιπροσωπεύει την εποχιακό στοιχείο το οποίο μεταβάλλεται κατά την περίοδο του m και αναπαριστά μία στοχαστική διαδικασία. Σημειώνεται πως οι όροι ω_t και v_t είναι μη συσχετιζόμενες διαδικασίες λευκού θορύβου. Η εφαρμογή μίας διαφοράς εποχικότητας στην συνάρτηση της χρονοσειράς έχει ως αποτέλεσμα ένα μοντέλο MA(1) με μία μέγιστη αυτοσυσχέτιση ίση με την καθυστέρηση της περιόδου m .

$$(1 - B^m)x_t = x_t - x_{t-m} = v_t + w_t - w_{t-m} \quad (42)$$

Επομένως ένα SARIMA μοντέλο μπορεί να εκφραστεί με τη μαθηματική έκφραση:

$$\Phi_P(B^m)\phi(B)\nabla_m^D\nabla^d x_t = \delta + \Theta_Q(B^m)\theta(B)w_t \quad (43)$$

Στο σημείο αυτό πρέπει να σημειωθεί πως για την επίλυση προβλημάτων πρόβλεψης χρονοσειράς, εκτός από τα στατιστικά μοντέλα που περιεγράφηκαν παραπάνω, χρησιμοποιούνται πληθώρα τεχνικών μηχανικής μάθησης, όπως έχει ήδη αναφερθεί στο προηγούμενο κεφάλαιο. Τέτοιες τεχνικές μπορεί να είναι μοντέλα βαθιάς μηχανικής μάθησης όπως τα συνελκτικά νευρωνικά δίκτυα (CNN) και LSTM (όπως έχει ήδη περιγραφεί σε προηγούμενα κεφάλαια), καθώς και μοντέλα όπως Random Forests, Support Vector Regression και Multiple Perceptron.

Κεφάλαιο 5 - Προσομοιώσεις

5.1 Εισαγωγή

Το ερευνητικό ενδιαφέρον της παρούσης διπλωματικής εργασίας επικεντρώθηκε στη μελέτη της πρόβλεψης της κίνησης στα οπτικά δίκτυα. Πιο συγκεκριμένα, η κίνηση αυτή αφορά την κίνηση μεταξύ δύο κόμβων στο επίπεδο IP ενός οπτικού δικτύου.

Η μελέτη της πρόβλεψης της κίνησης υλοποιήθηκε με τη χρήση τριών μοντέλων και με δεδομένα που αφορούν είτε την κίνηση στη διάρκεια της ημέρας είτε στη διάρκεια της εβδομάδας.

5.2 Κριτήρια επιλογής μοντέλων

Για την ημερήσια πρόβλεψη κίνησης χρησιμοποιήθηκε ένα συνελκτικό νευρωνικό μοντέλο CNN, ενώ για την εβδομαδιαία πρόβλεψη κίνησης χρησιμοποιήθηκε ένα στατιστικό μοντέλο SARIMA καθώς και ένα Vanilla LSTM νευρωνικό δίκτυο.

Το μοντέλο CNN επιλέχθηκε για την πρόβλεψη της κίνησης σε δεδομένα με διάρκεια μίας ημέρας καθώς ενδείκνυται για short term προβλέψεις. Οι προβλέψεις πραγματοποιήθηκαν για το test set.

Το στατιστικό μοντέλο SARIMA επιλέχθηκε έναντι ενός απλού στατιστικού μοντέλου ARIMA λόγω της εποχικότητας που παρουσιάζει η χρονοσειρά (S - Seasonality). Επίσης, επιλέχθηκε διότι παρέχει τη δυνατότητα πραγματοποίησης προβλέψεων για όλο το μήκος του dataset και όχι μόνο του test set. Οι προβλέψεις πραγματοποιήθηκαν για δεδομένα διάρκειας μίας εβδομάδας.

Το μοντέλο LSTM επιλέχθηκε για την πρόβλεψη κίνησης σε δεδομένα διάρκειας μίας εβδομάδας καθώς το μοντέλο αυτό ενδείκνυται για long term προβλέψεις. Επίσης, οι προβλέψεις πραγματοποιήθηκαν για το test set.

5.3 Software

Η παρούσα εργασία υλοποιήθηκε μέσω της γλώσσας προγραμματισμού Python [40]. Η Python αποτελεί μία ιδιαίτερα δημοφιλή γλώσσα προγραμματισμού και γνωρίζει ιδιαίτερη άνθιση στην εφαρμογή τεχνικών μηχανικής μάθησης για την επίλυση προβλημάτων [41]. Είναι μία γλώσσα γενικού σκοπού με κύριο γνώρισμα την ευκολία της στην κατανόηση εξαιτίας της ευαναγνωσιμότητάς της. Επίσης, είναι μία δυναμική γλώσσα και παρέχει τη δυνατότητα έρευνας και ανάπτυξης εφαρμογών στην ίδια γλώσσα, απλουστεύοντας τη μετάβαση από την έρευνα στην εκτέλεση.

Ο λόγος που επιλέχθηκε η συγκεκριμένη γλώσσα προγραμματισμού για την υλοποίηση της εργασίας έγκειται στην πληθώρα έτοιμων βιβλιοθηκών που διαθέτει και που

υποστηρίζουν προβλήματα μηχανικής μάθησης. Ένα πλεονέκτημα που την καθιστά ως την πιο δημοφιλή γλώσσα στον τομέα της μηχανικής μάθησης και γενικότερα της επιστήμης των δεδομένων [42].

Για την ανάγκη των προσομοιώσεων χρησιμοποιήθηκαν οι εξής βιβλιοθήκες:

- pandas και numpy για τη διαχείριση των datasets
- keras για τη δημιουργία των νευρωνικών μοντέλων
- statsmodels για τη δημιουργία του στατιστικού μοντέλου SARIMA
- scikit-learn για την εισαγωγή και διαχείριση των μετρικών
- matplotlib για την δημιουργία των γραφικών παραστάσεων

5.4 Συλλογή δεδομένων

Όπως αναφέρθηκε στα προηγούμενα κεφάλαια, ένας από τους πιο σημαντικούς παράγοντες της μηχανικής μάθησης είναι τα δεδομένα με τα οποία θα εκπαιδευτεί το μοντέλο μάθησης. Ιδανικά, το dataset αυτό θα πρέπει να προέρχεται από πραγματικά δεδομένα κίνησης. Η χρησιμοποίηση όμως πραγματικών δεδομένων κίνησης δικτύου πρακτικά είναι αρκετά δύσκολη [43]. Ο λόγος έγκειται στο γεγονός πως τα δεδομένα αυτά είναι σε μεγάλο βαθμό απόρρητα καθώς δεν δημοσιεύονται από τους πάροχους επικοινωνιακών υπηρεσιών. Για το λόγο αυτό, δηλαδή της μη διαθεσιμότητας πραγματικών δεδομένων κίνησης δικτύου, αποκτά ιδιαίτερη σημασία η δυνατότητα προσομοίωσης της κίνησης αυτής.

Αν και η δημιουργία της κίνησης αποτελεί εύκολη διαδικασία [44], είναι πολύ πιο δύσκολο το εγχείρημα παραγωγής κίνησης που εμφανίζει πραγματικά χαρακτηριστικά, όπως αυτά που παρατηρούνται στο Διαδίκτυο. Κατά το παρελθόν έχουν αναπτυχθεί αρκετές μέθοδοι παραγωγής κίνησης οι οποίες όμως επικεντρώνονται κυρίως στη δημιουργία αντιπροσωπευτικής IP κίνησης για δίκτυα που βασίζονται σε πακέτα ή στην άφιξη σύνδεσης με βάση τη κατανομή Poisson για δίκτυα μεταγωγής κυκλώματος [45]. Ωστόσο, για την αξιολόγηση της απόδοσης των αρχιτεκτονικών δικτύου απαιτείται η ενδιάμεση δημιουργία κίνησης (μεταξύ δημιουργίας πακέτων και της μοντελοποίησης άφιξης σύνδεσης) για τη δημιουργία συνεχούς κίνησης.

Στα πλαίσια της παρούσας εργασίας, στο πρώτο μέρος των προσομοιώσεων ακολουθήθηκε η διαδικασία αυτή, δηλαδή τα δεδομένα των μοντέλων που χρησιμοποιήθηκαν αποτελούν προϊόν προσομοιωμένης κίνησης. Στο δεύτερο μέρος χρησιμοποιήθηκε ένα dataset το οποίο παρέχεται δημόσια από την WAND Group.

5.4.1 Δημιουργία κίνησης και παραγωγή συνθετικών δεδομένων

Υπάρχουν διάφοροι τρόποι για την δημιουργία της κίνησης που περιγράψαμε. Ακολουθώντας την μέθοδο των συγγραφέων [46], θεωρήσαμε μία συνάρτηση $f(t; T)$ ως μία περιοδική συνάρτηση με περίοδο T που επιστρέφει τη μέση τιμή του μοντέλου συναρτήσεως του χρόνου κίνησης.

Επίσης, παρατηρούνται συνήθως ορισμένες τυχαίες τιμές γύρω από τη μέση τιμή ως αποτέλεσμα, μεταξύ άλλων, της διαδικασίας παρακολούθησης και καταγραφής. Αυτή η τυχαία συνάρτηση ε_t μπορεί να μοντελοποιηθεί ως μια συνάρτηση κατανομής πιθανότητας, όπως η κανονική (Gaussian) κατανομή.

Στην περίπτωση αυτή, $\varepsilon_t \sim (\mu, \sigma^2)$ ορίζεται από τη μέση τιμή μ και την τυπική απόκλιση σ . Επομένως, το περίπλοκο προφίλ κίνησης $Y(t; T)$ που θα αναπαραχθεί, μπορεί να δημιουργηθεί ως:

$$Y(t; T) = f(t; T) + \varepsilon_t \quad (44).$$

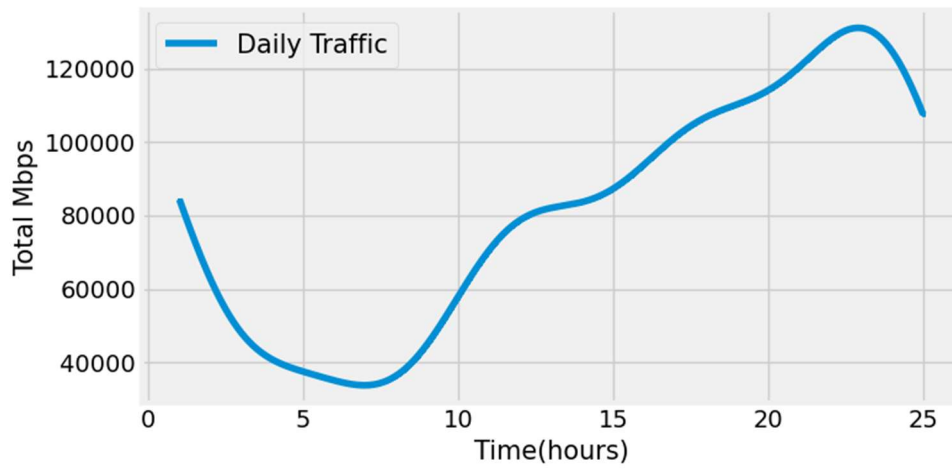
Η συνάρτηση $f(t; T)$ μπορεί να δημιουργηθεί με πολλούς τρόπους. Στην περίπτωσή μας τη μοντελοποιήσαμε ως άθροισμα ημιτόνων. Η συγκεκριμένη μέθοδος χρησιμοποιείται για την μοντελοποίηση περιοδικής κίνησης λόγω του επαναλαμβανόμενου μοτίβου της [47]:

$$f(t; T) = a_0 + \sum_{k=1}^K a_k \cdot \sin \left[2 \cdot \pi \cdot k \frac{t-t_k}{T} \right] \quad (45),$$

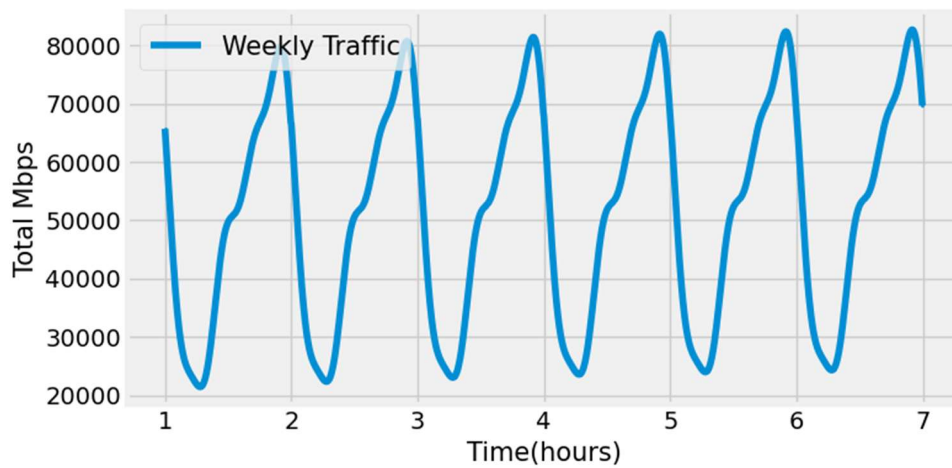
όπου k είναι ο δείκτης για τον αντίστοιχο όρο στην έκταση των ημιτονοειδών σειρών, a_0 είναι η μέση ένταση της κυκλοφορίας, a_k το πλάτος, t_k η μετατόπιση χρόνου (σε δευτερόλεπτα) και T η ημερήσια περίοδος (σε δευτερόλεπτα).

Για την μοντελοποίηση της παραπάνω περιοδικότητας για περισσότερες από μία ημέρες, υποθέτουμε πως το προφίλ των επόμενων ημερών είναι ίδιο με της πρώτης. Ωστόσο, η ένταση της κίνησης συνήθως αυξάνεται, ειδικά όταν αναφερόμαστε σε περιόδους μεγάλης διάρκειας.

Στις παρακάτω γραφικές παραστάσεις απεικονίζεται η κίνηση για διάστημα μίας ημέρας και η κίνηση μίας εβδομάδας:



Εικόνα 44: Γραφική παράσταση κίνησης συνθετικών δεδομένων διάρκειας 1 ημέρας

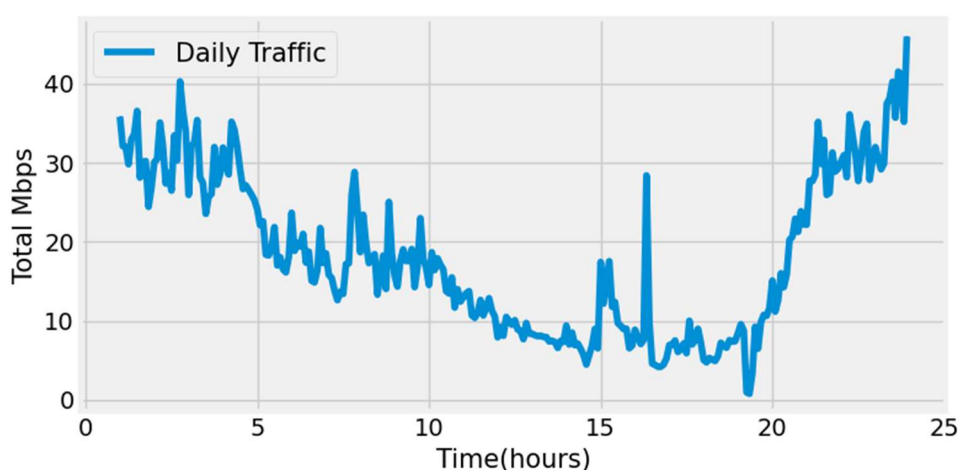


Εικόνα 45: Γραφική παράσταση κίνησης συνθετικών δεδομένων διάρκειας 1 εβδομάδας

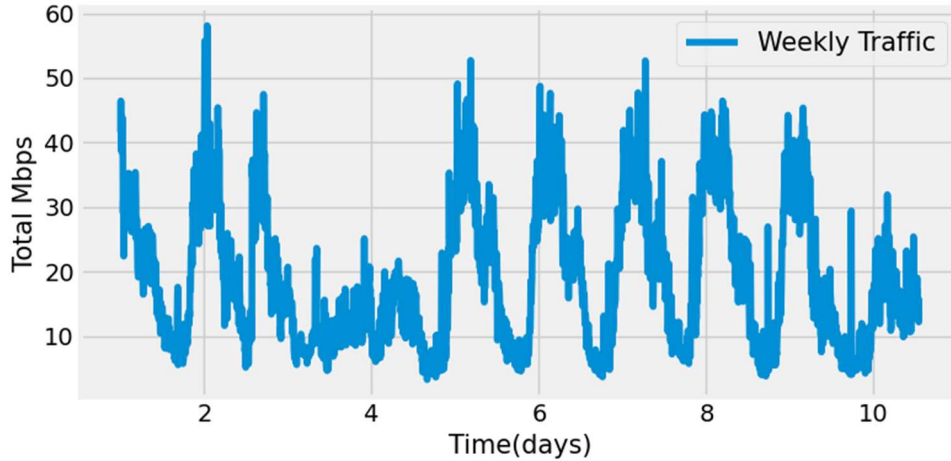
5.4.2 Πραγματικά δεδομένα κίνησης

Εκτός από τα συνθετικά δεδομένα, για την παρούσα εργασία χρησιμοποιήθηκε ένα σετ δεδομένων από το αποθετήριο WITS (Waikato Internet Traffic Storage) [48] του Wand Group. Η Wand είναι μία ερευνητική ομάδα του πανεπιστημίου του Γουαϊκάτο της Νέας Ζηλανδίας και συγκεκριμένα του τμήματος Επιστήμης των Υπολογιστών. Η ομάδα αυτή ασχολείται με μια σειρά πρότζεκτ σχετικά με τα δίκτυα υπολογιστών που εστιάζονται κυρίως στις μετρήσεις του δικτύου. Στόχος του αποθετηρίου WITS είναι η συλλογή και η καταγραφή όλων των στοιχείων κίνησης στο διαδίκτυο τα οποία διαθέτει και επεξεργάζεται η ομάδα Wand με σκοπό να τα καταστήσουν διαθέσιμα για πραγματοποίηση ερευνών.

Τα δεδομένα του συγκεκριμένου dataset αφορούν δεδομένα κίνησης 11 ημερών. Στις παρακάτω γραφικές παραστάσεις απεικονίζεται η κίνηση της πρώτης μέρας καθώς επίσης και η κίνηση για όλη τη διάρκεια των δεδομένων, δηλαδή την κίνηση 11 ημερών:



Εικόνα 46:Γραφική παράσταση πραγματικής κίνησης δεδομένων διάρκειας 1 ημέρας



Εικόνα 47: Γραφική παράσταση πραγματικής κίνησης δεδομένων διάρκειας 11 ημερών

5.5 Εκτίμηση απόδοσης μοντέλων

Η εκτίμηση της απόδοσης των μοντέλων έγινε μέσω των μετρικών RMSE (root mean squared error) και RRSE (root relevant squared error). Ο λόγος επιλογής των μετρικών αυτών έγκειται στο γεγονός πως η μετρική RMSE είναι η πιο κοινή μετρική εκτίμησης μοντέλων. Παρόλα αυτά όμως, εξαιτίας του γεγονότος πως στον υπολογισμό της λαμβάνεται υπόψιν και το ύψος των παρατηρήσεων, για τη σύγκριση των αποτελεσμάτων των διαφορετικών dataset χρησιμοποιήθηκε η μετρική RRSE η οποία εκτιμά το μοντέλο ανεξαρτήτως του ύψους των παρατηρήσεων.

RMSE (Root Mean Squared Error)

Το RMSE, δηλαδή η ρίζα μέσου τετραγωνικού σφάλματος, υπολογίζεται από τη μαθηματική έκφραση:

$$E_i = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_{(ji)} - T_j)^2} \quad (46),$$

όπου $P_{(ji)}$ η τιμή που προβλέπεται και T_j είναι ο στόχος. Ιδανική τιμή για την μετρική αυτή είναι η τιμή 0.

RRSE (Root Relevant Squared Error)

Το RRSE, δηλαδή η ρίζα του σχετικού τετραγωνικού σφάλματος, σχετίζεται με το αποτέλεσμα που θα προέκυπτε από την εφαρμογή μίας απλής πρόβλεψης. Πιο συγκεκριμένα, αυτή η απλή πρόβλεψη είναι ο μέσος όρος όλων των πραγματικών τιμών. Με απλά λόγια, το RRSE είναι ο λόγος της ρίζας του τετραγωνικού σφάλματος

της μεθόδου που χρησιμοποιήθηκε προς τη ρίζα του τετραγωνικού σφάλματος της απλής πρόβλεψης.

Επομένως, το σχετικό τετραγωνικό σφάλμα παίρνει το συνολικό τετράγωνο σφάλμα και το κανονικοποιεί διαιρώντας με το συνολικό τετράγωνο σφάλμα της απλής πρόβλεψης. Λαμβάνοντας την τετραγωνική ρίζα του σχετικού τετραγώνου σφάλματος μειώνεται το σφάλμα στις ίδιες διαστάσεις με την ποσότητα που προβλέπεται. Ιδανική τιμή για τη μετρική αυτή θεωρείται η τιμή 0, κατά την οποία όλες οι προβλέψεις ταυτίζονται με τις πραγματικές τιμές. Οποιαδήποτε τιμή μικρότερη της μονάδας θεωρείται καλύτερη πρόβλεψη από την βασική πρόβλεψη. Είναι προφανές πως όσο μικρότερη είναι η τιμή του RRSE τόσο καλύτερη είναι η πρόβλεψη.

Η μαθηματική έκφραση της μετρικής (E_i), είναι:

$$E_i = \sqrt{\frac{\sum_{j=1}^n (P_{(ij)} - T_j)^2}{\sum_{j=1}^n (T_j - \bar{T})^2}} \quad (47),$$

όπου $P_{(ij)}$ είναι η τιμή που προβλέπεται, T_j ο στόχος και ο όρος \bar{T} δίνεται από τη σχέση:

$$\bar{T} = \frac{1}{n} \sum_{j=1}^n T_j \quad (48).$$

5.6 Πραγματοποίηση Προσομοιώσεων

Όπως ήδη αναφέρθηκε, η πρόβλεψη κίνησης πραγματοποιήθηκε με χρήση δύο ξεχωριστών dataset, ενός πραγματικού και ενός συνθετικού βάσει της διαδικασίας που περιεγράφηκε. Και στις δύο περιπτώσεις dataset εφαρμόστηκε η ίδια αρχιτεκτονική μοντέλου.

Στην περίπτωση του συνθετικού dataset πραγματοποιήθηκε δειγματοληψία με σκοπό την καλύτερη εκπαίδευση του μοντέλου καθώς επίσης και λόγω του πολύ μεγάλου αριθμού παρατηρήσεων του συνθετικού dataset. Τα δεδομένα αυτά, αφορούν παρατηρήσεις ανά δευτερόλεπτο συνεπώς και για τη διάρκεια της ημέρας καθώς και για τη διάρκεια της εβδομάδας πρόκειται τεράστιος αριθμός παρατηρήσεων. Για την διαχείριση ενός τέτοιου όγκου dataset απαιτείται και τεράστια υπολογιστική ισχύ. Για το λόγο αυτό πραγματοποιήθηκε δειγματοληψία στις παρατηρήσεις ανά 300 παρατηρήσεις, δηλαδή 300 δευτερόλεπτα. Συνεπώς η περίοδος δειγματοληψίας είναι 5 λεπτά.

Η διαδικασία της δειγματοληψίας προέκυψε με χρήση των εντολών:

```
df = pd.read_csv('my_data.csv')
df.columns= ['Traffic']
series = df.loc[df.index % 300 == 0]
data = series.values
```

Στην περίπτωση του dataset που αφορά πραγματικά δεδομένα κίνησης, δεν πραγματοποιήθηκε δειγματοληψία καθώς η κάθε παρατήρηση απείχε 5 λεπτά από την επόμενη, οπότε πρόκειται η αναλογία με τα συνθετικά δεδομένα μετά τη δειγματοληψία.

Επίσης, για τα μοντέλα CNN και LSTM, και τα δύο dataset χωρίστηκαν σε ποσοστό 80-20. Το 80% αφορά το training set και το 20% αφορά το test set. Η εκπαίδευση των μοντέλων πραγματοποιήθηκε για στο πρώτο 80% του dataset, και οι προβλέψεις πραγματοποιήθηκαν στο τελευταίο 20%. Η διαδικασία αυτή πραγματοποιήθηκε μέσω των εντολών:

```
#split percentage of training-test set
split_prc = 0.8
# split dataset
train, test = data[:-round(len(data)*split_prc)], data[round(len(data)*split_prc):]
```


5.6.1 Πρόβλεψη κίνησης με χρήση CNN Νευρωνικού Δικτύου

Ένα CNN μοντέλο μπορεί να χρησιμοποιηθεί για μονοδιάστατα δεδομένα, όπως ακολουθίες κειμένου και χρονοσειρές. Σε αυτήν την περίπτωση, θα χρησιμοποιηθεί για την πρόβλεψη της κίνησης που θα ανακτήθηκαν από δεδομένα μίας ημέρας.

Οι παράμετροι του μοντέλου που πρέπει να ρυθμιστούν είναι οι εξής:

- `n_input`: Ο αριθμός των καθυστερημένων παρατηρήσεων που χρησιμοποιεί το μοντέλο ως είσοδο. Στο μοντέλο αυτό, η παράμετρος `n_input` ορίστηκε `n_input=12` το οποίο σημαίνει πως κάθε παρατήρηση λαμβάνει υπόψη την κίνηση της προηγούμενης ώρας ($12 \times 5 \text{min} = 1 \text{h}$).
- `n_filters`: Ο αριθμός των παράλληλων φίλτρων.
- `n_kernel`: Ο αριθμός των χρονικών βημάτων που λαμβάνονται υπόψη σε κάθε ανάγνωση της ακολουθίας εισόδου.
- `n_epochs`: Το πλήθος των επαναλήψεων για την έκθεση του μοντέλου σε ολόκληρο το training set.
- `n_batch`: Ο αριθμός των δειγμάτων μέσα σε μία επανάληψη μετά από τον οποίον ενημερώνονται τα βάρη.

Συνεπώς για το μοντέλο, επιλέχθηκε η εξής ρύθμιση:

```
# define config
config = [12, 256, 3, 10, 100]
```

Για τον ορισμό του μοντέλου με σκοπό την εκπαίδευση του στα training data, επιλέχθηκε μία συνάρτηση κόστους MSE, μία συνάρτηση βελτιστοποίησης ADAM και συνάρτηση ενεργοποίησης ReLU.

```
# define model
model = Sequential()
model.add(Conv1D(filters=n_filters, kernel_size=n_kernel,
activation='relu', input_shape=(n_input, 1)))
model.add(Conv1D(filters=n_filters, kernel_size=n_kernel,
activation='relu'))
model.add(MaxPooling1D(pool_size=2))
model.add(Flatten())
model.add(Dense(1))
model.compile(loss='mse', optimizer='adam')

# fit
model.fit(train_x, train_y, epochs=n_epochs,
batch_size=n_batch, verbose=0)
```

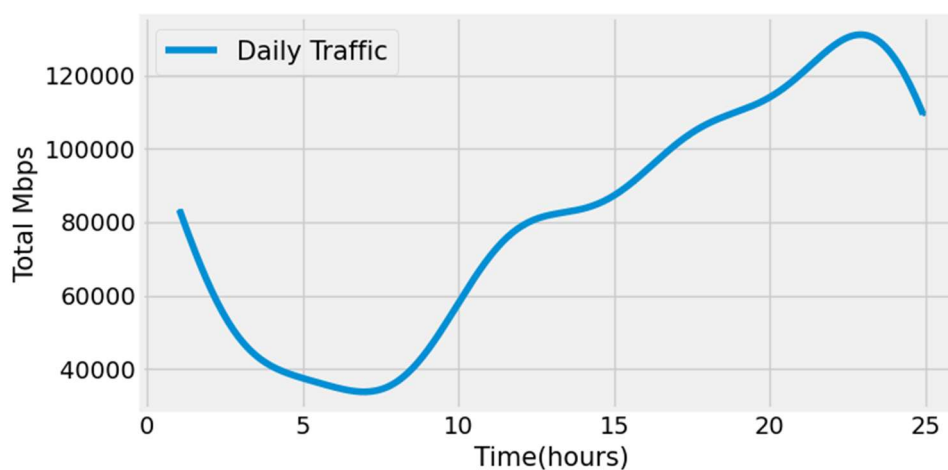
Για την δημιουργία των προβλέψεων, κλήθηκε η συνάρτηση:

```
# forecast with a pre-fit model
def model_predict(model, history, config):
    # unpack config
    n_input, _, _, _, _ = config
    # prepare data
    x_input = array(history[-n_input:]).reshape((1, n_input,
1))
    # forecast
    yhat = model.predict(x_input, verbose=0)
    return yhat[0]

# seed history with training dataset
history = [x for x in train]
# step over each time-step in the test set
for i in range(len(test)):
    # fit model and make forecast for history
    yhat = model_predict(model, history, config)
    # store forecast in list of predictions
    predictions.append(yhat)
    # add actual observation to history for the next loop
    history.append(test[i])
```

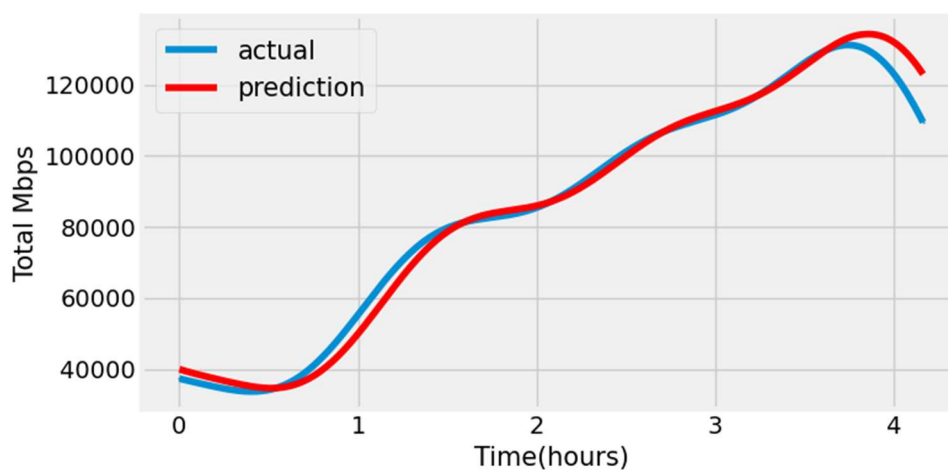
Συνθετικά δεδομένα

Η μορφή της κίνησης που προέκυψε μετά τη δειγματοληψία, απεικονίζεται στην παρακάτω γραφική παράσταση:



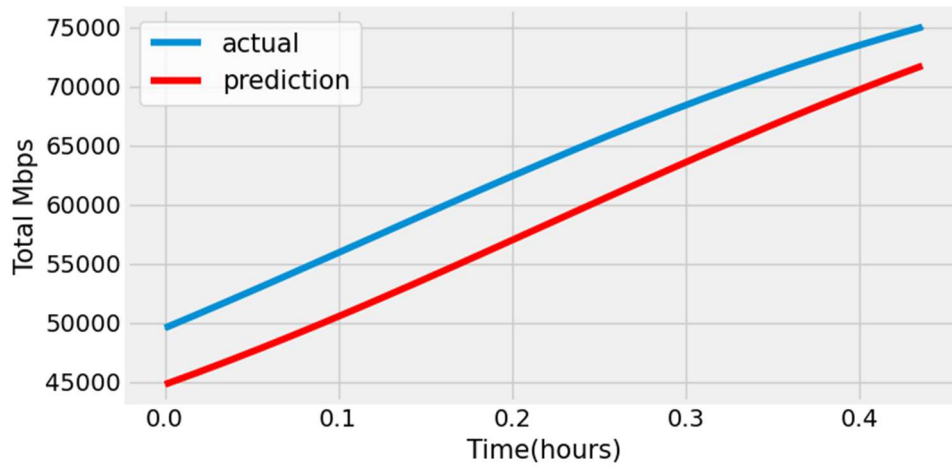
Εικόνα 48: Γραφική παράσταση κίνησης δεδομένων διάρκειας 1 ημέρας μετά τη δειγματοληψία

Με δεδομένα εισόδου που προήλθαν από την προσομοίωση της κίνησης, παρουσιάζεται στο παρακάτω γράφημα η πρόβλεψη της κίνησης για το σύνολο των παρατηρήσεων στο test set:



Εικόνα 49: Πρόβλεψη κίνησης για δεδομένα εισόδου μίας μέρας

Στην παρακάτω γραφική παράσταση παρουσιάζεται τμήμα της κίνησης με μεγαλύτερη λεπτομέρεια:



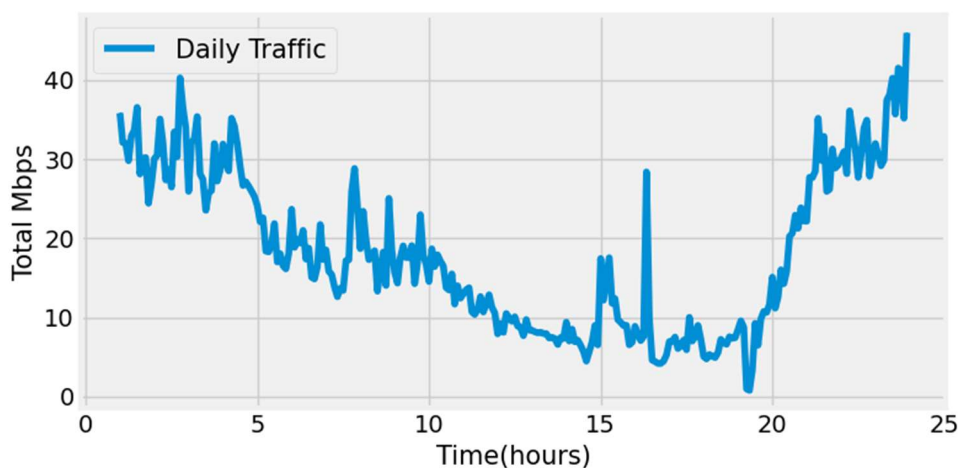
Εικόνα 50: Πρόβλεψη κίνησης

Για την αξιολόγηση των μοντέλων εφαρμόστηκαν οι συναρτήσεις των μετρικών. Τα αποτελέσματα των μετρικών παρουσιάζονται παρακάτω:

Test RMSE: 3430.134
Test RRSE: 0.107

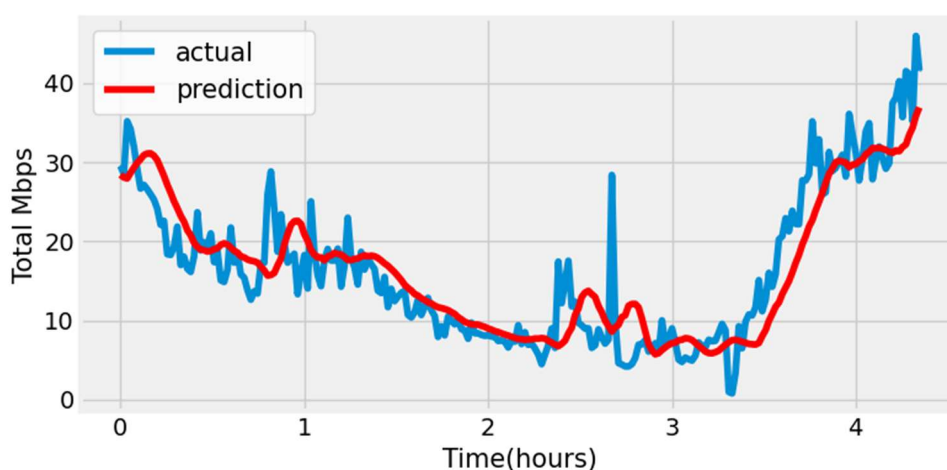
Πραγματικά δεδομένα

Η ημερήσια κίνηση με βάση τα πραγματικά δεδομένα παρουσιάζεται στην παρακάτω γραφική παράσταση:



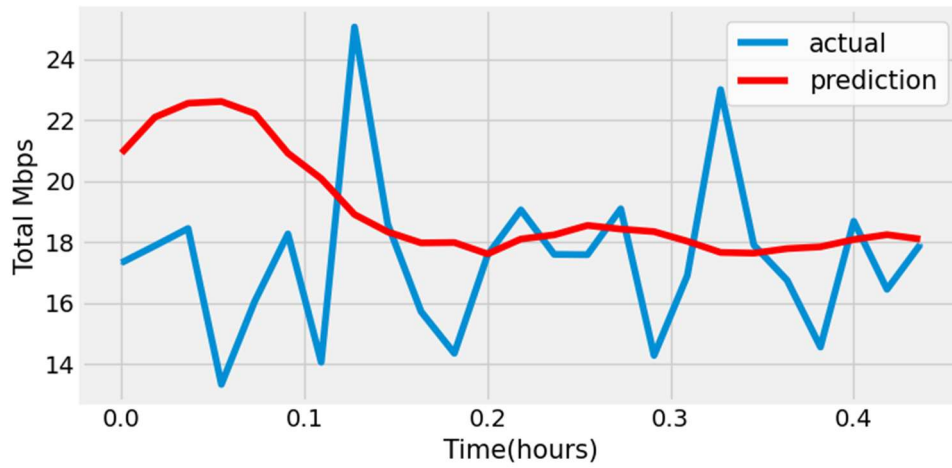
Εικόνα 51: Γραφική παράσταση κίνησης δεδομένων διάρκειας 1 ημέρας

Εφαρμόζοντας την ίδια αρχιτεκτονική του μοντέλου CNN, πραγματοποιήθηκαν προβλέψεις σε όλο το μήκος του test set. Οι προβλέψεις παρουσιάζονται στην παρακάτω γραφική παράσταση:



Εικόνα 52: Πρόβλεψη κίνησης για δεδομένα εισόδου μίας μέρας

Με μεγαλύτερη ακρίβεια, παρουσιάζονται οι προβλέψεις για μικρότερο τμήμα της κίνησης:



Εικόνα 53: Πρόβλεψη κίνησης

Εφαρμόζοντας τις συναρτήσεις των μετρικών, λήφθηκαν οι επιδόσεις αξιολόγησης των μοντέλων:

Test RMSE: 4.370
Test RRSE: 0.460

5.6.2 Πρόβλεψη κίνησης με χρήση SARIMA μοντέλου

Η πρόβλεψη της κίνησης με χρήση SARIMA μοντέλου αφορά την κίνηση σε δεδομένα κίνησης μίας εβδομάδας.

Για το μοντέλο αυτό, αρχικά εξασφαλίστηκε πως είναι στάσιμο (stationary), απαραίτητη προϋπόθεση για την πρόβλεψη της χρονοσειράς.

Για την εξασφάλιση της στασιμότητας υπάρχουν δύο βασικές μέθοδοι:

- Rolling Statistics: Γραφικά, με το σχεδιασμό του κινούμενου μέσου (rolling mean) και κινούμενης τυπικής απόκλισης (rolling standard deviation), με στόχο να είναι κατά το δυνατόν παράλληλα με τον άξονα X.
- ADF Test: Κατά τη μέθοδο ADF (Augmented Dickey-Fuller) Test, μία χρονοσειρά θεωρείται στάσιμη όταν η τιμή p-value είναι μικρότερη του threshold (0.05) και οι κρίσιμες τιμές 1%, 5% και 10% να είναι κοντά στην τιμή ADF Statistics.

Η συνάρτηση που πραγματοποιεί την εξής διαδικασία περιγράφεται από την συνάρτηση:

```
def get_stationarity(timeseries):  
  
    # rolling statistics  
    rolling_mean = timeseries.rolling(window=12).mean()  
    rolling_std = timeseries.rolling(window=12).std()  
  
    # rolling statistics plot  
    original = plt.plot(timeseries, color='blue', label='Original')  
    mean = plt.plot(rolling_mean, color='red', label='Rolling Mean')  
    std = plt.plot(rolling_std, color='black', label='Rolling Std')  
    plt.legend(loc='best')  
    plt.title('Rolling Mean & Standard Deviation')  
    plt.show(block=False)  
  
    # Dickey-Fuller test:  
    result = adfuller(timeseries['Traffic'])  
    print('ADF Statistic: {}'.format(result[0]))
```

```

print('p-value: {}'.format(result[1]))
print('Critical values:')
for key, value in result[4].items():
    print('\t{}: {}'.format(key, value))

```

Το πλεονέκτημα ενός SARIMA μοντέλου είναι πως ακόμα και να μην είναι στάσιμη μία χρονοσειρά, έχει τη δυνατότητα με εφαρμογή κατάλληλων συναρτήσεων να την μετατρέψει σε στάσιμη. Η μετατροπή της χρονοσειράς σε στάσιμη μπορεί να γίνει μέσω της εκθετικής αποσύνθεσης (exponential decay) και κατ' επέκταση αφαιρώντας τον κινούμενο μέσο (rolling mean).

Η διαδικασία της εκθετικής αποσύνθεσης περιγράφεται με τις εντολές:

```

rolling_mean_exp_decay = df.ewm(halflife=12, min_periods=0,
adjust=True).mean()
df_exp_decay = df - rolling_mean_exp_decay
df_exp_decay.dropna(inplace=True)
fig, axes = plt.subplots(figsize=(12,3))
get_stationarity(df_exp_decay)

```

Η διαδικασία της αφαίρεσης του κινούμενου μέσου περιγράφεται εντολές:

```

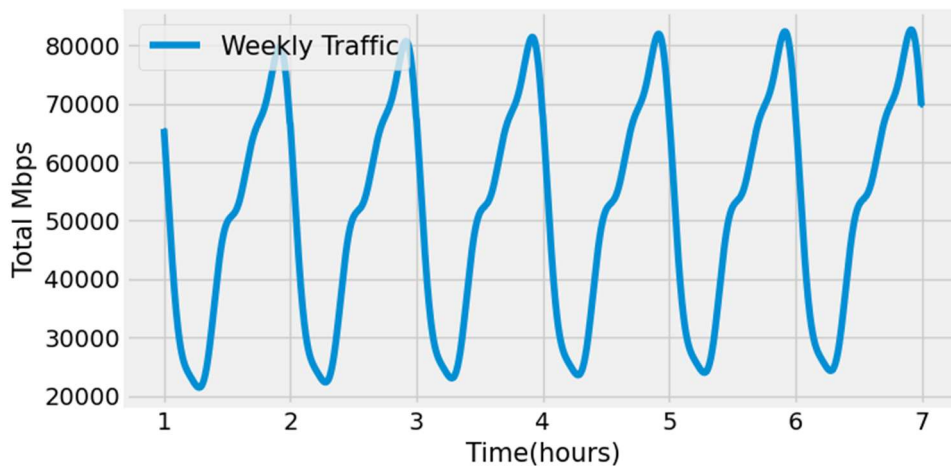
df_shift = df - df.shift()
df_shift.dropna(inplace=True)
fig, axes = plt.subplots(figsize=(12,3))
get_stationarity(df_shift)

```

Η διαδικασία ελέγχου στασιμότητας θα περιγραφεί για κάθε dataset ξεχωριστά.

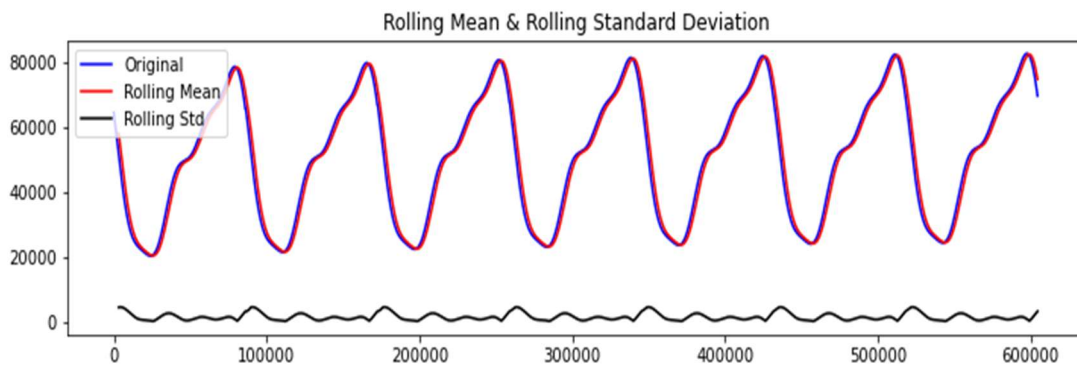
Συνθετικά δεδομένα

Η μορφή της κίνησης μετά τη δειγματοληψία παρουσιάζεται στην παρακάτω γραφική παράσταση:



Εικόνα 54: Γραφική παράσταση κίνησης δεδομένων διάρκειας 1 εβδομάδας μετά τη δειγματοληψία

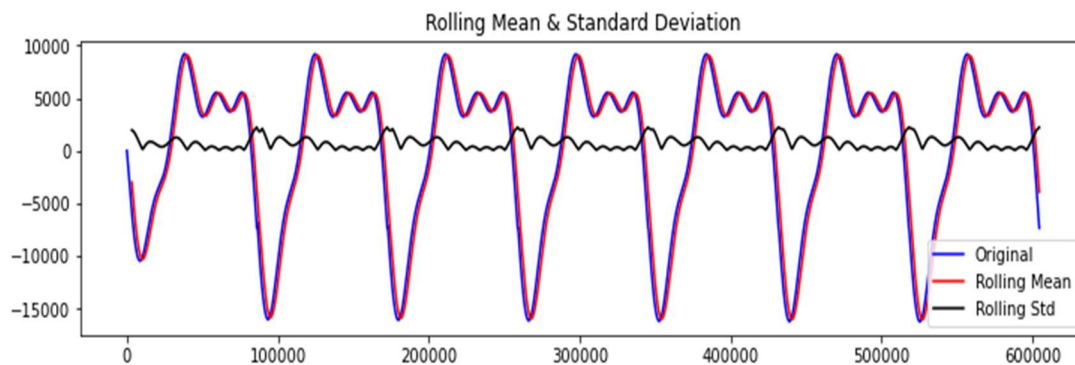
Για την διερεύνηση της στασιμότητας της χρονοσειράς, ακολουθήθηκε η διαδικασία γραφικού σχεδιασμού και ADF Test:



ADF Statistic: -7.830817002375212
 p-value: 6.300255821390691e-12
 Critical Values:
 1%: -3.4336419856688276
 5%: -2.862994217611322
 10%: -2.567544164035933

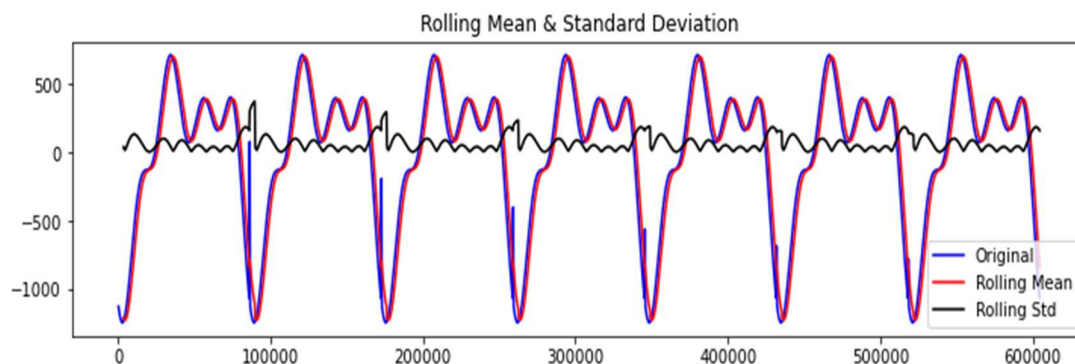
Όπως παρατηρείται, ενώ τη γραφική παράσταση της rolling std είναι σχετικά παράλληλη στον άξονα των X, η γραφική παράσταση του rolling mean ακολουθεί γραφικά την χρονοσειρά. Επίσης, παρόλο που η τιμή της p-value είναι αρκετά χαμηλότερη από το κατώφλι (0.05), οι κρίσιμες τιμές απέχουν από την τιμή ADF Statistic. Συνεπώς συμπεραίνεται πως η χρονοσειρά δεν είναι στάσιμη και θα εφαρμοστούν τεχνικές μετατροπής της.

Συνεχίζοντας με τη διαδικασία της εκθετικής αποσύνθεσης, προκύπτει:



ADF Statistic: -5.454838106488585
 p-value: 2.5955989920450155e-06
 Critical Values:
 1%: -3.4336419856688276
 5%: -2.862994217611322
 10%: -2.567544164035933

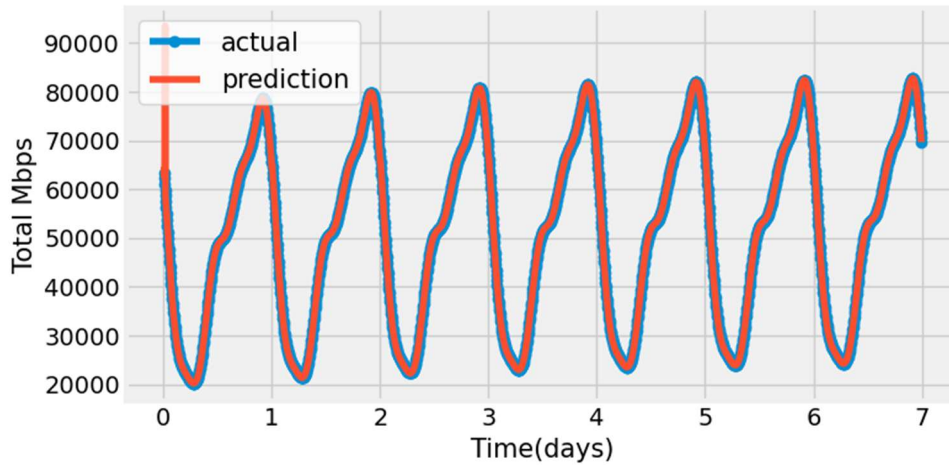
Παρατηρείται πως η χρονοσειρά παραμένει μη στάσιμη, οπότε η διαδικασία προχωρά με την αφαίρεση του rolling mean, οπότε προκύπτει:



ADF Statistic: -3.5396360441086627
 p-value: 0.007025629967796359
 Critical Values:
 1%: -3.433643643742798
 5%: -2.862994949652858
 10%: -2.5675445538118042

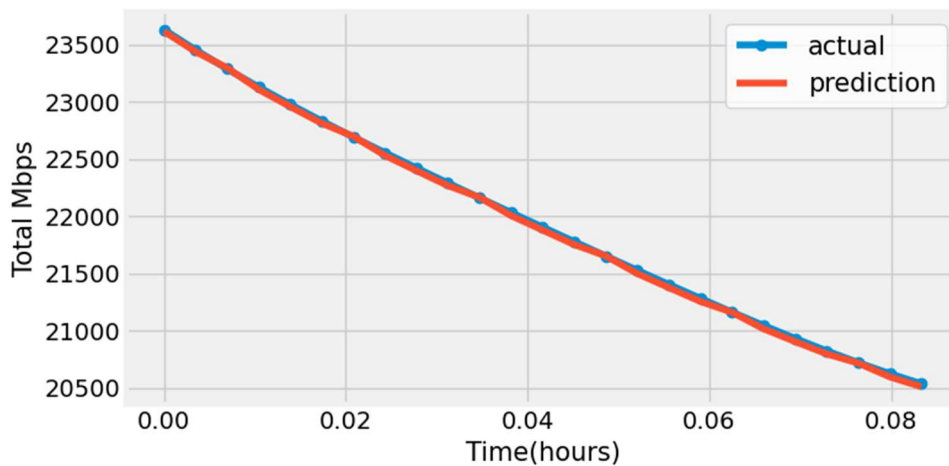
Όπως παρατηρείται, η τιμή p-value παραμένει κάτω από το κατώφλι που έχει οριστεί, καθώς και οι κρίσιμες τιμές έχουν συγκλίνει στην τιμή ADF Statistic. Συνεπώς πληρείται το κριτήριο στασιμότητας και η χρονοσειρά μπορεί να θεωρηθεί στάσιμη.

Εν συνεχεία, ακολουθήθηκε η διαδικασία δημιουργίας του μοντέλου και ρύθμισης των υπερπαραμέτρων, όπως αυτή περιεγράφηκε, οπότε προέκυψαν οι προβλέψεις για όλη τη διάρκεια της κίνησης:



Εικόνα 55: Πρόβλεψη κίνησης για δεδομένα εισόδου μίας εβδομάδας

Οι προβλέψεις παρουσιάζονται πιο αναλυτικά σε μικρότερο τμήμα της κίνησης:



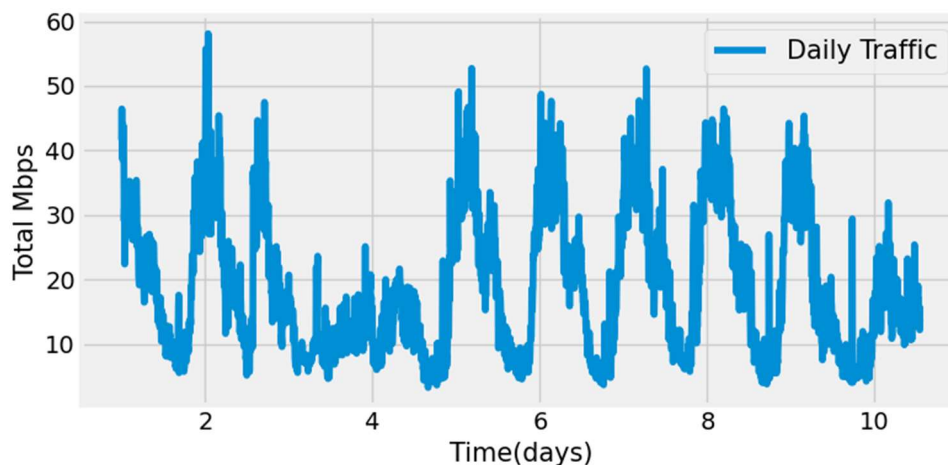
Εικόνα 56: Πρόβλεψη κίνησης

Από την εφαρμογή των συναρτήσεων των μετρικών, προέκυψαν τα εξής αποτελέσματα στην αξιολόγηση του μοντέλου:

Test RMSE: 1622.190
Test RRSE: 0.085

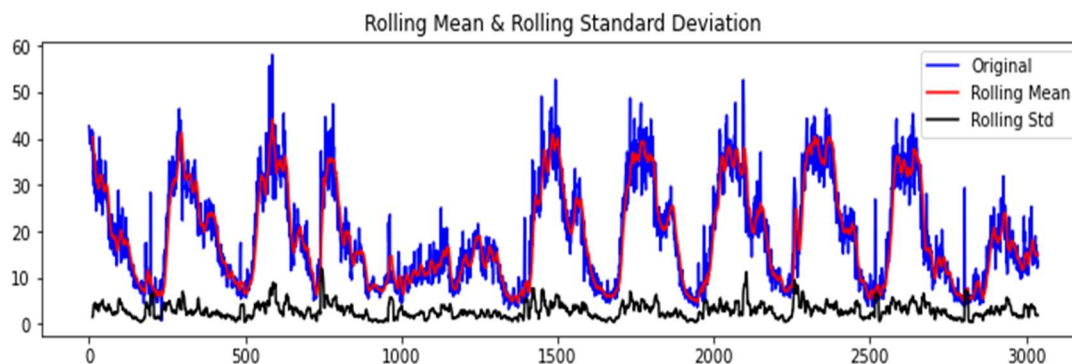
Πραγματικά δεδομένα

Η κίνηση για όλο το μήκος των δεδομένων του πραγματικού dataset παρουσιάζεται στην παρακάτω γραφική παράσταση:



Εικόνα 57: Γραφική παράσταση κίνησης δεδομένων διάρκειας 11 ημερών

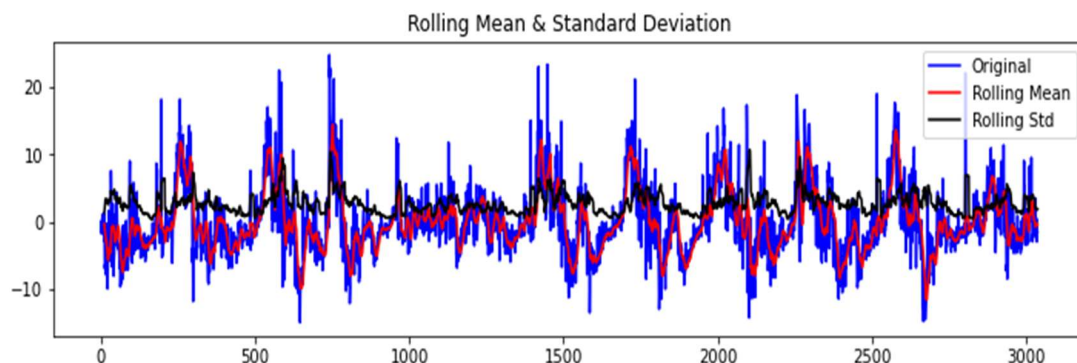
Κάνοντας έλεγχο του κριτηρίου στασιμότητας, προκύπτει:



```
ADF Statistic: -4.340070386110291
p-value: 0.00037834943862740365
Critical Values:
 1%: -3.432509302788687
 5%: -2.8624940419851748
10%: -2.567277861021724
```

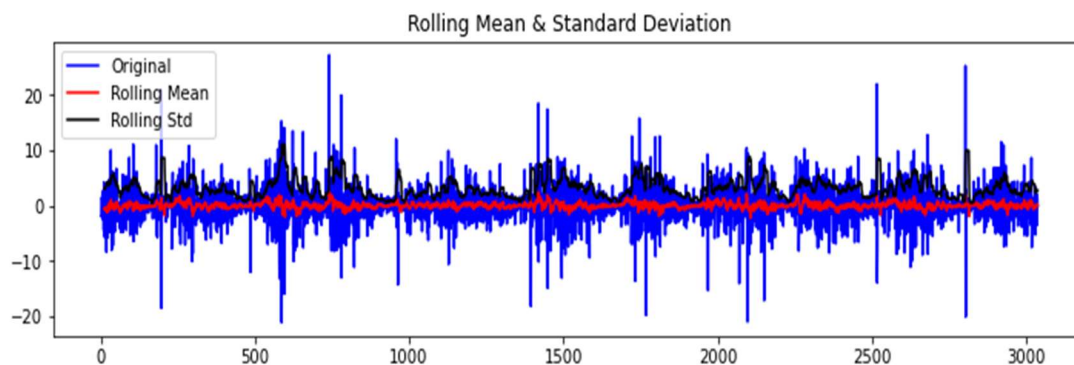
Παρά το γεγονός πως το κριτήριο ADF Test πληρείται καθώς η τιμή της p-value είναι μικρότερη του threshold και οι κρίσιμες τιμές τείνουν στην τιμή ADF Statistic, θα πραγματοποιηθεί η περαιτέρω μετατροπή της χρονοσειράς σε στάσιμη.

Συνεχίζοντας τη διαδικασία της εκθετικής αποσύνθεσης, προκύπτει:



```
ADF Statistic: -7.20860182116591
p-value: 2.265698122031941e-10
Critical Values:
1%: -3.432512873738261
5%: -2.862495619160489
10%: -2.5672787006900757
```

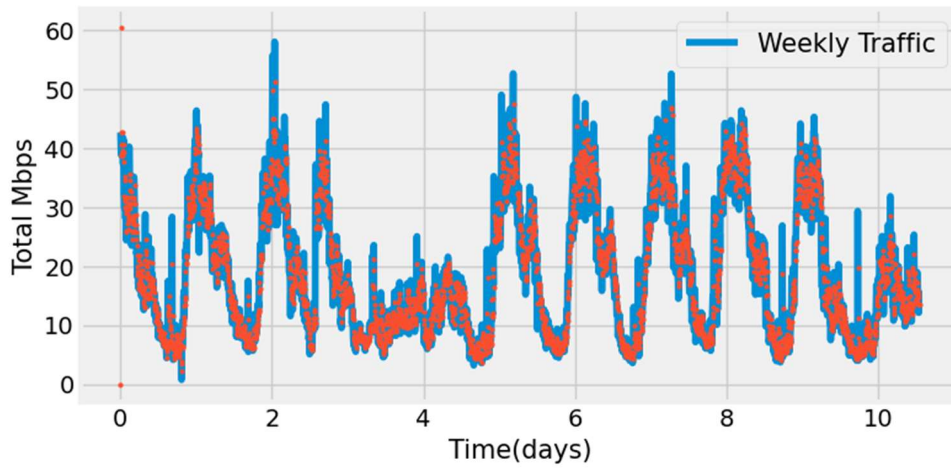
Στο σημείο αυτό παρατηρείται πως ενώ πληρείται το κριτήριο ADF Test σε μικρότερο βαθμό μετά την εφαρμογή της εκθετικής αποσύνθεσης, οι κυματομορφές της rolling mean και rolling std τείνουν να γίνουν παράλληλες στον άξονα X. Συνεπώς η διαδικασία προχώρησε στην αφαίρεση του rolling mean:



```
ADF Statistic: -2.94621923547664
p-value: 0.0
Critical Values:
1%: -3.432509302788687
5%: -2.8624940419851748
10%: -2.567277861021724
```

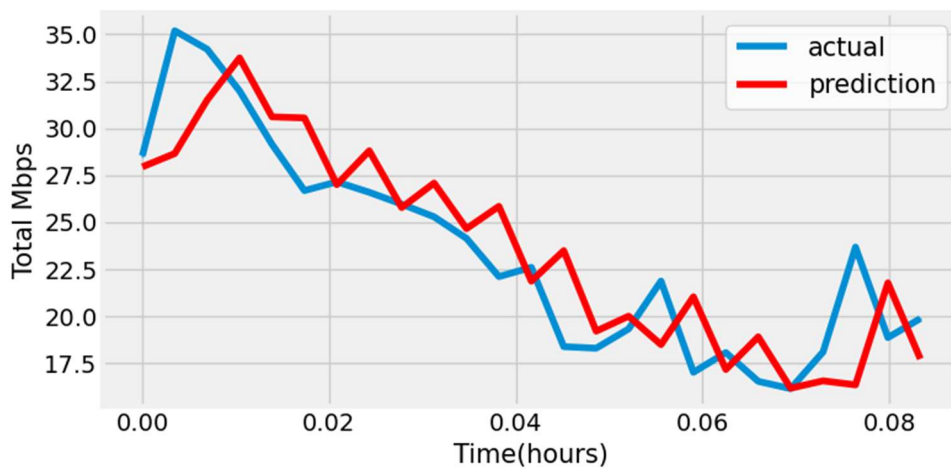
Όπως παρατηρείται πληρούνται και τα δύο κριτήρια και η χρονοσειρά πλέον θεωρείται στάσιμη.

Εν συνεχεία, ακολουθήθηκε η διαδικασία ρύθμισης των υπερπαραμέτρων, όπως αυτή περιεγράφηκε, οπότε προέκυψαν οι προβλέψεις για όλη τη διάρκεια της κίνησης:



Εικόνα 58: Πρόβλεψη κίνησης για δεδομένα εισόδου 11 ημερών

Οι προβλέψεις παρουσιάζονται πιο αναλυτικά σε μικρότερο κομμάτι της κίνησης:



Εικόνα 59: Πρόβλεψη κίνησης

Για την αξιολόγηση των μοντέλων εφαρμόστηκαν οι συναρτήσεις των μετρικών. Σύμφωνα με τις μετρικές, προκύπτουν τα αποτελέσματα:

Test RMSE: 3.382
 Test RRSE: 0.307

5.6.3 Πρόβλεψη κίνησης με χρήση LSTM Νευρωνικού Δικτύου

Για την πρόβλεψη κίνησης χρησιμοποιήθηκε μια ειδική μορφή LSTM δικτύου, το Vanilla LSTM νευρωνικό δίκτυο, το οποίο διαθέτει μόνο ένα εσωτερικό «κρυμμένο» επίπεδο, καθώς και ένα ακόμα επίπεδο εξόδου για τις προβλέψεις.

Η πρόβλεψη πραγματοποιήθηκε με χρήση συνθετικών δεδομένων για διάρκεια μιας εβδομάδας έπειτα από δειγματοληψία, καθώς και για το πλήρες dataset διάρκειας 11 ημερών.

Για τον ορισμό του μοντέλου με σκοπό την εκπαίδευση του στα training data, επιλέχθηκε μία συνάρτηση κόστους MSE, μία συνάρτηση βελτιστοποίησης ADAM και συνάρτηση ενεργοποίησης ReLU.

```
# define model
model = Sequential()
n_features = 1
model.add(LSTM(50, activation='relu', return_sequences=True,
input_shape=(n_steps, n_features)))
model.add(LSTM(50, activation='relu'))
model.add(Dense(1))
model.compile(optimizer='adam', loss='mse')
```

Επίσης, το μοντέλο ρυθμίστηκε έτσι ώστε για κάθε πρόβλεψη να λαμβάνονται υπόψη 5 προηγούμενες παρατηρήσεις:

```
# define input sequence
raw_seq = df.Traffic.values
# choose a number of time steps
n_steps = 5
# split into samples
x, y = split_sequence(list(raw_seq), n_steps)
```

Όπως ήδη αναφέρθηκε, και τα δύο dataset χωρίστηκαν σε ποσοστό 80-20. Το μοντέλο εκπαιδεύτηκε στο πρώτο 80% του dataset και οι προβλέψεις αφορούν το τελευταίο 20%.

```
# define split percentage
split_prc = 0.8
```

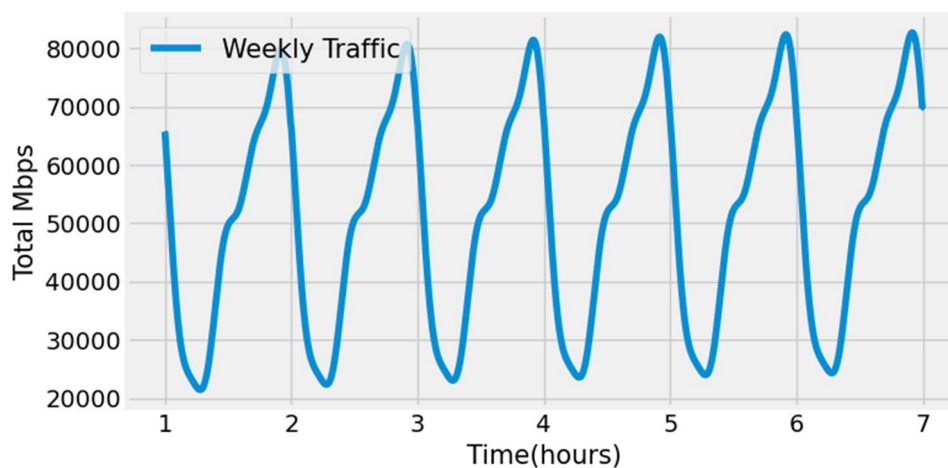
```
n_train_time = int(len(X) * split_prc)
# split into input and outputs
train_X, train_y = X[:n_train_time], y[:n_train_time]
test_X, test_y = X[n_train_time:], y[n_train_time:]
# reshape input to be 3D [samples, timesteps, features]
train_X = train_X.reshape((train_X.shape[0], n_steps, n_features))
test_X = test_X.reshape((test_X.shape[0], n_steps, n_features))
print(train_X.shape, train_y.shape, test_X.shape, test_y.shape)
#The input is reshaped into the 3D format as expected by LSTMs, namely
[samples, timesteps, features].
```

Για την εκπαίδευση του μοντέλου χρησιμοποιήθηκαν 10 epochs, δηλαδή 10 επαναλήψεις.

```
n_epochs = 10
model.fit(train_X, train_y, epochs=n_epochs)
```

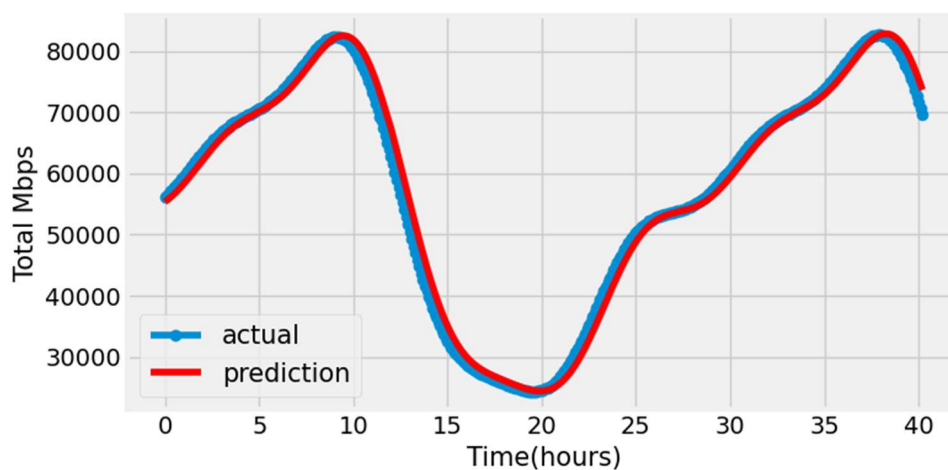

Συνθετικά δεδομένα

Η γραφική παράσταση μετά τη δειγματοληψία παρουσιάζεται στην παρακάτω γραφική παράσταση:



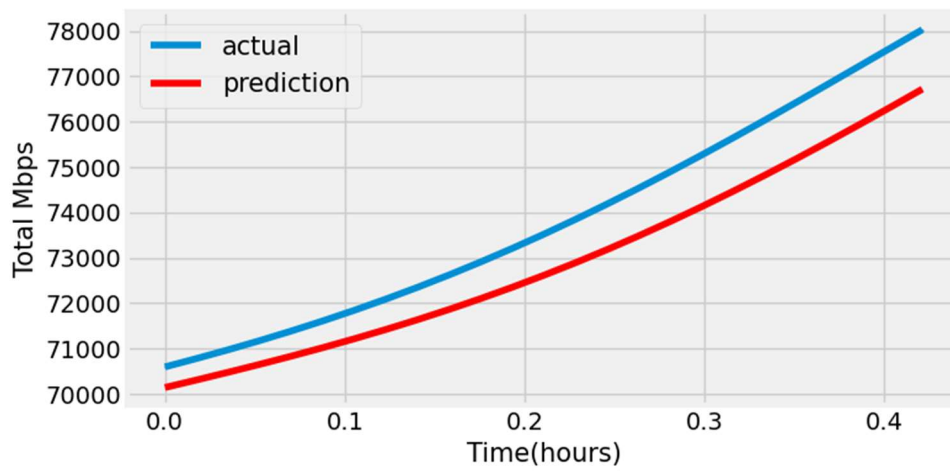
Εικόνα 60: Γραφική παράσταση κίνησης δεδομένων διάρκειας 1 εβδομάδας μετά τη δειγματοληψία

Έχοντας ρυθμίσει το μοντέλο με τον τρόπο που περιεγράφηκε, προκύπτουν οι προβλέψεις:



Εικόνα 61: Πρόβλεψη κίνησης για δεδομένα εισόδου μίας εβδομάδας

Οι προβλέψεις παρουσιάζονται αναλυτικότερα σε μικρότερο τμήμα της κίνησης:



Εικόνα 62: Πρόβλεψη κίνησης

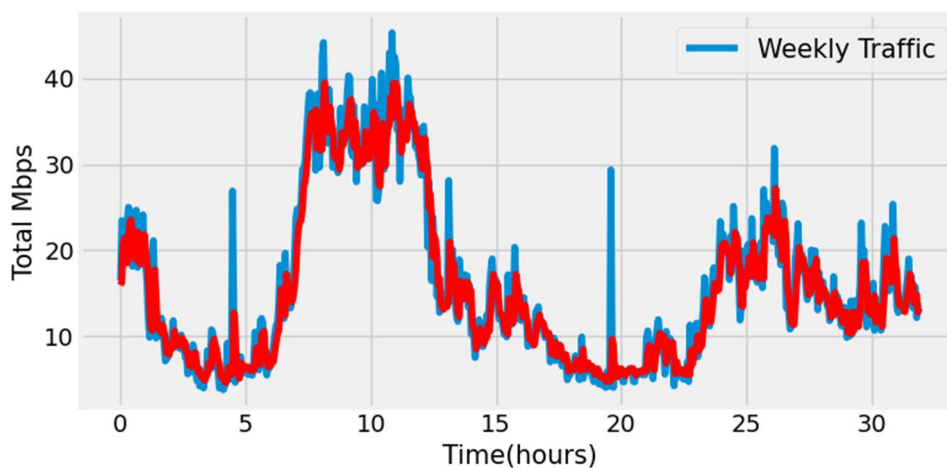
Για την αξιολόγηση των μοντέλων εφαρμόστηκαν οι συναρτήσεις των μετρικών. Τα αποτελέσματα των μετρικών παρουσιάζονται παρακάτω:

Test RMSE: 1804.484

Test RRSE: 0.096

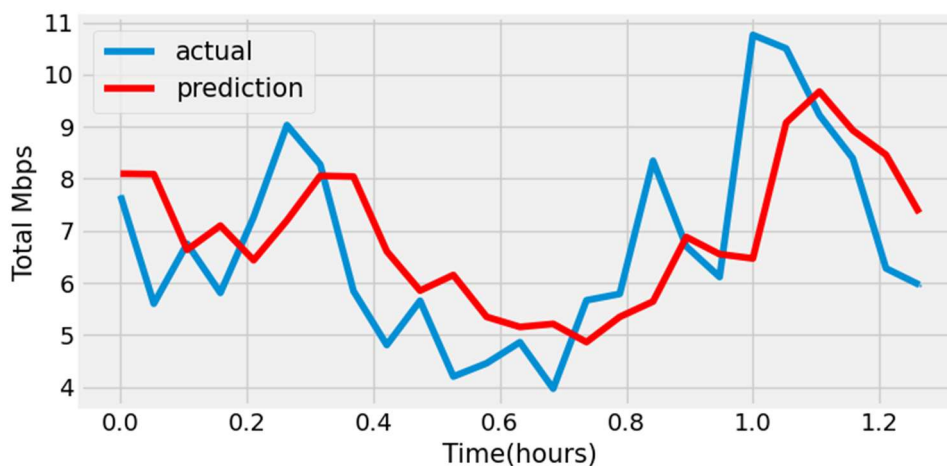
Πραγματικά δεδομένα

Ακολουθώντας την ίδια αρχιτεκτονική και διαδικασία ρύθμισης του μοντέλου, προκύπτουν οι προβλέψεις:



Εικόνα 63: Πρόβλεψη κίνησης για δεδομένα εισόδου 11 ημερών

Στην παρακάτω γραφική παράσταση παρουσιάζονται οι προβλέψεις για μικρότερο τμήμα της κίνησης:



Εικόνα 64: Πρόβλεψη κίνησης

Για την αξιολόγηση των μοντέλων εφαρμόστηκαν οι συναρτήσεις των μετρικών. Τα αποτελέσματα των μετρικών παρουσιάζονται παρακάτω:

Test RMSE: 3.162
Test RRSE: 0.327

5.7 Αποτελέσματα

Στον παρακάτω πίνακα παρουσιάζονται συγκεντρωτικά τα αποτελέσματα των προσομοιώσεων:

Data type	Model	RMSE	RRSE
Synthetic Data	CNN	3430.134	0.107
Real Data	CNN	4.370	0.460
Synthetic Data	SARIMA	1622.190	0.085
Real Data	SARIMA	3.382	0.307
Synthetic Data	Vanilla LSTM	1804.484	0.096
Real Data	Vanilla LSTM	3.162	0.327

Όπως γίνεται φανερό, τα μοντέλα μηχανικής μάθησης πραγματοποίησαν ιδιαίτερα καλές προβλέψεις. Στόχος μας κατά την προσομοίωση ήταν τα μοντέλα που χρησιμοποίησαν συνθετικά dataset να επιτύχουν τιμές για τη μετρική RRSE χαμηλότερες της τιμής 0.1. Όντως, ο στόχος αυτός επετεύχθη, καθώς για το μοντέλο CNN η τιμή ήταν 0.107, για το μοντέλο SARIMA ήταν 0.085 και για το μοντέλο Vanilla LSTM ήταν 0.096. Η επίδοση αυτή σημαίνει πως τα μοντέλα αυτά ήταν περισσότερο από 90% καλύτερα από ένα βασικό μοντέλο πρόβλεψης. Η μετρική RMSE χρησιμοποιήθηκε γιατί αποτελεί την πιο κοινά χρησιμοποιούμενη μετρική, συνεπώς μπορεί να χρησιμοποιηθεί για σύγκριση με άλλες μελέτες και αρχιτεκτονικές. Όπως έχει ήδη αναλυθεί, η μετρική RMSE συμπεριλαμβάνει στο αποτέλεσμα της και το ύψος των μετρήσεων, συνεπώς και για το λόγο αυτό εμφανίζει υψηλότερες τιμές από την μετρική RRSE. Παρόλα αυτά, παρατηρείται αναλογία στις επιδόσεις τους.

Το γεγονός πως τα μοντέλα που χρησιμοποίησαν dataset βασισμένα σε πραγματικά δεδομένα είχαν υψηλότερες τιμές, δηλαδή είχαν λιγότερο καλές προβλέψεις, οφείλεται ακριβώς στο γεγονός πως τα δεδομένα στηρίζονταν σε πραγματικές μετρήσεις, οι οποίες περιλάμβαναν αρκετά μεγαλύτερο θόρυβο καθώς και αρκετές ακραίες τιμές (outliers).

Επίσης, μέσω των δύο διαφορετικών dataset, επιχειρήθηκε αρχικά η δημιουργία και εκπαίδευση μοντέλων σε συνθετικά δεδομένα με στόχο την κατά τον δυνατόν καλύτερη επιλογή αρχιτεκτονικής, ρύθμισης και κατ' επέκταση απόδοσης των μοντέλων.

Τέλος, αφότου εξασφαλίστηκε η υψηλή επίδοση των μοντέλων, χρησιμοποιήθηκαν οι ίδιες αρχιτεκτονικές σε πραγματικά dataset αποδίδοντας επίσης αποτελεσματικά με τιμές RRSE χαμηλότερη από 0.5 για το μοντέλο CNN, και οριακά πάνω από 0.3 για τα μοντέλα SARIMA και LSTM.

Επίλογος

Στόχος της παρούσας διπλωματικής εργασίας ήταν η μελέτη των εφαρμογών της Μηχανικής Μάθησης για την πρόβλεψη κίνησης στα Οπτικά Δίκτυα. Επομένως, στα πλαίσια αυτά λοιπόν, μελετήθηκε αρχικά ο κλάδος των οπτικών δικτύων καθώς παρουσιάστηκε η εξέλιξη τους, η αρχιτεκτονική τους καθώς έγινε και περιγραφή της τεχνολογίας WDM στην οποία βασίζεται η μετάδοση στο οπτικό επίπεδο. Εν συνεχεία πραγματοποιήθηκε εκτενή αναφορά και ανάλυση στον κλάδο της Μηχανικής Μάθησης, δίνοντας ιδιαίτερη βαρύτητα στον τομέα της μάθησης με επίβλεψη και των Νευρωνικών Δικτύων. Δόθηκε ιδιαίτερη έμφαση στα κίνητρα και τις προκλήσεις εφαρμογής τεχνικών μηχανικής μάθησης στα οπτικά δίκτυα και παρουσιάστηκαν τομείς στους οποίους συναντάμε τέτοιες εφαρμογές. Τέλος παρουσιάστηκε το πρόβλημα της πρόβλεψης της κίνησης και πραγματοποιήθηκαν προσομοιώσεις με σκοπό την μελέτη του προβλήματος αυτού.

Η εφαρμογή τεχνικών μηχανικής μάθησης για την πρόβλεψη κίνησης έρχεται να δώσει λύση στον αντίκτυπο που προκαλεί η ολοένα αυξανόμενη κίνηση στο διαδίκτυο. Με την εκ των προτέρων γνώση της συμπεριφοράς των καταναλωτών κατ' επέκταση της συμπεριφοράς του δικτύου, δίνεται η δυνατότητα βέλτιστης διαχείρισης των πόρων του δικτύου και κατ' επέκταση την παροχή της αντίστοιχης χωρητικότητας που απαιτείται. Τα αποτελέσματα των προβλέψεων υπόσχονται ευελιξία ακόμα και σε εφαρμογές όχι μόνο στο επίπεδο δικτύου, αλλά και στο οπτικό επίπεδο, όπως για παράδειγμα στην βέλτιστη προσαρμογή των παραμέτρων των πομποδεκτών.

Όπως φάνηκε και από τα αποτελέσματα των προσομοιώσεων που πραγματοποιήθηκαν για τις ανάγκες της παρούσας εργασίας, κατέστη εφικτή η πρόβλεψη της κίνησης για συγκεκριμένα κομμάτια του δικτύου, δοθέντος δεδομένων είτε πραγματικών είτε συνθετικών. Με τα δίκτυα τηλεφωνίας 5^{ης} γενιάς να καταφθάνουν, η πρόβλεψη της κίνησης υπόσχεται ευελιξία και προσαρμογή στις αυξημένες και συνεχώς μεταβαλλόμενες ανάγκες.

Κατά τον Peter Drucker *«ο καλύτερος τρόπος για να προβλέψεις το μέλλον είναι να το δημιουργήσεις»*. Ίσως η Μηχανική Μάθηση να διαψεύδει τον εν λόγο ισχυρισμό καθώς, όπως φαίνεται, είναι σε θέση να προβλέπει το μέλλον και αντί να το δημιουργεί, να προσαρμόζεται σε αυτό..

Βιβλιογραφία

- [1] statista, «statista,» [Ηλεκτρονικό]. Available: <https://www.statista.com/topics/1145/internet-usage-worldwide/>.
- [2] Cisco Networks, «Cisco,» 2020. [Ηλεκτρονικό]. Available: https://www.cisco.com/c/en/us/td/docs/optical/15000r8_0/dwdm/reference/guide/d80refrn/d80txmxp.html.
- [3] P. T. M. O. A. P. G.I. Papadimitriou, ΟΠΤΙΚΑ ΔΙΚΤΥΑ ΤΕΧΝΟΛΟΓΙΑΣ WDM, ΤΟΠΙΚΑ ΚΑΙ ΜΗΤΡΟΠΟΛΙΤΙΚΑ ΔΙΚΤΥΑ, Κλειδάριθμος, 2005.
- [4] Γ. Μπαλαμώτης, Πολυπλεξία Μήκους Κύματος με Ολοκληρωμένους Οπτικούς Μικροδακτυλίους Πυριτίου, Θεσσαλονίκη: Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Σχολή Θετικών Επιστημών, Τμήμα Πληροφορικής, 2014.
- [5] [Ηλεκτρονικό]. Available: <http://www.fundinguniverse.com/company-histories/ciena-corporation-history/>.
- [6] K. S. G. S. R. Ramaswami, Optical Networks: A Practical Perspective, 3rd Edition, Morgan Kaufmann, 2009.
- [7] S. Pearlstein, «The Puzzling Allure of David Huber, The Washington,» *Washington Post*, 7 January 2005.
- [8] P. Winzer, «Optical Networking Beyond WDM,» *IEEE Photonics Journal*, Vol 4, Issue 2, pp. 647-651, 2012, 2012.
- [9] Γ. Π. Χ. Βασιλόπουλος, Συστήματα Πολυπλεξίας Μήκους Κύματος (WDM).
- [10] B. Mukherjee, Optical WDM Networks, Springer, 2006.
- [11] M. Academy, Introduction to Dense Wavelength Division Multiplexing.
- [12] C. S. Inc., Introduction to DWDM for Metropolitan Networks.
- [13] G. Agrawal, Fiber-Optic Communication Systems, 4th ed επιμ., Wiley, 2010.
- [14] F. C. G. Alliance, FTTH Council – Definition of Terms, Version 4.0 επιμ., FTTH Council, February 2015.
- [15] F. C. Europe, «Point-to-Point FTTH and Ethernet Access,» *Architecture and technology Committee*, February 2004.
- [16] G. P. G. Kramer, «Ethernet Passive Optical Network (EPON): Building a Next- Generation Optical Access Network,» *IEEE Communications Magazine*, February 2002.
- [17] A. Burkov, The Hundred-Page Machine Learning Book, 2019.
- [18] T. M. Mitchel, Machine Learning, McGraw-Hill Science/Engineering/Math, 1997.
- [19] «colah's blog,» 27 August 2015. [Ηλεκτρονικό]. Available: <https://colah.github.io/posts/2015-08-Understanding->

LSTMs/?fbclid=IwAR00FbdLGK5qdD7XAu83M1o8RGIfX8K0HsKQGcFnR71ygaMw0rNeZ5HNISg.

- [20] [Ηλεκτρονικό]. Available: <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>.
- [21] J. S. Sepp Hochreiter, «LONG SHORT-TERM MEMORY,» *Neural Computation* 9(8):1735{1780, 1997.
- [22] «BOOSTEDML,» 14 April 2020. [Ηλεκτρονικό]. Available: <https://boostedml.com/2020/04/1-d-convolutional-neural-networks-for-time-series-basic-intuition.html>.
- [23] J. P. a. M. K. J. Han, *Data mining: concepts and techniques*, Elsevier, 2011.
- [24] B. S. a. A. Z. O. Chapelle, *Semi-supervised learning*, MIT Press, 2006.
- [25] R. S. S. a. A. G. Barto, *Reinforcement learning: An introduction.*, Cambridge: MIT press, 1998.
- [26] M. I. C. R. M. I. A. N. M. I. I. M. D. Z. M. I. M. R. S. M. I. a. M. T. S. M. I. Francesco Musumeci, «An Overview on Application of Machine Learning Techniques in Optical Networks,» 1 December 2018.
- [27] C. DeCusatis, «Optical interconnect networks for data communications,» *Journal of Lightwave Technology*, vol. 32, no. 4, p. pp. 544–552.
- [28] B.-W. K. a. B. M. H. Song, «Long-reach optical access networks: A survey of research challenges, demonstrations, and bandwidth assignment mechanisms,» *IEEE communications surveys & tutorials*, vol. 12, no. 1., 2010.
- [29] K. Zhu and B. Mukherjee, «Traffic grooming in an optical WDM mesh network,» *IEEE Journal on Selected Areas in Communications*, vol. 20, no 1, p. pp. 122–133, Jan 2002.
- [30] S. R. a. B. Mukherjee, «Survivable WDM mesh networks. Part I-Protection,» *Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM) 1999*, vol.2, p. pp. 744–751, Mar. 1999.
- [31] [Ηλεκτρονικό]. Available: <https://www.lightreading.com/artificial-intelligence-machine-learning/csp-embbrace-machine-learning-and-ai/a/d-id/737973>.
- [32] [Ηλεκτρονικό]. Available: <http://www-file.huawei.com/-/media/CORPORATE/PDF/white%20paper/White-Paper-on-Technological-Developments-of-Optical-Networks.pdf>.
- [33] M. J. A. L. a. S. B. Y. O. Gerstel, «Elastic optical networking: A new dawn for the optical layer?,» *IEEE Communications Magazine*, vol. 50, no. 2, February 2012.
- [34] N. S. a. E. O. B. C. Chatterjee, «Routing and spectrum allocation in elastic optical networks: A tutorial,» *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3., p. pp. 1776–1800, 2015.
- [35] F. A. I. K. M. K. R. S. a. G. N. R. S. Talebi, «management techniques for elastic optical networks: A survey,» *Optical Switching and Networking*, vol. 13, p. pp. 34–48, 2014.
- [36] J. T. M. P. R. J. a. D. Z. J. Wass, «Gaussian Process Regression for WDM System Performance Prediction,» *Optical Fiber Communication Conference (OFC)*, March 2017.
- [37] J. Brownlee, «Machine Learning Mastery,» 2 Decemper 2016. [Ηλεκτρονικό]. Available: <https://machinelearningmastery.com/time-series-forecasting/>.

- [38] A. Rajbhoj, «towards data science,» 26 September 2019. [Ηλεκτρονικό]. Available: <https://towardsdatascience.com/arima-simplified-b63315f27ebc>.
- [39] K. Foo, «medium.com,» 3 January 2018. [Ηλεκτρονικό]. Available: <https://medium.com/@kfoofw/seasonal-lags-sarima-model-fa671a858729>.
- [40] «python.org,» [Ηλεκτρονικό]. Available: <https://www.python.org/>.
- [41] «stackoverflow,» [Ηλεκτρονικό]. Available: <https://insights.stackoverflow.com/survey/2016>.
- [42] J. Brownlee, *Intriduction to Time Series Forecasting with Python, Machine Learning Mastery*, 2019.
- [43] K. L. J. H. J. B. a. J. P. Kamath, «Generation of high bandwidth network traffic traces,» *Proc. IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer Systems*, 2002.
- [44] A. V. a. N. Larrieu, «How to generate realistic network traffic,» σε *IEEE 38th Annual International Computers, Software & Applications Conference (COMPSAC)*, 2014..
- [45] D. W. a. J. W. A. Burks, «An analysis of a logical machine using parenthesis-free notation,» *Mathematical Tables and Other Aids to Computation*, vol. 8, pp. pp. 53-57, 1954.
- [46] A. V. F. M. M. R. a. L. V. Alba P. Vela, «Traffic Generation for Telecom Cloud-Based Simulation,» 2016. [Ηλεκτρονικό]. Available: http://personals.ac.upc.edu/lvelasco/docs/research/2016_ICTON-6.pdf.
- [47] L. V. M. R. M. K. J. P. F.-P. a. D. C. A. Castro, «Dynamic routing and spectrum (re)allocation in future flexgrid optical networks, vol. 56,» *Computer Networks*, pp. pp. 2869-2883, 2012.
- [48] «WAND Network Research Group,» [Ηλεκτρονικό]. Available: <https://wand.net.nz/wits/index.php>.