



ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΣΧΟΛΕΣ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ
ΠΜΣ «ΕΦΑΡΜΟΣΜΕΝΑ ΜΑΘΗΜΑΤΙΚΑ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗ»

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

**ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ
ΦΥΣΙΚΗΣ ΓΛΩΣΣΑΣ**

Σικαλόπουλος Δημήτριος

Ιωάννινα, 2019

Η παρούσα Μεταπτυχιακή Διατριβή εκπονήθηκε στο πλαίσιο των Σπουδών για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στα Εφαρμοσμένα Μαθηματικά και Πληροφορική που απονέμει το Τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων.

Εγκρίθηκε την 27/6/2019 από την Εξεταστική Επιτροπή:

Όνοματεπώνυμο **Βαθμίδα**

Επιβλέπων:

Σωκράτης Μπαλτζής Λέκτορας

τα Μέλη ΔΕΠ:

Νικόλαος Γλυνός Επίκουρος Καθηγητής

Ιωάννης Σταματίου Καθηγητής

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ

«Δηλώνω υπεύθυνα ότι η παρούσα διατριβή εκπονήθηκε κάτω από τους διεθνείς ηθικούς και ακαδημαϊκούς κανόνες δεοντολογίας και προστασίας της πνευματικής ιδιοκτησίας. Σύμφωνα με τους κανόνες αυτούς, δεν έχω προβεί σε ιδιοποίηση ξένου επιστημονικού έργου και έχω πλήρως αναφέρει τις πηγές που χρησιμοποίησα στην εργασία αυτή»

Δημήτριος Σικαλόπουλος

Ευχαριστίες

Ολοκληρώνοντας τον μεταπτυχιακό κύκλο σπουδών και της διατριβής μου επιθυμώ να εκφράσω τις ευχαριστίες μου σε όλους όσους με βοήθησαν να ολοκληρώσω το έργο μου αυτό.

Αρχικά, θα ήθελα να ευχαριστήσω το Τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων. από το οποίο απεφοίτησα, που μου έδωσε την ευκαιρία να συμμετάσχω και να ολοκληρώσω τις μεταπτυχιακές σπουδές μου στο αντικείμενο που επέλεξα.

Ευχαριστώ τον κ. Μπαλτζή Σωκράτη, Λέκτορα του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων, για την ανάθεση της παρούσας μεταπτυχιακής διατριβής, την επίβλεψη και επιστημονική του καθοδήγηση, τη στήριξη του, καθώς και για την ευκαιρία που μου έδωσε προκειμένου να αποκτήσω γνώσεις, εμπειρίες και εξοικείωση σε ένα πολύ ενδιαφέρον και σημαντικό αντικείμενο όπως το παρών τόσο κατά την εκπόνηση της εργασίας αυτής, όσο και καθ' όλη τη διάρκεια του μεταπτυχιακού κύκλου σπουδών.

Επίσης, ευχαριστώ θερμά τον κ. Γλυνό Νικόλαο, Επίκουρο Καθηγητή του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων και τον κ. Σταματίου Ιωάννη, Καθηγητή του Τμήματος Διοίκησης Επιχειρήσεων του Πανεπιστημίου Πατρών, μέλη της Τριμελούς Επιτροπής για την αξιολόγηση και κρίση της Διατριβής, για την πρόθυμη και σημαντική βοήθεια τους αλλά και για τον πολύτιμο χρόνο που αφιέρωσαν για την αξιολόγηση της παρούσας διατριβής.

Τέλος, θα ήθελα να ευχαριστήσω από τα βάθη της καρδιάς μου την οικογένεια μου καθώς και τους στενούς μου ανθρώπους που με παρότρυναν αλλά και με στήριξαν ώστε να ακολουθήσω και να πραγματοποιήσω τα όνειρα μου.

Ιωάννινα, Ιούνιος 2019.

Δημήτριος Σικαλόπουλος

ΠΕΡΙΛΗΨΗ

Τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) είναι ένα υπολογιστικό μοντέλο η έμπνευση του οποίου προήλθε από τα Βιολογικά Νευρωνικά Δίκτυα. Τα ΤΝΔ έχουν δώσει αισιόδοξα μηνύματα τα τελευταία χρόνια για την αποτελεσματικότητα στην έρευνα και στις αξιόπιστες εφαρμογές τους στην Αναγνώριση Ομιλίας, στην Υπολογιστική Όραση, στην Επεξεργασία Κειμένου κ.ά.

Σκοπός της διατριβής αυτής είναι η παρουσίαση της συμβολής των ΤΝΔ στην εξέλιξη και αξιοποίησή τους στον τομέα της Επεξεργασίας Φυσικής Γλώσσας (ΕΦΓ).

Στο Κεφάλαιο 1 παρουσιάζονται ο σημαντικός ρόλος των μαθηματικών μοντέλων των Γραμματικών στην ΕΦΓ και οι σημαντικότεροι σταθμοί της εξέλιξης των Γραμματικών που οδήγησαν στις περαιτέρω μεθόδους και τεχνικές. Αναπτύσσονται τα θεμελιώδη εργαλεία των ΤΝΔ τα οποία είναι πρωτοποριακοί αλγόριθμοι, σε αντίθεση με τους μέχρι τώρα παραδοσιακούς αλγορίθμους της βιβλιογραφίας. Οι αλγόριθμοι αυτοί ορίζονται αφενός ως Μέθοδοι Εκμάθησης Μηχανής, δηλαδή αλγόριθμοι εκμάθησης της μηχανής των συνθηκών για την αντιμετώπιση των προδιαγραφών/δεδομένων συγκεκριμένου προβλήματος και αφετέρου ως Μέθοδοι Εκμάθησης Βάθους, δηλαδή αλγόριθμοι εκπαίδευσης της μηχανής για την αντιμετώπιση παρόμοιων προδιαγραφών προβλημάτων· επίσης, παραθέτονται παραδείγματα και συγκρίσεις αυτών με τους παραδοσιακούς αλγόριθμους.

Το Κεφάλαιο 2 συμπληρώνεται με την παρουσίαση των Μεθόδων των ΤΝΔ, το οποίο πλαισιώνεται από ορισμούς, ταξινομήσεις ΝΔ, και παραδείγματα εφαρμογών τους.

Στο Κεφάλαιο 3 αναφέρονται οι πολύ πρόσφατες μέθοδοι και παραδείγματα που στοιχειοθετούν τις σύγχρονες εξελίξεις και εφαρμογές των ΤΝΔ στην ΕΦΓ.

Στο Κεφάλαιο 4 παραθέτονται οι προτάσεις μας για περαιτέρω έρευνα που αφορά στα ΤΝΔ και την ΕΦΓ. Συγκεκριμένα στην υλοποίηση της Μεθόδου Επεξεργασίας της Νέας Ελληνικής Γλώσσας που προτείνεται από: τις Μητροειδείς και τις Σπονδυλωτές Μητροειδείς Γραμματικές – Χαρακτηριστικοί Εκθέτες, το Βασικό Νεοελληνικό Υπολογιστικό Πολυλεξικό (ΒΝΠ) και των αλγορίθμων του.

Τέλος, στα Παρατήματα περιλαμβάνονται 2 Γλωσσάρια ελληνοαγγλικών και αγγλοελληνικών όρων της διατριβής αντίστοιχα.

Στην αρχή κάθε Κεφαλαίου παρατίθενται λέξεις κλειδιά που αναφέρονται στο Κεφάλαιο και στο τέλος κάθε Κεφαλαίου η αντίστοιχη βιβλιογραφία του.

Abstract

Artificial Neural Networks (ANN) is a computational model inspired by Biological Neural Networks. ANNs have provided optimistic messages over recent years about research effectiveness and their reliable applications in Speech Recognition, Computer Vision, Text Processing, and so on.

The purpose of this dissertation is to present the contribution of ANNs to their development and exploitation in the field of Natural Language Processing (NLP).

Chapter 1 presents the important role of Grammar mathematical models in NLP and the most important stages in the evolution of Grammars that led to further methods and techniques. We develop the fundamental tools of ANNs, which are pioneering algorithms, as opposed to the traditional algorithms of the literature so far. These algorithms are defined on the one hand as Machine Learning Methods, that is, machine learning algorithms of the condition engine for dealing with specifications/data of a particular problem, and on the other hand as Deep Learning Methods, i.e. Machine Learning algorithms to deal with similar problem specifications. In addition, examples and comparisons of these with traditional algorithms are given.

Chapter 2 is complemented by the presentation of the ANNs Methods, which is framed by definitions, Neural Network classifications, and examples of their applications.

Chapter 3 lists the very recent methods and examples of the modern developments and applications of ANNs in NLP.

Chapter 4 lists our suggestions for further research on ANNs and NLP. Specifically in the implementation of the Method of Modern Greek Language Processing proposed by: Template Grammars and Characteristic Exponents, Basic Modern Greek Language Computational Polylexicon and its algorithms.

Finally, in the Appendices are included two glossaries of a Greek-English and an English-Greek basic dissertation terms correspondingly.

At the beginning of each Chapter, keywords referred to the Chapter are listed and at the end of each Chapter, the corresponding bibliography is listed as well.

ΠΕΡΙΕΧΟΜΕΝΑ

Περίληψη	7
Abstract	9
1 Προαπαιτούμενες Έννοιες και Ορισμοί	
1.1 Εισαγωγή	15
1.2 Γραμματικές για την Επεξεργασία Φυσικής Γλώσσας	19
1.2.1 Εισαγωγή	19
Λέξεις Κλειδιά	
1.2.2 Γενετικές και Μετασχηματιστικές Γραμματικές της Ιεραρχίας Chomsky, Επαυξημένα Δίκτυα Μετάβασης και Γραμματικές των Συνδεδεμένων Δέντρων	20
1.2.3 Γραμματικές Ενοποίησης - Λεξικές Λειτουργικές Γραμματικές	25
1.2.4 Ετικετοποίηση	27
1.2.5 Στοχαστικές Γραμματικές	28
1.2.6 Μητροειδείς και Σπονδυλωτές Μητροειδείς Γραμματικές - Χαρακτηριστικοί Εκθέτες και το Βασικό Νεοελληνικό Πολυλεξικό	29
1.2.7 Βιβλιογραφικές Αναφορές	34
1.3 Εκμάθηση Μηχανής	
1.3.1 Εισαγωγή	39
Λέξεις Κλειδιά	
1.3.2 Βασικές Αρχές, Έννοιες και Ορισμοί	39
1.3.3 Η λειτουργία, ο ρόλος της Εκμάθησης Μηχανής και οι διαφορές της από τον Συμβατικό Προγραμματισμό	40
1.3.4 Ταξινόμηση των Αλγορίθμων Εκμάθησης Μηχανής, εφαρμογές και παραδείγματα	42
1.3.5 Παράδειγμα Σύγκρισης Συμβατικού Προγραμματισμού και Εκμάθησης Μηχανής	45
1.3.5.1 1 ^η προσέγγιση με κώδικα Συμβατικού Προγραμματισμού	46
1.3.5.2 2 ^η προσέγγιση με κώδικα Εκμάθησης Μηχανής	46
1.3.5.3 Συμπεράσματα	50
1.3.6 Εφαρμογές της Εκμάθησης Μηχανής	51

1.3.7 Βιβλιογραφικές Αναφορές	53
1.4 Εκμάθηση Βάθους	
1.4.1 Εισαγωγή	55
Λέξεις Κλειδιά	
1.4.2 Βασικές Αρχές, Έννοιες και Ορισμοί	55
1.4.3 Λειτουργία της Εκμάθησης Βάθους	57
1.4.4 Διαφορά Εκμάθησης Βάθους και Εκμάθησης Μηχανής	59
1.4.5 Εφαρμογές της Εκμάθησης Βάθους	60
1.4.6 Επιτάχυνση μοντέλων Εκμάθησης Βάθους μέσω Μονάδας Επεξεργασίας Γραφικών	64
1.4.7 Παράδειγμα Αλγορίθμου Εκμάθησης Βάθους για την Επεξεργασία Φυσικής Γλώσσας και πλεονεκτήματά του	65
1.4.8 Βιβλιογραφικές Αναφορές	68
2 Τεχνητά Νευρωνικά Δίκτυα	
2.1 Εισαγωγή	69
Λέξεις Κλειδιά	
2.2 Ιστορική Αναδρομή	70
2.3 Βασικές Έννοιες και Ορισμοί των Τεχνητών Νευρωνικών Δικτύων	73
2.3.1 Βιολογικά Νευρωνικά Δίκτυα και Τεχνητά Νευρωνικά Δίκτυα	73
2.3.1.1 Βιολογικός Νευρώνας	73
2.3.1.2 Αναλογία Βιολογικών Νευρωνικών Δικτύων και Τεχνητών Νευρωνικών Δικτύων	74
2.3.2 Το μοντέλο του Τεχνητού Νευρώνα και η Συνάρτηση ενεργοποίησης	74
2.3.3 Αρχιτεκτονικές και Πλεονεκτήματα των Τεχνητών Νευρωνικών Δικτύων	78
2.3.3.1 Δίκτυα Πρόσθιας Τροφοδότησης	79
2.3.3.1.1 Αισθητήρας ενός επιπέδου	79
2.3.3.1.2 Αισθητήρας Πολλών Επιπέδων	80
2.3.3.1.3 Δίκτυα Ακτινικών Συναρτήσεων	81
2.3.3.1.4 Συνελκτικά Νευρωνικά Δίκτυα	82
2.3.3.2 Επαναληπτικά Νευρωνικά Δίκτυα	85
2.3.3.2.1 Αναδρομικά Νευρωνικά Δίκτυα	86

2.3.3.2.2	Δίκτυα Kohonen	87
2.3.3.2.3	Δίκτυα Hopfield	87
2.3.3.3	Πλεονεκτήματα των Τεχνητών Νευρωνικών Δικτύων	88
2.3.4	Διαδικασίες Εκμάθησης των Τεχνητών Νευρωνικών Δικτύων	90
2.3.4.1	Εισαγωγή	90
2.3.4.2	Επιβλεπόμενη Εκμάθηση	90
2.3.4.3	Μη Επιβλεπόμενη Εκμάθηση	91
2.3.5	Εκπαίδευση των Τεχνητών Νευρωνικών Δικτύων	92
2.3.5.1	Εκπαίδευση Τεχνητών Νευρωνικών Δικτύων με χρήση του αλγόριθμου BackPropagation	92
2.3.5.2	Μια εναλλακτική περιγραφή της εκπαίδευσης των Τεχνητών Νευρωνικών Δικτύων	95
2.3.6	Βασικές διαφορές της λειτουργίας των Τεχνητών Νευρωνικών Δικτύων, του Παραδοσιακού Υπολογισμού και των Ειδικών Συστημάτων	96
2.4	Συμπεράσματα	98
2.5	Βιβλιογραφικές Αναφορές	100

3 Νευρωνικά Δίκτυα και Επεξεργασία Φυσικής Γλώσσας

3.1	Εισαγωγή	103
	Λέξεις κλειδιά	
3.2	Μέθοδοι των Τυπικών Νευρωνικών Δικτύων για την Επεξεργασία Φυσικής Γλώσσας	104
3.2.1	Εισαγωγή	104
3.2.2	Μέθοδος του Αποδεκτού Συνδυασμού Λέξεων με Νευρωνικά Δίκτυα Word2Vec	105
3.2.2.1	Νευρώνες για τον (αποδεκτό) συνδυασμό λέξεων	107
3.2.2.2	Παραδείγματα Κατανόησης Word2Vec	110
3.2.2.3	N-grams και Skip-grams (βήμα - βήμα αναζήτηση)	112
3.2.2.4	Βελτιώσεις στην Επεξεργασία Φυσικής Γλώσσας: ELMO, BERT και GPT-2	112
3.2.2.5	Ειδικές Περιπτώσεις του Word2Vec	113
3.2.2.6	Η ευρεσιτεχνία της Google για το Word2Vec	113
3.2.2.7	Χρήση του Word2Vec στις ξένες γλώσσες	113

3.2.2.8 Ακολουθιακά Διανύσματα	114
3.2.2.9 Χαρακτηριστικά του Word2Vec στο deeplearning4j	114
3.3 Άλλες Μέθοδοι Νευρωνικών Δικτύων για την Επεξεργασία Φυσικής Γλώσσας	115
3.3.1 Τα Συνελκτικά Νευρωνικά Δίκτυα για την Επεξεργασία Φυσικής Γλώσσας	115
3.3.2 Τα Επαναληπτικά και Αναδρομικά Νευρωνικά Δίκτυα για την Επεξεργασία Φυσικής Γλώσσας	116
3.4 Συμπεράσματα	118
3.5 Βιβλιογραφικές Αναφορές	120
4 Συμπεράσματα και Προτάσεις για περαιτέρω έρευνα	123
5 Παραρτήματα	
Παράρτημα 1: Γλωσσάρι Ελληνοαγγλικών όρων της Διατριβής	127
Παράρτημα 2: Γλωσσάρι Αγγλοελληνικών όρων της Διατριβής	137

ΚΕΦΑΛΑΙΟ 1

Προσπαιτούμενες Έννοιες και Ορισμοί

1.1 Εισαγωγή

Η Επεξεργασία Φυσικής Γλώσσας¹ (ΕΦΓ) είναι η δυνατότητα αυτόματης ή με τη βοήθεια του Η/Υ επεξεργασία της Φυσικής Γλώσσας (ΦΓ), δηλαδή:

1. Αναγνώριση των συστατικών της πρότασης της ΦΓ ή του εντοπισμού,
 - της ενέργειας ή του ρήματος της πρότασης,
 - του δράστη της ενέργειας ή του υποκειμένου της πρότασης,
 - του αποδέκτη της ενέργειας ή του αντικειμένου της πρότασης και
 - όλων των υπολοίπων όρων της πρότασης που έχουν σχέση με τις παραπάνω οντότητες, για περαιτέρω αξιοποίηση καθώς και
2. Παραγωγή αποδεκτών προτάσεων της ΦΓ.

Η ΕΦΓ αξιοποιείται στην:

- Μηχανική Μετάφραση
- Ενεργοποίηση Ρομπότ
- Εκπαίδευση
- Εξαγωγή Νοήματος
- Διαχείριση Εγγράφων κ.ά.

Η ΕΦΓ περιλαμβάνεται σε ποικίλα άλλα επικαλυπτόμενα πεδία στα διάφορα Τμήματα, όπως:

- Υπολογιστική Γλωσσολογία στη Γλωσσολογία,

¹ Είναι η γλώσσα που χρησιμοποιούν οι άνθρωποι για την καθημερινή επικοινωνία τους

- Επεξεργασία Φυσικής Γλώσσας στην Επιστήμη Η/Υ,
- Αναγνώριση Λόγου στους Ηλεκτρολόγους Μηχανικούς και τέλος
- Υπολογιστική Ψυχολογία στους Ψυχολόγους της Γνωσιακής Επιστήμης.

Η ΕΦΓ ως πεδίο της Τεχνητής Νοημοσύνης αναπτύχθηκε και αποτέλεσε μαζί με την Επεξεργασία του Λόγου ανεξάρτητο επιστημονικό πεδίο, αυτό της Γλωσσικής Τεχνολογίας.

Η Επεξεργασία του Λόγου (ΕΛ) είναι η δυνατότητα του Η/Υ για:

1. Αναγνώριση του Λόγου (Αναγνώριση Φωνής) και της Γραπτής Αποτύπωσης των λέξεων, των προτάσεων και των κειμένων μίας Φυσικής Γλώσσας
2. Παραγωγή του Λόγου από τη Γραπτή Αποτύπωσή του, όπως προαναφέρθηκε, των λέξεων, των προτάσεων και των κειμένων της υπό μελέτη ΦΓ.

Η Γλωσσική Τεχνολογία αποτελείται από διαδικασίες/μεθόδους αυτόματης επεξεργασίας που είναι διαδικασίες/μέθοδοι επίπονες και εξελισσόμενες.

Σημειώνεται ότι η ΕΦΓ είναι ένα πεδίο διεπιστημονικό που απαιτεί για την εκάστοτε ΦΓ:

1. τη μελέτη, την ταξινόμηση, την κωδικοποίηση και αρχειοθέτηση αφενός των γλωσσολογικών δεδομένων της αφετέρου των αλγοριθμοποιημένων δομολειτουργικών κανόνων της,
2. τον σχεδιασμό των μαθηματικών προτύπων (μοντέλων) αναγνώρισης και παραγωγής των συστατικών της και τέλος
3. την κωδικοποίηση και υλοποίηση των διαδικασιών αυτόματης επεξεργασίας της, όπως ορίστηκε παραπάνω.

Για την ΕΦΓ προτάθηκαν μέθοδοι και τεχνικές που στηρίχτηκαν, καταρχάς, σε μαθηματικά πρότυπα παραγωγής και τεχνολόγησης αλυσίδων χαρακτήρων, τις Γραμματικές. Στην προκειμένη περίπτωση Γραμματικές που αντιστοιχούσαν στην παραγωγή και τεχνολόγηση λέξεων, προτάσεων και κατ' επέκταση κειμένων της υπό επεξεργασία ΦΓ.

Η εξέλιξη των Μεθόδων ΕΦΓ, δηλαδή των Γραμματικών, οδήγησε σε έναν πυρήνα μεθόδων ΕΦΓ που κυρίαρχο ρόλο είχαν οι προσεγγίσεις γραμμικών μοντέλων για να ελέγχεται η Εκμάθηση και επικεντρωνόταν, για περισσότερο από μία δεκαετία, σε αλγορίθμους που στηρίζονταν:

1. σε πρότυπα αναγνώρισης και διάκρισης, τους Αισθητήρες,
2. σε πρότυπα μηχανών γραμμικής διανυσματικής υποστήριξης, υποστήριξης δηλαδή πολυδιάστατων δεδομένων,
3. στη λογιστική παλινδρόμηση και τέλος
4. στην εκπαίδευση σε διανύσματα πολύ αραιών δεδομένων αλλά συγχρόνως πολύ μεγάλων διαστάσεων.

Οι μέθοδοι αυτές ΕΦΓ συμπεριλήφθηκαν στους αλγορίθμους Εκμάθησης Μηχανής (MM), δηλαδή αλγορίθμων Εκμάθησης των προδιαγραφών αντιμετώπισης συγκεκριμένου προβλήματος, και κατ' επέκταση στους αλγορίθμους Εκμάθησης Βάθους (MB), δηλαδή αλγορίθμων εκπαίδευσης για την αντιμετώπιση εν γένει συγγενών - παρόμοιων προδιαγραφών προβλημάτων.

Το 2014 ξεκινούν κάποιες επιτυχημένες προσπάθειες ΕΦΓ, ενσωματώνοντας τους παραπάνω αλγορίθμους MM και MB που εντάσσονταν σε Μεθόδους TN και μετατρέπουν τα γραμμικά μοντέλα με αραιά δεδομένα μεγάλων διαστάσεων σε πρότυπους αλγορίθμους που περιέγραφαν τις δομολειτουργίες Τεχνητών Νευρωνικών Δικτύων (TNΔ) που μπορούσαν να χειριστούν πυκνά δεδομένα μεγάλων διαστάσεων. Κάποιες από τις μεθόδους των TNΔ απλά γενίκευαν τα γραμμικά μοντέλα και σχεδόν αντικαθιστούσαν τις γραμμικές διαδικασίες ταξινόμησης, δηλαδή τους γραμμικούς ταξινομητές. Άλλες πάλι, οι πιο καινοτόμες εισήγαγαν νέες νοοτροπίες προσέγγισης παρέχοντας νέες ευκαιρίες προτυποποίησης. Στις παρακάτω ενότητες θα αναλυθούν οι Γραμματικές, η MM, η MB και τα TNΔ γενικά αλλά και ειδικά για την ΕΦΓ. Επίσης σε κάποιες περιπτώσεις γίνεται σύγκριση των διεθνώς καθιερωμένων μεθόδων για την Επεξεργασία της ΑΓ με προταθείσες μεθόδους Επεξεργασίας της Νεοελληνικής (NEΓ). Σε άλλες πάλι περιπτώσεις προτείνεται ως θέμα για περαιτέρω έρευνα η εφαρμογή της προταθείσας μεθόδου Επεξεργασίας της NEΓ σε σύγχρονα περιβάλλοντα υλοποίησης των TNΔ.

1.2 Γραμματικές για την Επεξεργασία Φυσικής Γλώσσας

1.2.1 Εισαγωγή

Η ΕΦΓ έγκειται στην ανάλυση του κειμένου της, η οποία προϋποθέτει τη συντακτική ανάλυσή του ή την τεχνολόγησή του, κατά τη διάρκεια της διαδικασίας αυτής κάθε πρόταση του «απογραμματικοποιείται», δηλαδή από τις λέξεις της πρότασης εξάγεται μια δενδροειδής δομή. Αυτή η δομή περιγράφει τον συντακτικό ρόλο κάθε λέξης της πρότασης. Η δυνατότητα ή μη εξαγωγής της δενδροειδούς δομής επιβεβαιώνει αν η πρόταση ανήκει ή όχι στη γλώσσα. Στη διαδικασία αυτή τον κεντρικό ρόλο τον παίζει η Γραμματική, η οποία είναι ένα σχήμα που περιγράφει τα χαρακτηριστικά των προτάσεων της γλώσσας. Στην περιγραφή των φυσικών γλωσσών οι Γραμματικές Chomsky, καθώς και διάφορες τροποποιήσεις και επαυξήσεις που οφείλονται είτε στον Chomsky είτε σε άλλους ερευνητές παρουσίασαν κάποιες ανεπάρκειες. Πρέπει όμως να σημειώσουμε ότι η ανεπάρκεια αυτή παρατηρήθηκε και καταγράφηκε κυρίως στην περίπτωση της ΑΓ, η οποία διαφέρει σημαντικά από τη ΝΕΓ και από άποψη πληροφοριακού περιεχομένου των λέξεων (γένη, πτώσεις, κλπ.).

Οι ανεπάρκειες αυτές για την ΑΓ εντοπίζονται κυρίως στο ότι στις προταθείσες γραμματικές δεν ενσωματώνονται, σε ανάλογο βαθμό, μορφολογικές και κυρίως σημασιολογικές πληροφορίες, οι οποίες ως γνωστόν είναι αλληλοεξαρτώμενες με αυτές της σύνταξης. Επίσης στο ότι το μοντέλο τους δεν είναι γενικό, «καθολικό», για όλη τη γλώσσα για την οποία σχεδιάστηκαν, αλλά προτείνονται τόσες περιπτώσεις όσες για κάθε περίπτωση πρότασης ξεχωριστά. Χαρακτηριστικό παράδειγμα είναι οι γραμματικές Ενοποίησης όπου γίνεται μία προσπάθεια ενσωμάτωσης μορφολογικών και σημασιολογικών πληροφοριών, τα μοντέλα που προτάθηκαν είναι δύσκολα στην υλοποίηση εξαιτίας της πολυπλοκότητας του σχεδιασμού τους, η οποία επιδεινώνεται καθώς δεν προβλέπεται, και πάλι, ένα καθολικό μοντέλο που να καλύπτει όλες τις περιπτώσεις. Για την ΕΦΓ έχουν προταθεί κατά καιρούς και άλλα μοντέλα που χρησιμοποιούν πιθανότητες. Στις παρακάτω ενότητες θα παρουσιαστούν ενδεικτικά οι κυριότερες γραμματικές.

Λέξεις Κλειδιά: Γραμματικές, Ιεραρχία Chomsky, Επαυξημένα Δίκτυα Μετάβασης, Γραμματικές των Συνδεδεμένων Δέντρων, Γραμματικές Ενοποίησης, Λεξικές Λειτουργικές Γραμματικές, Ετικετοποίηση, Στοχαστικές Γραμματικές, Μητροειδείς Γραμματικές - Χαρακτηριστικοί Εκθέτες και Σπονδυλωτές Μητροειδείς Γραμματικές - Χαρακτηριστικοί Εκθέτες, Μορφολογικοί Αναλυτές, Σημασιολογικοί Αναλυτές,

Βασικό Νεοελληνικό Πολυλεξικό, Νεοελληνικό Μορφολογικό Υπολογιστικό Λεξικό, Νεοελληνικό Συντακτικό Υπολογιστικό Λεξικό, Νεοελληνικό Σημασιολογικό Λεξικό

1.2.2 Γενετικές και Μετασχηματιστικές Γραμματικές της Ιεραρχίας Chomsky, Επαυξημένα Δίκτυα Μετάβασης και Γραμματικές των Συνδεδεμένων Δέντρων

Το 1957 ο Avram Noam Chomsky δημοσιεύει τις Συντακτικές Δομές (Syntactic Structures), βιβλίο που αποτελεί σύνοψη της διδακτορικής του διατριβής (1955), η οποία είχε τίτλο «Η Λογική Δομή της Γλωσσολογικής Θεωρίας», η οποία δημοσιεύτηκε τελικά το 1975. Με τη μελέτη αυτή ο δομιστής γλωσσολόγος Noam Chomsky εισάγει τις «Γενετικές Μετασχηματιστικές Γραμματικές» για την περιγραφή της ΑΓ. Η έννοια της γραμματικής όπως την εισάγει ο Chomsky είναι ένα μαθηματικό μοντέλο, γεννήτρια γλώσσας, που παράγει αλυσίδες χαρακτήρων βάσει συγκεκριμένων κανόνων κάθε φορά. Οι κανόνες αυτοί είναι αναδρομικοί κανόνες, ορίζονται ως κανόνες παραγωγής ή φρασεοδομής και περιγράφουν τις δομές βάθους της ΑΓ.

Είναι αξιοσημείωτο ότι έχουν προηγηθεί: 1. ο Alan Matheson Turing που δίνει τη μεγάλη ώθηση στις εξελίξεις για τη δημιουργία του πρώτου Η/Υ με τη γνωστή «Θέση Turing», ότι: «κάθε εσωτερική διεργασία φυσικού φαινομένου που μπορεί να περιγραφεί σαν αποτελεσματική διαδικασία, δηλαδή σαν αλγόριθμος, μπορεί να γίνει αποδεκτή από μια «Μηχανή Turing² (MT)» καθώς και 2. ο Von Neumann που αναπτύσσει τη θεωρία των «Πεπερασμένων Αυτομάτων», δηλαδή, μαθηματικών μοντέλων - Αποδεκτών Γλωσσών - που αναγνωρίζουν αλυσίδες χαρακτήρων οι οποίες έχουν παραχθεί βάσει συγκεκριμένων κανόνων κάθε φορά.

Σύμφωνα με τη θεωρία που αναπτύχθηκε η Γραμματική είναι ένα μαθηματικό

² Η MT είναι η υλοποίηση ενός μαθηματικού μοντέλου, που μπορεί να προσαρμοστεί ώστε να προσομοιώνει τη λογική οποιουδήποτε αλγορίθμου, δηλαδή είναι μία απλή μηχανή η οποία θα μπορούσε να υπολογίσει - μηχανικά ή αυτόματα - οτιδήποτε μπορεί να υπολογιστεί, όπως για παράδειγμα την αναγνώριση ή μη αλυσίδων χαρακτήρων που έχουν προκύψει με συγκεκριμένους κανόνες, κ.λπ. Επίσης, η MT είναι ιδιαίτερα χρήσιμη στο να εξηγήει τις λειτουργίες μιας κεντρικής μονάδας επεξεργασίας στο εσωτερικό του Η/Υ, αφού η υλοποίησή της MT είναι ο σύγχρονος Η/Υ. Ας σημειωθεί ότι η εξέλιξη μιας MT είναι η ίδια MT. Με την αποφασιστικής σημασίας προσφορά του ο Turing, στο γνωστικό πεδίο της θεωρίας και όχι μόνο, θεωρείται ο πατέρας των Η/Υ. Στο τέλος του Β' παγκοσμίου πολέμου ο Turing προσκαλείται από το Εθνικό Εργαστήριο Φυσικής του Λονδίνου να σχεδιάσει έναν Υπολογιστή. Με τη σχετική αναφορά του τον Μάρτιο του 1946 έχει ήδη ριφθεί ο κύβος για τη μετέπειτα κατασκευή του πρώτου Η/Υ. Σημειώνεται ότι ο Turing με τη βοήθεια ενός άλλου μαθηματικού του W G Welchman, δημιούργησε τη «μηχανή Bombe», που στηρίχτηκε σε προγενέστερη δουλειά Πολωνών Μαθηματικών. Με τη μηχανή αυτή, κατά το Β' παγκόσμιο πόλεμο και από τα τέλη του 1940, οι σύμμαχοι αποκωδικοποιούσαν όλα τα μηνύματα που έστελναν οι Γερμανοί με τη βοήθεια της «μηχανής Enigma» της Luftwaffe.

σύστημα $G = (V_M, V_T, \Pi, S)$, όπου τα σύμβολα V_M, V_T, Π και S παριστούν αντίστοιχα, το πεπερασμένο σύνολο των μεταβλητών ή μη τερματικών στοιχείων, το πεπερασμένο σύνολο των τερματικών συμβόλων, το πεπερασμένο σύνολο των κανόνων παραγωγής ή αντικατάστασης ή φρασεοδομής που είναι αναδρομικοί κανόνες και τέλος το αρχικό σύμβολο. Είναι επίσης $V_M \cap V_T = \emptyset$.

Οι κανόνες παραγωγής, τα στοιχεία του συνόλου Π , έχουν τη μορφή $\alpha \rightarrow \beta$, όπου τα σύμβολα α, β παριστούν αλυσίδες αποτελούμενες από στοιχεία των συνόλων V_M , και V_T , δηλαδή $\alpha, \beta \in (V_M \cup V_T)^+$. Διακρίνονται τρεις τύποι Γραμματικών που αποτελούν την ιεραρχία Chomsky ανάλογα με ορισμένους περιορισμούς στους οποίους υποβάλλονται οι αλυσίδες α και β . Όταν τα μήκη των αλυσίδων αυτών υπόκεινται στον περιορισμό $|\alpha| < |\beta|$, η Γραμματική λέγεται τύπου 1 ή ευπαθής στα συμφραζόμενα. Εάν κάθε κανόνας μιας Γραμματικής G είναι της μορφής $A \rightarrow \alpha$, όπου $A \in V_M$ και $\alpha, \beta \in (V_M \cup V_T)^+$, τότε η Γραμματική λέγεται τύπου 2 ή ανεξάρτητη συμφραζομένων. Εάν η Γραμματική έχει κανόνες μόνο της μορφής $A \rightarrow aB$ ή της μορφής $A \rightarrow a$, όπου $\{A, B\} \in V_M$ και $a \in V_T$, τότε λέγεται τύπου 3 ή κανονική. Όταν οι κανόνες μιας Γραμματικής δεν υπόκεινται σε κανέναν περιορισμό τότε η Γραμματική είναι τύπου 0 και παράγει αναδρομικά απαριθμήσιμα σύνολα. Η γλώσσα L που παράγει μια Γραμματική G ορίζεται ως $L(G) = \{x / x \in V_T^* \ \& \ S \Rightarrow x \text{ βάσει των κανόνων της Γραμματικής και με όσες εφαρμογές τους χρειαστεί}\}$.

Στην περίπτωση των ΦΓ επεκράτησε οι μεταβλητές της Γραμματικής δηλαδή τα στοιχεία του V_M να είναι τα: <ΠΡ> (πρόταση), <ΟΦ> (ονοματική φράση), ΡΦ (ρηματική φράση), <ΑΡΘΡ> (άρθρο), <ΟΝ> (όνωμα), <ΕΠ> (επίθετο), <Ρ> (ρήμα), <ΕΠΡ> (επίρρημα) κ.λπ., ενώ τα τερματικά σύμβολα, δηλαδή τα στοιχεία του είναι οι λέξεις της ΦΓ. Από τις λέξεις αυτές και με την εφαρμογή των κανόνων φρασεοδομής κατασκευάζονται σωστές (συνακτικικά) προτάσεις που αντιστοιχούν στη δομή βάθους, όπως ορίζεται από αυτή τη φάση παραγόμενη φρασεοδομή. Ο μετασχηματισμός σε προτάσεις αποδεκτές από την εκάστοτε ΦΓ ως προς τη μορφολογία και τη σημασιολογία της υλοποιείται με τη βοήθεια των μετασχηματιστικών κανόνων. Η προσέγγιση αυτή εξασφαλίζει την ορθότητα της συντακτικής δομής των προτάσεων, σε ορισμένες δε περιπτώσεις γλωσσών, όπως π.χ. της Αγγλικής, εξασφαλίζεται και ο συντακτικός ρόλος των των λέξεων, ανάλογα με τη σειρά τους στην πρόταση. Για παράδειγμα η <ΟΦ> που προηγείται της <ΡΦ> είναι το υποκείμενο του <Ρ>, ενώ η <ΟΦ> που έπεται του <Ρ> είναι το αντικείμενο κ.λπ. Η ιδιότητα που μόλις

αναφέρθηκε δεν αποτελεί σε καμία περίπτωση ικανή συνθήκη εντοπισμού των συντακτικών ρόλων των λέξεων σε Νεοελληνικές προτάσεις,. Επίσης οι κανόνες της μορφολογίας και της σημασιολογίας δεν ενσωματώνονταν στους κανόνες της Γραμματικής.

Παράδειγμα 1

Το παράδειγμα με το οποίο θα ασχοληθούμε αποτελεί ένα υποσύνολο της Γλώσσας των Μαθηματικών. Συγκεκριμένα θέλουμε να εισάγουμε μια γραμματική που να παράγει το σύνολο των πολυωνύμων με ακέραιους συντελεστές:

$$p(x) = \alpha_n x^n + \alpha_{n-1} x^{n-1} + \dots + \alpha_2 x^2 + \alpha_1 x + \alpha \quad \text{ή}$$

$$p(x) = \alpha x^n + \alpha x^{n-1} + \dots + \alpha x^2 + \alpha x + \alpha \quad \text{ή}$$

σε μορφή Horner

$$p(x) = (\dots ((\alpha * x + \alpha) * x + \dots + \alpha * x + \alpha) * x) + \alpha, \alpha \in \mathbb{Z}$$

Η Γραμματική G που παράγει αυτόν τον συγκεκριμένο μικρόκοσμο, δηλαδή αυτό το υποσύνολο δείγματος της Γλώσσας των Μαθηματικών αποτελείται από:

$V_M = \{S\}$, το πεπερασμένο σύνολο των μη τερματικών συμβόλων ή μεταβλητών

$V_T = \{a, x, +, *, (,)\}$, το αλφάβητο ή πεπερασμένο σύνολο τερματικών συμβόλων

$\Pi = \{S \rightarrow (S)*x+\alpha \mid \alpha*x+\alpha\}$, το πεπερασμένο σύνολο κανόνων παραγωγής και

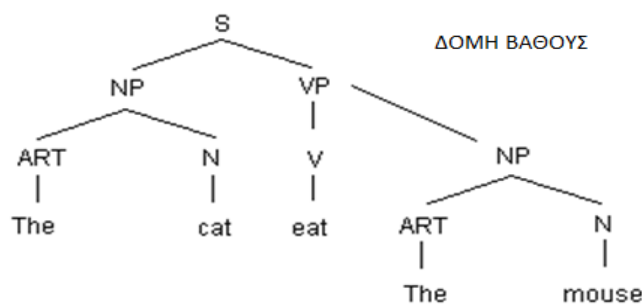
S : το αρχικό σύμβολο

Παράδειγμα 2

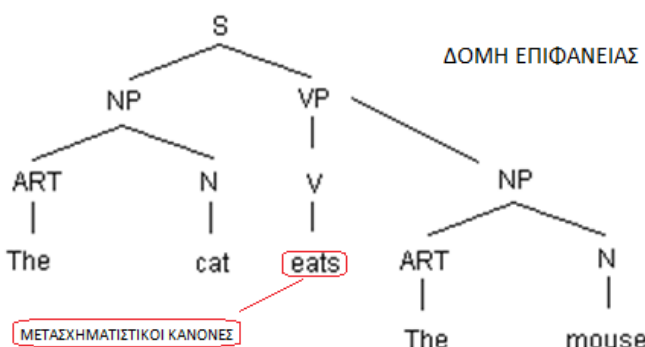
Έστω η Γραμματική $G = (V_M, V_T, \Pi, S)$ που αποτελείται από τα σύνολα:

- $V_M = \{S, NP, VP, V, N, Art\}$,
- $V_T = \{cat, eat, mouse, the\}$
- $\Pi = \{S \rightarrow NP VP,$
 $NP \rightarrow Art N$
 $VP \rightarrow V NP,$
 $V \rightarrow eat$
 $N \rightarrow cat \mid mouse,$
 $Art \rightarrow the\},$
- S

Η δένδροειδής δομή που περιγράφει τον συντακτικό ρόλο της πρότασης είναι:



Μετά την εφαρμογή των μετασχηματιστικών κανόνων προκύπτει η Δομή Επιφανείας:



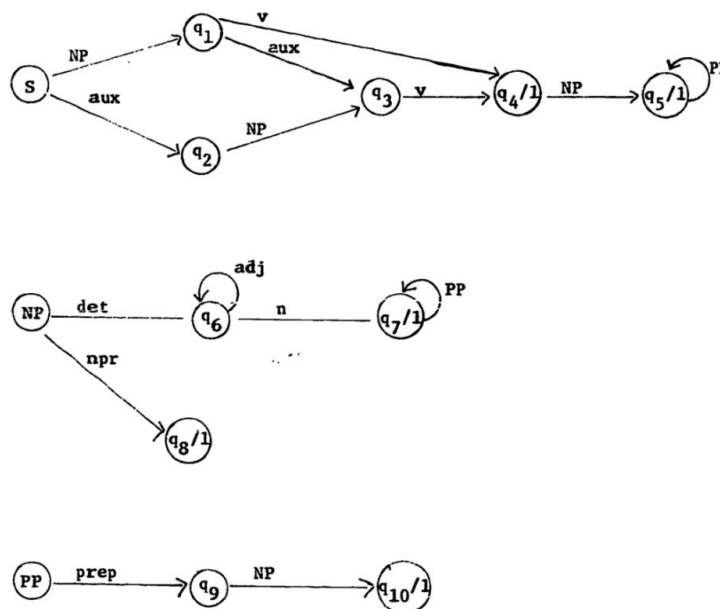
Οι Γραμματικές της Ιεραρχίας Chomsky παρουσιάζουν ανεπάρκειες για την περιγραφή της ΑΓ, που συνεπάγονται ανεπάρκειες και στην αυτόματη παραγωγή και αναγνώριση, των προτάσεων ή δομών επιφανείας της. Το γεγονός αυτό οδήγησε τις Γραμματικές της Ιεραρχίας Chomsky σε αδιέξοδο στην περιγραφή της ΑΓ, όμως ταυτόχρονα έδωσε μεγάλη ώθηση στην ανάπτυξη των Τυπικών Γλωσσών³. Με τις Γραμματικές της Ιεραρχίας Chomsky ορίζονται οι γλώσσες προγραμματισμού με τις οποίες γράφονται τα προγράμματα - κώδικες. Δηλαδή, οι διαδικασίες επίλυσης προβλημάτων ή αλγόριθμοι κωδικοποιούνται σε γλώσσα που καταλαβαίνουν οι Η/Υ, και με τα πεπερασμένα αυτόματα ελέγχονται, κατανοούνται και ενεργοποιούνται οι ανάλογες εσωτερικές διαδικασίες των Η/Υ για την υλοποίησή τους. Τα παραπάνω μοντέλα των Γεννητριών Γλωσσών και των αποδεκτών τους έπαιξαν σημαντικότατο ρόλο στην ανάπτυξη των Η/Υ, των εφαρμογών τους, κατά συνέπεια της Επεξεργασίας Φυσικής Γλώσσας (ΕΦΓ) και κατ' επέκταση της Γλωσσικής Τεχνολογίας.

Τα Αναδρομικά και τα Επαυξημένα δίκτυα Μετάβασης εισάγονται από τον Woods για να περιγράψουν καλύτερα τους κανόνες φρασεοδομής του Chomsky. Έτσι ένα

³ Γλώσσες που επινοήθηκαν για συγκεκριμένους σκοπούς κάθε φορά, όπως για παράδειγμα είναι οι γλώσσες που έχουν σχέση με τους Η/Υ κι είναι: οι γλώσσες περιγραφής αλγορίθμων, οι γλώσσες προγραμματισμού και οι μεταγλώσσες, δηλαδή γλώσσες που περιγράφουν γλώσσες. Στις τυπικές γλώσσες, ο ορισμός των κανόνων προηγείται και ό,τι προκύπτει από την εφαρμογή τους είναι στοιχείο της (τυπικής) γλώσσας, σε αντίθεση με τις φυσικές γλώσσες, που προϋπάρχουν των κανόνων, οι οποίοι απλά τις περιγράφουν. Σημειώνουμε επίσης ότι, στις τυπικές γλώσσες δεν υπάρχει μορφολογία, η σύνταξη ταυτίζεται με τη σημασιολογία.

Αναδρομικό Δίκτυο Μετάβασης (ΑΔΜ) είναι ένα κατευθυνόμενο γράφημα με επισημασμένες καταστάσεις και τόξα, μια ανεξάρτητη κατάσταση που ονομάζεται κατάσταση έναρξης και ένα διακεκριμένο σύνολο καταστάσεων που ονομάζεται τελική κατάσταση. Στην ουσία μοιάζει αρκετά με ένα Μη Αιτιοκρατικό Πεπερασμένο Μεταβατικό Διάγραμμα, με μόνη διαφορά ότι εδώ οι ετικέτες στα τόξα μπορεί να είναι ονόματα καταστάσεων καθώς και τερματικά σύμβολα.

Στο παρακάτω σχήμα της Εικόνας 1 περιγράφεται χαρακτηριστικά ένα παράδειγμα της λειτουργίας των ΑΔΜ για ένα μικρό υποσύνολο Αγγλικών. Το Δίκτυο του σχήματος αποδέχεται προτάσεις όπως για παράδειγμα “John washed the car” και “Did the red barn collapse?” κ.λπ. και για να αναγνωριστεί π.χ. η δεύτερη πρόταση, το δίκτυο θα ξεκινήσει τη λειτουργία του από την κατάσταση S.



Εικόνα 1: Δείγμα Αναδρομικού Δικτύου Μετάβασης όπου το S είναι η αρχική κατάσταση και τα q4, q5, q7, q8, q10 είναι οι τελικές καταστάσεις

Η πρώτη μετάβαση που κάνει το δίκτυο είναι η μετάβαση “aux” η οποία μεταβαίνει στην κατάσταση q2 λόγω του βοηθητικού ρήματος “did” της αλυσίδας χαρακτήρων. Έπειτα, από την κατάσταση q2 μπορεί να μεταβεί στην κατάσταση q3 εάν και μόνο αν η επόμενη λέξη της αλυσίδας χαρακτήρων της εισόδου είναι ουσιαστικό, όπου για να διαπιστωθεί εάν συμβαίνει αυτό, θα πρέπει να κληθεί η κατάσταση NP. Από την κατάσταση NP θα προχωρήσει το δίκτυο στο τόξο που είναι επισημασμένο με την λέξη “det” για να πάει στην κατάσταση q6, λόγω του προσδιοριστή “the”. Στην κατάσταση που βρίσκεται τώρα το δίκτυο, θα δημιουργηθεί ένας βρόχος στην

κατάσταση q_6 που θα προκληθεί από το επίθετο “red” και το επόμενο ουσιαστικό “barn” θα προκαλέσει τη μετάβαση στην κατάσταση q_7 . Δεδομένου ότι η κατάσταση q_7 είναι μια τελική κατάσταση, καθίσταται εφικτό να εμφανιστεί από τον NP υπολογισμό και να συνεχίσει τον υπολογισμό του ανώτατου επιπέδου S αρχίζοντας από την κατάσταση q_3 η οποία βρίσκεται στο τέλος του τόξου NP. Από την κατάσταση q_3 , το ρήμα “collapse” επιτρέπει τη μετάβαση στην κατάσταση q_4 και εφόσον αυτή η κατάσταση είναι και η τελική και η λέξη “collapse” είναι η τελευταία λέξη της αλυσίδας χαρακτήρων, τότε προκύπτει ότι η αλυσίδα χαρακτήρων γίνεται δεκτή ως πρόταση.

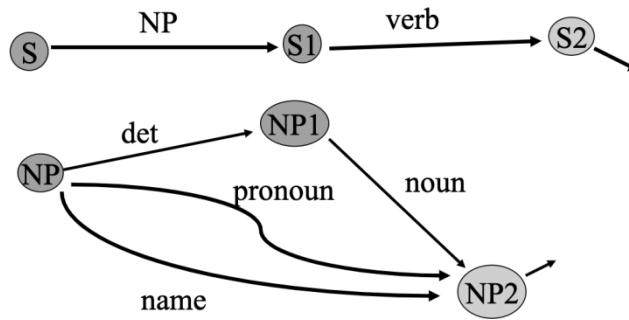
Ένα Επαυξημένο Δίκτυο Μετάβασης (ΕΔΜ) μπορεί να θεωρηθεί ως ένας τύπος γραφήματος της θεωρητικής δομής που χρησιμοποιείται στις Κανονικές Γλώσσες και έχει ευρεία εφαρμογή στον τομέα της TN. Ένα ΕΔΜ μπορεί, θεωρητικά, να αναλύσει τη δομή οποιασδήποτε πρότασης, όσο περίπλοκη και αν είναι. Τα ΕΔΜ αποτελούν στην ουσία μια τροποποιημένη εκδοχή των Αναδρομικών Δικτύων Μετάβασης.

Τα ΕΔΜ βασίζονται στην ιδέα της χρήσης των μηχανών πεπερασμένων καταστάσεων (μοντέλο Markov) για την ανάλυση προτάσεων. Ο W.A.Woods στο βιβλίο του με τίτλο “Transition Network Grammars for Natural Language Analysis” ισχυρίζεται ότι προσθέτοντας έναν αναδρομικό μηχανισμό σε ένα μοντέλο πεπερασμένων καταστάσεων, έχει σαν αποτέλεσμα η ανάλυση να μπορεί να επιτευχθεί πιο αποτελεσματικά.

Τα ΕΔΜ κατασκευάζουν μια μερική διαρθρωτική περιγραφή της πρότασης καθώς προχωρούν από κατάσταση σε κατάσταση μέσα στο δίκτυο. Τα κομμάτια αυτής της μερικής περιγραφής κρατούνται μέσα σε καταχωρητές οι οποίοι μπορούν να περιέχουν οποιοδήποτε ριζωμένο δέντρο ή έναν κατάλογο από ριζωμένα δέντρα και επίσης ωθούνται προς τα κάτω αυτόματα όταν μια αναδρομική εφαρμογή του Επαυξημένου Δικτύου καλείται.

Ένα χαρακτηριστικό παράδειγμα της λειτουργίας ενός ΕΔΜ περιγράφεται στο σχήμα της Εικόνας 2 που ακολουθεί.

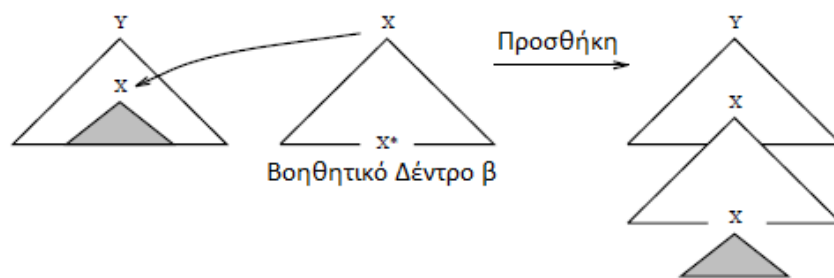
Στο σχήμα της Εικόνας 2 το επάνω δίκτυο καθορίζει τη δομή της πρότασης ενώ το κάτω δίκτυο αντιστοιχεί στην NP μετάβαση και θα χρησιμοποιηθεί για να επεξεργαστεί μια φράση ουσιαστικού ως αντικείμενο της πρότασης.



Εικόνα 2: Παράδειγμα λειτουργίας ενός Επαυξημένου Δικτύου Μετάβασης

Ο φορμαλισμός των Γραμματικών των Συνδεδεμένων Δέντρων (ΓΣΔ) (Joshi, 1987) θεωρείται ιδιαίτερα κατάλληλος για την περιγραφή πολλών γλωσσικών φαινομένων. Μια ΓΣΔ δίνεται από ένα σύνολο στοιχειωδών δέντρων τα οποία έπειτα χωρίζονται σε δύο κατηγορίες δέντρων: τα αρχικά δέντρα και τα βοηθητικά δέντρα. Οι εσωτερικοί κόμβοι των δέντρων επισημαίνονται από μη-τερματικά σύμβολα και τα φύλλα τους από μη τερματικά ή και τερματικά. Κάθε βοηθητικό δέντρο έστω β έχει ένα διακεκριμένο φύλλο, το οποίο λέγεται και πόδι και επισημαίνεται από μη τερματικά σύμβολα, όπως ακριβώς δηλαδή και ο κόμβος της ρίζας του β .

Για την παραγωγή/εξαγωγή των δέντρων από τα στοιχειώδη δέντρα μπορούν αν γίνουν δύο πράξεις. Η πρώτη πράξη που μπορεί να εκτελεστεί, ονομάζεται αντικατάσταση και στην ουσία κατά τη διάρκεια της πράξης αυτής αντικαθίσταται ένας κόμβος φύλλο που έχει επισημανθεί από ένα μη τερματικό σύμβολο x από ένα αρχικό δέντρο « α » του οποίου η ρίζα είναι επίσης επισημασμένη με το σύμβολο x .



Εικόνα 3: Απεικόνιση της πράξης της πρόσθεσης

Η δεύτερη πράξη, που ονομάζεται προσθήκη/συμπλήρωση απεικονίζεται χαρακτηριστικά στο σχήμα της Εικόνας 3 που ακολουθεί. Στην πράξη αυτή, ένα βοηθητικό δέντρο β , του οποίου η ρίζα φέρει την ένδειξη x , μπορεί να προστεθεί σε οποιονδήποτε κόμβο Y επισημασμένο με το σύμβολο x . Επομένως, το υποδέντρο που είναι ριζωμένο στο v είναι εμβολιασμένο πλέον στο πόδι του β .

1.2.3 Γραμματικές Ενοποίησης – Λεξικές Λειτουργικές Γραμματικές

Οι Γραμματικές Ενοποίησης αποτέλεσαν ένα σημαντικό κεφάλαιο στην έρευνα για την ΕΦΓ καθώς κατάφεραν να επιτρέψουν μια υψηλού επιπέδου περιγραφή των Γραμματικών σχέσεων. Στις Γραμματικές αυτές οι δομές των χαρακτηριστικών χρησιμοποιούνται για την κωδικοποίηση κάθε είδους πληροφορίας που σχετίζεται με οποιοδήποτε δεδομένο γλωσσικό/γλωσσολογικό αντικείμενο, για παράδειγμα τις συντακτικές ιδιότητες ή το σημασιολογικό περιεχόμενο. Οι κανόνες των Γραμματικών Ενοποίησης ορίζουν τη σχέση μεταξύ των επιφανειακών αλυσίδων χαρακτήρων μιας γλώσσας με τις πληροφορίες που σχετίζονται με αυτήν.

Οι Γραμματικές Ενοποίησης ταξινομούνται σε πέντε μεγάλες κατηγορίες:

- Στις Κατηγορηματικές Γραμματικές οι οποίες χρησιμοποιούν μια ακολουθία αλυσίδων χαρακτήρων για το σχηματισμό των συστατικών,
- στις Γραμματικές Γενικευμένης Φρασεοδομής, οι οποίες αποτελούν ένα πλαίσιο για την περιγραφή της σύνταξης και της σημασιολογίας των ΦΓ,
- στις Γραμματικές Οδηγούμενες από την Κεφαλή, οι οποίες αποτελούν μια επέκταση των ΓΓΦ και εισάγουν τη δυνατότητα περιτύλιξης της κεφαλής των αλυσίδων χαρακτήρων και την κατάργηση των περιορισμών του συστήματος των χαρακτηριστικών,
- στις Λεξικές Λειτουργικές Γραμματικές, οι οποίες αποτελούν ένα γραμματικό πλαίσιο το οποίο βασίζεται σε περιορισμούς της θεωρητικής γλωσσολογίας και τέλος
- στις Γραμματικές FUG, οι οποίες επιτρέπουν τη σύζευξη ως μια υποχρεωτική λειτουργία για τον σχηματισμό των συστατικών.

Στην παρούσα εργασία θα εστιάσουμε και θα αναλύσουμε περαιτέρω τις Λεξικές Λειτουργικές Γραμματικές καθώς είναι οι πιο διαδοσμένες Γραμματικές σε σύγκριση με τις άλλες αλλά και οι πιο ευρέως χρησιμοποιούμενες.

Οι Λεξικές Λειτουργικές Γραμματικές (ΛΛΓ) είναι μια θεωρία που έχει να κάνει με τη δομή της γλώσσας και το πώς σχετίζονται οι διαφορετικές πτυχές της γλωσσικής δομής. Όπως ακριβώς υποδηλώνει και το όνομα τους, η θεωρία αυτή είναι λεξική καθώς: το λεξικό είναι πλούσια δομημένο, με λεξικές σχέσεις και όχι με μετασχηματισμούς ή λειτουργίες πάνω σε δέντρα φρασεοδομής ως κάποιο μέσο καταγραφής των γλωσσικών γενικεύσεων. Επιπρόσθετα, η θεωρία αυτή είναι και Λειτουργική καθώς: οι Γραμματικές Λειτουργίες όπως είναι το υποκείμενο και το

αντικείμενο αποτελούν θεμελιώδη στοιχεία της θεωρίας, τα οποία δεν ορίζονται από άποψη διαμόρφωσης της φρασεοδομής ή των σημασιολογικών ρόλων.

Παράδειγμα

Στο σχήμα που ακολουθεί απεικονίζεται χαρακτηριστικά ο φορμαλισμός των Γραμματικών Ενοποίησης όπου μπορεί να παρατηρηθεί ότι έχουν ενσωματωθεί και μορφολογικές και σημασιολογικές πληροφορίες με πιο σύνθετο όμως τρόπο συγκριτικά με τις προηγούμενες Γραμματικές που αναφέρθηκαν. Πιο συγκεκριμένα περιγράφονται πιο σύνθετες έννοιες όπως είναι το Υποκείμενο (Subject), το Άμεσο Αντικείμενο (Direct Object), η ετικέτα Head που αναφέρεται στο Ουσιαστικό της Ονοματικής Φράσης που ανήκει, κ.ά.

$$\left[\begin{array}{l} \text{CAT} = \text{S} \\ \text{SUBJ} = \text{he} \\ \text{DOBJ} = \left[\begin{array}{l} \text{CAT} = \text{NP} \\ \text{HEAD} = \text{books} \\ \text{MOD} = \left[\begin{array}{l} \text{CAT} = \text{PRES} \\ \text{LEX} = \text{WRITE} \end{array} \right] \end{array} \right] \\ \text{VERB} = \text{LIKE} \\ \text{TENSE} = \text{PRES} \\ \text{VOICE} = \text{ACTIVE} \end{array} \right]$$

Οι ΛΛΓ υποθέτουν ότι υπάρχουν δύο συντακτικά επίπεδα τα οποία είναι σημαντικά για την ανάλυση της γλωσσικής δομής. Τα δύο αυτά επίπεδα είναι τα εξής: 1. Η δομή-F, η οποία αντιπροσωπεύει τις αφηρημένες γραμματικές λειτουργίες όπως το υποκείμενο και το αντικείμενο, καθώς επίσης και τα αφηρημένα χαρακτηριστικά όπως τον χρόνο του ρήματος και την πτώση και 2. Η δομή-C, η οποία αντιπροσωπεύει τη συμπαγή φραστική δομή αυτών των σχέσεων που αναφέρθηκαν προηγουμένως, η οποία διέπεται από περιορισμούς της γλώσσας.

Οι κανονικότητες που προκύπτουν από την σχέση της δομής-F και της δομής-C προκύπτουν από λειτουργίες που σχετίζουν τα τμήματα της μιας δομής με τα τμήματα της άλλης. Για παράδειγμα, η φράση του υποκειμένου στο δέντρο της δομής-C σχετίζεται με το υποκείμενο της δομής-F μέσω μιας συνάρτησης η οποία συνδέει τους κόμβους του δέντρου της δομής-C με τμήματα της δομής-F για μια πρόταση. Επίσης, έχουν διερευνηθεί και οριστεί και άλλες σχέσεις μεταξύ των δύο αυτών δομών και άλλων γλωσσικών επιπέδων από την άποψη των λειτουργικών χαρτογραφήσεων από τα υποτμήματα της μιας δομής στα αντίστοιχα υποτμήματα άλλων δομών.

1.2.4 Ετικετοποίηση

Στη Γλωσσολογία των Σωμάτων Κειμένων, η Ετικετοποίηση με μέρη του λόγου είναι η διαδικασία μέσω της οποίας οι λέξεις ενός κειμένου λαμβάνουν μια ετικέτα/σήμανση, η οποία σχετίζεται με ένα συγκεκριμένο μέρος του λόγου βάσει τόσο του ορισμού της όσο και των συμφραζομένων της, δηλαδή, με βάση τη σχέση της με τις γειτονικές και συναφείς λέξεις σε μια φράση, πρόταση ή παράγραφο. Τα μέρη του λόγου στην ουσία εξηγούν τον τρόπο με τον οποίο μια λέξη χρησιμοποιείται σε μια πρόταση. Υπάρχουν 8 κύρια μέρη του λόγου τα οποία είναι τα εξής: Ουσιαστικό, Αντωνυμίες, Επίθετα, Ρήματα, Επιρρήματα, Προθέσεις, Συζεύξεις και Επιφωνήματα. Στην Ετικετοποίηση το κάθε ένα από αυτά τα 8 μέρη του λόγου εμφανίζεται με κάποιον μοναδικό κωδικό όταν μια λέξη ετικετοποιείται, οι κωδικοί αυτοί είναι οι ακόλουθοι: Ουσιαστικό (N), Ρήμα (V), Επίθετο (ADJ), Επίρρημα (ADV), Πρόθεση (P), Σύζευξη (CON), Αντωνυμία (PRO) και Επιφώνημα (INT).

Καθοριστικό ρόλο για την ομαλή λειτουργία και χρήση της Ετικετοποίησης παίζουν οι Ετικετοποιητές οι οποίοι χωρίζονται σε δύο κατηγορίες:

1. Ετικετοποιητές που βασίζονται σε κανόνες και
2. Ετικετοποιητές που βασίζονται σε στοχαστικές μεθόδους ή Στοχαστικοί Ετικετοποιητές.

Οι Ετικετοποιητές που βασίζονται σε κανόνες χρησιμοποιούν τις πληροφορίες των συμφραζομένων μιας λέξης για την εκχώρηση ετικετών σε άγνωστες ή διαφορούμενες λέξεις. Η αποσαφήνιση εδώ γίνεται μέσω της ανάλυσης των γλωσσικών χαρακτηριστικών της λέξης, της προηγούμενης λέξης, της επόμενης λέξης και άλλων πτυχών. Ωστόσο, η αυτόματη ετικετοποίηση των μερών του λόγου είναι ένας τομέας ΕΦΓ όπου οι στατιστικές τεχνικές αποδείχθηκαν ότι είναι πιο επιτυχημένες από τις τεχνικές που βασίζονται σε κανόνες.

Οι Ετικετοποιητές που βασίζονται σε στοχαστικές μεθόδους αποσαφηνίζουν τις λέξεις με βάση μόνο την πιθανότητα να εμφανιστεί μια λέξη με μια συγκεκριμένη ετικέτα. Με άλλα λόγια, η ετικέτα που συναντάται πιο συχνά να έχει εκχωρηθεί σε κάποια λέξη στο σύνολο εκπαίδευσης είναι και αυτή που έχει εκχωρηθεί σε μια διαφορούμενη περίπτωση αυτής της λέξης. Το πρόβλημα όμως που συναντάται σε αυτή τη προσέγγιση είναι ότι ενώ μπορεί να αποδώσει μια έγκυρη ετικέτα για μια δεδομένη λέξη, υπάρχει ο κίνδυνος να αποδώσει με λάθος τρόπο κάποιες ακολουθίες ετικετών.

Για το λόγο αυτό, εισήχθη μια εναλλακτική λύση για την προσέγγιση των

συχνοτήτων των λέξεων, η οποία υπολογίζει την πιθανότητα μιας δεδομένης ακολουθίας ετικετών με βάση τη συχνότητα εμφάνισής τους. Αυτή η προσέγγιση αναφέρεται και ως προσέγγιση n-gram η οποία είναι πιο ισχυρή από αυτή που καθορίστηκε προηγουμένως, καθώς θεωρεί ότι οι ετικέτες είναι μεμονωμένες λέξεις που βασίζονται στα συμφραζόμενα.

Τέλος, η καλύτερη και πιο διαδεδομένη προσέγγιση που μπορεί να εισαχθεί σε έναν στοχαστικό Ετικετοποιητή και συνδυάζει τις προηγούμενες δύο προσεγγίσεις που αναφέρθηκαν είναι τα Κρυφά Μοντέλα Markov. Για την μοντελοποίηση οποιουδήποτε προβλήματος, τα Μοντέλα Markov χρειάζονται ένα σύνολο παρατηρήσεων και ένα σύνολο πιθανών καταστάσεων, όπου οι παρατηρήσεις είναι οι ίδιες οι λέξεις στη δεδομένη ακολουθία και οι καταστάσεις θα είναι οι ετικέτες των μερών του λόγου για τις λέξεις.

Ωστόσο, αν και η Ετικετοποίηση βοήθησε σημαντικά στην σωστή και γρήγορη απόδοση των μερών του λόγου παρουσίασε κάποιες ανεπάρκειες στην χρήση της. Οι ανεπάρκειες αυτές σημειώθηκαν κυρίως στην ΑΓ όπου ένας Ετικετοποιητής δεν ήταν σε θέση να καταλάβει εάν κάποιες συγκεκριμένες λέξεις έπρεπε να ετικετοποιηθούν ως Επίθετα ή Ουσιαστικά, γεγονός το οποίο μετέτρεπε τον αλγόριθμο της Ετικετοποίησης σε έναν εξαιρετικά πολύπλοκο αλγόριθμο. Αυτή ήταν και η αιτία που ώθησε τους επιστήμονες της Υπολογιστικής Γλωσσολογίας στην ανακάλυψη νέων μεθόδων προσέγγισης των Γραμματικών – τις Στοχαστικές Γραμματικές – οι οποίες θα μελετηθούν παρακάτω.

1.2.5 Στοχαστικές Γραμματικές

Μια Στοχαστική Γραμματική ή Στατιστική Γραμματική είναι ένα πλαίσιο Γραμματικής με μια πιθανοτική έννοια. Η ΕΦΓ που χρησιμοποιεί μεθόδους στατιστικής χρησιμοποιεί στοχαστικές, πιθανοτικές και στατιστικές μεθόδους για να αντιμετωπίσει τις δυσκολίες που προκύπτουν από τις μεγάλες προτάσεις που είναι εξαιρετικά διαφορούμενες αποδίδοντας χιλιάδες ή εκατομμύρια πιθανές αναλύσεις.

Μια Στοχαστική Γραμματική G_s μπορεί να θεωρηθεί ως το ζεύγος $\langle G_c, D \rangle$, όπου G_c είναι μια Γραμματική τύπου- i (με $i \in \{0, 1, 2, 3\}$) και D είναι ένας εκχωρητής πιθανοτήτων. Οι πιθανότητες ορίζονται ως οι πραγματικές τιμές στο διάστημα $[0, 1]$ και το G_c καλείται ως η Χαρακτηριστική Γραμματική του G_s .

Ένας άλλος τύπος Στοχαστικών Γραμματικών που αξίζει να αναφερθεί είναι οι

Απεριόριστες Στοχαστικές Γραμματικές. Μια Απεριόριστη Στοχαστική Γραμματική G_u εκφράζεται με το ζεύγος $\langle G_c, D \rangle$, όπου G_c – που ονομάζεται Χαρακτηριστική Γραμματική του G_u – είναι μια Γραμματική Ανεξάρτητη Συμφραζομένων και D είναι ένας εκχωρητής πιθανοτήτων.

Τέλος, υπάρχει και ένας άλλος τύπος Στοχαστικών Γραμματικών, οι Σταθερές Στοχαστικές Γραμματικές, οι οποίες χρησιμοποιούνται στην περίπτωση που το μέτρο της πιθανότητας p που επηρεάζεται από το G_s στη γλώσσα $L(G_c)$ πληροί την ακόλουθη σχέση: $\sum_{x \in L(G)} p(x) = 1$, διαφορετικά αν δεν ισχύει η εξίσωση αυτή τότε η Γραμματική θα καλείται μη σταθερή.

1.2.6 Μητροειδείς και Σπονδυλωτές Μητροειδείς Γραμματικές-Χαρακτηριστικοί Εκθέτες και το Βασικό Νεοελληνικό Πολυλεξικό

Οι ανεπάρκειες που εντοπίζονται στις προταθείσες Γραμματικές:

1. αφορούν κυρίως την άκλιτη ή ημίκλιτη ΑΓ και πολλαπλασιάζονται για κλιτές ή πολύκλιτες γλώσσες, όπως για παράδειγμα είναι η ΝΕΓ.
2. προκαλούνται διότι δεν ενσωματώνονται στις Γραμματικές, σε ανάλογο βαθμό, μορφολογικές και κυρίως σημασιολογικές πληροφορίες, οι οποίες ως γνωστόν είναι αλληλοεξαρτώμενες με αυτές της σύνταξης, επίσης,
3. δεν είναι το μοντέλο τους γενικό, «καθολικό», για όλη τη γλώσσα για την οποία σχεδιάστηκαν, αλλά προτείνονται τόσες περιπτώσεις όσες για κάθε περίπτωση πρότασης ξεχωριστά. Χαρακτηριστικό παράδειγμα είναι οι Γραμματικές Ενοποίησης, όπου γίνεται μία προσπάθεια ενσωμάτωσης μορφολογικών και σημασιολογικών πληροφοριών. Τα μοντέλα που προτάθηκαν είναι δύσκολα στην υλοποίηση εξαιτίας της πολυπλοκότητας του σχεδιασμού τους, η οποία επιδεινώνεται καθώς δεν προβλέπεται, και πάλι, ένα καθολικό μοντέλο που να καλύπτει όλες τις περιπτώσεις.

Η έρευνα έδειξε ότι οι δομολειτουργικές διεργασίες ακόμα και μίας πολύκλιτης ΦΓ, όπως είναι η ΝΕΓ, μπορούν να περιγραφούν ως αλγόριθμοι. Επιπλέον η έρευνα έδειξε ότι, οι αλγόριθμοι αυτοί σε συνδυασμό με τη μελέτη, ταξινόμηση, κωδικοποίηση και αρχειοθέτηση των πολυδιάστατων γλωσσολογικών δεδομένων της ή διανυσμάτων της, κατάλληλα σε βάσεις δεδομένων, επιτρέπουν να προσεγγιστεί η επεξεργασία της αναλυτικά και με στόχους ακρίβειας 100%.

Η παρατήρηση ότι στη μη κλιτή ή ημίκλιτη ΑΓ οι πληροφορίες που φέρουν οι

λέξεις στην πρόταση σχετικά με τον συντακτικό τους ρόλο καθορίζεται από την θέση τους (πρβλ SVO) ενώ στην πολύκλιτη NEG από τις πτώσεις τους οδηγεί στο συμπέρασμα ότι οι κανόνες φρασεοδομής των Γραμματικών της Ιεραρχίας Chomsky, που είναι και η βάση όλων των επεκτάσεων ή επαυξήσεων αυτών και εκφράζουν τη συντακτική δομή των προτάσεων της ΑΓ, δεν μπορούν να εκφράσουν επαρκώς τη συντακτική δομή των προτάσεων της NEG. Για παράδειγμα η πρόταση: «η γάτα τρώει τον ποντικό» στη NEG μπορεί να διατυπωθεί διατηρώντας το ίδιο νόημα με 6 διαφορετικούς τρόπους αλλάζοντας τη θέση των λέξεων. Στην ΑΓ όμως: «the cat eats the mouse» διατυπώνεται μόνο με έναν τρόπο. Η πρόταση: «the mouse eats the cat», η μοναδική επιτρεπτή αλλαγή λέξεων, σημαίνει κάτι τελείως διαφορετικό. Επισημαίνουμε το πόσο σημαντική είναι η παθητική φωνή για την ΑΓ, αφού είναι η μοναδική δυνατότητα που έχει να αντιστρέψει τη θέση υποκειμένου και αντικειμένου, χωρίς να αλλάξει το νόημα της πρότασης, εισάγοντας το ποιητικό αίτιο, μία πρακτική η οποία στη NEG δεν είναι διαδεδομένη.

Η διατύπωση λοιπόν Γραμματικής που παράγει τις συντακτικές δομές των προτάσεων της NEG συνεπάγεται την παραγωγή άρα και τη δυνατότητα επεξεργασίας των συνταγματικών αξόνων τους, καθώς και του ορισμού και ποικιλίας των παραδειγματικών αξόνων τους σε κάθε περίπτωση.

Η παρατήρηση ότι οι συντακτικές κατηγορίες της NEG που συνθέτουν τους συνταγματικούς άξονες είναι τα σημεία όπου εντοπίζεται η αλληλεξάρτηση της σύνταξης, της μορφολογίας και της σημασιολογίας οδήγησε στη διατύπωση των Μητροειδών Γραμματικών (μ. γραμματικών) – Χαρακτηριστικοί Εκθέτες (χ.ε.), μίας τροποποιημένης εκδοχής των Γραμματικών Chomsky. Η τροποποίηση δεν αφορά τους τύπους των Γραμματικών οι οποίοι παραμένουν ως έχουν, αλλά τους κανόνες παραγωγής, οι οποίοι είναι κανόνες συντακτικής δομής και όχι φρασεοδομής, και τους ομαδοποιούν σε μήτρες συντακτικών κανόνων, οι οποίες παράγουν συντακτικές δομές συνήθων και εύχρηστων προτάσεων της NEG. Δηλαδή, στο προτεινόμενο τροποποιημένο σχήμα $G = (V_M, V_T, \Pi, S)$, Π : είναι το πεπερασμένο σύνολο μητρών συντακτικών κανόνων, και όχι φρασεοδομής, επιπλέον V_T : είναι το πεπερασμένο σύνολο των συντακτικών κατηγοριών της NEG και όχι λέξεις της. Για παράδειγμα, οι συντακτικές δομές ή συνταγματικοί άξονες των απλών προτάσεων που παράγονται από τη μ.γραμματική $G_1 = (V_{M1}, V_{T1}, \Pi_1, S)$, όπου $\Pi_1 = \{ \pi_{11}, \pi_{12} \}$ με π_{11} παράγει τη συντακτική δομή: <υποκειμενο><κατηγορηματικο_ρημα> και π_{12} παράγει τη συντακτική

δομή: <υποκειμενο><συνδετικο_ρημ><κατηγορουμενο>. Με ανάλογο τρόπο οι G_2 , G_3 και G_4 παράγουν αντίστοιχα τις συντακτικές δομές των σύνθετων, των επαυξημένων και των σύνθετων και επαυξημένων προτάσεων της ΝΕΓ. Οι μ. γραμματικές παράγουν μία ενδιάμεση γλώσσα, τη γλώσσα των συντακτικών κατηγοριών ή γλώσσα βάση της ΝΕΓ, η οποία προσομοιάζει με τη ΝΕΓ ως προς τη συντακτική δομή της. Το πλεονέκτημα είναι ότι είναι απαλλαγμένη από γραμματικούς τύπους και έννοιες και μελετάται εύκολα η συντακτική συμπεριφορά της αφού είναι μία τυπική γλώσσα που εύκολα προγραμματίζεται. Επίσης κάθε παραγόμενη συντακτική δομή της γλώσσας αναγνωρίζεται ευκολότερα αφού οι κανόνες αυτοί εντοπίζονται άμεσα με τη βοήθεια των χ.ε. Οι χ.ε. είναι ακολουθίες αριθμητικών τιμών των οποίων η σύνθεση και οι τιμές τους αντιστοιχούν αμφιμονοσήμαντα σε κάθε μήτρα και στο αντίστοιχο περιεχόμενο κάθε παραγόμενης συντακτικής δομής της. Με τον τρόπο αυτόν αποφεύγονται οι δαπανηρές ευρεστικές μέθοδοι εντοπισμού των κανόνων του συνόλου P των Γραμματικών της Ιεραρχίας Chomsky.

Με ανάλογο τρόπο ορίζονται και οι επεκτάσεις των μ. γραμματικών, οι Σπονδυλωτές Μητροειδείς Γραμματικές (σ.μ. γραμματικές) – Χαρακτηριστικοί Εκθέτες για την παραγωγή και αναγνώριση των συντακτικών δομών κειμένων ή κειμενοποιημένων προτάσεων της ΝΕΓ. Έστω η μ. γραμματική: $G = (V_M, V_T, \Pi, S)$, ονομάζουμε σ.μ. γραμματική τη $G_\Sigma = (V_{M\Sigma}, V_{T\Sigma}, \Pi_\Sigma, S)$, για την οποία ισχύουν $V_{M\Sigma} \equiv V_N$, $V_{T\Sigma} = V_T \cup \{ u, v, w \}$ και $\Pi_\Sigma = \Pi \cup \Pi_{K\Sigma 1} \cup \Pi_{K\Sigma 2} \cup \Pi_{K\Sigma 3}$, το πεπερασμένο σύνολο κανόνων παραγωγής και μητρών κανόνων σπονδυλωτής σύνδεσης. Εξ ορισμού μία σ.μ. γραμματική πέραν της καταγραφής, κατονομασίας των δυνατών συνδυασμών των κανόνων που παράγουν τις μη κειμενοποιημένες (απλές) συντακτικές δομές της γλώσσας, δίνει τη δυνατότητα να καταγραφούν και να κατονομαστούν οι κανόνες που παράγουν τις κειμενοποιημένες (σύνθετες) συντακτικές δομές της γλώσσας. Σημειώνουμε ότι με τις σ.μ. γραμματικές: 1. δεν αυξάνονται και δεν διατυπώνονται πολυπλοκότερες μήτρες κανόνων παραγωγής συντακτικών δομών, με αποτέλεσμα κάθε παραγόμενη σύνθετη συντακτική δομή της γλώσσας να αναγνωρίζεται ευκολότερα αφού οι κανόνες αυτοί εντοπίζονται άμεσα με τους χ.ε., όπως ακριβώς και στις μ. γραμματικές, αποφεύγοντας τις δαπανηρές μεθόδους εντοπισμού τους με διαδοχικές δοκιμές κανόνων του συνόλου Π και 2. ισχύει και στις σ.μ. γραμματικές η Ιεραρχία τροποποιημένης εκδοχής Γραμματικών Chomsky.

Αν οι μ. γραμματικές παράγουν την ενδιάμεση γλώσσα των συντακτικών κατηγοριών ή γλώσσα βάση της ΝΕΓ, η επέκτασή τους, δηλαδή οι σ.μ. γραμματικές παράγουν κειμενοποιημένες τις συνήθειες και εύχρηστες προτάσεις της, δηλαδή συνήθειες και εύχρηστες συντακτικές δομές σωμάτων κειμένων της ΝΕΓ.

Με τον τρόπο αυτόν: 1. γίνεται η προσέγγιση από πάνω προς τα κάτω η μελέτη της δομής των σωμάτων κειμένων σε αντίθεση με την Ετικετοποίηση που αποτελεί προσέγγιση από κάτω προς τα επάνω και 2. καλύπτεται η γενετική διάσταση της Γραμματικής με την παραγωγή των συντακτικών δομών βάθους προτάσεων και κειμένων της ΝΕΓ.

Σε κάθε συντακτική δομή της ΝΕΓ ή συντακτική δομή βάθους (κειμενοποιημένη ή μη) αντιστοιχεί ένας μεγάλος αριθμός ποικίλων αντίστοιχων δομών επιφανείας (κειμενοποιημένων ή μη) αποδεκτών προτάσεων της ΝΕΓ. Η αντιστοίχιση αυτή ή μετασχηματιστική διάσταση υλοποιείται με τη βοήθεια του Βασικού Νεοελληνικού Πολυλεξικού (ΒΝΠ) και των αλγορίθμων που το συνοδεύουν. Δηλαδή, η σύνδεση των συνταγματικών αξόνων με τη μορφολογία και τη σημασιολογία στην πραγματικότητα γίνεται με ένα σύστημα Ηλεκτρονικών Υπολογιστικών Λεξικών της ΝΕΓ και των αλγορίθμων που τα συνοδεύουν. Με τον τρόπο αυτόν στους κανόνες της συντακτικής δομής των μ. γραμματικών και κατ' επέκταση των σ.μ. γραμματικών ενσωματώνονται και όλες οι άλλες αλληλοεξαρτώμενες γλωσσολογικές πληροφορίες. Το ΒΝΠ αποτελείται από:

1. Το Βασικό Νεοελληνικό Λεξικό (ΒΝΛ) περιέχει τις λέξεις, σε αλφαβητική σειρά⁴ τις οποίες το προτεινόμενο σύστημα αναγνωρίζει και επεξεργάζεται. Στα υπάρχοντα λήμματα είναι καταχωρημένα αντιπροσωπευτικά δείγματα όλων των κλιτών και άκλιτων μερών του λόγου. Το ΒΝΛ εύκολα συντηρείται και μπορεί να επεκταθεί στον χειρισμό 60.000 και πλέον λημμάτων.

2. Το Νεοελληνικό Μορφολογικό Υπολογιστικό Λεξικό (ΝΜορφΥπΛ), το οποίο περιέχει τα λήμματα του ΒΝΛ και ελέγχει όλα τα είδη της μορφολογικής πληροφορίας, μπορεί: 1. να παράγει και να αναγνωρίζει⁵ με ακρίβεια 100%, τα άκλιτα μέρη του λόγου, τα κλιτά μέρη του λόγου και τους παραγόμενους τύπους τους, λαμβάνοντας υπόψη ότι κάθε ουσιαστικό έχει 8 διακριτούς κλιτούς τύπους, κάθε

⁴ περίπου 7.000 λήμματα, μεταξύ των οποίων τουλάχιστον 200 πολύσημες λέξεις.

⁵ να παράγει και να αναγνωρίζει: 3000 ουσιαστικά, 1600 επίθετα, 2100 ρήματα και τις αντίστοιχες μετοχές τους, 50 αντωνυμίες και 250 μη κλιτές λέξεις, καθώς και να επεκταθεί στον χειρισμό 60.000 και πλέον λημμάτων.

επίθετο έχει 120, κάθε ρήμα έχει 76 κ.λπ. 2. να χειριστεί περιφραστικούς αλλά και πολλαπλούς τύπους και να ελέγχει μορφολογικές αμφισημίες, έτσι ώστε να αποδίδει με σαφήνεια και ακρίβεια τις οποιεσδήποτε πληροφορίες και τέλος 3. είναι εύκολα επεκτάσιμο και ευκολοσυντήρητο.

3. Το Νεοελληνικό Συντακτικό Υπολογιστικό Λεξικό (ΝΣυνΥπΛ), το οποίο ελέγχει τις συμφωνίες των μορφολογικών τύπων των λέξεων της πρότασης ανάλογα με τον συντακτικό ρόλο τους στην πρόταση, τέλος,

4. Το Νεοελληνικό Σημασιολογικό Υπολογιστικό Λεξικό (ΝΣημΥπΛ) αποτελείται από το Λογικό Υπολογιστικό Λεξικό των Βασικών Εννοιών (ΛΥπΛΒΕ) και το Υπολογιστικό Λεξικό των Συμφραζομένων της ΝΕΓ (ΛΥπΛΣΝΓ).

Το ΛΥπΛΒΕ περιέχει αναταξινομημένες τις λέξεις του ΒΝΛ με προδιαγραφές «Λογικού Λεξικού». Ενώ το ΛΥπΛΣΝΓ ομαδοποιεί τις λέξεις που ανήκουν στο ίδιο μέρος του λόγου και που μπορούν να αντικαταστήσουν η μία την άλλη σε συνταγματικές σχέσεις. Έτσι το ΛΥπΛΒΕ με τους αλγόριθμους μπορεί για κάθε λέξη που περιέχει να δώσει τα συνώνυμα και τα αντώνυμά της, εάν υπάρχουν, καθώς και τις διαφορετικές έννοιες εάν οι λέξεις αυτές είναι πολύσημες. Επίσης, μπορεί να αναγνωρίσει ή να υποδείξει αποδεκτούς ή μη σημασιολογικούς συνδυασμούς με άλλες λέξεις στα πλαίσια συγκεκριμένων συντακτικών ρόλων που οι λέξεις αυτές μπορούν να έχουν.

Σημειώνουμε ότι οι κωδικοποιήσεις των απαιτούμενων για επεξεργασία πολυδιάστατων γλωσσολογικών πληροφοριών ή διανυσμάτων όπως έχουν οριστεί και αρχειοθετηθεί μπορούν να αξιοποιηθούν και επιμέρους πολλαπλά.

1.2.7 Βιβλιογραφικές Αναφορές

1. Adetuyi, C. & Fidelis, O. O. (2015). Key Concepts in Transformational Generative Grammar. *International Journal of English Language and Communication Studies*, 1(8).
2. Aho, A.V., Sethi, R. and Ullman, J.D. (1986). *Compilers: principles, techniques, and tools*. Mass: Addison-Wesley Longman Publishing Co.
3. Austin, P. K. (2003). Lexical functional grammars. In: Smelser, N. J. & Baltes, P., (Eds.), *International Encyclopedia of the Social and Behavioral Sciences* (pp. 8748-8754). London: Pergamon.
4. Baldzis S.D., Eumeridou E. & Kolalas S. (2007). Efficient NLP Tools for the Generation and Recognition of Greek word Dominant Semantic Combinations. In: *Conference Proceedings II, 3rd Balkan Conference in Informatics: Research in Informatics and Information Technology*, 3-16.
5. Baldzis, S.D., Kolalas S. & Eumeridou E. (2005a). The Computational Modern Greek Morphological Lexicon - An Efficient and Comprehensive System for Morphological Analysis and Synthesis. *Literary and Linguistic Computing*, 19(4), 1-35.
6. Baldzis S.D., Kolalas S. & Eumeridou E. (2005b). Computational Morphological Lexicons: Current tools for Knowledge and Communication Management. In: *Conference Proceedings, 2nd Balkan Conference in Informatics*. Ohrid, 346 - 356.
7. Baldzis, S.D., Savranidis, Ch. & Kolalas S. (2001a). Written Modern Greek Sentences Corrector and Content Analyser based on Intelligent Computer Methods. In: *Proceedings of 1st Panhellenic Conference with International Participation on Human – Computer Interaction*. Patras: University of Patras, Advances in Human-Computer Interaction I, 182-191.
8. Baldzis, S.D., Eumeridou E. & Kolalas S. (2001b). A Complete and Comprehensive System for Modern Greek Language Processing Proposed as a Modern Greek Language Call Method Developer. *Literary and Linguistic Computing*, 17(4), 373-400.
9. Baldzis, S.D. (1999). *The Modern Greek Language Semantic Basis for Processing*. In *Transscientific Semiotics: III-V*, vol.11-1-3: 9-22.

10. Baldzis, S. D. (1998). *Generative and Transformational Grammars in Modern Greek Language Processing*, Technical Report No 267(b). Ioannina: Dept of Maths, University of Ioannina.
11. Barthlemy, F., Cnam, C., Boullier, P. & Kaouane, L. (2003). Tools and resources for Tree Adjoining Grammars. *Proceedings of the ACL 2001 Workshop for Sharing Tools and Resources*. Doi: <https://www.aclweb.org/anthology/W01-1509>
12. Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory* (2), 113–124.
13. Chomsky, N. (2002). *Syntactic Structures*. Berlin-New York: Mouton de Gruyter.
14. DeRose, S. J. (1988). Grammatical Category Disambiguation by Statistical Optimization. *Computational Linguistics*, 14(1), pp. 31-39.
15. Hopcroft, J. E., Motwani, R. & Ullman, J.D. (2001). *Introduction to Automata Theory, Languages, and Computation*. USA: Addison Wesley.
16. Jäger, G. & Rogers, J. (2012). Formal language theory: Refining the Chomsky hierarchy. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1598), pp. 1956-1970.
17. Jurafsky, D. & Martin, J.H. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. New Jersey: Prentice Hall.
18. Jurafsky, D. (1996). A Probabilistic Model of Lexical and Syntactic Access and Disambiguation. *Cognitive Science*, 20(2), 137-194.
19. Petrov, S., Das, D. & McDonald, R. (2011). A Universal Part-of-Speech Tagset. *Computing Research Repository – CORR*, pp. 2089-2096.
20. Shieber, S. M. (1986). *An Introduction to Unification-Based Approaches to Grammar*. Mass: Microtome Publishing.
21. Wanner, E. & Maratsos, M. (1978). An ATN Approach to Comprehension. In: Halle, M., Bresnan, J. & Miller, G. A. (Eds.), *Linguistic Theory and Psychological Reality* (pp. 119-161). Cambridge: M.I.T. Press.
22. Woods, W. A. (1969). *Augmented Transition Networks for Natural Language Analysis*. Rep. No. CS-1. Aiken Computation Laboratory, Harvard University, Cambridge, Massachussets.

23. Μπαλτζής, Δ. Σ. (2018). Υπολογιστικά Λεξικά της Νεοελληνικής το μέσο για τη συνέχεια της γλώσσας στη σύγχρονη παγκοσμιοποιημένη ηλεκτρονική εποχή. Υπό δημοσίευση στα Πρακτικά ΣΤ' Ευρωπαϊκού Συνεδρίου Νεοελληνικών Σπουδών της ΕΕΝΣ: *Ο ελληνικός κόσμος σε περιόδους κρίσης και ανάκαμψης, 1204-2018*, Lund.
24. Μπαλτζής, Δ. Σ., Βυζάς Θ. & Ευμοιρίδου Ε. (2017). *Εναλλακτική αξιοποίηση της έννοιας της οντολογίας στη σημασιολογική κωδικοποίηση μίας φυσικής γλώσσας για αυτόματη επεξεργασία της - Παραδείγματα από τη Νέα Ελληνική Γλώσσα*. It will appear in Conference Proceedings of the 6th International Conference on Modern Greek Language (ICGL 13), University of Westminster- London.
25. Μπαλτζής, Δ. Σ., Βυζάς, Θ. & Ευμοιρίδου Ε. (2014). Η σημασιολογική συμπεριφορά του ρήματος ως καθοριστικός παράγοντας στην κατασκευή των σημασιολογικών κανόνων της νεοελληνικής για την αυτόματη επεξεργασία της. Πρακτικά Ε' Ευρωπαϊκού Συνεδρίου Νεοελληνικών Σπουδών της ΕΕΝΣ: *Συνέχειες, ασυνέχειες, ρήξεις στον ελληνικό κόσμο (1204-2014): οικονομία, κοινωνία, ιστορία, λογοτεχνία*. Θεσσαλονίκη: Τόμος Δ' 611-650,
http://www.eens.org/EENS_congresses/2014/books/tomo4.pdf.
26. Μπαλτζής, Δ. Σ., Μάντζιου Ό. & Αποστόλου Κ. (2010). Μια επέκταση του καινοτόμου Ηλεκτρονικού Υπολογιστικού Λεξικού της Νέας Ελληνικής Γλώσσας - ΒΝΠ ως Διαδικτυακό Πολυμεσικό Εκπαιδευτικό Λογισμικό Φωνολογικής Ενημερότητας πρώτης σχολικής ηλικίας. Πρακτικά Δ' Ευρωπαϊκού Συνεδρίου ΝΣ της ΕΕΝΣ: *Ταυτότητες στον ελληνικό κόσμο (από το 1204 έως σήμερα)*. Γρανάδα: Τόμος Ε' 649-888,
http://www.eens.org/wordpress/wp-content/uploads/2012/05/Identities-in-the-Greek-world-Granada-2010-Congress-Vol_5-2011-isbn_978-960-99699-7-01.pdf.
27. Μπαλτζής, Δ. Σ., Κολαλάς Σ., Ευμοιρίδου Ε., Αλεξιάκης Α. & Δούκα Δ. (2006). Ένα Καινοτόμο Ηλεκτρονικό Λεξικό της Νέας Ελληνικής Γλώσσας - Πρώτο μέρος: Μορφολογικό. Πρακτικά Γ' Ευρωπαϊκού Συνεδρίου ΝΣ της ΕΕΝΣ: *Ο ελληνικός κόσμος ανάμεσα στην εποχή του Διαφωτισμού και στον εικοστό αιώνα*. Βουκουρέστι: Τόμος Β' 341-354.

28. Μπαλτζής, Δ. Σ. (1997). *Σπονδυλωτές Μητροειδείς Γραμματικές και Αυτόματη Επεξεργασία της Νεοελληνικής Γλώσσας*. Πρακτικά του 6ου Πανελληνίου Συνεδρίου Πληροφορικής. Αθήνα (ΕΠΥ): Τόμος I, 213-227.
29. Μπαλτζής, Δ. Σ. (1995). *Μητροειδείς Γραμματικές και Αυτόματη Επεξεργασία της Νεοελληνικής*. Πρακτικά του 5ου Πανελληνίου Συνεδρίου Πληροφορικής. Αθήνα (ΕΠΥ): Τόμος II, 699-712.

1.3 Εκμάθηση Μηχανής

1.3.1 Εισαγωγή

Η Εκμάθηση Μηχανής (MM) είναι η επιστημονική μελέτη των αλγορίθμων και στατιστικών μοντέλων που χρησιμοποιούν τα Υπολογιστικά Συστήματα για την αποτελεσματική εκτέλεση μιας συγκεκριμένης και προκαθορισμένης εργασίας χωρίς να χρησιμοποιούν ρητές οδηγίες από τον προγραμματιστή. Έτσι, οι αλγόριθμοι MM κατασκευάζουν ένα μαθηματικό μοντέλο που βασίζεται σε κάποια δεδομένα εκπαίδευσης προκειμένου να προβεί σε προβλέψεις ή αποφάσεις χωρίς να έχει προγραμματιστεί αυστηρά. Οι αλγόριθμοι MM χρησιμοποιούνται σε μια ευρεία ποικιλία εφαρμογών όπου μεταξύ άλλων έχουν καταξιωθεί και αναγνωριστεί στον τομέα των ΝΔ τα οποία θα περιγραφούν αναλυτικά παρακάτω.

Λέξεις Κλειδιά: Εκμάθηση Μηχανής, Τεχνητή Νοημοσύνη, Συμβατικός Προγραμματισμός, Επιτηρούμενη Εκμάθηση, Μη Επιτηρούμενη Εκμάθηση.

1.3.2 Βασικές Αρχές, Έννοιες και Ορισμοί

Η MM αποτελεί έναν σημαντικό υποκλάδο της Τεχνητής Νοημοσύνης (TN), ο οποίος έχει γνωρίσει μια τεράστια άνθηση τα τελευταία χρόνια. Η MM σχεδιάστηκε ώστε να έχει την ικανότητα να μαθαίνει μέσα από παραδείγματα και να αυτοβελτιώνεται από αυτά χωρίς να χρειάζεται κάποια ρητή κωδικοποίηση από τον προγραμματιστή.

Αναλυτικότερα η MM συνδυάζει τα δεδομένα με στατιστικά εργαλεία προκειμένου να προβλέψει το αποτέλεσμα ως έξοδο. Αυτό το αποτέλεσμα χρησιμοποιείται στη συνέχεια για να παράξει εσωτερικά πληροφορίες/δεδομένα. Στη συνέχεια η μηχανή δέχεται τα δεδομένα αυτά στην είσοδο της και χρησιμοποιεί κάποιον αλγόριθμο για τη διατύπωση των απαντήσεων που θα δώσει. Ένα κλασσικό παράδειγμα της λειτουργίας της MM είναι η παροχή συστάσεων/συμβουλών σε όσους έχουν λογαριασμό στην διαδικτυακή υπηρεσία παροχής ταινιών/σειρών Netflix. Έτσι, μέσω της υπηρεσίας αυτής, όλες οι συστάσεις/συμβουλές ταινιών/σειρών που διατίθενται, ανά τακτά χρονικά διαστήματα, βασίζονται στα ιστορικά δεδομένα που αποθηκεύονται με βάση την παρακολούθηση του χρήστη. Οι διάφορες εταιρίες τεχνολογίας χρησιμοποιούν τη MM για να βελτιώσουν την εμπειρία των χρηστών μέσω της εξατομίκευσης των συστάσεων/συμβουλών.

Η MM βρίσκει πλέον εφαρμογή σε διάφορα επιστημονικά πεδία, και όχι μόνο, με

απαιτήσεις υπολογιστικών διεργασιών, των οποίων ο σχεδιασμός και προγραμματισμός των αντίστοιχων αλγορίθμων τους με ικανοποιητική απόδοση είναι δύσκολος ή ακόμη και αδύνατος. Κάποια χαρακτηριστικά παραδείγματα τέτοιων εφαρμογών είναι το φιλτράρισμα του ηλεκτρονικού ταχυδρομείου, η ανίχνευση επιβλαβών εισβολών σε δίκτυα, η οπτική αναγνώριση χαρακτήρων, οι εφαρμογές στην υπολογιστική όραση, η προγραμματισμένη συντήρηση των ιστοσελίδων κ.λπ..

1.3.3 Η λειτουργία, ο ρόλος της Εκμάθησης Μηχανής και οι διαφορές της από τον Συμβατικό Προγραμματισμό

Η MM προσομοιάζει με τον τρόπο που ο ανθρώπινος εγκέφαλος μαθαίνει. Είναι γνωστό ότι οι άνθρωποι μαθαίνουν από την εμπειρία τους, δηλαδή, όσο πιο πολλά γνωρίζουν τόσο πιο εύκολα αναγνωρίζουν/προβλέπουν. Έτσι, όταν αντιμετωπίζουν μια άγνωστη κατάσταση, η πιθανότητα να προβλέψουν με επιτυχία οτιδήποτε αφορά αυτήν είναι μικρότερη, από ότι αν η κατάσταση ήταν γνωστή. Οι μηχανές εκπαιδεύονται με τον ίδιο τρόπο, δηλαδή, για να κάνουν μια ακριβή πρόβλεψη, έχει προηγηθεί η επεξεργασία/Εκμάθηση ενός παραδείγματος. Στη συνέχεια δίνοντας στο μηχανήμα ένα παρόμοιο παράδειγμα, αυτό είναι σε θέση να προβλέψει το αποτέλεσμα. Εάν όμως η μηχανή τροφοδοτηθεί με ένα άγνωστο παράδειγμα, τότε θα αντιμετωπίσει κάποιες δυσκολίες στην πρόβλεψη του αναμενόμενου αποτελέσματος.

Έτσι οι βασικοί στόχοι της MM είναι αφενός η εκμάθηση, αφετέρου η εξαγωγή συμπεράσματος. Αρχικά, η μηχανή μαθαίνει μέσα από την ανακάλυψη προτύπων, που πραγματοποιείται χάρη στα δεδομένα. Κατά συνέπεια είναι απαραίτητο να επιλέγονται και να τροφοδοτούνται προσεκτικά τα κατάλληλα δεδομένα στο μηχανήμα. Ο κατάλογος των χαρακτηριστικών που χρησιμοποιούνται για την επίλυση ενός προβλήματος ονομάζεται διάνυμα χαρακτηριστικών, το οποίο είναι ένα υποσύνολο δεδομένων που χρησιμοποιείται για την αντιμετώπιση ενός προβλήματος. Είναι αξιοσημείωτο επίσης, το γεγονός ότι η μηχανή χρησιμοποιεί κάποιους εξειδικευμένους αλγορίθμους για να απλοποιήσει την πραγματικότητα και να μετατρέψει την ανακάλυψη που έχει κάνει σε κάποιο μοντέλο. Επομένως, το στάδιο εκμάθησης χρησιμοποιείται για να περιγράψει η μηχανή τα δεδομένα και να τα συνοψίσει σε ένα μοντέλο. Στο σχήμα που ακολουθεί απεικονίζεται παραστατικά η φάση της εκμάθησης (βλ. Εικόνα 4).

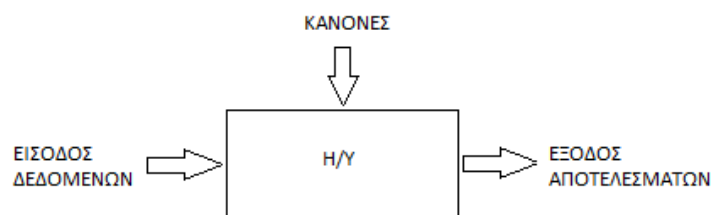
Ένα αντιπροσωπευτικό παράδειγμα για το πως δουλεύει η φάση της εκμάθησης είναι η μηχανή που προσπαθεί να κατανοήσει τη σχέση μεταξύ του μισθού ενός

ατόμου και της πιθανότητας το άτομο αυτό να επισκεφθεί κάποιο καλό και ακριβό εστιατόριο. Αποδεικνύεται ότι, η μηχανή τελικά θα βρει μια θετική σχέση μεταξύ του μισθού και του εστιατορίου η οποία θα είναι και το ζητούμενο μοντέλο.



Εικόνα 4: Η φάση της εκμάθησης

Ο Συμβατικός Προγραμματισμός (ΣΠ) διαφέρει σημαντικά από τη ΜΜ. Στον ΣΠ ο προγραμματιστής κωδικοποιεί όλους τους κανόνες σε συνεννόηση με έναν ειδικό του κλάδου που το συγκεκριμένο λογισμικό αναπτύσσεται. Κάθε κανόνας για τον οποίο το μηχανήμα θα εκτελεί μια έξοδο ακολουθώντας τη λογική δήλωση βασίζεται σε αντίστοιχη θεωρία του υπό μελέτη κλάδου. Όταν όμως το σύστημα γίνει πολυπλοκότερο, επιβάλλεται να γραφούν περισσότεροι κανόνες, με πιθανό αποτέλεσμα το σύστημα να γίνει μη βιώσιμο και να μην μπορεί πλέον να διατηρηθεί (βλ. Εικόνα 5).



Εικόνα 5: Γράφημα Συμβατικού Προγραμματισμού



Εικόνα 6: Γράφημα Εκμάθησης Μηχανής

Αντιθέτως η ΜΜ «είναι σε θέση» να ξεπεράσει αυτά τα ζητήματα που δημιουργούνται στον ΣΠ. Η μηχανή μαθαίνει να συσχετίζει τα δεδομένα εισόδου και εξόδου και να διατυπώνει έναν αντίστοιχο κανόνα κάθε φορά. Οι προγραμματιστές δεν χρειάζεται να γράφουν νέους κανόνες όταν προκύπτουν νέα δεδομένα, καθώς οι αλγόριθμοι αυτοπροσαρμόζονται στα νέα δεδομένα και στις νέες εμπειρίες, με αποτέλεσμα να βελτιώνεται η αποτελεσματικότητα της μηχανής με την πάροδο του

χρόνου (βλ. Εικόνα 6).

Από τη στιγμή που το μοντέλο είναι έτοιμο δοκιμάζεται με νέα δεδομένα για να επιβεβαιωθεί η ισχύς του. Κατά τη δοκιμασία τα νέα δεδομένα μετασχηματίζονται σε διάλυμα χαρακτηριστικών και οδηγούν στην πρόβλεψη. Σημειώνουμε το μεγάλο πλεονέκτημα της MM είναι το γεγονός ότι για τη εξαγωγή συμπερασμάτων για τα νέα δεδομένα χρησιμοποιείται αυτό καθαυτό το ήδη εκπαιδευμένο μοντέλο, χωρίς να χρειάζεται περαιτέρω ενημέρωση των κανόνων του ή η επανεκπαίδευση του (βλ. Εικόνα 7).



Εικόνα 7: Η φάση της δοκιμής

1.3.4 Ταξινόμηση των Αλγορίθμων Εκμάθησης Μηχανής, εφαρμογές και παραδείγματα

Θα ήταν παράλειψη να μην αναφερθεί η ταξινόμηση της MM ανάλογα με τους τύπους των αλγορίθμων της. Έτσι, οι τύποι αυτοί των αλγορίθμων ταξινομούν τη MM σε Επιτηρούμενη Εκμάθηση Μηχανής και σε Μη Επιτηρούμενη Εκμάθηση Μηχανής, οι οποίες και αυτές ανάλογα υποδιαιρούνται περαιτέρω. Αναλυτικότερα:

1.3.4.1 Επιτηρούμενη Εκμάθηση (EMM): είναι ο αλγόριθμος που χρησιμοποιεί τα δεδομένα εκπαίδευσης και πληροφορίες που τροφοδοτούνται από τον ανθρώπινο παράγοντα, προκειμένου ο αλγόριθμος να μάθει τη σχέση που συνδέει τα δεδομένα εισόδου με τα αποτελέσματα εξόδου. Για παράδειγμα, ένας επαγγελματίας μπορεί να αξιοποιήσει τις πληροφορίες που προκύπτουν από τα έξοδα του μάρκετινγκ και την πρόγνωση του καιρού ως δεδομένα εισόδου, για να προβλέψει τις πωλήσεις που θα κάνει από τις κονσέρβες που πουλάει η εταιρεία του.

Η EMM μπορεί να χρησιμοποιηθεί όταν είναι γνωστά τα αποτελέσματα εξόδου και δίνεται η δυνατότητα στον αλγόριθμο να προβλέψει νέα δεδομένα. Υπάρχουν δύο βασικές κατηγορίες της EMM, αυτή της Ταξινόμησης και αυτή της Παλινδρόμησης.

1.3.4.1.1 Ταξινόμηση (EMMT): Για παράδειγμα, έστω ότι θέλουμε να προβλέψουμε το φύλο ενός πελάτη για μια εμπορική διαφήμιση. Η πρόβλεψη αυτή θα ξεκινήσει από τη συλλογή δεδομένων της βάσης δεδομένων των πελατών,

σχετικά με το ύψος, το βάρος, την εργασία, τον μισθό, το καλάθι αγορών του κ.λπ.. Ο σκοπός του ταξινομητή είναι να συμπληρώσει την ετικέτα του γένους του πελάτη, αν δηλαδή είναι αρσενικού ή θηλυκού γένους, βάσει των πληροφοριών - χαρακτηριστικών που έχουν συλλεχτεί και διατίθενται. Όταν το μοντέλο έχει πλέον μάθει πώς να αναγνωρίζει το γένος, μπορούν να χρησιμοποιηθούν νέα δεδομένα για να κάνει μια νέα πρόβλεψη.

1.3.4.1.2 Παλινδρόμηση (EMMΠ): είναι ο αλγόριθμος που απαιτείται όταν η έξοδος είναι μια συνεχής τιμή. Για παράδειγμα, ένας οικονομικός αναλυτής μπορεί να χρειαστεί να προβλέψει την αξία αποθέματος με ένα εύρος χαρακτηριστικών όπως: τις μετοχές, τις προηγούμενες αποδόσεις των αποθεμάτων, τον μακροοικονομικό δείκτη. Επομένως, το μοντέλο θα εκπαιδευτεί για την εκτίμηση της τιμής των αποθεμάτων με το μικρότερο δυνατό λάθος. Η EMMΠ περιλαμβάνει πολλούς αλγορίθμους για την επιτυχή πρόβλεψη και οι οποίοι:

1.3.4.1.2.1 Γραμμική Παλινδρόμηση (EMMΓΠ): είναι EMM αλγόριθμος με στόχο να βρίσκει τρόπο συσχετισμού κάθε χαρακτηριστικού με την έξοδο ώστε να βοηθήσει την πρόβλεψη των μελλοντικών τιμών.

1.3.4.1.2.2 Λογιστική Παλινδρόμηση (EMMΛΠ): είναι αλγόριθμος προέκτασης της EMMΓΠ ο οποίος μπορεί να χρησιμοποιηθεί για ταξινομήσεις. Η μεταβλητή εξόδου θα είναι δυαδική (π.χ. μόνο μαύρο ή λευκό) και όχι συνεχής (π.χ. μια άπειρη λίστα πιθανών χρωμάτων).

1.3.4.1.2.3 Δέντρο Απόφασης (EMMΔΑ): είναι αλγόριθμος τόσο μοντέλου EMMT όσο και EMMΠ που διασπά τα δεδομένα που αντιστοιχούν σε τιμές χαρακτηριστικών σε κλάδους στους κόμβους απόφασης (π.χ. εάν ένα χαρακτηριστικό είναι χρώμα, τότε κάθε πιθανό χρώμα μετατρέπεται σε ένα νέο κλάδο) μέχρι να εξαχθεί μία τελική απόφαση.

1.3.4.1.2.4 Naive Bayes: η Μπαεζιανή Μέθοδος είναι αφενός τόσο EMMT όσο και EMMΠ αλγόριθμος αφετέρου μία μέθοδος ταξινόμησης που χρησιμοποιεί το Μπαεζιανό θεώρημα. Το θεώρημα ενημερώνει την προηγούμενη γνώση ενός γεγονότος με την

ανεξάρτητη πιθανότητα κάθε χαρακτηριστικού που μπορεί να επηρεάσει το γεγονός.

1.3.4.1.2.5 **Μηχανές Διανυσμάτων Υποστήριξης:** είναι αλγόριθμοι EMMPI (όχι τόσο κοινοί) και EMMT. Τυπικά χρησιμοποιείται για ταξινομήσεις βρίσκοντας ένα υπερεπίπεδο που διαιρεί κατά βέλτιστο τρόπο τις κλάσεις. Ο αλγόριθμος ενδείκνυται για μη γραμμικές επιλύσεις.

1.3.4.1.2.6 **Δάσος με Τυχαιότητα:** είναι EMMT και EMMPI αλγόριθμοι που βασίζονται σε ένα Δέντρο Απόφασης για να βελτιώσει δραστικά την ακρίβεια. Το Δάσος με Τυχαιότητα παράγει πολλαπλά απλά Δέντρα Απόφασης και χρησιμοποιεί την «πλειοψηφική μέθοδο» για να αποφασίσει σε ποιά ετικέτα θα επιστρέψει. Για Ταξινομήσεις, η τελική πρόβλεψη είναι αυτή που πλειοψήφησε, ενώ για Παλινδρομήσεις, η τελική πρόβλεψη είναι η μέση πρόβλεψη όλων των δέντρων.

1.3.4.1.2.7 **AdaBoost:** είναι EMMPI και EMMT αλγόριθμος. Η μέθοδος αυτή αποτελεί μια τεχνική Ταξινόμησης/Παλινδρόμησης η οποία για να αποφασίσει χρησιμοποιεί μια πληθώρα μοντέλων, αξιολογώντας τα βάσει της ακρίβειάς τους στην πρόβλεψη του αποτελέσματος.

1.3.4.1.2.8 **Δέντρα που ενισχύουν τη διαβάθμιση:** είναι EMMT και EMMPI αλγόριθμοι. Η τεχνική αυτή τελευταίας τεχνολογίας επικεντρώνεται στο σφάλμα που προκύπτει από τα προηγούμενα δέντρα και προσπαθεί να το διορθώσει.

1.3.2.1.3 Μη Επιτηρούμενη Εκμάθηση

Στη Μη Επιτηρούμενη Εκμάθηση (MEMM), ο αλγόριθμος διερευνά τα δεδομένα εισόδου χωρίς να του δοθεί μια ρητή μεταβλητή εξόδου (π.χ. διερευνά τα δημογραφικά δεδομένα του πελάτη για να ταυτοποιήσει τα πρότυπα). Η MEMM μπορεί να χρησιμοποιηθεί όταν δεν γνωρίζει ο χρήστης πως να ταξινομήσει τα δεδομένα και θέλει να εντοπίσει ο αλγόριθμος τα πρότυπα και να ταξινομήσει τα δεδομένα. Η MEMM περιλαμβάνει πολλούς αλγορίθμους για την επιτυχή πρόβλεψη της, οι οποίοι περιγράφονται παρακάτω:

- 1.3.2.1.3.1 **K-Ομαδοποίηση:** είναι αλγόριθμος MEMM που ομαδοποιεί τα δεδομένα σε K-ομάδες, όπου σε κάθε ομάδα περιέχονται δεδομένα με παρόμοια χαρακτηριστικά τα οποία καθορίζονται από το μοντέλο και όχι εκ των προτέρων από τον άνθρωπο.
- 1.3.2.1.3.2 **Μοντέλο μίγματος του Γκάους:** είναι MEMM αλγόριθμος που είναι γενίκευση του αλγορίθμου K-Ομαδοποίησης, ο οποίος παρέχει μεγαλύτερη ευελιξία στο μέγεθος και το σχήμα των ομάδων που περιγράφηκαν στην προηγούμενη μέθοδο.
- 1.3.2.1.3.3 **Ιεραρχική Ομαδοποίηση:** είναι αλγόριθμος MEMM που διασπά τις ομάδες κατά μήκος ενός ιεραρχικού δέντρου για να σχηματιστεί ένα σύστημα ταξινόμησης.
- 1.3.2.1.3.4 **Προτεινόμενο Σύστημα:** είναι αλγόριθμος MEMM που βοηθά στο καθορισμό των σχετικών δεδομένων για να γίνει μια πρόταση.
- 1.3.2.1.3.5 **PCA/T-SNE:** είναι αλγόριθμος MEMM που χρησιμοποιείται κυρίως για τη μείωση της διάστασης των δεδομένων. Οι αλγόριθμοι που μειώνουν τον αριθμό των χαρακτηριστικών σε 3 ή 4 διανύσματα με τις υψηλότερες διακυμάνσεις.

1.3.5 Παράδειγμα Σύγκρισης Συμβατικού Προγραμματισμού και Εκμάθησης Μηχανής

Το παράδειγμα που παραθέτομε αφορά στο πρόβλημα το οποίο λαμβάνει ως είσοδο έναν αριθμό και προσπαθεί να τον διαιρέσει με το 3 και το 5. Εάν ο αριθμός διαιρείται με το 3, τότε θα εκτυπώσει τη λέξη «fizz», εάν διαιρείται με το 5, τότε θα εκτυπώσει τη λέξη «buzz» και εάν διαιρείται και με τους δύο αυτούς αριθμούς, τότε θα εκτυπώσει τη λέξη «fizzbuzz». Αντίθετα, εάν δεν διαιρείται με κανέναν από τους αριθμούς 3 και 5, τότε θα εκτυπώσει τη λέξη «Other». Το παιχνίδι αυτό ονομάζεται Fizzbuzz.

Το παράδειγμα υλοποιείται αφενός με «πρόγραμμα ΣΠ» και αφετέρου με «πρόγραμμα ΜΜ», με στόχο να φανούν και να υπογραμμιστούν οι διαφορές μεταξύ αυτών των δύο προγραμματιστικών προσεγγίσεων. Το παράδειγμα το ονομάζομε «Κώδικα Other».

Αρχικά, η προσέγγιση του ΣΠ είναι η τροφοδότηση του υπολογιστή με ένα σύνολο δεδομένων για ένα καθορισμένο σύνολο σεναρίων. Μετά από αυτή τη διαδικασία ο υπολογιστής θα αξιοποιήσει τις δυνατότητές του προκειμένου να βοηθήσει τον

άνθρωπο για την επεξεργασία των δεδομένων με πιο γρήγορο και αποτελεσματικό τρόπο. Όσον αφορά τη MM, ένα τεράστιος όγκος δεδομένων εισέρχεται στον υπολογιστή, ο οποίος με τη σειρά του επεξεργάζεται όλα αυτά τα δεδομένα και στη συνέχεια περνά στη διαδικασία Εκμάθησης.

1.3.5.1 1^η προσέγγιση με κώδικα ΣΠ

Είναι εξαιρετικά εύκολο για τον ΣΠ να τροφοδοτεί τον υπολογιστή με ένα σύνολο οδηγιών και αυτό διότι στο συγκεκριμένο παράδειγμα έχουμε μόνο 4 σενάρια που πρέπει να επαληθευτούν και να εκτυπώσουν την έξοδο με βάση αυτό. Ο κώδικας Python μπορεί να γραφεί όπως φαίνεται παρακάτω:

➤ Συνθήκη κώδικα

```
def fizzbuzz(n): # if the number is divisible by 3 as well as by 5 and returns #
"FizzBuzz"
    if n % 3 == 0 and n % 5 == 0:
        return 'FizzBuzz'# If the first condition is not satisfied then it checks if it is #
divisible by 3 and return "Fizz"
    elif n % 3 == 0:
        return 'Fizz'# If both of the above tests are not satisfying, then it will check #
whether it is divisible by 5 and return "Buzz"
    elif n % 5 == 0:
        return 'Buzz'# If all the conditions above do not satisfy then it returns "Other"
    else:
        return 'Other'
```

1.3.5.2 2^η προσέγγιση με κώδικα MM

Για τη MM, υποθέτουμε ότι έχουμε ήδη πολλούς αριθμούς των οποίων τα αποτελέσματα είναι ήδη γνωστά, δηλαδή οι λέξεις «fizz» ή «buzz» ή «fizzbuzz». Επομένως, το μόνο που χρειάζεται να κάνουμε τώρα είναι να γράψουμε ένα κώδικα MM και να τροφοδοτήσουμε/εκπαιδεύσουμε τα τρέχοντα δεδομένα. Στη συνέχεια να επαληθεύσουμε εάν έχουμε δημιουργήσει με επιτυχία το μοντέλο ελέγχοντάς το με κάποια άγνωστα δεδομένα. Εάν το μοντέλο δώσει την έξοδο χρησιμοποιώντας το εκπαιδευμένο μοντέλο, χωρίς να υπολογίζει πραγματικά το αποτέλεσμα, τότε ο σκοπός έχει επιτευχθεί.

Για τον κώδικα, θα χρησιμοποιηθεί η βιβλιοθήκη Tensorflow της Google. Παρακάτω

δίνονται μερικά σημαντικά αποσπάσματα του κώδικα:

1.3.5.2.1 Κατασκευή του προτύπου (μοντέλου)

```
# Placeholders are the type of variables nodes where data can be
# fed from outside when we actually run the model
#Placeholder for input data
inputTensor = tf.placeholder(tf.float32, [None, 10])
# Placeholder for output data
outputTensor = tf.placeholder(tf.float32, [None, 4])
# The number of neurons which 1st hidden neurons will have i.e. 1000
NUM_HIDDEN_NEURONS_LAYER_1 = 1000
```

➤ Ρυθμός/Ποσοστό Εκμάθησης

```
# Learning rate, which will be later used to optimize for optimizer function
# Learning rate defines at what rate the optimizer function will move towards
minima per iteration
# Less than optimum learning rate will slow down the process where as higher
learning rate will have chances of
# skipping the minima all together. So it will never converge rather it will keep on
going back and forth.
LEARNING_RATE = 0.05
```

➤ Ορισμός αρχικών τιμών Βαρών

```
# Initializing the weights to Normal Distribution
# The weights will keep on adjusting towards optimum values in each iteration.
Conceptually, weights determine the
# discrimination factor of a particular variable in the neural network. More the
weight, more will be it's
# contribution in determining the solution.
def init_weights(shape):
    return tf.Variable(tf.random_normal(shape,stddev=0.01))

# Initializing the input to hidden layer weights
# We will need a total of 10(input layer number) * 100(number of hidden layers)
input_hidden_weights = init_weights([10,
```

```
NUM_HIDDEN_NEURONS_LAYER_1])
```

```
# Initializing the hidden to output layer weights
```

```
# In this case we will need 100(number of hidden neurons layers) * 4(output  
neurons, we have only 4 categories)
```

```
hidden_output_weights = init_weights([NUM_HIDDEN_NEURONS_LAYER_1,  
4])
```

➤ **Υπολογισμός των τιμών στα κρυφά επίπεδα**

```
# Computing values at the hidden layer
```

```
# Matrix multiplication is done and then rectifier neural network activation  
function is used
```

```
# for regularization of the resulting multiplication
```

```
hidden_layer = tf.nn.relu(tf.matmul(inputTensor, input_hidden_weights))
```

➤ **Υπολογισμός των τιμών στα επίπεδα εξόδου**

```
# Computing values at the output layer
```

```
# Matrix multiplication of hidden layer and output weights it done.
```

```
output_layer = tf.matmul(hidden_layer, hidden_output_weights)
```

➤ **Καθόρισμος της συνάρτησης σφάλματος**

```
# Defining Error Function
```

```
# Error function computes the difference between actual output and model output.
```

```
# Here we are calculating the error in the output as compared to the output label
```

```
error_function =
```

```
tf.reduce_mean(tf.nn.softmax_cross_entropy_with_logits(logits=output_layer,  
labels=outputTensor))
```

➤ **Καθορισμός του αλγορίθμου εκμάθησης και των παραμέτρων εκπαίδευσης**

```
# Defining Learning Algorithm and Training Parameters
```

```
# We are using Gradient Descent function to optimize the error or to reach the  
minima.
```

```
training =
```

```
tf.train.GradientDescentOptimizer(LEARNING_RATE).minimize(error_function)
```

➤ **Συνάρτηση πρόβλεψης**

```
# Prediction Function
prediction = tf.argmax(output_layer, 1)
```

1.3.5.2.2 Εκπαίδευση του προτύπου (μοντέλου)

```
NUM_OF_EPOCHS = 5000
BATCH_SIZE = 128
training_accuracy = []

with tf.Session() as sess:

    # Set Global Variables ?
    # We had only defined the model previously. To run the model, all the variables
    need to be initialized and run.

    # Actual computation can only start after the initialization.
    tf.global_variables_initializer().run()

    for epoch in tqdm_notebook(range(NUM_OF_EPOCHS)):

        #Shuffle the Training Dataset at each epoch
        #Shuffling is done to have even more randomized data, which adds to the
        generalization of model even more.

        p = np.random.permutation(range(len(processedTrainingData)))
        processedTrainingData = processedTrainingData[p]
        processedTrainingLabel = processedTrainingLabel[p]

        # Start batch training
        # With batch size of 128, there will be total of 900/128 runs in each epoch
        where 900 is the total

        # training data.
        for start in range(0, len(processedTrainingData), BATCH_SIZE):
            end = start + BATCH_SIZE
            sess.run(training, feed_dict={inputTensor:
                processedTrainingData[start:end],
                outputTensor: processedTrainingLabel[start:end]})

        # Training accuracy for an epoch
        # We are checking here the accuracy of model after each epoch
```

```

training_accuracy.append(np.mean(np.argmax(processedTrainingLabel,
axis=1) == sess.run(prediction, feed_dict={inputTensor: processedTrainingData,
outputTensor: processedTrainingLabel})))
# Testing
predictedTestLabel = sess.run(prediction, feed_dict={inputTensor:
processedTestingData})

```

1.3.5.2.3 Δοκιμή του προτύπου (μοντέλου)

```

wrong = 0
right = 0

```

```

predictedTestLabelList = []

```

➤ Σύγκριση της προβλεπόμενης τιμής με την πραγματική ετικέτα

```

#Comparing the predicted value with the actual label in training data
for i,j in zip(processedTestingLabel,predictedTestLabel):
    predictedTestLabelList.append(decodeLabel(j))

```

```

if np.argmax(i) == j:

```

```

    right = right + 1

```

```

else:

```

```

    wrong = wrong + 1

```

```

print("Errors: " + str(wrong), " Correct : " + str(right))

```

```

print("Testing Accuracy: " + str(right/(right+wrong)*100))

```

1.3.5.2.4 Αποτελέσματα

```

Errors: 2 Correct :98

```

```

Testing Accuracy: 98.0 %

```

Παρατήρηση: Τα ίδια δεδομένα τροφοδοτήθηκαν στο μοντέλο 5000 φορές, με τυχαία σειρά κάθε φορά. Μετά από κάθε εκτέλεση μετρήθηκε η ακρίβεια του μοντέλου, η οποία συνίσταται στη μέτρηση του ποσοστού του σφάλματος στα δεδομένα δοκιμών. Τελικά, η ακρίβεια του μοντέλου άγγιξε το 98%.

1.3.5.3 Συμπεράσματα

Εν κατακλείδι:

1. Τα βήματα των προγραμμάτων MM είναι απλά και μπορούν να συνοψιστούν ως:
 - 1.1 Ορισμός μιας ερώτησης,
 - 1.2 Συλλογή τα δεδομένων,
 - 1.3 Οπτικοποίηση των δεδομένων,
 - 1.4 Εκπαίδευση του αλγορίθμου,
 - 1.5 Έλεγχος του αλγορίθμου,
 - 1.6 Συλλογή των σχολίων,
 - 1.7 Βελτίωση του αλγορίθμου,
 - 1.8 Επανάληψη των βημάτων 4-7 μέχρι τα αποτελέσματα να είναι ικανοποιητικά/αποδεκτά,
 - 1.9 Επαναχρησιμοποίηση του μοντέλου για τη ζητούμενη πρόβλεψη, τέλος,
 - 1.10 Εφαρμογή της υπάρχουσας γνώσης σε νέα σύνολα δεδομένων.
 2. Η MM αποδεικνύεται το πιο χρήσιμο εργαλείο που υπάρχει μέχρι σήμερα για την ανάλυση, την κατανόηση και την αναγνώριση ενός προτύπου στα δεδομένα.
 3. Μια από τις βασικές ιδέες που διέπουν τη MM είναι ότι ο υπολογιστής μπορεί να εκπαιδευτεί για να αυτοματοποιεί εξαντλητικές διαδικασίες ή πολύ δύσκολες για να εκτελεστούν από έναν άνθρωπο.
 4. Η ξεκάθαρη διαφορά από την συμβατική ανάλυση είναι ότι η MM μπορεί να πάρει αποφάσεις με ελάχιστη ανθρώπινη παρέμβαση.
- ### 1.3.6 Εφαρμογές της Εκμάθησης Μηχανής
- Η MM είναι μια συλλογή αλγορίθμων και μεθόδων με τις οποίες βελτιώνεται σημαντικά η αποδοτικότητα μιας μηχανής, για παράδειγμα ο υπολογιστής, στην εκτέλεση «ευφυών» εργασιών. Κάποιες από τις βασικές κατηγορίες των ευφυών εργασιών που καλείται να εκτελέσει η μηχανή είναι:
- **Ενίσχυση/Βελτίωση/Επαύξηση:** Βοηθά τους ανθρώπους στις καθημερινές τους εργασίες, προσωπικές ή εμπορικές χωρίς να έχει τον πλήρη έλεγχο της παραγωγής. Αυτού του είδους η MM χρησιμοποιείται με διάφορους τρόπους, όπως είναι η Εικονική Βοηθός, η Ανάλυση Δεδομένων, οι Λύσεις Λογισμικών, κ.λπ.. Η κύρια χρήση της είναι να μειώνει τα σφάλματα που προκλήθηκαν από την ανθρώπινη προκατάληψη.

- **Αυτοματισμός:** Η MM λειτουργεί εντελώς αυτόνομα σε οποιονδήποτε τομέα χωρίς την ανάγκη για οποιαδήποτε ανθρώπινη παρέμβαση. Χαρακτηριστικό παράδειγμα είναι τα ρομπότ τα οποία εκτελούν τα βασικά βήματα επεξεργασίας σε εργοστάσια παραγωγής.
- **Χρηματοοικονομική Βιομηχανία:** Η δημοτικότητα της MM ολοένα και αυξάνεται στη χρηματοπιστωτική βιομηχανία. Οι τράπεζες χρησιμοποιούν κυρίως τη MM για να βρουν πρότυπα μέσα στα δεδομένα αλλά και για να αποτρέψουν την όποια απάτη ήθελε προκύψει.
- **Κυβερνητικές Οργανώσεις:** Πολλές κυβερνήσεις ανά τον κόσμο χρησιμοποιούν τη MM για τη διαχείριση της δημόσιας ασφάλειας και των υπηρεσιών κοινής ωφέλειας. Χαρακτηριστικό παράδειγμα αυτού είναι η κυβέρνηση της Κίνας, η οποία χρησιμοποιεί μαζικά την αναγνώριση προσώπου.
- **Βιομηχανία Υγείας:** Η υγειονομική περίθαλψη ήταν ένας από τους πρώτους κλάδους που υιοθέτησε τη χρήση της MM για την ανίχνευση εικόνων.
- **Εμπόριο:** Η ευρεία χρήση της TM γίνεται πλέον στο μάρκετινγκ χάρη στην άφθονη πρόσβαση σε δεδομένα. Με την άνθηση των δεδομένων το τμήμα μάρκετινγκ βασίζεται στην TN για τη βελτιστοποίηση της σχέσης πελατών και της εκστρατείας μάρκετινγκ.

1.3.7 Βιβλιογραφικές Αναφορές

1. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Singapore: Springer.
2. Duda, R. O., Hart, P. E. & Stork, D. G. (2002). *Pattern Classification*. New York: Wiley-Interscience.
3. Harrington, P. (2012). *Machine Learning in Action*. USA: Manning Publications Co.
4. Raschka, S. (2016). *Python Machine Learning*. Birmingham: Packt Publishing Ltd.
5. Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. New York: Cambridge University Press.
6. Shai Shalev, S. & Shai Ben, D. (2014). *UNDERSTANDING MACHINE LEARNING: From Theory to Algorithms*. New York: Cambridge University Press.
7. Sutton, R. S. & Barto, A. G. (2017). *Reinforcement Learning: An Introduction*. Mass: MIT Press.

1.4 Εκμάθηση Βάθους

1.4.1 Εισαγωγή

Η Εκμάθηση Βάθους (MB) είναι μια τεχνική MM που διδάσκει τους υπολογιστές να ενεργούν, όπως ακριβώς διδάσκονται ανάλογα και οι άνθρωποι μέσα από παραδείγματα.

Τα μοντέλα της MB είναι αόριστα, εμπνευσμένα από τις επεξεργασίες πληροφοριών και τα πρότυπα επικοινωνίας των βιολογικών νευρικών συστημάτων, αλλά παρουσιάζουν ποικίλες διαφορές από τις δομικές και λειτουργικές ιδιότητες των βιολογικών εγκεφάλων (κυρίως των ανθρώπινων εγκεφάλων), που στην ουσία τα καθιστούν ασυμβίβαστα σε σχέση με τη Νευρολογία. Η MB είναι μια εξειδικευμένη μορφή της MM και βρίσκει όλο και περισσότερα πεδία εφαρμογής της με πολύ καλά αποτελέσματα που φτάνουν σε πολύ μεγάλους βαθμούς ακρίβειας. .

Λέξεις κλειδιά: Εκμάθηση Βάθους, Εποπτευόμενη Εκμάθηση Βάθους, Ημιεποπτευόμενη Εκμάθηση Βάθους, Μη Εποπτευόμενη Εκμάθηση Βάθους, Νευρωνικά Δίκτυα, Επεξεργασία Φυσικής Γλώσσας, Εκμάθηση Μηχανής.

1.4.2 Βασικές Αρχές, Έννοιες και Ορισμοί

Οι διαδικασίες MB διαιρούνται στις εξής κατηγορίες: 1. Εποπτευόμενη (EMM), 2. Ημιεποπτευόμενη (HEMM) και 3. Μη Εποπτευόμενη (MEMM). Η MB αποτελεί μια κλάση αλγορίθμων MM που:

1. Χρησιμοποιεί μια αλληλουχία πολλαπλών επιπέδων μη - γραμμικών μονάδων επεξεργασίας για την εξαγωγή και τον μετασχηματισμό χαρακτηριστικών, όπου κάθε διαδοχικό επίπεδο χρησιμοποιεί την έξοδο από το προηγούμενο επίπεδο ως είσοδο του.
2. Μαθαίνει με τρόπους επίβλεψης, για παράδειγμα ταξινόμηση, ή και με τρόπους που δεν απαιτούν την επίβλεψη, για παράδειγμα ανάλυση προτύπων.
3. Μαθαίνει σε πολλαπλά επίπεδα από αναπαραστάσεις που αντιστοιχούν σε διαφορετικά επίπεδα αφαίρεσης, όπου αυτά αποτελούν μια ιεραρχία εννοιών.

Στη MB, κάθε επίπεδο μαθαίνει να μετασχηματίζει τα δεδομένα της εισόδου του σε μια ελαφρώς πιο αφηρημένη και σύνθετη αναπαράσταση. Στην αναγνώριση εικόνας

για παράδειγμα, η ακατέργαστη είσοδος μπορεί να είναι ένας πίνακας εικονοστοιχείων. Για παράδειγμα, έστω ότι έχουμε σαν εικόνα ένα πρόσωπο, όπου: το πρώτο επίπεδο αναπαράστασης μπορεί να αφαιρέσει τα εικονοστοιχεία και να κωδικοποιήσει τις ακμές, το δεύτερο επίπεδο μπορεί να συνθέσει και να κωδικοποιήσει τις διατάξεις των ακμών, το τρίτο επίπεδο μπορεί να κωδικοποιήσει τη μύτη και τα μάτια, και τέλος το τέταρτο επίπεδο μπορεί να αναγνωρίσει ότι η εικόνα περιέχει ένα πρόσωπο. Επίσης, είναι άξιο αναφοράς το γεγονός ότι μια διαδικασία MB είναι σε θέση από μόνη της να μάθει ποιά χαρακτηριστικά και σε ποιο επίπεδο μπορεί να τοποθετήσει βέλτιστα.

Επομένως, στη MB, ένα μοντέλο υπολογιστή μαθαίνει να εκτελεί διάφορες εργασίες ταξινόμησης απευθείας μέσω των εικόνων, των κειμένων ή των ήχων που του δίνονται ως πληροφορία. Τα μοντέλα MB μπορούν να επιτύχουν εξαιρετική ακρίβεια, η οποία μερικές φορές μπορεί να ξεπεράσει ακόμα και τις επιδόσεις του ανθρώπου, καθώς τα μοντέλα αυτά εκπαιδεύονται χρησιμοποιώντας ένα μεγάλο σύνολο ετικετοποιημένων δεδομένων και αρχιτεκτονικές ΝΔ.

Το χαρακτηριστικό της MB είναι η ακρίβεια. Αυτό συμβαίνει διότι η MB μπορεί και επιτυγχάνει τεράστια ποσοστά ακρίβειας στην αναγνώριση, για παράδειγμα των εικόνων. Το γεγονός αυτό βοηθά τα ηλεκτρονικά είδη ευρείας κατανάλωσης να ανταποκρίνονται στις προσδοκίες/προτιμήσεις των χρηστών και επίσης είναι ένα ζωτικής σημασίας εργαλείο για εργασίες που έχουν να κάνουν με την ασφάλεια των ανθρώπων όπως είναι η οδήγηση με χρήση του αυτόματου πιλότου. Οι τελευταίες εξελίξεις στη MB έχουν βελτιωθεί σε σημείο που μπορεί και υπερβαίνει τους ανθρώπους σε ορισμένες ενέργειες όπως η ταξινόμηση εικόνων αντικειμένων.

Ενώ η MB αναπτύχθηκε αρχικά στη δεκαετία του '80, υπήρξαν δύο κύριοι λόγοι που μόλις πρόσφατα έγινε χρήσιμη και άρχισε να χρησιμοποιείται ευρέως:

1. Η MB απαιτεί μεγάλα ποσά ετικετοποιημένων δεδομένων. Για παράδειγμα, η οδήγηση με αυτόματο πιλότο απαιτεί ως πληροφορία για την MB εκατομμύρια εικόνες και χιλιάδες ώρες βίντεο.
2. Η MB απαιτεί σημαντική υπολογιστική ισχύ. Οι Μονάδες Επεξεργασίας Γραφικών (GPU) υψηλής απόδοσης διαθέτουν μια παράλληλη αρχιτεκτονική που είναι αποτελεσματική για τη MB. Όταν συνδυαστεί με κλάσεις ή με το υπολογιστικό νέφος, δίνει τη δυνατότητα στις ομάδες ανάπτυξης λογισμικού να μειώσουν τον χρόνο της εκπαίδευσης για ένα ΝΔ

Βάθους⁶ από εβδομάδες σε ώρες ή και λιγότερο.

1.4.3 Λειτουργία της Εκμάθησης Βάθους

Οι περισσότερες μέθοδοι MB χρησιμοποιούν αρχιτεκτονικές ΝΔ (βλ. Ενότητα 2.2.3) και γι' αυτόν τον λόγο τα μοντέλα των ΝΔ συχνά καλούνται και ΝΔ Βάθους. Τα μοντέλα MB εκπαιδεύονται χρησιμοποιώντας μεγάλα σύνολα ετικετοποιημένων δεδομένων και αρχιτεκτονικές ΝΔ που μαθαίνουν τις λειτουργίες απευθείας από τα δεδομένα χωρίς την ανάγκη κάποιας χειροκίνητης εξαγωγής χαρακτηριστικών.

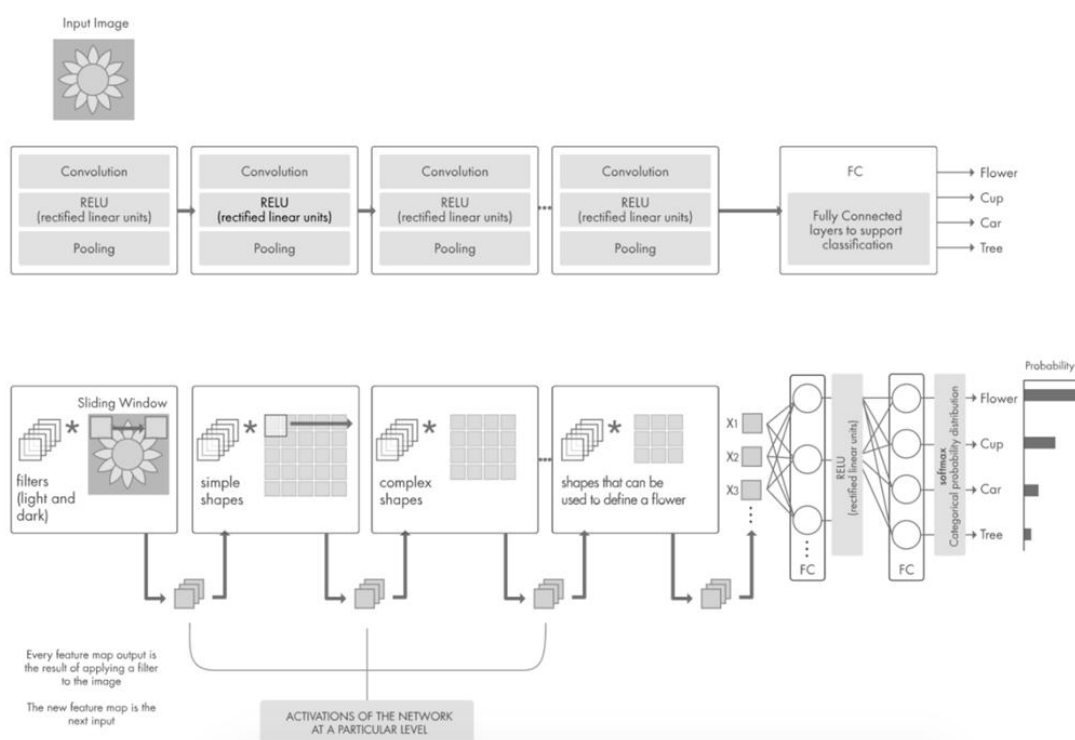
Αναφορικά με την παραδοσιακή MM, η διαδικασία Εκμάθησης είναι EMM και ο προγραμματιστής θα πρέπει να είναι πολύ συγκεκριμένος όταν λέει στον υπολογιστή ποιά πράγματα θα πρέπει να αναζητά για να αποφασίσει αν μια εικόνα, για παράδειγμα, περιέχει έναν σκύλο ή όχι. Αυτή είναι μια επίπονη διαδικασία που ονομάζεται εξαγωγή χαρακτηριστικών και το ποσοστό επιτυχίας του υπολογιστή εξαρτάται εξ' ολοκλήρου από την ικανότητα του προγραμματιστή να ορίσει με ακρίβεια ένα σύνολο χαρακτηριστικών για τον «σκύλο». Ενώ αντίθετα, το πλεονέκτημα της MB, η οποία είναι MEMM, είναι ότι το πρόγραμμα «χτίζει» το χαρακτηριστικό που έχει τεθεί από μόνο του χωρίς κάποια επίβλεψη, κάνοντάς την όχι μόνο ταχύτερη αλλά συνήθως και πιο ακριβή.

Αρχικά, το πρόγραμμα του υπολογιστή μπορεί να είναι εφοδιασμένο με δεδομένα εκπαίδευσης, δηλαδή με ένα σύνολο εικόνων που ο άνθρωπος έχει ετικετοποιήσει, με τη χρήση «μεταετικετών», κάθε εικόνα του ως «σκύλο» ή «όχι σκύλο». Το πρόγραμμα χρησιμοποιεί στη συνέχεια τις πληροφορίες που δέχεται από τα δεδομένα εκπαίδευσης για να δημιουργήσει ένα σύνολο χαρακτηριστικών για τον «σκύλο» και για να δημιουργήσει ένα πρότυπο πρόβλεψης. Σε αυτή την περίπτωση, το πρότυπο που δημιουργεί ο υπολογιστής για πρώτη φορά μπορεί να προβλέψει ότι οτιδήποτε σε μια εικόνα που έχει τέσσερα πόδια και μια ουρά θα πρέπει να φέρει την ετικέτα «σκύλος». Φυσικά, το πρόγραμμα δεν γνωρίζει τις ετικέτες που φέρουν τις ενδείξεις «τέσσερα πόδια» ή «ουρά» και γι' αυτόν τον λόγο θα αναζητήσει απλά πρότυπα «εικονοστοιχεία» στα ψηφιακά δεδομένα. Ωστόσο, με κάθε επανάληψη, το πρότυπο πρόβλεψης που δημιουργεί ο υπολογιστής καθίσταται όλο και πιο πολύπλοκο αλλά

⁶ Ο όρος «βάθος» συνήθως αναφέρεται στον αριθμό των κρυφών επιπέδων που περιέχονται σε ένα ΝΔ, καθώς, τα τυπικά ΝΔ περιέχουν μόνο 2-3 κρυφά επίπεδα, ενώ αντίθετα τα ΝΔ Βάθους μπορεί να περιέχουν έως και 150 κρυφά επίπεδα.

και πιο ακριβές.

Ένας από τους πιο δημοφιλείς τύπους ΝΔ Βάθους είναι τα Συνελικτικά Νευρωνικά Δίκτυα (ΣΝΔ) (βλ. Ενότητα 2.3.3.1.4). Ένα ΣΝΔ περιγράφει τις λειτουργίες που έχει μάθει μέσω των δεδομένων της εισόδου του και χρησιμοποιεί ένα δισδιάστατο «συνελικτικό επίπεδο» προκειμένου να καταστήσει την αρχιτεκτονική του ΣΝΔ κατάλληλη για την επεξεργασία δισδιάστατων δεδομένων, όπως για παράδειγμα τις εικόνες.



Εικόνα 8: Παράδειγμα δικτύου με πολλά «συνελικτικά επίπεδα». Τα φίλτρα εφαρμόζονται σε κάθε εικόνα εκπαίδευσης με «διαφορετικές αναλύσεις» η κάθε μία και η έξοδος κάθε «συνελικτικής εικόνας» χρησιμεύει ως είσοδος στο επόμενο επίπεδο

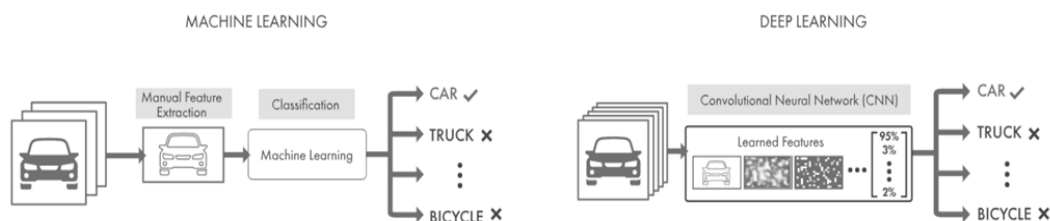
Τα ΣΝΔ επί της ουσίας εξαλείφουν την ανάγκη της χειροκίνητης εξαγωγής χαρακτηριστικών, με αποτέλεσμα να μην χρειάζεται ο προσδιορισμός των χαρακτηριστικών για την ταξινόμηση εικόνων. Τα ΣΝΔ λειτουργούν εξάγοντας τα χαρακτηριστικά απευθείας από τις εικόνες. Τα σχετικά χαρακτηριστικά δεν έχουν προεκπαιδευτεί, αλλά μαθαίνονται όταν το δίκτυο εκπαιδευτεί από μια συλλογή εικόνων. Αυτή η αυτοματοποιημένη εξαγωγή χαρακτηριστικών καθιστά τα μοντέλα MB ιδιαίτερα ακριβή για διεργασίες που αφορούν την υπολογιστική όραση, όπως για παράδειγμα είναι η ταξινόμηση αντικειμένων.

Τα ΣΝΔ μαθαίνουν να ανιχνεύουν τα διαφορετικά χαρακτηριστικά μιας εικόνας χρησιμοποιώντας δεκάδες ή εκατοντάδες κρυφά επίπεδα (βλ. Εικόνα 8). Κάθε κρυφό επίπεδο αυξάνει την πολυπλοκότητα των χαρακτηριστικών των εικόνων που αναγνωρίζει το δίκτυο. Για παράδειγμα, το πρώτο κρυφό επίπεδο θα μπορούσε να ανιχνεύει και να αναγνωρίζει τις άκρες, ενώ το τελευταίο κρυφό επίπεδο θα μπορούσε να ανιχνεύει και να αναγνωρίζει πιο πολύπλοκα σχήματα, ειδικά για το σχήμα του αντικειμένου που εξετάζει.

1.4.4 Διαφορά Εκμάθησης Βάθους και Εκμάθησης Μηχανής

Όπως αναφέρθηκε την αρχή της ενότητας, η MB είναι μια εξειδικευμένη μορφή της MM. Η ροή εργασίας της MM ξεκινά με τα σχετικά χαρακτηριστικά που εξάγονται χειροκίνητα από τις εικόνες. Τα χαρακτηριστικά αυτά στη συνέχεια χρησιμοποιούνται για τη δημιουργία ενός προτύπου που κατηγοριοποιεί τα αντικείμενα της εικόνας. Ενώ με τη ροή εργασίας της MB, τα σχετικά χαρακτηριστικά εξάγονται αυτόματα από τις εικόνες. Επιπλέον, η MB «πραγματοποιεί τη Εκμάθηση από άκρο σε άκρο», δίνοντας σε ένα δίκτυο ακατέργαστα δεδομένα και μια διαδικασία προς εκτέλεση, για παράδειγμα μια διαδικασία ταξινόμησης δεδομένων, την οποία μαθαίνει να εκτελεί με αυτόματο τρόπο.

Μια άλλη βασική διαφορά είναι ότι οι αλγόριθμοι MB προσαρμόζονται με τον όγκο των δεδομένων και συγκλίνουν στο προσδοκώμενο αποτέλεσμα συγκεκριμένου ενδιαμέσου επιπέδου καθώς προστίθενται παραδείγματα για εκπαίδευση του δικτύου. Με άλλα λόγια ένα βασικό πλεονέκτημα των δικτύων MB είναι ότι αυτά συχνά συνεχίζουν να βελτιώνονται όσο αυξάνεται το μέγεθος των δεδομένων τους.



Εικόνα 9: Σύγκριση μιας προσέγγισης MM όσον αφορά την κατηγοριοποίηση των οχημάτων (αριστερά) με τη MB (δεξιά)

Συμπερασματικά επισημαίνουμε ότι στη MM, επιλέγονται χειροκίνητα οι λειτουργίες και ο ταξινομητής για την ταξινόμηση των εικόνων, ενώ αντίθετα στη MB, τα βήματα εξαγωγής και μοντελοποίησης των χαρακτηριστικών γίνονται αυτόματα.

1.4.5 Εφαρμογές της Εκμάθησης Βάθους

Όλο και περισσότερο βρίσκει εφαρμογές η MB, τις περισσότερες φορές ως διαδικασία στις όποιες επεξεργασίες των ΝΔ (βλ. Ενότητα 2). Παρακάτω παραθέτομε κάποιες αντιπροσωπευτικές εφαρμογές της MB:

1.4.5.1 Αυτόματη Αναγνώριση Ομιλίας: Η μεγάλης κλίμακας αυτόματη αναγνώριση ομιλίας είναι η πρώτη και πιο πειστική επιτυχημένη περίπτωση της MB. Τα Επαναληπτικά Νευρωνικά Δίκτυα (ΕΝΔ) Αμφίδρομης Μακράς Βραχυπρόθεσμης Μνήμης (ΕΝΔ-ΑΜΒΜ) έχουν την ικανότητα να μαθαίνουν τις διεργασίες με «Εκμάθηση πολλαπλού βάθους» που περιλαμβάνουν διαστήματα πολλαπλών δευτερολέπτων, τα οποία περιέχουν συμβάντα ομιλίας χωρισμένα από χιλιάδες διακριτά βήματα χρόνου, όπου κάθε χρονικό βήμα αντιστοιχεί σε περίπου 10 ms.

1.4.5.2 Αναγνώριση Εικόνας: Ένα κοινό σύνολο αξιολογήσεων για την ταξινόμηση των εικόνων είναι το σύνολο δεδομένων της βάσης δεδομένων MNIST⁷. Το MNIST αποτελείται από χειρόγραφα ψηφία και περιλαμβάνει 60.000 παραδείγματα εκπαίδευσης και 10.000 παραδείγματα δοκιμών. Όπως και με το TIMIT⁸, του οποίου το μικρό του μέγεθος επιτρέπει στους χρήστες να δοκιμάζουν πολλαπλές διαμορφώσεις. Υπάρχει πλήρης κατάλογος αποτελεσμάτων για αυτό το τεστ.

Η αναγνώριση εικόνας που βασίζεται στη MB «έχει ξεπεράσει την ανθρώπινη προσπάθεια», καθώς παράγει πιο ακριβή αποτελέσματα από αυτά που προκύπτουν από τους ανθρώπους που το προσπαθούν. Η πρώτη αναγνώριση με χρήση αυτής της μεθόδου πραγματοποιήθηκε το 2011.

1.4.5.3 Επεξεργασία εικονικής τέχνης: Σχετική με την πρόοδο που έχει σημειωθεί στην αναγνώριση της εικόνας είναι η αυξανόμενη εφαρμογή τεχνικών MB σε διάφορες διεργασίες οπτικής τέχνης. Τα ΝΔ βάθους έχουν αποδειχθεί κατάλληλα, π.χ. για: 1. την αναγνώριση του στυλ κάποιας δοθείσας ζωγραφικής, 2. τη «σύλληψη» του στυλ μιας δοθείσας ζωγραφικής και της εφαρμογής της σε τρόπο οπτικά ευχάριστο σε μια αυθαίρετη φωτογραφία και 3. εικόνες που βασίζονται σε τυχαία οπτικά πεδία εισαγωγής.

⁷ Η βάση δεδομένων MNIST (Modified National Institute of Standards and Technology database) είναι μια μεγάλη βάση δεδομένων με χειρόγραφα ψηφία που χρησιμοποιείται συνήθως για την εκπαίδευση διαφόρων συστημάτων επεξεργασίας εικόνας αλλά και για την εκπαίδευση και δοκιμή στον τομέα της MM.

⁸ Το σώμα TIMIT της αναγνώρισης λόγου έχει σχεδιαστεί για να παρέχει δεδομένα ομιλίας για την απόκτηση ακουστικής-φωνητικής γνώσης και για την ανάπτυξη και την αξιολόγηση των αυτόματων συστημάτων αναγνώρισης ομιλίας.

1.4.5.4 Επεξεργασία φυσικής γλώσσας: Τα ΝΔ έχουν χρησιμοποιηθεί για την εφαρμογή γλωσσικών προτύπων από τις αρχές της δεκαετίας του 2000. Η AMBM βοήθησε στη βελτίωση της MM και της μοντελοποίησης των γλωσσών.

Άλλες βασικές τεχνικές σε αυτό το πεδίο είναι η αρνητική δειγματοληψία και η ενσωμάτωση των λέξεων. Η ενσωμάτωση λέξεων, όπως η μέθοδος Word2Vec (βλ. Ενότητα 3.2.2), μπορεί να θεωρηθεί ως ένα επίπεδο αναπαράστασης σε μια αρχιτεκτονική MB που μετατρέπει μια ατομική λέξη σε μια θέση αναπαράστασης της λέξης σε σχέση με άλλες λέξεις στο σύνολο δεδομένων, η θέση αυτή αντιπροσωπεύεται ως ένα σημείο ενός διανυσματικού χώρου. Η χρήση της ενσωμάτωσης λέξεων ως επίπεδο εισόδου για ένα ENΔ (βλ. Ενότητα 2.3.3.2) επιτρέπει στο δίκτυο να αναλύει προτάσεις και φράσεις χρησιμοποιώντας μια αποτελεσματική Γραμματική διανυσματικών συνθέσεων. Μια Γραμματική συνθετικού διανύσματος μπορεί να θεωρηθεί ως μια Γραμματική Ανεξάρτητη Συμφραζομένων η οποία χρησιμοποιεί πιθανότητες και που αξιολογείται από ένα ENΔ. Οι αναδρομικοί αυτόματοι κωδικοποιητές μπορούν να αξιολογήσουν την ομοιότητα της πρότασης και να ανιχνεύσουν την παράφραση. Οι νευρωνικές αρχιτεκτονικές βάθους παρέχουν τα καλύτερα αποτελέσματα όσον αφορά την ανάλυση συναισθημάτων, την ανάκτηση πληροφοριών, την κατανόηση της ομιλίας, τη μηχανική μετάφραση, την αλληλεπίδραση και τα συμφραζόμενα, την στιλιστική αναγνώριση γραφής, την ταξινόμηση κειμένου και άλλα. Είναι αξιοσημείωτο επίσης το γεγονός ότι οι πρόσφατες εξελίξεις γενικεύουν την ενσωμάτωση λέξεων στην ενσωμάτωση προτάσεων.

Το γνωστό σε όλους μας Google Translate, χρησιμοποιεί μεγάλα δίκτυα AMBM. Η νευρωνική μετάφραση της μηχανής Google χρησιμοποιεί μια μέθοδο μεταφραστικής μηχανής που βασίζεται σε παραδείγματα, στην οποία το σύστημα «μαθαίνει μέσα από εκατομμύρια παραδείγματα» και βάσει αυτού μας δίνει τα αποτελέσματα των μεταφράσεων που αναζητούμε. Επιπλέον, το Google Translate μεταφράζει ολόκληρες προτάσεις ταυτόχρονα, και όχι κομμάτια αυτών, καθώς επίσης, υποστηρίζει τη μετάφραση σε εκατό κι πλέον γλώσσες. Το δίκτυο κωδικοποιεί τη σημασιολογία της φράσης και όχι απλώς την απομνημόνευση μεταφράσεων από φράση σε φράση. Τέλος το Google Translate χρησιμοποιεί τα Αγγλικά ως διαμεσολαβητή μεταξύ των περισσότερων ζευγών γλωσσών.

1.4.5.5 Ανακάλυψη φαρμάκων και τοξικολογία: Ένα μεγάλο ποσοστό φαρμάκων αποτυγχάνει να εγκριθεί. Αυτό οφείλεται στην ανεπαρκή με το στόχο τους αποτελεσματικότητα, τις ανεπιθύμητες, εκτός στόχων αλληλεπιδράσεις ή τις απρόβλεπτες τοξικές επιδράσεις. Η έρευνα στον τομέα αυτόν έχει ανακαλύψει ότι η χρήση της MB μπορεί να προβλέψει τους βιομοριακούς στόχους ή αστοχίες, τις τοξικές επιδράσεις των περιβαλλοντικών χημικών ουσιών στα θρεπτικά συστατικά και τα οικιακά προϊόντα και φάρμακα.

Επίσης, το AtomNet είναι ένα σύστημα MB για τον ορθολογικό σχεδιασμό φαρμάκων που βασίζεται σε δομές και χρησιμοποιήθηκε για την πρόβλεψη νέων υποψήφιων βιομορίων για στόχους ασθενειών όπως είναι ο ιός Ebola και η πολλαπλή σκλήρυνση.

1.4.5.6 Διαχείριση σχέσεων πελατών: Η ενισχυμένη MB έχει χρησιμοποιηθεί για την άμεση προσέγγιση της αξίας των πιθανών δράσεων μάρκετινγκ, που καθορίζονται από την άποψη των μεταβλητών RFM⁹. Η συνάρτηση της εκτιμώμενης τιμής έχει δείξει ότι έχει μια φυσική ερμηνεία όσο η αξία της ζωής του πελάτη.

1.4.5.7 Σύστημα συστάσεων: Τα συστήματα συστάσεων έχουν χρησιμοποιήσει τη MB για να εξάγουν σημαντικά χαρακτηριστικά για ένα πρότυπο λανθανόντων παραγόντων για συστάσεις που αφορούν τη μουσική βάση του περιεχομένου. Η MB Multiview έχει εφαρμοστεί για την Εκμάθηση των προτιμήσεων των χρηστών αναφορικά με πολλούς και διάφορους τομείς. Το πρότυπο χρησιμοποιεί μια υβριδική προσέγγιση συνεργασίας και περιεχομένου και βελτιώνει τις συστάσεις σε πολλαπλές διεργασίες.

1.4.5.8 Βιοπληροφορική: Ένας αυτόματος κωδικοποιητής τύπου ΝΔ χρησιμοποιήθηκε στη Βιοπληροφορική, προκειμένου να προβλέψει τα σχόλια περί της οντολογίας των γονιδίων και την σχέση μεταξύ γονιδίου και λειτουργίας τους.

Στην ιατρική πληροφορική, η MB χρησιμοποιήθηκε για την πρόβλεψη της ποιότητας του ύπνου με βάση τα δεδομένα που αντλήθηκαν από τις φορητές συσκευές που φοριούνται τύπου smart watch αλλά και για τις προβλέψεις των επιπλοκών της υγείας από τα δεδομένα των ηλεκτρονικών ιατρικών εφαρμογών. Η MB είναι επίσης αποτελεσματική και στην υγειονομική περίθαλψη.

⁹ Το RFM είναι μια μέθοδος που χρησιμοποιείται για την ανάλυση της αξίας ενός πελάτη. Χρησιμοποιείται συνήθως στις βάσεις δεδομένων του μάρκετινγκ και στο άμεσο μάρκετινγκ και έχει προσεχθεί ιδιαίτερα από τις βιομηχανίες λιανικών και επαγγελματικών υπηρεσιών.

1.4.5.9 TN και MB σε σχέση με την Υγεία: Τα τελευταία χρόνια ο τομέας της υγειονομικής περίθαλψης έχει ωφεληθεί σημαντικά από την εξάπλωση των μηχανών που η λειτουργία τους υποστηρίζεται από την TN, όπως για παράδειγμα είναι η μέτρηση της αρτηριακής πίεσης, κ.ά. Το επίτευγμα αυτό αποτελεί μια από τις πολλές εξελίξεις που θα μπορούσε να βελτιώσει δραστικά την εν γένει διάγνωση στην ιατρική, με καλύτερα ποσοστά επιβίωσης, αφού θα είναι εφικτή η αποτελεσματικότερη πρόληψη της νόσου και όχι η εκ των υστέρων θεραπεία της. Αυτό με το κατάλληλο λογισμικό TN και MB όπου απαιτείται, που αντίστοιχα «μαθαίνει να αναγνωρίζει τα όποια μοτίβα σε ξεχωριστά επίπεδα, ώστε κάθε επίπεδο του ΝΔ να λειτουργεί ανεξάρτητα και συντονισμένα, χωρίζοντας πτυχές όπως το χρώμα, το μέγεθος και το σχήμα πριν την ενσωμάτωση των αποτελεσμάτων» της ιατρικής απεικόνισης. Σημειώνεται ότι, αυτή η πρόοδος στη χρήση οπτικών εργαλείων είναι ζωτικής σημασίας για την πρόοδο της ιατρικής διάγνωσης, αφού τα προηγούμενα χρόνια χρησιμοποιήθηκε η TN μόνο στο πιο βασικό της επίπεδο - από το να εκπαιδεύονται οι μηχανές προκειμένου να εκτελούν ανθρώπινες δεξιότητες, με αισιόδοξα επιτυχή αποτελέσματα και σε μελλοντικές ανακαλύψεις.

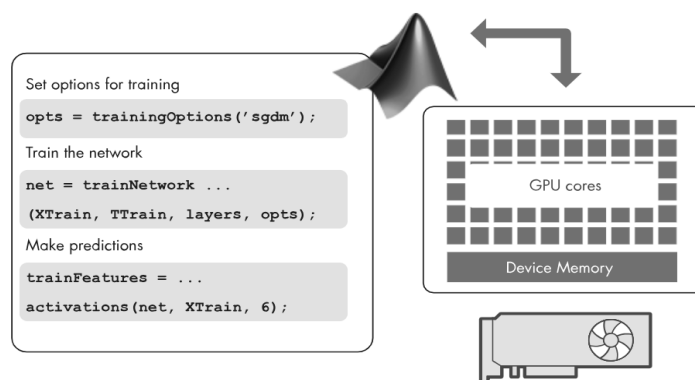
1.4.5.10 Διαφημίσεις στα κινητά τηλέφωνα: Ο εντοπισμός του κατάλληλου κοινού που χρησιμοποιεί το διαδίκτυο στα κινητά τηλέφωνα, προκειμένου να προωθούνται οι διαφημίσεις στις συσκευές τους είναι πάντα μια δύσκολη υπόθεση, αφού θα πρέπει πρώτα να εξεταστούν και να αφομοιωθούν πολλά συναφή δεδομένα για να χρησιμοποιηθούν στην προβολή των διαφημίσεων από οποιονδήποτε διακομιστή διαφημίσεων. Η MB έχει χρησιμοποιηθεί για την ερμηνεία και συλλογή μεγάλων, πολυδιάστατων συνόλων δεδομένων, κατά τη διάρκεια του κύκλου διαφήμισης στο διαδίκτυο. Αυτές οι πληροφορίες μπορούν να αποτελέσουν τη βάση της Εκμάθησης των μηχανών για τη βελτίωση της επιλογής των διαφημίσεων.

1.4.5.11 Επαναφορά εικόνας: Η MB έχει εφαρμοστεί με επιτυχία σε αντίστροφα προβλήματα, όπως η αποκόλληση, η υπερανάλυση, η αποτύπωση και ο χρωματισμός του φιλμ. Αυτές οι εφαρμογές περιλαμβάνουν μεθόδους Εκμάθησης όπως είναι τα «Πεδία Συρρίκνωσης για αποτελεσματική αποκατάσταση της εικόνας». Δηλαδή, εφαρμογές που εκπαιδεύονται από ένα σύνολο δεδομένων για την επαναφορά της προηγούμενης μορφής σε βάθος της εικόνας που καταστράφηκε. Το πρόγραμμα «Deep Image Prior», είναι ένα παράδειγμα προγράμματος που εκπαιδεύεται για να επαναφέρει την εικόνα που χρειάζεται επιδιόρθωση.

1.4.5.12 Ανίχνευση οικονομικής απάτης: Η MB εφαρμόζεται με επιτυχία στην ανίχνευση της οικονομικής απάτης και κατά του ξεπλύματος χρημάτων. Το σύστημα ανίχνευσης βάθους της νομιμοποίησης των εσόδων από παράνομες δραστηριότητες μπορεί να εντοπίσει και να αναγνωρίσει τις σχέσεις και τις ομοιότητες μεταξύ των δεδομένων και, παραπέρα, να μάθει να ανιχνεύει ανωμαλίες ή να ταξινομεί και να προβλέπει συγκεκριμένα γεγονότα. Η λύση αξιοποιεί και τις τεχνικές EMM, όπως είναι η ταξινόμηση των ύποπτων συναλλαγών και τη Εκμάθηση χωρίς επίβλεψη, π.χ. την ανίχνευση ανωμαλιών.

1.4.6 Επιτάχυνση μοντέλων Εκμάθησης Βάθους μέσω Μονάδας Επεξεργασίας Γραφικών

Η εκπαίδευση ενός μοντέλου MB μπορεί να διαρκέσει πολύ καιρό, από μέρες έως εβδομάδες. Η χρήση ενός επιταχυντή GPU μπορεί να επιταχύνει σημαντικά αυτή τη διαδικασία. Η χρήση του MATLAB με GPU μπορεί να μειώσει αισθητά το χρόνο που απαιτείται για την εκπαίδευση ενός δικτύου και μπορεί επίσης να μειώσει το χρόνο για ένα πρόβλημα ταξινόμησης εικόνας, από μέρες σε ώρες. Κατά την εκπαίδευση μοντέλων MB, το MATLAB χρησιμοποιεί τη Μονάδα Επεξεργασίας Γραφικών (όποτε αυτή είναι διαθέσιμη) χωρίς να χρειάζεται ο προγραμματιστής να καταλάβει πώς να τις προγραμματίσει ρητά.



Εικόνα 10: Εργαλειοθήκη εντολών MB για την εκπαίδευση ενός ιδιόκτητου ΣΝΔ από το μηδέν ή χρησιμοποιώντας ένα προεκπαιδευμένο μοντέλο για τη μεταφερόμενη Εκμάθηση

1.4.7 Παράδειγμα Αλγορίθμου Εκμάθησης Βάθους για την Επεξεργασία Φυσικής Γλώσσας και πλεονεκτήματά του

Η MB έχει χρησιμοποιηθεί εκτεταμένα στην ΕΦΓ και αυτό διότι η MB είναι

κατάλληλη για την Εκμάθηση της σύνθετης υποκείμενης δομής μιας φράσης και της σημασιολογικής εγγύτητας διαφόρων λέξεων.

Αξίζει να σημειωθεί επίσης ότι η MB εμφανίζει αρκετά πλεονεκτήματα έναντι άλλων αλγορίθμων για την ΕΦΓ και αυτά φαίνονται παρακάτω:

- 1. Ευέλικτα Πρότυπα (Μοντέλα):** Τα πρότυπα MB είναι πολύ πιο ευέλικτα από άλλα πρότυπα δομές, προσθέτοντας και αφαιρώντας επίπεδα ανάλογα με τις ανάγκες του. Ένα άλλο προτέρημα των μοντέλων MB είναι το γεγονός ότι επιτρέπουν τη δημιουργία προτύπων με ευέλικτες εξόδους. Αυτή η ευελιξία είναι στην ουσία το κλειδί για την ανάπτυξη προτύπων τα οποία είναι κατάλληλα για την κατανόηση σύνθετων γλωσσικών δομών. Η MB είναι επίσης απαραίτητη και για την ανάπτυξη διαφόρων εφαρμογών ΕΦΓ όπως μεταφράσεις, chatbots και εφαρμογές κειμένου σε ομιλία.
- 2. Απαιτείται λιγότερη γνώση επάνω στον τομέα:** Ενώ κάποιος χρειάζεται σίγουρα κάποια γνώση και διαίσθηση για να αναπτύξει ένα καλό μοντέλο MB, η ικανότητα των αλγορίθμων MB να μαθαίνουν ιεραρχίες χαρακτηριστικών από μόνοι τους σημαίνει ότι ένας προγραμματιστής δεν χρειάζεται να έχει τόσο βαθιά γνώση του χώρου του προβλήματος για να αναπτύξει αλγορίθμους MB στην ΕΦΓ. Για ένα χώρο προβλημάτων τόσο πολύπλοκο όσο η φυσική γλώσσα, αυτό είναι ένα πολύ ευπρόσδεκτο πλεονέκτημα.
- 3. Εύκολη συνεχής Εκμάθηση:** Οι αλγόριθμοι MB είναι εύκολο να εκπαιδευτούν καθώς εισέρχονται νέα δεδομένα. Ενώ από την άλλη μερικοί αλγόριθμοι MM απαιτούν την αποστολή ολόκληρου του συνόλου δεδομένων μέσω του προτύπου για ενημέρωση, γεγονός που θα δημιουργούσε πρόβλημα για μεγάλα σύνολα δεδομένων.

Ένα χαρακτηριστικό παράδειγμα της εφαρμογής της MB περιγράφεται παρακάτω όπου και θα χρησιμοποιηθεί το ευρέως γνωστό σύνολο δεδομένων MNIST.

- **Σε πρώτη φάση στον κώδικα που ακολουθεί θα εισαχθούν οι βιβλιοθήκες που απαιτούνται:**

```
import nltk
from nltk import word_tokenize, sent_tokenize
import gensim
from gensim.models.word2vec import Word2Vec
```

```
from sklearn.manifold import TSNE
import pandas as pd
from bokeh.io import output_notebook
from bokeh.plotting import show, figure
```

- **Στη συνέχεια θα φορτωθεί το πρότυπο (μοντέλο) και τα δεδομένα:**

```
# Punctuation and Tokenizer module
nltk.download('punkt')
# The Gutenberg dataset. A set of 18 books we can use
# to train upon.
nltk.download('gutenberg')
from nltk.corpus import gutenberg
```

- **Εάν ο χρήστης θέλει να ρίξει μια ματιά στα βιβλία που είναι διαθέσιμα στο σύνολο δεδομένων Gutenberg, αρκεί να εκτελέσει την παρακάτω γραμμή κώδικα:**

```
print(gutenberg.fileids())
```

- **Για να φορτωθούν οι προτάσεις χρειάζεται να εισαχθούν οι ακόλουθοι κώδικες:**

```
# Due to lack of resources, I'm not working with the full Gutenberg
# dataset (18 books). If you got a GPU, you can just omit the
# 'fields' parameter and all 18 books will be loaded.
gutenberg_sents = gutenberg.sents(fileids=['bible-kjv.txt',
                                           'austen-emma.txt',
                                           'austen-persuasion.txt',
                                           'austen-sense.txt',
                                           'carroll-alice.txt'])
```

- **Εάν ο χρήστης επιθυμεί να μάθει πόσες λέξεις υπάρχουν στο σύνολο που φορτώθηκε, αρκεί να εκτελέσει την παρακάτω γραμμή κώδικα:**

```
print(len(gutenberg.sents(fileids=['bible-kjv.txt',
                                   'austen-emma.txt',
                                   'austen-persuasion.txt',
                                   'austen-sense.txt',
```

'carroll-alice.txt']))

1.4.8 Βιβλιογραφικές Αναφορές

1. Aggarwal, C. C. (2018). *Neural Networks and Deep Learning*. Switzerland: Springer.
2. Chollet, F. (2017). *DEEP LEARNING with Python*. USA: Manning Publications Co.
3. Goodfellow, I., Bengio, Y. & Courville, A. (2016). *DEEP LEARNING*. Mass: MIT Press.
4. Patterson, J. & Gibson, A. (2017). *Deep Learning*. Sebastopol: O'Reilly Media.
5. Rodrigues, W. (2017). Deep Learning for Natural Language Processing – Part I. Retrieved January 10, 2019, from <https://medium.com/cityai/deep-learning-for-natural-language-processing-part-i-8369895ffb98>.
6. Rodrigues, W. (2017). Deep Learning for Natural Language Processing – Part II. Retrieved January 12, 2019, from <https://medium.com/cityai/deep-learning-for-natural-language-processing-part-ii-8b2b99b3fa1e>.
7. Rodrigues, W. (2017). Deep Learning for Natural Language Processing – Part III. Retrieved January 12, 2019, from <https://medium.com/cityai/deep-learning-for-natural-language-processing-part-iii-96cfc6acfcc3>.

ΚΕΦΑΛΑΙΟ 2

Τεχνητά Νευρωνικά Δίκτυα

2.1 Εισαγωγή

Τα ΤΝΔ στο εξής ΝΔ αποτελούν έναν νέο κλάδο στον τομέα των φυσικών επιστημών και έχουν κάνει αισθητή την παρουσία τους τα τελευταία χρόνια. Η δομή και η λειτουργία των ΝΔ βασίστηκε στο ανθρώπινο κεντρικό νευρικό σύστημα και γι' αυτόν τον λόγο μιμούνται τη λειτουργία των βιολογικών νευρώνων του. Τα ΝΔ είναι χτισμένα σε ένα μεγάλο αριθμό απλών και προσαρμόσιμων μονάδων επεξεργασίας που είναι συνδεδεμένες με τρόπο τέτοιο ώστε να μπορούν να αποθηκεύουν βιωματική γνώση μέσω της Εκμάθησης από παραδείγματα και όπως τα βιολογικά συστήματα, έχουν την ικανότητα να λαμβάνουν ασαφής πληροφορίες από τον έξω κόσμο και να τις επεξεργάζονται χωρίς κάποιο ρητό σύνολο κανόνων. Η προσέγγισή τους αυτή είναι σε αντίθεση με την παραδοσιακή προσέγγιση των υπολογισμών, η οποία επεξεργάζεται πληροφορίες διαδοχικά σύμφωνα με ένα σύνολο κανόνων με ακρίβεια.

Έτσι η πηγή έμπνευσης για την έρευνα που έγινε σχετικά με τα ΝΔ αφορούσε στη δομή και τη λειτουργία του εγκεφάλου. Είναι γνωστό ότι το κύριο δομικό στοιχείο του εγκεφάλου είναι οι νευρώνες, δηλαδή τα νευρωνικά κύτταρα τα οποία δημιουργούν ένα πυκνό δίκτυο επικοινωνίας μεταξύ τους. Ο στόχος των μελετών του νευρώνα και των ΝΔ είναι να επινοηθεί ένα νέο μοντέλο υπολογισμού με δομή παρόμοια με αυτή του εγκεφάλου. Η προσέγγιση αυτή υπολογισμού πιστεύεται να είναι καταλληλότερη για την ανάπτυξη «ευφυών αλγορίθμων» και γενικότερα διαδικασιών που σχετίζονται με την ΤΝ, όπως για παράδειγμα η Επεξεργασία

Φυσικής Γλώσσας, η Ομαδοποίηση Προτύπων κ.λπ. Οι ομοιότητες των ΝΔ και του εγκεφάλου εντοπίζονται αφενός στη διαδικασία Εκμάθησης των ΝΔ, που τους παρέχει την απαιτούμενη γνώση για να εκτελέσουν τις όποιες διεργασίες τους ζητηθούν, αφετέρου στην αποθήκευση αυτής της γνώσης στις συνδέσεις των νευρώνων, δηλαδή τα συναπτικά βάρη, τα οποία αποτελούν αναπόσπαστο κομμάτι της ομαλής και σωστής λειτουργίας των ΝΔ.

Λέξεις Κλειδιά: Τεχνητά Νευρωνικά Δίκτυα, Νευρωνικά Δίκτυα, Βιολογικά Νευρωνικά Δίκτυα, Νευρώνας, Βιολογικός Νευρώνας, Αισθητήρας, Κρυφό Επίπεδο, Αρχιτεκτονική Νευρωνικών Δικτύων, Δίκτυα Πρόσθιας Τροφοδότησης, Συνελκτικά Νευρωνικά Δίκτυα, Αναδρομικά Νευρωνικά Δίκτυα, Επαναληπτικά Νευρωνικά Δίκτυα, Συναπτικά Βάρη, Σταθμισμένο Άθροισμα, Συνάρτηση Ενεργοποίησης, Κατώφλι, Εκπαίδευση Νευρωνικών Δικτύων.

2.2 Ιστορική Αναδρομή

Η μελέτη του ανθρώπινου εγκεφάλου ξεκίνησε χιλιάδες χρόνια πριν. Στα πλαίσια της σύγχρονης ηλεκτρονικής προσέγγισης οι άνθρωποι προσπάθησαν να αξιοποιήσουν τη διαδικασία σκέψης του εγκεφάλου. Το πρώτο βήμα προς τα ΝΔ έγινε το 1943 όταν ο Warren McCulloch, ένας νευροφυσιολόγος και ένας νεαρός τότε μαθηματικός, ο Walter Pitts, έγραψαν μία εργασία σχετικά με τον τρόπο με τον οποίο οι νευρώνες θα μπορούσαν να λειτουργήσουν. Η εργασία αυτή θεωρείται ότι είναι η πρώτη ιστορικά προσέγγιση ενός ΝΔ. Επιπλέον, οι συγγραφείς παρομοίασαν τους νευρώνες και τις συνδέσεις τους με το ηλεκτρικό κύκλωμα γι' αυτό και το μοντέλο του απλού νευρώνα που σχεδίασαν είχε ηλεκτρικά κυκλώματα στην αρχιτεκτονική του.

Ενισχύοντας αυτή την έννοια των νευρώνων και τον τρόπο λειτουργίας τους γράφτηκε το 1949 το βιβλίο με τίτλο «The Organization of Behavior» με συγγραφέα τον Donald Hebb. Σε αυτό το βιβλίο επισημάνθηκε ότι τα νευρικά μονοπάτια ενισχύονται κάθε φορά που χρησιμοποιούνται.

Καθώς οι υπολογιστές εξελίχθηκαν στις αρχές της δεκαετίας του 1950, δόθηκε η ευκαιρία να αρχίσουν να μοντελοποιούνται τα βασικά στοιχεία αυτών των θεωριών σχετικά με την ανθρώπινη σκέψη. Ο Nathaniel Rochester ηγήθηκε της πρώτης προσπάθειας προσομοίωσης ενός ΝΔ στα ερευνητικά εργαστήρια της IBM, η οποία ωστόσο απέτυχε. Οι προσπάθειές του άρχισαν να αποδίδουν όταν και η παραδοσιακή πληροφορική άρχισε να ανθίζει. Όπως όμως ήταν αναμενόμενο η τότε έμφαση στη

χρήση των υπολογιστών άφηνε την ουσιαστική έρευνα των ΝΔ στο παρασκήνιο.

Ωστόσο, καθ' όλη αυτή την περίοδο, οι υποστηρικτές των «σκεπτόμενων μηχανών» συνέχισαν να συζητούν τα ζητήματα που είχαν ήδη τεθεί. Το 1956 το ερευνητικό πρόγραμμα Dartmouth Summer για την ΤΝ έδωσε την ώθηση να συνεχιστεί το ενδιαφέρον τόσο στην ΤΝ όσο και στα ΝΔ. Ένα από τα αποτελέσματα αυτής της διαδικασίας ήταν η τόνωση της έρευνας τόσο στην ΤΝ, όσο και στο ευρύ φάσμα της επεξεργασίας των νευρώνων του εγκεφάλου.

Στα χρόνια που ακολούθησαν μετά το πρόγραμμα Dartmouth, ο John von Neumann πρότεινε τη μίμηση των απλών λειτουργιών των νευρώνων χρησιμοποιώντας τηλεγραφήματα ή σωλήνες κενού. Επίσης, ο Frank Rosenblatt, ένας νευροβιολόγος του Cornell, άρχισε να εργάζεται επάνω στον Αισθητήρα (Perceptron). Ήταν ενθουσιασμένος με τη λειτουργία του ματιού της μύγας και πιο συγκεκριμένα με το γεγονός ότι ένα μεγάλο μέρος της επεξεργασίας που δίνει την εντολή σε μια μύγα να πετάξει μακριά γίνεται από το μάτι της. Ο Αισθητήρας, που προέκυψε από αυτή την έρευνα, κατασκευάστηκε με τη βοήθεια του μηχανικού μέρους του υπολογιστή και είναι το παλαιότερο ΝΔ που χρησιμοποιείται ακόμη και σήμερα. Ο Αισθητήρας αυτός, ενός μόνο επιπέδου, αποδείχτηκε χρήσιμος για την κατάταξη ενός συνόλου δεδομένων συνεχών τιμών σε μία από δύο κλάσεις. Δηλαδή, ο Αισθητήρας μπορεί και υπολογίζει ένα άθροισμα βαρών δεδομένων, αφαιρεί στη συνέχεια ένα κατώφλι (2.3.2) και αποδίδει μία από τις δύο πιθανές τιμές ως αποτέλεσμα. Δυστυχώς όμως, ο Αισθητήρας είχε και περιορισμούς που αποδείχτηκαν κατά τα χρόνια της δυστοκίας του χώρου και αυτό αποτυπώθηκε στο βιβλίο με τίτλο «Perceptrons» των Marvin Minsky και Seymour Papert το 1969. Με άλλα λόγια αποδείχτηκε με αναλυτικές μαθηματικές μεθόδους ότι κάποιοι συγκεκριμένοι περιορισμοί που προκύπτουν από τη χρήση του Αισθητήρα, δεν καθιστούν, για παράδειγμα, εφικτή τη λύση του σχετικά απλού προβλήματος X-OR¹⁰.

Οι αρχικές προσδοκίες ωστόσο που είχαν ήδη δημιουργηθεί από την εξέλιξη και τις δυνατότητες των ΝΔ φάνηκε ότι δεν επαληθεύτηκαν, με αποτέλεσμα τα ΝΔ να χάσουν την δημοτικότητα τους και ο κόσμος να στραφεί σε μια νέα περιοχή προσέγγισης της νοημοσύνης, αυτή της ΤΝ.

¹⁰ Το πρόβλημα X-OR, είναι ένα κλασικό πρόβλημα στην έρευνα των ΤΝΔ. Είναι ένα πρόβλημα που χρησιμοποιεί τα ΝΔ προκειμένου να προβλέψει τις εξόδους των Χοr λογικών πυλών δοθέντων δύο δυαδικών εισόδων. Η συνάρτηση Χοr θα πρέπει να επιστρέψει μια πραγματική τιμή εάν οι δύο εισόδοι δεν είναι ίσες και μια τιμή ψεύδους εάν είναι ίσες.

Την ίδια περίπου εποχή, το 1959 οι Bernard Widrow και Marcian Hoff του Stanford ανέπτυξαν τα πρότυπα (μοντέλα) Adaline¹¹ και Madaline¹². Μάλιστα αυτά τα πρότυπα (μοντέλα) ήταν τα πρώτα πρότυπα (μοντέλα) που χρησιμοποιήθηκαν επιτυχώς για πρακτικά προβλήματα, δηλαδή για να εξαλείψουν την ηχώ στις τηλεφωνικές γραμμές.

Με την πρόοδο που επιτελέστηκε, σε πολλαπλά επίπεδα, στον εν γένει χώρο της ΤΝ παρατηρείται την τελευταία δεκαετία μια επανεμφάνιση και εξέλιξη των ΝΔ σε σημείο που να αποτελέσουν ανεξάρτητο πεδίο της ΤΝ. Η πρόοδος αυτή είχε ως συνέπεια την εξέλιξη στην ορολογία, την τεκμηρίωση της δομολειτουργίας και των εφαρμογών τους, τις υλοποιήσεις τους και τον μεγάλο αριθμό των επιστημόνων που ασχολούνται αποκλειστικά με τη νέα αυτή περιοχή.

Πιο συγκεκριμένα, από το 1985 άρχισαν να οργανώνονται από την American Physical Society και από την IEEE τα πρώτα συνέδρια με θέμα τα ΝΔ με περισσότερες από χίλιες συμμετοχές. Ταυτόχρονα δημιουργήθηκαν ποικίλες ειδικές επαγγελματικές-επιστημονικές εταιρίες ΝΔ με χιλιάδες μέλη, όπου μεταξύ άλλων είναι η International Neural Network Society που έχει έδρα τις εξής 3 χώρες: Αμερική (με διευθυντή τον Grossberg), Ευρώπη (Kohonen) και Ιαπωνία (Amari).

Επίσης, στα τέλη της δεκαετίας του 80 πρωτοεμφανίστηκαν τουλάχιστον πέντε νέα περιοδικά που είχαν ως θέμα τους αποκλειστικά τα ΝΔ, μετά το 1990 εκδίδονται και άλλα 3-4 νέα περιοδικά και τελικά σήμερα υπάρχουν περίπου 10 επιστημονικά περιοδικά αφιερωμένα στα ΝΔ. Επιπλέον, μια πλειάδα άρθρων με νέα αποτελέσματα που αφορούν τα ΝΔ δημοσιεύονται κάθε μήνα στα γνωστά περιοδικά της Επιστήμης των Ηλεκτρικών Υπολογιστών, της Φυσικής, και των Ηλεκτρολόγων Μηχανικών. Παρατίθενται ενδεικτικά μερικά από τα εξειδικευμένα νέα περιοδικά:

- Neural Networks: The Official Journal of the International Neural Network Society (Pergamon Press).
- Network: Computation in Neural Systems (Institute of Physics Publishing).
- International Journal of Neural Systems (World Scientific).
- Neural Computation.

¹¹ Adaline: Παίρνει το όνομα του από το **Adaptive Linear Element**. Είναι ένα απλό σύστημα αισθητήρα (perceptron-line system) που επιτυγχάνει την ταξινόμηση τροποποιώντας τα βάρη με τέτοιο τρόπο ώστε να μειώνεται το MSE (Mean Square Error) σε κάθε επανάληψη. Αυτό μπορεί να επιτευχθεί χρησιμοποιώντας ένα προσαρμοστικό γραμμικό στοιχείο κλίσης (gradient adaptive linear element).

¹² Madaline: Παίρνει το όνομα του από το Multiple Adalines Connected. Ένα Madaline αποτελείται από πολλά Adalines. Αυτός ο αλγόριθμος επιτρέπει επίσης στο δίκτυο να επιλύει μη-διαχωρίσιμα προβλήματα (non-separable problems).

- Connection Science: Journal of Neural Computing, Artificial Intelligence and Cognitive Research (Carfax Publishing).
- Neural Network World: Neural and Massively Parallel Computing and Information Systems (Computer World, Prague).

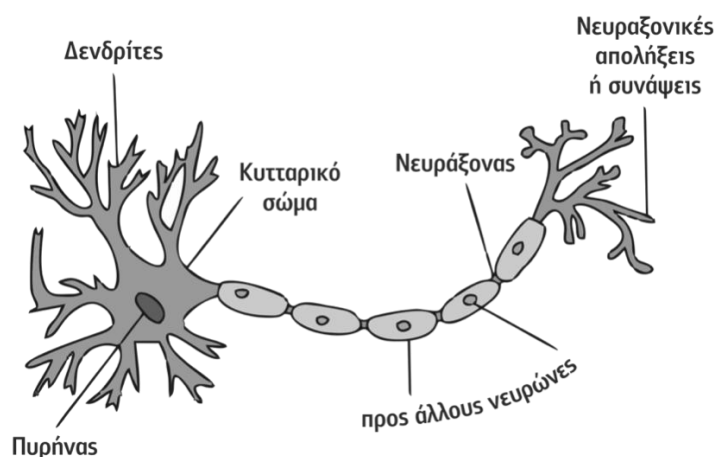
Πολύ σημαντικό είναι επίσης το γεγονός ότι τα τελευταία δεκαπέντε χρόνια ιδρύθηκαν και οι πρώτες εμπορικές εταιρίες οι οποίες ασχολούνται αποκλειστικά με τα ΝΔ. Βρίσκονται σχεδόν όλες στις ΗΠΑ, συνήθως έχουν μικρό αριθμό εργαζομένων (π.χ. 20 άτομα), και παράγουν εξειδικευμένα προγράμματα για τη λύση συγκεκριμένων προβλημάτων. Μερικά από αυτά έχουν επιτυχία, ενώ σε άλλα διαφαίνεται με το πέρασμα του χρόνου ότι οι αρχικές προσδοκίες για ραγδαία αύξηση των εμπορικών εφαρμογών δεν επαληθεύονται, αντιθέτως, έχουν φθάσει πλέον σε ένα σταθερό επίπεδο ανάπτυξης, με μικρή ετήσια αύξηση.

2.3 Βασικές έννοιες και ορισμοί των Τεχνητών Νευρωνικών Δικτύων

2.3.1 Βιολογικά Νευρωνικά Δίκτυα και Τεχνητά Νευρωνικά Δίκτυα

Είναι γνωστό ότι ο Βιολογικός Νευρώνας (BNN) είναι το βασικό δομικό στοιχείο του νευρικού συστήματος του ανθρώπου, η λειτουργία του θα παρουσιαστεί συνοπτικά στη συνέχεια για την καλύτερη κατανόηση της λειτουργίας των τεχνητών νευρώνων αλλά και της αναλογίας μεταξύ ΝΔ και Βιολογικών Νευρωνικών Δικτύων (ΒΝΔ).

2.3.1.1 Βιολογικός Νευρώνας



Εικόνα 11: Βασικός Βιολογικός Νευρώνας με τρεις κύριες λειτουργικές μονάδες: τους δενδρίτες, το κυτταρικό σώμα και τον άξονα

Είναι γνωστό ότι το ανθρώπινο νευρικό σύστημα αποτελείται από

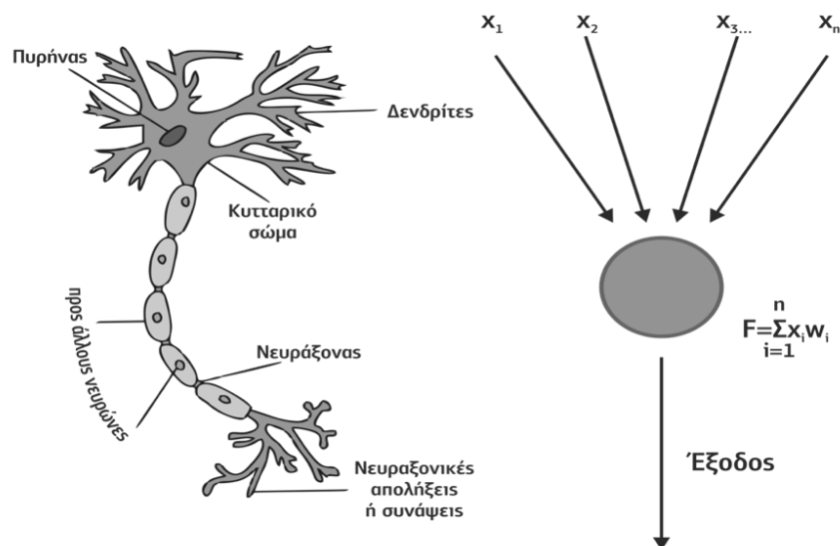
δισεκατομμύρια νευρώνες διαφόρων τύπων και μηκών που έχουν άμεση σχέση με τη θέση τους στο σώμα. Το σχήμα της Εικόνας 11 που ακολουθεί απεικονίζει έναν υπεραπλουστευμένο BNN με τρεις κύριες λειτουργικές μονάδες: τους δενδρίτες, το

κυτταρικό σώμα και τον άξονα.

Το σώμα του κυττάρου έχει έναν πυρήνα που περιέχει τις πληροφορίες σχετικά με τα χαρακτηριστικά κληρονομικότητας και έχει και ένα πλάσμα που κρατά τον μοριακό εξοπλισμό ο οποίος χρησιμοποιείται για την παραγωγή του υλικού που χρειάζεται ο νευρώνας. Οι δενδρίτες στη συνέχεια λαμβάνουν τα σήματα από άλλους νευρώνες και τους μεταφέρουν στο σώμα των κυττάρων. Η συνολική περιοχή λήψης των δενδριτών ενός τυπικού νευρώνα είναι περίπου 0,25 χιλιοστά. Ο άξονας ο οποίος διακλαδίζεται, λαμβάνει σήματα από το κυτταρικό σώμα και τα μεταφέρει διαμέσου της συνάψεως στους δενδρίτες των γειτονικών νευρώνων.

2.3.1.2 Αναλογία Βιολογικών Νευρωνικών Δικτύων και Τεχνητών Νευρωνικών Δικτύων

Η αναλογία μεταξύ του Τεχνητού Νευρώνα (TNN) και του BNN είναι ότι οι συνδέσεις μεταξύ των κόμβων αντιπροσωπεύουν τους άξονες και τους δενδρίτες, τα βάρη σύνδεσης αντιπροσωπεύουν τις συνάψεις και το κατώφλι προσεγγίζει τη δραστηριότητα στο σώμα. Η Εικόνα 12 που ακολουθεί απεικονίζει τον BNN με διάφορα σήματα x και συναπτικά βάρη w που τροφοδοτούν έναν νευρώνα με ένα κατώτατο όριο b , το οποίο καλείται πόλωση, και δίπλα απεικονίζεται το ισοδύναμο σύστημα του TNN.



Εικόνα 12: Η αναλογία Βιολογικού Νευρώνα και Τεχνητού Νευρώνα

2.3.2 Το μοντέλο του Τεχνητού Νευρώνα και η Συνάρτηση Ενεργοποίησης

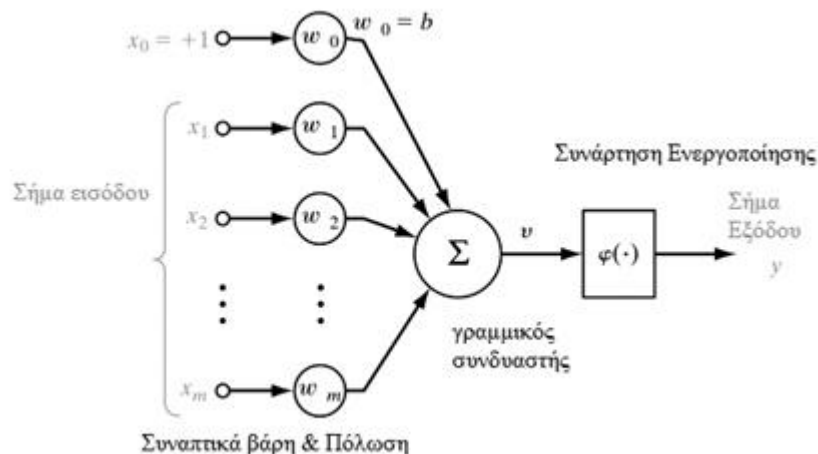
Ο TNN είναι ένα υπολογιστικό πρότυπο (μοντέλο) όπου τα μέρη που το απαρτίζουν παρουσιάζουν αρκετές ομοιότητες με αυτά ενός BNN. Ο TNN απεικονίζεται

χαρακτηριστικά στην Εικόνα 13.

Ένας TNN απαρτίζεται από m εισόδους x_1, x_2, \dots, x_n , από m συναπτικά βάρη w_1, w_2, \dots, w_m , οι τιμές των βαρών μπορεί να είναι είτε αρνητικές είτε θετικές, καθώς και από μια σταθερά την πόλωση b . Δηλαδή, το πρότυπο (μοντέλο) του TNN περιλαμβάνει στην αρχιτεκτονική του μία σταθερά, την πόλωση, που εφαρμόζεται στο εξωτερικό μέρος του και συμβολίζεται με το σύμβολο b . Η πόλωση μπορεί να αυξήσει ή να μειώσει αντίστοιχα τη δικτυακή διέγερση της συνάρτησης ενεργοποίησης, το οποίο εξαρτάται από το εάν είναι θετική ή αρνητική. Η πόλωση b εκφράζεται μέσω ενός συναπτικού βάρους $w_0 = b$ το οποίο βάρους αναφέρεται σε μια είσοδο $x_0 = -1, 1$. Έτσι, κάθε μια από αυτές τις εισόδους πολλαπλασιάζεται με τα αντίστοιχα βάρη και στην απλούστερη περίπτωση μόλις γίνει ο πολλαπλασιασμός αυτός, ο νευρώνας θα κληθεί να υπολογίσει το σταθμισμένο άθροισμα των εισόδων $\sum_i w_i x_i$ και έπειτα να εφαρμόσει σε αυτό τη μη γραμμική συνάρτηση ενεργοποίησης. Το κανονικοποιημένο εύρος των τιμών πλάτους που μπορεί να πάρει η έξοδος ενός νευρώνα κυμαίνεται συνήθως στα διαστήματα $[0,1]$ ή $[-1,1]$. Οι μαθηματικές εξισώσεις που περιγράφουν τον νευρώνα είναι οι ακόλουθες:

$$v = \sum_{i=1}^m x_i w_i \text{ και } y = \varphi(v)$$

όπου η συνάρτηση v συμβολίζει την έξοδο των δεδομένων που είναι στην πραγματικότητα το αποτέλεσμα του αθροίσματος του γραμμικού συνδυασμού των $w_i x_i$ συν την αντίστοιχη πόλωση, το $\varphi(\cdot)$ συμβολίζει τη συνάρτηση ενεργοποίησης και τέλος η συνάρτηση y το σήμα της εξόδου του νευρώνα.



Εικόνα 13: Μοντέλο του Τεχνητού Νευρώνα

Η έξοδος λοιπόν ενός νευρώνα ορίζεται από τη συνάρτηση ενεργοποίησης $\varphi(\cdot)$, η οποία εξαρτάται από την v , τιμή ενεργοποίησης, όπως περιγράφεται στις εξισώσεις

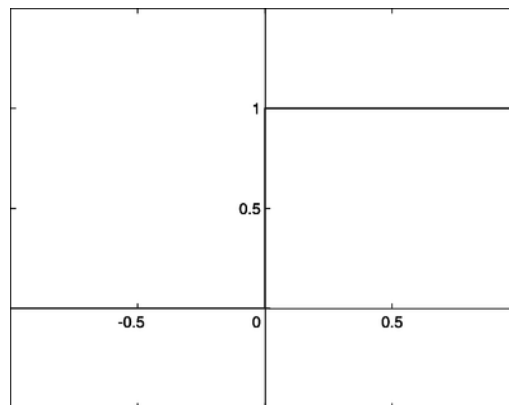
παρακάτω.

Υπάρχουν πολλές και διάφορες **συναρτήσεις ενεργοποίησης** προς χρήση, ενδεικτικά οι πιο κοινώς χρησιμοποιούμενες είναι οι εξής:

1. **Συνάρτηση Κατώφλι:** Η συνάρτηση αυτή ενεργοποίησης είναι η πιο απλή μορφή και ορίζεται με τον τύπο που ακολουθεί:

$$\varphi(v) = 1, \text{ εάν } v \geq 0 \text{ ή } \varphi(v) = 0, \text{ εάν } v < 0, \text{ όπου } v \text{ η τιμή ενεργοποίησης}$$

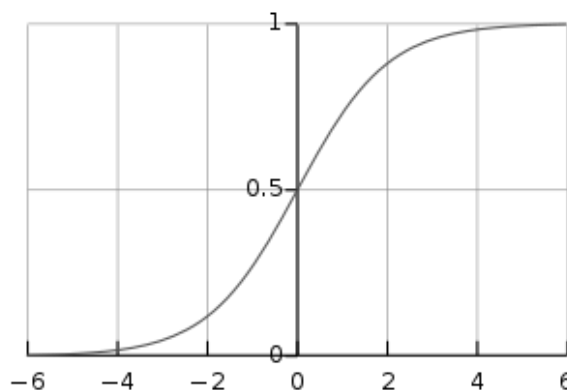
Σημειώνουμε ότι, η Συνάρτηση Κατώφλι μπορεί να συναντηθεί και με την ονομασία Heaviside.



Εικόνα 14: Γραφική αναπαράσταση της Συνάρτησης Κατώφλι

2. **Σιγμοειδής Συνάρτηση:** Η συνάρτηση αυτή είναι η πιο κοινή μορφή συνάρτησης ενεργοποίησης καθώς χρησιμοποιείται πολύ τον τελευταίο καιρό για την κατασκευή ΝΔ και ορίζεται με τον τύπο:

$$\varphi(v) = \frac{1}{1+e^{-bv}}$$



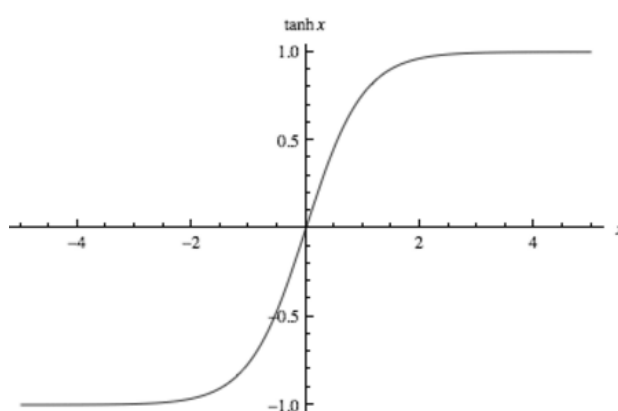
Εικόνα 15: Γραφική αναπαράσταση της Σιγμοειδούς Συνάρτησης

Οι τιμές της Σιγμοειδούς Συνάρτησης από ένα συνεχές πεδίο τιμών κυμαίνονται από 0 έως 1. Όταν οι τιμές εισόδου είναι πολύ μεγάλες και

αρνητικές τότε η έξοδος τείνει στο 0, ενώ αντίθετα εάν είναι πολύ μεγάλες και θετικές τότε τείνει στο 1. Η συνάρτηση αυτή παρουσιάζει δύο σημαντικές ιδιότητες οι οποίες και την ξεχωρίζουν από τις άλλες συναρτήσεις ενεργοποίησης, η πρώτη είναι ότι η συνάρτηση αυτή είναι διαφορίσιμη και αυτό έχει μεγάλη αξία στη θεωρία των ΝΔ και η δεύτερη είναι πως η κλίση της για πολύ μικρές ή πολύ μεγάλες αντίστοιχα τιμές εισόδου είναι σχεδόν μηδενική.

3. **Υπερβολική Εφαπτομένη:** Οι τιμές της εξόδου της συγκεκριμένη συνάρτησης κυμαίνονται στο διάστημα $(-1,1]$ και ο τύπος της ορίζεται ως εξής:

$$\varphi(v) = \tanh(v)$$

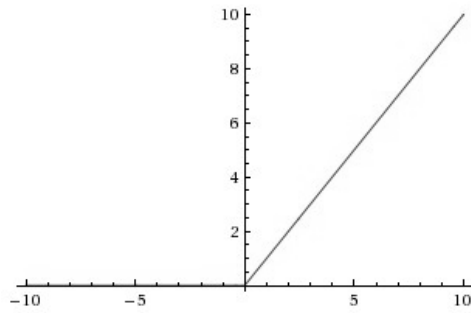


Εικόνα 16: Γραφική αναπαράσταση της Υπερβολικής Εφαπτομένης

Οι τιμές εξόδου της υπερβολικής εφαπτομένης είναι κεντραρισμένες γύρω από το μηδέν και αυτό το γεγονός καθιστά τη συγκεκριμένη συνάρτηση ενεργοποίησης καλύτερη ως προς τις προτιμήσεις των χρηστών από ότι τη Σιγμοειδή Συνάρτηση.

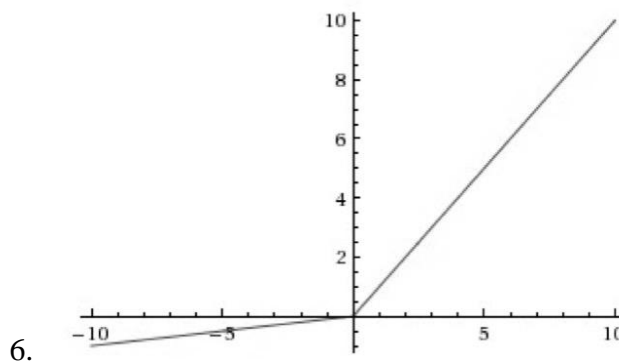
4. **Ανορθωμένη Γραμμική ή Συνάρτηση Ράμπας:** Η συνάρτηση αυτή αποτελεί μια κλασσική επιλογή για την εκπαίδευση των ΣΝΔ (βλ. Ενότητα 2.3.3.1.4), αφού η έξοδος της είναι μη γραμμική, γεγονός που βοηθάει στην επιτάχυνση της εκπαίδευσης. Παρόλα αυτά, η Συνάρτηση Ράμπας παρουσιάζει ένα σημαντικό μειονέκτημα, ορισμένες φορές μπορεί να οδηγήσει τους νευρώνες σε κάποιες τιμές βαρών που να μην ενεργοποιηθούν. Η συνάρτηση ράμπας περιγράφεται από την παρακάτω εξίσωση:

$$\varphi(v) = \max(0,v)$$



Εικόνα 17: Γραφική αναπαράσταση της Ανορθωμένης Γραμμικής ή Συνάρτησης Ράμπας

5. **Παραμετροποιημένη Συνάρτηση Ράμπας:** Η συγκεκριμένη συνάρτηση αποτελεί επίσης μια κλασσική επιλογή για την εκπαίδευση των ΣΝΔ (βλ. Ενότητα 2.3.3.1.4). Έχει ως στόχο την επιδιόρθωση του προβλήματος των νεκρών νευρώνων που μπορεί να προκαλέσει η συνάρτηση Ανορθωμένη Γραμμική ή Συνάρτηση Ράμπας πολλαπλασιάζοντας την εξόδο με μια μικρή τιμή α όταν η είσοδος είναι αρνητική.



Εικόνα 18: Γραφική αναπαράσταση Παραμετροποιημένης Συνάρτησης Ράμπας

Στην περίπτωση που η τιμή α είναι ίση με 0, τότε η συνάρτηση παίρνει τη μορφή της Ανορθωμένης Γραμμικής ή Συνάρτησης Ράμπας. Αν όμως το α πάρει μια μικρή και σταθερή τιμή, τότε η συνάρτηση θα μετονομαστεί σε μη Στεγανή Ανορθωμένη Γραμμική ή μη Στεγανή Συνάρτηση Ράμπα. Η Παραμετροποιημένη Συνάρτηση Ράμπας περιγράφεται από τον εξής τύπο:

$$\varphi(u) = u, \text{ εάν } u > 0 \text{ ή } \alpha u, \text{ αλλιώς}$$

2.3.3 Αρχιτεκτονικές και Πλεονεκτήματα των Τεχνητών Νευρωνικών Δικτύων

Τα βασικά χαρακτηριστικά που απαρτίζουν ένα ΝΔ είναι η αρχιτεκτονική του, οι λειτουργίες που επιτελεί και η μέθοδος εκπαίδευσής του. Η αρχιτεκτονική των ΝΔ

αποτελεί ένα αναπόσπαστο κομμάτι τους αφού καθορίζει τη διάταξη, τον αριθμό και τον τύπο των νευρώνων τους. Οι δύο βασικές κατηγορίες στις οποίες διαιρείται η αρχιτεκτονική των ΝΔ είναι: τα Δίκτυα Πρόσθιας Τροφοδότησης και τα Αναδρομικά Δίκτυα.

2.3.3.1 Δίκτυα Πρόσθιας Τροφοδότησης

Ένα ΝΔ Πρόσθιας Τροφοδότησης (ΝΔΠΤ) είναι ένα ΝΔ του οποίου η κατεύθυνση των πληροφοριών που μεταδίδει έχει αποκλειστικά μόνο μια φορά, την προς τα εμπρός, δηλαδή από τους κόμβους εισόδου, μέσω των κρυφών κόμβων, εάν υπάρχουν, μεταδίδονται στους κόμβους εξόδου. Παρακάτω παρουσιάζονται συνοπτικά τα διάφορα είδη ΝΔΠΤ.

2.3.3.1.1 Αισθητήρας Ενός Επιπέδου

Το πιο απλό είδος ΝΔ είναι το δίκτυο του Αισθητήρα ενός Επιπέδου (ΝΔΑιΕΕ) (βλ. Εικόνα 19), του οποίου η αρχιτεκτονική είναι πολύ απλή καθώς αποτελείται από ένα μόνο επίπεδο εξόδου και γι' αυτό οι εισοδοί του τροφοδοτούν απευθείας τις εξόδους του μέσω των βαρών.

Ένας ΝΔΑιΕΕ δέχεται πολλαπλά σήματα εισόδου, σε περίπτωση που το άθροισμα των σημάτων αυτών υπερβαίνει ένα συγκεκριμένο όριο κατωφλίου τότε είτε επιστρέφει κάποιο σήμα είτε παραμένει «σιωπηλό».

Η προσέγγιση του Frank Rosenblatt να εκπαιδεύσει τον Αισθητήρα με κανόνες οδήγησε στο να εισάγει τον αλγόριθμο της MM. Ο αλγόριθμος του Αισθητήρα μαθαίνει τα βάρη για τα σήματα εισόδου με σκοπό να σχεδιάσει γραμμικά όρια απόφασης που θα μας επιτρέπουν να διακρίνομε μεταξύ δύο γραμμικά διαχωρίσιμων κλάσεων +1 και -1.

Ο βασικός κανόνας του Rosenblatt για τον Αισθητήρα είναι αρκετά απλός και μπορεί να συνοψιστεί στα παρακάτω βήματα:

1. Ορίζει τις αρχικές τιμές των βαρών με την τιμή 0 ή με τυχαίες μικρές αριθμητικές τιμές,
2. Για κάθε δείγμα εκπαίδευσης $x^{(i)}$:
 - 1.1 Υπολογίζει την τιμή εξόδου και
 - 1.2 Ενημερώνει τα βάρη

Το αποτέλεσμα (τιμή εξόδου) είναι η ετικέτα της κλάσης που προβλέφθηκε από τη

συνάρτηση του μοναδιαίου βήματος την οποία ορίσαμε προηγουμένως ως (output =g(z)) και η ενημέρωση των βαρών μπορεί να τυποποιηθεί ως $w_j:=w_j+\Delta w_j$.

Η τιμή για την ενημέρωση των βαρών σε κάθε επαύξηση υπολογίζεται από τον κανόνα Εκμάθησης:

$$\Delta w_j = \eta(\text{target}^{(i)} - \text{output}^{(i)})x_j^{(i)}$$

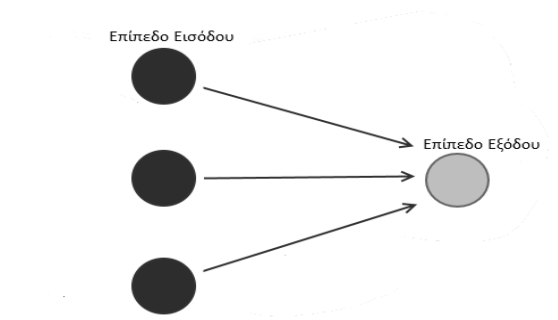
όπου το η είναι ο ρυθμός εκμάθησης, η οποία είναι μια σταθερά μεταξύ των τιμών 0,0 και 1,0, «target» είναι η πραγματική ετικέτα κλάσης και «output» είναι η προβλεπόμενη ετικέτα κλάσης.

Σημειώνεται ότι όλα τα βάρη στο διάνυσμα βάρους ενημερώνονται ταυτόχρονα. Πιο συγκεκριμένα, για ένα δισδιάστατο σύνολο δεδομένων, η ενημέρωσή τους γράφεται ως εξής:

$$\Delta w_0 = \eta(\text{target}^{(i)} - \text{output}^{(i)})$$

$$\Delta w_1 = \eta(\text{target}^{(i)} - \text{output}^{(i)})x_1^{(i)}$$

$$\Delta w_2 = \eta(\text{target}^{(i)} - \text{output}^{(i)})x_2^{(i)}$$



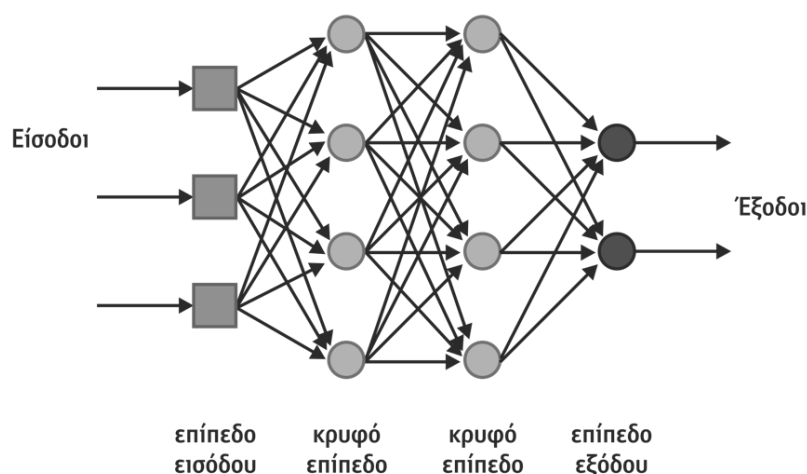
Εικόνα 19: Η αρχιτεκτονική του Αισθητήρα ενός Επιπέδου ή η δομή ενός απλού νευρώνα

2.3.3.1.2 Αισθητήρας Πολλών Επιπέδων

Αυτή η κατηγορία δικτύων, δηλαδή, των Αισθητήρων Πολλών Επιπέδων (ΑΙΠΕ), αποτελείται από πολλαπλά επίπεδα υπολογιστικών μονάδων, τα οποία συνήθως είναι διασυνδεδεμένα μέσω Πρόσθιας Τροφοδότησης. Η κατεύθυνση των συνδέσεων του κάθε νευρώνα είναι προς τους νευρώνες του επόμενου επιπέδου. Οι μονάδες των ΑΙΠΕ (βλ. Εικόνα 20) για να ενεργοποιηθούν σε ποικίλες εφαρμογές εφαρμόζουν μια Σιγμοειδή Συνάρτηση.

Για την εκπαίδευσή των ΑΙΠΕ χρησιμοποιούνται πολλές και διάφορες μέθοδοι. Από τις πιο δημοφιλείς μεθόδους ΜΜ που εφαρμόζονται είναι η Οπισθοδιάδοση. Σε αυτή τη μέθοδο, γίνεται μια σύγκριση των τιμών εξόδου, δηλαδή των πραγματικών

τιμών με την προβλεπόμενη τιμή, δηλαδή τη σωστή απάντηση, προκειμένου να είναι εφικτός ο υπολογισμός της τιμής κάποιας προκαθορισμένης συνάρτησης σφάλματος. Στη συνέχεια, μέσω διαφόρων τεχνικών, το σφάλμα που προκύπτει διοχετεύεται μέσω του δικτύου. Έτσι ο αλγόριθμος χρησιμοποιεί αυτές τις πληροφορίες και προσαρμόζει τα βάρη της κάθε σύνδεσης με σκοπό τη μείωση της τιμής της συνάρτησης σφάλματος. Εφόσον αυτή η διαδικασία εκπαίδευσης επαναληφθεί αρκετές φορές, το δίκτυο θα είναι σε θέση να συγκλίνει σε κάποια τιμή με το σφάλμα των υπολογισμών να είναι εμφανώς μικρότερο. Αναφορικά με τη σωστή ενημέρωση των βαρών, εφαρμόζεται μια γενική μέθοδος για τη μη γραμμική βελτιστοποίηση η οποία ονομάζεται βαθμολογημένη κλίση. Για τη βελτιστοποίηση αυτή το δίκτυο υπολογίζει την παράγωγο της συνάρτησης σφάλματος συναρτήσει των βαρών του δικτύου και αλλάζει τα βάρη με σκοπό τη μείωση του σφάλματος.



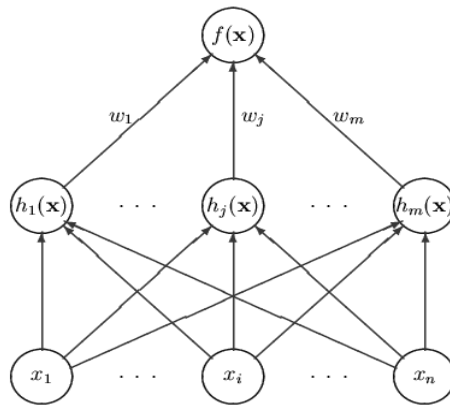
Εικόνα 20: Αισθητήρας πολλαπλών επιπέδων

Σημειώνεται ότι η εκπαίδευση ενός δικτύου για να αποδίδει σωστά ακόμη και με δεδομένα που δεν έχουν ξαναχρησιμοποιηθεί απαιτεί αρκετές επιπλέον τεχνικές. Επίσης, σε περιπτώσεις περιορισμένου αριθμού δεδομένων υπάρχει ο κίνδυνος το δίκτυο να μην εκπαιδευτεί σωστά αφού λόγω των περιορισμένων δεδομένων δεν θα καταγράφεται η πραγματική στατιστική διαδικασία.

2.3.3.1.3 Δίκτυα Ακτινικών Συναρτήσεων

Τα Δίκτυα Ακτινικών Συναρτήσεων (ΔΑΣ) αποτελούν μια ιδιαίτερη κατηγορία ΝΔ (βλ. Εικόνα 21).

Η αρχιτεκτονική αυτών των δικτύων είναι αρκετά απλή καθώς αποτελούνται μόνο από ένα επίπεδο εισόδου, ένα κρυφό επίπεδο με μια μη γραμμική Συνάρτηση



Εικόνα 21: Αναπαράσταση ενός ΔΑΣ

Ενεργοποίησης και ένα επίπεδο εξόδου. Τα δίκτυα αυτά εφαρμόζονται σε ποικίλους τομείς, με τους πιο σημαντικούς και άξιους αναφοράς να είναι η προσέγγιση λειτουργιών, η πρόβλεψη χρονοσειρών, η ταξινόμηση και ο έλεγχος του συστήματος.

Ένα χαρακτηριστικό γνώρισμα των ΔΑΣ είναι ότι παρουσιάζουν μια ανομοιογένεια στον τρόπο λειτουργίας τους, καθώς στο κρυφό επίπεδο η συνάρτηση ενεργοποίησης είναι ακτινικού τύπου και οι παράμετροι προς εκπαίδευση είναι τα κέντρα c_i και τα εύροι σ_i των κρυφών νευρώνων, όπου $i = 1, 2, \dots, K$. Όσον αφορά για το επίπεδο εξόδου, η συνάρτηση ενεργοποίησης είναι γραμμική και οι παράμετροι προς εκπαίδευση είναι στην ουσία οι τιμές των συναπτικών βαρών w_{ij} . Εξαιτίας αυτής της ανομοιογένειας το πρόβλημα της εκπαίδευσης του κρυφού επιπέδου αντιμετωπίζεται ξεχωριστά από αυτό της εκπαίδευσης του επιπέδου εξόδου.

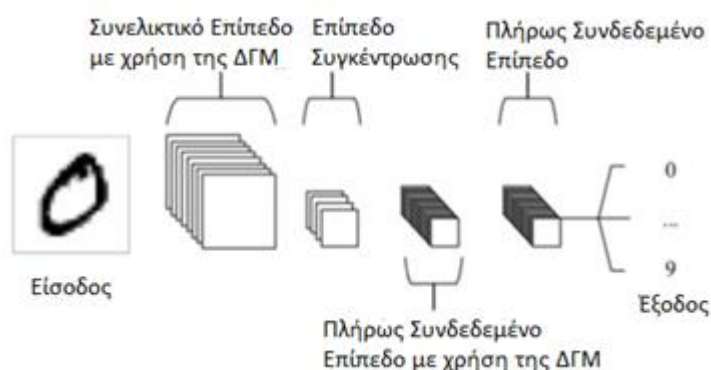
2.3.3.1.4 Συνελικτικά Νευρωνικά Δίκτυα

Μια από τις πιο εντυπωσιακές μορφές αρχιτεκτονικής ΝΔ είναι τα ΣΝΔ. Τα δίκτυα αυτά χρησιμοποιούνται κυρίως για την επίλυση πολύπλοκων προβλημάτων. Η διαφορά των ΣΝΔ και των κλασικών ΝΔ είναι ότι χρησιμοποιούνται κυρίως στο πεδίο της Αναγνώρισης Προτύπων, όπως κατά κύριο λόγο είναι οι εικόνες και μόλις πρόσφατα άρχισαν να εφαρμόζονται και στην ΕΦΓ που ήδη τα ΝΔ ξεχωρίζουν. Το γεγονός αυτό επιτρέπει στον χρήστη να κωδικοποιεί τις ιδιαιτερότητες του προτύπου - εικόνας ή κειμένου - στην αρχιτεκτονική του δικτύου, καθιστώντας το κατάλληλο για την επίλυση τέτοιων προβλημάτων. Δηλαδή, τα εισερχόμενα διανύσματα, του εκάστοτε υπό μελέτη προτύπου, μέχρι και την τελική έξοδο, που είναι η βαθμολογία/αξιολόγηση του σε ποσοστό επί τις 100, το δίκτυο θα εξακολουθεί να εκφράζει μια ενιαία συνειδητή συνάρτηση βαθμολόγησης/αξιολόγησης, μειώνοντας τις παραμέτρους που απαιτούνται για την ομαλή λειτουργία του μοντέλου με τη χρήση

διάφορων τεχνικών.

Την τελευταία δεκαετία, οι πρόοδοι που σημειώθηκαν στον τομέα της Υπολογιστικής Όρασης είναι πραγματικά εκπληκτικές και άνευ προηγουμένου. Οι μηχανές μπορούν πλέον να αναγνωρίζουν εικόνες και καρέ σε βίντεο με ακρίβεια (98%) που υπερβαίνει το ποσοστό των ανθρώπων (97%). Παλαιότερα, οι νευρολόγοι διεξήγαγαν πειράματα σε γάτες όταν ανακάλυψαν ότι τα παρόμοια μέρη μιας εικόνας μπορούν και προκαλούν την ενεργοποίηση των αντίστοιχων μερών του εγκεφάλου της γάτας (άλφα ζώνη). Τα αποτελέσματα τους λοιπόν κατέληξαν στο συμπέρασμα ότι ο εγκέφαλος του ζώου περιέχει μια ζώνη νευρώνων που αντιδρά στα ιδιαίτερα χαρακτηριστικά μιας εικόνας. Εμπνευσμένοι από τη λειτουργία του εγκεφάλου, οι μαθηματικοί δημιούργησαν ένα πρότυπο για την προσομοίωση διαφορετικών ομάδων νευρώνων που ενεργοποιούνται όταν αναγνωρίζουν μια διαφορετική όψη της δοθείσας εικόνας και επικοινωνούν μεταξύ τους για να σχηματίσουν μια μεγαλύτερη εικόνα.

Η αρχιτεκτονική των ΣΝΔ κατασκευάζεται έτσι ώστε να μπορούν να αντιμετωπίσουν τους συγκεκριμένους τύπους δεδομένων με αποτελεσματικό τρόπο. Τα ΣΝΔ αποτελούνται από τρεις τύπους επιπέδων, τα Συνελκτικά Επίπεδα, τα Επίπεδα Συγκέντρωσης και τα Πλήρως Συνδεδεμένα Επίπεδα. Στο Σχήμα της Εικόνας 22 που ακολουθεί περιγράφεται μια απλοποιημένη αρχιτεκτονική ενός ΣΝΔ για την ταξινόμηση της βάσης δεδομένων MNIST (βλ. ενότητα 1.4.5, ιδιότητα 1.4.5.2).



Εικόνα 22: Μια απλοποιημένη αρχιτεκτονική Συνελκτικού Νευρωνικού Δικτύου που αποτελείται από 5 επίπεδα

Η βασική διαδικασία του παραπάνω παραδείγματος μπορεί να αναλυθεί ως εξής:

1. Το **Επίπεδο Εισόδου**: θα περιέχει τις τιμές των εικονοστοιχείων της εικόνας.
2. Το **Συνελκτικό Επίπεδο**: θα καθορίσει την έξοδο των νευρώνων που συνδέονται με τοπικές περιοχές της εισόδου υπολογίζοντας τη βαθμωτή τιμή των βαρών τους και

της περιοχής που συνδέεται με το Επίπεδο Εισόδου. Σημαντικό ρόλο στα ΣΝΔ παίζει η Διορθωμένη Γραμμική Μονάδα (ΔΓΜ) η οποία εφαρμόζει μια στοιχειώδη Σιγμοειδή Συνάρτηση Ενεργοποίησης στην παραγόμενη από το προηγούμενο επίπεδο έξοδο ενεργοποίησης και συντελεί στην περαιτέρω βελτίωση της απόδοσης του μοντέλου.

3. Το **Επίπεδο Συγκέντρωσης**: θα πραγματοποιήσει στη συνέχεια μια Υποδειγματοληψία κατά μήκος της χωρικής διαστασιολογίας της εισόδου, μειώνοντας σημαντικά τον αριθμό των παραμέτρων/διαστάσεων.

4. Τα **Πλήρως Συνδεδεμένα Επίπεδα**: θα εκτελέσουν τις ίδιες διεργασίες που εκτελούνται και στα κλασσικά ΝΔ και θα προσπαθήσουν να παράξουν κλάσεις βαθμολογιών για την ταξινόμηση. Αξίζει να σημειωθεί ότι και σε αυτό το επίπεδο συστήνεται η χρήση της ΔΓΜ.

Παρά την μεγάλη επιτυχία των τυπικών ΝΔ σε εφαρμογές ποικίλων τεχνολογικών τομέων, εμπόδιο στην εξέλιξη τους στάθηκε η αύξηση της πολυπλοκότητας των δεδομένων εισόδου, καθώς ο μεγάλος αριθμός των δεδομένων εισόδου απαιτούσε και πολύ μεγαλύτερα ΝΔ με τεράστιους αριθμούς συνδέσεων και νευρώνων. Ειδικότερα για προβλήματα Αναγνώρισης Εικόνας, οι πρώτες προσπάθειες που έγιναν για την εξαγωγή των χαρακτηριστικών τους όπως χρώμα, υφή, συχνότητα και ακμές με χρήση των ΝΔ, αποδείχθηκε ότι δεν ήταν επιτυχείς. Για το λόγο αυτό προβλήματα με πολλές διαστάσεις στα δεδομένα εισόδου αντιμετωπίστηκαν μέσω των τεχνικών ΜΒ και πιο συγκεκριμένα μέσω των ΣΝΔ.

Ο όρος ΣΝΔ χρησιμοποιήθηκε για να περιγράψει την αρχιτεκτονική των ΝΔ τα οποία εφαρμόζονται σε δισδιάστατους πίνακες (κυρίως εικόνες) και βασίζονται σε μια χωρικά οριοθετημένη νευρωνική είσοδο. Η συγκεκριμένη αρχιτεκτονική έχει αναφερθεί και ως η τεχνική των μοιρασμένων βαρών ή των τοπικών πεδίων υποδοχής. Οι LeCun και Bengio σημειώνουν τρεις αρχιτεκτονικές αρχές που είναι κοινές σε όλα τα ΣΝΔ:

1. Τα **τοπικά πεδία υποδοχής**,
2. Τα **κοινά βάρη** και
3. Η **χωρική υποδειγματοληψία**.

Τέλος, αναφέρονται ευρέως γνωστά ΣΝΔ τα οποία χρησιμοποιούνται στο πεδίο της ΜΜ και επιλέγονται ανάλογα με τον τύπο του προβλήματος που επιδέχεται λύση:

1. **LeNet**: Αποτελεί την πρώτη επιτυχημένη εφαρμογή ΣΝΔ που κατασκευάστηκε από τον Yann LeCun την δεκαετία του 1990. Η συγκεκριμένη αρχιτεκτονική

χρησιμοποιήθηκε κυρίως για την αναγνώριση κωδικών, ψηφίων κ.λπ..

2. **AlexNet:** Το συγκεκριμένο ήταν το πρώτο δίκτυο που έκανε τα ΣΝΔ διάσημα στο χώρο της Υπολογιστικής Όρασης. Το δίκτυο αυτό είχε μια αρκετά παρόμοια αρχιτεκτονική με αυτή του LeNet, ωστόσο, ήταν βαθύτερο, μεγαλύτερο και αποτελούταν από πολλά συνελκτικά επίπεδα, στοιβαγμένα το ένα πάνω στο άλλο. Το δίκτυο αυτό αποτέλεσε μια πρωτοποριακή τεχνική την εποχή που κατασκευάστηκε.

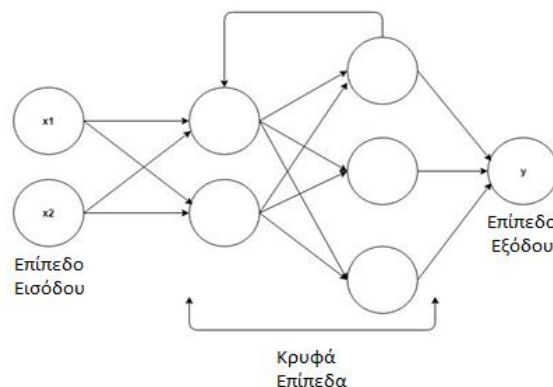
3. **ZF Net:** Το δίκτυο αυτό κατασκευάστηκε για να βελτιώσει το AlexNet, διορθώνοντας κάποιες υπερπαραμέτρους της αρχιτεκτονικής του AlexNet και πιο συγκεκριμένα επεκτείνοντας το μέγεθος του μεσαίου συνελκτικού επιπέδου.

4. **GoogleNet:** Το συγκεκριμένο δίκτυο αναπτύχθηκε από την Google. Το βασικό πλεονέκτημα του έναντι των προηγούμενων μοντέλων έγκειται στην δραματική μείωση των παραμέτρων του δικτύου. Ειδικότερα, ο αριθμός των παραμέτρων μειώθηκε κατά 4 εκατομμύρια σε σύγκριση με το AlexNet που διέθετε 60 εκατομμύρια παραμέτρους.

5. **VGGNet:** Το δίκτυο αυτό, αν και απέσπασε τη δεύτερη θέση στο διαγωνισμό ILSVRC 2014, ήταν εξίσου αποτελεσματικό με το GoogleNET το οποίο ήταν ο νικητής στο διαγωνισμό. Το δίκτυο αυτό εισήγαγε μια καινοτομία στην αρχιτεκτονική του, το βάθος, το οποίο αποδείχθηκε στη συνέχεια ότι συντέλεσε στην πολύ καλή απόδοση του ΣΝΔ.

6. **ResNet:** Το δίκτυο αυτό ήταν το νικητήριο δίκτυο στον διαγωνισμό ILSVRC 2015. Εισήγαγε για πρώτη φορά τις παραλειπόμενες συνδέσεις και την εκτεταμένη χρήση της κανονικοποίησης. Είναι αξιοσημείωτο επίσης το γεγονός ότι η αρχιτεκτονική του δικτύου αυτού δεν περιλάμβανε κανένα πλήρες συνδεδεμένο επίπεδο.

2.3.3.2 Επαναληπτικά Νευρωνικά Δίκτυα



Εικόνα 23: Αναπαράσταση ενός Επαναληπτικού Νευρωνικού Δικτύου

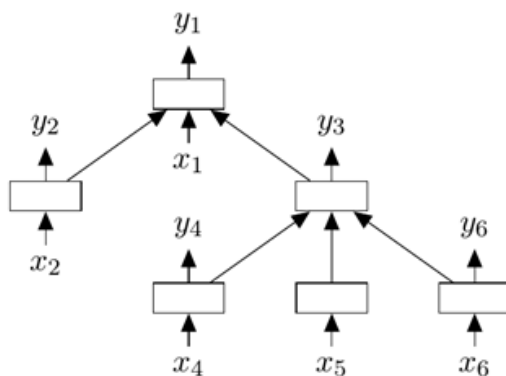
Τα ΕΝΔ (βλ. Εικόνα 23) αποτελούν μια οικογένεια ΝΔ που εξειδικεύεται στην επεξεργασία διαδοχικών δεδομένων, δηλαδή στην επεξεργασία μιας ακολουθίας $x^{(1)}, \dots, x^{(r)}$.

Η πρωταρχική ιδέα που ώθησε την επέκταση των ΑιΠΕ στα ΕΝΔ ξεκίνησε από τη ΜΜ και τα στατιστικά μοντέλα τη δεκαετία του 1980 αφού ο κύριος προβληματισμός τους ήταν η εναλλαγή των παραμέτρων στα διάφορα τμήματα του προτύπου. Η κοινή χρήση λοιπόν των παραμέτρων έθεσε τις βάσεις για την επέκταση και εφαρμογή του προτύπου αυτού σε προβλήματα διαφόρων μορφών και τη γενίκευση σε αυτά.

Το βασικό γνώρισμα που κάνει τα ΕΝΔ ξεχωρίζουν έναντι των τυπικών Δικτύων Πρόσθιας Τροφοδότησης είναι ότι συμπεριλαμβάνουν στην αρχιτεκτονική τους τουλάχιστον μια ανάδραση ανάμεσα στους κόμβους των επιπέδων τους. Για το λόγο αυτό τα ΕΝΔ μπορούν να θεωρηθούν ως δυναμικά συστήματα τα οποία μπορούν να μοντελοποιούν και να αναπαριστούν μοντέλα που μεταβάλλονται στο χρόνο. Τα δίκτυα αυτά μπορούν να χρησιμοποιηθούν αποδοτικά σε πολλούς και διάφορους τομείς όπως για: την ταξινόμηση των δεδομένων σε κλάσεις, το φιλτράρισμα, τις τηλεπικοινωνίες, την αναγνώριση ομιλίας, την ρομποτική, την μετάφραση. Μόλις πρόσφατα όπως θα δούμε στο Κεφάλαιο 3 άρχισε να δοκιμάζεται η απόδοση τους στον τομέα της ΕΦΓ και τα αποτελέσματα αυτού φάνηκαν να είναι ελπιδοφόρα.

2.3.3.2.1 Αναδρομικά Νευρωνικά Δίκτυα

Τα Αναδρομικά Νευρωνικά Δίκτυα (ΑΝΔ) (βλ. Εικόνα 24) αποτελούν μια γενίκευση των ΕΝΔ και αποτελούνται από ένα διαφορετικό είδος υπολογιστικού γραφήματος, το οποίο είναι δομημένο ως ένα βαθύ δέντρο γι' αυτό και η δομή του διαφέρει από την αλυσιδωτή δομή η οποία εμφανίζεται στα ΕΝΔ.



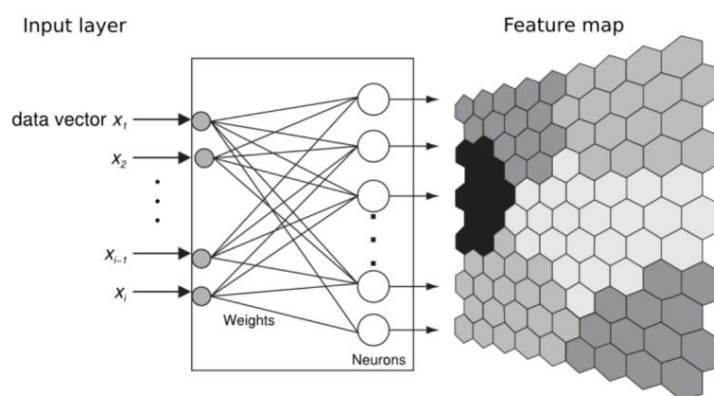
Εικόνα 24: Αναδρομικό Νευρωνικό Δίκτυο

Τα ΑΝΔ εισήχθησαν από τον Pollack (1990) και η πιθανή χρήση τους για εκμάθηση

στη λογική περιγράφηκε από τον Bottou (2011). Τα ANΔ έχουν εφαρμοστεί επιτυχώς στην επεξεργασία των δομών δεδομένων (Frasconi et al., 1997, 1998), στην ΕΦΓ (Socher et al., 2011, 2013), καθώς και στην Υπολογιστική Όραση (Socher, 2011). Ένας σαφές πλεονέκτημα που προκύπτει από τα ANΔ έναντι των ENΔ είναι ότι για μια ακολουθία μήκους τ , το βάθος (που μετριέται ως ο αριθμός των συνθέσεων των μη γραμμικών λειτουργιών) μπορεί να μειωθεί από τ σε $O(\log \tau)$, γεγονός που μπορεί να βοηθήσει στην αντιμετώπιση μακροπρόθεσμων εξαρτήσεων.

2.3.3.2 Δίκτυα Kohonen

Τα δίκτυα Kohonen ή Αυτοργανωτικοί Χάρτες Χαρακτηριστικών, είναι δίκτυα δύο διαστάσεων και αποτελούν μια υπολογιστική μέθοδο για την απεικόνιση και ανάλυση των δεδομένων μεγάλης διάστασης, ειδικά των πληροφοριών που αποκτήθηκαν πειραματικά (βλ. Εικόνα 25). Επίσης, ορίζουν μια ταξινομημένη χαρτογράφηση, ένα είδος προβολής δηλαδή από ένα σύνολο στοιχείων δεδομένων σε ένα κανονικό, συνήθως δισδιάστατο πλέγμα. Ένα στοιχείο λοιπόν σε αυτά τα δίκτυα θα χαρτογραφηθεί στον κόμβο του οποίου το μοντέλο είναι περισσότερο παρόμοιο με το στοιχείο δεδομένων, δηλαδή, θα έχει τη μικρότερη απόσταση από το στοιχείο δεδομένων προς κάποια μετρική.



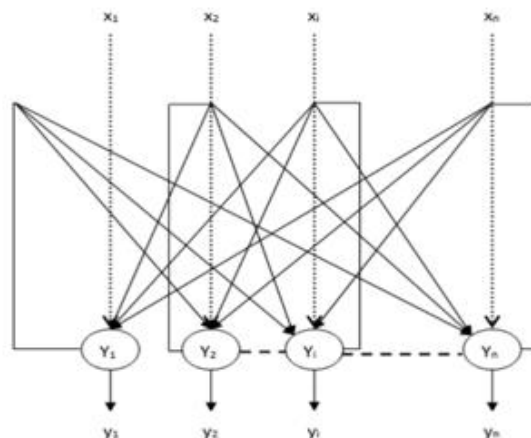
Εικόνα 25: Δίκτυο Kohonen ή Αυτοργανωτικοί Χάρτες Χαρακτηριστικών

Τα δίκτυα αυτά εκπαιδεύονται κατά τρόπο MEMM ώστε να σχηματίζουν συστάδες εντός των δεδομένων, δηλαδή, ομαδοποίηση δεδομένων.

2.3.3.3 Δίκτυα Hopfield

Ένα δίκτυο Hopfield είναι μια μορφή ANΔ που κατασκευάστηκε από τον John Hopfield το 1982, είχε όμως περιγραφεί νωρίτερα το 1974 από τον Little. Τα δίκτυα

αυτά χρησιμεύουν ως συστήματα διευθυνσιοδοτούμενης από το περιεχόμενο μνήμης με δυαδικούς κόμβους της Συνάρτησης Κατώφλι. Τα Δίκτυα Hopfield συγκλίνουν τοπικά ελάχιστα και ως εκ τούτου, μπορεί να συγκλίνουν σε ένα ψεύτικο πρότυπο, δηλαδή λανθασμένο τοπικό ελάχιστο, και όχι στο αποθηκευμένο πρότυπο, δηλαδή το αναμενόμενο τοπικό ελάχιστο. Τέλος, τα δίκτυα Hopfield αποτελούν ένα μοντέλο για την κατανόηση της ανθρώπινης μνήμης (βλ. Εικόνα 26).



Εικόνα 26: Δίκτυο Hopfield

2.3.3.3 Πλεονεκτήματα των Τεχνητών Νευρωνικών Δικτύων

Η συμπαγής παράλληλα κατανεμημένη δομή και η ικανότητα εκπαίδευσης των ΝΔ, καθιστούν δυνατή την επίλυση πολύπλοκων προβλημάτων, τα οποία διαιρούνται σε μικρότερα υποπροβλήματα που καλούνται να επιλύσουν τα ΝΔ ανάλογα με τις δυνατότητες που έχουν.

Η χρήση των ΝΔ προσδίδει τις ακόλουθες χρήσιμες ιδιότητες :

1. **Μη-γραμμικότητα:** Αυτή οφείλεται στο γεγονός ότι η δομή ενός ΝΔ πραγματοποιείται μέσω της σύνδεσης των νευρώνων που περιέχονται σε αυτό, οι οποίοι νευρώνες είναι μη-γραμμικές συσκευές. Η μη-γραμμικότητα είναι πολύ σημαντική ιδιότητα, ιδιαίτερα αν ο φυσικός μηχανισμός για την παραγωγή των σημάτων εισόδου είναι μη-γραμμικός.
2. **Αντιστοίχιση Εισόδου-Εξόδου:** Ένα συνηθισμένο παράδειγμα Εκμάθησης που χρησιμοποιεί την Επιτηρούμενη Εκμάθηση, στηρίζεται στη μεταβολή των συναπτικών βαρών των ΝΔ, εφαρμόζοντας ένα σύνολο παραδειγμάτων εκπαίδευσης. Κάθε παράδειγμα αποτελείται από ένα μοναδικό σήμα εισόδου και μια αντίστοιχη επιθυμητή απόκριση. Η εκπαίδευση του δικτύου επαναλαμβάνεται στη συνέχεια για πολλά παραδείγματα, μέχρι το δίκτυο να

φτάσει σε μια σταθερή κατάσταση, όπου πλέον δεν μπορούν να γίνουν άλλες μεταβολές στα βάρη. Αυτό έχει σαν αποτέλεσμα το δίκτυο να μάθει μέσα από τα παραδείγματα, κατασκευάζοντας ένα πρότυπο εισόδου-εξόδου για πρόβλημα που πιθανόν να μην μπορεί να υλοποιηθεί αυτόματα.

3. **Προσαρμοστικότητα:** Τα ΝΔ έχουν τη δυνατότητα να προσαρμόζουν τα συναπτικά τους βάρη ανάλογα με τις μεταβολές που γίνονται στο περιβάλλον τους. Ένα ΝΔ που έχει εκπαιδευτεί να λειτουργεί σε ένα συγκεκριμένο περιβάλλον, μπορεί εύκολα να επανεκπαιδευτεί ώστε να μπορεί να χειρίζεται τις μεταβολές στο νέο περιβάλλον λειτουργίας του.
4. **Παράλληλος τρόπος λειτουργίας:** Τα ΝΔ λειτουργούν με παράλληλο τρόπο διότι ένα πρόβλημα μοιράζεται στα διάφορα τμήματα του δικτύου, δηλαδή σε όλους τους επιμέρους νευρώνες. Γι' αυτόν τον λόγο τα ΝΔ είναι συστήματα «παράλληλων κατανεμημένων διεργασιών». Το γεγονός αυτό έχει σαν αποτέλεσμα τα ΝΔ να λειτουργούν με μεγάλες ταχύτητες καθώς είναι σαν να έχουμε ταυτόχρονα πολλούς επεξεργαστές.
5. **Συναφής Πληροφορία:** Η γνώση αναπαριστάνεται από την πολύ δομημένη και ενεργή κατάσταση των ΝΔ.
6. **Αντοχή σε σφάλματα:** Ένα ΝΔ υλοποιημένο σε μηχανική διάσταση έχει τη σημαντική ιδιότητα να είναι ανεκτικό στην αντιμετώπιση σφαλμάτων. Δηλαδή, κάτω από αντίξοες συνθήκες λειτουργίας του, αυτή δεν διακόπτεται, αλλά μειώνει την απόδοσή του βαθμιαία και ομαλά. Έτσι για παράδειγμα αν ένας νευρώνας ή οι συνδέσεις του καταστραφούν, τότε η ποιότητα της πληροφορίας της εξόδου απλά θα μειωθεί. Ωστόσο, για να αρχίσει να μειώνεται σοβαρά η συνολική απόδοση του δικτύου, η βλάβη θα πρέπει να λάβει μεγάλη έκταση και αυτό γιατί η πληροφορία που αποθηκεύεται στο δίκτυο είναι κατανεμημένη σε όλη τη δομή του.
7. **Δυνατότητα Υλοποίησης σε VLSI¹³:** Η συμπαγής παράλληλη φύση του ΝΔ, καθιστά δυνατή την υλοποίησή του σε VLSI τεχνολογία, έτσι ώστε τα ΝΔ να μπορούν να χρησιμοποιηθούν σε εφαρμογές πραγματικού χρόνου.
8. **Ομοιομορφία Ανάλυσης και Σχεδιασμού:** Η ιδιότητα αυτή έγκειται ότι σε όλα

¹³Η τεχνολογία VLSI (very-large-scale integration ή αλλιώς πολύ μεγάλης κλίμακας ολοκλήρωση) είναι η διαδικασία δημιουργίας ενός ολοκληρωμένου κυκλώματος η οποία κατέχει την ικανότητα να συνδυάζει εκατοντάδες χιλιάδες τρανζίστορ ή συσκευές σε ένα ενιαίο τσιπ. Η τεχνολογία αυτή ξεκίνησε τη δεκαετία του 1970 όταν αναπτύχθηκαν σύνθετες τεχνολογίες ημιαγωγών και επικοινωνιών. Χαρακτηριστικό παράδειγμα μιας τεχνολογίας VLSI είναι ο μικροεπεξεργαστής.

τα πεδία εφαρμογής των ΝΔ με κοινό συστατικό τους νευρώνες χρησιμοποιείται ο ίδιος συμβολισμός,. Αυτή η προσέγγιση επιτρέπει τη χρήση των ίδιων θεωριών και αλγορίθμων εκμάθησης σε διαφορετικές εφαρμογές των ΝΔ.

9. **Αναλογία με τη Νευροβιολογία:** Είναι γνωστό ότι η σχεδίαση των ΝΔ δανείζεται στοιχεία από τη λειτουργία του ανθρώπινου εγκεφάλου, ο οποίος είναι η ζωντανή απόδειξη ότι η εύρωστη, παράλληλη επεξεργασία δεν είναι μόνο φυσικά εφικτή, αλλά επίσης γρήγορη και ισχυρή. Για τον λόγο αυτόν, τα ΝΔ, όπως και τα βιολογικά ΝΔ, έχουν μεγάλη ανοχή σε σφάλματα. Με άλλα λόγια, η κακή λειτουργία ή η καταστροφή ενός νευρώνα ή κάποιων συνδέσεων δεν είναι ικανή να διαταράξει σημαντικά τη λειτουργία τους καθώς η πληροφορία που εσωκλείουν δεν είναι εντοπισμένη σε κάποιο συγκεκριμένο σημείο αλλά είναι διάχυτη σε όλο το δίκτυο (βλ ιδιότητα 6).

2.3.4 Διαδικασίες Εκμάθησης των Νευρωνικών Δικτύων

2.3.4.1 Εισαγωγή

Η διαδικασία Εκμάθησης των ΝΔ είναι μια πολύ σημαντική διαδικασία για την αποτελεσματική και ομαλή λειτουργία τους. Μόλις το δίκτυο δομηθεί για να επιλύσει ένα συγκεκριμένο πρόβλημα τότε αυτό είναι έτοιμο να εκπαιδευτεί. Για να ξεκινήσει αυτή η διαδικασία θα πρέπει τα αρχικά βάρη να επιλεγούν τυχαία και στη συνέχεια να αρχίσει η διαδικασία της Εκμάθησης.

Υπάρχουν δύο προσεγγίσεις για τη διαδικασία της Εκμάθησης, η πρώτη είναι η Επιβλεπόμενη Εκμάθηση ή Εκμάθηση Με Επίβλεψη (EM) και η δεύτερη η Μη Επιβλεπόμενη Εκμάθηση ή Εκμάθηση Χωρίς Επίβλεψη (MEM). Η EM περιλαμβάνει ένα μηχανισμό όπου παρέχεται στο δίκτυο η επιθυμητή έξοδος είτε μέσω χειροκίνητης «ταξινόμησης» της απόδοσης του δικτύου είτε παρέχοντας χειροκίνητα «τις επιθυμητές εξόδους και εισόδους». Από την άλλη, στη MEM θα πρέπει το δίκτυο από μόνο του να ορίσει τις εισόδους χωρίς κάποια εξωτερική βοήθεια. Αναλυτικά η EM και MEM παρουσιάζονται στις παρακάτω ενότητες.

2.3.4.2 Επιβλεπόμενη Εκμάθηση

Στην EM, παρέχονται τόσο οι εισόδοι όσο και οι εξόδοι στο δίκτυο. Το δίκτυο στη συνέχεια επεξεργάζεται τις εισόδους και συγκρίνει τις προκύπτουσες εξόδους με τις

επιθυμητές εξόδους του. Τα σφάλματα μεταδίδονται προς τα πίσω μέσω του συστήματος, ρυθμίζοντας τα βάρη που ελέγχουν το δίκτυο. Αυτή η διαδικασία επαναλαμβάνεται τόσο όσο τα βάρη τροποποιούνται. Το σύνολο των δεδομένων που επιτρέπει την εκπαίδευση του συστήματος ονομάζεται “σύνολο εκπαίδευσης”.

Εάν ένα δίκτυο δεν μπορεί να λύσει το ζητούμενο πρόβλημα, τότε ο σχεδιαστής του θα πρέπει να αναθεωρήσει:

1. τις εισόδους και τις εξόδους,
2. τον αριθμό των επιπέδων,
3. τον αριθμό των νευρώνων ανά επίπεδο,
4. τις συνδέσεις μεταξύ των επιπέδων,
5. τις συναρτήσεις ενεργοποίησης, μεταφοράς και εκπαίδευσης και
6. τα ίδια τα αρχικά βάρη.

Οι αλλαγές που απαιτούνται για τη δημιουργία ενός επιτυχημένου δικτύου αποτελούν μια διαδικασία στην οποία φαίνεται ξεκάθαρα η “τέχνη” της νευρωνικής δικτύωσης.

Ένα άλλο μέρος της δημιουργικότητας που θα πρέπει να έχει ο σχεδιαστής αφορά τους κανόνες Εκμάθησης. Υπάρχουν πολλοί κανόνες (αλγόριθμοι) που χρησιμοποιούνται για την εφαρμογή της προσαρμοστικής ανάδρασης που απαιτείται για την ενημέρωση των βαρών κατά τη διάρκεια της εκπαίδευσης. Ο πιο συνηθισμένος αλγόριθμος που χρησιμοποιείται για την ενημέρωση των βαρών στη διαδικασία της εκπαίδευσης είναι αυτός της Οπισθοδιάδοσης.

Ωστόσο, η εκπαίδευση δεν είναι απλά μια τεχνική, αλλά χρειάζεται ο σχεδιαστής του ΝΔ να έχει μια «αίσθηση» και «συνειδητή ανάλυση», προκειμένου να διασφαλιστεί ότι το δίκτυο δεν είναι «υπερβολικά εκπαιδευμένο». Αρχικά, ένα ΝΔ διαμορφώνεται σύμφωνα με τις γενικές στατιστικές τάσεις των δεδομένων του και αργότερα, συνεχίζει να «μαθαίνει» για άλλες πτυχές των δεδομένων του, οι οποίες μπορεί και να είναι ψευδείς από γενικής άποψης.

Όταν τελικά το σύστημα έχει εκπαιδευτεί σωστά και δεν χρειάζεται κάποια περαιτέρω εκπαίδευση, τότε τα βάρη μπορούν, εάν είναι επιθυμητό, να “παγώσουν”.

2.3.4.3 Μη Επιβλεπόμενη Εκμάθηση

Στη MEM παρέχονται στο δίκτυο οι εισοδοί αλλά όχι οι επιθυμητές εξοδοί. Εδώ, το ίδιο το σύστημα θα πρέπει να αποφασίσει ποια χαρακτηριστικά θα χρησιμοποιήσει για

να ομαδοποιήσει τα δεδομένα εισόδου, το οποίο συχνά αναφέρεται ως αυτό-οργάνωση ή προσαρμογή.

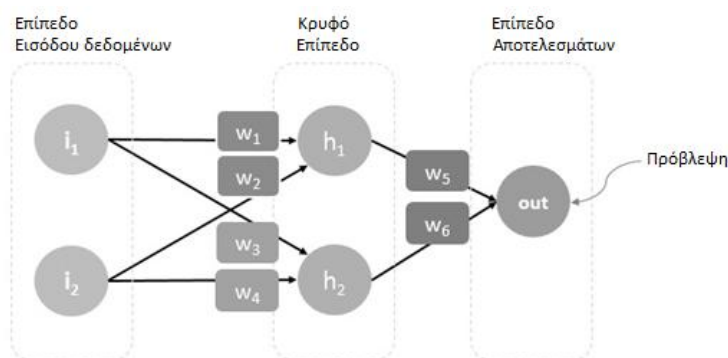
Προς το παρόν, η συγκεκριμένη Εκμάθηση δεν είναι ευρέως κατανοητή. Αυτή η δυνατότητα της Εκμάθησης, δηλαδή να προσαρμόζεται στο περιβάλλον της θα έχει άμεση επιρροή στα ρομπότ του μέλλοντος και θα τα επιτρέψει να μαθαίνουν συνεχώς από μόνα τους, καθώς θα συναντούν νέες καταστάσεις και νέα περιβάλλοντα.

Ένας από τους κορυφαίους ερευνητές στη MEM είναι ο Tueeno Kohonen, ηλεκτρολόγος μηχανικός στο Τεχνολογικό Πανεπιστήμιο του Ελσίνκι, ο οποίος έχει αναπτύξει ένα αυτό-οργανωτικό δίκτυο, που μαθαίνει από μόνο του χωρίς να έχει το όφελος της γνώσης της σωστής απάντησης. Το συγκεκριμένο δίκτυο είναι ασυνήθιστο ως προς την αρχιτεκτονική του καθώς περιέχει μόνο ένα επίπεδο με πολλές συνδέσεις, όπου τα βάρη των συνδέσεων αυτών θα πρέπει να αρχικοποιηθούν και οι εισοδοί να εξομαλυνθούν. Ο Kohonen έχει επισημάνει ότι η έλλειψη της τοπολογίας σε μοντέλα ΝΔ κάνουν τα σημερινά ΝΔ να αποτελούν αφηρημένες έννοιες των ΝΔ του εγκεφάλου. Καθώς αυτή η έρευνα συνεχίζεται, καθίσταται εφικτό να γίνουν στο μέλλον πιο ισχυρά τα δίκτυα αυτοεκπαίδευσης. Ωστόσο, επί του παρόντος, αυτός ο τομέας παραμένει ένας τομέας που βρίσκεται ακόμη σε πειραματικό στάδιο εργαστηριακού περιβάλλοντος.

2.3.5 Εκπαίδευση των Τεχνητών Νευρωνικών Δικτύων

2.3.5.1 Εκπαίδευση των Τεχνητών Νευρωνικών Δικτύων με χρήση του αλγορίθμου της Οπισθοδιάδοσης

Ο προγραμματιστής που πρόκειται να κατασκευάσει το δικό του ΝΔ, θα πρέπει να γνωρίζει εκ των προτέρων τον τρόπο να το εκπαιδεύσει. Ο αλγόριθμος της Οπισθοδιάδοσης είναι μια κοινώς χρησιμοποιούμενη τεχνική για την εκπαίδευση των

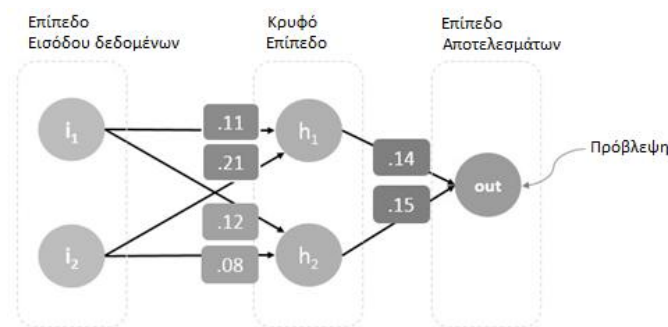


Εικόνα 27: Παράδειγμα Νευρωνικού Δικτύου

ΝΔ. Παρακάτω θα εξηγηθεί ο τρόπος λειτουργίας της Οπισθοδιάδοσης με ένα παράδειγμα. Έστω ότι θέλουμε να χτίσουμε ένα ΝΔ που αποτελείται από τρία επίπεδα, εκ των οποίων το πρώτο επίπεδο θα αποτελείται από δύο νευρώνες εισόδου, το κρυφό επίπεδο θα αποτελείται επίσης από δύο κρυφούς νευρώνες και το επίπεδο εξόδου θα αποτελείται από έναν μόνο νευρώνα όπως φαίνεται στο σχήμα που ακολουθεί.

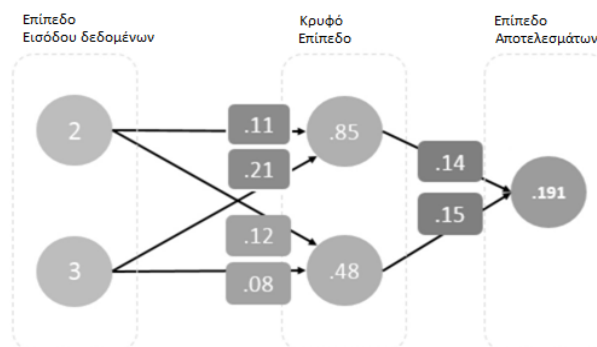
Αναφορικά με τα βάρη σε ένα ΝΔ είναι γνωστό ότι παίζουν καθοριστικό ρόλο στην αρχιτεκτονική, λειτουργία και εκπαίδευση του. Η εκπαίδευση ενός ΝΔ λοιπόν, αφορά στην εύρεση βαρών που ελαχιστοποιούν το σφάλμα πρόβλεψης. Η εκπαίδευση ξεκινά συνήθως με την εισαγωγή ενός τυχαίου συνόλου παραγόμενων βαρών. Στη συνέχεια, η Οπισθοδιάδοση χρησιμοποιείται για την ενημέρωση των βαρών αυτών προκειμένου να χαρτογραφηθούν σωστά οι αυθαίρετες εισοδοί στις εξόδους.

Για το ΝΔ της Εικόνας 27 θα υποθέσουμε ότι τα αρχικά βάρη που θα εισαχθούν θα έχουν τις εξής τιμές: $w_1 = 0.11$, $w_2 = 0.21$, $w_3 = 0.12$, $w_4 = 0.08$, $w_5 = 0.14$ και $w_6 = 0.15$



Εικόνα 28: Παράδειγμα Νευρωνικού Δικτύου με εισαγωγή των βαρών

Επίσης, το σύνολο δεδομένων του ΝΔ που κατασκευάζουμε θα περιέχει ένα δείγμα με δύο εισόδους και μια έξοδο, όπου οι τιμές αυτών θα είναι: είσοδος = [2, 3] και έξοδος = [1].



Εικόνα 29: Παράδειγμα Νευρωνικού Δικτύου με εισαγωγή των βαρών και νευρώνων

Προχωρώντας στη διαδικασία της πρόσθιας τροφοδότησης θα πρέπει να προβλεφθεί η έξοδος από το ΝΔ χρησιμοποιώντας τα δεδομένα βάρη και τις εισόδους που του έχουν δοθεί, στη συνέχεια οι εισοδοί θα πολλαπλασιαστούν με τα βάρη και τα αποτελέσματα αυτής της πράξης θα μεταφερθούν προς το επόμενο επίπεδο, όπως φαίνεται στο σχήμα που ακολουθεί.

Οι υπολογισμοί που γίνονται στο σχήμα της Εικόνας 29 είναι οι εξής:

$$[2, 3] \times \begin{bmatrix} 0.11 & 0.12 \\ 0.21 & 0.08 \end{bmatrix} = [0.85 \ 0.48] \times \begin{bmatrix} 0.14 \\ 0.15 \end{bmatrix} = [0.191]$$

Το επόμενο βήμα του αλγορίθμου της Οπισθοδιάδοσης αφορά στον υπολογισμό του σφάλματος. Επομένως, σκοπός αυτού του βήματος είναι η εκτίμηση της απόδοσης του ΝΔ, η οποία μπορεί να πραγματοποιηθεί με τον υπολογισμό της διαφοράς μεταξύ της πραγματικής εξόδου και της προβλεπόμενης. Παρατηρείται λοιπόν ότι η τιμή της εξόδου που παράγει το δίκτυο ή η πρόβλεψή του (0.191) δεν είναι καθόλου κοντά στην πραγματική έξοδο (1). Ο υπολογισμός της διαφοράς αυτής ή του σφάλματος διαφορετικά γίνεται ως εξής:

$$\text{Σφάλμα} = \frac{1}{2} (\text{πρόβλεψη} - \text{πραγματική})^2 \quad \text{ή} \quad \text{Σφάλμα} = \frac{1}{2} (0.191 - 1.0)^2 = 0.327$$

Ο κύριος στόχος λοιπόν του αλγορίθμου της Οπισθοδιάδοσης είναι η μείωση του σφάλματος ή της διαφοράς μεταξύ της προβλεπόμενης τιμής και της πραγματικής. Οπότε δεδομένου ότι, η πραγματική έξοδος είναι μια σταθερά δεν μπορεί να αλλάξει, επομένως ο μόνος τρόπος να μειωθεί το σφάλμα είναι μέσω της αλλαγής της προβλεπόμενης τιμής. Με άλλα λόγια, για να αλλάξει η προβλεπόμενη τιμή θα πρέπει να αλλάξουν οι τιμές των βαρών.

Για την αλλαγή των βαρών σημαντικό ρόλο παίζει ο ακόλουθος μηχανισμός που θα περιγραφεί, μέσω του οποίου θα ενημερωθούν τα βάρη χρησιμοποιώντας την μαθηματική κλίση. Η κλίση αυτή υπολογίζει τη συνάρτηση σφάλματος σε σχέση με τα βάρη του ΝΔ. Η σχέση αυτή είναι η εξής:

$$W_x^* = W_x - a \left(\frac{\partial \text{σφάλμα}}{\partial W_x} \right), \text{ όπου } a \text{ είναι ο ρυθμός εκμάθησης}$$

Για παράδειγμα, για να ενημερωθεί το βάρος W_6 , παίρνουμε το τρέχον βάρος και αφαιρούμε τη μερική παράγωγο της συνάρτησης σφάλματος σε σχέση με το w_6 και προαιρετικά πολλαπλασιάζουμε την παράγωγο της συνάρτησης σφάλματος με έναν επιλεγμένο αριθμό για να βεβαιωθούμε ότι το νέο ενημερωμένο βάρος ελαχιστοποιεί τη συνάρτηση σφάλματος, όπου ο επιλεγμένος αριθμός ονομάζεται ρυθμός

εκμάθησης. Με άλλα λόγια για την ενημέρωση του βάρους W_6 χρησιμοποιείται ο τύπος:

$$W_6^* = W_6 - a \left(\frac{\partial \sigma \alpha \lambda \mu \alpha}{\partial W_6} \right)$$

Στην εξίσωση που προαναφέρθηκε όσον αφορά την παράγωγο της συνάρτησης του σφάλματος, αυτή μπορεί να υπολογιστεί με τον τύπο Δh_2 ο οποίος προκύπτει από τον κανόνα της αλυσίδας και όπου Δ είναι ο γνωστός κανόνας Δέλτα ο οποίος ισούται με

$$\Delta = \text{πρόβλεψη} - \text{πραγματική τιμή και όπου } h_2 = i_1 W_3 + i_2 W_4.$$

Συνεπώς, για την ενημέρωση του βάρους W_6 χρησιμοποιείται ο τύπος:

$$W_6^* = W_6 - a \Delta h_2.$$

Τέλος, ο κατάλογος με τις εξισώσεις για την ενημέρωση των υπόλοιπων βαρών ακολουθώντας την ίδια φιλοσοφία με το βάρος W_6 είναι:

$$W_5^* = W_5 - a \Delta h_1, \text{ όπου } h_1 = i_1 W_1 + i_2 W_2$$

$$W_4^* = W_4 - a(i_2 \Delta W_6)$$

$$W_3^* = W_3 - a(i_1 \Delta W_6)$$

$$W_2^* = W_2 - a(i_2 \Delta W_5)$$

$$W_1^* = W_1 - a(i_1 \Delta W_5).$$

2.3.5.2 Μια Εναλλακτική περιγραφή της εκπαίδευσης των Τεχνητών Νευρωνικών Δικτύων

Έστω ένας ορειβάτης στην κορυφή ενός βουνού, και η νύχτα έχει πέσει. Πρέπει να φτάσει στο στρατόπεδο βάσης το οποίο βρίσκεται στο κάτω μέρος του βουνού, αλλά στο σκοτάδι με μόνο εφόδιο το φακό του δεν μπορεί να δει παρά μόνο μερικά εκατοστά μπροστά του. Οπότε γεννάται το ερώτημα, πως θα κατέβει; Μια στρατηγική είναι να κοιτάξει προς όλες τις κατευθύνσεις προκειμένου να δει προς ποιά κατεύθυνση το έδαφος πάει προς τα κάτω περισσότερο και εν συνεχεία να προχωρήσει προς αυτή τη κατεύθυνση. Στη συνέχεια επαναλαμβάνει αυτή τη διαδικασία πολλές φορές και σταδιακά θα καταφέρει να πάει μακρύτερα και πιο κάτω. Μπορεί μερικές φορές να κολλήσει σε κάποια μικρή κοιλάδα, όπου καλό θα ήταν να ακολουθήσει την απόφαση της στιγμής προκειμένου να βγει από αυτή τη κατάσταση. Με κάθε επιφύλαξη, η στρατηγική αυτή θα τον οδηγούσε τελικά στο πάτο του βουνού.

Αυτό το σενάριο που περιγράφηκε παραπάνω σε πρώτη φάση μπορεί να φαίνεται

αποσυνδεδεμένο από τα ΝΔ, αλλά αποδεικνύεται ότι είναι μια καλή αναλογία για τον τρόπο με τον οποίο αυτά εκπαιδεύονται. Είναι τόσο καλή αυτή η αναλογία, που η κύρια τεχνική για την εκπαίδευση, ταυτίζεται αρκετά με αυτό που μόλις περιγράψαμε. Ας θυμηθούμε ότι η εκπαίδευση έχει να κάνει με τον καθορισμό του καλύτερου συνόλου βαρών για τη μεγιστοποίηση της ακρίβειας του ΝΔ.

Είναι αξιοσημείωτο το γεγονός ότι τα ΝΔ μπορούν να χρησιμοποιηθούν χωρίς να γνωρίζουμε με ακρίβεια πως λειτουργεί η εκπαίδευση τους, όπως ακριβώς μπορεί κανείς να χειριστεί έναν φακό χωρίς να γνωρίζει πως λειτουργούν τα ηλεκτρονικά μέσα του. Οι περισσότερες σύγχρονες βιβλιοθήκες MM έχουν αυτοματοποιήσει σε μεγάλο βαθμό τη διαδικασία εκπαίδευσης.

2.3.6 Βασικές διαφορές της λειτουργίας των Τεχνητών Νευρωνικών Δικτύων, του Παραδοσιακού Υπολογισμού και των Ειδικών Συστημάτων

Τα ΝΔ προσφέρουν ένα διαφορετικό τρόπο ανάλυσης δεδομένων και αναγνώρισης προτύπων από τις παραδοσιακές μεθόδους υπολογισμού, αν και δεν αποτελούν λύση για όλα τα προβλήματα της Πληροφορικής. Οι παραδοσιακές μέθοδοι υπολογισμού λειτουργούν καλά για τα προβλήματα τα οποία μπορούν να χαρακτηριστούν ως «σωστά». Για παράδειγμα η εξισορρόπηση των βιβλίων επιταγών, η διατήρηση της καθολικότητας κ.α. είναι καλά καθορισμένα προβλήματα και δεν χρειάζονται τα ιδιαίτερα χαρακτηριστικά των ΝΔ. Ο πίνακας 1 παρακάτω προσδιορίζει τις βασικές διαφορές μεταξύ των δύο αυτών υπολογιστικών προσεγγίσεων.

Οι Παραδοσιακοί Υπολογισμοί αποτελούν την ιδανική λύση για πολλές και διάφορες εφαρμογές. Η προσέγγιση λοιπόν αυτή μπορεί, να επεξεργαστεί τα δεδομένα, να ανιχνεύσει τα αποθέματα, να προστατέψει τον εξοπλισμό και πολλά άλλα, όπου αυτές οι εφαρμογές που αναφέρθηκαν δεν χρειάζονται τα ειδικά χαρακτηριστικά των ΝΔ.

Τα Ειδικά Συστήματα (ΕΣ) αποτελούν μια επέκταση του Παραδοσιακού Υπολογισμού και πολλές φορές ονομάζονται ως η πέμπτη γενιά υπολογισμού. Σημειώνουμε ότι, η πρώτη γενιά υπολογισμού χρησιμοποιούσε διακόπτες και σύρματα, η δεύτερη γενιά προέκυψε από την ανάπτυξη των τρανζίστορ, η τρίτη γενιά περιελάμβανε τη τεχνολογία Solid State, η τέταρτη γενιά αφορούσε τις υψηλού επιπέδου γλώσσες όπως τη COBOL, τη Fortran και τη C και τέλος η πέμπτη γενιά περιλαμβάνει τη TN.

Πίνακας 1: Σύγκριση Παραδοσιακού Υπολογισμού και ΝΔ

Χαρακτηριστικά	Παραδοσιακός υπολογισμός (συμπεριλαμβανομένων και των ειδικών συστημάτων)	Τεχνητά Νευρωνικά Δίκτυα
Στυλ επεξεργασίας	Ακολουθιακό	Παράλληλο
Συναρτήσεις	Λογικές μέσω κανόνων, εννοιών και υπολογισμών	Μέσω εικόνων, φωτογραφιών και ρυθμίσεων
Μέθοδος Εκμάθησης	Μέσω κανόνων	Μέσω παραδειγμάτων
Εφαρμογές	Λογιστική, επεξεργασία λέξεων, μαθηματικά, ψηφιακές επικοινωνίες	Επεξεργασία αισθητήρα, αναγνώριση ομιλίας, αναγνώριση προτύπου, αναγνώριση κειμένου

Πίνακας 2: Σύγκριση ΕΣ και Νευρωνικών Δικτύων

Χαρακτηριστικά	Αρχιτεκτονική Von Neumann που χρησιμοποιείται για τα ΕΣ	Νευρωνικά Δίκτυα
Επεξεργαστές	VLSI	Ποικιλία τεχνολογιών
Προσέγγιση επεξεργασίας	Ένας κανόνας κάθε φορά ακολουθιακά	Πολλαπλά, ταυτόχρονα
Συνδέσεις	Εξωτερικά προγραμματισμένες	Προγραμματίζονται δυναμικά από μόνα τους
Αυτοδιδασκαλία	Μόνο αλγοριθμικά τροποποιημένες παράμετροι	Συνεχής προσαρμογή
Ανοχή σε σφάλματα	Καμία με ειδικούς επεξεργαστές	Σημαντική ανοχή λόγω της φύσης των διασυνδεδεμένων νευρώνων
Χρήση της νευροβιολογίας στο σχεδιασμό	Καμία	Μετριασμένη
Προγραμματισμός	Μέσω ενός κελύφους βασισμένου σε κανόνες	Προγραμματίζεται αυτόματα αλλά το δίκτυο πρέπει να ρυθμιστεί σωστά
Ικανότητα να είναι γρήγορο	Απαιτεί μεγάλους επεξεργαστές	Απαιτεί πολλαπλά προσαρμοσμένα chip

Τυπικά, ένα ΕΣ αποτελείται από δύο μέρη: έναν μηχανισμό συμπερασμάτων και μια βάση γνώσεων. Ο μηχανισμός συμπερασμάτων είναι γενικός και διαχειρίζεται το περιβάλλον του χρήστη, τα εξωτερικά αρχεία, τη πρόσβαση στο πρόγραμμα και τον προγραμματισμό. Η βάση γνώσεων περιέχει από μόνη της τις συγκεκριμένες πληροφορίες που απαιτούνται για ένα συγκεκριμένο πρόβλημα και επιτρέπει σε έναν εμπειρογνώμονα να καθορίσει τους κανόνες που διέπουν μια διαδικασία. Είναι άξιο αναφοράς το γεγονός ότι αυτός ο εμπειρογνώμονας δεν χρειάζεται να γνωρίζει καλά τον παραδοσιακό προγραμματισμό, καθώς πρέπει απλά να καταλάβει τι θέλει να κάνει για αυτόν ο υπολογιστής και το πώς λειτουργεί ο μηχανισμός των κελυφών του ΕΣ.

Οι προσπάθειες που πραγματοποιήθηκαν για να γενικευτούν περισσότερο τα ΕΣ συνάντησαν έναν αριθμό προβλημάτων καθώς η πολυπλοκότητα του συστήματος αυξανόταν και έτσι το σύστημα απαιτούσε πολλούς υπολογιστικούς πόρους με αποτέλεσμα να γίνει υπερβολικά αργό. Έχει αποδειχθεί ότι τα ΕΣ είναι εφικτά μόνο όταν περιορίζονται στενά.

Τα ΝΔ από την άλλη προσφέρουν μια εντελώς διαφορετική προσέγγιση για την επίλυση προβλημάτων και μερικές φορές καλούνται ως η έκτη γενιά υπολογισμών. Αυτά, παρέχουν ένα εργαλείο όπου το πρόγραμμα μπορεί να μάθει από μόνο του. Τα ΝΔ είναι δομημένα έτσι ώστε να παρέχουν την δυνατότητα να λύνονται προβλήματα χωρίς τα οφέλη ενός εμπειρογνώμονα και χωρίς να χρειάζεται κάποιος προγραμματισμός. Μπορούν επίσης να αναζητήσουν πρότυπα σε δεδομένα που κανείς δεν ξέρει ότι υπάρχουν.

Μια σύγκριση του Παραδοσιακού Προγραμματισμού και των ΝΔ περιγράφεται στον Πίνακα 1 και των ΕΣ και ΝΔ περιγράφεται στον Πίνακα 2 που ακολουθεί.

Τα ΕΣ έχουν καταφέρει να επιδείξουν σημαντικές επιτυχίες. Ωστόσο, η ΤΝ έχει αντιμετωπίσει προβλήματα σε τομείς όπως η όραση, η συνεχής αναγνώριση ομιλίας, η ΜΜ κ.α. και είναι επίσης όμηρος της ταχύτητας του επεξεργαστή στον οποίο τρέχει με αποτέλεσμα να περιορίζεται στο θεωρητικό όριο ενός μόνο επεξεργαστή. Τέλος, η ΤΝ επιβαρύνεται επίσης λόγω του γεγονότος ότι οι εμπειρογνώμονες δεν μιλούν πάντοτε μέσω των κανόνων.

2.4 Συμπεράσματα

Καταλήγομε λοιπόν στο συμπέρασμα ότι αν και ο Παραδοσιακός Υπολογισμός και τα ΕΣ είναι δύο υπολογιστικές μέθοδοι που συνεισέφεραν σημαντικά ως η Πέμπτη γενιά υπολογισμών στο παρελθόν, η χρήση αυτών σε βάθος χρόνου παρουσίασε κάποια

σημαντικά προβλήματα. Για το λόγο αυτό δημιουργήθηκε η έκτη γενιά υπολογισμών όπου ανήκουν τα ΝΔ, η χρήση των οποίων συντέλεσε στην σημαντική μείωση των προβλημάτων που δημιουργήθηκαν από τις προσεγγίσεις της πέμπτης γενιάς. Επίσης, τα ΝΔ υπερίσχυσαν έναντι των προηγούμενων προσεγγίσεων καθώς ήταν λιγότερο πολύπλοκα αλλά και επειδή δεν απαιτούσε η αρχιτεκτονική τους από τον κατασκευαστή να έχει πολλές γνώσεις για το αντικείμενο αυτό.

2.5 Βιβλιογραφικές Αναφορές

1. Anderson, D. & McNeill, G. (1992). *Artificial Neural Networks Technology*. New York: Kaman Sciences Corporation.
2. Atlas, L. E., Marks, R. J. & Toshiro, H. (1987). An artificial neural network for spatio-temporal bipolar patterns: Application to phoneme classification. In D. Anderson (Eds.), *Neural Information Processing Systems* (pp.31–40). Maryland: American Institute of Physics.
3. Basheer, I. A. & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. *Journal of microbiological methods*, 43(1), pp.3-31.
4. Bishop, J. M. (2015). HISTORY AND PHILOSOPHY OF NEURAL NETWORKS. In Ishibuchi, H. (Eds.), *Computational Intelligence* (pp.22-96). France: Eolss Publishers.
5. Broomhead, D. S. & Lowe, D. (1988). *Radial basis functions, multi-variable functional interpolation and adaptive networks*. London: Controller HMSO.
6. China, A. (2009). Understanding the Principles of Recursive Neural Networks: A Generative Approach to Tackle Model Complexity. In: Alippi, C., Polycarpou, M., Panayiotou, C., Ellinas, G. (Eds.), *ICANN. LNCS 5768*, (pp. 952-963). Heidelberg: Springer.
7. Fergus, R. & Zeiler, M.D. (2014). Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B. & Tuytelaars, T. (Eds.), *Computer Vision – ECCV 2014* (pp.818-833). Zurich: Springer.
8. Fukushima, K. (1980). Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cybernetics*, 36, pp. 193-202.
9. Goldberg, Y. (2016). A Primer on Neural Network Models for Natural Language Processing. *Journal of Artificial Intelligence Research*, 57(1), pp.345-420.
10. Graves, A. (2012). *Supervised Sequence Labelling with Recurrent Neural Network*. New York: Springer.
11. Gurney, K. (1997). *An introduction to neural networks*. London: UCL Press.

12. Hubel, D. H. & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1), pp.106-154.2.
13. Jain, A. K., Mao, J. & Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *Computer*, 29(3), pp.31-44.
14. Kohonen, T. (1982). Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, 43 (1), pp. 59-69.
15. Kriesel, D. (2005). *A Brief Introduction to Neural Networks*. Germany: Kriesel.
16. Kshirsagar, P. & Rathod, N. (2012). Artificial neural network. *International Journal of Computer Applications*, 2, pp.12-16.
17. Manavazhahan, M. (2017). *A Study of Activation Functions for Neural Networks*. Fayetteville: University of Arkansas.
18. McCulloch, W. & Pitts, W. (1943). A Logical Calculus of Ideas Immanent in Nervous Activity. In Hastings, A. & Laubenbacher, R. (Eds.), *Bulletin of Mathematical Biophysics* (pp.115–133). USA: Springer.
19. Minsky, M. & Papert, S. (1969). *Perceptrons: An Introduction to Computational Geometry*. Mass: MIT Press.
20. O'Shea, K. & Nash, R. (2015). An Introduction to Convolutional Neural Networks. ArXiv e-prints, pp.1-11. Retrieved January 27, 2019, from <http://arxiv.org/abs/1511.08458>
21. Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model For Information Storage And Organization In The Brain. *Psychological Review*, 65 (6), pp.386–408.
22. Ruder, S. (2018). A review of the Neural History of Natural language processing. Retrieved January 21, 2019, from <https://blog.aylien.com/a-review-of-the-recent-history-of-natural-language-processing/>
23. Sathasivam, S. & Abdullah, W. (2008b). Logic Learning in the hopfield Networks. *Modern Applied Science* 2(3), pp.1-8.
24. Suzuki, K. (2011). *Artificial Neural Networks – Methodological Advances and Biomedical Applications*. Croatia: InTech.
25. Weng, J., Ahuja, N. & Huang, T.S. (1993). Learning Recognition and Segmentation Using the Cresceptron. *International Journal of Computer Vision*, 25(2), pp.109-143.

ΚΕΦΑΛΑΙΟ 3

Νευρωνικά Δίκτυα και Επεξεργασία Φυσικής Γλώσσας

3.1 Εισαγωγή

Τα τελευταία χρόνια, τα ΝΔ επανεμφανίστηκαν ως μοντέλα ΜΜ, αποδίδοντας πολύ καλά αποτελέσματα σε διάφορα πεδία, μεταξύ αυτών στην Αναγνώριση Εικόνας και την Αναγνώριση του Λόγου. Τα μοντέλα αυτά των ΝΔ τον τελευταίο καιρό άρχισαν να εφαρμόζονται και στον τομέα της Επεξεργασίας Φυσικής Γλώσσας (ΕΦΓ) με πολλά υποσχόμενα αποτελέσματα. Οι προσεγγίσεις των μεθόδων ΕΦΓ ξεκίνησαν βάσει των Γραμματικών και επεκτάθηκαν με τις προσεγγίσεις των μεθόδων ΜΜ που ήταν οι κυρίαρχες τεχνικές ΕΦΓ για εξίσου μεγάλο χρονικό διάστημα. Οι προσεγγίσεις με μεθόδους ΜΜ χρησιμοποιούσαν γραμμικά μοντέλα όπως είναι για παράδειγμα οι Μηχανές Διανυσμάτων Υποστήριξης ή η Λογιστική Παλινδρόμηση. Όμως για την εκπαίδευσή τους απαιτούσαν αραιά διανύσματα, υψηλών διαστάσεων χαρακτηριστικών. Η πρόσφατη εξέλιξη της ΕΦΓ συντελέστηκε με τη μετάβαση των γραμμικών μοντέλων στα μη γραμμικά μοντέλα των ΝΔ, παρόλο που δεν είναι όλες οι αρχιτεκτονικές των ΝΔ κατάλληλες για ΕΦΓ. Αυτό διότι ενώ οι περισσότερες τεχνικές των ΝΔ είναι εύκολο να εφαρμοστούν, ωστόσο παρεμβάλλεται ένα σημαντικό εμπόδιο στο οποιοδήποτε επίπεδο εισόδου τους, λόγω της μη διαχειρίσιμης κατάστασης που προκαλεί η πολυπλοκότητα των δεδομένων που είναι συνηθισμένη στην ΕΦΓ, δυσκολίες παρόμοιες με εκείνες που αντιμετωπίζονται και στην επεξεργασία εικόνας.

Όπως έχει προαναφερθεί σκοπός της ΕΦΓ είναι η αυτόματη επεξεργασία κειμένων

της για την ανάλυσή τους και την εξαγωγή πολυεπίπεδων πληροφοριών, με πυκνά και πολυδιάστατα χαρακτηριστικά, για περαιτέρω χρήση τους, όπως είναι αυτά που απαιτούνται στη μηχανική μετάφραση, την εξαγωγή νοήματος, τη διαχείριση εγγράφων, την ενεργοποίηση ρομπότ, την Ετικετοποίηση¹⁴, κ.ά.

Αξίζει να αναφερθεί ότι τα ΝΔ εφαρμόζονται στην ΕΦΓ με καλά αποτελέσματα, για την ώρα, προσελκύνοντας όλο και περισσότερο το ενδιαφέρον των ερευνητών για την ανακάλυψη νέων μοντέλων, αφού αποδεικνύεται ότι είναι αποτελεσματικά και καθόλου δαπανηρά. Επίσης σημειώνεται ότι τα ΝΔ που προτιμήθηκαν για την ΕΦΓ, πέραν των τυπικών ΝΔ, είναι τα Συνελικτικά Νευρωνικά Δίκτυα και τα Επαναληπτικά Νευρωνικά Δίκτυα. Αυτές οι τρεις κατηγορίες ΝΔ είναι το αντικείμενο αυτής της ενότητας.

Λέξεις Κλειδιά: Επεξεργασία Φυσικής Γλώσσας, Νευρωνικά Δίκτυα, Word2Vec, Αποδεκτός Συνδυασμός Λέξεων, deeplearning4j, Επαναληπτικά Νευρωνικά Δίκτυα, Συνελικτικά Νευρωνικά Δίκτυα, Επίπεδο Συνέλιξης, Επίπεδο Συγκέντρωσης

3.2 Μέθοδοι Τυπικών Νευρωνικών Δικτύων και Επεξεργασία Φυσικής Γλώσσας

3.2.1 Εισαγωγή

Είναι γνωστό ότι η αρχιτεκτονική και οι αλγόριθμοί της ΜΒ ώθησαν την εξέλιξη της τεχνολογίας εν γένει, όπως της υπολογιστικής όρασης (computer vision) και της αναγνώρισης προτύπων (pattern recognition). Ακολουθώντας αυτή την εξέλιξη η ΕΦΓ εστίασε την έρευνα στη χρήση των νέων μεθόδων ΜΒ. Με την υιοθέτηση των ΝΔ που ήταν κατάλληλα σχεδιασμένα για τη διαχείριση διανυσμάτων (δεδομένων), με πυκνά και πολλές διαστάσεις χαρακτηριστικά αντιμετωπίστηκαν τα όποια προβλήματα των προηγούμενων μεθόδων. Έτσι υλοποιήθηκαν αφενός ο αποδεκτός συνδυασμός λέξεων και αφετέρου οι μέθοδοι ΜΒ. Η ΜΒ έκανε εφικτή την αυτόματη Εκμάθηση πολυεπίπεδων εκφράσεων (βασικών) χαρακτηριστικών, δηλαδή πολυδιάστατων διανυσμάτων. Σε αντίθεση με τις παραδοσιακές μεθόδους ΕΦΓ οι οποίες δύσκολα μπορούσαν να συνδυάσουν ακόμα και (βασικά) χαρακτηριστικά που δεν είχαν κωδικοποιηθεί αυτόματα, τα οποία σε πολλές περιπτώσεις είχαν προσεγγιστεί χειρονακτικά και με ανεπάρκεια. Στην επόμενη παράγραφο θα αναπτυχθεί εύληπτα ως παράδειγμα αυτής της εξέλιξης η μέθοδος Word2Vec των τυπικών ΝΔ η οποία συνδυάζει αποδεκτά λέξεις.

¹⁴ Επισήμανση των λέξεων του κειμένου με ετικέτες που αντιστοιχούν στο μέρος του λόγου που οι λέξεις αυτές ανήκουν.

3.2.2 Μέθοδος του Αποδεκτού Συνδυασμού Λέξεων με Νευρωνικά Δίκτυα Word2Vec

Η Μέθοδος Word2Vec αποτελείται από ένα σύνολο προτύπων (μοντέλων) τα οποία χρησιμοποιούνται για την παραγωγή αποδεκτού συνδυασμού λέξεων και έχουν προσελκύσει μεγάλο ενδιαφέρον τα τελευταία χρόνια. Οι αναπαραστάσεις των συναρτήσεων των λέξεων που έχουν διδαχθεί στα πρότυπα (μοντέλα) του Word2Vec έχουν δείξει ότι φέρουν σημασιολογικές πληροφορίες, χρήσιμες σε ποικίλα προβλήματα της ΕΦΓ.

Το Word2Vec είναι ένα τυπικό ΝΔ δύο επιπέδων το οποίο κατασκευάστηκε από μια ομάδα ερευνητών με επικεφαλής τον Tomas Mikolov στη Google και έχει την ιδιότητα να επεξεργάζεται κείμενα. Η είσοδος του είναι ένα σώμα κειμένου και η έξοδος του είναι ένα σύνολο από διανύσματα, δηλαδή, διανύσματα (δεδομένα) πολυδιάστατων χαρακτηριστικών των λέξεων του σώματος κειμένου που δέχεται ως είσοδο. Ενώ το Word2Vec δεν είναι ένα βαθύ ΝΔ, μπορεί και μετατρέπει, δηλαδή, κωδικοποιεί το κείμενο σε μια αριθμητική μορφή/έκφραση την οποία τα δίκτυα βάθους μπορούν να κατανοήσουν. Το `deeplearning4j`¹⁵ υλοποιεί μια κατανεμημένη μορφή Word2Vec η οποία λειτουργεί στο Spark¹⁶ με χρήση της Μονάδας Επεξεργασίας Γραφικών (GPU).

Οι εφαρμογές της μεθόδου αυτής αξιοποιούνται πέραν από την ανάλυση των προτάσεων και σε διάφορους άλλους τομείς όπως: στα γονίδια του DNA, στους κώδικες κάθε μορφής, στις λίστες αναπαραγωγής ψηφιοποιημένων τραγουδιών και βίντεο (playlists), στα γραφήματα των κοινωνικών μέσων δικτύωσης κ.α. Ο σκοπός και η χρησιμότητα της μεθόδου αυτής είναι να ανιχνεύει ομοιότητες με μαθηματικές μεθόδους και τεχνικές, ομαδοποιώντας διανύσματα παρόμοιων λέξεων στον αντίστοιχο διανυσματικό χώρο. Αυτό ισχύει διότι οι λέξεις θεωρούνται διακριτές καταστάσεις, όπως είναι τα υπόλοιπα δεδομένα που προαναφέρθηκαν, και απλά αναζητούνται οι προβλέψιμες πιθανότητες αποδεκτής συνύπαρξης αυτών των καταστάσεων. Έτσι οι διαδικασίες `gene2vec`, `2vec` και `follower2vec`, είναι όλες ενεργοποιήσιμες. Με τον τρόπο αυτόν δημιουργούνται αποδεκτοί συνδυασμοί λέξεων με τη χρήση των νευρώνων για κάθε ομάδα διακριτών καταστάσεων που μπορούν αποδεκτά να συνυπάρξουν.

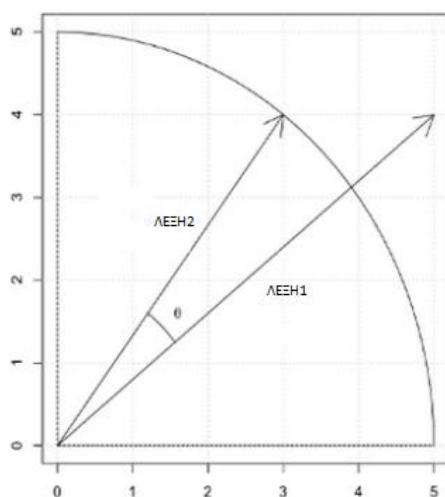
Ο σκοπός και η χρησιμότητα του Word2Vec είναι να ομαδοποιήσει τα διανύσματα

¹⁵ Το `deeplearning4j` είναι μια βιβλιοθήκη προγραμματισμού ΒΜ της Java που είναι συμβατή με γλώσσες προγραμματισμού όπως η Scala, Clojure, Kotlin.

¹⁶ Το Spark είναι ένα πολύ ισχυρό, κλιμακωτό εργαλείο ανάλυσης δεδομένων σε πραγματικό χρόνο.

των όμοιων λέξεων σε έναν διανυσματικό χώρο, μέθοδος που μπορεί να θεωρηθεί ότι ανιχνεύει τις ομοιότητες με μαθηματική προσέγγιση. Εμφατίζεται ότι το Word2Vec δημιουργεί διανύσματα χωρίς ανθρώπινη ή παρέμβαση. Τα διανύσματα αυτά είναι κατανεμημένες αριθμητικές παραστάσεις χαρακτηριστικών λέξεων, όπως είναι τα συμφραζόμενα μεμονωμένων λέξεων.

Τροφοδοτώντας το Word2Vec με έναν ικανό αριθμό δεδομένων, δηλαδή λέξεων, μπορεί να προβλέψει με μεγάλη ακρίβεια τις έννοιες των λέξεων με βάση την προηγούμενη αποκτηθείσα εμπειρία. Αυτές οι προβλέψεις μπορούν να χρησιμοποιηθούν για να οριστούν σχέσεις μιας λέξης με άλλες (π.χ. ο «άνδρας» με το «αγόρι» και η «γυναίκα» με το «κορίτσι») ή για την διαχείριση/ομαδοποίηση εγγράφων, δηλαδή για να γίνει συλλογή και ταξινόμησή τους με βάση τις θεματικές τους. Αυτές οι ταξινομήσεις μπορούν να τυποποιήσουν τη βάση για περαιτέρω έρευνα, όπως για την ανάλυση συναισθήματος και να οδηγήσουν σε διαφορετικά πεδία όπως είναι αυτό που αφορά στις διάφορες πτυχές της νομικής διαδικασίας, το ηλεκτρονικό εμπόριο και τη διαχείριση των σχέσεων πελατών.



Εικόνα 30: Η μέτρηση της ομοιότητας των λέξεων που εκφράζεται με τη βοήθεια του ορισμού του συνημίτονου και υπολογίζει τη γωνιακή απόσταση μεταξύ των διανυσμάτων τους

Η έξοδος αποτελεσμάτων της μεθόδου Word2Vec των ΝΔ είναι ένα λεξιλόγιο στο οποίο σε κάθε στοιχείο του αντιστοιχίζεται ένα διάνυσμα που το χαρακτηρίζει και το οποίο μπορεί να τροφοδοτήσει ένα δίκτυο ΜΒ ή απλά όταν ερωτηθεί να υποδείξει σχέσεις μεταξύ λέξεων.

Με τη μέτρηση της ομοιότητας που εκφράζεται με τη βοήθεια του συνημίτονου¹⁷, καμία ομοιότητα δεν εκφράζεται ως γωνία 90 μοιρών, ενώ αντίθετα η ολική ομοιότητα που συμβολίζεται με 1 αντιστοιχεί τεκμηριωμένα στη γωνία των 0 μοιρών.

Για παράδειγμα η Σουηδία ισούται με τη Σουηδία, ενώ η Νορβηγία έχει απόσταση που υπολογίζεται με το συνημίτονο της γωνίας - «γωνιακή απόσταση» - ίση με 0,760124 από τη Σουηδία και βρίσκεται υψηλότερα στο χάρτη από οποιαδήποτε άλλη χώρα της Ευρώπης. Στην εικόνα παρακάτω (βλ. Εικόνα 31) ακολουθεί μια λίστα με λέξεις που σχετίζονται με τη “Σουηδία” με χρήση της μεθόδου Word2Vec με σειρά εγγύτητας, όπου παρατηρείται ότι μεταξύ των πρώτων 9 χωρών συγκαταλέγονται τα έθνη της Σκανδιναβίας και αρκετές πλούσιες, βόρειες Ευρωπαϊκές, Γερμανικές χώρες.

Λέξη	Γωνιακή Απόσταση
norway	0.760124
denmark	0.715460
finland	0.620022
switzerland	0.588132
belgium	0.585835
netherlands	0.574631
iceland	0.562368
estonia	0.547621
slovenia	0.531408

Εικόνα 31: Λίστα με λέξεις που σχετίζονται με τη “Σουηδία”

3.2.2.1 Νευρώνες για τον (αποδεκτό) συνδυασμό λέξεων

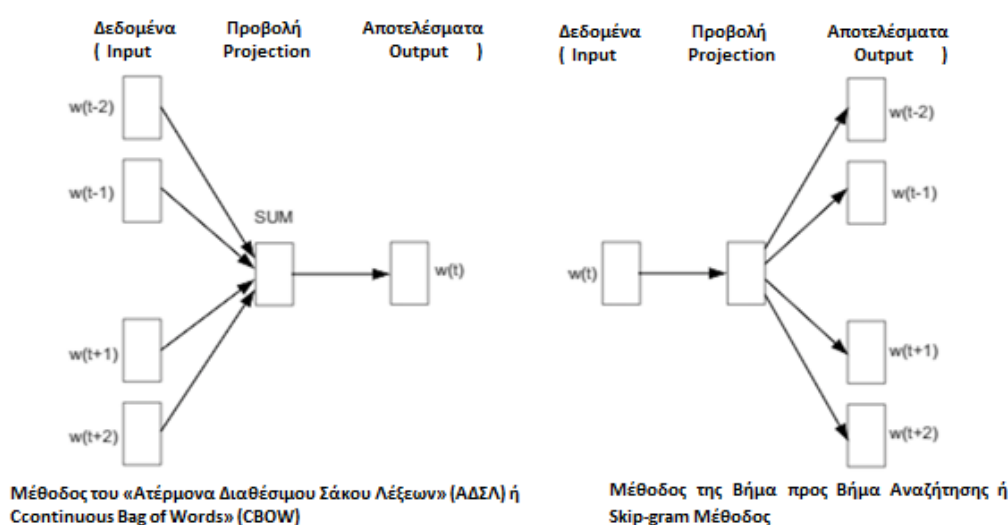
Τα διανύσματα που χρησιμοποιούμε για να συμβολίσουμε λέξεις ορίζονται ως νευρώνες ενσωμάτωσης λέξεων ή νευρώνες για τον (αποδεκτό) συνδυασμό λέξεων, με ιδιαίτερο συμβολισμό. Το Word2Vec μετατρέπει τις λέξεις σε διανύσματα ώστε η Φυσική Γλώσσα να είναι αναγνωρίσιμη από τον υπολογιστή, με αυτόν τον τρόπο δίνεται η δυνατότητα εκτέλεσης αποτελεσματικών μαθηματικών πράξεων για την ανίχνευση των ομοιοτήτων των λέξεων.

Επομένως, ένας νευρώνας για τον (αποδεκτό) συνδυασμό λέξεων στην ουσία παριστά μια λέξη με αριθμούς, είναι μία διαδικασία που μπορεί να θεωρηθεί ως μια απλή μετάφραση ή κωδικοποίηση. Δηλαδή, το Word2Vec προσομοιάζει με έναν

¹⁷ Είναι γνωστό ότι οι τριγωνομετρικές συναρτήσεις του ημιτόνου, του συνημίτονου και της εφαπτομένης γωνίας υπολογίζονται σε συνάρτηση με τα μήκη των πλευρών ορθογωνίου τριγώνου. Στην προκειμένη περίπτωση η μέτρηση ομοιότητας της προτεινόμενης μεθόδου εκφράζεται με τη βοήθεια του ορισμού του συνημίτονου υπολογίζοντας τη γωνιακή απόσταση μεταξύ των διανυσμάτων των λέξεων, με τρόπο ανάλογο με τον οποίο στην αρχαιότητα οι ναυτικοί μετρούσαν την απόσταση των άστρων με τον εξάντα.

αυτόματο κωδικοποιητή, ο οποίος αντιστοιχεί την κάθε λέξη σε ένα διάνυσμα. Το Word2Vec εκπαιδεύει στον αποδεκτό συνδυασμό λέξεων με άλλες λέξεις που γειτονεύουν στο σώμα κειμένου εισόδου, αντί να εκπαιδεύει λέξεις με αναδομήσεις, όπως μία μηχανή «Boltzmann με Περιορισμούς¹⁸».

Η εκπαίδευση του Word2Vec υλοποιείται με δύο τρόπους: 1. χρησιμοποιεί συμφραζόμενα για να προβλέψει μια υπό αναζήτηση λέξη, γνωστή ως Μέθοδος του «Ατέρμονα Διαθέσιμου Σάκου Λέξεων» (ΑΔΣΛ) ή Continuous Bag of Words (CBOW) και 2. χρησιμοποιεί την υπό μελέτη λέξη για να προβλέψει τα υπό αναζήτηση συμφραζόμενα, γνωστή ως Μέθοδος της «Βήμα προς Βήμα Αναζήτησης» ή Skip-gram Μέθοδος (βλ. Εικόνα 32).



Εικόνα 32: Αναπαράσταση των μεθόδων CBOW και Skip-gram

Όταν το διάνυσμα των χαρακτηριστικών που έχει αντιστοιχηθεί σε μία λέξη δεν μπορεί να χρησιμοποιηθεί για την πρόβλεψη με ακρίβεια των συμφραζόμενων της, τότε αυτά τα χαρακτηριστικά, συστατικά του διανύσματος, επιβάλλεται να ρυθμιστούν. Στην πράξη, η διαδικασία που ρυθμίζει το διάνυσμα των χαρακτηριστικών στέλνοντας πίσω τα σήματα των σφαλμάτων είναι τα συμφραζόμενα της λέξης στο σώμα κειμένων. Έτσι τα διανύσματα των λέξεων που κρίνονται όμοιες από τα συμφραζόμενα τους ομαδοποιούνται αναπροσαρμόζοντας τις τιμές των συστατικών στο διάνυσμά τους.

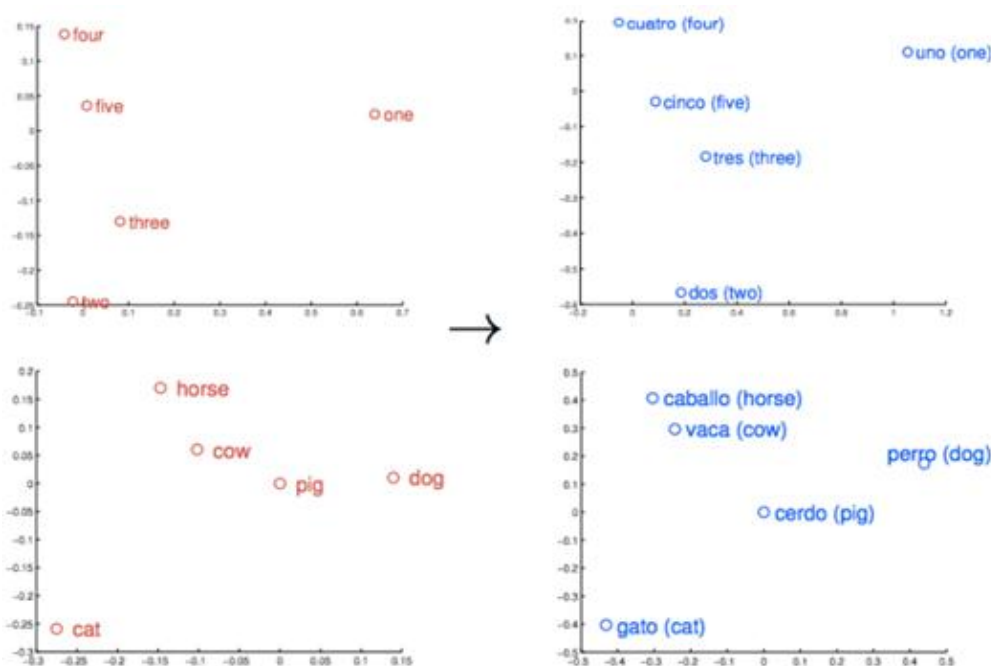
Όπως ο πίνακας του Βαν Γκογκ «τα ηλιοτρόπια» είναι μία δισδιάστατη μίξη λαδιού σε ζωγραφικό καμβά που απεικονίζει φυτική ύλη σε τρισδιάστατο χώρο στο Παρίσι στα τέλη της δεκαετίας του 1880, έτσι και οι 500 αριθμοί ως συνιστώσες ενός

¹⁸ Μέθοδος που προτάθηκε από τον Geoffrey Hinton. Η Restricted Boltzmann machine είναι ένας αλγόριθμος που εφαρμόζεται για πράξεις όπως είναι η μείωση διαστάσεων, η ταξινόμηση, η παλινδρόμηση, το φιλτράρισμα, η εκμάθηση χαρακτηριστικών και προτύπων.

διανύσματος απεικονίζουν μία λέξη ή μία ομάδα λέξεων. Αυτοί οι αριθμοί ορίζουν, θεωρητικά, τη θέση της λέξης ως ένα σημείο στον 500διάστατο διανυσματικό χώρο, όσο αυτό και αν είναι δύσκολο να στο να αποτυπωθεί ως εικόνα πρακτικά.

Ένα καλά εκπαιδευμένο σύνολο διανυσμάτων λέξεων τοποθετεί τις όμοιες λέξεις τη μία κοντά στην άλλη. Έτσι οι λέξεις: oak (βελανιδιά), elm (σημύδα) και birch (φτελιά) ομαδοποιούνται και καταλαμβάνουν ένα σημείο του χώρου, ανάλογα ομαδοποιούνται και καταλαμβάνουν ένα άλλο σημείο του χώρου οι λέξεις war (πόλεμος), conflict (σύγκρουση) και strife (διαμάχη).

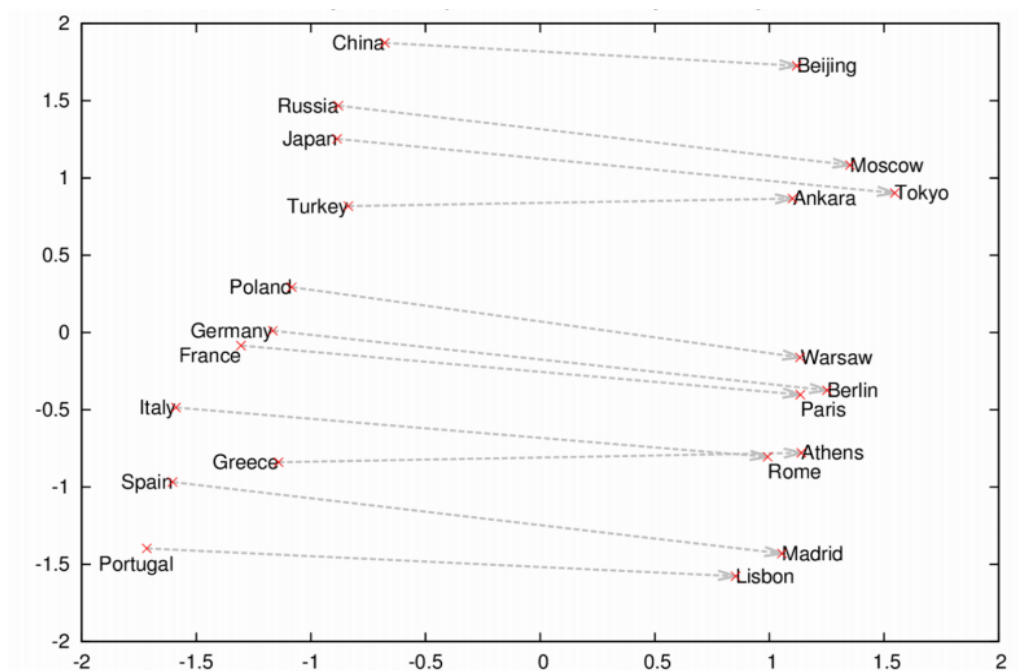
Αποδεικνύεται ότι τα αντίστοιχα σημεία σε παρόμοια πράγματα και ιδέες βρίσκονται κοντά. Οι έννοιες τους έχουν μετατραπεί σε μετρήσιμες αποστάσεις. Οι ποιοτικές τους ιδιότητες μετατρέπονται σε ποσοτικά μεγέθη, έτσι που οι αλγόριθμοι να μπορούν να λειτουργήσουν. Σημειώνεται ότι η σχέση ομοιότητας είναι η μόνη βάση για την εκπαίδευση του Word2Vec για πολλούς αποδεκτούς συνδυασμούς λέξεων. Για παράδειγμα μπορεί να μετρηθούν οι σχέσεις μεταξύ λέξεων μίας γλώσσας και να απεικονιστούν αυτές με λέξεις άλλης γλώσσας (βλ. Εικόνα 33).



Εικόνα 33: Χαρτογράφηση των σχέσεων των λέξεων μιας γλώσσας σε άλλη γλώσσα

Αυτά τα διανύσματα αποτελούν τη βάση για μια πιο ολοκληρωμένη γεωμετρία των λέξεων. Για παράδειγμα όχι μόνο η Ρώμη, το Παρίσι, το Λονδίνο, η Βιέννη, η Αθήνα κ.ά. ταξινομούνται η μία πλησίον της άλλης, αλλά στον διανυσματικό τους χώρο

ισαπέχουν από τις χώρες που αντίστοιχα είναι πρωτεύουσες τους. Δηλαδή, η Ρώμη, το Παρίσι, το Λονδίνο, η Βιέννη, η Αθήνα ισαπέχουν αντίστοιχα από την Ιταλία, τη



Εικόνα 34: Διανύσματα Χωρών και Πρωτευουσών αυτών

Γαλλία, τη Βρετανία, την Αυστρία, την Ελλάδα, κ.ο.κ. Έτσι, αν γνωρίζαμε την πρωτεύουσα της Ελλάδας και αναζητούσαμε την πρωτεύουσα της Γαλλίας, τότε η σχέση Ελλάδα - Αθήνα + Γαλλία θα μας επέστρεφε το Παρίσι (βλ. Εικόνα 34).

3.2.2.2 Παραδείγματα Κατανόησης Word2Vec

Στη συνέχεια θα αναφερθούν μερικές άλλες συσχετίσεις που η μέθοδος Word2Vec μπορεί να παράξει. Επομένως, τα μαθηματικά σύμβολα +, - και = θα αντικατασταθούν με το : που θα μεταφράζεται ως «βρίσκεται/ανήκει» και το :: που θα μεταφράζεται ως «όπως». Χαρακτηριστικό παράδειγμα αυτού αποτελεί η πρόταση «Η Αθήνα βρίσκεται στην Ελλάδα, όπως το Παρίσι βρίσκεται στην Γαλλία», η οποία θα μεταφραστεί με τα νέα πλέον πλέον σύμβολα σε Αθήνα:Ελλάδα::Παρίσι:Γαλλία.

Στον κώδικα που ακολουθεί ο οποίος υλοποιήθηκε σε πρόγραμμα Python, αντί να δίνεται σαν αποτέλεσμα η πραγματική απάντηση, αντ' αυτού θα δίνεται η λίστα των λέξεων που προτείνει το μοντέλο Word2Vec, όταν του δίνονται τα πρώτα τρία στοιχεία ως είσοδος:

```
China:Taiwan::Russia:[Ukraine, Moscow, Moldova, Armenia]
```

```
//Δύο μεγάλες χώρες και οι μικρές αποξενωμένες γειτονικές τους χώρες.
```


New York Times:Sulzberger::Fox:[Murdoch, Chernin, Bancroft, Ailes]

//Η οικογένεια Sulzberger-Ochs διοικεί τη New York Times.

//Η οικογένεια Murdoch διευθύνει την εταιρία News Corp., στην οποία ανήκει το ειδησεογραφικό πρακτορείο Fox News.

//Ο Peter Chernin ήταν ο Επικεφαλής της εταιρίας News Corp. για 13 χρόνια.

//Ο Roger Ailes είναι ο πρόεδρος του πρακτορείου Fox News.

//Οι οικογένεια Bancroft πούλησε τη Wall Street Journal στη News Corp.

Donald Trump:Republican::Barack Obama:[Democratic, GOP, Democrats, McCain]

//Αξίζει να σημειωθεί ότι, όπως ο Obama ήταν αντίπαλος με τον McCain, έτσι και το Word2Vec θεωρεί ότι ο Trump έχει μια αντιπαλότητα με τους Ρεπουμπλικανούς (ενώ ο ίδιος είναι Ρεπουμπλικανός).

Το πρότυπο Word2Vec εκπαιδεύτηκε σύμφωνα με το λεξικό Google News, το οποίο μπορεί να εισάγει οποιοσδήποτε θέλει στο αντίστοιχο πρότυπο του και να πειραματιστεί αναλόγως. Είναι αξιοσημείωτο επίσης το γεγονός ότι το Word2Vec δεν έχει καμία γνώση/πληροφορία για τον κόσμο και δεν συσχετίζεται με οποιαδήποτε συμβολική λογική ή γνώση που βασίζεται σε κανόνες. Ωστόσο, μπορεί και μαθαίνει περισσότερο, με ευέλικτο και αυτοματοποιημένο τρόπο, από ότι τα περισσότερα γραφήματα γνώσης θα μάθουν μετά από πολλά χρόνια ανθρώπινης παρέμβασης. Τέλος, αξίζει να σημειωθεί ότι το Word2Vec στα πρώιμα στάδια του εμφανίζεται στα Έγγραφα του Google News ως μια λευκή σελίδα και μέχρι το τέλος της εκπαίδευσης του είναι σε σημείο να μπορεί να υπολογίσει πολύπλοκες αναλογίες που έχουν νόημα για τον άνθρωπο.

Το πρότυπο Word2Vec μπορεί επίσης να χρησιμοποιηθεί και για άλλων ειδών συσχετίσεις, καθώς δεν είναι απαραίτητο να υπάρχουν δύο αναλογίες που αντικατοπτρίζουν η μια την άλλη. Οι εναλλακτικές συσχετίσεις που μπορεί να ζητήσει κάποιος από το Word2Vec είναι οι εξής:

- Γεωπολιτική: Ιράκ – Βία = Ιορδανία
- Διακρίσεις: Άνθρωπος – Ζώο = Ηθική
- Πρόεδρος – Δύναμη = Πρωθυπουργός
- Βιβλιοθήκη – Βιβλία = Μεγάλη Αίθουσα
- Αναλογία: Χρηματιστήριο \approx Θερμόμετρο

Καταλήγουμε λοιπόν στο συμπέρασμα ότι, χτίζοντας την αίσθηση της εγγύτητας μιας λέξης σε σχέση με άλλες παρόμοιες με αυτήν λέξεις, οι οποίες δεν είναι

υποχρεωτικό να περιέχουν τα ίδια γράμματα με αυτήν, έχουν ξεπεραστεί τα σκληρά tokens που εμφανιζόντουσαν σε μεγάλο βαθμό παλαιότερα (βλέπε Ετικετοποίηση), καθώς πλέον χρησιμοποιείται μια πιο ομαλή και γενική έννοια για τα νοήματα των λέξεων.

3.2.2.3 N-grams και Skip-grams (βήμα-βήμα αναζήτηση)

Είναι γνωστό ότι οι λέξεις διαβάζονται μέσα από το διάνυσμα που τις περιέχει μια φορά κάθε στιγμή και σαρώνεται ένα συγκεκριμένο εύρος αυτών στη συνέχεια προς τα εμπρός και προς τα πίσω. Αυτά τα εύρη είναι τα n-grams τα οποία αποτελούν μια συνεχή ακολουθία από n στοιχεία μιας δεδομένης γλωσσικής ακολουθίας και είναι η n-η στη σειρά έκδοση μετά τα unigram, bigram, trigram, four-gram και five-gram.

Αναφορικά με τα skip-grams, αυτά έγιναν δημοφιλή από τον Mikolov και υλοποιήθηκαν στην βιβλιοθήκη `deeplearning4j` αποδεικνύοντας ότι είναι πιο ακριβή από ότι άλλα μοντέλα, όπως το μοντέλο του «Ατέρμονα Διαθέσιμου Σάκου Λέξεων» (ΑΔΣΛ) ή Continuous Bag of Words λόγω των πιο γενικευμένων συμφοραζομένων που μπορούν και παράγουν.

Το n-gram στη συνέχεια τροφοδοτείται σε ένα ΝΔ για να μάθει την έννοια ενός δεδομένου διανύσματος λέξεων, όπου η έννοια ορίζεται να είναι η χρησιμότητα του ως δείκτης ορισμένων μεγαλύτερων εννοιών ή ετικετών.

3.2.2.4 Βελτιώσεις στην Επεξεργασία Φυσικής Γλώσσας: ELMO, BERT και GPT-2

Τα διανύσματα λέξεων αποτελούν την βάση των πιο πρόσφατων προόδων που σημειώθηκαν στην ΕΦΓ, συμπεριλαμβανομένων των γλωσσικών μοντέλων: ELMO, ULMFit και BERT. Ωστόσο, αυτά τα γλωσσικά μοντέλα αλλάζουν τον τρόπο με τον οποίο εκφράζονται οι λέξεις, όπως ακριβώς δηλαδή τα διανύσματα αντιπροσωπεύουν τις αλλαγές που γίνονται σε μια πρόταση.

Το Word2Vec είναι ένας αλγόριθμος που χρησιμοποιείται για την παραγωγή κατανομημένων ενσωματωμένων λέξεων ή των τύπων των λέξεων, δηλαδή οποιαδήποτε λέξη σε ένα λεξιλόγιο όπως οι λέξεις «grab» ή «go» ή «get» ενσωματώνονται σε ένα δικό τους διάνυσμα που αποθηκεύεται αποτελεσματικά σε έναν πίνακα αναζήτησης ή σε ένα λεξικό. Δυστυχώς όμως, αυτή η προσέγγιση της ενσωμάτωσης των λέξεων δεν αποδίδει την πολυσημία ή την συνύπαρξη πολλών πιθανών εννοιών για μια δεδομένη λέξη ή φράση. Για παράδειγμα, η λέξη «go» δηλώνει ένα ρήμα αλλά επίσης είναι και ένα επιτραπέζιο παιχνίδι, η λέξη «get» δηλώνει επίσης ένα ρήμα αλλά είναι και απόγονος

ενός ζώου. Η έννοια λοιπόν ενός συγκεκριμένου τύπου λέξης όπως των «get» και «go» ποικίλει ανάλογα των συμφραζομένων τους, δηλαδή τις λέξεις που τις περιβάλλουν.

Επομένως ένα πράγμα που επιδεικνύουν τα ELMO και BERT είναι ότι κωδικοποιώντας τα συμφραζόμενα μιας δεδομένης λέξης, μπορούν να επιτευχθούν καλύτερα αποτελέσματα για προβλήματα που αφορούν την ΕΦΓ.

Σε διάφορες δοκιμές που έγιναν στο τεστ SWAG, διαπιστώθηκε ότι η μείωση του σφάλματος μειώθηκε κατά 5% σε σχέση με τα διανύσματα λέξεων που είναι ανεξάρτητα συμφραζομένων λόγω της χρήση του ELMO, ενώ το BERT σημείωσε πρόσθετη μείωση κατά 66%. Ενώ, πιο πρόσφατα το GPT-2 σημείωσε εκπληκτικά αποτελέσματα στη παραγωγή Φυσικών Γλωσσών.

3.2.2.5 Ειδικές Περιπτώσεις του Word2Vec

Ο Kenny Hensens, ερευνητής δεδομένων στο Βέλγιο, εφάρμοσε μια υλοποίηση του Word2Vec και πιο συγκεκριμένα το deeplearning4j στη βάση δεδομένων OMB (Online Mendelian Inheritance In Man) της NCBI. Έπειτα, έψαξε τις λέξεις που σχετίζονται περισσότερο με την λέξη «alk» που είναι ένα γνωστό ογκογονίδιο του μη μικροκυτταρικού καρκινώματος του πνεύμονα και το Word2Vec του επέστρεψε ως απάντηση «nonsmall, carcinomas, carcinoma, mapdkd». Λόγω αυτής της ανακάλυψης ίδρυσε νέες αναλογίες μεταξύ άλλων φαινοτύπων καρκίνου και των γονότυπων τους. Αυτό είναι ένα μόνο παράδειγμα των συσχετίσεων που είναι ικανό να μάθει το Word2Vec από ένα μεγάλο σώμα κειμένου. Ένα άλλο παράδειγμα της εφαρμογής του Word2Vec αφορά τους εμπόρους οι οποίοι ενδέχεται να επιδιώξουν τη δημιουργία σχέσεων μεταξύ των προϊόντων για να κατασκευάσουν έναν μηχανισμό συστάσεων.

3.2.2.6 Η ευρεσιτεχνία της Google για το Word2Vec

Όπως προαναφέρθηκε προηγουμένως το Word2Vec είναι μια υπολογιστική μέθοδος που εισήχθη από μια ομάδα ερευνητών στη Google και φιλοξενεί μια έκδοση ανοιχτού κώδικα του Word2Vec που έχει κυκλοφορήσει υπό την άδεια Apache 2.0 με επικεφαλής τον Tomas Mikolov. Το 2014 ωστόσο ο Mikolov εγκατέλειψε τη Google για να πάει στο Facebook με αποτέλεσμα το Μάιο του 2015, να δοθεί στην Google μια πατέντα/ευρεσιτεχνία για τη μέθοδο αυτή, η οποία δεν κατήργησε την άδεια Apache υπό την οποία είχε ανακοινωθεί αρχικά.

3.2.2.7 Χρήση του Word2Vec στις ξένες γλώσσες

Είναι γνωστό ότι οι λέξεις οποιασδήποτε γλώσσας μπορούν να ενσωματωθούν σε διανύσματα χρησιμοποιώντας το Word2Vec και το deeplearning4j. Όμως, η προεπεξεργασία της Φυσικής Γλώσσας που χρειάζεται να γίνει μπορεί να απαιτεί πολύπλοκα γλωσσικά δεδομένα και εργαλεία που δεν συμπεριλαμβάνονται στις υπάρχουσες βιβλιοθήκες. Για το λόγο αυτό, η ομάδα Επεξεργασίας της Φυσικής Γλώσσας στο πανεπιστήμιο του Stanford διέθεσε μια σειρά εργαλείων Java που ειδικεύονται στην Ετικετοποίηση των μερών του λόγου και στην αναγνώριση της ονομαστικής οντότητας για γλώσσες όπως τα Κινέζικα, τα Αραβικά, τα Γαλλικά, τα Γερμανικά και τα Ισπανικά, ενώ για τα Ιαπωνικά υπάρχει το εργαλείο Kuromoji το οποίο είναι ένας μορφολογικός αναλυτής ανοικτού κώδικα που γράφεται σε γλώσσα Java και χρησιμοποιείται ευρέως.

3.2.2.8 Ακολουθιακά Διανύσματα

Το deeplearning4j περιλαμβάνει την εντολή Sequence Vectors, η οποία στην ουσία είναι μια εντολή αφαίρεσης που επιτρέπει στον χρήστη την εξαγωγή χαρακτηριστικών από οποιαδήποτε ακολουθία, συμπεριλαμβανομένων των προφίλ στα μέσα κοινωνικής δικτύωσης, των συναλλαγών, των πρωτεϊνών κ.λπ. Εάν λοιπόν τα δεδομένα μπορούν να εκφραστούν ως ακολουθίες, οι ακολουθίες αυτές μπορούν να εκπαιδευτούν χρησιμοποιώντας τα εργαλεία skip-gram και hierarchic softmax με την εντολή Abstract Vectors.

3.2.2.9 Χαρακτηριστικά του Word2Vec στο deeplearning4j

- Μόλις προστεθεί στο μοντέλο η σειριοποίηση τότε τα βάρη μπορούν να ενημερωθούν. Δηλαδή, η κατάσταση του μοντέλου μπορεί να ενημερωθεί με, για παράδειγμα, 200GB νέου κειμένου, με τη χρήση των εντολών LoadFullModel, TokenizerFactory, SentenceIterator και με το κάλεσμα της εντολής fit() στο αποκατεστημένο μοντέλο.
- Προστέθηκε η δυνατότητα πολλαπλών πηγών δεδομένων που δίνει τη δυνατότητα κατασκευής λεξιλογίου.
- Οι εποχές και οι επαναλήψεις μπορούν να προσδιοριστούν ξεχωριστά, παρόλο που και οι δύο εμφανίζονται τυπικά με τον κωδικό “1”.

- Η εντολή `Word2Vec.Builder` προσφέρει τη χρήση της εντολής `HugeModelExpected`, η οποία εάν οριστεί ως αληθής, τότε το λεξιλόγιο θα περιορίζεται περιοδικά κατά τη διάρκεια της κατασκευής.
- Δύο νέες εντολές έχουν εισαχθεί επίσης, το `WriteFullModel` και το `LoadFullModel`, οι οποίες μπορούν να αποθηκεύουν και να φορτώνουν αντίστοιχα την κατάσταση ενός πλήρους προτύπου.
- Ένας αξιοπρεπής χώρος εργασίας θα πρέπει να είναι σε θέση να χειριστεί ένα λεξιλόγιο που περιέχει μερικά εκατομμύρια λέξεις. Η υλοποίηση του `Word2Vec` στο `deeplearning4j` μπορεί να μοντελοποιήσει μερικά terabyte δεδομένων σε ένα μόνο μηχάνημα. Κατά προσέγγιση, τα μαθηματικά που χρειάζονται για αυτήν την ενέργεια εκφράζονται από την εντολή `vectorSize * 4 * 3 * vocab.size()`.

3.3 Άλλες Μέθοδοι Νευρωνικών Δικτύων για την Επεξεργασία Φυσικής Γλώσσας

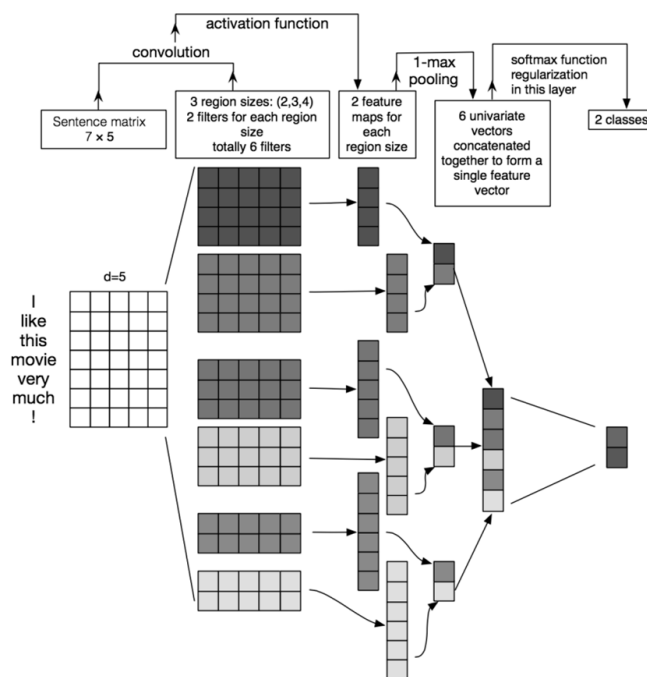
3.3.1 Τα Συνελικτικά Νευρωνικά Δίκτυα για την Επεξεργασία Φυσικής Γλώσσας

Τα τελευταία χρόνια τα ΣΝΔ έχουν γνωρίσει μια τεράστια επιτυχία και φήμη με αποτέλεσμα να εφαρμόζονται πλέον σε πολλούς τομείς και πιο πρόσφατα στην ΕΦΓ με ελπιδοφόρα αποτελέσματα.

Αναφορικά με την εφαρμογή των ΣΝΔ στον τομέα της ΕΦΓ αξίζει να σημειωθεί πως αντί των εικονοστοιχείων μιας εικόνας, η είσοδος στα περισσότερα αυτά προβλήματα είναι προτάσεις ή έγγραφα που εκφράζονται με τη μορφή πίνακα. Κάθε σειρά του πίνακα λοιπόν αντιστοιχεί σε ένα token, το οποίο είναι συνήθως μια λέξη, αλλά θα μπορούσε να είναι και ένας χαρακτήρας, με άλλα λόγια κάθε σειρά του πίνακα είναι ένα διάνυσμα που ενσωματώνει μια λέξη. Συνήθως, αυτά τα διανύσματα είναι στην ουσία νευρώνες για τον συνδυασμό λέξεων όπως τα `Word2Vec` και `GloVe` που είδαμε προηγουμένως, αλλά θα μπορούσαν επίσης να είναι και διανύσματα τα οποία κατατάσσουν τις λέξεις σε λεξιλόγια. Για παράδειγμα, για μια πρόταση 10 λέξεων και χρησιμοποιώντας μια διάσταση μεγέθους 100 θα είχαμε έναν πίνακα 10 x 100 ως είσοδο στο ΣΝΔ, το οποίο στην ουσία είναι η «εικόνα» του δικτύου.

Στο σχήμα της Εικόνας που ακολουθεί περιγράφεται αναλυτικά ο τρόπος που λειτουργούν τα ΣΝΔ για την ΕΦΓ και πιο συγκεκριμένα όταν χρησιμοποιούνται για δεδομένα κειμένων. Παρατηρείται στην διαδικασία αυτή ότι το αποτέλεσμα κάθε συνέλιξης θα πυροδοτηθεί όταν εντοπιστεί από το δίκτυο ένα ειδικό μοτίβο.

Επομένως, αλλάζοντας το μέγεθος των πυρήνων και συγκαλύπτοντας τις εξόδους τους, επιτρέπεται στο δίκτυο να ανιχνεύει πρότυπα πολλαπλών μεγεθών, δηλαδή 2,3 ή 5 παρακείμενες λέξεις. Τα πρότυπα αυτά θα μπορούσαν να είναι εκφράσεις όπως για παράδειγμα «I hate», «very good» και τα όποια θα μπορούσαν να αναγνωριστούν από ένα ΣΝΔ ανεξάρτητα από τη θέση τους στην πρόταση.



Εικόνα 35: Απεικόνιση της αρχιτεκτονικής ενός ΣΝΔ για την ταξινόμηση προτάσεων.

Αναλυτικά στην εικόνα 35 παρουσιάζονται τρία μεγέθη, 2, 3, 4, των περιοχών των φίλτρων, καθένα από τα οποία περιέχει 2 φίλτρα. Κάθε φίλτρο εκτελεί την πράξη της Συνέλιξης στον πίνακα της πρότασης και έπειτα παράγει τους χάρτες χαρακτηριστικών οι οποίοι έχουν μεταβλητό μήκος. Στη συνέχεια, πραγματοποιείται η πράξη της συγκέντρωσης σε κάθε χάρτη, δηλαδή καταγράφεται ο μεγαλύτερος αριθμός από κάθε χάρτη χαρακτηριστικών. Επομένως, παράγεται ένα μονοδιάστατο διάνυσμα χαρακτηριστικών και από τους έξι χάρτες και στη συνέχεια τα χαρακτηριστικά αυτά συνδυάζονται για να σχηματίσουν ένα διάνυσμα χαρακτηριστικών για το προτελευταίο επίπεδο. Το τελευταίο επίπεδο λαμβάνει τελικά αυτό το διάνυσμα χαρακτηριστικών ως είσοδο και το χρησιμοποιεί για να ταξινομήσει την πρόταση. Στο συγκεκριμένο σχήμα χρησιμοποιείται η δυαδική ταξινόμηση και ως εκ τούτου απεικονίζονται δύο πιθανές καταστάσεις εξόδου.

3.3.2 Τα Επαναληπτικά και τα Αναδρομικά Νευρωνικά Δίκτυα για την Επεξεργασία Φυσικής Γλώσσας

Όπως τα ΣΝΔ, έτσι και τα ΕΝΔ άρχισαν να εφαρμόζονται πολύ τα τελευταία χρόνια στον τομέα της ΕΦΓ παρουσιάζοντας ελπιδοφόρα για το μέλλον αποτελέσματα. Αναφορικά λοιπόν με την εφαρμογή των ΕΝΔ στον τομέα της ΕΦΓ, παρατηρείται ότι αξιοποιώντας τη γνώση των νευρώνων για τον συνδυασμό λέξεων ενισχύεται η ακρίβεια του ΕΝΔ λόγω της ενσωμάτωσης των νέων πληροφοριών (λεξικογραφικές και σημασιολογικές) σχετικά με τις λέξεις αυτές, πληροφορία η οποία έχει εκπαιδευτεί σε ένα πολύ μεγάλο σύνολο δεδομένων. Η υπεροχή των ΕΝΔ μπορεί να παρατηρηθεί στο ακόλουθο παράδειγμα: εάν ζητούσαμε από ένα μοντέλο MM να διαβάσει τις ακόλουθες προτάσεις «Πήγα στο Νεπάλ το 2009» και «Το 2009, πήγα στο Νεπάλ» και να εξάγει το έτος κατά το οποίο ο αφηγητής επισκέφθηκε το Νεπάλ, θα θέλαμε να αναγνωρίσει το έτος 2009 ως μια σχετική πληροφορία επί του θέματος, είτε αυτή εμφανίζεται στην έκτη θέση είτε στη δεύτερη θέση της πρότασης. Εάν υποθέσουμε λοιπόν ότι το δίκτυο που εκπαιδεύτηκε είναι τύπου Πρόσθιας Τροφοδότησης και επεξεργάζεται προτάσεις σταθερού μήκους τότε αυτό θα έχει ξεχωριστές παραμέτρους για κάθε χαρακτηριστικό εισόδου και θα πρέπει να μάθει όλους τους κανόνες της γλώσσας ξεχωριστά. Εν αντιθέσει, ένα ΕΝΔ είναι ικανό να μοιράζεται τα ίδια βάρη στα διάφορα χρονικά βήματα, γεγονός που το καθιστά ανώτερο από τα τυπικά ΝΔ για την επίλυση τέτοιων προβλημάτων.

Για την Ταξινόμηση του κειμένου με χρήση των ΕΝΔ θα πρέπει σε πρώτη φάση στον κώδικα να εισαχθεί η εντολή “Beautiful Soup” για την αφαίρεση κάποιων ετικετών HTML αλλά και ορισμένων ανεπιθύμητων χαρακτήρων ως ακολούθως:

```
def clean_str(string):
    string = re.sub(r"\\", "", string)
    string = re.sub(r"\"", "", string)
    string = re.sub(r"\"", "", string)
    return string.strip().lower()

texts = [ ]; labels = [ ]

for i in range(df.message.shape[0]):
    text = BeautifulSoup(df.message[i])
    texts.append(clean_str(str(text.get_text().encode())))
```

```
for i in df['class']:
```

```
    labels.append(i)
```

Στη συνέχεια, θα πρέπει να χρησιμοποιηθεί η βιβλιοθήκη ανοικτού κώδικα Keras στα δεδομένα κειμένου, αλλά πρώτα είναι αναγκαίο να γίνει η προεπεξεργασία της. Για αυτό το σκοπό, μπορεί να χρησιμοποιηθεί η κλάση του Ετικετοποιητή Keras', η οποία δέχεται την εντολή `num_words` που είναι ο μέγιστος αριθμός λέξεων που αποθηκεύονται μετά την ετικετοποίηση με βάση τη συχνότητα τους.

```
MAX_NB_WORDS = 20000
```

```
tokenizer = Tokenizer (num_words=MAX_NB_WORDS)
```

```
tokenizer.fit_on_texts(texts)
```

Μόλις λοιπόν ο Ετικετοποιητής τοποθετηθεί στα δεδομένα, μπορεί να χρησιμοποιηθεί για τη μετατροπή των αλυσίδων των χαρακτήρων του κειμένου σε ακολουθίες αριθμών, όπου αυτοί οι αριθμοί αντιπροσωπεύουν τη θέση της κάθε λέξης στο λεξικό (η διαδικασία αυτή μπορεί να θεωρηθεί ως χαρτογράφηση).

Τέλος, αναφορικά με την επέκταση των ΕΝΔ, τα ΕΝΔ δηλαδή, αξίζει να σημειωθεί ότι αν και βρίσκονται σε πρώιμο στάδιο αναφορικά με την εφαρμογή τους στον τομέα της ΕΦΓ, φάνηκε να επιδεικνύουν ελπιδοφόρα αποτελέσματα για το μέλλον που τους επιφυλάσσεται. Αυτό συμβαίνει διότι όταν επεξεργάζονται προτάσεις φυσικής γλώσσας, η δομή του δέντρου στην αρχιτεκτονική τους μπορεί να προσαρμοστεί με βάση τη δομή που ανέλυσε τη πρόταση και που παρέχεται από έναν αναλυτή φυσικής γλώσσας.

3.4 Συμπεράσματα

Στο παρόν κεφάλαιο μελετήθηκαν με σαφήνεια οι προσεγγίσεις της ΕΦΓ και πιο συγκεκριμένα, τα τυπικά ΝΔ, τα ΣΝΔ και τα ΕΝΔ. Συγκρίνοντας λοιπόν αυτές τις τρεις μεθόδους και ξεκινώντας με τα τυπικά ΝΔ μπορεί να παρατηρηθεί ότι μέχρι και σήμερα αποτελούν την πιο διαδεδομένη και με σίγουρα αποτελέσματα προσέγγιση της ΕΦΓ και γι' αυτό το λόγο έχουν κατασκευαστεί τόσες μέθοδοι για την επίλυση προβλημάτων ΕΦΓ όπως είδαμε προηγουμένως. Τέλος, αναφορικά με τα ΣΝΔ και τα ΕΝΔ αν και η σχέση τους με την Φυσική Γλώσσα βρίσκεται σε ένα πρώιμο στάδιο, φάνηκε να επιδεικνύουν ελπιδοφόρα για το μέλλον αποτελέσματα, γεγονός που ωθεί τους ερευνητές ανά τον κόσμο να εστιάζουν τις έρευνες τους ολοένα και περισσότερο σε αυτό το πολύ καινούργιο και καινοτόμο πεδίο, των ΣΝΔ και ΕΝΔ δηλαδή στην

ΕΦΓ.

Συμπερασματικά λοιπόν καταλήγουμε ότι τα ΝΔ και οι διάφορες κατηγορίες αυτών έχουν συντελέσει σημαντικά τα τελευταία χρόνια στην γρηγορότερη και αποτελεσματικότερη επίλυση προβλημάτων σε ποικίλους τομείς. Ειδικότερα στην ΕΦΓ, τα ΝΔ χρησιμοποιούνται ολοένα και περισσότερο για την επεξεργασία προτάσεων, ταξινόμηση κειμένων κ.α. αλλά βρίσκονται ακόμη σε πειραματικό στάδιο. Ωστόσο, παρατηρείται ότι οι ερευνητές και οι προγραμματιστές έχουν εστιάσει την προσοχή τους στην περαιτέρω βελτίωση αυτών των νέων υπολογιστικών μεθόδων καθότι έχουν αποδείξει ότι, με την συνεισφορά τους μπορούν να ενισχύσουν σημαντικά τη ζωή των ανθρώπων τα επόμενα χρόνια επιλύοντας προβλήματα σε αναρίθμητους τομείς που μέχρι πρότινος δεν έχει υπάρξει άλλη μέθοδος που να το έχει καταφέρει αυτό.

3.5 Βιβλιογραφικές Αναφορές

1. Angeline, P. J., Saunders, G. M. & Pollack, J.B. (1994). An evolutionary algorithm that constructs recurrent neural networks. *IEEE Transactions on Neural Networks*, 5(1), pp. 54–65.
2. Chinea, A. (2009). Understanding the Principles of Recursive Neural Networks: A Generative Approach to Tackle Model Complexity. In: Alippi, C., Polycarpou, M., Panayiotou, C., Ellinas, G. (Eds.), *ICANN. LNCS 5768*, (pp. 952-963). Heidelberg: Springer.
3. Conneau, A., Schwenk, H., Le Cun, Y. & Barrault, L. (2016). Very Deep Convolutional Networks for Natural Language Processing. *KI - Künstliche Intelligenz*, 26 (1), pp.1107-1116.
4. Fergus, R. & Zeiler, M.D. (2014). Visualizing and understanding convolutional networks. In Fleet, D., Pajdla, T., Schiele, B. & Tuytellers, T. (Eds.), *Computer Vision – ECCV 2014* (pp.818-833). Zurich: Springer.
5. Goldberg, Y. & Omer, L. (2014). *word2vec Explained: Deriving Mikolov et al.'s Negative-Sampling Word-Embedding Method*. Retrieved January 20, 2019, from <https://arxiv.org/abs/1402.3722>
6. Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., Szegedy, C. & Liu, W. (2015, June 7-12). *Going deeper with convolutions*. Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Doi: <https://arxiv.org/abs/1409.4842>
7. Medsker, L. & Jain, L. C. (1999). *Recurrent Neural Networks: Design and Applications*. USA: CRC Press.
8. Mikolov, T., Corrado, G. S., Chen, K. & Dean, J. (2013, January). *Efficient Estimation of Word Representations in Vector Space*. Paper presented at the Proceedings of the International Conference on Learning Representations (ICLR). Doi: https://www.researchgate.net/publication/319770439_Efficient_Estimation_of_Word_Representations_in_Vector_Space
9. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. & Dean, J. (2013, December 5-10). *Distributed representations of words and phrases and their compositionality*. Paper presented at the Proceedings of the 26th

- International Conference on Neural Information Processing Systems. Doi:
<https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf>
10. Rong, X. (2016). Word2vec Parameter Learning Explained. *ArXiv e-prints*, pp.1-21.
 11. Sutskever, I., Krizhevsky, A. & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*, 25(2), pp.1097-1105.
 12. Trupti, K. (2018). Recurrent Neural Network and its Various Architecture Types. *International Journal of Research and Scientific Innovation (IJRSI)* 5(3), pp.124-129.
 13. Xiangang, L. & Xihong, W. (2014, April). *Constructing Long Short-Term Memory based Deep Recurrent Neural Networks for Large Vocabulary Speech Recognition*. Paper presented at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Doi:
https://www.researchgate.net/publication/308851552_Constructing_long_short-term_memory_based_deep_recurrent_neural_networks_for_large_vocabulary_speech_recognition
 14. Zisserman, A. & Simonyan, K. (2015). *Very deep convolutional networks for largescale image recognition*, pp.1-14. Retrieved January 27, 2019, from <https://arxiv.org/abs/1409.1556>

ΚΕΦΑΛΑΙΟ 4

Συμπεράσματα και Προτάσεις για περαιτέρω έρευνα

Συμπερασματικά καταλήγουμε ότι τα ΝΔ, τα οποία είναι εννοιολογικά και δομικά εμπνευσμένα από τα ανθρώπινα και ζωικά νευρικά συστήματα, παρουσιάζουν τεράστιο ενδιαφέρον σε συνδυασμό με τη Εκμάθηση Βάθους λόγω των μεγάλων ποσοστών ακρίβειας που έχουν σημειώσει σε διάφορους τομείς συμπεριλαμβανομένης και της ΕΦΓ όπως περιγράφηκε αναλυτικά στο Κεφάλαιο 3.

Ωστόσο, τα ΝΔ μόλις πρόσφατα άρχισαν να εφαρμόζονται στον τομέα της ΕΦΓ με ελπιδοφόρα αποτελέσματα. Μια σημαντική πρόταση για περαιτέρω έρευνα αφορά την εφαρμογή των ΝΔ στη μέθοδο για την επεξεργασία της ΝΕΓ που αποτελείται από

1. τις Μητροειδείς και τις Σπονδυλωτές Μητροειδείς Γραμματικές – Χαρακτηριστικούς Εκθέτες, οι οποίες ελέγχουν, δηλαδή παράγουν και αναγνωρίζουν με αιτιοκρατικές διαδικασίες τη δομή των προτάσεων και κατ' επέκταση των κειμένων της ΝΕΓ ,
2. το Ηλεκτρονικό Υπολογιστικό Βασικό Νεοελληνικό Πολυλεξικό (ΒΝΠ) που περιέχει κωδικοποιημένα, αρχειοθετημένα και υπολογιστικά αξιοποιήσιμα, αναλυτικά τα γλωσσολογικά δεδομένα της ΝΕΓ και
3. τους αλγορίθμους του ΒΝΠ που περιγράφουν τη λειτουργία της ΝΕΓ και μετασχηματίζουν τις δομές της ΝΕΓ σε αποδεκτές προτάσεις και κείμενα της ΝΕΓ, όπως επίσης και τις αναγνωρίζουν για περαιτέρω αξιοποίηση.

Η όλη προσπάθεια για την επεξεργασία της ΝΕΓ έγινε αξιοποιώντας την παραδοσιακή περιγραφή και μελέτη της, αξιοποιώντας επίσης υπολογιστικά τις

εσωτερικές διεργασίες της και αναλυτικά τα γλωσσολογικά δεδομένα της, αποφεύγοντας τις ξένες προς τη δομολειτουργία της ΝΕΓ ερμηνείες και περιγραφές τους καθώς και τις μεθόδους για την παραγωγή με τη βοήθεια ευρεστικών ή προσεγγιστικών στατιστικών μεθόδων βιομηχανικών υπολογιστικών γλωσσολογικών προϊόντων. Σημειώνεται ότι οι παραδοσιακοί τύποι αλγόριθμων που σχεδιάστηκαν ήταν μικροί, ευέλικτοι, ταχύτατοι, γραμμικής πολυπλοκότητας με ακρίβεια αποτελεσμάτων 100%. Η πρόταση είναι η υλοποίηση της παραπάνω μεθόδου με τις τεχνικές των ΝΔ για περαιτέρω βελτίωση.

5. ΠΑΡΑΡΤΗΜΑΤΑ

**ΠΑΡΑΡΤΗΜΑ 1: ΓΛΩΣΣΑΡΙ ΕΛΛΗΝΟΑΓΓΛΙΚΩΝ
ΟΡΩΝ ΤΗΣ ΔΙΑΤΡΙΒΗΣ**

A

Αβαθή Νευρωνικά Δίκτυα	Shallow Neural Networks
Αισθητήρας	Perceptron
Αιτιοκρατικό	Deterministic
Ακολουθία	Sequence
Αλγόριθμος	Algorithm
Αμφίδρομη Μακρά Βραχυπρόθεσμη Μνήμη	Long Short Term Memory
Αναγνώριση Ομιλίας	Speech Recognition
Αναγνώριση Κειμένου	Text Recognition
Αναγνώριση Προτύπων	Pattern Recognition
Αναδρομικά Νευρωνικά Δίκτυα	Recurrent Neural Networks
Ανακάλυψη Προτύπων	Pattern Recognition
Ανάκληση	Recall
Ανάλυση	Analysis
Αναπαράσταση	Representation
Ανασυγκρότηση	Reconstruction
Ανεξαρτήτων Συμφραζομένων	Context Free
Ανίχνευση	Detection
Αντικατάσταση	Substitution
Αντικείμενο	Object
Αντωνυμία	Pronoun
Απαιτήσεις Συστήματος	System Requirements
Απόκτηση	Acquisition
Αρχικά Δέντρα	Initial Trees
Αρχιτεκτονικές	Architectures
Αυτοματισμός	Automation
Αυτόματος Κωδικοποιητής	Automatic Encoder

B

Βαθιά Νευρωνικά Δίκτυα	Deep Neural Networks
Βάθος	Depth
Βάρος	Weight
Βιβλιοθήκη	Library
Βιολογικά Νευρωνικά Δίκτυα	Biological Neural Networks
Βιομηχανία Υγείας	Healthcare Industry

Βοηθητικά Δέντρα

Auxiliary Trees

Γ

Γλώσσα

Language

Γλώσσες Ανεξαρτήτων Συμφραζομένων

Context-Free Languages

Γλώσσες Ευαίσθητες Συμφραζομένων

Context-Sensitive Languages

Γλωσσολογία

Linguistics

Γραμματικές Γενικευμένης Φρασεοδομής

Generalized Phrase Structure
Grammars

Γραμματικές Οδηγούμενες από την Κεφαλή

Head Driven Grammars

Γραμματικές Συνδεδεμένων Δέντρων

Tree Adjoining Grammars

Γραμματική

Grammar

Γραμμική Παλινδρόμηση

Linear Regression

Γραμμικό Μοντέλο

Linear Model

Γράφημα

Graph

Δ

Δάσος με Τυχαιότητα

Random Forest

Δεδομένα

Data

Δεκτικά Πεδία

Receptive Fields

Δέντρα Απόφασης

Decision Trees

Δέντρα που ενισχύουν τη διαβάθμιση

Gradient-Boosting Trees

Διάνυσμα Χαρακτηριστικών

Features Vector

Διάσταση

Dimension

Δίκτυα Hopfield

Hopfield Networks

Δίκτυα Kohonen

Kohonen Networks

Δίκτυα πρόσθιας τροφοδότησης

Feedforward Networks

Δίκτυο

Networks

Δομή

Structure

Ε

Εγκατάσταση

Install

ΕιδικάΣυστήματα

Expert Systems

Εικονοστοιχεία

Pixels

Είσοδος

Input

Εκπαίδευση

Training

Εμπόριο

Marketing

Ενεργητικές Γνώσεις

Actionable Insights

Ενεργοποίηση	Activation
Ενίσχυση/Βελτίωση	Augmentation
Εντολή	Command
Εξαγωγή Χαρακτηριστικών	Feature Extraction
Έξοδος	Output
Εμπόριο	Marketing
Επαναλαμβανόμενα Νευρωνικά Δίκτυα	Recursive Neural Networks
Επαυξημένα Δίκτυα Μετάβασης	Augmented Transition Networks
Επεξεργασία	Processing
Επεξεργασία Φυσικής Γλώσσας	Natural Language Processing
Επεξεργαστής	Processor
Επίθετο	Adjective
Επίπεδο	Layer
Επίπεδο Εισόδου	Input Layer
Επίπεδο Εξόδου	Output Layer
Επίρρημα	Adverb
Επιφώνημα	Interjection
Εποπτευόμενη Εκμάθηση	Supervised Learning
Εργαλείο	Tool
Ετικετοποίηση	Tagging
Εφαρμογές	Applications
	I
Ιεραρχία	Hierarchy
Ιεραρχική Ομαδοποίηση	Hierarchical clustering
	K
Καθολικότητα	Universality
Κανόνας	Rule
Κανόνες Παραγωγής	Production Rules
Κανονικές Γλώσσες	Regular Languages
Καρέ	Frame
Κατανόηση	Understanding/Comprehension
Κατάσταση	State
Κιβωτισμοί Λέξεων	Word Embeddings
Συνδυασμός/Ενσωμάτωση	Embedding
Αποδεκτός Συνδυασμός Λέξεων	Word Embedding

Κ-Ομαδοποίηση	K-means Clustering
Κόμβος	Node
Κρυφό	Hidden
Κρυφό Επίπεδο	Hidden Layer
Κυβερνητικές Οργανώσεις	Government Organizations
Κωδικοποίηση	Coding
Κωδικοποιητής	Encoder
Λ	
Λειτουργικές Γραμματικές	Functional Grammars
Λέξη	Word
Λεξικές Λειτουργικές Γραμματικές	Lexical Functional Grammars
Λεξικό	Dictionary
Λίστα Αναπαραγωγής	Playlist
Λογιστική Παλινδρόμηση	Logistic Regression
Μ	
Εκμάθηση Βάθους	Deep Learning
Εκμάθηση Μηχανής	Machine Learning
Μέθοδος	Method
Μέρη του Λόγου	Part of speech
Μετασχηματισμός	Transformation/Modification
Μετασχηματιστικές Γραμματικές	Transformational Grammars
Μη - Εποπτευόμενη Εκμάθηση	Unsupervised Learning
Μηχανές Γραμμικής Διανυσματικής Υποστήριξης	Linear Support Vector Machines
Μηχανές Διανυσμάτων Υποστήριξης	Support Vector Machines
Μηχανή	Machine
Μηχανική Μετάφραση	Machine Translation
Μνήμη	Memory
Μοίρα	Degree
Μοντέλο /Πρότυπο	Model
Μοτίβο	Pattern
Ν	
Νευρώνας	Neuron
Νευρωνικά Δίκτυα	Neural Networks

Ο

Όγκοι Δεδομένων	Data Volumes
Όγκος Εισόδου	Input Volume
Όγκος Εξόδου	Output Volume
Ομαδοποίηση	Clustering
Ομοιότητα	Similarity
Οπισθοδιάδοση	BackPropagation
Ουσιαστικό	Noun

Π

Παλινδρόμηση	Regression
Παράγω	Produce
Παραδοσιακός Προγραμματισμός	Traditional Programming
Παραδοσιακός Υπολογισμός	Traditional Computing
Πεπερασμένος	Finite
Περιβάλλον Εκμάθησης	Learning Environment
Περιγραφή	Description
Πλεονέκτημα	Advantage
Πλήρως Συνδεδεμένο Νευρωνικό Δίκτυο	Fully Connected Neural Network
Πραγματικές Πληροφορίες	Actionable Insights
Πρόθεση	Preposition
Προσθήκη/Συμπλήρωση	Adjunction
Πρόταση/Κείμενο	Sentence
Προτεινόμενο Σύστημα	Recommended System
Πυρήνας	Kernel

Ρ

Ρήμα	Verb
Ρίζα	Root

Σ

Σιγμοειδής Συνάρτηση	Sigmoid Function
Στοχαστικές Γλώσσες	Stochastic Languages
Στοχαστικές Γραμματικές	Stochastic Grammars
Σύζευξη	Conjunction
Συμβατός	Compatible
Συμβατός Προγραμματισμός	Traditional Programming

Σύμβολο	Symbol
Αλυσίδα Χαρακτήρων	String
Συμπεράσματα	Inferring
Συνάρτηση	Function
Συνάρτηση Ενεργοποίησης	Activation Function
Συνάρτηση Κατωφλίου	Threshold Function
Συνάρτηση Ράμπας	Rectified Linear Unit
Συνεδρία	Conference
Συνελικτικά Νευρωνικά Δίκτυα	Convolutional Neural Network
Συνέλιξη	Convolution
Συνεχής τσάντα λέξεων	Continuous bag of words
Σύνολο	Set
Σύστημα	System
Σφάλμα	Error
Σχεδιασμός	Design
Σώμα Κειμένων	Corpus

T

Ταξινόμηση	Classification
Τεχνητά Νευρωνικά Δίκτυα	Artificial Neural Networks
Τεχνητή Νοημοσύνη	Artificial Intelligence
Τοπικά Συνδεδεμένα Δίκτυα	Locally Connected Networks
Τοπολογία	Topology
Τροποποιημένο	Modified
Τυπική Γλώσσα	Formal Language
Τυπική Γραμματική	Formal Grammar
Τυπικό	Typical

Y

Υπερβολική Εφαπτομένη	Hyberbolic Tangent
Υπερπαράμετρος	Hyperparameter
Υποδειγματοληψία	Pooling
Υποκείμενο	Subject
Υπολογισμός	Computation
Υπολογιστική Γλωσσολογία	Computational Linguistics
Υπολογιστική Όραση	Computer Vision

X

Χάρτης Χαρακτηριστικών
Χρηματοοικονομική Βιομηχανία

Feature Map
Finance Industry

**ΠΑΡΑΡΤΗΜΑ 2: ΓΛΩΣΣΑΡΙ ΑΓΓΛΟΕΛΛΗΝΙΚΩΝ
ΟΡΩΝ ΤΗΣ ΔΙΑΤΡΙΒΗΣ**

A

Acquisition	Απόκτηση
Actionable Insights	Ενεργητικές Γνώσεις
Actionable Insights	Πραγματικές Πληροφορίες
Activation	Ενεργοποίηση
Activation Function	Συνάρτηση Ενεργοποίησης
Adjective	Επίθετο
Adjunction	Προσθήκη/Συμπλήρωση
Advantage	Πλεονέκτημα
Adverb	Επίρρημα
Algorithm	Αλγόριθμος
Analysis	Ανάλυση
Applications	Εφαρμογές
Architectures	Αρχιτεκτονικές
Artificial Intelligence	Τεχνητή Νοημοσύνη
Artificial Neural Networks	Τεχνητά Νευρωνικά Δίκτυα
Augmentation	Ενίσχυση/Βελτίωση
Augmented Transition Networks	Επαυξημένα Δίκτυα Μετάβασης
Automatic Encoder	Αυτόματος Κωδικοποιητής
Automation	Αυτοματισμός
Auxiliary Trees	Βοηθητικά Δέντρα

B

BackPropagation	Οπισθοδιάδοση
Biological Neural Networks	Βιολογικά Νευρωνικά Δίκτυα

C

Classification	Ταξινόμηση
Clustering	Ομαδοποίηση
Coding	Κωδικοποίηση
Command	Εντολή
Compatible	Συμβατός
Computation	Υπολογισμός
Computational Linguistics	Υπολογιστική Γλωσσολογία
Computer Vision	Υπολογιστική Όραση
Conference	Συνεδρία

Conjunction	Σύζευξη
Context Free	Ανεξαρτήτων Συμφραζομένων
Context-Free Languages	Γλώσσες Ανεξαρτήτων Συμφραζομένων
Context-Sensitive Languages	Γλώσσες Ευαίσθητες Συμφραζομένων
Continuous bag of words	Συνεχής τσάντα λέξεων
Convolution	Συνέλιξη
Convolutional Neural Networks	Συνελικτικά Νευρωνικά Δίκτυα
Corpus	Σώμα Κειμένων

D

Data	Δεδομένο
Data Volumes	Όγκοι Δεδομένων
Decision Trees	Δέντρα Απόφασης
Deep Learning	Εκμάθηση Βάθους
Deep Neural Networks	Βαθιά Νευρωνικά Δίκτυα
Degree	Μοίρα
Depth	Βάθος
Description	Περιγραφή
Design	Σχεδιασμός
Detection	Ανίχνευση
Deterministic	Αιτιοκρατικό
Dictionary	Λεξικό
Dimension	Διάσταση

E

Embedding	Κιβωτισμός
Encoder	Κωδικοποιητής
Error	Σφάλμα
Expert Systems	Ειδικά Συστήματα

F

Feature Map	Χάρτης Χαρακτηριστικών
Features Vector	Διάνυσμα Χαρακτηριστικών
Feature Extraction	Εξαγωγή Χαρακτηριστικών
Feedforward Networks	Δίκτυα πρόσθιας τροφοδότησης
Finance Industry	Χρηματοοικονομική Βιομηχανία
Finite	Πεπερασμένος
Formal Grammar	Τυπική Γραμματική

Formal Language	Τυπική Γλώσσα
Frame	Καρέ
Fully Connected Neural Network	Πλήρως Συνδεδεμένο Νευρωνικό Δίκτυο
Function	Συνάρτηση
Functional Grammars	Λειτουργικές Γραμματικές
G	
Generalized Phrase Structure Grammars	Γραμματικές Γενικευμένης Φρασεοδομής
Government Organization	Κυβερνητικές Οργανώσεις
Gradient-Boosting Trees	Δέντρα που ενισχύουν τη διαβάθμιση
Grammar	Γραμματική
Graph	Γράφημα
H	
Head Driven Grammars	Γραμματικές Οδηγούμενες από την Κεφαλή
Hidden	Κρυφό
Hidden Layer	Κρυφό Επίπεδο
Hierarchical Clustering	Ιεραρχική Ομαδοποίηση
Hierarchy	Ιεραρχία
Hopfield Networks	Δίκτυα Hopfield
Hyperbolic Tangent	Υπερβολική Εφαπτομένη
Hyperparameter	Υπερπαράμετρος
I	
Inferring	Συμπεράσματα
Initial Trees	Αρχικά Δέντρα
Input	Είσοδος
Input Layer	Επίπεδο Εισόδου
Input Volume	Όγκος Εισόδου
Install	Εγκατάσταση
Interjection	Επιφώνημα
K	
K-means Clustering	Κ-Ομαδοποίηση
Kernel	Πυρήνας
Kohonen Networks	Δίκτυα Kohonen

L

Language	Γλώσσα
Layer	Επίπεδο
Learning Environment	Περιβάλλον Εκμάθησης
Lexical Functional Grammars	Λεξικές Λειτουργικές Γραμματικές
Library	Βιβλιοθήκη
Linear Model	Γραμμικό Μοντέλο
Linear Regression	Γραμμική Παλινδρόμηση
Linear Support Vector Machines	Μηχανές Γραμμικής Διανυσματικής Υποστήριξης
Linguistics	Γλωσσολογία
Locally Connected Networks	Τοπικά Συνδεδεμένα Δίκτυα
Logistic Regression	Λογιστική Παλινδρόμηση
Long Short Term Memory	Αμφίδρομη Μακρά Βραχυπρόθεσμη Μνήμη

M

Machine	Μηχανή
Machine Learning	Εκμάθηση Μηχανής
Machine Translation	Μηχανική Μετάφραση
Marketing	Εμπόριο
Memory	Μνήμη
Method	Μέθοδος
Model	Μοντέλο
Modified	Τροποποιημένο

N

Natural Language Processing	Επεξεργασία Φυσικής Γλώσσας
Networks	Δίκτυο
Neural Networks	Νευρωνικά Δίκτυα
Neuron	Νευρώνας
Node	Κόμβος
Noun	Ουσιαστικό

O

Object	Αντικείμενο
Output	Έξοδος
Output Layer	Επίπεδο Εξόδου

Output Volume

Όγκος Εξόδου

P

Part of speech

Μέρη του Λόγου

Pattern

Μοτίβο

Pattern Recognition

Αναγνώριση Προτύπων

Pattern Recognition

Ανακάλυψη Προτύπων

Perceptron

Αισθητήρας

Pixels

Εικονοστοιχεία

Playlist

Λίστα Αναπαραγωγής

Pooling

Υποδειγματοληψία

Preposition

Πρόθεση

Processing

Επεξεργασία

Processor

Επεξεργαστής

Produce

Παράγω

Production Rules

Κανόνες Παραγωγής

Pronoun

Αντωνυμία

R

Random Forest

Δάσος με Τυχαιότητα

Recall

Ανάκληση

Receptive Fields

Δεκτικά Πεδία

Recommended System

Προτεινόμενο Σύστημα

Reconstruction

Ανασυγκρότηση

Rectified Linear Unit

Συνάρτηση Ράμπας

Recurrent Neural Networks

Αναδρομικά Νευρωνικά Δίκτυα

Recursive Neural Networks

Επαναλαμβανόμενα Νευρωνικά Δίκτυα

Regression

Παλινδρόμηση

Regular Languages

Κανονικές Γλώσσες

Representation

Αναπαράσταση

Root

Ρίζας

Rule

Κανόνας

S

Sentence

Πρόταση/Κείμενο

Sequence

Ακολουθία

Set

Σύνολο

Shallow Neural Networks

Αβαθή Νευρωνικά Δίκτυα

Sigmoid Function	Σιγμοειδής Συνάρτηση
Similarity	Ομοιότητα
Speech Recognition	Αναγνώριση Ομιλίας
State	Κατάσταση
Stochastic Grammars	Στοχαστικές Γραμματικές
Stochastic Languages	Στοχαστικές Γλώσσες
String	Αλυσίδα Χαρακτήρων
Structure	Δομή
Subject	Υποκείμενο
Substitution	Αντικατάσταση
Supervised Learning	Εποπτευόμενη Εκμάθηση
Support Vector Machines	Μηχανές Διανυσμάτων Υποστήριξης
Symbol	Σύμβολο
System	Σύστημα
System Requirements	Απαιτήσεις Συστήματος
	T
Tagging	Ετικετοποίηση
Text Recognition	Αναγνώριση Κειμένου
Tool	Εργαλείο
Topology	Τοπολογία
Traditional Computing	Παραδοσιακός Υπολογισμός
Traditional Programming	Παραδοσιακός Προγραμματισμός
Traditional Programming	Συμβατός Προγραμματισμός
Training	Εκπαίδευση
Transformation/Modification	Μετασχηματισμός
Transformational Grammars	Μετασχηματιστικές Γραμματικές
Tree Adjoining Grammars	Γραμματικές Συνδεδεμένων Δέντρων
Threshold Function	Συνάρτηση Κατωφλίου
Typical	Τυπικό
	U
Understanding/Comprehension	Κατανόηση
Universality	Καθολικότητα
Unsupervised Learning	Μη - Εποπτευόμενη Εκμάθηση
	V
Verb	Ρήμα

	W
Weight	Βάρος
Word	Λέξη
Word Embedding	Αποδεκτός Συνδυασμός Λέξεων