



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Βαθιά μηχανική μάθηση για
κατηγοριοποίηση προτάσεων

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

ΚΟΚΚΙΝΟΥ ΦΙΛΙΠΠΟΥ

Επιβλέπων: Αλέξανδρος Ποταμιάνος
Αναπληρωτής Καθηγητής

Αθήνα, Ιούνιος 2017



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Τομέας Σημάτων, Ελέγχου και Ρομποτικής

Βαθιά μηχανική μάθηση για κατηγοριοποίηση προτάσεων

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

του

ΚΟΚΚΙΝΟΥ ΦΙΛΙΠΠΟΥ

Επιβλέπων: Αλέξανδρος Ποταμιάνος

Αναπληρωτής Καθηγητής

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 15η Ιουνίου 2017.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....

Αλέξανδρος Ποταμιάνος

Αναπληρωτής Καθηγητής

.....

Γεώργιος Στάμου

Επίκουρος Καθηγητής

.....

Γιων Ανδρουτσόπουλος

Αναπληρωτής Καθηγητής

Αθήνα, Ιούνιος 2017



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Τομέας Σημάτων, Ελέγχου και Ρομποτικής

Copyright ©–All rights reserved Φίλιππος Κόκκινος, 2017.

Με την επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα.

Το περιεχόμενο αυτής της εργασίας δεν απηχεί απαραίτητα τις απόψεις του Τμήματος, του Επιβλέποντα, ή της επιτροπής που την ενέκρινε.

Υπεύθυνη Δήλωση

Βεβαιώνω ότι είμαι συγγραφέας αυτής της πτυχιακής εργασίας, και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην πτυχιακή εργασία. Επίσης έχω αναφέρει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επίσης, βεβαιώνω ότι αυτή η πτυχιακή εργασία προετοιμάστηκε από εμένα προσωπικά ειδικά για τις απαιτήσεις του προγράμματος σπουδών του Τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου.

(Υπογραφή)

.....

Φίλιππος Κόκκινος

Περίληψη

Καθώς το πλήθος των αδόμητων δεδομένων στο διαδίκτυο μεγαλώνει, γίνεται πλέον επιτακτική ανάγκη η εύρεση τρόπων για συστηματική επεξεργασία τους και η εξαγωγή πολύτιμης γνώσης από αυτά. Το αντικείμενο της διπλωματικής εργασίας είναι η ανάπτυξη μοντέλων που μπορούν να εξαγάγουν αναπαραστάσεις από τη φυσική γλώσσα και συγκεκριμένα από τη συντακτική δομή και τη σημασία με σκοπό την επίλυση άλλων υψηλότερων προβλημάτων όπως είναι η ανάλυση συναισθήματος. Ο σκοπός είναι η ανάπτυξη μια τεχνολογία για επεξεργασία φυσικής γλώσσας που μπορεί να δημιουργήσει αναπαραστάσεις φράσεις και προτάσεις. Επιπλέον, για τη δημιουργία των αναπαραστάσεων το αρχικό κομμάτι κειμένου δεν υπόκειται καμία επεξεργασία και δεν γίνεται κάποια υπόθεση απλοποίησης όπως για παράδειγμα αναζαρησία των λέξων μεταξύ τους, σύνητες τεχνική για αλγορίθμους επεξεργασίας λέξεων.

Συγκεκριμένα, τα αναδρομικά νευρωνικά δίκτυα που αναπτύχθηκαν είναι μια κατηγορία νευρωνικών δικτύων που εφαρμόζονται σε δεδομένα με ιεραρχική δομή, πχ δυαδικά δέντρα. Η προσθήκη πληροφορίας για δομή έχει αποτελέσει καθοριστικό στοιχείο για την καλή απόδοση αυτών των δικτύων ειδικά όταν αυτά εφαρμόζονται πάνω σε προβλήματα φυσικής γλώσσας όπως είναι η ανάλυση συναισθήματος, η συντακτική ανάλυση και η μοντελοποίηση γλώσσας. Η δύναμη τους βρίσκεται στο γεγονός ότι μπορούν να δημιουργήσουν αναπαραστάσεις προτάσεων μέσω της σύνθεσης λεκτικών αναπαραστάσεων, επιτυγχάνοντας με αυτό τον τρόπο τη δημιουργία ενός πολυδιάστατου χώρου που μπορεί να ταξινομηθεί.

Στα κεφάλαια παρουσιάζεται η θεωρία και αναπτύσσουμε αρχιτεκτονικές αναδρομικών δικτύων με μνήμες που υπολογίζουν σε μονή (TreeGRU) και διπλή κατεύθυνση (TreeBiGRU) και η εφαρμογή τους πάνω στο πρόβλημα της ανάλυσης συναισθήμα-

τος. Επομένω, χρησιμοποιώντας τη συντακτική δομή μιας πρότασης, δημιουργούνται αναπαραστάσεις για τη πρόταση και τις επιμέρους υπο-προτάσεις οι οποίες μετά ταξινομούνται με βάση το συναίσθημα. Στη συνέχεια, προτάσσεται ο μηχανισμός προσοχής σε δομή (structural attention mechanism) και εφαρμόζεται πάνω στα δίκτυα που αναπτύχθηκαν. Ο μηχανισμός προσοχής πάνω σε δομή λειτουργεί στο πλαίσιο ότι οι αναπαραστάσεις υπο-προτάσεων δεν έχουν την ίδια βαρύτητα για τη σωστή αναπαράσταση μιας πρότασης. Τέλος, γίνεται ανάλυση διαφόρων παραλλαγών των δικτύων που αναπτύχθηκαν στη διπλωματική καθώς και δικτύων που παρουσιάζονται στη βιβλιογραφία.

Λέξεις Κλειδιά

Νευρωνικό δίκτυο, Αναδρομικό νευρωνικό δίκτυο, Μηχανισμοί Προσοχής, Ανάλυση Συναισθήματος

Abstract

As the number of unstructured data on the web grows, it becomes imperative to find ways to systematically process them and extract valuable knowledge from them. The subject of this diploma thesis is the development of models that can extract representations from the natural language and in particular from the syntactic structure and the meaning in order to solve other problems such as sentiment analysis. The purpose is to develop a technology for natural language processing that can create representations of phrases and sentences. In addition, for the induction of representations, the original piece of text is not subject to any processing and no simplification assumption is made, such as for instance the independence of words, a common technique for natural language processing algorithms.

In summary, the recursive neural networks developed are a class of neural networks that are applied to hierarchical structure data, e.g., binary trees. Adding structural information has been a key element in the good performance of these networks, especially when they are applied to natural language problems such as sentiment analysis, syntactic analysis and language modeling. Their power lies in the fact that they can create representations of sentences through the synthesis of word representations, thus achieving the creation of a multidimensional space that can be classified.

In the chapters, we present the theory and we are developing recursive networks with memories that work either in an unidirectional (en TreeGRU) and a bidirectional manner (en TreeBiGRU) and apply them to the problem of sentiment analysis. Therefore, using the syntactic structure of a sentence, representations are created for the sentence and sub-phrases that are then categorized based on senti-

ment. Furthermore, structural attention mechanism is introduced and implemented on the networks that are being developed. The mechanism of attention on a structure works in the context that the sub-sentence representations do not have the same weight for the proper representation of a sentence. Finally, we analyze various variants of the networks developed in the diploma thesis as well as the networks presented in the bibliography.

Keywords

Neural Networks, Recursive Neural Network, Attention Model, Sentiment analysis

στους γονείς μου

Ευχαριστίες

Θα ήθελα καταρχήν να ευχαριστήσω τον καθηγητή κ. Αλέξανδρο Ποταμιάνο για την επίβλεψη αυτής της διπλωματικής εργασίας και για την ευκαιρία που μου έδωσε να την εκπονήσω στο εργαστήριο Επεξεργασία Φωνής και Φυσικής Γλώσσας. Τέλος θα ήθελα να ευχαριστήσω τους γονείς μου για την καθοδήγηση και την ηθική συμπαράσταση που μου προσέφεραν όλα αυτά τα χρόνια.

Περιεχόμενα

Περίληψη	i
Abstract	iii
Ευχαριστίες	vii
Περιεχόμενα	x
Κατάλογος Σχημάτων	xiii
Κατάλογος Πινάκων	xv
1 Εισαγωγή	1
1.1 Οργάνωση του τόμου	3
2 Θεωρητικό υπόβαθρο Νευρωνικών Δικτύων	5
2.1 Νευρωνικά Δίκτυα	5
2.1.1 Ιστορική Αναδρομή	5
2.1.2 Το μοντέλο του τεχνητού νευρώνα	8
2.2 Αρχιτεκτονικές Νευρωνικών Δικτύων	12
2.3 Μάθηση και Εκπαίδευση	17
2.3.1 Αλγόριθμος Πίσω Διάδοσης Σφάλματος με Κάθοδο Κλίσης (Gradient Descent)	19
2.3.2 Αλγοριθμικές Τροποποιήσεις και επεκτάσεις της Καθόδου Κλίσης	
21	
2.4 Εφαρμογές Νευρωνικών Δικτύων	29

2.4.1	Πλεονεκτήματα Νευρωνικών Δικτύων	31
3	Θεωρητικό υπόβαθρο Ανάλυσης Συναισθήματος	33
3.1	Ανάλυση Συναισθήματος	33
3.1.1	Κατηγοριοποίηση Συναισθήματος	34
3.2	Τεχνικές Ανάλυσης Συναισθήματος	35
3.2.1	Τεχνικές Μηχανικής Μάθησης	35
3.2.2	Τεχνικές με λεξικά	42
3.3	Εφαρμογές Ανάλυσης Συναισθήματος	42
4	Αναδρομικά Νευρωνικά Δίκτυα για Επεξεργασία Φυσικής Γλώσσας	45
4.1	Αρχιτεκτονική Αναδρομικών Νευρωνικών Δικτύων	45
4.1.1	Αρχιτεκτονική Αναδρομικών Νευρωνικών Δικτύων με Μνήμη	51
4.2	Εκπαίδευση Αναδρομικών Δικτύων	56
4.3	Χρήση Αναδρομικών Δικτύων στην Επεξεργασία Φυσικής Γλώσσας	56
4.3.1	Αλγόριθμοι Glove και Word2vec	57
5	Σχεδίαση Αναδρομικού Νευρωνικού Δικτύου	65
5.1	Σχεδίαση Αναδρομικού Δικτύου	65
5.1.1	Αναδρομικό Νευρωνικό δίκτυο με GRU	65
5.1.2	Αμφίδρομο Αναδρομικό Νευρωνικό δίκτυο με GRU	66
5.1.3	Μηχανισμός προσοχής εφαρμοσμένος σε δομή	70
6	Ανάλυση και Πειράματα	73
6.1	Δεδομένα Stanford Sentiment Treebank	73
6.2	Υλοποίηση των Αναδρομικών Δικτύων	75
6.3	Αποτελέσματα	77
7	Επίλογος	85
7.1	Συμπεράσματα	85
7.2	Μελλοντικές Επεκτάσεις	86

Κατάλογος Σχημάτων

2.1	Το μη-γραμμικό μοντέλο ενός τεχνητού νευρώνα	9
2.2	Συναρτήσεις Ενεργοποίησης	10
2.3	Τεχνητό νευρωνικό δίκτυο εμπρόσθιας τροφοδότησης, αρχιτεκτονικής 3-4-2	14
2.4	Τεχνητό νευρωνικό δίκτυο εμπρόσθιας τροφοδότησης, αρχιτεκτονικής με 2 κρυφά επίπεδα	15
2.5	Απλό αναδρομικό δίκτυο εφαρμοσμένο σε δενδρική δομή	16
2.6	Επαναλαμβανόμενο δίκτυο εφαρμοσμένο σε τρεις διαδοχικές 'χρονικές' στιγμές	16
2.7	Τεχνητό συνελικτικό δίκτυο	17
2.8	Σύγκλιση αλγορίθμου καθόδου κλίσης χωρίς ορμή	23
2.9	Σύγκλιση αλγορίθμου καθόδου κλίσης με ορμή	24
3.1	Γραμμικές σχέσεις μεταξύ παραθετικών διαφορών λέξεων της αγγλικής γλώσσας	36
3.2	Γραμμικές σχέσεις μεταξύ ανδρών και γυναικών	37
3.3	Σχηματική αναπαράσταση του δικτύου MV-RNN για το σχηματισμό ενός διανύσματος στο επίπεδο πρότασης	39
3.4	Κανονικοποιημένο ιστόγραμμα συναισθηματικών ετικετών, για κάθε n-gram	40
3.5	Τροχός συναισθημάτων – Πηγή: (Gill, et al., 2008)	41
4.1	Παράδειγμα συντακτικού δέντρου που δημιουργείται από την ανάλυση της πρότασης A cat eats a mouse	46

4.2	Παράδειγμα του πιο απλού αναδρομικού δικτύου που εφαρμόζεται πάνω σε δυο κόμβους C1, C2 για τη δημιουργία της αναπαράστασης του πατέρα αυτών p1,2	47
4.3	Παράδειγμα του MV-RNN αναδρομικού δικτύου που εφαρμόζεται πάνω σε δυο κόμβους a, b, c για τη δημιουργία της αναπαράστασης του πατέρα αυτών p1 και στη συνέχεια p2	49
4.4	Ένα επίπεδο του Recursive Neural Tensor Network που εφαρμόζεται πάνω στη λέξεις - κόμβους b,c	50
4.5	Σχηματική απεικόνιση της δομής μια μονάδας Μακριά Βραχυπρόθεσμης Μνήμης - Long Short-Term Memory	53
4.6	Σχηματική απεικόνιση της δομής μια μονάδας Φραγμένης Επαναλαμβανόμενης Μονάδας - Gated Recurrent Unit	54
4.7	Κλασικό μοντέλο για μοντελοποίησης γλώσσας	59
4.8	Το μοντέλο C & W	61
4.9	Continuous bag-of-words	62
4.10	Skip-Gram	64
5.1	Σχηματική απεικόνιση της καθοδικής φάσης του αναδρομικού νευρωνικού δικτύου TreeBiGRU εφαρμοσμένο πάνω στο συντακτικό δέντρο της πρόταση "Best indie of the year, so far."	68
5.2	Σχηματική απεικόνιση της ανοδικής φάσης του αναδρομικού νευρωνικού δικτύου TreeBiGRU εφαρμοσμένο πάνω στο συντακτικό δέντρο της πρόταση "Best indie of the year, so far."	70
6.1	Σχηματική απεικόνιση ενός παραδείγματος από το Stanford Sentiment Treebank. Στην εικόνα παρουσιάζεται η συντακτική δομή και οι αντίστοιχες ετικέτες συναισθήματος για την πρόταση "Yet the act is still charming here."	74
6.2	Υλοποίηση του βασικού κώδικα για τον υπολογισμό της εμπρόσθιας φάσης για το δίκτυο TreeGRU	76
6.3	Σχηματική απεικόνιση της απόδοσης των συστημάτων με βάση το βάθος του δέντρου	81

- 6.4 Σχηματική απεικόνιση της απόδοσης των συστημάτων με βάση το μέγεθος της πρότασης 81
- 6.5 Σχηματική απεικόνιση της σύγκλισης και απόδοσης του απλού αναδρομικού δικτύου με χρήση διάφορων αλγορίθμων εκπαίδευσης. Με μπλε και κόκκινη γραμμή απεικονίζεται η ακρίβεια σε επίπεδο δέντρου και πρότασης σε σχέση με τα δεδομένα εκπαίδευσης, όμοια με πράσινη και κόκκινη γραμμή για τα δεδομένα (test). Με κίτρινη γραμμή συμβολίζεται το λάθος του δικτύου. 83

Κατάλογος Πινάκων

6.1	Ακρίβεια που επιτυγχάνεται Stanford Sentiment Treebank δατασετ. RNN, MV-RNN ανδ RNTN [44]. PVec: [31]. TreeLSTM [47]. DRNN [17]. DCNN [18].CNN-multichannel [19]. DMN [23]	77
6.2	Πίνακας με τα πιο θετικά και αρνητικά n-grams, όπως αυτά προβλέπονται από το σύστημα TreeBiGRU	79
6.3	Αναφορά Ταξινόμησης για το πρόβλημα 5 κλάσεων στο επίπεδο πρότασης του μοντέλου TreeBiGru με μηχανισμό προσοχής	80
6.4	Αναφορά Ταξινόμησης για το πρόβλημα 5 κλάσεων στο σύνολο των κόμβων του μοντέλου TreeBiGru με μηχανισμό προσοχής	80

Κεφάλαιο 1

Εισαγωγή

Καθημερινά στο διαδίκτυο προστίθονται μεγάλοι όγκοι αδόμητων κειμένων και δεδομένων από ανθρώπους σε όλη την υφήλιο. Η τάση αυτή κάνει επιτακτική την εύρεση ευφυών διαδικασιών που εξαγάγουν γνώση μέσα από αυτά. Ο σκοπός αυτής της διπλωματικής είναι η ανάπτυξη μοντέλων που αυτόματα εξαγάγουν αναπαραστάσεις της ανθρώπινης γλώσσας και του νοήματος με σκοπό την επίλυση άλλων υψηλότερων προβλημάτων, όπως η αναγνώριση συναισθήματος.

Έχει υπάρξει μεγάλη πρόοδος στις τεχνολογίες επεξεργασίας φυσικής γλώσσας όπως την εξαγωγή γνώσης μέσα από αδόμητα δεδομένα του διαδικτύου και αναγνώριση συναισθήματος στα κοινωνικά δίκτυα. Ένα σκοπός της Επεξεργασίας Φυσικής Γλώσσας είναι η ανάπτυξη γενικών αλγορίθμων που επιλύουν τέτοια προβλήματα και μαθαίνουν τις απαραίτητες ενδιάμεσες αναπαραστάσεις για γλωσσολογικά φαινόμενα. Παρόλα αυτά, οι συνήθεις τεχνικές για αυτό το σκοπό έχουν δύο βασικά προβλήματα.

- Απλές γλωσσολογικές υποθέσεις

Στη επεξεργασία φυσικής γλώσσας και στην αναγνώριση προτύπων, συχνά αναπτύσσονται αλγόριθμοι και μετά τροποποιούνται τα δεδομένα με σκοπό να είναι συμβατά με τον αλγόριθμο που αναπτύχθηκε. Για παράδειγμα, ένα πρώτο βήμα στην ταξινόμηση κειμένου είναι η δημιουργία αναπαραστάσεων από λέξεις των οποίων η σειρά εμφάνισης δεν αξιοποιείται (bag of words). Το γεγονός αυτό δημιουργεί προφανή προβλήματα όταν ο αλγόριθμος προσπαθεί να κατανοήσει την πρόταση. Μια άλλη απλοποίηση, είναι η μελέτη των λέξεων με βάση ένα σταθερό

πλήθος από γειτονικές λέξεις.

- Αναπαράσταση Χαρακτηριστικών

Ενώ μεγάλος χρόνος ξοδεύεται πάνω στην δημιουργία και ανάπτυξη μοντέλων, είναι κοινό μυστικό ότι η απόδοση των περισσότερων γλωσσικών μοντέλων σχετίζεται με τη συλλογή χαρακτηριστικών που δίνονται ως είσοδο στο σύστημα. Για παράδειγμα, σύγχρονα συστήματα που περιγράφονται στη βιβλιογραφία χρησιμοποιούν πέρα από λεκτικά ιστογράμματα, συντακτική πληροφορία, οντότητες και μεγάλες συλλογές από λεξικά που έχουν δημιουργηθεί από ειδικούς γλωσσολόγους. Κάθε ένα από αυτά τα χαρακτηριστικά χρειάστηκε αρκετό χρόνο για να αναπτυχθεί και να χρησιμοποιηθεί τελικά στην ανάπτυξη αλγορίθμων.

Τα μοντέλα που προτείνονται στην διπλωματική αντιμετωπίζουν αυτά τα δύο προβλήματα. Δημιουργούν αναπαραστάσεις για προτάσεις με ικανότητα γενίκευσης χρησιμοποιώντας τη σειρά των λέξεων και χωρίς χρήση πρότερης γνώσης. Περαιτέρω, παρουσιάζουν state of the art απόδοση χωρίς επιπλέον χαρακτηριστικά. Η έμπνευση για τη δημιουργία των μοντέλων αποτέλεσε η δουλειά του Richard Socher που πρώτος χρησιμοποίησε με επιτυχία τα αναδρομικά νευρωνικά δίκτυα για την επίλυση προβλημάτων επεξεργασίας φυσικής γλώσσας. Τα δίκτυα αυτά αποτελούν μέρος ενός ευρύτερου πεδίου που ονομάζεται Βαθιά Μηχανική Μάθηση (deep learning). Η φύση των δικτύων της διπλωματικής εργασίας είναι αναδρομική, καθώς συνδυάζουν τον αναδρομικό υπολογισμό αναπαραστάσεων πάνω σε γραμματικές γλωσσικές δομές.

Τα αναδρομικά νευρωνικά δίκτυα έχουν την ικανότητα να μαθαίνουν τους τρόπους με του οποίου οι λέξεις ενώνονται για να σχηματίσουν το νόημα μεγαλύτερων προτάσεων. Πέρα από την επίλυση του κλασσικού προβλήματος με το μήκος της πρότασης και το πλήθος των γειτονικών λέξεων που λαμβάνει κανείς υπόψιν, τα αναδρομικά δίκτυα μαθαίνουν πως να συνθέτουν το νόημα μια πρότασης. Θεωρείται ότι είναι αναγκαία προϋπόθεση για την πλήρη μοντελοποίηση της γλώσσας. Περαιτέρω, τα μοντέλα αυτά μαθαίνουν συναρτήσεις σύνθεσης (compositional functions) από τα δεδομένα εκπαίδευσης χωρίς παρέμβαση από ανθρώπινο παράγοντα.

1.1 Οργάνωση του τόμου

Η εργασία αυτή είναι οργανωμένη σε επτά κεφάλαια: Στο Κεφάλαιο 2 δίνεται το θεωρητικό υπόβαθρο των βασικών τεχνολογιών που σχετίζονται με τα νευρωνικά δίκτυα. Αρχικά, γίνεται μια λεπτομερής ιστορική αναδρομή στην έρευνα που οδήγησε σε αυτό που σήμερα ονομάζεται νευρωνικό δίκτυο και βαθιά μηχανική μάθηση. Στη συνέχεια, περιγράφονται με μαθηματικό φορμαλισμό τα νευρωνικά δίκτυα και σχετική θεωρία βελτιστοποίησης.

Πέρα από τα νευρωνικά δίκτυα, στο Κεφάλαιο 3 δίνεται το θεωρητικό υπόβαθρο των βασικών τεχνολογιών που σχετίζονται με την ανάλυση συναισθήματος με μεθόδους μηχανικής μάθησης καθώς και μεθόδους λεξικών. Το κεφάλαιο 3 αποτελεί βασικό πυλώνα για να καταλάβει ο αναγνώστης ποια είναι η αναγκαιότητα της χρήσης νευρωνικών δικτύων στην επεξεργασία φυσικής γλώσσας και συγκεκριμένα στο πρόβλημα της ανάλυσης συναισθήματος.

Στο Κεφάλαιο 4 παρουσιάζεται αναλυτικά η θεωρία και ο μαθηματικός φορμαλισμός των αναδρομικών δικτύων καθώς και οι τεχνολογίες κατενεμημένων αναπαραστάσεων λέξεων, που καταλαμβάνουν βασικό ολόκληρο για την εκπόνηση αυτής της διπλωματικής. Μετα τα κεφάλαια 2 και 3, γίνεται λεπτομερής επεξήγηση των αναδρομικών νευρωνικών δικτύων που αποτελούν τη βάση της διπλωματικής. Οι δυο βασικοί πυλώνες, δηλαδή η καταναμημένη αναπαράσταση λέξεων και ο αναδρομικός γράφος υπολογισμού, επεξηγούνται σε βάθος για να γίνει εύκολα αντιληπτή η προσθήκη της διπλωματικής εργασίας στο πεδίο.

Η πρόταση των καινούργιων αναδρομικών δικτύων που αναπτύχθηκαν στο πλαίσιο της διπλωματικής δίνεται στο Κεφάλαιο 5. Τα δυο βασικά δίκτυα TreeGRU και Tree-BiGRU επεξηγούνται με λεπτομερές τρόπο και γίνεται ιδιαίτερη αναφορά στους λόγους που οδήγησαν στην ανάπτυξη τους. Επίσης, στο ίδιο κεφάλαιο γίνεται και παρουσίαση ενός νέου μηχανισμού προσοχής που σχεδιάστηκε συγκεκριμένα για να εφαρμόζεται πάνω σε γλωσσικές συντακτικές δομές, γνωστός και ως Structural Attention Mechanism.

Στο Κεφάλαιο 6 παρουσιάζονται τα πειράματα και ανάλυση αυτών πάνω στις συλλογές δεδομένων που χρησιμοποιήθηκαν για εκπαίδευση και έλεγχο. Αναλύονται με

εμπειρικό τρόπο τα πλεονεκτήματα και τα μειονεκτήματα της κάθε δομής ξεχωριστά και συγκρίνονται με ανάλογες υλοποιήσεις που μπορεί κανείς να βρει στη βιβλιογραφία.

Τέλος στο Κεφάλαιο 7 δίνεται περιληπτικά η συνεισφορά αυτής της διπλωματικής εργασίας, καθώς και μελλοντικές επεκτάσεις.

Κεφάλαιο 2

Θεωρητικό υπόβαθρο Νευρωνικών Δικτύων

2.1 Νευρωνικά Δίκτυα

2.1.1 Ιστορική Αναδρομή

Η πρωτοποριακή δουλειά των McCulloch και Pitts [27] αποτέλεσε την έναρξη και άνθηση του τομέα των νευρωνικών δικτύων. Ο πρώτος ήταν ψυχίατρος και ο δεύτερος μαθηματικός και σύμφωνα με τον Rall (1990) η κλασική εργασία τους πραγματοποιήθηκε εντός του πανεπιστημίου του Σικάγου για περισσότερα από 5 χρόνια. Μέσα στην εργασία περιγράφεται ο λογικός λογισμός των νευρωνικών δικτύων που χρησιμοποιήθηκε από τον von Neumann για τον υπολογισμό των ιδεατών στοιχείων καθυστέρησης τα οποία οδήγησαν στην κατασκευή του EDVAC (Electronic Discrete Variable Automatic Computer) και στη συνέχεια του πρώτου γενικού σκοπού υπολογιστή γνωστού ως ENIAC.

Η επόμενη μεγάλη άνθηση στον τομέα των νευρωνικών δικτύων ήρθε το 1949 με την έκδοση του βιβλίου του Hebb με τίτλο "The Organization of Behavior" [15], μέσα στο οποίο έγινε για πρώτη φορά μια ιδιαίτερη δήλωση ενός φυσιολογικού κανόνα μάθησης για συνοπτικές τροποποιήσεις. Πιο λεπτομερώς ο Hebb πρότεινε τη θεωρία ότι οι συνδέσεις του εγκεφάλου δεν είναι σταθερές και συνεχώς τροποποιούνται καθώς ο οργανισμός μαθαίνει μέσα από διάφορες εργασίες. Ακολούθως, πρότεινε ότι η

αποτελεσματικότητα μιας σύναψης ανάμεσα σε δυο νευρώνες αυξάνεται από την επαναλαμβανόμενη ενεργοποίηση του ενός νευρώνα από τον άλλο κατά μήκος της σύναψης.

Η αναφορά των Rochester, Holland, Haibt και Duda [39] είναι ίσως η πρώτη προσπάθεια για χρήση υπολογιστή που χρησιμοποιεί την εξομοίωση για τον έλεγχο μια καλά σχηματισμένης και ολοκληρωμένης νευρωνικής θεωρίας βασισμένη στο αίτημα μάθηση του Hebb. Η εξομοίωση έδειξε ότι χρειαζόταν η προσθήκη παρεμπόδισης ώστε η θεωρία να δουλέψει. Τον ίδιο χρόνο ο Uttley [50] παρουσίασε την αποκαλούμενη "διαρρέουσα ολοκλήρωση" ή "νευρώνας φωτιάς" που αργότερα αναλύθηκε από τον Caianiello.

Το 1952 εκδόθηκε το βιβλίο του Ashby με τίτλο "Design for a brain: The Origin of Adaptive Behavior" [1], το οποίο ασχολήθηκε με την βασική έννοια ότι η προσαρμοζόμενη συμπεριφορά δεν είναι έμφυτη στον άνθρωπο αλλά μαθαίνεται κατά τη διάρκεια της ζωής του. Το 1954 ο Minsky έγραψε τη διδακτορική του διατριβή με τίτλο "Theory of Neural-Analog Reinforcement Systems and Its Application to the Brain-Model Problem", στη συνέχεια το 1961 έγραψε την εργασία με τίτλο "Steps Toward Artificial Intelligence" [32]. Εν παραλλήλω, το 1954 ο Gabor (πρωτοπόρος της θεωρίας επικοινωνιών και εφευρέτης του ολογραφήματος) εισήγαγε την έννοια των μη γραμμικών προσαρμοζόμενων φίλτρων την οποία υλοποίησε ως μια μηχανή που έπρεπε να αναπαράγει μια συνάρτηση-στόχο δοθέντος μιας τροφοδοσίας δειγμάτων στοχαστικών διαδικασιών.

Ένα θέμα ιδιαίτερου ενδιαφέροντος για τα νευρωνικά δίκτυα είναι αυτό της σχεδίασης ενός αξιόπιστου δικτύου με νευρώνες που μπορούν να θεωρηθούν σαν μη αξιόπιστα στοιχεία. Αυτό το σημαντικό πρόβλημα λύθηκε από τον von Neumann (1956), χρησιμοποιώντας την έννοια του πλεονασμού. Δεκαπέντε ολόκληρα χρόνια μετά την έκδοση της εργασίας των McCulloch και Pitts, έγινε από τον Risenblatt [40] μια νέα προσέγγιση πάνω στο πρόβλημα αναγνώρισης με αισθητήρια. Το ιδιαίτερο επίτευγμα του ήταν το θεώρημα σύγκλισης αισθητηρίου (perceptron convergence theorem). Το 1960 οι Widrow και Hoff [51] πρότειναν τον αλγόριθμο ελάχιστου μέσου τετραγώνου (least mean square - LMS) και τον χρησιμοποίησαν για να σχηματίσουν την μονάδα Adaline (adaptive linear element). Η διαφορά ανάμεσα στο αισθητήριο και στο Adaline βρίσκεται στον τρόπο μάθησης.

Κατά την διάρκεια της κλασικής περιόδου του perceptron (1960) υπήρχε η πεποίθηση ότι τα νευρωνικά δίκτυα μπορούσαν να λύσουν οποιοδήποτε πρόβλημα. Όμως, στο βιβλίο που εκδόθηκε από τους Minsky και Papert [33] αποδείχτηκε ότι υπάρχουν όρια στο τι μπορεί να υπολογιστεί και να λυθεί με χρήση αισθητηρίων. Ένα σημαντικό πρόβλημα πάνω στη σχεδίαση ενός πολυεπίπεδου αισθητηρίου είναι το πρόβλημα της ανάθεσης εμπιστοσύνης (credit assignment problem), το οποίο βρήκε την λύση του μόλις την δεκαετία του 1980.

Κατά την διάρκεια του 1970, το ερευνητικό ενδιαφέρον για τα νευρωνικά δίκτυα ελαττώθηκε, όμως παράλληλα σημειώθηκε μεγάλη πρόοδος στον τομέα των χαρτών αυτοοργάνωσης με ανταγωνιστική μάθηση. Το 1980 έγιναν πολλές εργασίες πάνω στην θεωρία αλλά και στον σχεδιασμό των νευρωνικών δικτύων. Ο Grossberg [13] ανέπτυξε μια καινούργια αρχή αυτοοργάνωσης που συνδυάζει φιλτράρισμα από 'κάτω προς τα πάνω' και αντίθετη αύξηση σε μικρή μνήμη με από 'πάνω προς τα κάτω' ταίριασμα προτύπων. Δεδομένης μια τέτοιας ικανότητας και αν το πρότυπο εισόδου ταιριάζει με την ανάδραση μάθησης, λαμβάνει χώρα μια δυναμική κατάσταση που καλείται adaptive resonance. Αυτό το φαινόμενο δίνει την βάση για μια νέα κατηγορία νευρωνικών δικτύων γνωστά σαν adaptive resonance theory (ART) [7]. το 1982 ο Hopfield χρησιμοποίησε την ιδέα μιας συνάρτησης ενέργειας για να φτιάξει ένα νέο τρόπο κατανόησης του υπολογισμού που γίνεται από τα δίκτυα με συμμετρικές συνοπτικές συνδέσεις. Επιπλέον καθιέρωσε τον ισομορφισμό ανάμεσα σε τέτοια περιοδικά δίκτυα και σε ένα Ising [6] μοντέλο που χρησιμοποιείται στη στατιστική. Αυτή η αναλογία άνοιξε τον δρόμο για ένα κατακλυσμό θεωριών για τα νευρωνικά δίκτυα. Αυτή η συγκεκριμένη τάξη νευρωνικών δικτύων με ανατροφοδότηση έτυχε ιδιαίτερης προσοχής κατά τη δεκαετία του 80 και με το χρόνο έγιναν γνωστά ως δίκτυα Hopfield [16].

Το 1983 οι Cohen και Grossberg [8] έδωσαν μια νέα αρχή για σχεδίαση μιας διευθυνσιοδοτούμενης μνήμης (content-addressable memory) που περιλαμβάνει την έκδοση συνεχούς χρόνου του δικτύου Hopfield σαν μια ιδιαίτερη περίπτωση. Μια ακόμα σημαντική ανάπτυξη το 1982 ήταν η έκδοση της εργασίας του Kohonen [22] πάνω στους χάρτες αυτοοργάνωσης, χρησιμοποιώντας μιας ή δύο διαστάσεων δικτυωτές δομές. Το 1983 οι Kirkpatrick, Gallat και Vecchi [21] περιέγραψαν μια νέα διαδικασία που λέγεται εξομοιωμένη αν όπτηση για λύση προβλημάτων συνδυαστικής βελτιστοποίησης. Η

εξομοιωμένη αν όπτηση χρησιμοποιείται στη στατιστική θερμοδυναμική και βασίζεται σε μια απλή τεχνική. Την ίδια χρονιά εκδόθηκε μια εργασία από τους Burto, Sutton και Anderson [2] πάνω στην ενισχυμένη μάθηση, η οποία δημιούργησε μεγάλο ενδιαφέρον πάνω στην ενισχυμένη μάθηση και την εφαρμογή της. Το 1984 ο Braintenberg [4] εξέδωσε ένα βιβλίο με τίτλο "Vehicles: Experiments in Synthetic Psychology" το οποίο περιγράφει διάφορες μηχανές με απλή εσωτερική αρχιτεκτονική και το οποίο περιγράφει διάφορες μηχανές με απλή εσωτερική αρχιτεκτονική που ενσωματώνει μερικές σημαντικές αρχές των αυτοοργανώμενων χαρτών. Το 1986 η ανάπτυξη του αλγορίθμου για πίσω διάδοση (back-propagation) παρουσιάστηκε για πρώτη φορά από τον Rumelhart [41].

Αυτός ο αλγόριθμος έγινε πολύ δημοφιλής και έδωσε νέα ώθηση στις εφαρμογές των νευρωνικών δικτύων. Το 1988 ο Linsker [25] περιέγραψε μια νέα αρχή για την αυτοοργάνωση σε ένα δίκτυο αισθητηρίων (perceptrons). Η αρχή αυτή σχεδιάστηκε ώστε να διατηρεί μέγιστη πληροφορία σχετικά με τα πρότυπα ενεργειών, που υπόκεινται σε περιορισμούς όπως συνοπτικές συνδέσεις και δυναμικές περιοχές σύναψης. Ο Linsker χρησιμοποίησε αφηρημένες έννοιες πάνω στην θεωρία πληροφοριών ώστε να σχηματίσει την αρχή της διατήρησης μέγιστης πληροφορίας.

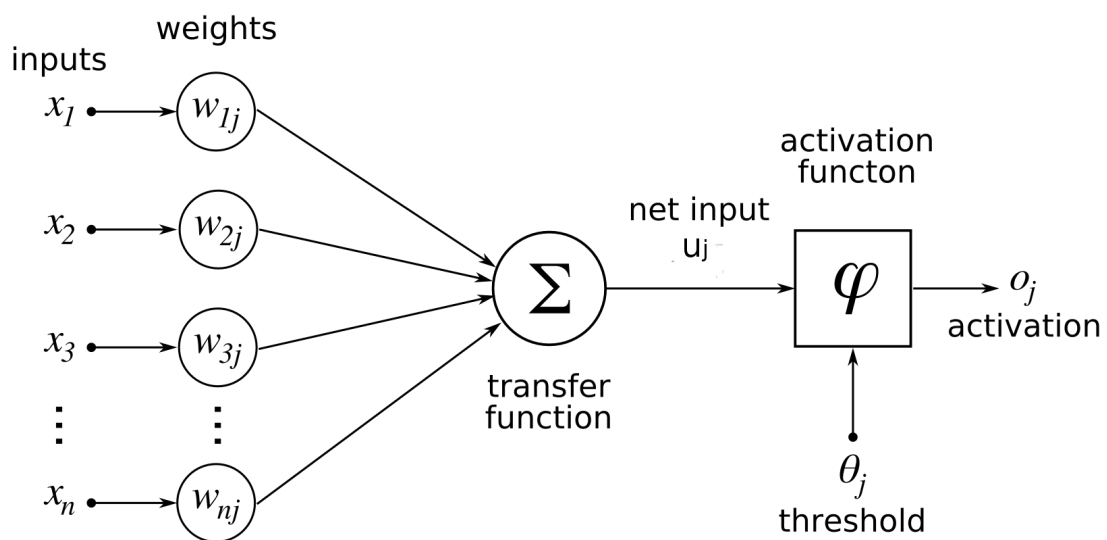
Αργότερα, το 1988 οι Broomhead και Lowe [5] περιέγραψαν μια διαδικασία για το σχεδιασμό των "προς τα εμπρός τροφοδότησης" (feedforward) δικτύων χρησιμοποιώντας συναρτήσεις ακτινικής βάσης (Radial Basis Functions - RBF) που είναι μια παραλλαγή των πολυεπίπεδων αισθητηρίων. Το 1989 εκδόθηκε το βιβλίο του Mead με τίτλο "VLSI and Neural Systems" [28]. Αυτό το βιβλίο δίνει μια ασυνήθιστη μίξη περιεχομένων από την νευροβιολογία και την τεχνολογία VLSI. Ίσως περισσότερο από κάθε άλλη έκδοση, η εργασία του Hopfield (1982) και το δίτομο βιβλίο των Rumelhart και McLelland (1986), να ήταν οι πιο σημαίνουσες εκδόσεις υπεύθυνες για την αναζωογόνηση του ενδιαφέροντος για τα νευρωνικά δίκτυα στην δεκαετία του 80.

2.1.2 Το μοντέλο του τεχνητού νευρώνα

Ένας νευρώνας είναι μια μονάδα επεξεργασίας πληροφορίας που είναι ο θεμελιώδης λίθος ενός τεχνητού ή μη νευρωνικού δικτύου. Το σχήμα 2.1 δείχνει το μοντέλο ενός

νευρώνα, του οποίου τα βασικά στοιχεία είναι:

1. Ένα σύνολο από συνάψεις ή συνδετικούς κρίκους, κάθε μια από τις οποίες χαρακτηρίζεται από ένα βάρος. Πιο συγκεκριμένα, ένα σήμα x_j στην είσοδο της τεχνητής συνάψης j που συνδέεται στον τεχνητό νευρώνα k , πολλαπλασιάζεται με το συναπτικό βάρος w_{kj} . Το βάρος w_{kj} είναι θετικό, αν η συνάψη είναι διεγερτική (δηλαδή ωθεί τον νευρώνα να ανταποκριθεί στη διέγερση που δέχεται) και αρνητικό, αν η συνάψη είναι απαγορευτική (δηλαδή αποτρέπει τον νευρώνα να ανταποκριθεί στη διέγερση).
2. Έναν αθροιστή για την πρόσθεση των σημάτων εισόδου, που πολλαπλασιάζονται με το βάρος της αντίστοιχης συνάψης.
3. Μια συνάρτηση ενεργοποίησης η οποία είναι γραμμική ή μη-γραμμική
4. Ένα εξωτερικά εφαρμοζόμενο κατώφλι θ_j που έχει επίδραση στην ελάττωση της εισόδου στην εφαρμοζόμενη συνάρτηση ενεργοποίησης. Το εν λόγω κατώφλι εμφανίζεται στη βιβλιογραφία και ως πόλωση (bias).



Σχήμα 2.1: Το μη-γραμμικό μοντέλο ενός τεχνητού νευρώνα

Σε μαθηματικούς όρους, ένας νευρώνας j περιγράφεται από τις εξισώσεις:

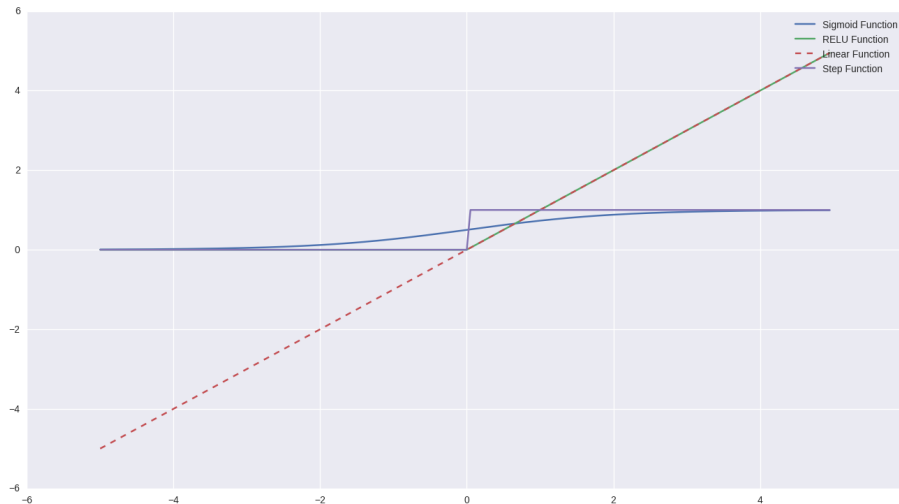
$$u_j = \sum_{i=1}^n w_{ij}x_i \quad (2.1)$$

$$o_j = \phi(u_j - \theta_j) \quad (2.2)$$

Το u_j είναι το άθροισμα του συνδυασμού βαρών και εισόδου του νευρώνα j , το θ_j είναι η πόλωση, $\phi(\cdot)$ είναι η συνάρτηση ενεργοποίησης και o_j είναι το σήμα εξόδου του νευρώνα. Η προσθήκη της πόλωσης θ_j έχει σαν αποτέλεσμα την εφαρμογή ενός εγγενούς (affine) μετασχηματισμού της εξόδου u_j όπως περιγράφεται από την παρακάτω:

$$u_j = W * x + \theta_j \quad (2.3)$$

Η συνάρτηση ενεργοποίησης $\phi(\cdot)$ ορίζει την έξοδο ενός νευρώνα συναρτήσει του επιπέδου ενεργοποίησης της εισόδου. Έχουμε 3 βασικούς τύπους συναρτήσεων ενεργοποίησης.



Σχήμα 2.2: Συναρτήσεις Ενεργοποίησης

1. Συνάρτηση Κατωφλίου

Για αυτό τον τύπο συνάρτησης ενεργοποίησης έχουμε:

$$\phi(u) = \begin{cases} 1, & \text{if } u \geq 1 \\ 0, & \text{otherwise} \end{cases} \quad (2.4)$$

Επομένως, η έξοδος ενός νευρώνα x , που έχει την παραπάνω συνάρτηση ενε-

γοποίησης, έχεις τη μορφή:

$$o_j = \begin{cases} 1, & \text{if } u_j \geq 1 \\ 0, & \text{otherwise} \end{cases} \quad (2.5)$$

όπου το u_j είναι το εσωτερικό επίπεδο ενεργοποίησης του νευρώνα και δίνεται από τη σχέση:

$$u_j = \sum_{k=0} n w_{kj} - \theta_j \quad (2.6)$$

Η παράγωγος είναι:

$$\phi'(u) = \begin{cases} ?, & \text{if } u_j \geq 1 \\ 0, & \text{otherwise} \end{cases} \quad (2.7)$$

Η παράγωγος δεν είναι γνωστή και για πρακτικούς λόγους ακολουθείται κάποιος μηχανισμός αριθμητικής σταθερότητας από το πεδίο του Αριθμητικής Ανάλυσης.

2. Γραμμική συνάρτηση

Για αυτό τον τύπο συνάρτησης ενεργοποίησης έχουμε:

$$\phi(u) = u \quad (2.8)$$

Επομένως, η έξοδος ενός νευρώνα x , που έχει την παραπάνω συνάρτηση ενεργοποίησης, έχεις τη μορφή:

$$o_j = u_j \quad (2.9)$$

όπου το u_j είναι το εσωτερικό επίπεδο ενεργοποίησης του νευρώνα και δίνεται από τη σχέση:

$$u_j = \sum_{k=0} n w_{kj} - \theta_j \quad (2.10)$$

Η παράγωγος είναι:

$$\phi'(u) = 1 \quad (2.11)$$

3. Σιγμοειδής συνάρτηση

Για αυτό τον τύπο συνάρτησης ενεργοποίησης έχουμε:

$$\phi(u) = \frac{1}{1 + e^{-\alpha u}} \quad (2.12)$$

Είναι η μια από τις πιο συνηθισμένες μορφές συνάρτησης ενεργοποίησης, που χρησιμοποιείται στην κατασκευή τεχνητών νευρωνικών δικτύων. Ορίζεται σαν αυστηρά αύξουσα συνάρτηση, η οποία παρουσιάζει εξομάλυνση και ασυμπτωτικές ιδιότητες. Μεταβάλλοντας την παράμετρο α παίρνουμε διαφορετικές συναρτήσεις, όπως φαίνεται στο σχήμα 2.2.

Η παράγωγος της σιγμοειδούς είναι:

$$\phi'(u) = \phi(u)(1 - \phi(u)) \quad (2.13)$$

4. Ανορθωμένη Γραμμική συνάρτηση

Για αυτό τον τύπο συνάρτησης ενεργοποίησης έχουμε:

$$\phi(u) = \begin{cases} u, & \text{if } u \geq 0 \\ 0, & \text{otherwise} \end{cases} = \max(0, u) \quad (2.14)$$

Αυτή η συνάρτηση ενεργοποίησης είναι γνωστή και ως συνάρτηση ράμπα και είναι ανάλογη της ανόρθωσης μισού-κύματος από τον τομέα των ηλεκτρονικών. Επίσης, είναι η πιο δημοφιλής συνάρτηση ενεργοποίησης στο πεδίο του deep learning ξεπερνώντας τη σιγμοειδή συνάρτηση που έχει για χρόνια την πρώτο καθεδρία στο χώρο.

Η παράγωγος της ανορθωμένης γραμμικής συνάρτησης είναι:

$$\phi'(u) = \begin{cases} 1, & \text{if } u \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (2.15)$$

Η συγκεκριμένη συνάρτηση ενεργοποίησης μπορεί να βρεθεί και σε παραμετροποιησιμη μορφή:

$$\phi(u) = \max(u, \alpha u) \quad (2.16)$$

με α να ανήκει στο εύρος μεταξύ 0 και 1.

2.2 Αρχιτεκτονικές Νευρωνικών Δικτύων

Ένα νευρωνικό δίκτυο είναι ένας κατευθυνόμενος γράφος, που αποτελείται από κόμβους με συναπτικές διασυνδέσεις και συνδέσεις ενεργοποίησης που χαρακτηρίζονται από τις ακόλουθες ιδιότητες:

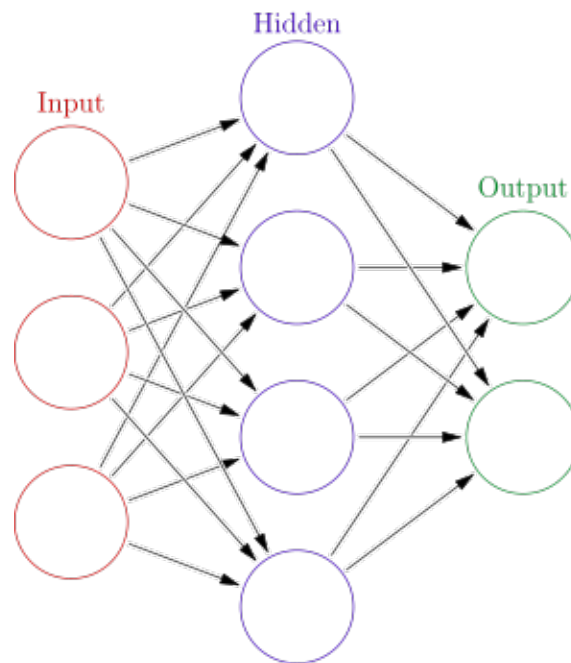
- Κάθε νευρώνας παριστάνεται από ένα σύνολο γραμμικών συναπτικών συνδέσεων, μια εξωτερικά εφαρμοζόμενη πόλωση και μια μη-γραμμική σύνδεση ενεργοποίησης. Η πόλωση παριστάνεται από συναπτικές συνδέσεις με σήμα εισόδου σταθερής τιμής -1 .
- Οι συναπτικές συνδέσεις ενός νευρώνα εφαρμόζουν κάποιο βάρος στα αντίστοιχα τμήματα εισόδου
- Το άθροισμα των βαρών των σημάτων εισόδου καθορίζει το συνολικό εσωτερικό επίπεδο ενεργοποίησης του νευρώνα
- Η σύνδεση ενεργοποίησης περιορίζει το εσωτερικό επίπεδο ενεργοποίησης, για την παραγωγή της εξόδου που παριστάνει την κατάσταση του νευρώνα.

Το τεχνητό νευρωνικό δίκτυο μπορεί να θεωρηθεί ως μια συλλογή από νευρώνες αλληλοσυνδεδεμένους μεταξύ τους. Τα πλέον ευρέως χρησιμοποιούμενα νευρωνικά δίκτυα είναι τα νευρωνικά δίκτυα πολλών στρωμάτων, τα οποία είναι νευρωνικά δίκτυα εμπρόσθιας τροφοδότησης (Artificial Neural Networks - ANN). Τα ANN αποτελούνται από ένα επίπεδο εισόδου και ένα επίπεδο εξόδου. Οι νευρώνες στο επίπεδο εισόδου είναι παθητικοί, δηλαδή ο καθένας απλά μεταδίδει μια τιμή μέσω των διασυνδεδετικών βαρών στους κρυφούς νευρώνες. Οι κρυφοί νευρώνες και οι νευρώνες εξόδου επεξεργάζονται τις εισόδους τους ως εξής: Κάθε κρυφός νευρώνας πολλαπλασιάζει κάθε είσοδο με το βάρος της, προσθέτει τα γινόμενα μαζί με την πόλωση και μετά εφαρμόζει μια συνάρτηση ενεργοποίησης για να παράγει το αποτέλεσμα. Στο σχήμα 2.3 φαίνεται ένα νευρωνικό με 3 νευρώνες στο επίπεδο εισόδου, 4 νευρώνες στο κρυφό επίπεδο και 2 νευρώνες στο επίπεδο εξόδου.

Το νευρωνικό αυτό έχει 12 συνδέσεις μεταξύ των νευρώνων στο επίπεδο εισόδου και των νευρώνων του κρυφού επιπέδου και 8 συνδέσεις μεταξύ των 4 νευρώνων του κρυφού επιπέδου και των 2 νευρώνων στο επίπεδο εξόδου. Συνολικά το δίκτυο έχει 20 συνδέσεις. Σε κάθε σύνδεση αντιστοιχεί μια τιμή βάρους w_{ij} που ενώνει το νευρώνα i με το νευρώνα j . Οι δύο νευρώνες βρίσκονται σε διαφορετικά επίπεδα καθώς το δίκτυο είναι εμπρόσθιας τροφοδότησης δηλαδή η έξοδος ενός επιπέδου είτε εισόδου είτε κρυφού γίνεται είσοδος για το επόμενο κατά σειρά επίπεδο. Οι πόλωσης του

τεχνητού νευρωνικού δικτύου στο σχήμα 2.3 είναι μηδενικές.

Ο τρόπος που δομείται ένα τεχνητό νευρωνικό δίκτυο είναι στενά συνδεδεμένος με τον αλγόριθμο εκμάθησης που χρησιμοποιείται για την εκπαίδευση του δικτύου.

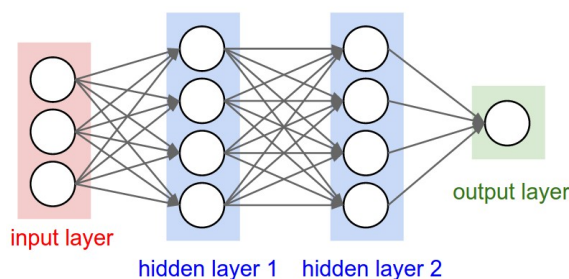


Σχήμα 2.3: Τεχνητό νευρωνικό δίκτυο εμπρόσθιας τροφοδότησης, αρχιτεκτονικής 3-4-2

Οι αρχιτεκτονικές νευρωνικών δικτύων δεν περιορίζονται στην εμπρόσθια τροφοδότηση ενός επιπέδου. Στη βιβλιογραφία μπορούμε να διακρίνουμε 3 διαφορετικές κλάσεις αρχιτεκτονικών δομών:

1. Πολλαπλών-επιπέδων εμπρόσθια τροφοδότησης δίκτυο

Σε ένα τέτοιο δίκτυο, οι νευρώνες είναι οργανωμένοι σε επίπεδα. Σε κάθε επίπεδο οι συνάψεις οδηγούν την πληροφορία προς τα μπροστά, δηλαδή η έξοδος ενός επιπέδου αποτελεί την είσοδο στο επόμενο επίπεδο. Το πλήθος των επιπέδων, πλην του επιπέδου εισόδου και εξόδου, μπορεί να είναι αρκετά μεγάλο με αντίστοιχη αύξηση στο πλήθος των παραμέτρων.



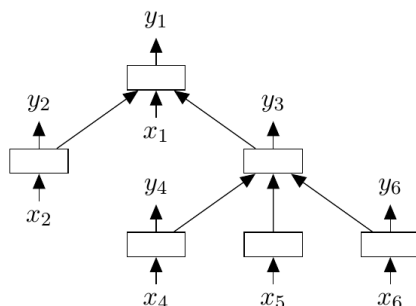
Σχήμα 2.4: Τεχνητό νευρωνικό δίκτυο εμπρόσθιας τροφοδότησης, αρχιτεκτονικής με 2 κρυφά επίπεδα

2. Αναδρομικό Δίκτυο

Η αρχιτεκτονική του Αναδρομικού δικτύου επιτρέπει την ύπαρξη ενός τουλάχιστον βρόχου ανάδρασης. Με αυτό το τρόπο, η έξοδος του νευρωνικού δικτύου γίνεται κομμάτι της εισόδου. Στο σχήμα 2.5 η είσοδος που συμβολίζεται ως x_i εφαρμόζεται στο αναδρομικό δίκτυο που συμβολίζεται ως παραλληλόγραμμο. Στα φύλλα της δενδρικής δομής μόνο τα διανύσματα x_i αποτελούν είσοδο στο αναδρομικό δίκτυο, καθώς όμως το ίδιο δίκτυο εφαρμόζεται σε κόμβους τότε κομμάτι της εισόδου αποτελεί η έξοδος από προηγούμενη φάση. Επομένως, η έξοδος σε προηγούμενη στιγμή του δικτύου γίνεται είσοδος σε μεταγενέστερη στιγμή, σχηματίζοντας έτσι ένα τουλάχιστον βρόχο ανάδρασης.

Τα αναδρομικά δίκτυα εφαρμόζονται σε προβλήματα όπου υπάρχει μια συγκεκριμένη τοπολογία στα δεδομένα, για παράδειγμα εφαρμόζονται στο πρόβλημα

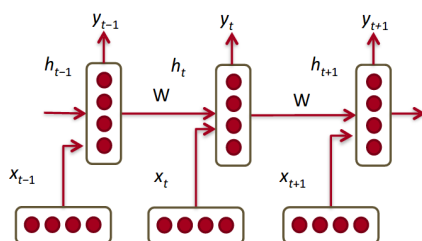
συντακτική ανάλυση μια πρότασης όταν αυτή αναπαρίσταται με δυαδικά δέντρα.



Σχήμα 2.5: Απλό αναδρομικό δίκτυο εφαρμοσμένο σε δενδρική δομή

3. Επαναλαμβανόμενο Δίκτυο

Η αρχιτεκτονική του Επαναλαμβανόμενου δικτύου [30] επιτρέπει την είσοδο ακολουθιών από διανύσματα ως είσοδο όπου ο νευρώνας δέχεται κάθε χρονική στιγμή ως είσοδο το ανάλογο διάνυσμα. Για μια ακολουθία εισόδου Q που περιέχει διανύσματα x_i , ο τεχνητός νευρώνας τη χρονική στιγμή i δέχεται ως είσοδο το διάνυσμα x_i καθώς και κρυφή πληροφορία h_i από την ακριβώς προηγούμενη χρονική στιγμή. Αντίστοιχα, ο τεχνητός νευρώνας εξαγάγει έξοδο y_i για την i χρονική στιγμή καθώς και κρυφή πληροφορία h_{i+1} για να δοθεί ως είσοδο στην $i + 1$ χρονική στιγμή. Κάθε σύνδεση του τεχνητού νευρώνα φέρει ένα πίνακα βαρών W με πραγματικές τιμές.

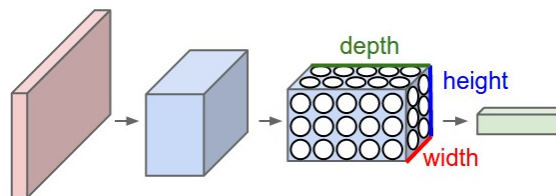


Σχήμα 2.6: Επαναλαμβανόμενο δίκτυο εφαρμοσμένο σε τρεις διαδοχικές 'χρονικές' στιγμές

4. Συνελικτικό Δίκτυο

Η αρχιτεκτονική του Συνελικτικού Δικτύου [24] είναι όμοια με την αρχιτεκτονική

ενός δικτύου Εμπρόσθιας Τροφοδότησης με διαφορετικό μοτίβο συνδέσεων το οποίο μιμείται τον οπτικό φλοιό των ζώων. Το Συνελικτικό Δίκτυο χρησιμοποιείται λόγω των ιδιοτήτων του στο πεδίο της Ανάλυσης Εικόνων και Βίντεο. Παρόλα αυτά, έχει χρησιμοποιηθεί με επιτυχία σε πολλά προβλήματα επεξεργασίας φυσικής γλώσσας όπως η ανάλυση συναισθήματος και μοντελοποίηση γλώσσας.



Σχήμα 2.7: Τεχνητό συνελικτικό δίκτυο

2.3 Μάθηση και Εκπαίδευση

Ένα νευρωνικό δίκτυο έχει δύο βασικές λειτουργίες την εκπαίδευση και την πρόβλεψη ή κατηγοριοποίηση. Οποιαδήποτε αρχιτεκτονική νευρωνικού δικτύου εκπαιδεύεται με τη βοήθεια παραδειγμάτων. Κάθε παράδειγμα αποτελείται από μια είσοδο, η οποία μοντελοποιείται με τη χρήση ενός διανύσματος ή πίνακα, και μια επιθυμητή έξοδο σε αυτή. Το σύνολο των παραδειγμάτων που δίνεται στο νευρωνικό δίκτυο ονομάζεται εκπαιδευτικό σύνολο. Τελικός σκοπός της εκπαίδευσης είναι ο υπολογισμός των βαρών του δικτύου που ελαχιστοποιούν το σφάλμα πρόβλεψης. Ένα απλοϊκός τρόπος υπολογισμού του σφάλματος είναι η διαφορά από την προβλεπόμενη τιμή εξόδου και της επιθυμητής τιμής όπως ορίζεται από το παράδειγμα εκπαίδευσης. Μόλις τελειώσει η εκπαίδευση του νευρωνικού δικτύου ελέγχεται η ικανότητα γενίκευσης του πάνω σε ένα σύνολο παραδειγμάτων που ονομάζεται σύνολο ελέγχου (test set). Αυτό το σύνολο ελέγχου αποτελείται από παραδείγματα που δεν δόθηκαν για εκπαίδευση στο νευρωνικό δίκτυο με σκοπό να αξιολογηθεί πόσο καλά μπορεί το δίκτυο να αναγνωρίσει καινούργια παραδείγματα τα οποία ήταν άγνωστα για αυτό κατά το στάδιο εκπαίδευσης.

Για τη βέλτιστη λειτουργία του νευρωνικού δικτύου, θα πρέπει κατά τα δύο στάδια της λειτουργίας του να ληφθούν υπόψη ορισμένοι βασικοί παράγοντες που το χαρακτη-

ρίζουν.

- Η δομή και ο καθορισμός της αρχιτεκτονικής του τεχνητού νευρωνικού δικτύου. Εξετάζεται ο αριθμός των κρυφών επιπέδων, η συνάρτηση ενεργοποίησης, ο αλγόριθμος εκπαίδευσης του νευρωνικού δικτύου, ο μέγιστος αριθμός επαναλήψεων του αλγόριθμου εκπαίδευσης, κτλ.
- Ο καθορισμός της κατάλληλης δομής των συνόλων εκπαίδευσης και ελέγχου του νευρωνικού δικτύου, όπως για παράδειγμα ο βέλτιστος αριθμός των προτύπων εκπαίδευσης.
- Η ικανότητα γενίκευσης του νευρωνικού δικτύου, η οποία διασφαλίζεται όταν το νευρωνικό δίκτυο παρουσιάζει μικρό σφάλμα πρόβλεψης τόσο στο σύνολο εκπαίδευσης όσο και στο σύνολο ελέγχου.

Η εκπαίδευση τεχνητών νευρωνικών δικτύων είναι μια χρονοβόρα διαδικασία, ιδίως όταν η αρχιτεκτονική των δικτύων είναι σύνθετη καθώς και το μέγεθος των κρυφών επιπέδων. Επίσης, το σύνολο των παραδειγμάτων εκπαίδευσης αποτελεί μέρος της χρονικής πολυπλοκότητας καθώς η διαδικασία εκπαίδευσης είναι μια επαναληπτική διαδικασία πάνω στο σύνολο τους. Ο αριθμός των νευρώνων στο επίπεδο εισόδου είναι ίσος με τον αριθμό των μεταβλητών εισόδου του προβλήματος. Παρόμοια, ο αριθμός των νευρώνων στο επίπεδο εξόδου είναι ίσος με τον αριθμό των μεταβλητών εξόδου του προβλήματος. Όμως, ο αριθμός των νευρώνων των κρυφών επιπέδων είναι αυθαίρετος και δεν υπάρχει γενική μέθοδος προσδιορισμού του. Προκειμένου, να βρεθεί ένας αριθμός νευρώνων για τα κρυφά στρώματα που αποδίδει ικανοποιητικά απαιτείται πειραματισμός σε ένα εύρος τιμών. Πιο συγκεκριμένα, το εύρος τιμών δεν θα πρέπει να είναι πολύ μικρό (μερικές δεκάδες) καθώς το νευρωνικό δίκτυο δεν θα μπορέσει να συγκλίνει σε λύση αλλά ούτε πολύ μεγάλος (μερικές χιλιάδες) καθώς το νευρωνικό δίκτυο μπορεί να χάσει την ικανότητα γενίκευσης του σε άγνωστα παραδείγματα (overfitting).

2.3.1 Αλγόριθμος Πίσω Διάδοσης Σφάλματος με Κάθοδο Κλίσης (Gradient Descent)

Η αναγνώριση προτύπων βασίζεται πάνω στο πρόβλημα της ελαχιστοποίησης μιας συνάρτησης κόστους που έχει τη μορφή αθροίσματος:

$$Q(w) = \frac{1}{n} \sum_{i=1}^n Q_i(w) \quad (2.17)$$

όπου η τιμή της παραμέτρου w που εκτιμάται πρέπει να ελαχιστοποιεί τη συνάρτηση κόστους $Q(w)$. Το άθροισμα σφάλματος $Q(w)$ αποτελείται από τα επιμέρους σφάλματα $Q(i)$ που αντιστοιχούν σε κάθε i δείγμα από τα σύνολα εκπαίδευσης (train set) ή ελέγχου (test set).

Για την εύρεση της τιμής της παραμέτρου που ελαχιστοποιεί το άθροισμα σφάλματος χρησιμοποιείται η μέθοδος κλίσης καθόδου (gradient descent):

$$w = w - l \frac{\sum_{i=1}^n \nabla Q_i(w)}{n} \quad (2.18)$$

όπου το l είναι το μέγεθος του βήματος (step size) ή αλλιώς γνωστό και ως βαθμός μάθησης (learning rate) στο χώρο της Αναγνώρισης Προτύπων.

Αφού ορίσουμε την αρχιτεκτονική του δικτύου, δηλαδή επιλέξουμε τον αριθμό των επιπέδων, το πλήθος των νευρώνων, το τρόπο σύνδεσης αυτών και τις συναρτήσεις ενεργοποίησης σε κάθε επίπεδο, καλούμαστε να το εκπαιδύσουμε χρησιμοποιώντας τον αλγόριθμο πίσω διάδοσης σφάλματος και της μεθόδου κλίσης καθόδου, όπως αυτή περιγράφηκε. Η εκπαίδευση του νευρωνικού δικτύου έχει ως στόχο την κατάλληλη επιλογή των συναπτικών βαρών και πολώσεων για όλους του νευρώνες ώστε να παράγονται οι σωστές έξοδοι για κάθε είσοδο. Η επιλογή αυτή γίνεται με βάση την ελαχιστοποίηση μια κατάλληλης συνάρτησης Q και η διόρθωση των βαρών κατευθύνεται από το επίπεδο εξόδου προς το επίπεδο εισόδου με τον κανόνα που ονομάζεται backpropagation (πίσω διάδοση σφάλματος). Έχοντας επεξηγήσει όλα τα βασικά μέρη, παρουσιάζεται ολόκληρη η διαδικασία εκπαίδευσης:

- Φάση εμπρόσθιας διάδοσης (forward propagation): Το διάνυσμα εισόδου εφαρμόζεται στο επίπεδο εισόδου και αφού γίνουν οι υπολογισμοί σε κάθε επίπεδο, παράγεται τελικά το διάνυσμα εξόδου.

- Φάση πίσω διάδοσης (backward propagation): Αφού υπολογιστεί το διάνυσμα εξόδου μπορεί να συγκριθεί με το αντίστοιχο 'χρυσό' διάνυσμα, δηλαδή το πραγματικό διάνυσμα εξόδου που αντιστοιχεί στην συγκεκριμένη είσοδο. Χρησιμοποιώντας τα δυο διανύσματα εξόδου και τη συνάρτηση σφάλματος Q , μπορούμε να εφαρμόσουμε τον κανόνα της καθόδου κλίσης για την ανανέωση των παραμέτρων. Μετά συνεχίζουμε με το επόμενο δείγμα εισόδου και τη φάση εμπρόσθιας διάδοσης

Η κυκλική παρουσίαση όλων των παραδειγμάτων στο δίκτυο κατά την εκπαίδευση ονομάζεται εποχή. Η εκπαίδευση μπορεί να τερματιστεί με διάφορα κριτήρια, όπως η επιλογή κατωφλίου για την ελαχιστοποίηση κάποιου κόστους ή ο μέγιστος αριθμός επαναλήψεων. Ακόμη μπορούμε για την αποφυγή της υπερπροσαρμογής (overfitting) του δικτύου μας στα δεδομένα εκπαίδευσης, φαινόμενο που σχετίζεται με την κακή ικανότητα γενίκευσης του μοντέλου μας, να χρησιμοποιήσουμε κάποια από τα δείγματα εκπαίδευσης, που λέμε ότι συνιστούν το σύνολο επικύρωσης (validation set) για την περιοδική εξέταση της απόδοσης σε άγνωστα δεδομένα στο τέλος κάθε εποχής. Η εκπαίδευση τότε μπορεί να τερματίζεται όταν η απόδοση πάνω στο validation set αδυνατεί να πέσει για κάποιο συνεχόμενο αριθμό φορών (Early Stopping).

Υπάρχουν 2 είδη μάθησης ανάλογα με τη συνάρτηση που ελαχιστοποιείται και το πότε γίνονται οι ανανεώσεις των βαρών. Παραπάνω περιγράψαμε την on-line μάθηση (on-line learning), όπου οι διορθώσεις των βαρών εκτελούνται σε βάση παράδειγμα προς παράδειγμα και συνεπώς η συνάρτηση κόστους προς ελαχιστοποίηση είναι το στιγμιαίο τετραγωνικό σφάλμα ή το στιγμιαίο κόστος διεντροπίας καθώς και άλλες. Υπάρχει όμως και η μαζική μάθηση (batch learning), όπου οι προσαρμογές στα συναπτικά βάρη του περσεπτρον πολλών επιπέδων εκτελούνται μετά την παρουσίαση του συνόλου των N παραδειγμάτων εκπαίδευσης και συνεπώς η συνάρτηση κόστους προς ελαχιστοποίηση είναι το μέσο τετραγωνικό σφάλμα ή το μέσο κόστος διεντροπίας. Η on-line μάθηση είναι απλή στην υλοποίηση και συγκλίνει πιο γρήγορα. Από την άλλη, η μαζική μάθηση έχει το πλεονέκτημα ότι μπορεί να παραλληλοποιηθεί, αλλά απαιτεί περισσότερο χώρο αποθήκευσης για την ανανέωση των βαρών και μεγαλύτερο αριθμό εποχών εκπαίδευσης. Οι βασικές διαφορές αυτών των παραλλαγών εκπαίδευσης παρουσιάζονται αναλυτικά στην επόμενη ενότητα

2.3.2 Αλγοριθμικές Τροποποιήσεις και επεκτάσεις της Καθόδου Κλίσης

Υπάρχουν τρεις παραλλαγές του αλγορίθμου καθόδου κλίσης οι οποίες διαφέρουν στο πλήθος των δεδομένων που χρησιμοποιούνται για τον υπολογισμό της κλίσης της συνάρτησης προς βελτιστοποίηση. Δεδομένου του πλήθους των δεδομένων, υπάρχει ένα αντιστάθμισμα μεταξύ της ακρίβειας της ενημέρωσης των παραμέτρων και του χρόνου που χρειάζεται ώστε να πραγματοποιηθεί μια ενημέρωση.

- Batch Gradient Descent:

Ο αλγόριθμος Batch Gradient Descent [3] υπολογίζει τις κλίσεις της συνάρτησης κόστους σε σχέση με τις παραμέτρους w για όλα το σύνολο δεδομένων εκπαίδευσης:

$$w = w - l \nabla_w Q(w) \quad (2.19)$$

Καθώς χρειάζεται ο υπολογισμός των κλίσεων για ολόκληρο το σύνολο δεδομένων εκπαίδευσης για να γίνει μια ενημέρωση των παραμέτρων, η μέθοδος αυτή μπορεί να γίνει πολύ αργή και είναι απαγορευτική για σύνολα δεδομένων που δεν χωρούν στην μνήμη. Επίσης, αυτή η μέθοδος δεν επιτρέπει την ενημέρωση του μοντέλου με νέα δεδομένα εκπαίδευσης που προέκυψαν κατά την διάρκεια εκπαίδευσης. Παρόλα αυτά, ο αλγόριθμος Batch Gradient Descent είναι σίγουρο ότι θα συγκλίνει σε ένα καθολικό ελάχιστο αν χρησιμοποιηθεί σε μια κυρτή επιφάνεια λάθους και σε κάποιο τοπικό ελάχιστο για μη κυρτές επιφάνειες.

- Stochastic Gradient Descent:

Ο αλγόριθμος Stochastic Gradient Descent σε αντίθεση με τον αλγόριθμο Batch Gradient Descent ενημερώνει τις παραμέτρους για κάθε δείγμα εκπαίδευσης x_i και επισήμανση y_i :

$$w = w - l \nabla_w Q(w; x_i; y_i) \quad (2.20)$$

Ο αλγόριθμος Batch Gradient Descent εκτελεί ένα πλήθος από αχρείαστους υπολογισμούς για μεγάλες συλλογές δεδομένων εκπαίδευσης, καθώς υπολογίζει συχνά κλίσεις για όμοια παραδείγματα πριν από κάθε ενημέρωση παραμέτρων. Από την άλλη πλευρά, ο αλγόριθμος Stochastic Gradient Descent ξεφεύγει από την ανάγκη για περίσσιους υπολογισμούς καθώς ενημερώνει τις παραμέτρους σε κάθε δείγμα. Για το λόγο αυτό, η σύγκλιση είναι πιο γρήγορη και μπορεί να χρησιμοποιηθεί και για online learning.

Από την άλλη πλευρά, οι συχνές ενημερώσεις των παραμέτρων έχουν μεγάλη διασπορά στις τιμές με αποτέλεσμα να δημιουργούνται διακυμάνσεις στην συνάρτηση κόστους. Το φαινόμενο αυτό, μπορεί να οδηγήσει τον αλγόριθμο να συγκλίνει σε άλλα και εν δυνάμει καλύτερα τοπικά ελάχιστα, σε αντίθεση με τον αλγόριθμο Batch Gradient Descent που συγκλίνει πάντα στο ίδιο ελάχιστο. Παρόλα αυτά, η διαδικασία περιπλέκει αρκετά την σύγκλιση σε ένα ελάχιστο καθώς ο Stochastic Gradient Descent θα έχει συχνές διακυμάνσεις. Θεωρητικά όμως, οι δυο αλγόριθμοι φέρουν παρόμοια συμπεριφορά σύγκλισης και μπορούν να βελτιστοποιήσουν κυρτές και μη-κυρτές επιφάνειες.

- Mini-Batch Gradient Descent:

Χρησιμοποιώντας τα καλύτερα στοιχεία από τους δυο προαναφερθέντες αλγορίθμους, δημιουργείται ο αλγόριθμος Mini-Batch Gradient Descent που ενημερώνει τις παραμέτρους για σύνολα δεδομένων μεγέθους n :

$$w = w - l \nabla_w Q(w; x_{i:i+n}; y_{i:i+n}) \quad (2.21)$$

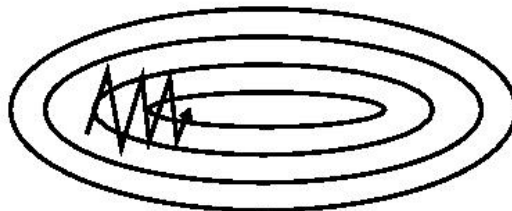
Με αυτό τον τρόπο, 1) ελαχιστοποιείται η διακύμανση των ενημερώσεων των παραμέτρων που οδηγεί σε μια σταθερή σύγκλιση και 2) μπορεί να υλοποιηθεί με πράξεις πινάκων που έχουν βελτιστοποιηθεί για το υλιστικό. Ο αλγόριθμος αυτός αποτελεί την πρώτη επιλογή για την εκπαίδευση ενός νευρωνικού δικτύου.

Παρόλα αυτά, η απλή εκδοχή του αλγορίθμου Mini-batch Gradient Descent δεν εγγυάται την καλή σύγκλιση και προσφέρει μια σειρά από προκλήσεις που πρέπει να αντιμετωπισθούν. Η εύρεση καλής τιμής για το ρυθμό μάθησης μπορεί να είναι δύσκολο να βρεθεί. Μια πολύ μικρή τιμή αναγκάζει τη διαδικασία εκπαίδευσης να συγκλίνει πολύ αργά ενώ μια μεγάλη τιμή μπορεί να προκαλέσει διακυμάνσεις στην τιμή της συνάρτησης κόστους ή να την αδυναμία σύγκλισης του δικτύου. Επίσης, ο ρυθμός μάθησης πρέπει να προσαρμόζεται κατά την διάρκεια εκπαίδευσης ενός δικτύου με κάποιο πρόγραμμα μείωσης με βάση το χρόνο ή τον αριθμό ενημερώσεων [38, 9].

Πέρα από τα προηγούμενα, χαρακτηριστικά με διαφορετικές συχνότητες δεν θα πρέπει να ενημερώνουν τις παραμέτρους με τον ίδιο τρόπο, δηλαδή χαρακτηριστικά που εμφανίζονται σπάνια θα πρέπει να ασκούν μια εντονότερη ενημέρωση των παραμέτρων, ρυθμιζόμενη πάντα από το ρυθμό μάθησης [10]. Για να αντιμετωπισθούν αυτά τα προβλήματα εισήχθησαν μια σειρά από τροποποιήσεις του αλγορίθμου Mini-batch Gradient Descent που χρησιμοποιούν είτε κάποια πρότερη γνώση είτε μαθητικές ιδιότητες των προβλημάτων σύγκλισης.

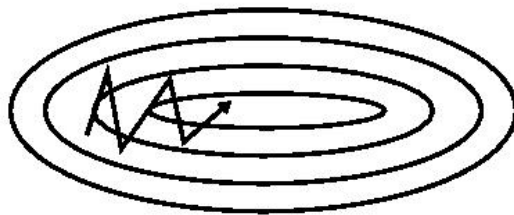
- Mini-batch Gradient Descent με ορμή

Ο αλγόριθμος Mini-batch Gradient Descent δεν μπορεί να αντιμετωπίσει επιφάνειες όπου οι καμπύλη είναι απότομη σε μια διάσταση που είναι σύνηθες φαινόμενο κοντά σε τοπικά ελάχιστα. Σε αυτές τις περιπτώσεις ο αλγόριθμος ταλαντώνεται μεταξύ των 'πλαγιών' σημειώνοντας πολύ μικρή πρόοδο προς το κέντρο.



Σχήμα 2.8: Σύγκλιση αλγορίθμου καθόδου κλίσης χωρίς ορμή

Η ορμή [38] είναι μια μέθοδος που επιταχύνει την σύγκλιση προς μια κατεύθυνση και ταυτοχρόνως μειώνει την ταλάντωση, όπως φαίνεται και στο σχήμα. Αυτό



Σχήμα 2.9: Σύγκλιση αλγορίθμου καθόδου κλίσης με ορμή

επιτυγχάνεται προστίθοντας μια τιμή γ στο διάνυσμα ενημέρωσης της προηγούμενης χρονικής στιγμής από το παρών διάνυσμα εκπαίδευσης:

$$v_t = \gamma v_{t-1} - l \nabla_w Q(w) \quad (2.22)$$

$$w = w - v_t \quad (2.23)$$

Διαισθητικά, όταν χρησιμοποιείται η ορμή, είναι ανάλογο της ρίψης μια μπάλα από ένα λόφο. Η μπάλα συσσωρεύει ορμή καθώς τσουλά στο λόφο και επιταχύνει συνεχώς μέχρι να φτάσει τη μέγιστη ταχύτητα. Το ίδιο πράγμα συμβαίνει και με τις ενημερώσεις των παραμέτρων. Ο όρος της ορμής αυξάνεται για διαστάσεις των οποίων η κλίση είναι στην ίδια κατεύθυνση ενώ μειώνεται για διαστάσεις των οποία η κλίση αλλάζει κατευθύνσεις. Με αυτό τον τρόπο επιτυγχάνεται γρήγορη σύγκλιση και λιγότερη ταλάντωση.

- Mini-batch Gradient Descent με Nesterov επιτάγχυση

Η Nesterov [34] επιτάγχυση είναι ένας τρόπος να βελτιώνεται ακόμη περισσότερο η ταχύτητα σύγκλισης για περιοχές κοντά στο τοπικό ή ολικό ελάχιστο. Για να επηρεαστεί η τιμή των παραμέτρων w θα χρησιμοποιηθεί ο όρος της ορμής γv_{t-1} . Επομένως υπολογίζοντας τον όρο $w - \gamma v_{t-1}$ μας δίνεται μια εκτίμηση της επόμενης θέσης των παραμέτρων, που με τη σειρά του επιτρέπει την εξερεύνηση της καινούργιας θέσης με βάση της εκτίμησης.

$$v_t = \gamma v_{t-1} - l \nabla_w Q(w - \gamma v_{t-1}) \quad (2.24)$$

$$w = w - v_t \quad (2.25)$$

Η προσθήκη της Nesterov επιτάχυνσης επιτρέπει την προσαρμογή των ενημερώσεων στη συνάρτηση κόστους και επιταχύνει με την σειρά της τον αλγόριθμο Mini-batch Gradient Descent. Με καινούργιες προσθήκες θα γίνει δυνατή και η προσαρμογή των ενημερώσεων με βάση κάθε παράμετρο για να γίνεται μικρή ή μεγάλη ενημέρωση των τιμών τους σε σχέση με την σπουδαιότητα τους.

- Αλγόριθμος Adagrad

Ο αλγόριθμος Adagrad [11] είναι ένας αλγόριθμος βελτιστοποίησης με βάση την κλίση που προσαρμόζει το ρυθμό μάθησης για κάθε παράμετρο, εφαρμόζοντας μεγάλες ενημερώσεις για σπάνιες παραμέτρους και μικρές ενημερώσεις για συχνές. Για το λόγο αυτό η χρήση του ενδείκνυται για αραιά δεδομένα. Έχει βρεθεί ότι ο αλγόριθμος Adagrad βελτιώνει την ευρωστία του αλγόριθμου καθόδου κλίσης.

Μέχρι τώρα, οι ενημερώσεις των παραμέτρων w γινόντουσαν μαζικά χρησιμοποιώντας τον ίδιο ρυθμό μάθησης l . Καθώς ο αλγόριθμος Adagrad χρησιμοποιεί διαφορετικό ρυθμό μάθησης για κάθε παράμετρο w_i τη χρονική στιγμή t , δείχνουμε πρώτα τη ανά παράμετρο ενημέρωση. Ορίζεται ως $g_{t,i}$ η κλίση της συνάρτησης κόστους σε σχέση με την παράμετρο w_i τη χρονική στιγμή t :

$$g_{t,i} = \nabla_w J(w_i) \quad (2.26)$$

Η ενημέρωση της κάθε παραμέτρου w_i κάθε χρονική στιγμή t γίνεται:

$$w_{t+1,i} = w_{t,i} - l \cdot g_{t,i} \quad (2.27)$$

Στον κανόνα ενημέρωσης, ο αλγόριθμος Adagrad τροποποιεί το βασικό ρυθμό μάθησης l σε κάθε χρονική στιγμή t για κάθε παράμετρο w_i βασισμένο πάνω στις προηγούμενες κλίσεις που υπολογίστηκαν για το w_i :

$$w_{t+1,i} = w_{t,i} - \frac{\eta}{\sqrt{G_{t,ii} + \epsilon}} \cdot g_{t,i} \quad (2.28)$$

όπου $G_t \in \mathbb{R}^{d \times d}$ είναι ο διαγώνιος πίνακας που κάθε διαγώνιο στοιχείο του είναι το άθροισμα των τετραγώνων των κλίσεων σε σχέση με την αντίστοιχη παράμετρο τη χρονική στιγμή t . Ο παράγοντας ϵ είναι ο όρος ομαλοποίησης και χρησιμοποιείται για την αποφυγή διαίρεσης με το μηδέν.

Καθώς το G_t περιέχει το άθροισμα των τετραγώνων των παλιών κλίσεων σε σχέση με τις παραμέτρους w , μπορεί να γίνει μια υλοποίηση με διανύσματα και πράξεων μεταξύ αυτών.

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{G_t + \epsilon}} \odot g_t \quad (2.29)$$

Ένα από τα πλεονεκτήματα του αλγορίθμου Adagrad είναι ότι προσαρμόζει το ρυθμό μάθησης, με αποτέλεσμα να μην χρειάζεται ρύθμιση με το χέρι. Παρόλα αυτά, το βασικό μειονέκτημα είναι η συσσώρευση των τετραγώνων των κλίσεων στον παρονομαστή. Επειδή ο κάθε όρος που προστίθεται είναι θετικός, τότε το συσσωρευμένο άθροισμα διαρκώς μεγαλώνει κατά την εκπαίδευση. Αυτό με τη σειρά του προκαλεί σμίκρυνση του ρυθμού μάθησης και τελικά γίνεται αδιάφορα μικρός. Για την επίλυση αυτού του προβλήματος, επινοήθηκε ο παρακάτω αλγόριθμος

- Αλγόριθμος Adadelta

Ο αλγόριθμος Adadelta [52] είναι μια επέκταση του αλγορίθμου Adagrad που διορθώνει την μονότονη μείωση του ρυθμού μάθησης. Αντί να συσσωρεύει το άθροισμα όλων των προηγούμενων τετραγώνων των κλίσεων, ο αλγόριθμος Adadelta περιορίζει το άθροισμα σε ένα παράθυρο μήκους w .

Για ένα παράθυρο w , ο αλγόριθμος υπολογίζει αναδρομικά το άθροισμα των κλίσεων ως ένα μέσο όρο όλων των προηγούμενων κλίσεων. Ο παρών μέσος όρος $E[g^2]_t$ τη χρονική στιγμή t εξαρτάται μόνο από το προηγούμενο μέσο όρο και τη τωρινή κλίση:

$$E[g^2]_t = \gamma E[g^2]_{t-1} + (1 - \gamma)g_t^2 \quad (2.30)$$

Επομένως, η νέα εξίσωση ενημέρωσης των παραμέτρων γίνεται:

$$\Delta w_t = -\frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} g_t \quad (2.31)$$

Καθώς ο παρονομαστής είναι η ρίζα των μέσων τετραγώνων μπορούμε να αλλάξουμε το κριτήριο ως εξής:

$$\Delta w_t = -\frac{\eta}{RMS[g]_t} g_t \quad (2.32)$$

Οι συντάχτες σημειώνουν ότι οι παράμετροι σε αυτήν την ενημέρωση δεν ταιριάζουν, διότι η ενημέρωση δεν έχει τις ίδιες υποθετικές οντότητες όπως των παραμέτρων. Για να γίνει πιο κατανοητό, γίνεται αλλαγή από μέσο όρο σε εκθετικά φθίνον μέσο όρο:

$$E[\Delta w^2]_t = \gamma E[\Delta w^2]_{t-1} + (1 - \gamma)\Delta w_t^2 \quad (2.33)$$

Τότε η ρίζα μέσων τετραγώνων γίνεται:

$$RMS[\Delta w]_t = \sqrt{E[\Delta w^2]_t + \epsilon} \quad (2.34)$$

Αφού η $RMS[\Delta w]_t$ είναι άγνωστη, προσεγγίζουμε την τιμή της με την RMS των ενημερώσεων των παραμέτρων μέχρι την προηγούμενη χρονική στιγμή. Η αντικατάσταση του ρυθμού μάθησης l με τον καινούργιο κανόνα ενημέρωσης βασισμένο στο $RMS[\Delta w]_{t-1}$ δίνει την ενημέρωση του αλγόριθμου Adadelta:

$$\Delta w_t = -\frac{RMS[\Delta w]_{t-1}}{RMS[g]_t} g_t \quad (2.35)$$

$$w_{t+1} = w_t + \Delta w_t \quad (2.36)$$

Αξίζει να σημειωθεί ότι με τη χρήση του αλγορίθμου Adadelta δεν χρειάζεται να οριστεί κάποιος αρχικός ρυθμός μάθησης.

- Αλγόριθμος RMRProp

Ο αλγόριθμος RMSProp αποτελεί έναν αδημοσίευτο προσαρμοστικό αλγόριθμο που προτάθηκε από τον George Hinton σε μια διάλεξη του. Ο αλγόριθμος αυτός είναι παρόμοιος με τον Adadelta και δημιουργήθηκαν την ίδιο χρονική περίοδο. Για την ακρίβεια ο RMSProp είναι ίδιος με το πρώτο διάνυσμα ενημέρωσης του αλγορίθμου Adadelta:

$$E[g^2]_t = 0.9E[g^2]_{t-1} + 0.1g_t^2 \quad (2.37)$$

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} g_t \quad (2.38)$$

- Αλγόριθμος Adam

Ο αλγόριθμος Adam (Adaptive Moment Estimation) [20] είναι μια άλλη μέθοδος υπολογισμού προσαρμόσιμων ρυθμών μάθησης για κάθε παράμετρο. Εκτός από την αποθήκευση των εκθετικά φθίνων μέσω των προηγούμενων κλίσεων v_t , όπως στην περίπτωση των Adadelta και RMSProp, ο αλγόριθμος Adam αποθηκεύει και τους εκθετικά φθίνων μέσους όρους των προηγούμενων κλίσεων m_t , παρόμοια με την ορμή:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (2.39)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (2.40)$$

Οι όροι m_t , v_t είναι εκτιμήσεις των πρώτων ορμών (μέσος όρος) και των δεύτερων ορμών (διακύμανση) των κλίσεων αντίστοιχα. Καθώς τα m_t , v_t είναι αρχικοποιημένα μηδενικά διανύσματα, οι συγγραφείς παρατηρούν ότι είναι πολωμένα προς

το μηδέν, ειδικά κατά τις πρώτες χρονικές στιγμές. Για την αντιμετώπιση αυτής της πόλωσης χρησιμοποιούνται μέθοδοι διόρθωσης για τον πρώτο και δεύτερο όρο:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (2.41)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (2.42)$$

Στο επόμενο στάδιο, γίνεται χρήση των νέων όρων κατά τον κανόνα ενημέρωσης των παραμέτρων:

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t \quad (2.43)$$

2.4 Εφαρμογές Νευρωνικών Δικτύων

Η ευρεία και επιτυχή εφαρμογή των νευρωνικών δικτύων σε μια ποικιλία προβλημάτων, σε πολλά διαφορετικά πεδία, τα κάνει πολύ ελκυστικά. Επίσης, η ταχεία πρόοδος της τεχνολογίας των υπολογιστών, ο νόμος του Moore καθώς και η ανάπτυξη προηγμένων αλγορίθμων εκπαίδευσης έχουν καταστήσει δυνατή την γρήγορη εκπαίδευση νευρωνικών δικτύων που γενικεύουν και σε άγνωστα δεδομένα. Από την πρώτη εφαρμογή τους σε προσαρμοζόμενους εξισωτές καναλιών το 1984 μέχρι σήμερα έχουν εφαρμοστεί σε μεγάλη ποικιλία εφαρμογών όπως φαίνεται παρακάτω:

- **Μεταφορές:** Συστήματα αυτόματης οδήγησης, συστήματα αυτόματου παρκαρίσματος, συστήματα διάγνωσης βλαβών φρένων, συστήματα δρομολόγησης
- **Τηλεπικοινωνίες:** Συμπύεση εικόνας και δεδομένων, αυτοματοποιημένες υπηρεσίες πληροφοριών, μετάφραση πραγματικού χρόνου, κωδικοποίηση πληροφορίας
- **Χρηματιστηριακές εφαρμογές:** Ανάλυση αγοράς, πρόβλεψη τιμών μετοχών

- **Επεξεργασία Φωνής:** Αναγνώριση φωνής, αναγνώριση ομιλητή, συμπίεση φωνής, σύνθεση φωνής από κείμενο, εύρεση λεκτικού όρου εντός ομιλίας
- **Ρομποτική:** Έλεγχος τροχιάς και σύστημα όρασης ρομπότ
- **Γεωλογικές έρευνες:** Εντοπισμός πετρελαίου και φυσικού αερίου, γεωλογική ανάλυση εδάφους από δορυφορικές εικόνες
- **Ιατρική:** Ανάλυση καρκινικών κυττάρων, ανάλυση ηλεκτροεγκεφαλογραφήματος και ηλεκτροκαρδιογραφήματος, βελτιστοποίηση χρόνου νοσηλείας, μείωση νοσοκομειακού κόστους, βελτίωση ποιότητας νοσοκομείων
- **Βιομηχανία:** Βιομηχανικός έλεγχος διεργασιών, ανάλυση και σχεδίαση προϊόντων, συστήματα ποιοτικού ελέγχου, διάγνωση βλαβών, διεργασιών και μηχανών, ανάλυση σχεδιασμού χημικών προϊόντων, δυναμικό μοντελάρισμα συστημάτων χημικών διεργασιών, σχεδιασμός και διοίκηση
- **Κοινωνική Ασφάλιση:** Αξιολόγηση εφαρμοζόμενης πολιτικής, βελτιστοποίηση παραγωγής
- **Οικονομία:** Οικονομική Ανάλυση, πρόβλεψη τιμών συναλλάγματος
- **Ηλεκτρονική:** Πρόβλεψη ακολουθίας κωδικών, μορφοποίηση ολοκληρωμένων κυκλωμάτων, έλεγχος διεργασιών, διάγνωση βλαβών ολοκληρωμένων κυκλωμάτων
- **Άμυνα:** Πλοήγηση όπλων, ανίχνευση στόχων, νέα είδη αισθητήρων, σόναρ, ραντάρ, ψηφιακή επεξεργασία σημάτων
- **Τραπεζικές εφαρμογές:** Αναγνώστες επιταγών και άλλων παραστατικών, συστήματα αξιολόγησης αιτήσεων δανειοδότησης
- **Αεροπορία:** Αυτόματοι πιλότοι αεροπλάνων, προσομοιωτές πτήσης, συστήματα αυτομάτου ελέγχου αεροπλάνων, συστήματα ανίχνευσης βλαβών

2.4.1 Πλεονεκτήματα Νευρωνικών Δικτύων

Η χρήση και ανάπτυξη Τεχνητών Νευρωνικών δικτύων παρέχει τις ακόλουθες ιδιότητες:

- **Μη-γραμμικότητα:** Αυτή οφείλεται στο γεγονός ότι ένα νευρωνικό δίκτυο δομείται από τη σύνδεση νευρώνων, οι οποίοι είναι μη-γραμμικές μονάδες. Η μη-γραμμικότητα είναι πολύ σημαντική ιδιότητα, ιδιαίτερα αν ο φυσικός μηχανισμός για την παραγωγή των σημάτων εισόδου είναι μη-γραμμικός
- **Προσαρμοστικότητα:** Τα νευρωνικά δίκτυα έχουν τη δυνατότητα να προσαρμόζουν τα βάρη τους στις αλλαγές του περιβάλλοντός τους. Μερικές φορές οι προσαρμογές οδηγούν σε μείωση της απόδοσης του συστήματος, γιατί θα πρέπει να είναι επαρκείς δίλημμα σταθερότητας.
- **Αποδεικτική Απόκριση:** Ένα τεχνητό νευρωνικό δίκτυο σχεδιάζεται για να παρέχει πληροφορία όχι μόνο το για το συγκεκριμένο παράδειγμα που επιλέχθηκε για αξιολόγηση αλλά και για εμπιστοσύνη στην απόφαση που παίρνεται. Δίνει, δηλαδή, ως έξοδο ποσοστό εμπιστοσύνης στην απόφαση που έβγαλε το σύστημα.
- **Σχεδιασμός Εισόδου-Εξόδου:** Ένα συνηθισμένο παράδειγμα μάθησης που καλείται επιβλεπόμενη μάθηση, εμπλέκει μεταβολή των βαρών των συνάψεων του Νευρωνικού Δικτύου, εφαρμόζοντας ένα σύνολο δειγμάτων εξάσκησης ή παραδείγματα έργων. Κάθε παράδειγμα αποτελείται από ένα μοναδικό σήμα εισόδου και την επιθυμητή απόκριση. Η εξάσκηση του δικτύου επαναλαμβάνεται για πολλά παραδείγματα, μέχρι το δίκτυο να φτάσει σε μια σταθερή κατάσταση, όπου πλέον δεν γίνονται αλλαγές στα βάρη. Έτσι το δίκτυο μαθαίνει από τα παραδείγματα, κατασκευάζοντας ένα σχεδιασμό εισόδου-εξόδου για το πρόβλημα στο χέρι.
- **Αντοχή σε σφάλματα:** Ένα νευρωνικό δίκτυο, υλοποιημένο σε υλικό (hardware) έχει τη σημαντική ιδιότητα να είναι “ανεκτικό σε σφάλματα”, με την έννοια ότι η απόδοσή του μειώνεται αλλά όχι καθοριστικά κάτω από αντίξοες λειτουργικές συνθήκες.

- **Υλοποιησιμότητα σε υλιστικό:** Η συμπαγής παράλληλη φύση του Νευρωνικού Δικτύου, κάνει δυνατή την υλοποίηση του σε hardware τεχνολογία (κυκλώματα τύπου VLSI, FPGA), έτσι ώστε τα νευρωνικά δίκτυα να μπορούν να χρησιμοποιηθούν σε εφαρμογές πραγματικού χρόνου.
- **Αναλογία με Νευροβιολογία:** Ο σχεδιασμός νευρωνικών δικτύων γίνεται σε αναλογία με τον ανθρώπινο εγκέφαλο και εγκεφάλους άλλων ανεπτυγμένων πλασμάτων. Οι νευροβιολόγοι βλέπουν τα νευρωνικά δίκτυα σαν αντικείμενο έρευνας για την εξήγηση νευροβιολογικών φαινομένων. Παράλληλα, οι μηχανικοί βλέπουν στη νευροβιολογία για νέες ιδέες για την επίλυση πολύπλοκων προβλημάτων.

Κεφάλαιο 3

Θεωρητικό υπόβαθρο Ανάλυσης Συναισθήματος

3.1 Ανάλυση Συναισθήματος

Η Συναισθηματική Ανάλυση είναι ο τομέας σύμφωνα με τον οποίο αποσαφηνίζεται η συναισθηματική πρόθεση του γράφοντος. Με άλλα λόγια, επισημαίνεται το κείμενο επάνω στο οποίο γίνεται η ανάλυση, με μία ετικέτα που προδίδει το συναισθηματικό του περιεχόμενο, είτε αυτό έγκειται σε μια κλίμακα θετικού-αρνητικού, είτε αναφέρεται σε συγκεκριμένο συναίσθημα. Η Συναισθηματική Ανάλυση, αποτελεί έναν σχετικά νέο και ανερχόμενο τομέα της Βαθιάς Μηχανικής Μάθησης. Παρόλα αυτά, υπάρχει πλούσιο υλικό εργασιών στη βιβλιογραφία με τεχνικές που επιχειρούν με τον έναν ή τον άλλον τρόπο να επιλύσουν το πρόβλημα της συναισθηματικής κατηγοριοποίησης κειμένων. Αξίζει να σημειωθεί ότι μέχρι και τη στιγμή που γράφτηκε αυτή η εργασία, δεν υπάρχει τεχνική που να έχει επιτύχει εξαιρετικά μεγάλο ποσοστό ακρίβειας (πάνω από 90%) και αυτό γιατί σχετίζεται με κείμενα γραμμένα σε φυσική γλώσσα. Ο θόρυβος αποτελεί τροχοπέδη για τη μεγιστοποίηση της ακρίβειας και συνήθως στη βιβλιογραφία γίνονται αποδεκτές μέθοδοι που κυμαίνονται ακόμα και μεταξύ 65%-70% (χωρίς φυσικά αυτό να εμποδίζει το να αναπτυχθούν μέθοδοι που να προσδοκούν σε μεγαλύτερα ποσά). Στη συνέχεια θα αναλυθούν οι δύο μεγάλες κατηγορίες της Συναισθηματικής Ανάλυσης, καθώς και οι πιο δημοφιλείς προσεγγίσεις τους, σε τεχνικό τομέα. Η Συ-

ναισθηματική Ανάλυση έχει δύο διαφορετικές κατηγοριοποιήσεις: ως προς το είδος της πληροφορίας που παράγεται και ως προς την τεχνική ανάπτυξής τους. Στην πρώτη ανήκουν το Opinion Mining [26] και το Sentiment Analysis, ενώ στη δεύτερη οι τεχνικές Μηχανικής Μάθησης και οι τεχνικές Λεξικογραφικής Ανάλυσης. Η πιο αποδεκτή κατηγοριοποίηση είναι αυτή του τεχνικού τομέα, καθώς η πρώτη απλά ανιχνεύει το κατά πόσον μια μέθοδος ανιχνεύει συναισθήματα ή διπολικές κλάσεις.

3.1.1 Κατηγοριοποίηση Συναισθήματος

Η κατηγοριοποίηση συναισθήματος χωρίζεται σε δυο κατηγορίες: την εξόρυξη γνώμης (Opinion Mining) και την Ανάλυση Συναισθήματος (Affective Analysis).

Σύμφωνα με την πρώτη κατηγορία, τα κείμενα αναλύονται ώστε να αναπαριστούν την κλίση στις απόψεις των ανθρώπων και τυπικά ταξινομούνται σε μια κλίμακα θετικού-αρνητικού, ενώ είναι σύνηθες να προστίθεται και μια ακόμα κατηγορία, αυτή του ουδέτερου. Τα δυο άκρα της κλίμακας δηλαδή το θετικό και το αρνητικό επιδέχονται περαιτέρω κατηγοριοποίηση σε πολύ θετικό και θετικό και αντίστοιχα πολύ αρνητικό και αρνητικό.

Η δεύτερη κατηγορία, επικεντρώνεται περισσότερο στα συναισθήματα των ανθρώπων, με το να εντοπίζει και να αναγνωρίζει την ακριβή πρόθεση τους, αναφορικά με τη σημασία του κειμένου. Τα συναισθήματα αυτά, τις περισσότερες φορές, βασίζονται ενδεικτικά στη θεωρία του Ekman [12], όπου αναφέρεται πως η ανάλυση μπορεί να εφαρμοστεί σε 6 βασικά συναισθήματα:

- Χαρά
- Λύπη
- Φόβος
- Θυμός
- Αγάπη
- Μίσος

Οι παραπάνω κατηγορίες δεν είναι μοναδικές και μπορούν να διαχωριστούν περαιτέρω, π.χ η λύπη μπορεί να διαχωριστεί σε μελαγχολία και θλίψη καθώς και να προστεθούν καινούργιες που υποδηλώνουν άλλα συναισθήματα απέχθεια και έκπληξη. Είτε η μία προσέγγιση, είτε η άλλη, μπορεί να παράγει την ακόλουθη έξοδο:

Μια μοναδική και μονοσήμαντη ετικέτα με το συναίσθημα αυτό καθ' αυτό, εάν η τεχνική το επιτρέπει (συνήθως με τεχνικές Μηχανικής Μάθησης). Ένα δεκαδικό αριθμό για κάθε μία απ' τις ετικέτες που ο αλγόριθμος φιλοδοξεί να κατατάξει τα κείμενα. Στην πρώτη περίπτωση, ένα κείμενο ή λέξη μπορεί να χαρακτηριστεί μονοσήμαντα ως «θετικό», «αρνητικό» ή «ουδέτερο», χωρίς να υπάρχει υπόνοια αμφισημίας της. Για παράδειγμα, η λέξη 'συγχαρητήρια' μπορεί να καταταχθεί ως απόλυτα «θετική» λέξη, η λέξη 'θάνατος' ως απόλυτα «αρνητική», ενώ η λέξη 'αεροπλάνο' μπορεί να επισημανθεί με την ετικέτα «ουδέτερη». Στη δεύτερη περίπτωση, εάν φέρ' ειπείν υπάρχουν 6 συναισθήματα, ο αλγόριθμος θα παράγει έναν δεκαδικό αριθμό για κάθε ένα απ' αυτά και στο τέλος, η μεγαλύτερη τιμή είναι και αυτή που επικρατεί. Συγκεκριμένες τεχνικές βελτιστοποίησης μπορούν να εφαρμοστούν επάνω στα αποτελέσματα, έτσι ώστε να γίνεται διάκριση από ουδέτερα κείμενα (κείμενα που έχουν σχεδόν ίδια τιμή σε όλες τις ετικέτες) ή κείμενα τα οποία έχουν επικρατούν συναίσθημα.

3.2 Τεχνικές Ανάλυσης Συναισθήματος

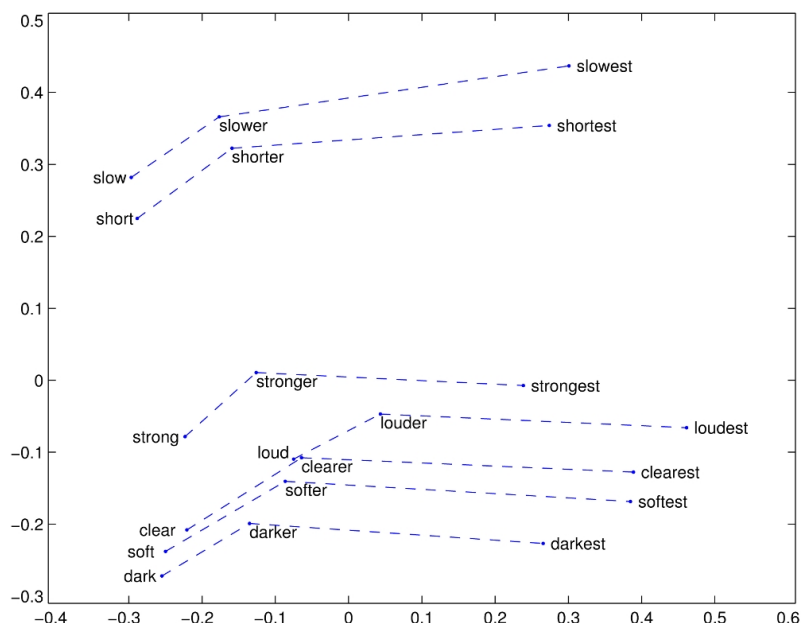
3.2.1 Τεχνικές Μηχανικής Μάθησης

Οι τεχνικές Μηχανικής Μάθησης εφαρμόζουν αλγορίθμους αυτής της ερευνητικής περιοχής και ακολουθούν σε υψηλό επίπεδο τη μεθοδολογία της εκπαίδευσης κάποιου ταξινομητή (classifier) σε κάποιο σύνολο δεδομένων που έχει ταξινομηθεί από ανθρώπους και κατόπιν προβλέπουν με βάση τα προαναφερθέντα ένα άγνωστο σύνολο. Οι τεχνικές αυτές χρησιμοποιούνται σε διάφορες παραλλαγές τους για τα προβλήματα του Opinion Mining και Sentiment Analysis.

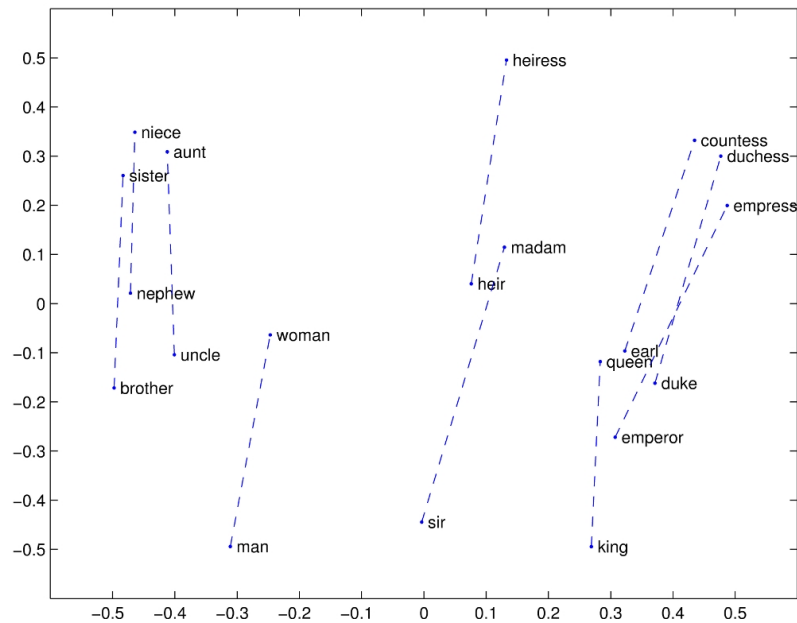
Μέσα σε αυτή την κατηγορία βρίσκονται και τα νευρωνικά δίκτυα που εκπαιδεύονται με βάση της αλγορίθμου πίσω διάδοσης σφάλματος και ένα σύνολο από ταξινομημένα κείμενα και προτάσεις από ανθρώπους. Μια τέτοια τεχνική είναι το Διάνυσμα Λέξης

(Word Vector) [29] και το Διάνυσμα Παραγράφου (Paragraph Vector) που δημιουργήθηκαν από τους Mikolov και Le. Παρόμοιας λογική είναι τα διανύσματα λέξεων Glove [37] που θα παρουσιάζονται σε μετέπειτα κεφάλαιο. Πρόκειται για αλγορίθμους μη επιβλεπόμενης μάθησης, όπου εκπαιδεύεται σταθερού μήκους αναπαράσταση από μεταβλητού μήκους τμήματα κειμένου όπως προτάσεις, παραγράφου και έγγραφα. Η δυνατότητα αυτών των αλγορίθμων να εκμεταλλεύονται αναπαραστάσεις, όπως προτάσεις, μεταβλητού μήκους είναι πολύ σημαντική καθώς ο γραπτός λόγος σπάνια περιέχει πλειοψηφία σταθερού μήκους τμήματα κειμένου.

Οι αλγόριθμοι αυτοί εκπαιδεύονται σε πρώτη φάση σε μεγάλες συλλογές κειμένων προκειμένου να δημιουργηθούν τα διανύσματα λέξεων, τα βάρη U , b του επιπέδου softmax. Ο τρόπος εκπαίδευσης είναι τέτοιος ώστε να μειώνεται η εντροπία των λεκτικών όρων δηλαδή το πόσο χαώδης είναι η αναπαράσταση τους στο χώρο. Με το πέρας της εκπαίδευσης, τα διανύσματα λέξεων έχουν κωδικοποιήσει την πληροφορία με τρόπο που καθίσταται εύκολο να βρεθούν λέξεις με μεγάλη σημασιολογική ομοιότητα, γραμμικές δομές όπως το καλός, καλύτερος, κάλλιστος καθώς και λεκτικές αναλογίες, δηλαδή ερωτήσεις του τύπου ότι είναι η γυναίκα στον άντρα είναι η Υ στον βασιλιά όπου Υ εννοείται η λέξη βασίλισσα.



Σχήμα 3.1: Γραμμικές σχέσεις μεταξύ παραθετικών διαφορών λέξεων της αγγλικής γλώσσας



Σχήμα 3.2: Γραμμικές σχέσεις μεταξύ ανδρών και γυναικών

Στη συνέχεια τα διανύσματα λέξεων μπορούν να χρησιμοποιηθούν συνδυαστικά για τη δημιουργία διανυσμάτων παραγράφων που μπορούν να εκπαιδευτούν σε προσημασμένες παραγράφους από ανθρώπους με σκοπό την επιτυχή επίλυση κάποιο προβλήματος Ανάλυσης Συναισθήματος και Εξαγωγής Γνώμης. Τα διανύσματα παραγράφων έχουν μια σειρά πλεονεκτημάτων. Πιο συγκεκριμένα, μπορούν να δημιουργούνται από μη προσημασμένα δεδομένα με αποτέλεσμα να αποδίδουν σε περιπτώσεις όπου υπάρχει έλλειψη δεδομένων. Επιπλέον, κληρονομούν την πολύ σημαντική ιδιότητα των διανυσμάτων λέξεων η οποία είναι η σημασιολογία. Η σημασιολογία μπορεί να προσδώσει βαρύτητα στην πληροφορία καθώς λέξεις υπό άλλη έννοια μπορούν να αλλάξουν και με αντίστοιχο τρόπο τη συναισθηματική κλίση των φράσεων που ανήκουν.

Περαιτέρω, τα διανύσματα παραγράφων λαμβάνουν υπόψιν τη σειρά των λέξεων μέσα σε μια πρόταση, όμως σε μεγάλα μεγέθη, με τον ίδιο τρόπο που ένα μοντέλο n-gram με μεγάλη τιμή n θα λάμβανε υπόψιν. Η διατήρηση της πληροφορίας της ιεραρχικής σειράς, που έχουν οι λέξεις, που επιτυγχάνεται διευκολύνει την ανάλυση διότι θα έπρεπε να σχηματιστούν σάκοι λέξεων όπου για κάθε φράση θα έπρεπε να περιλαμβάνουν έναν εξαιρετικά μεγάλο αριθμό όρων γεγονός που είναι υπολογιστικά ασύμφορο.

Ο Socher [43] και μια ομάδα ερευνητών χρησιμοποίησαν επίσης ένα Νευρωνικό Δίκτυο που διευθέτησαν τρία σημαντικά ζητήματα που δεν λήφθηκαν υπόψιν από τους

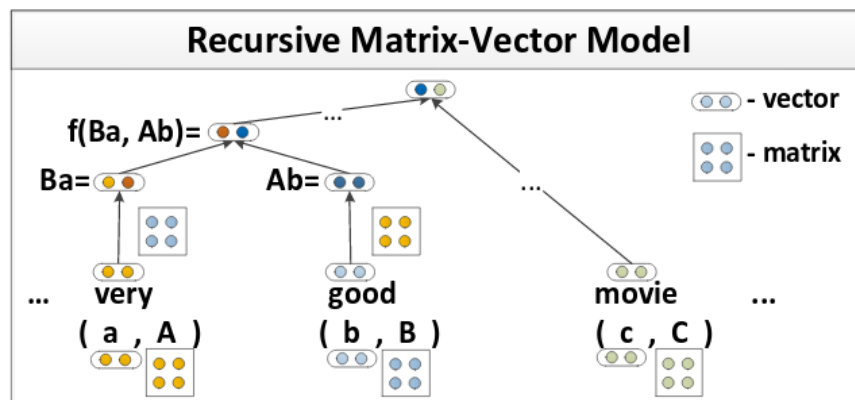
Le και Mikolon αλλά και παρόμοιες εργασίες. Αρχικά, αντί της προσέγγισης του σάκου λέξεων (bag-of-words), έκαναν χρήση της ιεραρχικής οργάνωσης (hierarchical structure) και της συνθετικής σημασιολογίας (compositional semantics) ώστε να γίνει αντιληπτό το συναίσθημα. Εν συνεχεία, το σύστημα που δημιούργησαν, χτίστηκε έτσι ώστε να μπορεί να γίνει εκπαίδευση και μέσω του τμήματος των μη προ-σημασμένων δεδομένων αλλά και με χρήση επιβλεπόμενων συναισθηματικά δεδομένων. Επιπλέον, δεν απαιτεί κανένα συναισθηματικό λεξικό που να βασίζεται στη γλώσσα του κειμένου. Τέλος, αντί να γίνει περιορισμός του συναισθήματος σε μια κλίμακα πόλωσης δηλαδή θετικού και αρνητικού, επιχειρήθηκε να προβλεφθούν πολυδιάστατες κατανομές επάνω σε περίπλοκα διασυνδεδεμένα συναισθήματα. Με άλλα λόγια, δε θεωρήθηκε δεδομένο πως τα συναισθήματα που θα προκύψουν θα ήταν αυστηρά διαχωρισμένα, αλλά το σενάριο ότι θα υπάρχουν αλληλοεπικαλύψεις.

Πρέπει να τονιστεί πως τα συναισθήματα δεν προκύπτουν απ' την κατανομή των 6 κλάσεων, οπότε η μέθοδος δε μπορεί να ταξινομηθεί στο τμήμα του Affective Analysis. Κατά κύριο λόγο, αυτό που γίνεται είναι να εκτίθεται η ισχύς της συναισθηματικής κλίσης (πόσο «θετική» ή «αρνητική» είναι μια συγκεκριμένη φράση κειμένου). Το μοντέλο που προτάθηκε, φιλοδοξούσε να βρει διανυσματικές αναπαραστάσεις φράσεων μεταβλητού μεγέθους, είτε σε περιβάλλοντα μη επιβλεπόμενης, είτε σε περιβάλλοντα ημι-επιβλεπόμενης μάθησης. Ωστόσο, πέραν του ότι το μοντέλο δεν πέτυχε σημαντικά υψηλά ποσοστά ακρίβειας, έπασχε στην περίπτωση που στο κείμενο προς εκπαίδευση εμφανιζόταν άρνηση.

Η επόμενη προσπάθεια που έγινε για την δημιουργία ενός καλύτερου μοντέλου ονομάστηκε Matrix-Vector Recursive Neural Network (MV-RNN) [42], το οποίο κάνει χρήση της έννοιας των διανυσμάτων λέξεων. Το μοντέλο είναι ικανό να εφαρμόσει άρνηση συναισθημάτων (sentiment negation), όταν τα επίθετα συνδυάζονται με τη λέξη "not". Επιγραμματικά, τα σημαντικότερα σημεία του αλγορίθμου είναι τα ακόλουθα:

Αρχικά μια λέξη αναπαρίσταται με δυο μορφές, η πρώτη είναι ένα συνεχές διάνυσμα και η δεύτερη ένας πίνακας παραμέτρων. Η διπλή αναπαράσταση έχει τη λογική του αφινικού μετασχηματισμού όπου κάθε λέξη που αναπαρίσταται ως ένα διάνυσμα, είναι ταυτοχρόνως και ένας τελεστής που επιδρά πάνω στο διάνυσμα με τη μορφή πίνακα. Η διαδικασία της εκπαίδευσης προσπαθεί να μάθει το διάνυσμα της λέξης καθώς και την

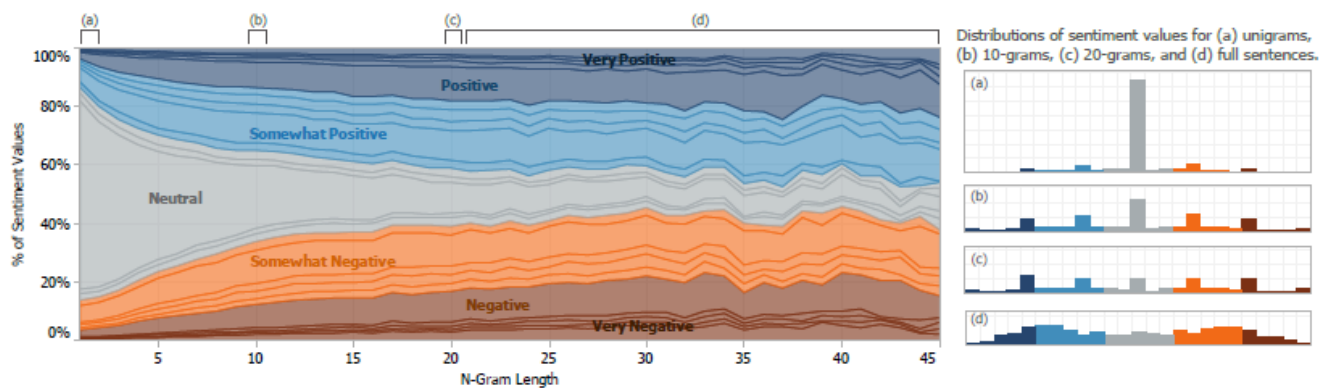
επίδραση που έχει αυτή η λέξη πάνω στις γειτονικές όπως φαίνεται στο σχήμα.



Σχήμα 3.3: Σχηματική αναπαράσταση του δικτύου MV-RNN για το σχηματισμό ενός διανύσματος στο επίπεδο πρότασης

Στην έρευνα που έκαναν, βρήκαν πως τα διαστήματα με διανύσματα σημασιολογικού περιεχομένου δε μπορούν να συλλάβουν την έννοια μεγαλύτερων φράσεων κατάλληλα. Αυτή τους η έρευνα, τους οδήγησε στη πολύ γνωστή συλλογή δεδομένων Stanford Sentiment Treebank [44] Στην πράξη, πρόκειται για ένα σημασιολογικά προ-σημασμένο σύνολο δεδομένων. Περιλαμβάνει ένα πλήθος από κριτικές ταινιών απ την ιστοσελίδα rottentomatoes.com, το οποίο, αρχικά, απαρτιζόταν από 10.662 προτάσεις, με τις μισές θετικές και τις μισές αρνητικές. Κάθε ετικέτα προέκυπτε από μια μεγαλύτερη κριτική ταινίας που αντανάχλούσε τη συνολική πρόθεση του γράφοντος για μια κριτική.

Στην Σχήμα 2.11, φαίνεται ένα κανονικοποιημένο ιστόγραμμα συναισθηματικών ετικετών για κάθε n-gram. Από το μέγεθος 20 και μετά, η πλειοψηφία τους μπορεί να θεωρηθεί ως πλήρεις προτάσεις. Παρατηρώντας το σχήμα αυτό που παρατηρεί κανείς είναι ότι το ισχυρότερο συναίσθημα δημιουργείται σε προτάσεις μεγαλύτερου μεγέθους, ενώ η πλειοψηφία των μικρών φράσεων έχουν ουδέτερο συναίσθημα. Στο δεξί τμήμα της εικόνας κανείς μπορεί να δει τέσσερα γραφήματα που υποδεικνύουν το επικρατούν αυτό συναίσθημα. Το πρώτο γράφημα επιδεικνύει το συναίσθημα των unigrams, δηλαδή των μεμονωμένων λέξεων του συνόλου, που κατά πλειοψηφία φαίνεται πως είναι το ουδέτερο. Τα επόμενα γραφήματα είναι συλλογές από 10, 20 και περισσότερες λέξεις.



Σχήμα 3.4: Κανονικοποιημένο ιστόγραμμα συναισθηματικών ετικετών, για κάθε n-gram

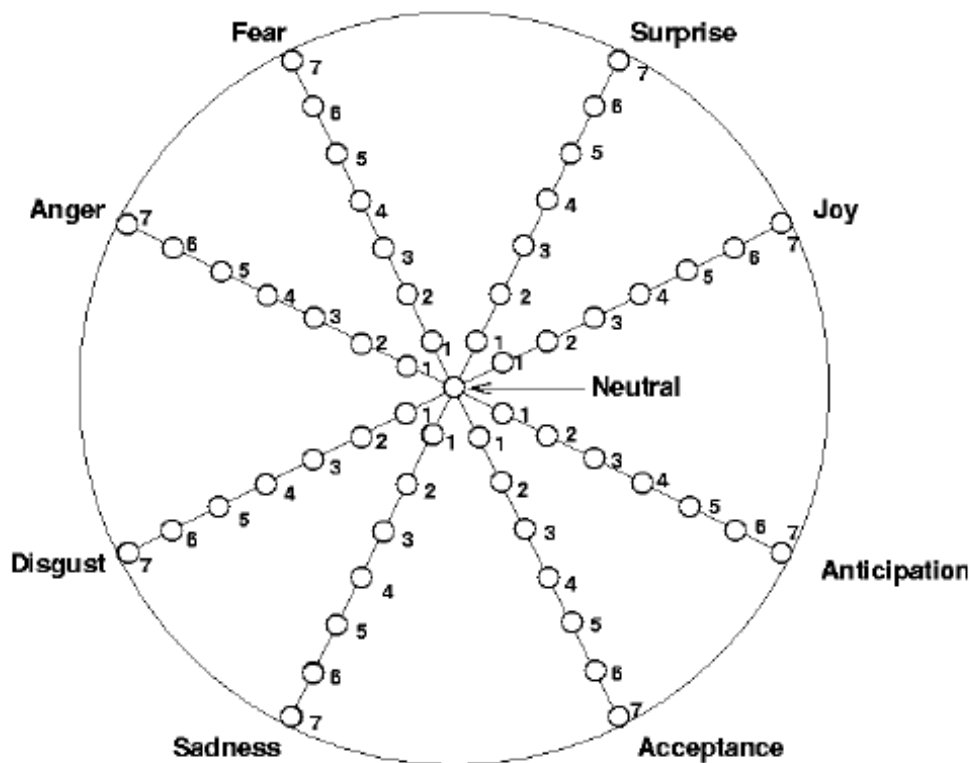
Όπως είναι αναμενόμενο απ' την αριστερή εικόνα, όσο αυξάνονται οι λέξεις, τόσο πιο έντονο γίνεται το συναίσθημα. Αυτό είναι φυσιολογικό, δεδομένου του ότι με περισσότερες λέξεις, ενισχύεται το νόημα και αποσαφηνίζεται καλύτερη η συναισθηματική κλίση του κειμένου.

Μία διαφορετική προσέγγιση ακολουθήθηκε από ερευνητή Gill. Αντίθετα με τις προηγούμενες τεχνικές, που κατέτασσαν ένα κείμενο σε μια διπολική κλίμακα συναισθημάτων, εν προκειμένω η ανάλυση επιχείρησε να προβλέψει το ακριβές συναίσθημα. Τα συναισθήματα που εξήχθησαν στα πλαίσια εκείνης της έρευνας, επικεντρώθηκαν σε 8 βασικά, που προήλθαν από ένα σύνολο συναισθηματικών ιδεών. Οργανώθηκαν, δε, σε τέσσερις κατηγορίες:

- Χαρά-Λύπη (Joy-Sadness)
- Αποδοχή-Απέχθεια (Accept-Disgust)
- Φόβος-Θυμός (Fear-Anger)
- Έκπληξη-Προσμονή (Surprise-Anticipation)

Η προτεινόμενη μέθοδος είχε τα ακόλουθα βήματα: Αρχικά, γινόταν η συλλογή των δεδομένων που συνήθως αποτελούν κείμενα σε φυσική γλώσσα. Εν συνεχεία, τα δεδομένα περνούσαν μια διαδικασία προετοιμασίας, χρησιμοποιώντας το πρόγραμμα ανάλυσης κειμένου LIWC (LIWC, 2015) [36]. Για να ελεγχθεί και η σημασιολογική

διάσταση των κειμένων, εφαρμόστηκε η Ανάλυση Λανθάνουσας Σημασιολογίας (Latent Semantic Analysis) καθώς και το Hyperspace Analogue to Language, ενώ οι 10 πιο συχνοί όροι κάθε κειμένου δόθηκαν για ανάλυση. Εν συνεχεία, με χρήση ενός πίνακα 7 υποδειγματικών λέξεων, υπολογίστηκε η σημασιολογική απόσταση των 10 πιο συχνών όρων, εφαρμόζοντας και τις δύο προαναφερθείσες τεχνικές. Τέλος, με την ολοκλήρωση της εκτέλεσης, τα κείμενα είχαν καταταχθεί στην πιο σωστή κατηγορία.



Σχήμα 3.5: Τροχός συναισθημάτων – Πηγή: (Gill, et al., 2008)

Το σχήμα 2.12 παρουσιάζει έναν τροχό συναισθημάτων, μια αναπαράσταση των κλάσεων στις οποίες η μέθοδος επιχείρησε να κατατάξει τα κείμενα. Όσο αυξάνεται η απόσταση από το κέντρο του κύκλου, τόσο πιο ισχυρό γίνεται το εκάστοτε συναίσθημα. Τιμές που πέφτουν στο κέντρο, θεωρούνται ουδέτερου συναισθήματος, ενώ κάθε ένας απ' τους άξονες που το τέμνουν, αντιπροσωπεύει και από ένα άλλο συναίσθημα (από τα 8 που χρησιμοποιήθηκαν). Η αναπαράσταση των συναισθημάτων με τροχό είναι μια θεωρείται μια πολύ καλή μέθοδος από το πεδίο της ψυχολογίας, παρόλα αυτά δεν είναι

σε θέση να απεικονίσει καλά συνδυασμούς συναισθημάτων που μπορεί να βιώνει ένα άνθρωπος όπως προσμονή (anticipation) και φόβος (fear).

3.2.2 Τεχνικές με λεξικά

Σ' αυτή την κατηγορία τεχνικών, ακολουθείται μια διαφορετική προσέγγιση από τη Μηχανική Μάθηση, όπου απαιτείται εκπαίδευση σε ένα προ-σημασμένο σύνολο δεδομένων. Αυτό που χρησιμοποιείται, είναι ένα «συναισθηματικό λεξικό» (Sentiment Dictionary) όπου υπάρχουν όλες οι λέξεις (συνήθως τα θέματά τους), με ένα σκορ που δείχνει είτε τη συναισθηματική τους κλίση για την κλίμακα αρνητικό-θετικό, είτε το σκορ σε κάθε ένα απ' τα 6 (ή περισσότερα) βασικά συναισθήματα.

Λεξικά για παρόμοιες προσεγγίσεις, μπορούν να δημιουργηθούν χειροκίνητα (Taboada, et al., 2011) [46] (Stone, et al., 1966) [45] (Tong, 2001), είτε αυτόματα, χρησιμοποιώντας λέξεις “seed”, που ενημερώνουν και επεκτείνουν το λεξικό (Hatzivassiloglou & McKeown, 1997) [14], (Turney, 2002) [48], (Turney & Littman, 2003) [49]. Πολλές από τις έρευνες που έχουν γίνει επικεντρώνονταν στη χρήση επιθέτων ως ενδεικτικές λέξεις για την κατανόηση του συναισθηματικού προσανατολισμού του κειμένου. Αρχικά, δημιουργείται μια λίστα με επίθετα και τις αντίστοιχες τιμές συναισθηματικού προσανατολισμού σε ένα λεξικό. Κατόπιν, για οποιοδήποτε κείμενο, εξάγονται όλα τα επίθετα και επισημαίνονται με την τιμή του σημασιολογικού τους προσανατολισμού, χρησιμοποιώντας τα σκορ του λεξικού. Τέλος, τα σκορ αυτά αθροίζονται σε ένα συνολικό για όλο το κείμενο προς ανάλυση. Η παραπάνω διαδικασία, βασίζεται στην ακόλουθη παραδοχή (Osgood, et al., 1957) [35]: οι μεμονωμένες λέξεις έχουν αυτό που αναφέρεται ως «πρότερη πολικότητα» (prior polarity), δηλαδή συναισθηματική κλίση που είναι ανεξάρτητη του κειμένου. Αυτή η συναισθηματική κλίση, μπορεί να εκφραστεί ως αριθμητική τιμή.

3.3 Εφαρμογές Ανάλυσης Συναισθήματος

Η ανάλυση συναισθήματος και εξόρυξη γνώμης αποτελούν τον κεντρικό πυλώνα αρκετών μικρών και μεγάλων επιχειρήσεων ανα τον κόσμο.

1. Εφαρμογές με βάση τις κριτικές

Οι δυνατότητες εφαρμογών που χρησιμοποιούν τις κριτικές σχετίζονται με την αναζήτηση και ανάλυση κριτικών στις διάφορες ιστοσελίδες, όπως για παράδειγμα το Amazon όπου ο χρήστης μπορεί να γράψει κριτική για το προϊόν που αγόρασε. Οι κριτικές μπορούν να ομαδοποιηθούν και να παραχθεί ένας δείκτης εκτίμησης της ποιότητας κάποιου προϊόντος ή υπηρεσίας καθώς και μια συνοπτική αναφορά για τις επικρατέστερες θετικές και αρνητικές απόψεις.

Επιπλέον, οι χρήστες αρκετές φορές κάνουν λάθη στις εκτιμήσεις του πατώντας χαμηλότερη βαθμολογία που δεν συνάδει με την αντίστοιχη κριτική που έγραψαν. Για την αντιμετώπιση αυτού του σφάλματος έχουν παραχθεί εφαρμογές που μετατρέπουν την κριτική σε βαθμολογία ή βελτιώνουν κάποια πόλωση που εμφανίζεται συχνά στις κριτικές.

2. Εφαρμογές ως μέρος άλλων εφαρμογών

Η ανάλυση συναισθήματος αποτελεί κομμάτι πολλών εμπορικών εφαρμογών που θέλουν να αξιολογήσουν την πορεία μιας συνεδρίας αλληλεπίδρασης με το χρήστη. Για παράδειγμα, αυτόματοι πράκτορες που επικοινωνούν με ανθρώπους μέσω φωνής περιλαμβάνουν συστήματα αναγνώρισης συναισθήματος καθώς τυχόν λάθη στην επικοινωνία των δυο μπορεί να επηρεάσουν τον χρήστη και κατά επέκταση να την καταστήσουν αναποτελεσματική. Άλλο παράδειγμα αποτελεί η συναισθηματική ανάλυση του οδηγού ενός αυτοκινήτου κατά τη διάρκεια μιας διαδρομής που αποτελεί κεντρικός πυλώνας στα σύγχρονα συστήματα των αυτοκινήτων, καθώς η συναισθηματική φόρτιση του οδηγού μπορεί να προκαλέσει κάποιο ατύχημα.

3. Εφαρμογές επιχειρηματικής ευφυΐας

Το πεδίο της ανάλυσης συναισθήματος και εξόρυξης γνώμης είναι δομημένο με γνώμονα διάφορες εφαρμογές τεχνητής νοημοσύνης. Μια από αυτές είναι η επιχειρηματική ευφυΐα που πραγματεύεται με την ανάλυση όγκο δεδομένων για να απαντηθούν διάφορα ερωτήματα που προκύπτουν στην εταιρεία. Χαρακτηριστικό παράδειγμα μια ερώτηση αποτελεί το "Γιατί οι καταναλωτές προτιμούν το ανταγωνιστικό προϊόν X από το δικό μας προϊόν Ψ;". Για να απαντηθεί μια τέτοια ερώτηση με χρήση συναισθηματικής ανάλυσης, αρκεί κανείς να μαζέψει με

αυτόματο τρόπο μια λίστα με τα μειονεκτήματα του προϊόντος Υ και μια λίστα από πλεονεκτήματα του προϊόντος Χ. Οι πληροφορίες αυτές θα παρθούν από τις κριτικές των δυο προϊόντων σε διάφορες ιστοσελίδες.

Κεφάλαιο 4

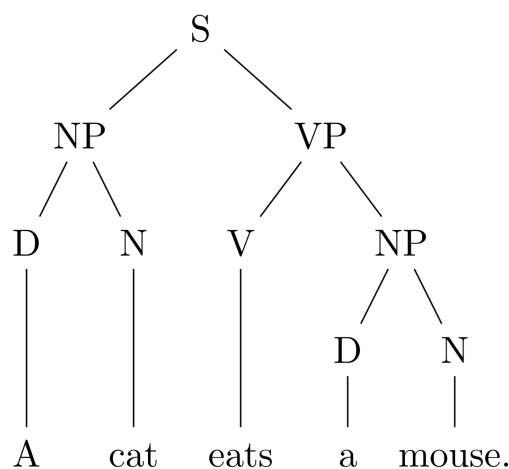
Αναδρομικά Νευρωνικά Δίκτυα για Επεξεργασία Φυσικής Γλώσσας

Στο κεφάλαιο αυτό γίνεται μια ενδελεχής και εις βάθος παρουσίαση της θεωρίας των αναδρομικών νευρωνικών δικτύων και σε επιμέρους παραλλαγές που εμφανίζονται στη βιβλιογραφία. Επιπροσθέτως, παρουσιάζονται αναλυτικά οι ιδιότητες που καθιστούν τα αναδρομικά νευρωνικά δίκτυα ένα δυνατό εργαλείο επεξεργασίας της φυσικής γλώσσας.

4.1 Αρχιτεκτονική Αναδρομικών Νευρωνικών Δικτύων

Ένα αναδρομικό νευρωνικό δίκτυο (recursive neural network - RNN) αποτελεί ένα είδος βαθύ νευρωνικού δικτύου το οποίο δημιουργείται χρησιμοποιώντας το ίδιο σύνολο βαρών αναδρομικά πάνω σε μια δομή, προκειμένου να δημιουργηθεί μια δομημένη πρόβλεψη πάνω σε δομές εισόδου μεταβλητού μεγέθους. Παράδειγμα τέτοιο δομών αποτελούν τα δένδρα αναπαράστασης της συντακτικής δομής μια πρότασης, όπως φαίνεται στο σχήμα. Η δομή εισόδου προσπελαίνεται με τοπολογική σειρά γεγονός που αναδεικνύει την υπεροχή των αναδρομικών δικτύων σε προβλήματα σχετικά με ιεραρχικές δομές όπως είναι η συντακτική ανάλυση μιας πρότασης. Ιστορικά τα αναδρομικά

νευρωνικά δίκτυα εισήχθησαν για την εκμάθηση κατανεμημένων αναπαραστάσεων που σχετίζονται με δομές όπως είναι οι λογικοί όροι. Για παράδειγμα, μια έκφραση άλγεβρας Boolean μπορεί να αναπαρασταθεί ως μια δενδρική δομή που περιλαμβάνει τελεστές και μεταβλητές και επομένως η χρήση αναδρομικών νευρωνικών δικτύων θα μπορούσε να υπολογίσει το αποτέλεσμα μια τέτοιας έκφρασης. Τα μοντέλα και τα γενικά θεωρητικά πλαίσια των αναδρομικών δικτύων ξεκίνησαν να αναπτύσσονται από τη δεκαετία του 90 και γνώρισαν ιδιαίτερη επιτυχία πρόσφατα ξεπερνώντας σε απόδοση άλλες δομές νευρωνικών δικτύων σε προβλήματα φυσικής γλώσσας.



Σχήμα 4.1: Παράδειγμα συντακτικού δέντρου που δημιουργείται από την ανάλυση της πρότασης A cat eats a mouse

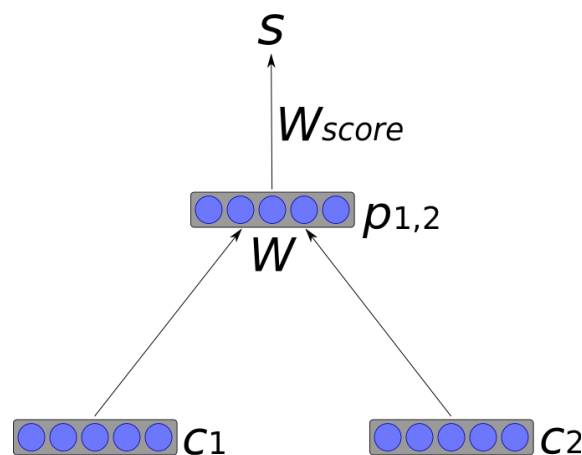
Στην πιο απλή αρχιτεκτονική του αναδρομικού νευρωνικού δικτύου, οι κόμβοι συνδυάζονται με τους πατέρες χρησιμοποιώντας μόνο ένα πίνακα βαρών για ολόκληρο το νευρωνικό δίκτυο και μια μη-γραμμική συνάρτηση όπως είναι η \tanh και άλλες που περιγράφηκαν στην ενότητα 2.1.2. Το νευρωνικό δίκτυο, όπως φαίνεται και από το σχήμα 3.2, δέχεται ως είσοδο δυο διανύσματα, εν προκειμένω τα c_1 , c_2 και δημιουργεί το διάνυσμα του πατέρα $p_{1,2}$ το οποίο μετά χρησιμοποιείται για να βγει κάποιο απόφαση για ολόκληρη τη δενδρική δομή. Τα μαθηματικά που διέπουν την αρχή λειτουργίας του δικτύου είναι:

$$p_{1,2} = \tanh(W[C_1; C_2]) \quad (4.1)$$

όπου ο πίνακας W είναι ο πίνακας παραμέτρων με διάσταση $n \times 2n$ και τα διανύσματα $p_{1,2}$, c_1, c_2 έχουν διάσταση n . Αυτή η αρχιτεκτονική, με μερικές προσθήκες, έχει

χρησιμοποιηθεί με επιτυχία στο πρόβλημα ανάλυση φυσικών σκηνών από εικόνες και στη συντακτική ανάλυση κειμένων και προτάσεων.

Χρησιμοποιώντας την απλή αρχιτεκτονική του βασικού αναδρομικού δικτύου, μπορεί κάποιος να αναπτύξει σύνθετα αναδρομικά δίκτυα που χρησιμοποιούν πληροφορία από διάφορους κόμβους πέρα από τους γειτονικούς, τροποποιώντας έτσι τη ροή της πληροφορίας στη δομή και τη τοπολογική σειρά που αυτή έχει. Η αμέσως επόμενη δομή που εισήχθηκε ήταν τα MV-RNN δηλαδή αναδρομικά δίκτυα που κάθε κόμβος αναπαρίσταται και από ένα διάνυσμα και από ένα πίνακα.



Σχήμα 4.2: Παράδειγμα του πιο απλού αναδρομικού δικτύου που εφαρμόζεται πάνω σε δυο κόμβους C1, C2 για τη δημιουργία της αναπαράστασης του πατέρα αυτών $p_{1,2}$

Η σύνηθες μεθοδολογία για την δημιουργία αναπαραστάσεων για φράσεις, που αποτελούνται από περισσότερες από μια λέξεις, ήταν η χρήση κάποιου γραμμικού συνδυασμού όπως το άθροισμα ή το άθροισμα με βάρη. Αυτό συμβαίνει στο information retrieval και σε διάφορες συναρτήσεις ομοιότητας κειμένου βασισμένου σε ομοιότητα λέξεων. Αυτές οι μεθοδολογίες αποδίδουν βέλτιστα όταν το νόημα μια πρότασης είναι κυριολεκτικά το άθροισμα των μερών της, αλλά παύουν να είναι αποτελεσματικές όταν οι λέξεις λειτουργούν ως τελεστές και αλλάζουν το νόημα των διπλανών τους λέξεων. Για παράδειγμα, η φράση 'απίστευτα ισχυρός' δεν μπορεί να δομηθεί από της διανυσματικές αναπαραστάσεις των επιμέρους λέξεων 'απίστευτα' και 'ισχυρός'. Παρόμοιο παράδειγμα αποτελεί και ο τελεστής 'δεν' που ασκεί μια Boolean λογική τύπου άρνησης (not) στη λέξη που ακολουθεί.

Το πρόβλημα αυτό επιλύεται με τη χρήση του μοντέλου αναδρομικού δικτύου όπως

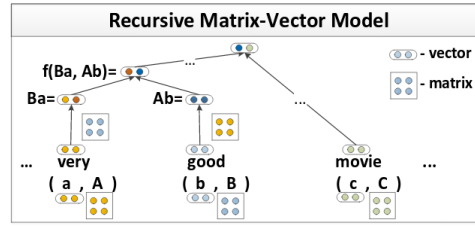
παρουσιάστηκε προηγουμένως καθώς είναι σε θέση να εκφράσει ένα σύνολο από μη γραμμικές συναρτήσεις. Χρησιμοποιώντας, όμως, μόνο έναν πίνακα παραμέτρων W είναι δύσκολο να ενσωματωθούν όλοι οι συνδυασμοί νοημάτων όλων των φαινομένων φυσικής γλώσσας όπως είναι η άρνηση, η απότομη άρνηση, η σύζευξη και άλλα. Παρόμοια έρευνα με το MV-RNN έχει δείξει ότι είναι εφικτό να μοντελοποιήσεις τελεστές της φυσικής γλώσσας χρησιμοποιώντας πίνακες οι οποίοι μπορούν να επηρεάζουν το νόημα άλλων λέξεων, δηλαδή στο παράδειγμα που αναφέρθηκε προηγουμένων η αναπαράσταση της λέξης 'απίστευτα' με πίνακα μπορεί να λειτουργήσει ως τελεστής στο διάνυσμα της λέξης 'ισχυρός'. Αυτές οι μέθοδοι σε εξαίρεση με το MV-RNN μπορούν να μοντελοποιήσουν γραμμικές συναρτήσεις μεταξύ ζευγών λέξεων, σε αντίθεση με το MV-RNN που χρησιμοποιεί μια μη-γραμμική λογική προκειμένου να υπολογίσει μια νοηματική αναπαράσταση μια ολόκληρης πρότασης ή μιας μεγάλης φράσης.

Επί της ουσίας, το MV-RNN συνδυάζει τη δύναμη της έρευνας πάνω σε τελεστές με αναπαράσταση από πίνακες και την ιδέα του αναδρομικού νευρωνικού δικτύου. Συγκεκριμένα, κάθε λέξη αναπαριστάται ως ένα εκπαιδύσιμο διάνυσμα και ένα εκπαιδύσιμο πίνακα που μετά το πέρας της διαδικασίας εκπαίδευσης έχουν μοντελοποιήσει πλέον συναρτήσεις για προτάσεις μεγάλου μεγέθους με οποιαδήποτε συντακτική δομή. Περαιτέρω ανάλυση στο δίκτυο έδειξε ότι λέξεις που αποτελούν κυρίως τελεστές όπως είναι τα επίθετα και τα επιρρήματα έχουν διάνυσμα που είναι κοντά στο μηδέν ενώ οι αντίστοιχοι πίνακες τους έχουν μορφή τελεστών, η οποία δεν έχει κάποια άμεση φυσική σημασία. Οι εξισώσεις που διέπουν το MV-RNN είναι:

$$p = f_{A,B}(a, b) = f(Ba, Ab) = g(W[Ba; Ab]) \quad (4.2)$$

Στην παραπάνω εξίσωση οι λέξεις a, b αναπαρίστανται από τα διανύσματα a, b και του αντίστοιχους πίνακες A, B στα οποία πολλαπλασιάζεται ο πίνακας βαρών W και η μη-γραμμική συνάρτηση f .

Παρόλο που η απόδοση του δικτύου MV-RNN είναι ικανοποιητική, ένα μεγάλο μειονέκτημα αποτελεί το πλήθος των εκπαιδύσιμων παραμέτρων που είναι πολλές τάξεις μεγαλύτερο από το απλό αναδρομικό δίκτυο αφού κάθε λέξη αναπαρίστανται και με διάνυσμα και πίνακα. Για την επίλυση αυτού του προβλήματος επιλέχθηκε η χρήση ενός μεγάλου τένσορα που είναι καθολικός για όλο το νευρωνικό δίκτυο. Το καινο-



Σχήμα 4.3: Παράδειγμα του MV-RNN αναδρομικού δικτύου που εφαρμόζεται πάνω σε δυο κόμβους a , b , c για τη δημιουργία της αναπαράστασης του πατέρα αυτών p_1 και στη συνέχεια p_2

ύργιο αυτό τεχνητό νευρωνικό δίκτυο ονομάστηκε Recursive Neural Tensor Network - RNTN που μεταφράζεται ως Αναδρομικό Νευρωνικό Δίκτυο με Τένσορα. Η χρήση του καθολικού τένσορα εισήχθη προκειμένου να μειωθεί ο όγκος των εκπαιδευσιμων παραμέτρων χρησιμοποιώντας μια αναπαράσταση σταθερού μεγέθους αλλά ικανής να μοντελοποιήσει όλες τις δυνατές συναρτήσεις και συνθέσεις που μπορεί κανείς να βρει στη φυσική γλώσσα. Ο συγκεκριμένος τένσορας επιδρά πάνω στα διανύσματα εισόδου του νευρωνικού δικτύου και εφαρμόζει μια συλλογή από συναρτήσεις όπως αυτές έχουν μοντελοποιηθεί κατά τη διάρκεια εκπαίδευσης.

Όπως φαίνεται και στην εικόνα 3.4, το γινόμενο του τένσορα h με τις λέξεις b, c , που αναπαρίστανται με διανύσματα, υπολογίζεται ως εξής:

$$h = [b; c]^T V [b; c]; h_i = [b; c]^T V^{[i]} [b; c] \quad (4.3)$$

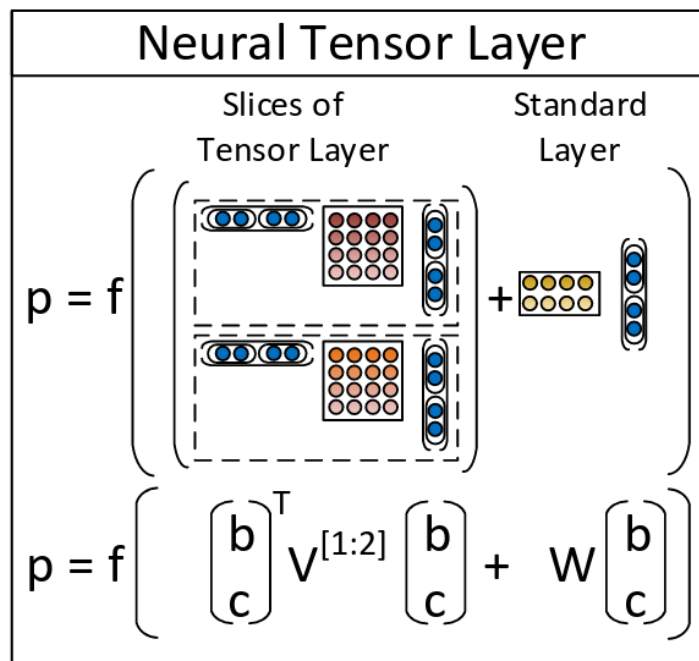
όπου το V είναι ο τένσορας που εφαρμόζεται πολλαπλούς διγραμμικούς μετασχηματισμούς στα διανύσματα εισόδου b και c . Η αναδρομική φύση του δικτύου φαίνεται με τις παρακάτω εξισώσεις που δείχνουν τον τρόπο υπολογισμού της διανυσματικής αναπαράστασης του κόμβου 'πατέρα' p_1 και στη συνέχεια p_2 .

$$p_1 = f([b; c]^T V [b; c] + W [b; c]) \quad (4.4)$$

$$p_1 = f([a; p_1]^T V [a; p_1] + W [a; p_1]) \quad (4.5)$$

όπου ο πίνακας W είναι ο κλασικός πίνακας βαρών που διέπει τη λειτουργία οποιουδήποτε νευρωνικού δικτύου αναδρομικού και μη. Το κύριο πλεονέκτημα του δικτύου

RNTN από τα άλλα αναδρομικά δίκτυα είναι ότι ο τένσορας μπορεί να συσχετίσει τα δυο διανύσματα εισόδου. Διαισθητικά, μπορούμε να θεωρήσουμε ότι κάθε υπό-πίνακας που βρίσκεται στον τένσορα καθώς και κάθε υπό-διάνυσμα μπορεί να μάθει μια συγκεκριμένη συνάρτηση που υπάρχει στη φυσική γλώσσα.



Σχήμα 4.4: Ένα επίπεδο του Recursive Neural Tensor Network που εφαρμόζεται πάνω στη λέξεις - κόμβους \mathbf{b}, \mathbf{c}

Μια ποιοτική ανάλυση του δικτύου RNTN αποκαλύπτει ότι το δίκτυο μαθαίνει από μόνο του φαινόμενα φυσικής γλώσσας όπως τους αντιπαραθετικούς συνδυασμούς, δηλαδή συνδυασμούς του τύπου "Ξ αλλά Υ". Ο πρώτος συνδυασμός X ερμηνεύεται ως ένα όρισμα για το δεύτερο συνδυασμό Y. Επιπλέον, το δίκτυο είναι σε θέση να αναγνωρίσει το φαινόμενο της άρνησης ακόμα και αν ο τελεστής άρνησης, για παράδειγμα η λέξη "δεν", βρίσκεται πολλές λέξεις μακριά από τη λέξη που εφαρμόζεται.

4.1.1 Αρχιτεκτονική Αναδρομικών Νευρωνικών Δικτύων με Μνήμη

Η αρχιτεκτονική αναδρομικών νευρωνικών δικτύων μπορεί να εμπλουτισθεί με ένα αρκετά σύνθετο υπολογιστικό μοτίβο, αυτό της μνήμης. Για να προστεθεί μνήμη σε ένα οποιοδήποτε νευρωνικό δίκτυο, πρέπει να εισαχθούν καινούργια εκπαιδευσιμα διανύσματα που δρουν ως γνώμονες της ροής πληροφορίας μέσα στο δίκτυο με σκοπό να εγκλωβίζουν ή να απελευθερώνουν πληροφορία ανάλογα την είσοδο. Η μνήμη στα νευρωνικά δίκτυα χωρίζεται σε δυο γενικές κατηγορίες 1) οι μονάδες Long Short-Term Memory (LSTM) δηλαδή οι μονάδες Μακριάς Βραχυπρόθεσμης Μνήμης και 2) οι μονάδες Gated Recurrent Units (GRU) δηλαδή οι μονάδες Φραγμένης Επαναλαμβανόμενης Μνήμης.

Για την λεπτομερή ανάλυση αυτών των δυο κατηγοριών θα επικεντρωθούμε για αρχή στην πιο απλή μορφή των αναδρομικών δικτύων, δηλαδή των Επαναλαμβανόμενων νευρωνικών δικτύων (Recurrent Neural Networks). Τα επαναλαμβανόμενα νευρωνικά δίκτυα, όπως αυτά περιγράφηκαν στο κεφάλαιο 2, μπορούν να επεξεργάζονται ακολουθίες εισόδου οποιοδήποτε μεγέθους χρησιμοποιώντας μια αναδρομική συνάρτηση μετάβασης πάνω στο διάνυσμα κρυφής αναπαράστασης h_t . Σε κάθε χρονικό βήμα t , η κρυφή αναπαράσταση h_t είναι μια συνάρτηση μεταξύ του διανύσματος εισόδου x_t που το νευρωνικό δίκτυο δέχεται τη στιγμή t και της προηγούμενης κρυφής αναπαράστασης h_{t-1} . Για παράδειγμα, το διάνυσμα εισόδου x_t μπορεί να είναι ένα διάνυσμα αναπαράστασης της t λέξης σε ένα κομμάτι κειμένου. Η κρυφή κατάσταση $h_t \in \mathbb{R}^d$ μπορεί να ερμηνευθεί ως μια κατανεμημένη αναπαράσταση διάστασης d της ακολουθίας από αντικείμενα μέχρι τη στιγμή t .

Όπως παρουσιάστηκε και στο κεφάλαιο 2 η σύννητες συνάρτηση μετάβασης του Επαναλαμβανόμενου Δικτύου είναι ένας αφινικός μετασχηματισμός ακολουθούμενος από μια μη-γραμμική συνάρτηση όπως είναι η υπερβολική εφαπτομένη:

$$h_t = \tanh(Wx_t + Uh_{t-1} + b) \quad (4.6)$$

Δυστυχώς, ένα πρόβλημα με τα Επαναλαμβανόμενα Δίκτυα με συναρτήσεις μετάβασης αυτής της μορφής είναι ότι κατά τη διάρκεια της εκπαίδευσης, μέρη των διανυσμάτων

κλίσης (gradient) μπορεί να μεγαλώσουν ή να μειωθούν εκθετικά για μεγάλες ακολουθίες εισόδου. Αυτό το πρόβλημα είναι γνωστό ως exploding or vanishing gradients και δημιουργεί προβλήματα κατά την εκμάθηση ενός Επαναλαμβανόμενου Δικτύου για συσχετίσεις μεγάλων αποστάσεων πάνω σε μια ακολουθία.

Η LSTM αρχιτεκτονική δημιουργήθηκε για την αντιμετώπιση του προβλήματος εκμάθησης συσχετίσεων μεγάλων αποστάσεων και χρησιμοποιεί τα λεγόμενα κύτταρα μνήμης. Τα κύτταρα μνήμης μπορούν να διατηρήσουν μια κατάσταση για ένα μεγάλο χρονικό διάστημα. Θεωρούμε ότι μια μονάδα LSTM τη χρονική στιγμή t είναι μια συλλογή από διανύσματα του χώρου \mathbb{R}^d : 1) ένα θύρα εισόδου (input gate) i_t , 2) μια θύρα λησμόνησης (forget gate) f_t , 3) μια θύρα εξόδου (output gate) o_t , 4) ένα κύτταρο μνήμης (memory cell) c_t και τέλος 5) μια κρυφή κατάσταση ή αναπαράσταση h_t . Το μέγεθος d ονομάζεται σε οποιαδήποτε αρχιτεκτονική μνήμης ως "διάσταση μνήμης".

Οι εξισώσεις μετάβασης που χαρακτηρίζουν μια μονάδα LSTM είναι οι εξής:

$$i_t = \sigma(W^i x_t + U^i h_{t-1} + b^i) \quad (4.7)$$

$$f_t = \sigma(W^f x_t + U^f h_{t-1} + b^f) \quad (4.8)$$

$$o_t = \sigma(W^o x_t + U^o h_{t-1} + b^o) \quad (4.9)$$

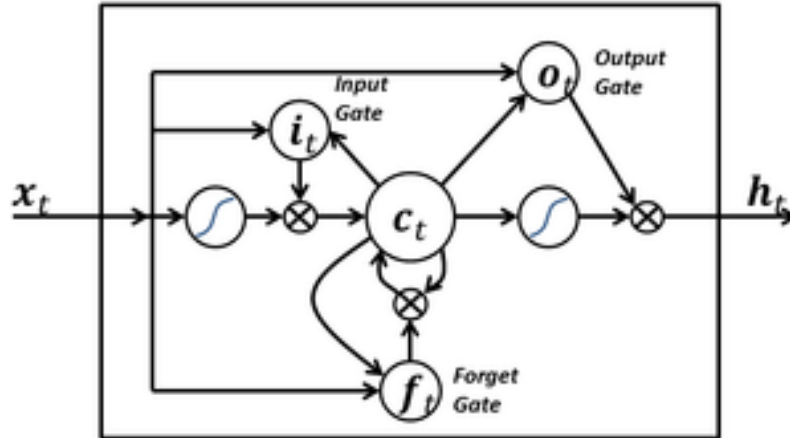
$$u_t = \tanh(W^u x_t + U^u h_{t-1} + b^u) \quad (4.10)$$

$$c_t = i_t \odot u_t + f_t \odot c_{t-1} \quad (4.11)$$

$$h_t = o_t \odot \tanh(c_t) \quad (4.12)$$

όπου το x_t είναι η είσοδος τη χρονική στιγμή t και το \odot συμβολίζει το γινόμενο Hadamard. Διαισθητικά, η θύρα λησμόνησης ελέγχει το βαθμό που θα ξεχαστεί το περιεχόμενο του προηγούμενου κυττάρου μνήμης, η θύρα εισόδου ελέγχει πόσο θα ενημερωθεί η κάθε μονάδα και η θύρα εξόδου ελέγχει την έκθεση της εσωτερικής

κατάστασης μνήμης. Το κρυφό διάνυσμα κατάστασης σε μια μονάδα LSTM είναι επί της ουσίας μια φραγμένη και μερική όψη της κατάστασης της κρυφής μνήμης. Αφού οι αριθμητικές τιμές των μεταβλητών των θυρών διαφέρουν μεταξύ τους, το μοντέλο είναι σε θέση να μοντελοποιήσει πληροφορία από διάφορες χρονικές κλίμακες.



Σχήμα 4.5: Σχηματική απεικόνιση της δομής μια μονάδας Μακριά Βραχυπρόθεσμης Μνήμης - Long Short-Term Memory

Η άλλη κατηγορία μνημών είναι αυτή της φραγμένης επαναλαμβανόμενης μονάδας (Gated Recurrent Unit - GRU). Όμοια με την LSTM μονάδα, η GRU μονάδα δημιουργήθηκε για να αιχμαλωτίζει με συσχετίσεις διαφορετικών χρονικών κλιμάκων προσαρμόζοντας τις εσωτερικές τις καταστάσεις. Η μονάδα GRU αποτελείται και αυτή από διάφορες θύρες που ρυθμίζουν τη ροή της πληροφορίας μέσα στη μονάδα, όμως, δεν έχει την ανάγκη ύπαρξης ενός ξεχωριστού κυττάρου μνήμης. Μια μονάδα GRU για μια χρονική στιγμή t αποτελείται από τέσσερα διανύσματα που ανήκουν στο χώρο \mathbb{R}^d 1) η υπολογισμένη αναπαράσταση (calculated activation) h_t , 2) η θύρα ενημέρωσης (update gate) z_t , 3) η θύρα επαναφοράς (reset gate) r_t και τέλος 4) η υποψήφια αναπαράσταση (candidate activation) \tilde{h}_t . Η θύρα ενημέρωσης ελέγχει το βαθμό που η μονάδα θα ενημερώσει το περιεχόμενο της με βάση την καινούργια υποψήφια αναπαράσταση. Επίσης, η θύρα επαναφοράς ελέγχει την κλίμακα κατά την οποία θα χρησιμοποιηθεί πληροφορία της προηγούμενης χρονικής στιγμής από αυτή που εξετάζουμε.

Οι εξισώσεις που διέπουν τη λειτουργία αυτής της μονάδας είναι οι εξής:

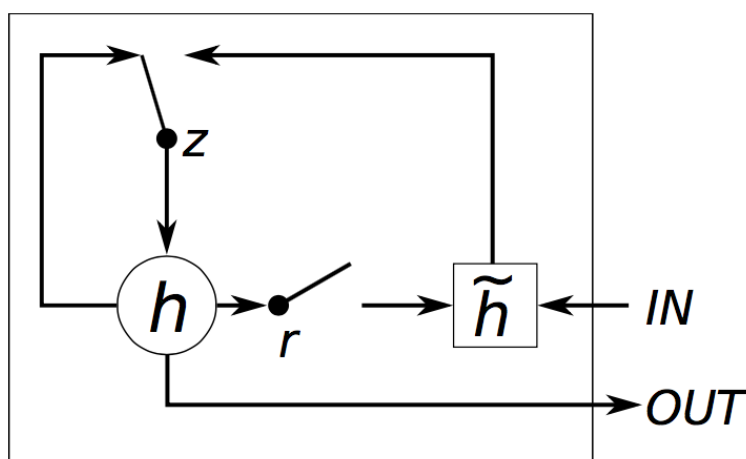
$$h_t = z_t \tilde{h}_t + (1 - z_t) * \tilde{h}_{t-1} \quad (4.13)$$

$$z_t = \sigma(W_z * x_t + U_z h_{t-1}) \quad (4.14)$$

$$\tilde{h}_t = \tanh(Wx_t + U(h_{t-1} \odot r_t)) \quad (4.15)$$

$$r_t = \sigma(W_r * x_t + U_r h_{t-1}) \quad (4.16)$$

Όπως στη μονάδα LSTM έτσι και στη GRU το x_t αποτελεί το διάνυσμα εισόδου και το \odot που δηλώνει γινόμενο Hadamard. Η διάσταση d δηλώνει τη διάσταση μνήμης του μονάδας. Από τις παραπάνω εξισώσεις καθίσταται φανερό ότι η υπολογισμένη αναπαράσταση είναι ο γραμμικός συνδυασμός μεταξύ της υπολογισμένης αναπαράστασης της προηγούμενης χρονικής στιγμής ($t-1$) καθώς και της υποψήφιας αναπαράστασης της χρονικής στιγμής t .



Σχήμα 4.6: Σχηματική απεικόνιση της δομής μια μονάδας Φραγμένης Επαναλαμβανόμενης Μονάδας - Gated Recurrent Unit

Με χρήση των προαναφερθέντων μνημών είναι δυνατό να δομηθούν εξελιγμένες και πολυπλοκότερες αρχιτεκτονικές νευρωνικών δικτύων. Η χρήση των μονάδων LSTM στα αναδρομικά δίκτυα παρουσιάζεται παρακάτω, ενώ η χρήση των μονάδων GRU για την ανάπτυξη αρχιτεκτονικών αναδρομικών νευρωνικών δικτύων παρουσιάζεται στο Κεφάλαιο 4 καθώς αποτελεί μέρος της παρούσας διπλωματικής εργασίας.

Τα αναδρομικά νευρωνικά δίκτυα ,γνωστά ως Tree-LSTM, είναι η τελευταία προσθήκη στις αρχιτεκτονικές των αναδρομικών δικτύων. Όπως παρουσιάστηκε στην

ενότητα 3.1, η τοπολογική τους μορφή είναι αυτή ενός δυαδικού δέντρου αφού εφαρμόζονται κυρίως πάνω στα συντακτικά δέντρα μια πρότασης ή ενός κομματιού κειμένου. Επομένως, και σε αυτή την αναδρομική αρχιτεκτονική διατηρείται η ιδιότητα της ιεραρχικής αναπαράστασης. Πιο συγκεκριμένα, για κάθε κόμβο j αναπαριστούμε την κρυφή κατάσταση και το κύτταρο μνήμης του k παιδιού ως h_{jk} και c_{jk} αντίστοιχα. Οι εξισώσεις μετάβασης τώρα διαμορφώνονται ως εξής:

$$i_j = \sigma(W^i x_j + \sum_{l=1}^N U_l^{(i)} h_{jl} + b^{(i)}) \quad (4.17)$$

$$f_{jk} = \sigma(W^{(f)} x_j + \sum_{l=1}^N U_{kl}^{(f)} h_{jl} + b^{(f)}) \quad (4.18)$$

$$o_j = \sigma(W^{(o)} x_j + \sum_{l=1}^N U_l^{(o)} h_{jl} + b^{(o)}) \quad (4.19)$$

$$u_j = \tanh(W^u x_t + U^u h_{t-1} + b^u) \quad (4.20)$$

$$c_j = i_j \odot u_j + \sum_{l=1}^N f_{jl} \odot c_{jl} \quad (4.21)$$

$$h_j = o_j \odot \tanh(c_j) \quad (4.22)$$

Η εισαγωγή διαφορετικού πίνακα βαρών για κάθε παιδί k επιτρέπει στο δίκτυο TreeLSTM να μάθει πιο λεπτές συνθήκες σχετικά με τη κατάσταση των μονάδων των παιδιών. Για παράδειγμα, σε ένα συντακτικό δέντρο όπου το αριστερό παιδί ενός κόμβου είναι μια ονομαστική φράση και το δεξί παιδί αντιστοιχεί σε μια ρηματική φράση. Σε αυτή την περίπτωση είναι προς όφελος του δικτύου να δώσει έμφαση στην ρηματική φράση κατά την κατασκευή του δικτύου. Για να επιτευχθεί αυτό, οι παράμετροι $U_{kl}^{(f)}$ μπορούν να εκπαιδευτούν με τέτοιο τρόπο ώστε το περιεχόμενο της f_{j1} να είναι κοντά στο μηδέν, δηλαδή να ξεχαστεί, ενώ το περιεχόμενο της f_{j2} να είναι κοντά στο 1, δηλαδή να διατηρηθεί. Στην εξίσωση 3.18, έχει εισαχθεί μια παραμετροποίηση για τη θύρα λησμόνησης f_{jk} του k -ιστού παιδιού. Αυτή η παραμετροποίηση επιτρέπει μεγαλύτερη ευελιξία στην ροή πληροφορίας από τους κόμβους παιδιά στον πατέρα.

4.2 Εκπαίδευση Αναδρομικών Δικτύων

Η εκπαίδευση των Αναδρομικών Νευρωνικών Δικτύων επιτυγχάνεται με τους αλγόριθμους που παρουσιάστηκαν στην ενότητα 2.3 και δεν διαφέρει καθόλου από το τρόπο που εκπαιδεύονται τα υπόλοιπα νευρωνικά δίκτυα.

Ο υπολογισμός χωρίζεται στις διακριτές φάσεις της πρόσθιας διάδοσης και της πίσω διάδοσης σφάλματος. Ο στόχος πάντα είναι ο υπολογισμός των κλίσεων και η ενημέρωση των παραμέτρων των αναδρομικών νευρωνικών δικτύων προκειμένου να μειωθεί το κόστος που υπολογίζεται με κάποια συνάρτηση κόστους. Οι πλέον σύνηθες συναρτήσεις κόστους αποτελούν η διεντροπία και το μέσο τετραγωνικό σφάλμα.

Παρόλα αυτά, η διαδικασία της εκπαίδευσης αναδρομικών νευρωνικών δικτύων καθίσταται δύσκολη κάποιες φορές λόγω των κλίσεων που 'εξαφανίζονται' ή αποκτούν πολύ μεγάλες τιμές (vanishing or exploding gradients). Το πρόβλημα αυτό είναι το πλέον σύνηθες κατά την διάρκεια εκπαίδευσης αναδρομικών, επαναλαμβανόμενων και πολύ μεγάλων δικτύων. Για το λόγο αυτό, προτάθηκαν τα αναδρομικά δίκτυα με μνήμη, επιλύοντας έτσι το πρόβλημα που οδηγεί τα δίκτυα σε κακή ή μέτρια επίδοση όσον αφορά τα προβλήματα που καλούνται να λύσουν.

4.3 Χρήση Αναδρομικών Δικτύων στην Επεξεργασία Φυσικής Γλώσσας

Η χρήση των Αναδρομικών Δικτύων στην Επεξεργασία Φυσικής Γλώσσας γίνεται εφικτή με τη χρήση των αλγορίθμων Glove και Word2Vec που αναφέρονται στην επόμενη ενότητα. Η ανάπτυξη κατανεμημένων αναπαραστάσεων για λέξεις σε συνδυασμό με τα αναδρομικά δίκτυα, έχει βελτιώσει κατά πολύ την επίδοση των νευρωνικών δικτύων σε διάφορα προβλήματα φυσικής γλώσσας όπως είναι η ανάλυση συναισθήματος, η συντακτική ανάλυση και η μοντελοποίηση γλώσσας.

4.3.1 Αλγόριθμοι Glove και Word2vec

Τα προβλήματα φυσικής γλώσσας δομούνται πάνω στο πρόβλημα αλληλεπίδρασης των λέξεων μεταξύ τους. Επομένως, καθίσταται αναγκαία η εύρεση τρόπου αναπαράστασης των λέξεων σε ένα σύστημα, προκειμένου να γίνει η επεξεργασία πάνω στη σημασία μια πρότασης, την ομοιότητα των λέξεων ή φράσεων και άλλα.

Ο πιο απλός τρόπος είναι η χρήση μια διακριτής αναπαράστασης όπου κάθε λέξη συμβολίζεται ως ένα μεγάλο διάνυσμα με 1 στην διάσταση που αντιστοιχεί στη λέξη και 0 στις υπόλοιπες. Έτσι για να συμβολίσει κανείς όλες τις λέξεις μια συλλογής δεδομένων που έχει 1 εκατομμύριο λέξεις θα χρειάζονται 1 εκατομμύριο διανύσματα με διάσταση του ίση με 1 εκατομμύριο. Η συγκεκριμένη μεθοδολογία ονομάζεται one-hot representation. Πέρα από το απαγορευτικές απαιτήσεις για την αναπαράσταση των λέξεων, η συγκεκριμένη μεθοδολογία δεν λαμβάνει υπόψιν το περιεχόμενο μέσα στο οποίο βρίσκεται κάποια λέξη καθώς δεν υπάρχει πληροφορία για το ποιες είναι οι λέξεις που ακολουθούν ή οι προηγούμενες λέξεις.

Ένας άλλος τρόπος είναι η διακριτή αναπαράσταση των λέξεων ως κόμβους σε ένα γράφο. Τέτοιοι γράφοι έχουν προταθεί αρκετοί στη βιβλιογραφία Wordnet, Sentinet και άλλοι. Το πρόβλημα όμως που προκύπτουν από τέτοιες μεθοδολογίες είναι η έλλειψη συσχετίσεων μεταξύ των λέξεων, η έλλειψη κάποιων λέξεων, η υποκειμενικότητα κατά την ανάπτυξη του γράφου. Παρόλα αυτά, το σπουδαιότερο πρόβλημα είναι ότι η συγκεκριμένη μεθοδολογία απαιτεί την ύπαρξη ανθρώπων που θα αναπτύξουν και θα συντηρήσουν μια τέτοια βάση δεδομένων σε βάθος χρόνων καθώς η γλώσσα θεωρείται ως ζωντανή ύπαρξη, δηλαδή με την πάροδο των χρόνων προσαρμόζεται, τροποποιείται και εξελίσσεται. Επίσης, εύκολα διαφαίνεται ότι η δημιουργία διακριτών αναπαραστάσεων δεν μπορεί να γίνει για το πλήθος των γλωσσών που υπάρχουν στον κόσμο καθώς το κόστος και οι εργατοώρες ενός τέτοιου εγχειρήματος θα ήταν απαγορευτικό.

Μια από τις πιο πετυχημένες ιδέες στη στατιστική επεξεργασία φυσικής γλώσσας είναι η χρήση των γειτονικών λέξεων για την αναπαράσταση μιας λέξης. Με αυτό τον τρόπο, μπορούν να δημιουργηθούν κατανομημένες αναπαραστάσεις που κωδικοποιούνται ως πυκνά διανύσματα που μπορούν να προβλέψουν άλλες λέξεις που εμφανίζονται στο ίδιο περιεχόμενο. Bengio et al. εισήγαγαν τον όρο αναπαράσταση

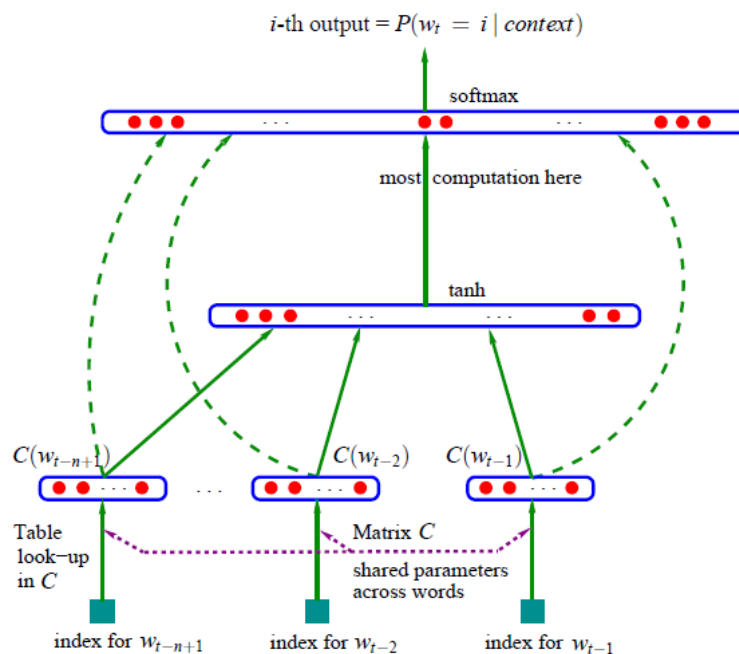
λέξης (word embeddings) το 2003 και τις εκπαίδευσαν μαζί με τις παραμέτρους ενός νευρωνικού μοντέλου για κωδικοποίηση γλώσσας. Οι πρώτοι που έδειξαν την χρήση προ-εκπαιδευμένων αναπαραστάσεων λέξεων ήταν οι Collobert και Weston το 2008. Η δημοσίευσή τους "A unified architecture for natural language processing" όχι μόνο παρουσιάζει τις αναπαραστάσεις λέξεων ως χρήσιμο εργαλείο για επιμέρους προβλήματα αλλά παρουσιάζουν και μια αρχιτεκτονική νευρωνικού δικτύου που θέτει την βάση για πολλές τωρινές προσεγγίσεις. Παρόλα αυτά, η μεγάλη ανάπτυξη για τις αναπαραστάσεις λέξεων επήλθε από τον Mikolov et al. το 2013 όταν δημιουργήθηκε το εργαλείο word2vec που μπορεί εύκολα να εκπαιδεύει τις αναπαραστάσεις των λέξεων. Το 2014, ο Pennington et al. δημοσίευσαν το Glove με μια συλλογή από προ-εκπαιδευμένες αναπαραστάσεις λέξεων. Αξίζει να σημειωθεί ότι οι αναπαραστάσεις λέξεων είναι από τις λίγες επιτυχημένες εφαρμογές της μη-εποπτευόμενης μάθησης. Το κύριο πλεονέκτημα τους είναι ότι δεν χρειάζονται τον ανθρώπινο παράγοντα όπως οι γράφοι αλλά μπορούν να εξαχθούν μέσα από μεγάλες συλλογές κειμένων όπως είναι το Wikipedia.

Κάθε εμπρόσθιο νευρωνικό δίκτυο που δέχεται λέξεις από ένα λεξιλόγιο σαν είσοδο και τις ενθέτει σαν διανύσματα σε ένα διανυσματικό χώρο μικρής διάστασης, ο οποίος μετά εκπαιδεύεται με βάση τον αλγόριθμο πίσω διάδοσης σφάλματος. Το επίπεδο που ενθέτει το δίκτυο τις λέξεις στο διανυσματικό χώρο μικρής διάστασης ονομάζεται Embedding Layer.

Η βασική διαφορά μεταξύ των δικτύων, που παράγουν τις ενθέσεις λέξεων word embeddings ως πάρα προϊόν, και των μεθόδου όπως word2vec, των οποίων ο βασικό στόχος είναι η δημιουργία των word embeddings, είναι η υπολογιστική πολυπλοκότητα. Για να παραχθούν τα word embeddings με κάποια βαθιά αρχιτεκτονική νευρωνικού δικτύου είναι υπολογιστικά ανέφικτο για πολύ μεγάλο λεξιλόγιο. Μια άλλη βασική διαφορά είναι ο στόχος της εκπαίδευσης, το Word2Vec και Glove είναι στοχευμένα στο να δημιουργούν word embeddings που κωδικοποιούν γενικές σημασιολογικές σχέσεις που είναι χρήσιμες για μικρότερα προβλήματα. Παρόλα αυτά, τα word embeddings που παράγονται δεν θα βοηθούσαν σε κάποιο πρόβλημα που δεν έχει συνάφεια με σημασιολογία. Σε αντίθεση, τα τυπικά νευρωνικά δίκτυα παράγουν word embeddings που έχουν άμεση συσχέτιση με το πρόβλημα που επιλύουν αλλά δεν είναι χρήσιμα σε προβλήματα.

Για το μαθηματικό φορμαλισμό υποθέτουμε μια συλλογή δεδομένων για εκπαίδευση με προτάσεις μήκους T και λέξεις $w_1, w_2, w_3, \dots, w_T$ που ανήκουν σε ένα λεξιλόγιο V με μέγεθος $|V|$. Συνήθως, χρησιμοποιείται περιεχόμενο μήκους n λέξεων και συσχετίζουμε κάθε λέξη με μια είσοδο v_w με διάσταση d και έξοδο v'_w . Το τελευταίο στάδιο είναι η βελτιστοποίηση μια συνάρτησης J_θ σε σχέση με τις παραμέτρους θ του μοντέλου και κάποιου σκορ εξόδου του μοντέλου $f_\theta(x)$ για κάθε είσοδο x .

Η πρώτη προσπάθεια για δημιουργία word embeddings έγινε από τον Bengio το 2003 όταν δημιούργησε ένα νευρωνικό μοντέλο για το πρόβλημα της μοντελοποίησης γλώσσας. Το μοντέλο αυτό αποτελείται από ένα εμπρόσθιο νευρωνικό δίκτυο ενός επιπέδου.



Σχήμα 4.7: Κλασικό μοντέλο για μοντελοποίησης γλώσσας

$$J_\theta = \frac{1}{T} \sum_{t=1}^T \log f(w_t, w_{t-1}, \dots, w_{t-n+1}) \quad (4.23)$$

όπου $f(w_t, w_{t-1}, \dots, w_{t-n+1})$ είναι η έξοδος του συστήματος, δηλαδή η πιθανότητα $p(w_t | w_{t-1}, \dots, w_{t-n+1})$ όπως αυτή υπολογίστηκε από τη συνάρτηση *softmax* με n να συμβολίζει τον αριθμό των προηγούμενων λέξεων που δόθηκαν στο μοντέλο.

Ο Bengio είναι ο πρώτος που δημιούργησε τη θεωρία αυτού που σήμερα αποκαλούμε

word embeddings, ένα διάνυσμα με πραγματικές τιμές. Οι βασικοί δομικοί λίθοι του συστήματος είναι οι ίδιοι που βρίσκονται και σε state-of-the art συστήματα στις μέρες μας. Αυτοί είναι:

- **Embedding Επίπεδο:**
ένα επίπεδο που δημιουργεί word embeddings πολλαπλασιάζοντας ένα διάνυσμα δείκτη με τον πίνακα των word embeddings.
- **Μεσαία επίπεδα:**
ένα ή περισσότερα επίπεδα που δημιουργούν μια αναπαράσταση για την είσοδο.
- **Επίπεδο Softmax:**
Ένα επίπεδο που δημιουργείται η πιθανότητα για ένα σύνολο λέξεων.

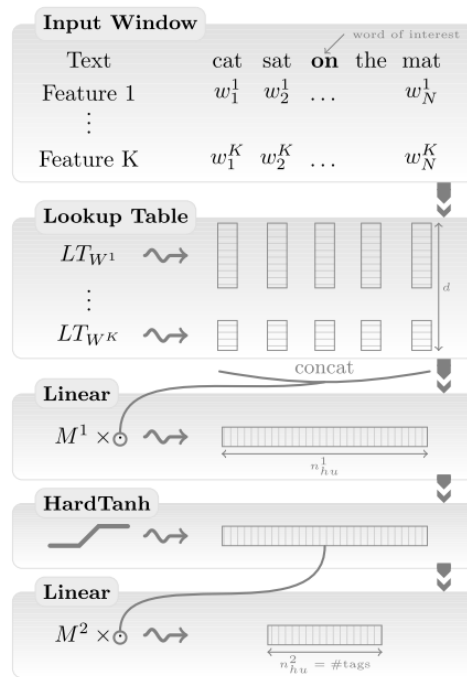
Μετά τις πρώτες απόπειρες του Bengio για το νευρωνικό γλωσσικό μοντέλο, η έρευνα πάνω στα word embeddings κόλλησε καθώς οι αλγόριθμοι και η υπολογιστική δύναμη δεν επέτρεπε την εκπαίδευση για μεγάλα λεξιλόγια.

Οι Collobert και Weston το 2008 έδειξαν ότι τα word embeddings όταν εκπαιδεύονται πάνω σε μια μεγάλη συλλογή δεδομένων μπορούν να κωδικοποιήσουν σημασιολογική και συντακτική πληροφορία, η οποία αποδεικνύουν ότι είναι και πολύ χρήσιμη σε αρκετά προβλήματα όπως η ανάλυση συναισθήματος. Η λύση τους για το πρόβλημα της υπολογιστικής πολυπλοκότητας είναι η αλλαγή της χρήσης της συνάρτησης εντροπίας. Στη θέση της χρησιμοποίησαν μια συνάρτηση που κατατάσσει ανάμεσα σε ζευγάρια:

$$J_{\theta} = \sum_{x \in X} \sum_{w \in V} \max\{0, 1 - f_{\theta}(x) + f_{\theta}(x^{(w)})\} \quad (4.24)$$

Το μοντέλο μοιράζει σε παράθυρα x που περιέχουν n λέξεις από το σύνολο X παραθύρων της συλλογής δεδομένων. Για κάθε παράθυρο x δημιουργείται μια λανθασμένη έκδοση $x^{(w)}$ αντικαθιστώντας της κεντρική λέξη του x με μια λέξη w από το V . Η συνάρτηση κόστους τώρα μεγιστοποιεί την απόσταση μεταξύ των σκορ που δίνει το μοντέλο για το σωστό και λανθασμένο παράθυρο. Η αρχιτεκτονική αναπαριστάται στην εικόνα 3.8.

Η γλωσσικό μοντέλο που παράγεται διαθέτει πολλές ιδιότητες των word embeddings για τις οποίες έγιναν γνωστά. Για παράδειγμα, οι χώρες ανήκουν στην ίδια γει-



Σχήμα 4.8: Το μοντέλο C & W

τονιά και συντακτικά όμοιες λέξεις κατέχουν παρόμοιες τοποθεσίες στο διανυσματικό χώρο.

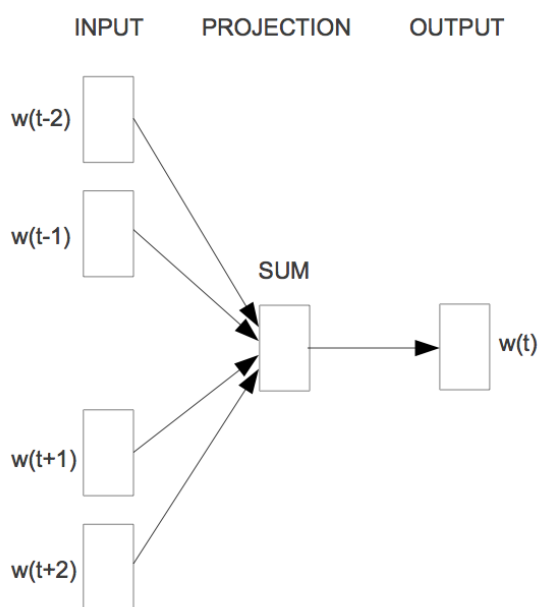
Στη συνέχεια εισήχθη ο αλγόριθμος Word2Vec που είναι ο πιο διάσημος αλγόριθμος για word embeddings. Καθώς τα word embeddings αποτελούν δομικό λίθο πολλών συστημάτων βαθιάς μηχανικής μάθησης, θεωρούνται ότι ανήκουν στον ίδιο τομέα. Τεχνικά, το Word2Vec δεν είναι ούτε σύστημα βαθιάς μηχανικής μάθησης ούτε χρησιμοποιεί μη-γραμμικές συναρτήσεις, σε αντίθεση με τα προηγούμενα συστήματα.

Στην πρώτη δημοσίευση ο Mikolov προτείνει δυο αρχιτεκτονικές για μάθηση των word embeddings οι οποίες υπολογιστικά εφικτές. Στη δεύτερη δημοσίευση, οι ιδέες αυτές εμπλουτίζονται με επιπλέον μεθοδολογίες και επιτυγχάνουν ακόμα καλύτερη επίδοση και ταχύτητα υπολογισμών.

Αυτές οι αρχιτεκτονικές έχουν δυο βασικά πλεονεκτήματα σε σχέση με τα προηγούμενα γλωσσικά μοντέλα.

- Δεν χρησιμοποιούν κρυφά επίπεδα
- Χρησιμοποιούν το περιεχόμενο μια λέξης

Η πρώτη αρχιτεκτονική ονομάζεται Continuous bag-of-words (CBOW). Ενώ τα γλωσσικά μοντέλα είναι σε θέση να χρησιμοποιήσουν μόνο τις προηγούμενες λέξεις για τις προβλέψεις, ένα μοντέλο που είναι συσχετισμένο με word embeddings δεν έχει αυτό τον περιορισμό. Για το λόγο αυτό, χρησιμοποιούνται οι πρώτες και επόμενες n λέξεις της λέξης στόχου w_t , όπως φαίνεται στο σχήμα. Το μοντέλο αυτό ονομάζεται συνεχές τσουβάλι από λέξεις (Continuous bag-of-words (CBOW)) γιατί χρησιμοποιεί συνεχής αναπαραστάσεις των οποίων η σειρά δεν έχει σημασία.



Σχήμα 4.9: Continuous bag-of-words

Η συνάρτηση προς βελτιστοποίηση του CBOW είναι λίγο διαφορετική από αυτή του γλωσσικού μοντέλου

$$J_{\theta} = \frac{1}{T} \sum_{t=1}^T \log p(w_t | w_{t-n}, \dots, w_{t-1}, w_{t+1}, \dots, w_{t+n}) \quad (4.25)$$

Αντί να περνάμε τις προηγούμενες n λέξεις στο μοντέλο, το μοντέλο δέχεται ένα παράθυρο από n λέξεις γύρω της λέξης στόχου w_t για κάθε χρονικό βήμα t .

Η δεύτερη αρχιτεκτονική είναι το Skip-Gram. Ενώ το CBOW μπορεί να θεωρηθεί σαν μοντελοποίηση γλώσσας με πρότερη πληροφορία, το skip-gram μεταβάλλει τη λογική της μοντελοποίησης γλώσσας. Αντί να χρησιμοποιούνται οι γειτονικές λέξεις για να προβλεφθεί η κεντρική λέξη, όπως στο CBOW, το skip-gram χρησιμοποιεί

την κεντρική λέξη για να προβλέψει τις γειτονικές λέξεις. Η λογική του skip-gram εμφανίζεται στην σχηματικό.

Ο στόχος του skip-gram είναι να αθροίσει το λογάριθμο των πιθανοτήτων των γειτονικών n λέξεων από τα αριστερά και τα δεξιά της λέξης στόχου w_t με σκοπό να βελτιστοποιήσει την εξής συνάρτηση κόστους:

$$J_{\theta} = \frac{1}{T} \sum_{t=1}^T \sum_{-n \leq j \leq n, j \neq 0} \log p(w_{t+j} | w_t) \quad (4.26)$$

Η αλλαγή που εισαγάγεται με το μοντέλο skip-gram έχει να κάνει με τον τρόπο που υπολογίζεται η πιθανότητα $p(w_{t+j} | w_t)$. Κανονικά η πιθανότητα στα γλωσσικά μοντέλα υπολογίζεται με χρήση της συνάρτησης softmax ως εξής:

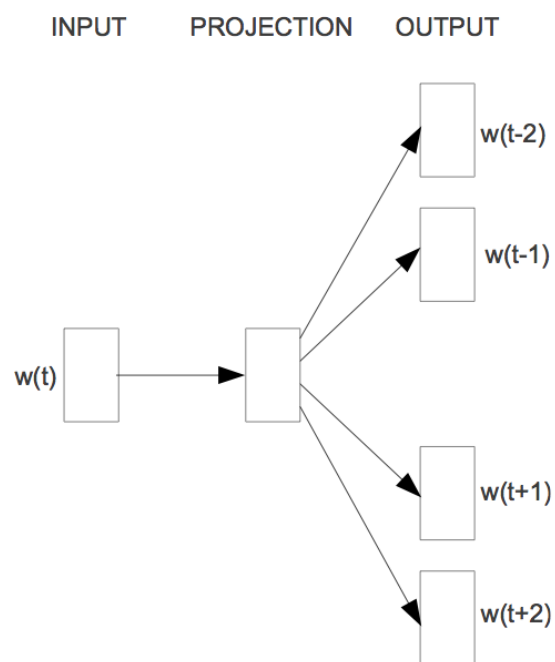
$$p(w_t | w_{t-1}, \dots, w_{t-n+1}) = \frac{\exp(h^{\top} v'_{w_t})}{\sum_{w_i \in V} \exp(h^{\top} v'_{w_i})} \quad (4.27)$$

Αντί να υπολογίζεται η πιθανότητα της λέξης στόχος w_t δοθέντος της προηγούμενης λέξης, το skip-gram υπολογίζει την πιθανότητα της γειτονικής λέξης w_{t+j} δοθέντος της w_t λέξης. Επομένως η εξίσωση μεταβάλλεται ως:

$$p(w_{t+j} | w_t) = \frac{\text{εξπ}(v_{w_t}^{\top} v'_{w_{t+j}})}{\sum_{w_i \in V} \text{εξπ}(v_{w_t}^{\top} v'_{w_i})} \quad (4.28)$$

Στο πλαίσιο αυτό, βασίστηκαν οι ερευνητές για την ανάπτυξη των αλγορίθμων Word2vec και Glove, για τις ανάγκες της πτυχιακής εργασίας θα περιγραφεί ο αλγόριθμος Word2vec καθώς οι δυο αλγόριθμοι έχουν ίδιο θεωρητικό κορμό και οι διαφορές τους προκύπτουν σε φορμαλισμό υψηλότερου επιπέδου.

Η βασική ιδέα του αλγόριθμου Word2vec είναι η πρόβλεψη μεταξύ κάθε λέξης και των λέξεων που μπορούν να βρεθούν μαζί στο περιεχόμενο. Οι δύο αλγόριθμοι που μπορούν να μοντελοποιήσουν το συγκεκριμένο πρόβλημα είναι ο Skip-Gram, ο οποίος προβλέπει τις λέξεις περιεχομένου δεδομένου μια λέξης στόχου, και ο Continuous Bag of Words που προβλέπει τη λέξη στόχο από το περιεχόμενο όταν αυτό συμβολίζεται ως συλλογή από λέξεις (bag-of-words).



Σχήμα 4.10: Skip-Gram

Κεφάλαιο 5

Σχεδίαση Αναδρομικού Νευρωνικού Δικτύου

Στην κεφάλαιο αυτή παρουσιάζονται οι αρχιτεκτονικές αναδρομικών νευρωνικών δικτύων που αναπτύχθηκαν στο πλαίσιο της διπλωματικής εργασίας. Στην αρχή ερευνάμε ένα απλό αναδρομικό νευρωνικό δίκτυο με μνήμη GRU και στη συνέχεια αναπτύσσουμε ένα νευρωνικό δίκτυο δύο κατευθύνσεων. Τέλος, εισάγεται η έννοια του μηχανισμού προσοχής πάνω σε δομή.

5.1 Σχεδίαση Αναδρομικού Δικτύου

5.1.1 Αναδρομικό Νευρωνικό δίκτυο με GRU

Τα αναδρομικά νευρωνικά δίκτυα με μνήμη GRU (Recursive GRU networks) πάνω σε δενδρικές δομές (TreeGRU) είναι η γενίκευση των ακολουθιακών επαναλαμβανόμενων δικτύων με μνήμες GRU (Recurrent GRU networks) και επιτρέπουν την διάδοση πληροφορίας μέσω της τοπολογίας του δικτύου.

Όμοια με τα αναδρομικά νευρωνικά δίκτυα με μνήμη LSTM πάνω σε δενδρικές δομές (TreeLSTM), για κάθε κόμβο του δέντρου, το TreeGRU έχει ένα μηχανισμό που ελέγχει τη ροή πληροφορίας μέσα στη μονάδα GRU χωρίς την ανάγκη για ξεχωριστό κύτταρο μνήμης. Η ενεργοποίηση h_j του TreeGRU για κάθε κόμβο j του δέντρου είναι ο γραμμικός συνδυασμός των προηγούμενων υπολογισμένων ενεργοποιήσεων h_{jk} του

k -οστού παιδιού-κόμβου από σύνολο N παιδιών και της υποψήφιας ενεργοποίησης \tilde{h}_j .

$$h_j = z_j * \sum_{k=1}^N h_{jk} + (1 - z_j) * \tilde{h}_j \quad (5.1)$$

όπου z_j είναι η συνάρτηση ενημέρωσης η οποία αποφασίζει το βαθμού της ενημέρωσης που θα συμβεί στη ενεργοποίηση με βάση το διάνυσμα εισόδου x_j και της προηγούμενης υπολογισμένης αναπαράστασης h_{jk} .

$$z_j = \sigma(U_z * x_j + \sum_{k=1}^N W_z^i * h_{jk}) \quad (5.2)$$

Η υποψήφια ενεργοποίηση \tilde{h}_j για ένα κόμβο j υπολογίζεται παρόμοια με την ενεργοποίηση του απλού αναδρομικού νευρωνικού δικτύου:

$$\tilde{h}_j = f(U_h * x_j + \sum_{k=1}^N W_h^k * (h_{jk} * r_j)) \quad (5.3)$$

όπου r_j είναι η θύρα λησμόνησης που επιτρέπει στο αναδρομικό δίκτυο να ξεχνά προηγούμενες υπολογισμένες αναπαραστάσεις όταν η τιμή είναι κοντά στο 0 και υπολογίζεται ως εξής:

$$r_j = \sigma(U_r * x_j + \sum_{k=1}^N W_r^k * h_{jk}) \quad (5.4)$$

Κάθε κομμάτι ενός Επαναλαμβανόμενο Δικτύου με Θύρες (Gate Recurrent Unit - GRU) $x_j, h_j, r_j, z_j, \tilde{h}_j \in \mathbb{R}^d$ όπου d είναι η διάσταση της διανύσματος εισόδου. σ είναι η σιγμοειδής συνάρτηση και f είναι η μη-γραμμική συνάρτηση της υπερβολικής εφαπτομένης (\tanh). Το σύνολο πινάκων $W^k, U \in \mathbb{R}^{d \times d}$ που χρησιμοποιείται στις παραπάνω εξισώσεις αποτελούν τις παραμέτρους βαρών προς εκπαίδευση που ενώνουν την k -ιοστή αναπαράσταση του παιδιού-κόμβου με τη j -ιοστή αναπαράσταση και του διανύσματος εισόδου x_j .

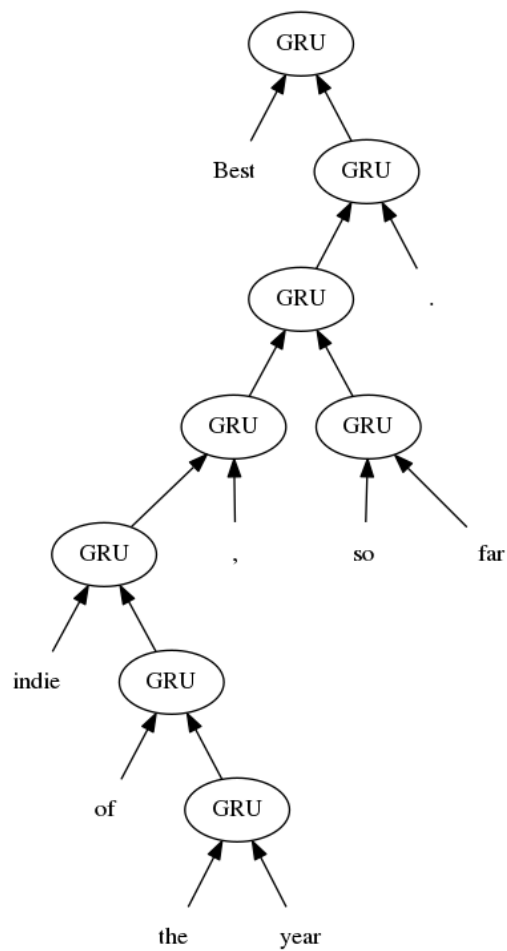
5.1.2 Αμφίδρομο Αναδρομικό Νευρωνικό δίκτυο με GRU

Μια φυσική επέκταση του Tree-Structured GRU είναι η προσθήκη μιας αμφίδρομης προσέγγισης. Το TreeGRU υπολογίζει την ενεργοποίηση του κόμβου j με τη χρήση των προηγούμενων υπολογισμένων ενεργοποιήσεων που βρίσκονται κάτω από

το κόμβο που εξετάζονται. Με άλλα λόγια το TreeGru για τον υπολογισμό της ενεργοποίησης ενός κόμβου χρησιμοποιεί πληροφορία που βρίσκεται στο δέντρο που έχει ως ρίζα τον κόμβο υπό εξέταση. Η αμφίδρομη προσέγγιση για μια δενδρική δομή χρησιμοποιεί πληροφορία που υπάρχει γύρω από έναν συγκεκριμένο κόμβο j , δηλαδή και σε υψηλότερους και χαμηλότερους κόμβους. Διαισθητικά, μια υπολογισμένη ενεργοποίηση περιλαμβάνει γνωσιακό περιεχόμενο από τους κόμβους-παιδιά καθώς και τον κόμβο-πατέρα, δηλαδή κάθε κόμβος χρησιμοποιεί πληροφορία που βρίσκεται γύρω του και όχι μόνο από κάτω του.

Το αμφίδρομο αναδρομικό δίκτυο μπορεί να χωριστεί σε δυο φάσεις: 1) η Ανοδική φάση (Upward phase) και 2) η Καθοδική φάση (Downward phase). Κατά την ανοδική φάση, η τοπολογία του νευρωνικού δικτύου είναι ίδια με αυτή του δικτύου TreeGRU, όπου κάθε ενεργοποίηση ή αλλιώς εσωτερική αναπαράσταση υπολογίζεται βασισμένη σε αναπαραστάσεις που υπολογίστηκαν προηγουμένως και βρίσκονται σε μεγαλύτερο βάθος από τον κόμβο που εξετάζουμε. Η μέθοδος αυτή είναι γνωστή στην Επιστήμη των Υπολογιστών ως μεταδιατεταγμένη διάσχιση (post-order), δηλαδή για κάθε κόμβο επισκεπτόμαστε πρώτα τους κόμβους του αριστερού του υποδένδρου, έπειτα τους κόμβους του δεξιού του υποδένδρου και στη συνέχεια τον ίδιο τον κόμβο.

Μετά το πέρας της ανοδικής φάσης, το νευρωνικό δίκτυο έχει υπολογίσει όλες τις εσωτερικές αναπαραστάσεις του δέντρου που αντιστοιχούν μια για κάθε κόμβο. Η αναπαράσταση της ρίζας του δέντρου χρησιμοποιείται ως είσοδος για την καθοδική φάση (Downward phase). Κατά τη διάρκεια της καθοδικής φάσης, το νευρωνικό δίκτυο υπολογίζει την ενεργοποίηση για κάθε παιδί-κόμβο χρησιμοποιώντας πληροφορία από το κόμβο-πατέρα. Η μέθοδος αυτή είναι γνωστή στην Επιστήμη των Υπολογιστών ως προδιατεταγμένη διάσχιση (pre-order, δηλαδή για κάθε κόμβο, επισκεπτόμαστε πρώτα τον ίδιο τον κόμβο, έπειτα τους κόμβους του αριστερού του υποδένδρου και στη συνέχεια τους κόμβους του δεξιού του υποδένδρου. Η διαδικασία για τον υπολογισμό των εσωτερικών αναπαραστάσεων μεταξύ των δυο φάσεων είναι σειριακή, επομένως η καθοδική φάση θα ξεκινήσει αμέσως μετά τη ανοδική φάση.



Σχήμα 5.1: Σχηματική απεικόνιση της καθοδικής φάσης του αναδρομικού νευρωνικού δικτύου TreeBiGRU εφαρμοσμένο πάνω στο συντακτικό δέντρο της πρόταση "Best indie of the year, so far."

Η ανοδική ενεργοποίηση h_j^\uparrow , ομοίως με το TreeGRU, για τον κόμβο j είναι η γραμμική παρεμβολή των προηγούμενων υπολογισμένων ενεργοποιήσεων h_{jk}^\uparrow του κ-ιστού παιδιού από ένα σύνολο N παιδιών-κόμβων και της υποψήφιας ενεργοποίησης \tilde{h}_j^\uparrow .

$$h_j^\uparrow = z_j^\uparrow * \sum_{k=1}^N h_{jk}^\uparrow + (1 - z_j^\uparrow) * \tilde{h}_j^\uparrow \quad (5.5)$$

Η θύρα ενημέρωσης, θύρα λησμόνησης και η υποψήφια ενεργοποίηση υπολογίζονται ως εξής:

$$z_j^\uparrow = \sigma(U_z * x_j^\uparrow + \sum_{k=1}^N W_z^k * h_{jk}^\uparrow) \quad (5.6)$$

$$r_j^\uparrow = \sigma(U_r * x_j^\uparrow + \sum_{k=1}^N W_r^k * h_{jk}^\uparrow) \quad (5.7)$$

$$\tilde{h}_j^\uparrow = f(U_h * x_j^\uparrow + \sum_{k=1}^N W_h^k * (h_{jk}^\uparrow * r_j^\uparrow)) \quad (5.8)$$

Η καθοδική ενεργοποίηση h_j^\downarrow για τον κόμβο j είναι η παρεμβολή των προηγούμενων υπολογισμένων ενεργοποιήσεων $h_{p(j)}^\downarrow$, όπου η συνάρτηση p επιστρέφει το δείκτη του κόμβου πατέρα, και της υποψήφιας ενεργοποίησης \tilde{h}_j^\downarrow .

$$h_j^\downarrow = z_j^\downarrow * h_{p(j)}^\downarrow + (1 - z_j^\downarrow) * \tilde{h}_j^\downarrow \quad (5.9)$$

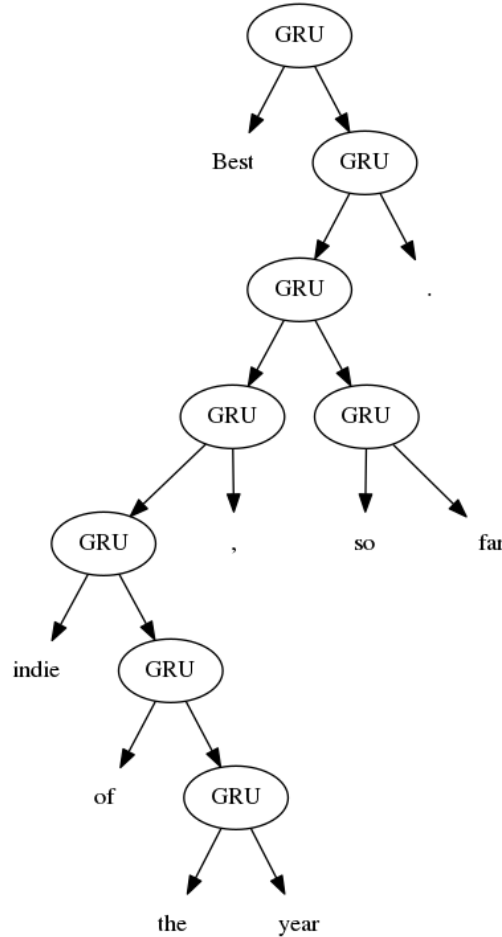
Η θύρα ενημέρωσης, θύρα λησμόνησης και η υποψήφια ενεργοποίηση για την καθοδική φάση υπολογίζονται ως εξής:

$$z_j^\downarrow = \sigma(U_z^d * h_j^\uparrow + W_z^d * h_{p(j)}^\downarrow) \quad (5.10)$$

$$r_j^\downarrow = \sigma(U_r^d * h_j^\uparrow + W_r^d * h_{p(j)}^\downarrow) \quad (5.11)$$

$$\tilde{h}_j^\downarrow = f(U_h^d * h_j^\uparrow + W_h^d * (h_{p(j)}^\downarrow * r_j^\downarrow)) \quad (5.12)$$

Κατά την καθοδική φάση, ο πίνακας $U^d \in \mathbb{R}^{d \times d}$ συνδέει την ανοδική αναπαράσταση του κόμβου j με την αντίστοιχη καθοδική αναπαράσταση, ενώ ο πίνακας $W^d \in \mathbb{R}^{d \times d}$ συνδέει την αναπαράσταση του πατέρα $p(j)$.



Σχήμα 5.2: Σχηματική απεικόνιση της ανοδικής φάσης του αναδρομικού νευρωνικού δικτύου TreeBiGRU εφαρμοσμένο πάνω στο συντακτικό δέντρο της πρότασης "Best indie of the year, so far."

5.1.3 Μηχανισμός προσοχής εφαρμοσμένος σε δομή

Ο μηχανισμός προσοχής εφαρμοσμένος σε δομή είναι μια γενίκευση του ακολουθιακού μοντέλου προσοχής και εξαγάγει του κόμβους με την περισσότερη πληροφορία από ένα συντακτικό δέντρο και στην συνέχεια συναθροίζει τις αναπαραστάσεις αυτών των κόμβων με σκοπό τη δημιουργία μια αναπαράστασής για τη ρίζα του δέντρου.

Η διαδικασία του μηχανισμού προσοχής ξεκινά με τροφοδότηση των κρυφών αναπαραστάσεων h_j κάθε κόμβου j σε ένα απλό νευρωνικό δίκτυο βάθους ένα με πίνακα βαρών $W_w \in \mathbb{R}^{d \times d}$ με σκοπό τη δημιουργία νέων κρυφών αναπαραστάσεων u_j .

$$u_j = \tanh(W_w * h_j) \quad (5.13)$$

Εν συνεχεία, χρησιμοποιώντας τη συνάρτηση softmax, παίρνουμε τα βάρη a_j για κάθε κόμβο βασισμένα στην ομοιότητα των κρυφών αναπαραστάσεων u_j και ενός καθολικού διανύσματος συμφραζομένων $u_w \in \mathbb{R}^d$. Τα κανονικοποιημένα βάρη a_j χρησιμοποιούνται για το σχηματισμό της αναπαράστασης της ρίζας $s \in \mathbb{R}^d$ και είναι η άθροιση με βάρη της αναπαράστασης κάθε κόμβου h_j .

$$a_j = \frac{u_j^\top * u_w}{\sum_{i=1}^N u_i^\top * u_w} \quad (5.14)$$

$$s = \sum_{i=1}^N a_i h_i \quad (5.15)$$

Η παραπάνω μέθοδος προσοχής εφαρμόζεται πάνω σε δομή αφού όλες οι αναπαραστάσεις κόμβων περιέχουν πληροφορία σχετικά με τη συντακτική δομή λόγω της αναδρομικής φύσης της συγκεκριμένης αρχιτεκτονικής νευρωνικών δικτύων. Η ιδιότητα της μεθόδου να εφαρμόζεται σε δομή της δίνει και το όνομα "Μηχανισμός προσοχής εφαρμοσμένος σε δομή".

Κεφάλαιο 6

Ανάλυση και Πειράματα

Στο κεφάλαιο αυτό περιγράφεται η υλοποίηση των νευρωνικών αρχιτεκτονικών, με βάση το φορμαλισμό που παρουσιάστηκε στο προηγούμενο κεφάλαιο. Αρχικά παρουσιάζεται η συλλογή δεδομένων για ανάλυση συναισθήματος και τα προγραμματιστικά εργαλεία που χρησιμοποιήθηκαν. Στη συνέχεια δίνονται οι λεπτομέρειες υλοποίησης για τους βασικούς αλγορίθμους. Τέλος παρουσιάζονται τα αποτελέσματα και ανάλυση των πειραμάτων και μια σύγκριση μεταξύ συστημάτων που αναφέρονται στη βιβλιογραφία.

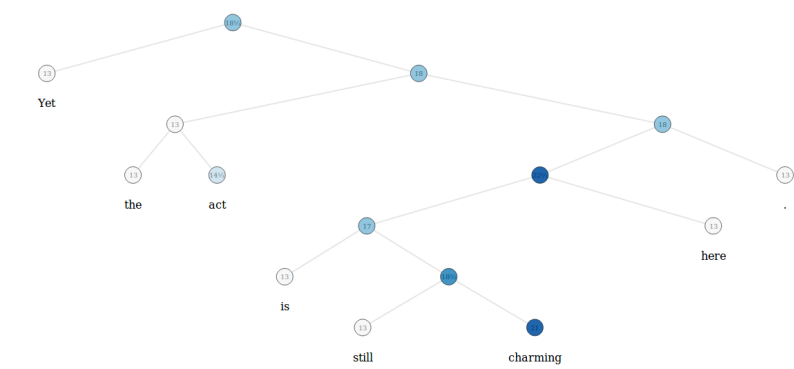
6.1 Δεδομένα Stanford Sentiment Treebank

Η συλλογή δεδομένων που χρησιμοποιήθηκε για την περάτωση όλων των πειραμάτων είναι η Stanford Sentiment Treebank που περιλαμβάνει ετικέτες (labels) για κάθε συντακτικά πιθανή φράση για χιλιάδες προτάσεις, επιτρέποντας με αυτό το τρόπο την εκπαίδευση και αξιολόγηση συνθετικών μοντέλων.

Τα δεδομένα προέρχονται από την ιστοσελίδα www.rottentomatoes.com που συλλέχθηκαν και δημοσιεύθηκαν αρχικά από τους Pang, Lee (2005). Η συλλογή δεδομένων περιέχει 10.662 προτάσεις από τις οποίες οι μισές θεωρούνται θετικές και οι άλλες μισές αρνητικές. Κάθε ετικέτα έχει παρθεί από μια μεγάλη κριτική για την ταινία και αντιπροσωπεύει την όλη πρόθεση του συγγραφέα για αυτή την κριτική. Ο Stanford Parser (Klein and Manning, 2003) χρησιμοποιείται για τη συντακτική ανάλυση των 10.662 προτάσεων. Η έξοδος που λαμβάνεται από τον Stanford Parser είναι το δυαδι-

κό δέντρο που αντιστοιχεί στην συντακτική δομή της πρότασης. Σε 1.100 περιπτώσεις τα κομμάτια κειμένου σπάνε σε επιμέρους προτάσεις. Μετά την προ-επεξεργασία των δεδομένων, δόθηκαν στην υπηρεσία Amazon Mechanical Turk οι προτάσεις με αποτέλεσμα τη λήψη ετικετών για 215.154 φράσεις.

Από την αρχική ανάλυση της συλλογής δεδομένων Stanford Sentiment Treebank βρέθηκε ότι οι περισσότερες προτάσεις μπορούν να θεωρηθούν ως ουδέτερες από την πλευρά του αναγνώστη. Επίσης, τα δυνατά συναισθήματα (πολύ αρνητικό, πολύ θετικό) συχνά χτίζονται σε μεγαλύτερες φράσεις και η πλειονότητα των σύντομων φράσεων είναι ουδέτερες. Άλλη παρατήρηση είναι ότι οι εξτρεμιστικές τιμές (πολύ αρνητικό, πολύ θετικό) εμφανίζονται σπάνια. Για τους λόγους αυτούς, ένα πρόβλημα ταξινόμησης με 5 κλάσεις στις κατηγορίες (πολύ αρνητικό, αρνητικό, ουδέτερο, θετικό, πολύ θετικό) είναι σε θέση να συλλάβει την μεταβλητότητα αυτών των ετικετών. Το πρόβλημα αυτό ονομάζεται *fine-grained sentiment classification* και τα παρακάτω πειράματα θα αποσκοπούν στην πρόβλεψη αυτών των πέντε ετικετών για όλες τις προτάσεις όλων των διαστάσεων.



Σχήμα 6.1: Σχηματική απεικόνιση ενός παραδείγματος από το Stanford Sentiment Treebank. Στην εικόνα παρουσιάζεται η συντακτική δομή και οι αντίστοιχες ετικέτες συναισθήματος για την πρόταση "Yet the act is still charming here."

Για την εκπαίδευση των αναδρομικών νευρωνικών αρχιτεκτονικών χρησιμοποιούμε το διαχωρισμό σε train/dev/test συλλογές που αποτελούνται από 6920/872/1821 για το δυαδικό πρόβλημα (θετικό, αρνητικό συναίσθημα) και 8544/1101/2210 για το *fine-grained* πρόβλημα. Για το δυαδικό πρόβλημα τα δεδομένα είναι λιγότερα καθώς προέρχονται από το *fine-grained* πρόβλημα με αφαίρεση όλων των προτάσεων που έχουν

ετικέτα διάφορη του θετικού και αρνητικού.

Πέρα από την αρχική προ-εξεργασία των δεδομένων, που έγινε από τους δημιουργούς, δεν έγινε κάποια περαιτέρω επεξεργασία.

6.2 Υλοποίηση των Αναδρομικών Δικτύων

Η υλοποίηση όλων των αρχιτεκτονικών αναδρομικών δικτύων TreeGRU, TreeBiGRU έγινε με χρήση της βιβλιοθήκης Theano και της γλώσσας προγραμματισμού Python. Η βιβλιοθήκη Theano επιτρέπει τον ορισμό, βελτιστοποίηση και επίλυση μαθηματικών εξισώσεων, ειδικά αυτών που βασίζονται σε πίνακες πολλαπλών διαστάσεων. Χρησιμοποιώντας τη βιβλιοθήκη Theano είναι εφικτό να επιτύχει κανείς ταχύτητα εκτέλεσης όμοια με προγράμματα που υλοποιήθηκαν εξαρχής σε γλώσσα C, γεγονός χρήσιμο για προβλήματα με μεγάλο όγκο δεδομένων. Επιπλέον, η βιβλιοθήκη επιτρέπει την περαιτέρω επιτάχυνση των υπολογισμών χρησιμοποιώντας τις τελευταίες τεχνολογίες των καρτών γραφικών GPU.

Οι κυριότερες δυσκολίες για την υλοποίηση νευρωνικών δικτύων είναι η υλοποίηση της εμπρόσθιας φάσης (Feed-Forward) και ο υπολογισμός των κλίσεων (Gradients). Η βιβλιοθήκη Theano παρέχει τη δυνατότητα αυτόματου υπολογισμού των κλίσεων με βάση την τοπολογία του γράφου υπολογισμού του εκάστοτε δικτύου, μειώνοντας έτσι πολύ το χρόνο υλοποίησης.

Για την υλοποίηση των δικτύων που περιγράφηκαν στο προηγούμενο κεφάλαιο χρειάστηκε να μοντελοποιηθεί η δενδρική δομή με τη μορφή πίνακα, καθώς οι κάρτες γραφικών δεν υποστηρίζουν άλλες δομές πέρα από πίνακες και διανύσματα. Για την μετατροπή των δενδρικών δομών σε πίνακες υλοποιήθηκε ένας αλγόριθμος όπου με διάσχιση κατά πλάτος δημιουργεί ένα πίνακα δεικτών με κάθε γραμμή να περιλαμβάνει το δείκτη του πατέρα, του αριστερού και τέλος του δεξιού παιδιού.

Η επόμενη δυσκολία που εμφανίστηκε ήταν η αδυναμία της βιβλιοθήκης Theano να υλοποιήσει αναδρομικές εξισώσεις. Σημειωτέον, ότι όλες οι βιβλιοθήκες παρόμοιας λογικής με το Theano αδυνατούν να υποστηρίξουν αναδρομικές εξισώσεις με αποτελεσματικό τρόπο. Η επίλυση του προβλήματος της αναδρομής έγινε με χρήση του αλγορίθμου μετατροπής αναδρομής από τέλος προς αρχής σε δομή επανάληψης. Η

μέθοδος αυτή ονομάζεται Tail Recursion Transformation και επι της ουσίας μετατρέπει την αναδρομή σε επανάληψη ξεκινώντας από το τέλος της αναδρομής προς την αρχή. Με αυτό τον τρόπο η αναδρομή υλοποιείται πολύ απλά με ένα for loop ή scan όπως το ονομάζει η βιβλιοθήκη Theano. Η παραπάνω υλοποίηση φαίνεται στο μικρό κομμάτι κώδικα που επισυνάπτεται. Η λογική του κώδικα είναι να παράγει τη κρυφή αναπαράσταση χρησιμοποιώντας του πίνακες εισόδου καθώς και προηγούμενες αναπαραστάσεις που έχουν υπολογιστεί.

```
def upstep(k, y_pred, s_t1, treeMatrix, mask, mask_emb):
    """
    Arguments:
    1) k steps : int
    2) y_prior: previous computation are not used using this Arguments
    3) h_vec : holds the computed parent vectors (dmatrix)
    4) treeMatrix : array with size (vocabulary size, 4) where columns are:
        1 - parent id
        2 - left child id
        3 - right child id
        4 - indices of the word (only for leaves of trees)

    Return:
    1) y_pred = Ws * h_t
    2) Save computed parent vector
    """
    left = treeMatrix[k, 1]
    right = treeMatrix[k, 2]
    parent = treeMatrix[k, 0]
    # parent = theano.printing.Print('Up Parent:')(treeMatrix[k, 0])
    leaf_index = treeMatrix[k, 3]

    z = ifelse(T.eq(treeMatrix[k, 1], -1),
              T.nnet.sigmoid(T.dot(self.U_z, L[:, leaf_index] * mask_emb) + self.bz),
              T.nnet.sigmoid(T.dot(self.W_z_l, s_t1[left]) + T.dot(self.W_z_r, s_t1[right]) + self.bz))
    r = ifelse(T.eq(treeMatrix[k, 1], -1),
              T.nnet.sigmoid(T.dot(self.U_r, L[:, leaf_index] * mask_emb) + self.br),
              T.nnet.sigmoid(T.dot(self.W_r_l, s_t1[left]) + T.dot(self.W_r_r, s_t1[right]) + self.br))

    h = ifelse(T.eq(treeMatrix[k, 1], -1),
              T.tanh(T.dot(self.U_h, L[:, leaf_index] * mask_emb) + self.bh),
              T.tanh(T.dot(self.W_h_l, s_t1[left] * r) + T.dot(self.W_h_r, s_t1[right] * r) + self.bh))

    s_t = ifelse(T.eq(treeMatrix[k, 1], -1),
              (1 - z) * h,
              (1 - z) * h + z * (s_t1[left] + s_t1[right]))
    s_t = ifelse(T.eq(treeMatrix[k, 0], 0), T.zeros_like(s_t), s_t)
    # return input space matrix
    out1 = T.set_subtensor(y_pred[k], T.dot(self.Ws, s_t * mask) + self.bs)
    return out1, T.set_subtensor(s_t1[parent], s_t), z, r, h
```

Σχήμα 6.2: Υλοποίηση του βασικού κώδικα για τον υπολογισμό της εμπρόσθιας φάσης για το δίκτυο TreeGRU

Με χρήση των εξόδων αυτής της συμβολικής συνάρτησης, η οποία τρέχει για όλους του κόμβους του δικτύου, μπορεί η βιβλιοθήκη Theano να υπολογίσει με αυτόματο τρόπο τις κλίσεις των παραμέτρων που θα χρησιμοποιηθούν από το αλγόριθμο εκπαίδευσης.

Όλες οι αρχιτεκτονικές που προτάθηκαν στο κεφάλαιο 5 δημιουργήθηκαν ως πα-

ραλλαγές της παραπάνω βασικής εξίσωσης με βάση τις εξισώσεις που διέπουν κάθε αρχιτεκτονική.

6.3 Αποτελέσματα

Τα πειραματικά αποτελέσματα παρουσιάζονται στο πίνακα 1 για όλα τα state-of-the-art μοντέλα που έχουν προταθεί στη βιβλιογραφία καθώς και για τα TreeGRU και TreeBiGRU που προτάθηκαν σε αυτή τη δουλειά. Ανάμεσα στις τεχνικές που έχουν προταθεί στη βιβλιογραφία, τα συστήματα DRNN και DMN έχουν τη μεγαλύτερη ακρίβεια για το δυαδικό πρόβλημα (binary scheme) (88.6), ενώ το DMN έχει την καλύτερη ακρίβεια και για το fine-grained πρόβλημα (52.1).

Σύστημα	Δυαδικό	Fine-grained
RNN	82.4	43.2
MV-RNN	82.9	44.4
RNTN	85.4	45.7
PVec	87.8	48.7
TreeLSTM	88.0	51.0
DRNN	86.6	49.8
DCNN	86.8	48.5
CNN-multichannel	88.1	47.4
DMN	88.6	52.1
TreeGRU		
- without attention	88.6	50.5
- with attention	89.0	51.0
TreeBiGRU		
- without attention	88.5	51.3
- with attention	89.5	52.4

Πίνακας 6.1: Ακρίβεια που επιτυγχάνεται Stanford Sentiment Treebank δατασετ. RNN, MV-RNN ανδ RNTN [44]. PVec: [31]. TreeLSTM [47]. DRNN [17]. DCNN [18]. CNN-multichannel [19]. DMN [23]

Παρατηρούμε ότι η καλύτερη επίδοση επιτυγχάνεται από το TreeBiGRU με το μηχανισμό προσοχής (attention mechanism) και για το δυαδικό πρόβλημα (89.5) και το fine-grained πρόβλημα (52.4) περνώντας τα αποτελέσματα που συναντώνται στη βιβλιογραφία.

Μεταξύ των αρχιτεκτονικών TreeGRU και TreeBiGRU παρατηρούμε ότι στα ανα-

δρομικά δίκτυα, η προσέγγιση δυο κατευθύνσεων, δηλαδή από κάτω προς τα πάνω και από πάνω προς τα κάτω, οδηγεί σε αποτελέσματα που είτε είναι καλύτερα ή όμοια με τα δίκτυα μιας κατεύθυνσης. Αυτό το εύρημα είναι σημαντικό καθώς η μέθοδος δυο κατευθύνσεων στα νευρωνικά δίκτυα έχει λιγότερες παραμέτρους από τις αντίστοιχες αρχιτεκτονικές σε επαναλαμβανόμενα δίκτυα (Recurrent Neural Networks).

Τέλος παρατηρούμε ότι ο μηχανισμός προσοχής βοηθά στη βελτίωση του μοντέλου και την ακρίβεια που επιτυγχάνει στο Stanford Sentiment Treebank. Όπως φαίνεται από τα σχήματα 6.3 και 6.4, η προσθήκη του μηχανισμού προσοχής βελτιώνει κατα πολύ την ικανότητα του δικτύου να ταξινομεί τα δέντρα με μεγάλο βάθος και αντίστοιχα τις προτάσεις με μεγάλο μήκος. Κατα κάποιον τρόπο, ο μηχανισμός προσοχής μπορεί να θεωρηθεί ως ένας διορθωτικός μηχανισμός καθώς επιτρέπει στο νευρωνικό δίκτυο να διορθώσει λάθη που έχουν προκύψει κατά την εφαρμογή του πάνω σε ένα δέντρο.

Περαιτέρω από τα σχήματα 6.3 και 6.4 φαίνεται ότι τα δίκτυα μονής κατεύθυνσης είναι ικανά να ταξινομήσουν με καλύτερη ακρίβεια τα μεγαλύτερα δέντρα σε αντίθεση με τα δίκτυα διπλής κατεύθυνσης. Παρόλα αυτά, τα δίκτυα διπλής κατεύθυνσης έχουμε μεγαλύτερη ακρίβεια σε δέντρα με μικρότερο βάθος τα οποία εμφανίζονται συχνότερα στη φυσική γλώσσα.

Στην επόμενη σελίδα ακολουθεί μια λίστα με τα πιο θετικά και αρνητικά n-gram που το νευρωνικό TreeBiGRU με μηχανισμό προσοχής, είναι σε θέση να αναγνωρίσει σωστά. Όλα τα n-grams προέρχονται από τη συλλογή δεδομένων Stanford Sentiment Treebank και τα οποία δεν εμφανίστηκαν κατά τη διάρκεια εκπαίδευσης του νευρωνικού δικτύου. Όσο αυξάνονται οι λέξεις εμφανίζονται όλο και πιο σύνθετες περιπτώσεις που εμφανίζονται στη φυσική γλώσσα, όπως η απλή και απότομη άρνηση ή η πρόσθεση επιθέτων με χρήση του 'και' ("and").

n	Most positive n-grams	Most negative n-grams
1	powerful; great; impressive; pretty	fails; stupid; boring; frustrating
2	sweetest movie; enormous fun; wonderful film; fantastic sets	the worst; a failure; sad-sack waste; bad film
3	lively and enjoyable; is just breathtaking; fascinating and fun; a charming triumph	extremely unfunny film; the worst thing; humourless and dull; another bad movie
5	Anchored by a terrific performance; Disturbing and brilliant documentary .; richly imagined and admirably mature; Resourceful and ingenious entertainment .	an embarrassment , a monotonous; overproduced and generally disappointing effort; heaviest , most joyless movie; is unrelentingly claustrophobic and unpleasant
8	An irresistible combination of a rousing good story; big , gorgeous , mind-blowing , breath-taking mess; It 's a terrific American sports movie and; the drumming and the marching are so excellent	of velocity and idiocy , this ruinous remake; Static , repetitive , muddy and blurry ; be the worst special-effects creation of the year; it 's a bad , embarrassing movie
9	Rich in detail , gorgeously shot and beautifully acted; part homage and part remake of the Italian masterpiece; Aside from being the funniest movie of the year; are performances to enjoy in a memorable ensemble piece	to come up with an irritatingly unimaginative retread concept; An instant candidate for worst movie of the year; A woozy , roisterous , exhausting mess , and; Contenders to create a completely crass and forgettable movie

Πίνακας 6.2: Πίνακας με τα πιο θετικά και αρνητικά n-grams, όπως αυτά προβλέπονται από το σύστημα TreeBiGRU

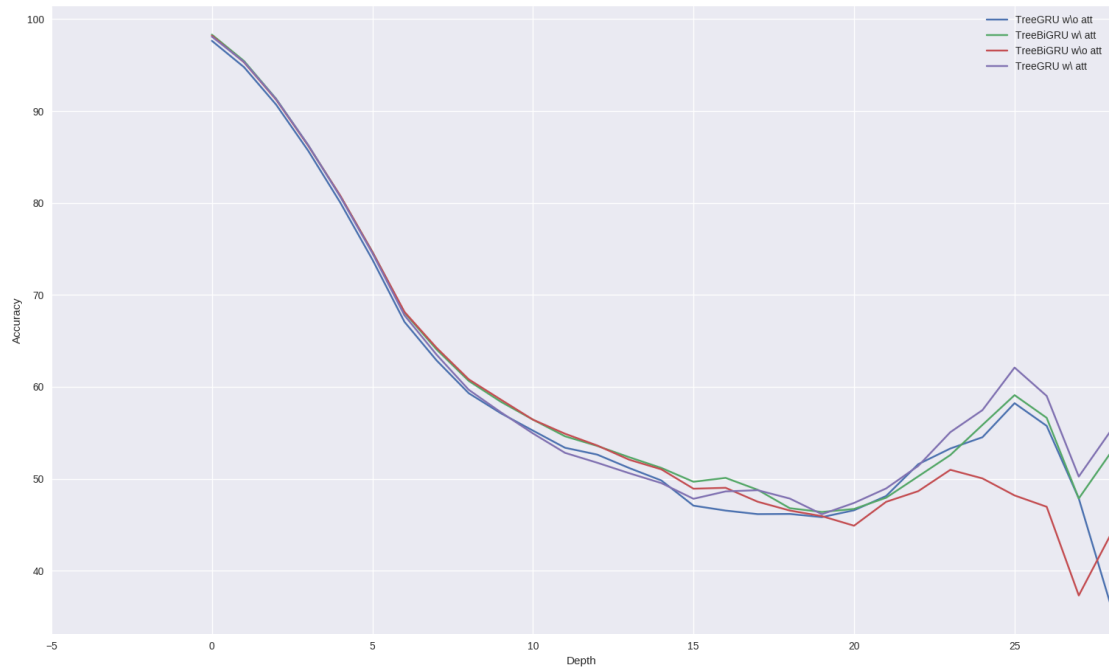
Επιπροσθέτως, οι πίνακες 6.3 και 6.4 αναφέρουν αναλυτικά τις ικανότητες ταξινόμησης του καλύτερου δικτύου από αυτά που εξετάσαμε, δηλαδή του TreeBiGRU με μηχανισμό προσοχής. Σε αυτούς παρατηρούμε ότι η ικανότητα του δικτύου να αναγνωρίζει προτάσεις με ουδέτερο συναίσθημα είναι χαμηλή καθώς η ουδέτερη (neutral) κλάση έχει μόλις 0.33 f1-score. Το γεγονός αυτό εν μέρει δικαιολογείται από την έλλειψη ουδέτερων προτάσεων παρόλο που οι περισσότερες υπο-προτάσεις φέρουν ουδέτερο συναίσθημα με βάση τη κατανομή του Stanford Sentiment TreeBank. Μια ακόμη παρατήρηση είναι ότι σε επίπεδο πρότασης και σε επίπεδο όλων των πιθανών υπο-προτάσεων, το δίκτυο είναι σε θέση να αναγνωρίζει με παρόμοια ακρίβεια τα ακραία συναισθήματα (πολύ θετικό, πολύ αρνητικό) καθώς και με σχετικά μικρή διαφορά τα μη ακραία (θετικό, αρνητικό).

	precision	recall	f1-score	support
very negative	0.54	0.43	0.48	279
negative	0.55	0.58	0.56	633
neutral	0.39	0.33	0.36	389
positive	0.46	0.71	0.56	510
very positive	0.74	0.40	0.52	399

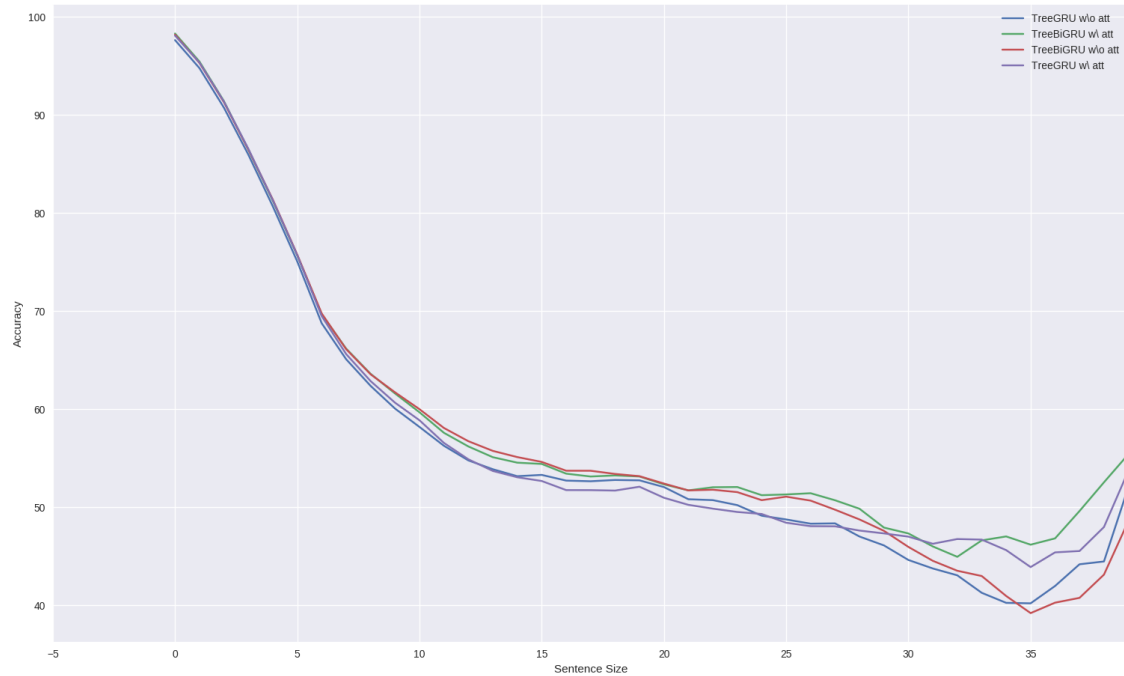
Πίνακας 6.3: Αναφορά Ταξινόμησης για το πρόβλημα 5 κλάσεων στο επίπεδο πρότασης του μοντέλου TreeBiGru με μηχανισμό προσοχής

	precision	recall	f1-score	support
very negative	0.66	0.36	0.47	2008
negative	0.66	0.62	0.64	9255
neutral	0.90	0.94	0.92	56548
positive	0.64	0.69	0.66	10998
very positive	0.85	0.39	0.53	3791

Πίνακας 6.4: Αναφορά Ταξινόμησης για το πρόβλημα 5 κλάσεων στο σύνολο των κόμβων του μοντέλου TreeBiGru με μηχανισμό προσοχής



Σχήμα 6.3: Σχηματική απεικόνιση της απόδοσης των συστημάτων με βάση το βάθος του δέντρου



Σχήμα 6.4: Σχηματική απεικόνιση της απόδοσης των συστημάτων με βάση το μέγεθος της πρότασης

Στο σχήμα 5.5 συγκρίνουμε τους 4 διαφορετικούς αλγορίθμους εκπαίδευσης που υλοποιήθηκαν στη βιβλιοθήκη Theano: 1) Adadelta, 2) Adagrad, 3) SGD και 4) Adam. Η σύγκριση γίνεται χρησιμοποιώντας ως μοντέλο την απλή μορφή νευρωνικού δικτύου και πάντα εκπαιδεύεται με τις ίδιες υπερ-παραμέτρους.

Με την πρώτη ματιά παρατηρούμε ότι ο αλγόριθμος Adam δεν μπορεί να συγκλίνει για το συγκεκριμένο πείραμα, γεγονός που επαναλήφθηκε πέντε φορές και σε κανένα δεν πέτυχε η ομαλή σύγκλιση, ενώ ταυτοχρόνως η ακρίβεια και για το train και για το test σύνολο πέφτει διαρκώς. Παρόλα αυτά, στις πρώτες επαναλήψεις του αλγορίθμου παρατηρούμε απότομη και γρήγορη σύγκλιση, οπότε δεν αποτελεί κάποιο σφάλμα κώδικα ή αριθμητική σταθερότητα. Επίσης, πειράματα με διαφορετικές υπερ-παραμέτρους δεν εμφάνιζαν προβλήματα σύγκλισης με χρήση του Adam. Η δεύτερη παρατήρηση που μπορεί να γίνει είναι ότι ο αλγόριθμος Adadelta συγκλίνει πιο γρήγορα από τον αλγόριθμο Adagrad, ενώ ο αλγόριθμος SGD συγκλίνει πολύ αργά και προς σε ένα κακό τοπικό ελάχιστο. Η ακρίβεια που επιτυγχάνεται με βάση τους αλγορίθμους Adagrad και Adadelta είναι παρόμοια, με μια μικρή παρατήρηση ότι η ακρίβεια σε επίπεδο πρότασης για τον αλγόριθμο Adagrad εμφανίζει μικρότερες διακυμάνσεις (μόβ γραμμή). Για το λόγο αυτό, ο αλγόριθμος Adagrad επιλέχθηκε ως ο κατάλληλος αλγόριθμος εκπαίδευσης γεγονός που διατυπώνεται και στη σχετική βιβλιογραφία με αναδρομικά νευρωνικά δίκτυα.



(α') Αλγόριθμος Adadelta



(β') Αλγόριθμος Adagrad



(γ') Αλγόριθμος SGD



(δ') Αλγόριθμος Adam

Σχήμα 6.5: Σχηματική απεικόνιση της σύγκλισης και απόδοσης του απλού αναδρομικού δικτύου με χρήση διάφορων αλγορίθμων εκπαίδευσης. Με μπλε και κόκκινη γραμμή απεικονίζεται η ακρίβεια σε επίπεδο δέντρου και πρότασης σε σχέση με τα δεδομένα εκπαίδευσης, όμοια με πράσινη και κόκκινη γραμμή για τα δεδομένα (test). Με κίτρινη γραμμή συμβολίζεται το λάθος του δικτύου.

Κεφάλαιο 7

Επίλογος

7.1 Συμπεράσματα

Σε αυτή τη διπλωματική εργασία, προτείνουμε μια εύλογη επέκταση των αναδρομικών δικτύων που ενσωματώνει μνήμες τύπου Φραγμένες Επαναλαμβανόμενες Μονάδες (Gated Recurrent Units - GRU). Με αυτή την επέκταση παρατηρούμε ότι το αντίστοιχο νευρωνικό δίκτυο που ενσωματώνει μονάδες LSTM πετυχαίνει καλύτερη ακρίβεια στα δεδομένα.

Στη συνέχεια, η επόμενη επέκταση στο αναδρομικό νευρωνικό δίκτυο TreeGRU αφορά την προσθήκη διπλής κατεύθυνσης υπολογισμό. Συγκεκριμένα, αποδεικνύουμε ότι η πληροφορία, που έχει καταλυτική σημασία για την δημιουργία της εσωτερικής αναπαράστασης, βρίσκεται γύρω από τον κόμβο και όχι κάτω από αυτόν όπως θεωρούν οι σχετικές εργασίες με νευρωνικά δίκτυα. Επίσης, παρατηρούμε ότι το πλήθος των παραμέτρων που χρειάζονται για τη δημιουργία ενός αναδρομικού δικτύου διπλής κατεύθυνσης είναι μόλις μισή φορά παραπάνω από το αντίστοιχο δίκτυο μια κατεύθυνσης. Σε αντίθεση, με τα επαναλαμβανόμενα δίκτυα που απαιτούν το διπλάσιο πλήθος παραμέτρων προκειμένου να τρέξουν με διπλή κατεύθυνση.

Επιπροσθέτως, εισαγάγουμε την έννοια του μηχανισμού προσοχής σε δομή. Χαρακτηριστικά, παρατηρούμε ότι δεν έχουν όλοι οι κόμβοι χρήσιμη πληροφορία για το σχηματισμό μια αναπαράστασης σε επίπεδο πρότασης. Από τα αποτελέσματα, βλέπουμε ότι η προσθήκη του μηχανισμού προσοχής σε δομή βοηθά τα δίκτυα να ταξινομούν

καλύτερα δέντρα που είναι βαθιά και έχουν περισσότερους κόμβους.

Τέλος, τα δίκτυα που υλοποιήθηκαν στο πλαίσιο αυτής της διπλωματικής εργασίας είχαν απόδοση που ξεπερνούσε τα άλλα συστήματα που συναντώνται στη βιβλιογραφία, ανεβάζοντας περαιτέρω την ακρίβεια στα δυο προβλήματα της συλλογής δεδομένων Stanford Sentiment Treebank.

7.2 Μελλοντικές Επεκτάσεις

Το σύστημα που αναπτύχθηκε στα πλαίσια αυτής της διπλωματικής εργασίας θα μπορούσε να βελτιωθεί και να επεκταθεί περαιτέρω, τουλάχιστον ως προς τρεις κατευθύνσεις. Συγκεκριμένα, αναφέρονται τα ακόλουθα:

- Ενσωμάτωση διαδικασίας μάθησης με πολιτική (curriculum learning). Όπως φάνηκε στο κεφάλαιο 6, τα δίκτυα δεν αποδίδουν τα βέλτιστα για μεγάλα και βαθιά δέντρα. Μια στρατηγική μάθησης, που εκπαιδεύει το δίκτυο σε όλο και πιο 'δύσκολα' δέντρα, μπορεί να οδηγήσει το σύστημα να συγκλίνει σε ένα καλύτερο τοπικό ελάχιστο. Η δυσκολία στα δέντρα μπορεί να μοντελοποιηθεί ως συνάρτησης του βάθους, του πλάτους ή του πλήθους των κόμβων.
- Εμβάθυνση στο μηχανισμό προσοχής πάνω σε δομή Ο μηχανισμός προσοχής πάνω σε δομή δείχνει σημάδια ότι λειτουργεί ως ένα μηχανισμό ανεκτικός σε λάθη. Μπορεί να θεωρηθεί ότι σφάλματα που γίνονται κατά τη διάρκεια υπολογισμού των εσωτερικών αναπαραστάσεων του δέντρου, διορθώνονται από το μηχανισμό προσοχής με αποτέλεσμα το δίκτυο να γίνεται περισσότερο εύρωστο.
- Αξιολόγηση του συστήματος σε άλλες συλλογές δεδομένων που δεν έχουν έτοιμη τη συντακτική ανάλυση των προτάσεων. Πέρα από τις πολλές άλλες συλλογές δεδομένων που υπάρχουν σχετικά με την ανάλυση συναισθήματος, θα είχε κάποιο ερευνητικό ενδιαφέρον να εφαρμοστεί το δίκτυο σε δεδομένα από μέσα κοινωνικής δικτύωσης.

- Προσθήκη μια δομής που λειτουργεί σαν μνήμη για το σύστημα. Πέρα από την τοπική μνήμη στη μορφή του GRU και LSTM, μια καθολική δομή μνήμης όπως αυτή αναλύεται στο δίκτυο DMN θα βοηθούσε στην περαιτέρω αύξηση της ακρίβειας στη συλλογή δεδομένων Stanford Sentiment Treebank.

Βιβλιογραφία

- [1] William Ashby. *Design for a brain: The origin of adaptive behaviour*. Springer Science & Business Media, 2013.
- [2] Andrew G Barto, Richard S Sutton και Charles W Anderson. Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE transactions on systems, man, and cybernetics*, (5):834–846, 1983.
- [3] Yoshua Bengio, Patrice Simard και Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166, 1994.
- [4] Valentino Braitenberg. *Vehicles: Explorations in synthetic psychology*, 1984.
- [5] David S Broomhead και David Lowe. Radial basis functions, multi-variable functional interpolation and adaptive networks. Τεχνική Αναφορά υπ. αριθμ., DTIC Document, 1988.
- [6] Gail A Carpenter και Stephen Grossberg. Art 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural networks*, 3(2):129–152, 1990.
- [7] Gail A Carpenter, Stephen Grossberg και David B Rosen. Fuzzy art: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural networks*, 4(6):759–771, 1991.
- [8] Michael A Cohen και Stephen Grossberg. Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE transactions on systems, man, and cybernetics*, (5):815–826, 1983.

- [9] Christian Darken, Joseph Chang και John Moody. Learning rate schedules for faster stochastic gradient search. Στο *Neural Networks for Signal Processing [1992] II., Proceedings of the 1992 IEEE-SP Workshop*, σελίδες 3–12. IEEE, 1992.
- [10] Yann N Dauphin, Razvan Pascanu, Caglar Gulcehre, Kyunghyun Cho, Surya Ganguli και Yoshua Bengio. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. Στο *Advances in neural information processing systems*, σελίδες 2933–2941, 2014.
- [11] John Duchi, Elad Hazan και Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(Θυλ):2121–2159, 2011.
- [12] Paul Ed Ekman και Richard J Davidson. *The nature of emotion: Fundamental questions*. Oxford University Press, 1994.
- [13] Stephen Grossberg. How does a brain build a cognitive code? *Psychological review*, 87(1):1, 1980.
- [14] Vasileios Hatzivassiloglou και Kathleen R McKeown. Predicting the semantic orientation of adjectives. Στο *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, σελίδες 174–181. Association for Computational Linguistics, 1997.
- [15] Donald Olding Hebb. *The organization of behavior: A neuropsychological theory*. Psychology Press, 2005.
- [16] John J Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the national academy of sciences*, 81(10):3088–3092, 1984.
- [17] Ozan Irsoy και Claire Cardie. Bidirectional recursive neural networks for token-level labeling with structure. *CoRR*, αβς/1312.0493, 2013.

- [18] Nal Kalchbrenner, Edward Grefenstette και Phil Blunsom. A convolutional neural network for modelling sentences. Στο *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, σελίδες 655–665, Baltimore, Maryland, 2014. Association for Computational Linguistics.
- [19] Yoon Kim. Convolutional neural networks for sentence classification. Στο *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, σελίδες 1746–1751, Doha, Qatar, 2014. Association for Computational Linguistics.
- [20] Diederik Kingma και Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [21] Scott Kirkpatrick, C Daniel Gelatt, Mario P Vecchi και others. Optimization by simulated annealing. *science*, 220(4598):671–680, 1983.
- [22] Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biological cybernetics*, 43(1):59–69, 1982.
- [23] Ankit Kumar, Ozan Irsoy, Jonathan Su, James Bradbury, Robert English, Brian Pierce, Peter Ondruska, Ishaan Gulrajani και Richard Socher. Ask me anything: Dynamic memory networks for natural language processing. *CoRR*, αβς/1506.07285, 2015.
- [24] Steve Lawrence, C Lee Giles, Ah Chung Tsoi και Andrew D Back. Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, 8(1):98–113, 1997.
- [25] Ralph Linsker. Self-organization in a perceptual network. *Computer*, 21(3):105–117, 1988.
- [26] Bing Liu. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167, 2012.

- [27] Warren S McCulloch και Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.
- [28] Carver Mead. Neuromorphic electronic systems. *Proceedings of the IEEE*, 78(10):1629–1636, 1990.
- [29] Tomas Mikolov, Kai Chen, Greg Corrado και Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [30] Tomas Mikolov, Martin Karafiát, Lukas Burget, Jan Cernocký και Sanjeev Khudanpur. Recurrent neural network based language model. Στο *Interspeech*, τόμος 2, σελίδα 3, 2010.
- [31] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado και Jeff Dean. Distributed representations of words and phrases and their compositionality. Στο *Advances in neural information processing systems*, σελίδες 3111–3119, 2013.
- [32] Marvin Minsky. Steps toward artificial intelligence. *Proceedings of the IRE*, 49(1):8–30, 1961.
- [33] Marvin Minsky και Seymour Papert. *Perceptrons*. 1969.
- [34] Yurii Nesterov. A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$. Στο *Doklady an SSSR*, τόμος 269, σελίδες 543–547, 1983.
- [35] Charles E Osgood. A behavioristic analysis of perception and language as cognitive phenomena. *Contemporary approaches to cognition*, σελίδες 75–118, 1957.
- [36] James W Pennebaker, Martha E Francis και Roger J Booth. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001, 2001.
- [37] Jeffrey Pennington, Richard Socher και Christopher D Manning. Glove: Global vectors for word representation. Στο *EMNLP*, τόμος 14, σελίδες 1532–1543, 2014.

- [38] Herbert Robbins και Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, σελίδες 400–407, 1951.
- [39] Nathaniel Rochester, J Holland, L Haibt και W Duda. Tests on a cell assembly theory of the action of the brain, using a large digital computer. *IRE Transactions on information Theory*, 2(3):80–93, 1956.
- [40] Frank Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.
- [41] David E Rumelhart, James L McClelland, PDP Research Group και others. Parallel distributed processing, vol. 1&2. *Cambridge, MA: The MIT Press*, 1986.
- [42] Richard Socher, Brody Huval, Christopher D. Manning και Andrew Y. Ng. Semantic compositionality through recursive matrix-vector spaces. Στο *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, EMNLP-CoNLL '12, σελίδες 1201–1211. Association for Computational Linguistics, 2012.
- [43] Richard Socher, Jeffrey Pennington, Eric H. Huang, Andrew Y. Ng και Christopher D. Manning. Semi-supervised recursive autoencoders for predicting sentiment distributions. Στο *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, σελίδες 151–161, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.
- [44] Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng και Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. Στο *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*, 2013.
- [45] Philip J Stone, Dexter C Dunphy και Marshall S Smith. The general inquirer: A computer approach to content analysis. 1966.

- [46] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll και Manfred Steede. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307, 2011.
- [47] Kai Sheng Tai, Richard Socher και Christopher D. Manning. Improved semantic representations from tree-structured long short-term memory networks. Στο *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, σελίδες 1556–1566, Beijing, China, 2015. Association for Computational Linguistics.
- [48] Peter D Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. Στο *Proceedings of the 40th annual meeting on association for computational linguistics*, σελίδες 417–424. Association for Computational Linguistics, 2002.
- [49] Peter D Turney και Michael L Littman. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS)*, 21(4):315–346, 2003.
- [50] AM Uttley. A theory of the mechanism of learning based on the computation of conditional probabilities. Στο *Proceedings of the First International Conference on Cybernetics, Namur, Gauthier-Villars*, 1956.
- [51] Bernard Widrow, Marcian E Hoff και others. Adaptive switching circuits. Στο *IRE WESCON convention record*, τόμος 4, σελίδες 96–104. New York, 1960.
- [52] Matthew D Zeiler. Adadelata: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.

Συντομογραφίες - Αρκτικόλεξα - - Ακρωνύμια

βλπ	βλέπε
κ.λπ.	και λοιπά
κ.ο.κ	και ούτω καθεξής
BPF	Band Pass Filter
Adam	Adaptive Moment Estimation
Word2Vec	Word to Vector
SGD	Stochastic Gradient Descent
Glove	Global Vector
LM	Language Model
Fine-grained	Five class problem
GRU	Gated Recurrent Units
LSTM	Long Short-Term Memory
CNN	Convolution Neural Network
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Network
RNN	Recursive Neural Network
DNN	Deep Neural Network
MLP	Multi Layer Perceptron
TreeGRU	Tree-Structure GRU network
TreeBiGRU	Bidirectional Tree-Structure GRU network
TreeLSTM	Tree-Structure LSTM network

DMN	Dynamic Memory network
DRNN	Deep Recursive Neural Network
MV-RNN	Matrix Vector Recursive Neural Network
DeepCNN	Deep Convolutional Neural Network
RNTN	Recursive Neural Tensor Network

Απόδοση ξενόγλωσσων όρων

Απόδοση

αδερφός

αμεταβλητότητα

ανάκτηση πληροφορίας

αντιμεταθετικότητα

απόγονος

γνώρισμα

διαπροσωπεία

διαφορά

δικτυακός κατάλογος

δικτυωτή δομή

δομικές σχέσεις

εγκυρότητα

ένωση

συνέλιξη

επαναλαμβανόμενος

αποτέλεσμα πολλαπλασιασμού

βελτιστοποίηση

εκπαίδευση

κλίση

κάθοδος

πρόβλημα εκθετικά μειωμένης κλίσης

Ξενόγλωσσος όρος

sibling

idempotency

information retrieval

commutativity

descendant

attribute

interface

difference

portal catalog

lattice

structural relationships

validity

union

convolution

recurrent

product

optimization

training

gradient

descent

vanishing gradient problem

