Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

Ανίχνευση μεταφορικών γλωσσικών φαινομένων με την χρήση Συνδυασμού Βαθιών Νευρωνικών δικτύων

Διπλωματικη Εργασια

του

ΡΟΛΑΝΔΟΥ ΑΛΕΞΑΝΔΡΟΥ ΠΟΤΑΜΙΑ

Επιβλέπων: Ανδρέας Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

Εργαστηριο Υψηλων Συστηματων
Αθήνα, Σεπτέμβριος 2018

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών
Εργαστήριο Ευφυών Συστημάτων

Ανίχνευση μεταφορικών γλωσσικών φαινομένων με την χρήση Συνδυασμού Βαθιών Νευρωνικών δικτύων

Διπλωματικη Εργασια

του

ΡΟΛΑΝΔΟΥ ΑΛΕΞΑΝΔΡΟΥ ΠΟΤΑΜΙΑ

Επιβλέπων: Ανδρέας Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 3η Οκτωβρίου 2018.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Ανδρέας Σταφυλοπάτης	Παναγιώτης Τσανάκας	Γεώργιος Στάμου
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Αναπληρωτής Καθηγητής Ε.Μ.Π.

Αθήνα, Σεπτέμβριος 2018

(Υπογραφή)

.....

Ρολανδος Αλεξανδρος Ποταμιας
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Περίληψη

Αντικείμενο της διπλωματικής εργασίας είναι η ανάπτυξη μοντέλων αναγνώρισης *μεταφορικών γλωσσικών φαινομένων(ΜΓΦ)* με τεχνικές βαθιάς μηχανικής μάθησης (Deep Learning). Το πρόβλημα της αναγνώρισης και κατάταξης ΜΦΓ αποτελεί ένα ανοιχτό πρόβλημα της Συναισθηματικής Ανάλυσης στο πεδίο της Επεξεργασίας Φυσικής Γλώσσας λόγω της *νοηματικής αντίθεσης* που περιέχεται σε αυτά. Το πρόβλημα αυτό αποτελείται από την αναγνώριση τριών αλληλένδετων ΜΓΦ: του σαρκασμού, της ειρωνείας και της μεταφοράς, τα οποία, στα πλαίσια της παρούσας εργασίας, αντιμετωπίζονται με προηγμένες τεχνικές βαθείας μηχανικής μάθησης (R, LSTM) και με τεχνικές μηχανισμών διανυσματικής υποστήριξης (SVM).

Αρχικά, διευρευενούνται μέσω εκτεταμένης βιβλιογραφικής έρευνας οι τεχνολογίες αιχμής (stat-of-the-art) και οι ερευνητικές εξελίξεις στην ανίχνευση και αναγνώριση ΜΓΦ και καταγράφονται οι σημαντικότερες προσεγγίσεις. Στην ανασκόπηση αυτή, δίνεται ιδιαίτερη έμφαση τόσο στην μέθοδο εξαγωγής χαρακτηριστικών όσο και στους αλγορίθμους μηχανικής μάθησης που χρησιμοποιούνται. Στην συνέχεια περιγράφονται συνοπτικά οι βασικές θεωρητικές αρχές πάνω στις οποίες βασίζεται η προτεινόμενη αντιμετώπιση του προβλήματος.

Στην συνέχεια αναπτύσσεται το πλαίσιο και το στάδιο προεπεξεργασίας των σχετικών δεδομένων (από κοινωνικά δίκτυα, -tweets)με σκοπό τη βέλτιστη προετοιμασία τους πριν εισαχθούν στα μοντέλα βαθιάς μηχανικής μάθησης. Επιπρόσθετα, εξάγονται από τα δεδομένα χαρακτηριστικά που μπορούν να διαχωριστούν σε τέσσερις κατηγορίες: τα συντακτικά, εκφραστικά, συναισθηματικά και ψυχολογικά, καθένα από τα οποία αποτυπώνει πτυχές για την μέθοδο γραφής και εκφοράς λόγου του χρήστη των κοινωνικών δικτύων.

Τέλος, δημιουργείται ένα πρωτότυπο μοντέλο **Deep Ensemble Soft Classifier-DESC**, που συνδυάζει αλγορίθμους βαθιάς μάθησης. Χρησιμοποιώντας τέσσερα διαφορετικά σύνολα δεδομένων αναφοράς (benchmark data), από γνωστά και διαδεδομένα συνέδρια και σχετικούς διαγωνισμούς (Semantic Evaluation-SemVal), και εξαντλητική αξιολόγηση της ικανότητας αναγνώρισης, διακρίνουμε πως το μοντέλο DESC επιτυγχάνει πολύ καλή συμπεριφορά, άξια σύγκρισης με σχετικές μεθοδολογίες και τεχνολογίες αιχμής στο προκλητικό πεδίο της αναγνώρισης ΜΓΦ.

Λέξεις Κλειδιά

Ανάλυση συναισθήματος, επεξεργασία φυσικής γλώσσας, μεταφορικά γλωσσικά φαινόμενα, σαρκασμός, ειρωνεία, μηχανική μάθηση, βαθιά μάθηση, τεχνητά νευρωνικά δίκτυα.

Abstract

The subject of the diploma thesis is the development of models for the recognition of *figurative language (FL)* utilizing *deep learning* techniques. The management, recognition and classification of FL is an open problem of *Sentiment analysis* in the broader field of *natural language processing (NLP)* due to the *contradictory meaning* contained in phrases with metaphorical content. The problem itself represent three interrelated FL recognition tasks: sarcasm, irony and metaphor which, in the present work, are dealt with advanced deep learning (Recurrent Neural Networks, LSTM) and support vector machine (SVM) techniques.

Initially, the state-of-the-art technologies in the field of FL detection and recognition are being explored through extensive bibliographical research, and the most important approaches are documented. The emphasis of the review is placed on both the feature extraction methodologies and the machine learning algorithms being utilized. In the sequel, the basic theoretical principles and techniques, on which the proposed approach is based, are presented.

Next, the prepossessing framework of the relevant social-media data (tweets) is presented. Data prepossessing aims towards efficient data representation formats so that to optimize the respective inputs to the deep learning models. In addition, special features are extracted from the data in order to characterize the syntactic, expressive, emotional and temper content reflected in the respective social media text references. These features aim to capture aspects of the social network user's writing method.

Finally, a prototype, **Deep Ensemble Soft Classifier-DESC** is created which, is based on the combination of different deep learning techniques. Using four different sets of benchmark data-sets, from well-known and widespread conferences and related contests, and based on the assessment of the performance of different FL recognition approaches, we conclude that the DESC model achieves a very good performance, worthy of comparison with relevant methodologies and state-of-the-art technologies in the challenging field of FL recognition.

Keywords

Sentiment analysis, natural language processing, figurative language, sarcasm, irony, machine learning, deep learning, neural networks.

Ευχαριστίες

Για την πραγματοποίηση της διπλωματικής μου εργασίας εργάστηκα στο Εργαστήριο Ευφυών Συστημάτων του τομέα Τεχνολογίας Πληροφορικής και Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου, με επιβλέποντα καθηγητή τον κ. Ανδρέα Σταφυλοπάτη. Έτσι το πρώτο άτομο που θα ήθελα να ευχαριστήσω είναι το καθηγητή μου κ. Ανδρέα Σταφυλοπάτη για την ευκαιρία που μου έδωσε να ασχοληθώ με τα πεδία της ταξινόμησης κειμένων σε κοινωνικά δίκτυα και προωθημένων τεχνικών Μηχανικής Μάθησης και να εκπονήσω τη διπλωματική εργασία μου στο εργαστήριο του.

Επιπλέον, θα ήθελα να ευχαριστήσω τον κύριο Δρ. Γεώργιο Σιόλα, ο οποίος μου προσέφερε στήριξη και βοήθεια σε όποιο πρόβλημα αντιμετώπιζα αλλά και καθοδήγηση για να ολοκληρώσω την διπλωματική μου. Θα ήθελα επίσης να τον ευχαριστήσω ιδιαιτέρως για τις πολύτιμες συμβουλές του οι οποίες βοήθησαν στην βελτίωση της έρευνας μου.

Στην συνέχεια θα ήθελα να ευχαριστήσω τους καθηγητές Γεώργιο Στάμου και Παναγιώτη Τσανάκα που συμμετέχουν στην τριμελή επιτροπή της διπλωματικής μου εργασίας.

Κλείνοντας θα ήθελα να ευχαριστήσω την οικογένεια μου και τους φίλους μου που είναι διπλά μου και με στηρίζουν όλα αυτά τα χρόνια, ώστε να επιτύχω τους στόχους μου.

Στους γονείς μου

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Κεφάλαιο 1

Εισαγωγή

Η χρήση των Μέσων Κοινωνικής Δικτύωσης αυξάνεται ραγδαία χρόνο με τον χρόνο μειώνοντας τις αποστάσεις μεταξύ των ανθρώπων με την δυνατότητα συνεχείς επικοινωνίας με κάθε μέρος του κόσμου. Η ταχύτητα μετάδοσης πληροφοριών γίνεται τεράστια και οι ειδήσεις πλέον μεταδίδονται ακαριαία σε κάθε πλευρά του πλανήτη. Η επικοινωνία μεταξύ των ανθρώπων αλλά και ο σχολιασμός γεγονότων της επικαιρότητας αποτελεί ιδιαίτερα χρήσιμες πληροφορίες που μπορούν να χρησιμοποιηθούν από αλγορίθμους Τεχνητής Νοημοσύνης. Τα δεδομένα αυτά μπορούν να αναλυθούν ώστε να κατηγοριοποιούνται τα γεγονότα της επικαιρότητας που αποτελούν πηγή ενδιαφέροντος ή να αξιολογείτε η επιρροή ενός πολιτικού κόμματος. Τα προβλήματα αυτά παρουσιάζουν ιδιαίτερες δυσκολίες όταν στους διαλόγους χρησιμοποιούνται ειρωνικά ή σαρκαστικά σχόλια τα οποία αναστρέφουν το νοηματικό περιεχόμενο του διαλόγου. Έτσι προκύπτει η ανάγκη για δημιουργία αλγορίθμων που θα συμβάλει στην αναγνώριση γλωσσικών φαινομένων που αλλοιώνουν την ουσία του διαλόγου.

1.1 Αντικείμενο της διπλωματικής

Στην διπλωματική εργασία αυτή διατυπώνουμε το πρόβλημα της αναγνώρισης μεταφορικών γλωσσικών φαινομένων(ΜΓΦ) (Figurative Language) όπως η ειρωνεία, ο σαρκασμός και οι μεταφορές. Το πρόβλημα αυτό αποτελεί ένα ανοικτό πρόβλημα στο ερευνητικό πεδίο της **Συναισθηματικής Ανάλυσης (Sentiment Analysis)** λόγω του διφορούμενου τρόπου ανάγνωσης και ερμηνείας της κάθε πρότασης. Σαν Συναισθηματική Ανάλυση ορίζουμε ορίζουμε το πρόβλημα προσδιορισμού του συναισθηματικού ενδιαφέροντος και κλίσης ή πόλωσης (inclination) μιας αναφοράς είτε/και μιας διαλογικής πράξης, εκφρασμένων σε μορφή ελεύθερου κειμένου . Αποτελεί πεδίο της υπολογιστικής γλωσσολογίας (computation lignuistics) και επικεντρώνεται στην αναζήτηση της συναισθηματικής πόλωσης του κειμένου. Οι κύριες μεθοδολογίες αντιμετώπισης των συγκεκριμένων προβλημάτων εστιάζονται σε λέξεις με που παρουσιάζουν υψηλό δείκτη συναισθηματικού ενδιαφέροντος όπως η χαρά, ο θυμός και η λύπη. Παρόλα αυτά υπάρχουν ιδιαιτερότητες στα παραπάνω προβλήματα όταν χρησιμοποιείται άρνηση ή κάποιο μεταφορικό γλωσσικό φαινόμενο που καθιστά την Συναισθηματική Ανάλυση σχεδόν αδύνατη. Η δυσκολία αύτη έγινε εμφανής στο Workshop Semantic Evaluation -

2014 [85] όπου στο σύνολο των δεδομένων προς Συναισθηματική Ανάλυση υπήρχαν αρκετές αναφορές με σαρκαστικό περιεχόμενο. Τα αποτελέσματα έδειξαν πως η ύπαρξη σαρκαστικών δεδομένων δυσκολεύει τις διαδικασίες της ταξινόμησης, παρότι στα μη σαρκαστικά δεδομένα οι αλγόριθμοι έχουν αξιόλογα αποτελέσματα. Προκύπτει λοιπόν η ανάγκη για δημιουργία ταξινομητών που θα αναγνωρίζουν τα παραπάνω γλωσσικά φαινόμενα. Τόσο η ειρωνεία όσο και ο σαρκασμός αποτελούν δημιουργικές χρήσης της γλώσσας αλλά τα τελευταία μόλις χρόνια απέσπασαν το επιστημονικό ενδιαφέρον της υπολογιστικής γλωσσολογίας. Συγκεκριμένα το ενδιαφέρον για αναγνώριση των μεταφορικών γλωσσικών φαινομένων ξεκίνησε με ένα tweet το 2014 από την αεροπορική εταιρία Ryanair όπως φαίνεται στην εικόνα 1.1 :

Σχήμα 1.1: Σαρκαστικό Tweet που δεν κατάλαβε ο υπεύθυνος της Ryanair

Παρότι ο σαρκασμός του Ryan Hand αναφορικά με την, μεταφορική έκφραση "emotional baggage" ήταν σχετικά ξεκάθαρος δεν έγινε αντιληπτό από την Ryanair η οποία αναζήτησε πληροφορίες για το γεγονός. Λίγα λεπτά αργότερα επανήλθε με νέο Tweet απολογούμενη: *"We apologize for temporary technical difficulties with our sarcasm detector"*. Παρότι οι έρευνες δείχνουν ότι τα παιδία ήδη από την ηλικία των 5 ετών αποκτούν, σε ένα βαθμό, την ιδιότητα να κατανοούν τον σαρκασμό [68], η αντιμετώπιση του στο πεδίο της υπολογιστικής γλωσσολογίας αλλά και από προσεγγίσεις Τεχνητής Νοημοσύνης (ΤΝ), έχει αποδειχτεί ιδιαίτερα δύσκολη. Τόσο η ειρωνεία όσο και ο σαρκασμός αποτελούν γλωσσικά φαινόμενα τα οποία διακρίνονται από την αντίθεση τους σε σχέση με τα συμφραζόμενα. Μερικές φορές ακόμα και τα πιο ειρωνικά και σαρκαστικά άτομα αδυνατούν να συλλάβουν την ειρωνεία του συνομιλητή τους. Η ανίχνευση των ΜΓΦ σε μικρού μήκους κείμενο όπως είναι τα tweet κάνουν την προσπάθεια ακόμα πιο σύνθετη. Σε αντίθεση με άλλες κατηγορίες προβλημάτων ανάλυσης φυσικής γλώσσας όπως η κατηγοριοποίηση με βάση το θέμα του κειμένου ή ακόμα και την συναισθηματική ανάλυση του, η ανίχνευση ΜΓΦ σε ένα κείμενο αποτελεί ένα δύσκολο έργο εξαιτίας της έλλειψης του τόνου, της χροιάς της φωνής που το διατυπώνει αλλά και της έκφρασης του προσώπου του. Ένας παράγοντας που κάνει το πρόβλημα ακόμα πιο δύσκολο είναι η πως ο μέσος άνθρωπος που διατυπώνει κάτι ειρωνικό στα Social Media αφιερώνει πε-

ρίπου τον διπλάσιο χρόνο για να καταγράψει την ειρωνεία ή τον σαρκασμό του σε σχέση με αυτόν που θα αφιερώνει σε μια προφορική ομιλία. Αυτό κάνει το ΜΓΦ ακόμα πιο 'βαθύ' και δυσνόητο και εκ' τούτου πιο δύσκολο στην υπολογιστική επεξεργασία του.

1.2 Κοινωνικό Δίκτυο Twitter

Το Twitter αποτελεί ένα μέσο κοινωνικής δικτύωσης όπου οι χρήστες εκφράζονται με κείμενα μικρού μήκους, το πολύ 120 χαρακτήρων, τόσο για την επικαιρότητα όσο και για την καταγραφή των σκέψεων τους. Το μεγαλύτερο ποσοστό των συγγενών εργασιών χρησιμοποιεί αντλεί δεδομένα από το Twitter. Ό λόγος είναι ιδιαίτερα προφανής καθότι το Twitter αποτελεί το αμιγές μέσο κοινωνικής για την ελεύθερη έκφραση των χρηστών του. Οι πολύ ελαστικές πολιτικές ορθής χρήσης του Twitter το καθιστά άμεσα συνδεδεμένο με την χρήση ειρωνικών, σαρκαστικών αλλά προσβλητικών σχολίων. Μερικά από τα χαρακτηριστικά του είναι η δυνατότητα του λογαριασμού να αποκτά άτομα που ακολουθούν της δραστηριότητες του (followers) αλλά και να ακολουθεί, ομοίως, άλλους χρήστες. Ταυτόχρονα δίνεται στον χρήστη η δυνατότητα να εκδηλώνει την ταύτιση του με κάποιο σχόλιο (Like) ή να αναπαράγει αυτό το σχόλιο προς τα άτομα που των ακολουθούν (Re-tweet). Ένα ακόμη σημαντικό χαρακτηριστικό του Twitter είναι η δυνατότητα που δίνει στους χρήστες του να χρησιμοποιούν hashtags (#tag) τα οποία διευκολύνουν και χρησιμοποιούνται για την ομαδοποίηση των tweets. Τα hashtag αποτελούν μια μορφή σύνοψης της πρόθεσης του χρήση για το tweet και πολύ εύχρηστο εργαλείο για την επιτυχή κατηγοριοποίηση και κατάταξη ενός τωεετ. Το API του Twitter δίνει την δυνατότητα στους χρήστες να κατεβάζουν ένα ποσοστό από τα tweet που καταγράφονται καθημερινά, με αποτέλεσμα χρησιμοποιώντας κάποια hashtag όπως για παράδειγμα τα #sarcasm , #irony , #not μπορούμε να αντλήσουμε ένα πλήθος από tweet για την δημιουργία ενός συνόλου μεταφορικών δεδομένων προς ανάλυση και σύγκριση συγκεκριμένων τεχνικών και αλγοριθμικών προσεγγίσεων.

1.3 Ορισμοί

Ως άνθρωποι χρησιμοποιούμε την γλώσσα για να επικοινωνούμε με τους συνανθρώπους μας αλλά και να εκφράζουμε τα συναισθήματα μας με αρκετούς τρόπους, με έναν από αυτούς να είναι τα ΜΓΦ. Η κατανόηση και η αντίληψη των ΜΓΦ, απαιτεί την ανάπτυξη μεγάλου εύρους γνωστικών ικανοτήτων που στους ανθρώπους ωριμάζουν προοδευτικά κατά τη διάρκεια της ζωής τους. Οι ικανότητες αυτές δίνουν την δυνατότητα στον άνθρωπο να αλληλοεπιδρά σε συνομιλίες και να αντιλαμβάνεται τόσο την κυριολεκτική έννοια μιας λέξης όσο και την μεταφορική της.

- **Ειρωνεία**: Ορίζεται ως εκείνο το μέρος αναφορών σε φυσική γλώσσα το οποίο διακατέχεται από δηκτικό περιεχόμενο και διάθεση και μπορεί να διακρίνει τη διαφορά της κυριολεκτικής από την μεταφορική έννοια της. Η ειρωνεία κατά τον Αριστοτέλη χαρακτηρίστηκε και αναδείχτηκε ως 'εξευγενισμένη προσβολή'. Η ειρωνεία μπορεί να διαχωριστεί σε δύο κατηγορίες την 'Καταστατική' (Situational) και την 'Προφορική' (Verbal)

με την οποία και θα ασχοληθούμε κυρίως. Η 'Καταστατική' ειρωνεία χαρακτηρίζεται από γεγονότα που έρχονται σε αντίθεση με τα αναμενόμενα όπως για παράδειγμα όταν ένας κωμικός πάσχει από κατάθλιψη ή όταν ο κρεοπώλης αποκαλύπτει πως είναι χορτοφάγος. Αντίθετα 'Προφορική' χαρακτηρίζεται η ειρωνεία κατά την οποία ο ομιλητής εκφράζει ακριβώς το αντίθετο από αυτό που εννοεί, όπως για παράδειγμα η διερώτηση του γονιού για το πόσο τυχερός είναι όταν ο γιός του ανακοινώνει πως τράκαρε το αμάξι του ή όταν ο μαθητής απαντά στην δασκάλα πως λατρεύει το εξώφυλλο του βιβλίο όταν τον ρωτά αν το διάβασε[18]. Η ειρωνεία χαρακτηρίζεται επίσης από ηχηρές λέξεις ώστε να γίνεται αντιληπτή χωρίς όμως να μπορεί να ακολουθηθεί κάποιο μοτίβο για την σύλληψη και την ερμηνεία της όπως εξηγούν οι Sperber&Wilson [88]. Συνήθως προβάλει την επικριτική στάση του ομιλητή σε κάποιο θέμα ή γεγονός.

- **Σαρκασμός**: Ο σαρκασμός σε αντίθεση με την ειρωνεία είναι δύσκολο να χαρακτηριστεί μέσω κάποιου ορισμού. Το Λεξικό Oxford[1] της χρήσης των Αγγλικών ορίζει τον σαρκασμό ως τη χρήση της ειρωνείας που αποσκοπεί στην κοροϊδία ή την περιφρόνηση. Αντίθετα το Merriam Webster[2] δίνει έναν ορισμό που δεν διακριτοποιεί ιδιαίτερα τον σαρκασμό από την ειρωνεία. Συγκεκριμένα αναφέρει πως *ο σαρκασμός είναι η χρήση λέξεων με νόημα αντίθετο από το κυριολεκτικό τους και αποσκοπεί στην προσβολή, την ένδειξη εκνευρισμού ή χιούμορ*. Πιθανόν μια λεπτή διαφορά μεταξύ της ειρωνείας είναι η αρνητική του προδιάθεση και η καυστικότητα του. Ψυχολόγοι χαρακτηρίζουν τον σαρκασμό ως μια μορφή λεκτικής άμυνας, επιθετικού χιούμορ αλλά και ως έναν τρόπο για να εκφραστούν συναισθήματα χωρίς να πληγώσουν άμεσα κάποιον [7]. Στην συνήθη πάντως μορφή του ο σαρκασμός αποτελεί μια επιθετική ειρωνεία με σκοπό την κριτική σε κάποιο συγκεκριμένο ζήτημα [2, 15, 29].

- **Μεταφορά**: Η μεταφορά αποτελεί ένα γλωσσικό φαινόμενο το οποίο ενυπάρχει τόσο στο σαρκασμό όσο και στην ειρωνεία. Η τροπική χρήση λέξεων που μπορούν να αποδώσουν διαφορετικό, από το προφανές, νόημα σε μια πρόταση είναι βασικό χαρακτηριστικό των μεταφορών. Ένα παράδειγμα μεταφορικής γλώσσας θα μπορούσε να είναι όταν αποκαλούμε *χρυσό παιδί τον φίλο μας* ή η γνωστή φράση του Bob Dylan *"Chaos is a friend of mine"*. Γλωσσολόγοι αναφέρουν πως για να χαρακτηριστεί μια πρόταση μεταφορική πρέπει το κυριολεκτικό της νόημα να μην μπορεί να σταθεί λογικά και χαρακτηρίζεται από από το ερμηνευτικό σχήμα '*κάτι είναι κάτι άλλο*'. Η μεταφορές χρησιμοποιούνται κυρίως στην λογοτεχνία, όμως παρατηρώντας κανείς καθημερινές συζητήσεις μπορεί να αντιληφθεί πως χαρακτηρίζουν και το προφορικό λόγο. Στην εργασία αυτή θα συμπεριλάβουμε και την **παρομοίωση** ως μεταφορικό σχήμα λόγου παρότι χαρακτηρίζεται από εκφράσεις της μορφής '*κάτι είναι σαν κάτι άλλο*'. Οι γλωσσολόγοι χαρακτηρίζουν πως η μεταφορά δεν αποτελεί παρομοίωση όμως η παρομοίωση αποτελεί μεταφορά. Ένα παράδειγμα παρομοίωσης θα μπορούσε να είναι όταν *παρατηρούμε πως το 'μπροστινό αμάξι πάει σαν χελώνα'* ή όταν *αναφέρουμε πως ο φίλος μας*

[1] https://en.oxforddictionaries.com/definition/sarcasm
[2] https://www.merriam-webster.com/dictionary/sarcasm

'λάμπει από την χαρά του'.

1.4 Χρησιμότητα

Όπως αναφέρθηκε παραπάνω, τόσο ειρωνεία όσο και ο σαρκασμός αποτελούν πεδίο της Συναισθηματικής Ανάλυσης που πλέον αποτελεί τομέα έρευνας για την κατανόηση από τον υπολογιστή των συναισθημάτων που εκφράζουν οι άνθρωποι. Όπως είναι εύκολα αντιληπτό ο βασικός τομέας στον οποίο μπορεί η αναγνώριση ΜΓΦ να εφαρμοστεί είναι για την Συναισθηματική Ανάλυση αλλά και την Ανάλυση Προφίλ Απόψεων (Opinion Mining). Τα μεταφορικά γλωσσικά φαινόμενα παρότι έχουν αναλυθεί σε μεγάλο βαθμό τόσο από την σκοπιά της γλωσσολογίας και της ψυχολογίας όσο και από την σκοπιά της γνωστικής επιστήμης[97, 24, 52] αποτελούν ένα ιδιαίτερα δύσκολο έργο του τομέα της υπολογιστικής γλωσσολογίας και της Τεχνητής Νοημοσύνης [69]. Η κατανόηση των ΓΜΦ δίνει την δυνατότητα σε τεχνικές Μηχανικής Μάθησης να προσαρμόζουν το προφίλ απόψεων κάθε χρήστη αποτελεσματικότερα αναγνωρίζοντας την αρνητικότητα σε γεγονότα και καταστάσεις στα λεγόμενα του [19]. Ταυτόχρονα η αναγνώριση ΜΓΦ αποτελεί ιδιαίτερα σημαντικό εργαλείο για εταιρείες παροχής υπηρεσιών που αναζητούν αυτοματοποιημένα συστήματα για την κατανόηση σχολίων και κριτικών αξιολόγησης. Είναι σαφές ότι στις φόρμες αξιολόγησης η εμφάνιση ΜΓΦ είναι πυκνή και τα καθιερωμένα συστήματα ανάλυσης δεν μπορούν να κατανοήσουν αξιολογήσεις που περιέχουν ειρωνικά ή σαρκαστικά σχόλια. Παρόμοια χρηστικότητα έχει στην πρόβλεψη μετακίνησης των δεικτών του χρηματιστηρίου αλλά και στην αξιολόγηση προϊόντων. Τέλος μια ακόμα χρήση της αναγνώρισης ΜΓΦ είναι στα συστήματα ασφαλείας τα οποία φέρεται να δέχονται καθημερινά δεκάδες απειλητικά μηνύματα τα οποία μερικά περιέχουν καθαρές μορφές ειρωνείας και σαρκασμού και θα μπορούσαν να παραλειφθούν κατά τη διαδικασία ελέγχου. Χαρακτηριστικό παράδειγμα ήταν το tweet ενός λογαριασμού κατά την διάρκεια του διαγωνισμού της Eurovision το 2015 όταν η Λιθουανία δεν έδωσε κάποια ψήφο στην Ρωσία : *"And 12 bombers depart now...to Lithiania"* προκαλώντας για μερικά λεπτά γενική αναταραχή. Παρόμοιο γεγονός είχε λάβει χώρα το 2010 όταν ένας νεαρός έκανα ένα σαρκαστικό tweet αναφέροντας πως θα ανατινάξει το αεροδρόμιο του Doncaster οδηγώντας τις αρχές στην σύλληψη του[3].

Αντίστοιχα γεγονότα έχουν αναφερθεί από τις μυστικές υπηρεσίες των ΗΠΑ και οδήγησε στη σχεδίαση και ανάπτυξη πανίσχυρων αλγορίθμων για την ανίχνευση σαρκασμού [4] ώστε να αποφευχθούν παρόμοια περιστατικά.

1.5 Οργάνωση της Διπλωματικής

Στην παρούσα διπλωματική αναπτύσσουμε την αντιμετώπιση των μεταφορικών γλωσσικών φαινομένων ως ένα πρόβλημα αναγνώρισης τριών αναπαραστάσεων τους: της ειρωνείας, του σαρκασμού και της μεταφοράς. Στο Κεφάλαιο 2 καταγράφουμε το σύνολο των συγγενών εργασιών με τις τρείς αναπαραστάσεις των ΜΓΦ, ενώ στο Κεφάλαιο 3 διατυπώνονται οι βα-

[3] https://www.theguardian.com/world/2010/jan/18/robin-hood-airport-twitter-arrest
[4] https://www.bbc.com/news/technology-27711109

σικές θεωρητικές αρχές και προσεγγίσεις που ακολουθούνται από τη παρούσα διπλωματική εργασία. Συνοπτικά, επεξηγούνται οι βασικές αρχές λειτουργίας των Νευρωνικών δικτύων και των τεχνικών βαθιάς μηχανικής μάθησης καθώς και οι τεχνικές επεξεργασίας κειμένου. Στο Κεφάλαιο 4 περιγράφονται οι βασικές αρχιτεκτονικές Νευρωνικών δικτύων που χρησιμοποιούνται για την επίλυση των προαναφερθέντων προβλημάτων. Οι αρχιτεκτονικές αυτές εφαρμόζονται σε τέσσερα διαφορετικά σύνολα δεδομένων στο Κεφάλαιο 5, όπου ταυτόχρονα καταγράφονται και οι τεχνικές προεπεξεργασίας τους καθώς και η σύγκριση αποτελεσμάτων με συγγενικές εργασίες. Τέλος, στο Κεφάλαιο 6 καταγράφουμε τα βασικά συμπεράσματα της εργασίας καθώς και τα μελλοντικά μας ερευνητικά σχέδια με το θέμα της αναγνώρισης μεταφορικών γλωσσικών φαινομένων.

Κεφάλαιο 2

Συγγενείς εργασίες

Στο παρόν κεφάλαιο παρουσιάζονται παρόμοιες και συγγενείς στον τομέα της αναγνώρισης Μεταφορικών Γλωσσικών Φαινομένων. Η εργασία αυτή μπορεί να χαρακτηριστεί από τρία πεδία ενδιαφέροντος: την αναγνώριση σαρκασμού, την αναγνώριση ειρωνείας καθώς και την αναγνώριση του συνδυασμού ΜΓΦ (ειρωνείας, σαρκασμού, μεταφοράς). Ελάχιστες εργασίες προσεγγίζουν πλήρως τα ΜΓΦ στο σύνολο τους, με εξαίρεση το διεθνή Workshop Semantic Evaluation 2015 που διατύπωσε την συνολική μορφή αυτού του προβλήματος, ως μια επέκταση της ανάλυσης συναισθήματος. Παρακάτω διατυπώνονται οι σχετικές εργασίες στα τρία αυτά προβλήματα.

2.1 Αναγνώριση Σαρκασμού

Το 2015 ο Ghosh [36] χρησιμοποιεί το #sarcasm για να αντλήσει δεδομένα από το Twitter. Εισάγει την έννοια της διαφωνίας των λέξεων μέσα σε ένα Tweet ως κομβικό στοιχείο για την κατηγοριοποίηση σαρκαστικών σχολίων. Ταυτόχρονα χρησιμοποιεί word2vec Embeddings και έναν κατάλληλα διαμορφωμένο πυρήνα ταξινομητή SVM που προσαρμόζεται ανάλογα με την μέτρηση ομοιότητας του tweet με ένα σαρκαστικό όριο. Ο Davidov[26] χρησιμοποιεί δύο σύνολα δεδομένων για την προσέγγιση του προβλήματος: δεδομένα από το Twitter με βάση εκ νέου το #sarcasm και δεδομένα από το Amazon[1]. Δημιουργεί χαρακτηριστικά βασισμένα σε πρότυπα (Pattern-Based) όπως η συχνότητα εμφάνισης λέξεων στο σύνολο των δεδομένων ή λέξεις περιεχομένου (content words). Ο τελικός πίνακας χαρακτηριστικών δημιουργείτε με έναν αλγόριθμο ταιριάσματος (Matching Algorithm) και τα δεδομένα ταξινομούνται με την χρήση κ-Πλησιέστερων Γειτόνων (k-Nearest Neighbours). Μια διαφορετική τεχνική ανίχνευσης σαρκασμού προτείνει ο Ibanez συγκρίνοντας τα σαρκαστικά σχόλια με θετικά και αρνητικά συναισθηματικά σχόλια (Positive-Negative Sentiment). Για την προσέγγιση αυτή χρησιμοποιούνται hashtag όπως #joy ,#happy , #lucky κλπ για την αναζήτηση θετικά συναισθηματικών σχολίων και #sad, #sadness, #angry για αρνητικά συναισθηματικά σχόλια. Το πρόβλημα λοιπόν προσεγγίζεται ως πρόβλημα δυαδικής ταξινόμησης, σαρκαστικό έναντι θετικών-αρνητικών σχολίων. Χρησιμοποιεί δύο είδη χαρακτηριστικών: Λεξιλογικά

[1] https://www.amazon.com/

χαρακτηριστικά με την χρήση του λεξιλογίου WordNet affect[92], του πίνακα καταμέτρησης λέξεων Linguistic Inquiry and Word Count-LWIC [73], σημεία στίξης και unigram μοντέλα και ¨πραγματιστικά' χαρακτηριστικά που αποτυπώνουν το συναίσθημα των emoji και των λέξεων. Χρησιμοποίει SVM με πυρήνα Sequential Minimal Optimization καθώς και Logistic Regression που δείχνει να αποδίδει καλύτερα. Παρόμοιες με τις προαναφερθείσες τεχνικές χρησιμοποιούνται από τον Liebrecht [56] για να προσεγγιστούν τα φαινόμενα σαρκασμού στο Γερμανικό Twitter. Τα δεδομένα συλλέγονται όμοια από το API του Twitter με το #sarcasme. Σε αντίθεση με τα παραπάνω επιλέγεται να διατηρηθούν τα κεφαλαία στα δεδομένα αφού μπορεί να προσδίδουν ένα είδος συναισθηματικής έντασης στο tweet. Τα χαρακτηριστικά δημιουργούνται με uni-gram και bi-gram μοντέλα ενώ χρησιμοποιείται ο ταξινομητής Balanced Window [58] που περιγράφεται στην εργασία Linguistic Classification System[2]. Μια διαφορετική οπτική χρησιμοποιείται από τον Rajadesingan [77] προσεγγίζοντας την αναγνώριση σαρκασμού από την σκοπιά της ανάλυσης συμπεριφοράς (Behavioral Analysis). Στην εργασία αυτή χρησιμοποιούνται χαρακτηριστικά που αναδεικνύουν την αντίθεση συναισθημάτων με το εργαλείο SentiStrength [93] αλλά και η εναλλαγή συναισθημάτων (θετικό - αρνητικό -θετικό) μέσα στο tweet. Ταυτόχρονα χρησιμοποιεί πληροφορίες για κάθε χρήστη όπως ο αριθμός των follower/following, το πλήθος retweet και tweet ώστε να εξάγει χαρακτηριστικά για το πόσο ενεργός χαρακτηρίζεται ο χρήστης στο Twitter. Επίσης διατηρώντας το ιστορικό του χρήστη δημιουργούνται μια σειρά από χαρακτηριστικά με βάση τα προηγούμενα tweet του χρήστη όπως το πλήθος λέξεων που χρησιμοποιεί κατά μέσο όρο αλλά και τον μέσο όρο των συναισθημάτων που χρησιμοποιεί στα tweet του. Δημιουργείτε έτσι με αυτή την μέθοδο ένα διάνυσμα χαρακτηριστικών που χαρακτηρίζει την συνολική συμπεριφορά του χρήστη το οποίο τροφοδοτείται σε ένα SVM. Αντίθετα, Βαθιά Νευρωνικά Δίκτυα χρησιμοποιούνται και από τον Kumar [53]. Συγκεκριμένα αποδίδει το σαρκαστικό περιεχόμενο των tweet με βάση τα αριθμητικά δεδομένα που περιέχει αλλά και την συντακτική του μορφή. Παρατηρείται πως ο σαρκασμός περιέχει έντονα αριθμητικά στοιχεία που τον διαφοροποιούν από την κυριολεξία. Παρόλα αυτά ο ταξινομητής που χρησιμοποιείται, που αποτελεί συνδυασμό LSTM και Συνελικτικών Νευρωνικά Δίκτυα(Convolutional Neural Networks), αποδίδει καλύτερα με την χρήση Word Embeddings. Οι Ghosh & Velae[34] χρησιμοποιούν επίσης βαθιά νευρωνικά δίκτυα για την αναγνώριση σαρκασμού στο Twitter. Συλλέγουν δεδομένα μέσω συγγενών hashtag που υποδεικνύουν σαρκασμό και χρησιμοποιούν τον συντακτικό αναλυτή του Stanford[3][21] αλλά και Bag Of Words για την δημιουργία χαρακτηριστικών στα tweet. Τα χαρακτηριστικά αυτά τροφοδοτούνται σε ένα βαθύ νευρωνικό δίκτυο που αποτελείται από ένα επίπεδο Embedding, δύο επίπεδα συνεκτικών δικτύων(CNN), δύο επίπεδα LSTM και στην έξοδο των οποίων συνδέεται ένα πυκνό νευρωνικό δίκτυο, αποδίδοντας πολύ καλά αποτελέσματα. Μια προσέγγιση βάση προτύπων γίνεται από τον Bouazizi [14] που βασίζεται σε συντακτικά πρότυπα αλλά και αλληλουχίες θετικά και αρνητικά φορτισμένων λέξεων. Επίσης λαμβάνονται υπόψιν η χρήση μη συνηθισμένων λέξεων αλλά και λέξεων που υποδηλώνουν χιούμορ. Συγκρίνονται τρείς ταξινομητές, Random Forest, Decision Trees, SVM, και παρατηρείται πως η χρήση Random

[2] http://www.phasar.cs.ru.nl/LCS/
[3] https://nlp.stanford.edu/software/lex-parser.shtml

Forest παρουσιάζει τα καλύτερα αποτελέσματα.

2.2 Αναγνώριση ειρωνείας

Η αναγνώριση ειρωνείας έχει χαρακτηριστεί από τις εργασίες του Reyes[80, 81] που αποτέλεσαν τις πρώτες ολοκληρωμένες μελέτες πάνω στα ΜΓΦ. Αρχικά προσεγγίστηκε η ειρωνεία ως ένα γλωσσικό φαινόμενο το οποίο περιέχει κάτι απροσδόκητο αλλά ταυτόχρονα ασαφές και αντιφατικό. Τα στοιχεία αυτά προσπάθησαν να προσεγγιστούν με την δημιουργία χαρακτηριστικών που καταμετρούν τις συναισθηματικές λέξεις σε συνδυασμό με λέξεις που προσδίδουν απροσδόκητο νόημα στο κείμενο. Η μελέτη του Reyes δεν διακρίνει την γλωσσολογική διαφορά μεταξύ ειρωνείας και σαρκασμού και τα δεδομένα αντλούνται από τα hastag #irony , #sarcasm ενώ χρησιμοποιεί Δένδρα Απόφασης για να ταξινομήσει τα δεδομένα αυτά. Στην μετέπειτα ερευνητική δουλεία του αυξάνει την διάσταση των χαρακτηριστικών δημιουργώντας πολυδιάστατα χαρακτηριστικά και εισάγει την έννοια της ¨ταυτότητας' του χρήστη με στοιχεία όπως τα σημεία στίξης, αντιφατικούς όρους και τα χρονικά επιρρήματα που χρησιμοποιούνται. Επιπλέον καταγράφετε ο παράγοντας του απροσδόκητου με τα αντιφατικά επιρρήματα ενώ προσδιορίζει και το στυλ γραφής λαμβάνοντας υπόψιν n-gram και skip-gram μοντέλα. Παρόμοια διαδικασία ακολουθείται και από τον Barbieri [5] με την διαφορά πως πλέον η ανίχνευση σαρκασμού γίνεται σε σύγκριση με tweet από συγκεκριμένα θέματα όπως η πολιτική, η εκπαίδευση και το χιούμορ. Συγκεκριμένα, αντλεί δεδομένα από το Twitter με βάση τα hashtag #sarcasm, #politics, #education, #humour και καταγράφει όμοια με το Reyes το απροσδόκητο των tweet. Αναλυτικά, χρησιμοποιεί σαν χαρακτηριστικά στοιχεία σπάνιων και συχνών λέξεων, εξετάζεται η χρήση λέξεων που χρησιμοποιούνται κυρίως στον προφορικό λόγο με βάση το American National Corpus Frequency Data[4]. Καταγράφεται επίσης η χρήση επιρρημάτων και επιθέτων, η μορφολογία του tweet, ο συναισθηματικός δείκτης του καθώς και πιθανές αμφισημίες του. Τα χαρακτηριστικά αυτά ταξινομεί με Random Forest και Δένδρα Απόφασης. Το στοιχείο του αναπάντεχου εισάγει και ο Buschmeir [17] από την οπτική της συναισθηματικής ανισορροπίας από την χρήση θετικά και αρνητικά ¨φορτισμένων' λέξεων. Επιπλέον, καταγράφει τον αριθμό από τα συνεχόμενα επιρρήματα του tweet και σε συνδυασμό με το Bag of Word δημιουργείται το σύνολο των χαρακτηριστικών της μεθόδου αυτής. Τα χαρακτηριστικά αυτά τροφοδοτούνται σε έναν αλγόριθμο Logistic Regression. Μια διαφορετική οπτική παρουσιάζει ο Huang [44] που χρησιμοποιεί Word Embeddings και αρχιτεκτονικές βαθιών νευρωνικών δικτύων. Συγκεκριμένα, παρουσιάζεται πως ένας μηχανισμός προσοχής(attention mechanism) σε Recurrent Neural Network παρουσιάζει τα καλύτερα αποτελέσματα σε σχέση με Συνελικτικά Νευρωνικά Δίκτυα. Μια προσπάθεια δημιουργίας προτύπων που χαρακτηρίζουν τα ειρωνικά σχόλια έγινε από τον Carvalho χρησιμοποιώντας n-grams και συνδυασμούς επιρρημάτων αλλά και αρκτικόλεξα που υποδηλώνουν χιούμορ για χαρακτηριστικά. Ταυτόχρονα χρησιμοποιήθηκαν μεγάλα κείμενα από την Wall Street Journal, υποσημειωμένα με μεταφορικές και κυριολεκτικές χρήσεις τις γλώσσας. Σε

[4] http://www.anc.org/data/anc-second-release/frequency-data/

ένα διαφορετικό μέσο κοινωνικής δικτύωσης, το Reddit[5], αναπτύχθηκε η ερευνητική δουλεία του Wallace όπου προτάθηκε η χρήση των συμφραζομένων για την ανίχνευση ειρωνείας. Λαμβάνετε υπόψιν το θέμα στο οποίο καταγράφηκε το σχόλιο, το συναίσθημα που περιέχεται αλλά και η ονοματική φράση του σχολίου. Το 2018 ο διεθνής διαγωνισμός Semantic Evaluation φέρει ως θέμα ενός διαγωνισμού την ανίχνευση ειρωνείας στο Αγγλικό Twitter[42]. Η ομάδα Thu-Ngn[101] που τερμάτισε στην πρώτη θέση του διαγωνισμού χρησιμοποιεί πλήρως συνδεδεμένα LSTM με προεκπαιδευμένα Word Embeddings, συντακτικά χαρακτηριστικά αλλά και συναισθηματικά στοιχεία του tweet μέσω του εργαλείου AffectiveTweet[67]. Όμοια, αμφίδρομα LSTM με την μορφή πλειοψηφικού συνδυασμού (voting ensemble) χρησιμοποιεί και η ομάδα Ntua-Slp τα οποία εκπαιδεύονται σε Word Embeddings. Η ομάδα WLV[83] χρησιμοποιεί επίσης πλειοψηφικό συνδυασμό των ταξινομητών Logistic Regression και SVM. Χρησιμοποιούνται ως χαρακτηριστικά Word-Emoji Embeddings αλλά και χαρακτηριστικά που αναδεικνύουν την συναισθηματική φόρτιση του tweet. Τέλος, στην τέταρτη θέση τερματίζει η ομάδα NLPRL-IITBHU[78] χρησιμοποιώντας έναν ταξινομητή XGBoost[6] με χαρακτηριστικά από το DeepMoji[7] αλλά και χαρακτηριστικά που βασίζονται στην αντίθεση και την δυσαρμονία των λέξεων.

2.3 Αναγνώριση Μεταφορικών Γλωσσικών Φαινομένων

Οι μελέτες πάνω στο σύνολο των μεταφορικών γλωσσικών φαινομένων είναι ελάχιστες και αυτό διότι η διαφορές τους είναι μικρές και μερικές φορές πολύ δύσκολα διαχωρίσιμες. Η πρώτη απόπειρα έγινε το 2009 από την Li[55, 89] για την αναγνώριση μη κυριολεκτικών φαινομένων με ημι-εποπτευόμενη μάθηση (semi-supervided learning). Χρησιμοποιήθηκε ένας ταξινομητής βασισμένος στις λεξιλογικές αλυσίδες αλλά και συνεκτικούς γράφους για την ταξινόμηση ιδιωμάτων από το Oxford Dictionary of Idiomatic English [25], εξετάζοντας την ύπαρξη συνεκτικότητας στις σχέσεις μεταξύ των λέξεων. Μετέπειτα προτάθηκε μια βελτίωση του μοντέλου αυτού που χρησιμοποιούσε ταξινομητή SVM στο δεύτερο στάδιο. Η Birke[12] δημιούργησε ένα μη επιβλεπόμενο ταξινομητή που υπολόγιζε την απόσταση ομοιότητας μια φράσης από ένα σύνολο κυριολεκτικών και μεταφορικών εκφράσεων. Η έρευνα αυτή αποσκοπούσε στην εύρεση των ρημάτων που χαρακτηρίζουν τις μη κυριολεκτικές εκφράσεις. Όμοια, μη επιβλεπόμενη μάθηση χρησιμοποιήθηκε και από την Bogdanova[13] όπου με τον k-means έγινε η προσπάθεια να ταξινομηθούν δεδομένα μεταφορικών γλωσσικών φαινομένων στα Ρωσικά, με βάση την απόσταση των λέξεων από τα συνήθη συμφραζόμενα τους. Η Feldman[31] προσεγγίζει την ανίχνευση μεταφορικών φαινομένων και αντλεί δεδομένα από προτάσεις του British National Corpus[8]. Το σύστημα βασίζεται στην Ανάλυση Κύριων Συνιστωσών(Principal Components Analysis-PCA) χρησιμοποιώντας όμως παράλληλα BoW και Tf-idf για την αναπαράσταση των δεδομένων. Το 2015 το διεθνές Workshop Semantic Evaluation - SemEval φιλοξένησε έναν θέμα με την ονομασία SemEval-2015 Task 11: Sentiment Analysis of Figurative Language

[5] https://www.reddit.com/
[6] https://xgboost.readthedocs.io/en/latest/parameter.html
[7] https://deepmoji.mit.edu/
[8] http://www.natcorp.ox.ac.uk/

in Twitter[35]. Το θέμα αυτό απαιτούσε την ταξινόμηση βάση συναισθήματος διάφορα από γλωσσικά φαινόμενα όπως η ειρωνεία, ο σαρκασμός και η μεταφορά. Τα δεδομένα προέρχονταν από το Twitter και ήταν επισημειωμένα με έναν ακέραιο αριθμό στην κλίμακα (-5,5) που υποδηλώνει το βαθμό θετικού-αρνητικού συναισθήματος του κάθε tweet. Η αξιοπιστία του ταξινομητή ελέγχεται με την μετρική ομοιότητας συνημίτονου (Cosine similarity) αλλά και με το μέσο τετραγωνικό σφάλμα (Mean Squared Error). Η ομάδα ClaC[70] τερμάτισε στην πρώτη θέση συνδυάζοντας τέσσερα λεξικά για την εξαγωγή χαρακτηριστικών από τα δεδομένα. Ταυτόχρονα χρησιμοποιήθηκαν οι συχνότητες εμφάνισης των όρων καθώς και χαρακτηριστικά του λόγου όπως τα emoji και τα μέρη του λόγου. Η ομάδα UPF[4], που τερμάτισε στην δεύτερη θέση, χρησιμοποίησε παλινδρόμηση, με τυχαίο δευτερεύον διάστημα, για την ταξινόμηση των χαρακτηριστικών. Για την εξαγωγή χαρακτηριστικών χρησιμοποιήθηκαν πόροι όπως το SentiWordNet[3], το DepecheMode[91] αλλά και το ANC[9]. Στην τρίτη θέση τερμάτισε η ομάδα LLT-PolyU[102] χρησιμοποιώντας ημι-εποπτευόμενη μάθηση. Τα χαρακτηριστικά εξάγονται από uni-gram και bi-gram μοντέλα ενώ γίνεται μια προσέγγιση στο συναίσθημα των tweet από την σκοπιά των λέξεων αλλά ταυτόχρονα λαμβάνεται υπόψιν και πιθανές αντιφάσεις λέξεων σε μικρή απόσταση. Τα χαρακτηριστικά ταξινομούνται με ένα regression μοντέλο Δέντρων Αποφάσεων. Έναν ταξινομητή βασιζόμενο στον SVM χρησιμοποιεί η ομάδα Elirf[37] που χρησιμοποιεί σαν χαρακτηριστικά n-grams και ένα Bag of Words μοντέλο του tf-idf συντελεστή για κάθε χαρακτηριστικό του n-gram. Επιπλέον, χρησιμοποιούνται λεξικά όπως το Affin[1], το Pattern[27] και του Jeffrey[10]. Τέλος, η ομάδα LT3 [96] χρησιμοποιεί έναν συνδυαστικό ταξινομητή, με ημι-εποπτευόμενη μάθηση, που αποτελείτε από μια συνάρτηση παλινδρόμησης και έναν SVM ταξινομητή. Τα χαρακτηριστικά αποτελούνται κυρίως από λεξιλογικά στοιχεία του κειμένου ενώ συνδυάζεται ένα σύστημα εξαγωγής ορολογίας. Ταυτόχρονα, αξιοποιούνται λεξικά όπως το WordNet[92] και το DBpedia[11]. Ο Rajani χρησιμοποίησε επίσης τα δεδομένα από παλαιότερο διαγωνισμό του SemEval[85] που περιέχουν ιδιώματα της Αγγλικής γλώσσας. Δημιουργεί ένα σύνολο χαρακτηριστικών με βάση το Bag of Words μοντέλο, προσθέτοντας χαρακτηριστικά του περιεχομένου τον λέξεων, όπως το WordNet:Similarity[72]. Ταυτόχρονα αξιοποιεί άλλο ένα λεξικό, το MRC Psycholinguistic Database Machine Usable Dictionary [100], σε μια προσπάθεια να καταγράψει λέξεις γενικού περιεχομένου αλλά και ειδικού σκοπού. Τέλος, τα χαρακτηριστικά αυτά ταξινομούνται με LIBLINEAR L2 regularized Logistic Regression (L2LR)[30]. Οι Piters&Wilks προσεγγίζουν το θέμα της αναγνώρισης μεταφορικής γλώσσας ως ένα πρόβλημα εξαγωγής σημασιολογικής συνήθειας στην χρήση της γλώσσας. Αξιοποιούν τόσο το λεξικό του WordNet[92] όσο και τα κείμενα του Semcor[16], που περιέχει επισημειωμένα σημασιολογικά χαρακτηριστικά που χρησιμοποιούνται για να εξάγουν την ομοιογένεια του κειμένου.

[9] http://www.anc.org/

[10] https://github.com/jeffreybreen/twitter-sentiment-analysis-tutorial-201107/tree/master/data/opinion-lexicon-English

[11] https://wiki.dbpedia.org/

Κεφάλαιο 3

Θεωρητικό υπόβαθρο

3.1 Μηχανική Μάθηση

Η Μηχανική Μάθηση (Machine Learning-ML) αποτελεί τομέα της Τεχνητής Νοημοσύνης (Artificial Intelligence-AI) και προσδοκεί την εκμάθηση υπολογιστικών μηχανών στην αυτόματη λήψη αποφάσεων. Ο κλάδος αυτό περιέχει στοιχεία τόσο από Στατιστική (Statistics) όσο από Θεωρία της Πληροφορίας (Information Theory) και την Γνωστική επιστήμη (Cognitive Science). Σκοπός του ML είναι η εκπαίδευση του υπολογιστή, με συγκεκριμένες μεθόδους και αλγορίθμους, έτσι ώστε κατά την λήψη αποφάσεων να επιτυγχάνει κάποιο συγκεκριμένο αποτέλεσμα. Το αποτέλεσμα αυτό ορίζεται ως απόδοση της μεθόδου ή του αλγορίθμου και αποτελεί μια μετρική που κρίνει την σωστή λειτουργία του. Βασικό κομμάτι του ML είναι τα δεδομένα από τα οποία αντλεί γνώση. Τα δεδομένα αποτελούν ως επί το πλείστων αριθμητικά δεδομένα τα οποία επεξεργάζονται και καταχωρούνται σε διανύσματα, τα οποία ονομάζονται διανύσματα χαρακτηριστικών (Feature Vectors), πριν τροφοδοτηθούν στον αλγόριθμο ML. Οι αλγόριθμοι ML χωρίζονται σε τρείς μεθόδους: Μάθηση με Επίβλεψη (Supervised Learning), Μάθηση χωρίς Επίβλεψη (Unsupervised Learning) και Ενισχυμένη Μάθηση (Reinforcement Learning). Παρακάτω θα γίνει μια προσπάθεια για την ανάλυση τους.

3.1.1 Μάθηση με Επίβλεψη

Κατά την Μάθηση με Επίβλεψη ο αλγόριθμος τροφοδοτείται με δεδομένα εκπαίδευσης (ή κατάρτισης, training data) τα οποία επισημειώνονται μια τιμή (label) που τα χαρακτηρίζει. Σε αυτό το είδος μάθησης τα δεδομένα που εισάγονται στον ML αλγόριθμο ονομάζονται δεδομένα εκπαίδευσης (Training Dataset) ενώ τα δεδομένα που χρησιμοποιούνται για την αξιολόγηση του αλγορίθμου ονομάζονται δεδομένα δοκιμής (Test Dataset). Τα προβλήματα Μάθησης με Επίβλεψη χωρίζονται σε:

- Προβλήματα παλινδρόμησης (Regression): Τα δεδομένα εισόδου του προβλήματος χαρακτηρίζονται από μια τιμή, συνήθως συνεχή, και ο αλγόριθμος αφού εκπαιδευτεί προσπαθεί να προβλέψει την τιμή αυτή για το κάθε νέο δεδομένο. Στην ουσία τα προβλήματα Παλινδρόμησης εστιάζουν την λειτουργία τους στην πρόβλεψη της εξόδου για κάθε

πρότυπο εισόδου. Χαρακτηριστικό πρόβλημα Παλινδρόμησης αποτελεί η πρόβλεψη των μετοχών του χρηματιστηρίου ή η πρόβλεψη του καιρού.

- Προβλήματα Ταξινόμησης (Classification): Σε αυτό το είδος Επιβλεπόμενης Μάθησης ο αλγόριθμος ML προσπαθεί να τοποθετήσει τα δείγματα των δεδομένων δοκιμής σε μια κατάλληλη κατηγορία. Τα δεδομένα στα προβλήματα ταξινόμησης φέρουν μια τιμή της διακριτής κατηγορίας ή κλάσης στην οποία ανήκουν, γεγονός που δικαιολογεί την ονομασία του προβλήματος, αφού η τιμή αυτή λειτουργεί ως 'Δάσκαλος' που επιβλέπει την διαδικασία της εκπαίδευσης του αλγορίθμου. Προβλήματα ταξινόμησης αποτελούν ο διαχωρισμός σκύλων και γατιών σε φωτογραφίες ή ο διαχωρισμός της Rock από την Jazz μουσική.

3.1.2 Μάθηση χωρίς Επίβλεψη

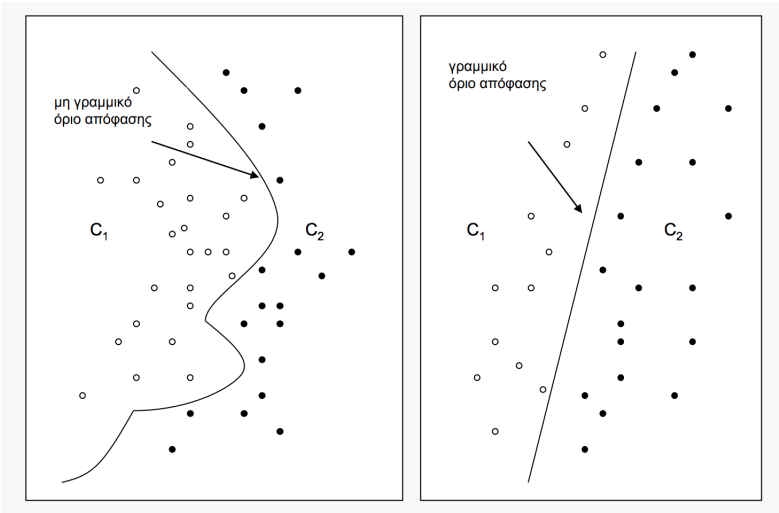
Σε αντίθεση με παραπάνω, στην συγκεκριμένη μέθοδο τα δεδομένα δεν διαθέτουν κάποια επισημειωμένη τιμή ώστε ο αλγόριθμος να εκπαιδευτεί από αυτή. Έτσι, τα προβλήματα χαρακτηρίζονται ως προβλήματα όμαδοποίησης' (Clustering) κατά τα οποία ο αλγόριθμος επιδιώκει να δημιουργήσει ομάδες ή συστάδες (Clusters) και να τοποθετήσει σε αυτές τα δεδομένα που παρουσιάζουν τις μεγαλύτερες, μεταξύ τους, ομοιότητες. Τις περισσότερες φορές το πρόβλημα αυτό δεν είναι σαφές αφού δεν γνωρίζουμε εκ των προτέρων το πλήθος των ομάδων που θέλουμε να δημιουργηθούν. Στην μέθοδο αυτή ανήκουν ένα σύνολο ιδιαίτερα σημαντικών αλγόριθμων, που εστιάζουν την λειτουργία τους στην παραγωγή δεδομένων, όπως οι γενετικοί αλγόριθμοι. Οι γενετικοί αλγόριθμοι χρησιμοποιούνται στην δημιουργία δεδομένων για την προσέγγιση συναρτήσεων και πήραν την ονομασία τους από γενετικούς μηχανισμούς κληρονόμησης τους οποίους, σε ένα βαθμό, προσομοιώνουν.

3.1.3 Ενισχυμένη Μάθηση

Η Ενισχυμένη μάθηση αποτελεί έναν συνδυασμό των δύο παραπάνω μεθόδων αφού παρότι ο αλγόριθμος τροφοδοτείται με δεδομένα τα οποία δεν χαρακτηρίζονται από κάποιο label διατίθεται ένα σύστημα τιμωρίας και ανταμοιβής (Punish and Reward Method) ώστε ο αλγόριθμος να εκπαιδεύεται στη σωστή κατεύθυνση. Η μέθοδος αυτή αποτελεί μια εξελίξιμη επιστημονική περιοχή και γίνεται ιδιαίτερα χρήσιμη διότι ο αλγόριθμος είναι σε θέση να αλληλοεπιδρά με το περιβάλλον του. Η μέθοδος Ενισχυμένης Μάθησης έχει χρησιμοποιηθεί ιδιαίτερα σε ηλεκτρονικά παιχνίδια όπως το σκάκι και έχει ιδιαίτερα υψηλές αποδόσεις.

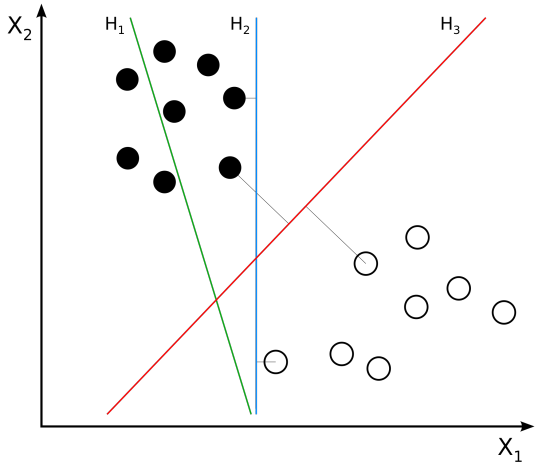
3.2 Μηχανές Διανυσματικής Υποστήριξης

Σε αυτό το σημείο είναι χρήσιμο να διατυπώσουμε το πρόβλημα γραμμικά διαχωρίσιμων δεδομένων για την βαθύτερη κατανόηση της λειτουργίας των Μηχανών Διανυσματικής Υποστήριξης. Ως γραμμικά διαχωρίσιμα δεδομένα ονομάζουμε τα δεδομένα για τα οποία υπάρχει ένα τουλάχιστον υπερεπίπεδο που διαχωρίζει πλήρως τις κατηγορίες που ανήκουν. Το υπερεπίπεδο αυτό, δηλαδή, πρέπει να 'αφήνει' τα πρότυπα της μιας κατηγορίας στο θετικό ημιχώρο



Σχήμα 3.1: Μη Γραμμικά(αριστερά) και Γραμμικά(δεξιά) Διαχωρίσιμα Δεδομένα

και της άλλης στον αρνητικό. Αντίθετα, τα σύνολα δεδομένων για τα οποία δεν υπάρχει διαχωριστικό υπερεπίπεδο ονομάζονται μη διαχωρίσιμα. Στα περισσότερα προβλήματα ταξινόμησης τα διανύσματα χαρακτηριστικών μεταξύ των κλάσεων δεν είναι γραμμικά διαχωρίσιμα δημιουργώντας προβλήματα στην απόδοση του ταξινομητή. Οι Μηχανές Διανυσματικής Υποστήριξης (Support Vector Machines-SVM) αποσκοπούν στην εύρεση του βέλτιστου υπερεπιπέδου διαχωρισμού με αποτέλεσμα να μπορούν να αποδίδουν αποτελεσματικά τόσο σε γραμμικά όσο και σε μη γραμμικά διαχωρίσιμα δεδομένα. Αναλυτικότερα, τα SVM αναζητούν ένα υπερεπίπεδο μέγιστου περιθωρίου διαχωρισμού των διανυσμάτων που ανήκουν σε διαφορετικές κλάσεις. Αυτό συνεπάγεται το διαχωριστικό υπερεπίπεδο να διέρχεται από σημεία που προσδίδουν το μέγιστο περιθώριο αποστάσεις μεταξύ των σημείων τις κάθε κλάσης. Η παραπάνω ιδιότητα μπορεί να γίνει αντιληπτή μέσω του σχήματος 3.2.



Σχήμα 3.2: Μηχανές Διανυσματικής Υποστήριξης (SVM)

Είναι προφανές πως η διαχωριστική ευθεία H_1 δεν διαχωρίζει πλήρως τα δεδομένα, όμως τόσο η H_2 όσο και η H_3 τα διαχωρίζουν. Παρόλα αυτά, η αξιοπιστία της ευθείας H_3 είναι

πολύ μεγαλύτερη αφού το περιθώριο (margin) που αφήνει από τις δύο κλάσεις είναι σαφώς μεγαλύτερο. Το βέλτιστό αυτό υπερεπίπεδο υπολογίζουν τα SVM αναζητώντας συντελεστές w, b που επιλύουν το παρακάτω πρόβλημα

$$\vec{w} \cdot \vec{x} - b = 0$$

διατηρώντας ταυτόχρονα τις συνθήκες:

$$\vec{w} \cdot \vec{x} - b = \begin{cases} 1, & \text{if} x \in Class_1 \\ -1, & \text{if} x \in Class_2 \end{cases} \tag{3.1}$$

που ονομάζονται αλλιώς φορείς υποστήριξης (support vectors) και η απόσταση τους είναι $\frac{2}{||w||}$. Μερικές φορές είναι επίσης χρήσιμο να μετασχηματίσουμε τα δεδομένα με κάποια μη γραμμική συνάρτηση δίνοντας την δυνατότητα στο SVM να ταξινομήσει αποτελεσματικότερα τα δεδομένα. Έτσι, για να αποκτηθούν μη γραμμικές σχέσεις στα δεδομένα χρησιμοποιούνται τα τρικ πυρήνα (kernel trick) που μετασχηματίζουν τα δεδομένα σε ένα μη γραμμικό χώρο. Με αυτόν τον τρόπο ο πυρήνας του SVM μέσω της συνάρτησης πυρήνα 3.2 μετασχηματίζει τα δεδομένα σε ένα χώρο που η εύρεση βέλτιστου υπερεπιπέδου είναι πολύ πιο απλή.

$$k(\vec{x}_i, \vec{x}_j) = \varphi(\vec{x}_i) \cdot \varphi(\vec{x}_j) \tag{3.2}$$

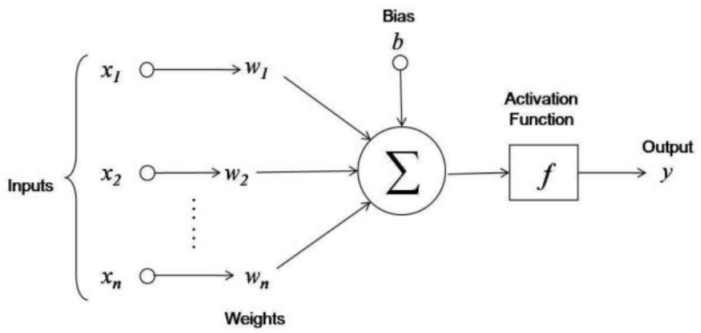
Μερικές από τις πιο διαδεδομένες συναρτσεις πυρήνα είναι:

- Ο πολυωνυμικός πυρήνας βαθμού d: $k(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \cdot \vec{x}_j + 1)^d$
- Ο Γκαουσιανός πυρήνας: $k(\vec{x}_i, \vec{x}_j) = \exp\left(-\frac{\|\vec{x}_i - \vec{x}_j\|^2}{2\sigma^2}\right)$
- Ο Πυρήνας Ακτινικής Βάσης (RBF) με παράμετρο γ: $k(\vec{x}_i, \vec{x}_j) = \exp\left(-\gamma \cdot \|\vec{x}_i - \vec{x}_j\|^2\right)$
- Ο Πυρήνας Υπερβολικής Εφαπτομένης με παράμετρο γ: $k(\vec{x}_i, \vec{x}_j) = tanh\left(\gamma(\vec{x}_i \cdot \vec{x}_j) + c\right)$

Ένα βασικό μειονέκτημα των SVM είναι η αδυναμία τους να αντιμετωπίσουν προβλήματα πολλών κλάσεων. Η βασική τεχνική που χρησιμοποιείται για να επεκταθεί η λειτουργία τους σε παραπάνω από δύο κατηγορίες ταξινόμησης ονομάζεται 'ένα εναντίων όλων', και δημιουργεί τόσα υπερεπεπίπεδα όσες και οι κατηγορίες διαχωρισμού. Παρόλα αυτά η μέθοδος αυτή, όπως γίνεται αντιληπτό, αντιμετωπίζει τόσα προβλήματα δυαδικής ταξινόμησης όσα και οι κατηγορίες του προβλήματος με αποτέλεσμα να αυξάνεται σε μεγάλο βαθμό την υπολογιστική πολυπλοκότητα της ταξινόμησης.

3.3 Τεχνητά Νευρωνικά Δίκτυα

Σε αυτό το τμήμα του κεφαλαίου παραθέτουμε τις βασικές αρχές λειτουργίες των τεχνιτών Νευρωνικών δικτύων(Artificial Neural Networks - ANN) ως αλγορίθμων ταξινόμησης. Συγκεκριμένα στο τμήμα αυτό αναλύονται βασικά μοντέλα ANNs όπως τα Multi Layer Perceptron, Βαθιά Νεωρωνικά Δίκτυα(Deep Neural Net works), Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (Long Short-Term Memory -LSTM).



Σχήμα 3.3: Διάταξη τεχνητού νευρώνα Perceptron

Τα ANNs δημιουργήθηκαν από τον άνθρωπο με σκοπό την επεξεργασία και την μετάδοση πληροφορίας βασισμένα στον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου. Η πρώτη απόπειρα για δημιουργία ενός τεχνητού νευρώνα έγινε από τους McCulloch και Pits το 1943[61] και αναπτύχθηκαν την δεκαετία του 1960 με κορύφωση το βιβλίο των Minsky και Papert το 1969 [65]. Σαφώς η λειτουργία του ανθρώπινου εγκεφάλου δεν είναι ούτε δυαδική ούτε σταθερή επομένως σε καμία περίπτωση δεν μπορεί να προσεγγιστεί επαρκώς από ANNs. Σε αυτή την λογική τα ANNs αποσκοπούν, κατά βάση, στην επίλυση προβλημάτων που δεν μπορούν να επιλυθούν από την παραδοσιακή υπολογιστική. Πολλές φορές είναι χρήσιμο να αντιμετωπίζουμε τα ANN ως κατευθυνόμενους γράφους με συνάψεις που διαθέτουν συναρτήσεις ενεργοποίησης. Η αρχιτεκτονική του ANN είναι άρρητα συνδεδεμένη με την εκπαίδευση του.

3.3.1 Τεχνητός Νευρώνας - Perceptron

Το βασικότερο στοιχείο των ANNs αποτελεί ο νευρώνας Perceptron ο οποίος περιεγράφηκε αρχικά από τον Rosenblatt το 1958[84]. Το Pereptron αποτελεί εμπρόσθια τροφοδοτούμενο (Feedforward) ANN. Στην γενική τους μορφή τα δίκτυα αυτά προσπαθούν να προσεγγίσουν μέσω της εξόδου τους μια συνάρτηση ή τιμή αναφοράς και ανήκουν στην κατηγορία γραμμικών ταξινομητών (linear classifiers). Η έξοδος ένος Feedforward δικτύου δίνεται από μια σχέση $y = f(x; \theta)$ με θ ένα σύνολο παραμέτρων της διάταξης.

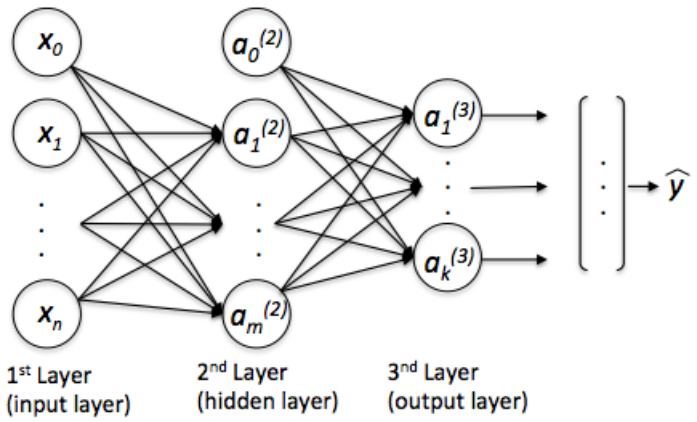
Το μοντέλο Perceptron δέχεται σαν είσοδο ένα διάνυσμα $x = [x_0, x_1, ..., x_n] \in \mathbb{R}^n$ και παράγει μια έξοδο $y \in \mathbb{R}$. Όπως μπορούμε να παρατηρήσουμε και από το Σχήμα 3.3 το διάνυσμα εισόδου πολλαπλασιάζεται με ένα διάνυσμα βαρών (βάρη σύνδεσης) $W \in \mathbb{R}^n$ και το αποτέλεσμα εισάγεται σε μια *μη γραμμική* συνάρτηση ενεργοποίησης (Activation Function) από την οποία παράγεται η έξοδος y του δικτύου. Αναλυτικά η έξοδος-πρόβλεψη του δικτύου δίνεται από την σχέση:

$$y = f(\sum_{i=0}^{n} x_i w_i + b) \tag{3.3}$$

με $f(.)$ την συνάρτηση ενεργοποίησης και b την πόλωση (bias) του δικτύου.

3.3.2 Πολυεπίπεδα Perceptron - Multy Layer Perceptron

Ο τεχνητός νευρώνας Perceptron αποτελεί έναν καθαρά *γραμμικό ταξινομητή* που δεν μπορεί να προσεγγίσει μη γραμμικά προβλήματα. Για την επίλυση μη γραμμικών προβλημάτων μπορούν να συνδυαστούν τεχνητοί νευρώνες και να δημιουργηθεί ένα πολυεπίπεδο δίκτυο νευρώνων Perceptron-Multy Layer Perceptron (MLP). Η δημιουργία πολυεπίπεδων νευρώνων σε συνδυασμό με την χρήση μη γραμμικών συναρτήσεων ενεργοποίησης μας φέρνει ένα βήμα πιο κοντά στην λειτουργία του ανθρώπινου εγκεφάλου που λειτουργεί με πολλά στρώματα διασυνδεδεμένων νευρώνων. Σε αυτή την αρχιτεκτονική κάθε νευρώνας συνδέεται με τους νευρώνες του προηγούμενου στρώματος και δεν υπάρχουν διασυνδέσεις μεταξύ νευρώνων του ίδιου στρώματος. Τα ενδιάμεσα από το αρχικό (εισόδου) στρώματα του MLP ονομάζονται *κρυφά στρώματα(Hidden Layers)* ενώ το τελικό στρώμα ονομάζεται *στρώμα εξόδου (Output Layer)*.



Σχήμα 3.4: Διάταξη Multy Layer Perceptron

Κάθε νευρώνας του κρυφού στρώματος αποτελεί ένα Perceptron έτσι η έξοδος του κρυφού στρώματος μπορεί να αναπαρασταθεί με συμπαγή τρόπο ως:

$$\boldsymbol{a}^{(j)} = \boldsymbol{W}\boldsymbol{y}^{(j-1)} + \boldsymbol{b} \tag{3.4}$$

με $W = [{w_1}^T, {w_2}^T, ..., {w_m}^T]^T$ τον πίνακα βαρών του στρώματος j που συνδέεται με τις εξόδους του στρώματος $j-1$ και $w_i = [w_{1i}, w_{2i}, ..., w_{Ni}]^T$ τα βάρη του νευρώνα Perceptron i στο στρώμα j. Η διάσταση του πίνακα W είναι $M \times N$ οπού M είναι το πλήθος των νευρώνων του στρώματος j και N το πλήθος των νευρώνων του στρώματος $j-1$.

Για να επιλύσουμε μη γραμμικά προβλήματα με την χρήση MLP είναι πλέον απαραίτητο ο αριθμός των νευρώνων εξόδου να ταυτίζεται με το πλήθος των κλάσεων-κατηγοριών του προβλήματος ταξινόμησης. Με βάση τα παραπάνω μπορούμε να υπολογίσουμε των **αριθμό των παραμέτρων** ενός MLP. Αν θεωρήσουμε ένα διάνυσμα εισόδου N_{in} διαστάσεων αυτό θα συνδέεται με το πρώτο στρώμα νευρώνων πλήθους N_1 και επομένως θα έχουμε $N_{in} \times N_1$

μεταβλητές. Με την ίδια λογική το δεύτερο στρώμα εξόδου θα δέχεται σαν είσοδο τις N_1 μεταβλητές του πρώτου στρώματος και θα υπολογίζει τον γραμμικό συνδυασμό τους με τις N_2 παραμέτρους του δεύτερου στρώματος, επομένως θα έχουμε $N_1 \times N_2$ μεταβλητές. Τέλος η ίδια λογική ισχύει και για το στρώμα εξόδου και αν θεωρήσουμε ότι αυτό αποτελείται από N_{out} νευρώνες τότε το σύνολο παραμέτρων του δικτύου υπολογίζεται ως:

$$card(MLP) = N_{in} \times N_1 + N_1 \times N_2 + ... + N_{n-1} \times N_{out} \tag{3.5}$$

Όταν το MLP αποτελείται από παραπάνω του ενός κρυφά στρώματα τότε αυτό ονομάζεται Βαθύ Νευρωνικό Δίκτυο (Deep Neural Network-DNN) και μπορεί να επιλύσει ακόμα πιο σύνθετα και μη γραμμικά προβλήματα.

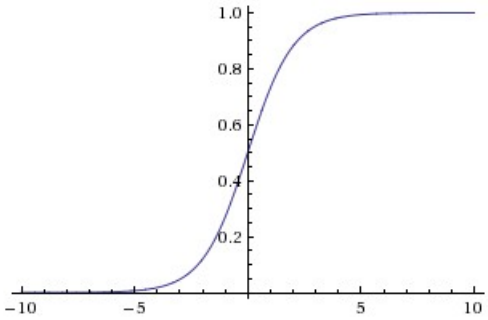
3.3.3 Εκπαίδευση Νευρωνικών Δικτύων

Είναι χρήσιμο να καθορίσουμε μερικά από τα βασικά χαρακτηριστικά που πρέπει να διέπουν ένα λειτουργικό και αποτελεσματικό νευρωνικό δίκτυο έτσι ώστε να γίνει αντιληπτή η σημασία της επιλογής κατάλληλων εργαλείων για την εκπαίδευση του. Ένα νευρωνικό δίκτυο πρέπει να έχει την δυνατότητα να περατώνει την εκπαίδευση σε πεπερασμένο χρόνο καταναλώνοντας όσο το δυνατόν λιγότερη υπολογιστική ισχύ, συνεπώς τόσο η επιλογή των ενδεδειγμένων εργαλείων όσο και η επιλογή του κατάλληλου πλήθους δεδομένων εκπαίδευσης είναι ιδιαίτερα σημαντική. Ταυτόχρονα πρέπει να διατηρεί την ιδιότητα του να γενικεύει δηλαδή να διατηρεί μικρό σφάλμα πρόβλεψης τόσο στα δεδομένα εκπαίδευσης όσο και στα δεδομένα ελέγχου. Η εκπαίδευση ενός νευρωνικού δικτύου αποτελεί μια επαναληπτική διαδικασία κατά την οποία οι παράμετροι του δικτύου προσαρμόζονται ώστε να έχουμε την επιθυμητή έξοδο/πρόβλεψη. Η κάθε επανάληψη της διαδικασία ονομάζεται εποχή (epoch) και το πλήθος των εποχών επηρεάζει σημαντικά τόσο την δυνατότητα κατηγοριοποίησης του μοντέλου όσο και την ικανότητα του να γενικεύει. Υπάρχουν δύο είδη εκπαίδευσης που διαχωρίζονται με βάση την ανανέωση των παραμέτρων του μοντέλου. Η πρώτη ονομάζεται **On-line Learning** και τα βάρη του δικτύου ανανεώνονται παράδειγμα με παράδειγμα από τα δεδομένα εκπαίδευσης. Αντίθετα στην εκπαίδευση με πακέτα (**Batch Learning**) τα βάρη του δικτύου ανανεώνονται μετά την είσοδο του συνόλου των δεδομένων εκπαίδευσης. Η διαδικασία της εκπαίδευσης λειτουργεί κυρίως με τον υπολογισμό διαφορών και παραγώγων ενώ εξαρτάται από διάφορες παραμέτρους όπως ο αριθμός των επαναλήψεων, η συνάρτηση ενεργοποίησης, η συνάρτηση κόστους και η συνάρτηση βελτιστοποίησης. Παρακάτω θα αναπτύξουμε μερικές από αυτές.

3.3.3.1 Συνάρτηση Ενεργοποίησης - Activation Function

Η συνάρτηση ενεργοποίησης είναι η βασική μη γραμμικότητα ενός νευρωνικού δικτύου και αποτελεί σημαντικό κομμάτι της εκμάθησης περίπλοκων προτύπων. Η συνάρτηση ενεργοποίησης στην ουσία αποτελεί έναν κόμβο 'απόφασης' που ενεργοποιείται και μεταφέρει το αποτέλεσμα στην έξοδο του δικτύου ή όχι ανάλογα με την τιμή που εισάγεται σε αυτόν. Μερικές από τις ποιο βασικές συναρτήσεις ενεργοποίησης παρουσιάζονται παρακάτω:

- *Σιγμοειδής Συνάρτηση (Sigmoid function)*: Αποτελεί μια από τις πρώτες συναρτήσεις που χρησιμοποιήθηκαν χρονικά. Η λειτουργία της εγγυείται στην αντιστοίχιση της τιμής εισόδου σε μια τιμή στο διάστημα (0,1) με την ιδιαιτερότητα όμως να μεταφέρει τις μικρότερες τιμές κοντά στο 0 και τις μεγαλύτερες κοντά στο 1, χωρίς όμως ποτέ να λαμβάνει τις τιμές αυτές. Το γεγονός πως οι μεγάλες και οι μικρές τιμές εισόδου κλιμακώνονται κοντά στο 0 και στο 1 αντίστοιχα, όπως φαίνεται και στο Σχήμα 3.5 μας οδηγεί σε πολύ μικρές τιμές κλίσης, ακόμα και μηδενικές, κάνοντας την διαδικασία της μάθησης αρκετά αργή ή ενδεχομένως να σταματήσει. Το φαινόμενο αυτό ονομάζεται εξασθένιση κλίσης(Vanishing Gradients) και μας δημιουργεί αρκετά προβλήματα στην διαδικασία μάθησης που θα αναπτύξουμε στην συνέχεια.

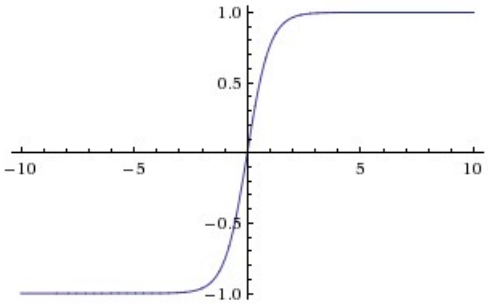


Σχήμα 3.5: Σιγμοειδής Συνάρτηση Ενεργοποίησης

Η σιγμοειδής συνάρτηση δίνεται από τον τύπο:

$$f(x) = \sigma(x) = \frac{1}{1 + e^{-x}} \tag{3.6}$$

- *Υπερβολική εφαπτομένη (Hyperbolic tangent)*: Όμοια με την σιγμοειδή η υπερβολική εφαπτομένη αντίστοιχη την είσοδο με βάση ένα κατώφλι στο διάστημα (-1,1).



Σχήμα 3.6: Υπερβολική Εφαπτομένη

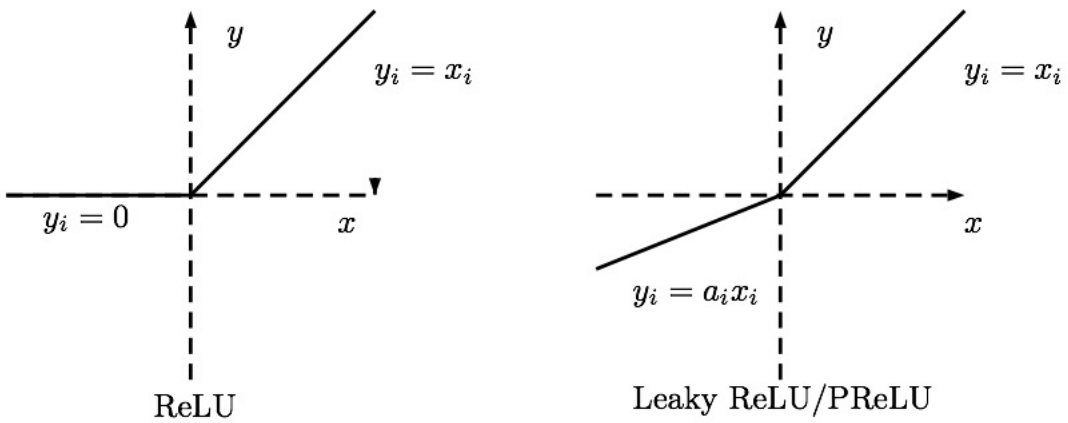
Το πλεονέκτημα της έναντι της σιγμοειδής συνάρτησης είναι δεν μεταβάλει αισθητά της τιμές που βρίσκονται κοντά στο 0 και επομένως βοηθούν τον επόμενο νευρώνα κατά την διαδικασία διάδοσης. Το παραπάνω πλεονέκτημα αποτελεί τον βασικό λόγο που χρησιμοποιείται σε επαναλαμβανόμενα νευρωνικά δίκτυα (Recurrent Neural Network).

Η υπερβολική εφαπτομένη υπολογίζεται ως :

$$f(x) = \tanh(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})} \tag{3.7}$$

Όπως μπορούμε να δούμε στο Σχήμα 3.6 οι μικρές τιμές μετατοπίζονται κοντά στο -1 ενώ οι μεγάλες στο 1. Όπως και στην περίπτωση της σιγμοειδής συνάρτησης αντιμετωπίζουμε προβλήματα Vanishing Gradient, σε πολύ μικρότερο όμως βαθμό.

- *Rectified Linear Unit (ReLU)*: Αποτελεί την πλέον ευρέως χρησιμοποιούμενη συνάρτηση ενεργοποίησης και όπως φαίνεται στο Σχήμα 3.7 οι τιμές εισόδου που είναι μικρότερες του μηδενός παύουν να λαμβάνουν μέρος στην διαδικασία της μάθησης. Ένα βασικό πρόβλημα της ReLu είναι πως για θετικές εισόδους δεν είναι οριοθετημένη με αποτέλεσμα για ιδιαίτερα μεγάλη είσοδο να λαμβάνουμε τεράστια έξοδο. Το γεγονός αυτό όμως υστερεί έναντι στο βασικό της πλεονέκτημα και την καθιστά βασική συνάρτηση ενεργοποίησης τόσο για MLP όσο και για DNN. Η συνάρτηση ενεργοποίησης



Σχήμα 3.7: Rectified Linear Unit (ReLU)

ReLU δίνεται από την εξίσωση:

$$f(x) = (0, max) \tag{3.8}$$

Η ReLu έχει ιδιαίτερα χαμηλή πυκνότητα (low sparsity) αφού 'νεκρώνει' το σύνολο των νευρώνων οι οποίοι έχουν αρνητικές τιμές. Το γεγονός αυτό είναι ιδιαίτερα χρήσιμο υπολογιστικά αφού μειώνει το χρόνο εκμάθησης του δικτύου κάνοντας το ιδιαίτερα αποδοτικό. Ταυτόχρονα όμως δημιουργούνται προβλήματα στους νευρώνες οι οποίοι έχουν 'νεκρωθεί' αφού η οριζόντια γραμμή για αρνητικές τιμές θα έχει σταθερή και μηδενική παράγωγο καθ' όλη την διαδικασία εκμάθησης, μη δίνοντας την δυνατότητα στον νευρώνα να αλλάξει τιμή. Το πρόβλημα αυτό αναφέρεται στην βιβλιογραφία ως 'dying ReLu' δηλώνοντας την αδυναμία της ReLu να επαναχρησιμοποιήσει κάποιο νευρώνα αφού λάβει μια αρνητική τιμή. Για να αντιμετωπιστεί το παραπάνω πρόβλημα χρησιμοποιούμε μια παραλλαγή της ReLU που αντικαθιστά την οριζόντια γραμμή των αρνητικών

τιμών με μια γραμμική συνάρτηση με πολύ μικρή κλίση ώστε να δίνει την δυνατότητα σε έναν νευρώνα που έχει λάβει αρνητική τιμή σε κάποιο διάστημα της μάθησης να επανέλθει(recover). Η παραλλαγή αυτή καλείται συνήθως ως Leaky ReLU και η τιμή της κλίσης αποτελεί παράμετρο της συνάρτησης.

- *Συνάρτηση Ενεργοποίησης (Softmax)*: Στην πραγματικότητα η συνάρτηση sofmax δεν αποτελεί κατ΄ ουσίαν συνάρτηση ενεργοποίησης αλλά αναφέρεται ως τέτοια βιβλιογραφικά. Εφαρμόζεται στο στρώμα εξόδου των περισσότερων ANN ανεξάρτητα από την συνάρτηση ενεργοποίησης που εφαρμόζεται στους νευρώνες του δικτύου. Οι νευρώνες στο στρώμα εξόδου θεωρητικά μπορούν να λάβουν οποιαδήποτε τιμή. Το γεγονός όμως πως αποτελούν πιθανότητες το τυχαίο δείγμα εισόδου να ανήκει σε μια από τις N-κλάσεις του προβλήματος δημιουργεί την ανάγκη να 'κανονικοποιήσουμε' της τιμές εξόδου έτσι ώστε να κατανεμηθούν στο διάστημα [0,1] και το άθροισμα τους να ισούται με 1. Στην ουσία θέλουμε η έξοδος κάθε νευρώνα του στρώματος εξόδου να ισούται με $\hat{y}_i = P(y = i|x)$. Η συνάρτηση sofmax εκτελεί ακριβώς αυτή την λειτουργία και δίνεται από τον τύπο:

$$Softmax(z_j) = \frac{e^{z_j}}{\sum_{j=0}^{N} e^{z_j}} \tag{3.9}$$

3.3.3.2 Συνάρτηση Κόστους - Cost Function

Η συνάρτηση κόστους αποσκοπεί στον έλεγχο της επαναληπτικής διαδικασίας εκπαίδευσης. Συνήθως την συμβολίζουμε με $J(\theta)$ και υπολογίζει πόσο κοντά βρίσκεται η έξοδος του δικτύου με την επιθυμητή τιμή για δεδομένες παραμέτρους. Η πιο γνωστή συνάρτηση κόστους χρησιμοποιεί την εντροπία και ονομάζεται **Απώλεια Διατροπικής Εντροπίας (Crossentropy Loss)**. Το κόστος υπολογίζεται πάνω στα βάρη-παραμέτρους του δικτύου όπως φαίνεται από την εξίσωση 3.10

$$J(\theta) = -H(y, p) = -\sum_{i=0}^{N} \hat{y}_i log(p_i j) \tag{3.10}$$

με N το συνολικό αριθμός κατηγοριών-κλάσεων των δεδομένων, y_i η εκτιμώμενη τιμή για την παρατήρηση i και $p_i j = p(\hat{y}_i = j|x)$ η *posterior* πιθανότητα το i δείγμα να ανήκει στην j κλάση. Η παραπάνω εξίσωση υπολογίζει το κόστος ενός δείγματος εισόδου επομένως για τον υπολογισμό του συνολικού κόστους των δεδομένων εισόδου αρκεί να υπολογίσουμε τον αριθμητικό μέσο όρο των επιμέρους σφαλμάτων. Η λειτουργία αυτής της συνάρτησης κόστους είναι η σύγκριση των δύο πιθανοτηκών κατανομών, της πρόβλεψης και της αναμενόμενης εξόδου. Συχνά στην βιβλιογραφία την συναντάμε ως Αρνητική Λογαριθμική Πιθανοφάνεια(Negative Logarithmic Likelihood).

Μια επίσης γνωστή συνάρτηση κόστους η οποία χρησιμοποιήθηκε στην μελέτη μας είναι το **Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error)** το οποίο υπολογίζει την τετραγωνική απόσταση μεταξύ της επιθυμητής τιμής και της πρόβλεψης όπως φαίνεται στην εξίσωση 3.11.

$$J(\theta) = \frac{1}{n}\sum_{i=0}^{M}(Y_i - \hat{y}_i)^2 \tag{3.11}$$

Όπως μπορεί να γίνει εύκολα αντιληπτό αυτή η συνάρτηση κόστους δεν υπολογίζει πιθανοτικές διαφορές αλλά καθαρά αριθμητικές. Το γεγονός αυτό, όπως μπορεί να γίνει εύκολα αντιληπτό, δημιουργεί αρκετά προβλήματα αφού οι κλάσεις ενός προβλήματος δεν αποτελούν απαραίτητα διατεταγμένους αριθμούς.

3.3.3.3 Αλγόριθμος Βελτιστοποίησης - Optimization Algorithm

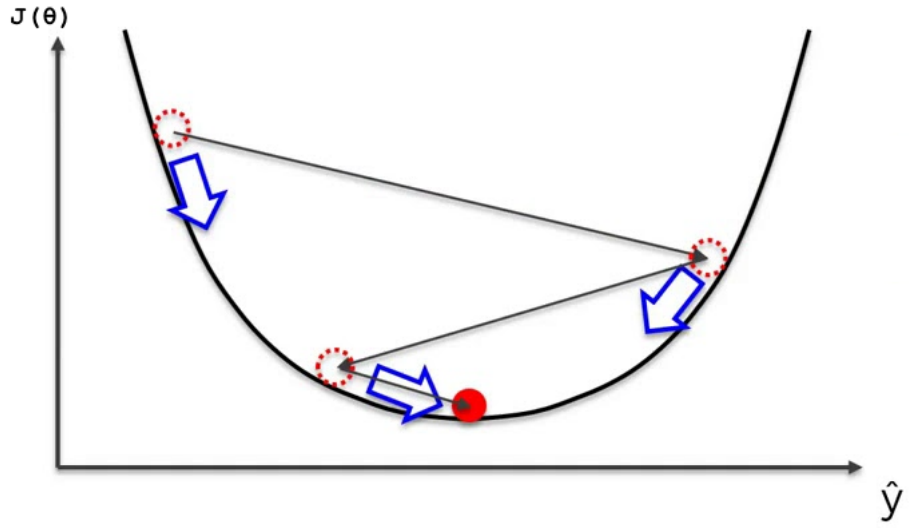
Υπάρχουν δύο κύριοι αλγόριθμοι βελτιστοποίησης οι οποίοι χρησιμοποιούνται ευρέως στα περισσότερα προβλήματα μηχανικής μάθησης. Ο πρώτος χρονικά είναι ο αλγόριθμος Οπίσθιας Ανατροφοδότησης Σφάλματος(Back Propagation Algorithm). Ο αλγόριθμος αυτός πρωτοδιατυπώθηκε την δεκαετία του 60, έλαβε την σύγχρονη μορφή του το 1970 από τον Φιλανδό φοιτήτη Linnainmaa[57] και πρώτο εφαρμόστηκε στα ANN το 1986 από τον Rumelhart [87]. Ο αλγόριθμος αυτός επαναυπολογίζει και προσαρμόζει τις τιμές βαρών του ANN με τον υπολογισμό της κλίσης του κόστους ως προς την κάθε παράμετρο του δικτύου. Η διαδικασία αυτή έπεται της εμπρόσθιας τροφοδότησης του δικτύου(Feed Forward) κατά την οποία τροφοδοτούμε το δικτύου με μια είσοδο x και υπολογίζουμε την προβλεπόμενη τιμή εξόδου y. Η προσαρμογή του κάθε βάρους γίνεται με την πρόσθεση της κλίσης της συνάρτησης κόστους ως προς το βάρος του νευρώνα k, $\frac{\partial J(\theta)}{\partial w_k}$. Η διαδικασία αυτή εκτελείται επαναληπτικά από την έξοδο του δικτύου προς την είσοδο και βασίζεται στον υπολογισμό μερικών παραγώγων με τον κανόνα της αλυσίδας. Το βασικό πλεονέκτημα του αλγορίθμου αυτού εγγυάται στον γρήγορο υπολογισμό του και στην άμεση αναπροσαρμογή τυχόν αυθαίρετης τιμής πόλωσης σε κάποιον νευρώνα.

Ένας άλλος, εξίσου διαδεδομένος, είναι ο **Στοχαστικός Αλγόριθμος Τάχιστης Κατάβασης (Stochastic Gradient Descent)**. Ανήκει στην μεγάλη κατηγορία αλγορίθμων Gradient Descent που βασίζουν την λειτουργία τους στην ελαχιστοποίηση ή μεγιστοποίηση μιας συνάρτησης κόστους με την χρήση της κλίσης(gradient) των παραμέτρων του προβλήματος. Ο αλγόριθμος αυτός εκτελείται επαναληπτικά μέχρι να επέλθει σύγκλιση ή να ολοκληρωθεί ένας ορισμένος αριθμός επαναλήψεων(Termination Criteria) και οι παράμετροι του δικτύου ανανεώνονται με βάση την παρακάτω εξίσωση:

$$\theta_{t+1} = \theta_t - \lambda \nabla_\theta J(\theta) \tag{3.12}$$

με θ το σύνολο των παραμέτρων του προβλήματος, λ τον ρυθμό μάθησης(learning rate) και $J(\theta)$ την συνάρτηση κόστους όπως αναλύσαμε παραπάνω. Στην πράξη χρησιμοποιείται με την μορφή μικρών πακέτων (Mini-batches) δεδομένων ώστε να έχει ταχύτερη σύγκληση και να μπορεί να εκτελεστεί παράλληλα σε όλους τους πυρήνες των υπολογιστικών συστημάτων. Έτσι κάθε επανάληψη του αλγορίθμου εκτελείται σε ένα τυχαίο δείγμα του συνόλου των δεδομένων εκπαίδευσης, προκαθορισμένου μεγέθους. Το μέγεθος αυτό υπολογίζεται εμπειρικά 16-64 δείγματα ανά επανάληψη. Ο αλγόριθμος αυτός έχει ορισμένα μειονεκτήματα που μας οδηγούν στην χρήση ορισμένων παραλλαγών του. Το γεγονός πως διατηρεί σταθερό τον ρυθμό

μάθησης λ καθώς και η πολύ αργή του σύγκληση σε προβλήματα με μεγάλο πλήθος δεδομένων λόγω της μεγάλης διασποράς της κλίσης(βλ. Σχήμα 3.8) οδήγησαν στην δημιουργία αλγορίθμων προσαρμοσμένων για προβλήματα μηχανικής μάθησης με μεγάλο πλήθος δεδομένων.



Σχήμα 3.8: Πρόβλημα Σύγκλισης SGD

Ένας από αυτούς είναι ο αλγόριθμος **Adam (adaptive moment estimation)**[50] που βασίζεται στην λειτουργία του κλασικού Stochastic Gradient Descent χρησιμοποιώντας όμως μεταβλητό ρυθμό μάθησης για κάθε παράμετρο-βάρος του δικτύου χρησιμοποιώντας τόσο την πρώτη όσο και την δεύτερη βαθμίδα κλίσης(First-Second Moment of Gradient) για την λειτουργία του. Ταυτόχρονα για να αποφύγει φαινόμενα ταλάντωσης όπως αυτό του Σχήματος 3.8 χρησιμοποιεί δυο μεταβλητές παράληψης (Forget Variables) που επιταχύνουν την διαδικασία σύγκλησης. Μεταβλητό ρυθμό μάθησης χρησιμοποιεί και ο αλγόριθμος **Adagrad** που ταυτόχρονα προσαρμόζει μεγάλες ενημερώσεις για σπάνιες παραμέτρους και μικρότερες για πιο συχνές. Το γεγονός αυτό κάνει τον αλγόριθμο ιδιαίτερα εύρωστο. Παρόλα αυτά παρουσιάζει σημαντική μείωση του ρυθμού μάθησης με την πάροδο των εποχών και η μάθηση γίνεται αδύνατη. Για την βελτίωση αυτού δημιουργήθηκε μια επέκταση του, ο **Adadelta** [103], που χρησιμοποιεί αποκλειστικά την δεύτερη βαθμίδα κλίσης ως ρυθμό μάθησης αλλά και τους προηγούμενους ρυθμούς μάθησης σε ένα ορισμένο χρονικό 'παράθυρο'. Έτσι ο ρυθμός μάθησης την τρέχουσα εποχή επηρεάζεται από τους ρυθμούς Δ προηγούμενους ρυθμούς σε αντίθεση με τον Adagrad που λάμβανε υπόψιν όλους τους προηγούμενους ρυθμούς μάθησης.

3.3.3.4 Κανονικοποίηση Δικτύου

Είναι σημαντικό σε αυτό το σημείο να καταγραφούν ορισμένες τεχνικές που χρησιμοποιούνται με σκοπό την μείωση του υπολογιστικού κόστους της εκπαίδευσης αλλά και της ικανότητα της γενίκευσης του μοντέλου.

Νόρμα L_2: Στο μεγαλύτερο μέρος των προβλημάτων μηχανική μάθησης κανονικοποιούμε τις παραμέτρους του μοντέλου με την L_2 νόρμα τους. Με την χρήση της κανονικοποίησης οι τιμές των παραμέτρων διατηρούνται κοντά στο 0 αποφεύγοντας μεγάλες τιμές και συνεπώς υψηλό υπολογιστικό κόστος. Ταυτόχρονα όπως αποδείχτηκε [39], η κανονικοποίηση οδηγεί σε τιμές παραμέτρων που μειώνουν αισθητά το κόστος και ταυτόχρονα τις αποτρέπει από το να λάβουν τιμές που δεν μειώνουν το κόστος. Με τον τρόπο αυτό μειώνεται η πιθανότητα ταλάντωσης του αλγορίθμου βελτιστοποίησης και έχουμε ταχύτερη σύγκληση. Η κανονικοποίηση L_2 γίνεται με την προσθήκη ενός τετραγωνικού όρου των παραμέτρων του δικτύου στη συνάρτηση κόστους όπως φαίνεται από την παρακάτω εξίσωση:

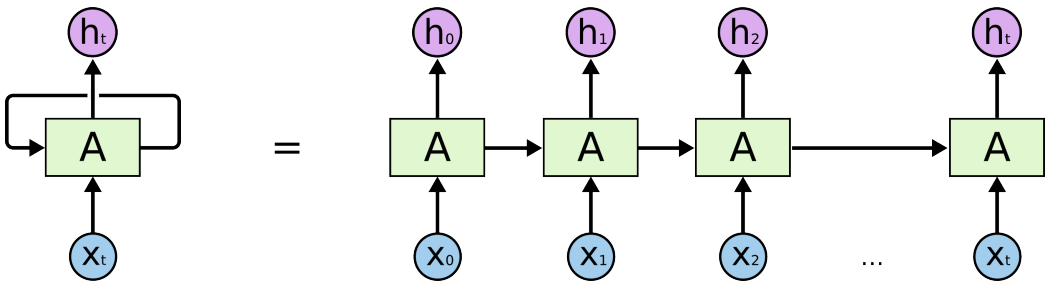
$$\boldsymbol{J'(\theta) = J(\theta) + \gamma \sum_i \theta_i} \tag{3.13}$$

όπου γ συντελεστής κανονικοποίησης που καθορίζει την επιρροή της κανονικοποίησης στη συνάρτηση κόστους.

Μια επίσης σημαντική τεχνική για την ιδιότητα των μοντέλων να γενικεύουν και να μην υπερπροσαρμόζονται (Overfit) στα δεδομένα εκπαίδευσης είναι ο τυχαίος μηδενισμός βαρών (**Dropout**) [90]. Η τεχνική αυτή μηδενίζει στοχαστικά ορισμένα από τα βάρη του δικτύου με σκοπό να 'νεκρώνονται' ορισμένοι νευρώνες του δικτύου κατά την εκπαίδευση. Αυτό αποσκοπεί στην εκπαίδευση νευρώνων σε ορισμένα χαρακτηριστικά ώστε η απόδοση του μοντέλου να είναι υψηλή για οποιοδήποτε δεδομένο εισόδου.

3.3.4 Επαναλαμβανόμενα Νευρωνικά Δίκτυα - Recurrent Neural Networks

Τα επαναλαμβανόμενα νευρωνικά δίκτυα αποτελούν μια κατηγορία Νευρωνικών Δικτύων που ειδικεύονται στα ακολουθιακά δεδομένα, όπως τα σήματα φωνής, δεδομένα κειμένων ή ακόμα και οπτικά δεδομένα με συνεχή κίνηση. Το βασικό χαρακτηριστικό τους είναι η ικανότητα να αποθηκεύουν πληροφορία για τις προηγούμενες καταστάσεις και να χρησιμοποιούν αυτή την πληροφορία (μνήμη) για την πρόβλεψη επόμενων καταστάσεων. Όπως φαίνεται και στο Σχήμα 3.9 η έξοδος του κάθε νευρώνα την χρονική στιγμή t εξαρτάται πλέον τόσο από την είσοδο x_t όσο και από την έξοδο του προηγούμενου νευρώνα, δηλαδή την h_{t-1}. Το γεγονός αυτό βοηθά το RNN να μπορεί να εξάγει αποτελέσματα με βάση τα συμφραζόμενα(context) που είναι ιδιαίτερα σημαντικό για ακολουθιακά δεδομένα. Όπως και στα παραδοσιακά ANN



Σχήμα 3.9: Επαναλαμβανόμενο Νευρωνικό Δίκτυο

έτσι και εδώ υπάρχει ένα σύνολο παραμέτρων θ τα οποία πρέπει να εκπαιδευτούν κατάλληλα με βάση τα δεδομένα. Σε αντίθεση όμως με τα ANN εκτός από τα βάρη W_i που συνδέουν την είσοδο Q_i με την έξοδο Y, υπάρχει και ένα διάνυσμα βαρών U_i το οποίο επιδρά πάνω στην έξοδο του προηγούμενου νευρώνα και συμβάλει με την σειρά του στον υπολογισμό της επόμενης κατάστασης. Οι εξισώσει που διέπουν την λειτουργία ενός RNN φαίνονται παρακάτω:

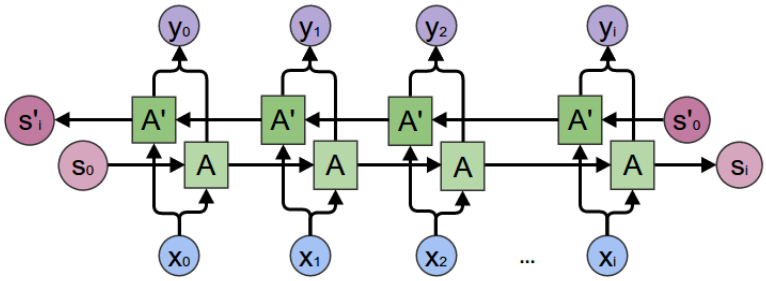
$$h_t = f_h(W_h x_t + U_h h_{t-1} + b_h) \tag{3.14}$$

$$y_t = f_y(W_y h_t + b_y) \tag{3.15}$$

με f_h, f_y τις συναρτήσεις ενεργοποίησης για τα h και y αντίστοιχα, W_h, W_y, U_h τα βάρη του δικτύου, x_t η είσοδος του δικτύου την χρονική στιγμή t, h_t η κρυφή έξοδος του δικτύου την χρονική στιγμή t και τέλος b_h, b_y οι τιμές πόλωσης (bias) για τις κρυφές κατάστασης και την έξοδο του δικτύου. Είναι εύκολο να γίνει κατανοητό πως για να δημιουργήσουμε βαθιές αρχιτεκτονικές αρκεί να στοιβάξουμε δύο ή περισσότερα RNN θέτοντας στην ουσία τις κρυφές καταστάσεις h_t ως είσοδο του επόμενου στρώματος. Τέλος αξίζει να σημειωθεί πως τα RNN έχουν την δυνατότητα να ανιχνεύουν χρονικές εξαρτήσεις και σε μη ακολουθιακά δεδομένα εισόδου.

3.3.4.1 Αμφίδρομα RNN

Μερικές φορές είναι χρήσιμο, αντί να προβλέπουμε μελλοντικές κατάστασης με βάση τις προηγούμενες εισόδους, να προβλέπουμε προγενέστερες καταστάσεις χρησιμοποιώντας τις μελλοντικές ανατρέχοντας τα δεδομένα προς τα πίσω. Σε επέκταση αυτού γεννήθηκε η ιδέα για την δημιουργία ενός νευρώνα ο οποίος θα λαμβάνει υπόψιν του και τις δύο πιθανές ροές τις πληροφορίας στα δεδομένα(ευθέως και αντίστροφα). Οι νευρώνες αυτοί ονομάστηκαν Αμφίδρομα Επαναλαμβανόμενα Νευρωνικά Δίκτυα (Bidirectional RNN) και η λειτουργία του φαίνεται στο Σχήμα 3.10.



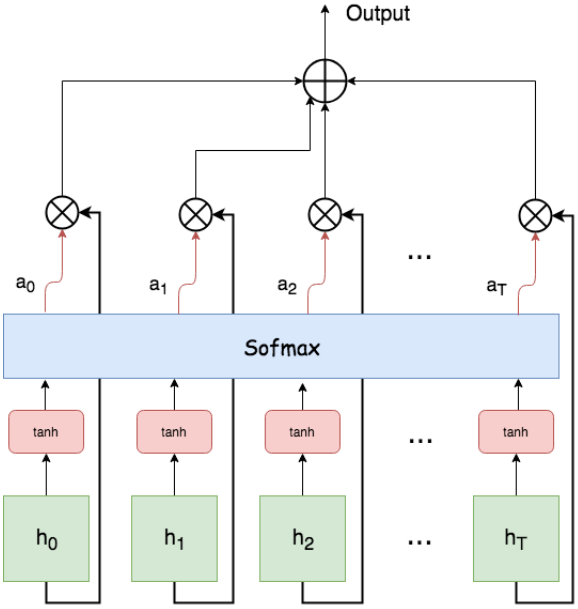
Σχήμα 3.10: Αμφίδρομο Επαναλαμβανόμενο Νευρωνικό Δίκτυο

Στην ουσία τα Αμφίδρομα Επαναλαμβανόμενα Νευρωνικά Δίκτυα αποτελούν συνδυασμό δύο RNN που επεξεργάζονται τα δεδομένα ανάποδα. Η κρυφή κατάσταση του δικτύου την χρονική στιγμή t αποτελεί την συνένωση των δύο κρυφών καταστάσεων, του ευθέως και του αντίστροφού: RNN:

$$h_t = \overrightarrow{h_t} || \overleftarrow{h_t} \tag{3.16}$$

3.3.4.2 Μηχανισμός Προσοχής (Attention Mechanism)

Ο Μηχανισμός προσοχής εφαρμόζεται πάνω στο ανώτερο στρώμα ενός RNN και εστιάζει στα σημαντικότερα τμήματα της ακολουθίας εισόδου αγνοώντας μέρη της ακολουθίας τα οποία δεν επιδρούν στην ταξινόμηση. Τα στοιχεία του διανύσματος α που προκύπτει από τον Μη-



Σχήμα 3.11: Μηχανισμός Προσοχής σε RNN

χανισμό Προσοχής αποτελούν τους συντελεστές επιρροής της κάθε κατάστασης στην τελική έξοδο και λόγω της χρήσης της συνάρτησης ενεργοποίησης $Softmax$ αθροίζουν στην μονάδα, δηλαδή $\sum_{t=0}^{T} \alpha_t = 1$. Αναλυτικότερα, ο μηχανισμός αυτός προσδίδει βάρη στις κρυφές καταστάσεις h_t ως εξής:

$$r_t = tanh(W_h h_t + b_t) \tag{3.17}$$

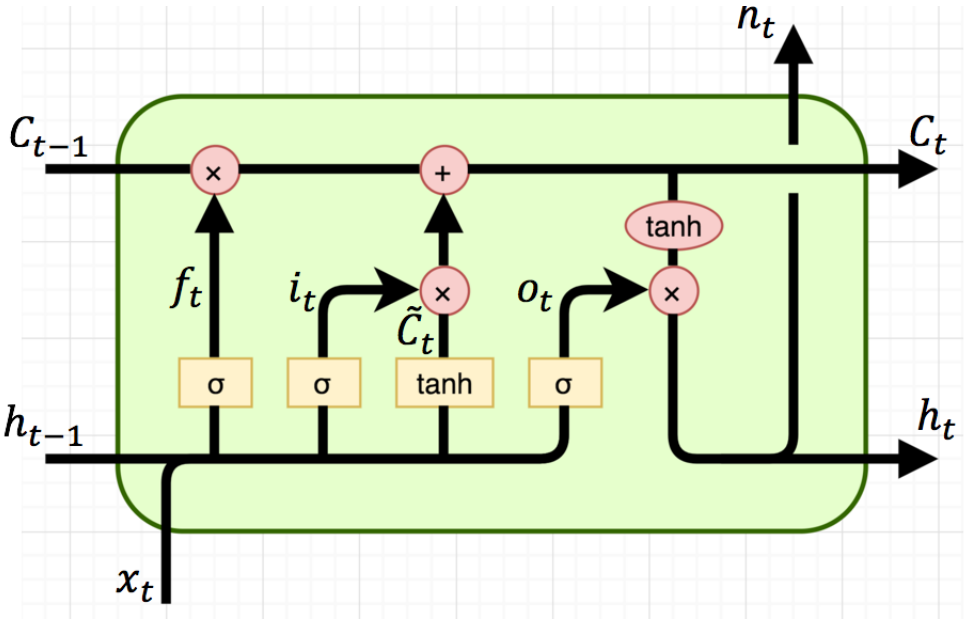
$$a_t = softmax(r_t) = \frac{e^{r_t}}{\sum_{j=0}^{T} e^{r_t}} \tag{3.18}$$

$$s = \sum_{t=0}^{T} a_t h_t \tag{3.19}$$

3.3.4.3 Μονάδα Μακράς Βραχυπρόθεσμης Μνήμης

Το βασικό μειονέκτημα των RNN είναι η αδυναμία τους να μοντελοποιήσουν χρονικές εξαρτήσεις μακράς διάρκειας. Οι αλγόριθμοι εκπαίδευσης χρησιμοποιούν παραγώγους για τον υπολογισμό και την ανανέωση των παραμέτρων του δικτύου όπως αναφέρθηκε στο κεφάλαιο 3.3.3. Οι κλίσεις όμως σε εξαρτήσεις μακράς διάρκειας αποτελούν συνθέσεις μερικών παραγώγων, σύμφωνα με τον κανόνα της αλυσίδας, που τείνουν να εξασθενούν τείνοντας στο μηδέν, και συνεπώς δεν η μάθηση των νευρώνων καθίσταται αδύνατη. Το φαινόμενο αυτό

ονομάζεται εξαφάνιση της κλίσης (Vanishing Gradients) και έχει σαν αποτέλεσμα την αδυναμία του μοντέλου να εκπαιδευτεί κατάλληλα. Το πρόβλημα αυτό μπορεί να αντιμετωπιστεί με την δημιουργία μια μονάδας Μακράς Βραχυπρόθεσμης Μνήμης (Long Short Term Memory). Το LSTM σε αντίθεση με τα κλασικά RNN, διαθέτει μια πύλη λήθης (forget gate) που δίνει την δυνατότητα στην μονάδα να αποκόπτει μικρές ή μεγάλες τιμές κλίσης ώστε να μην εμφανίζονται φαινόμενα εξαφάνισης ή έκρηξης κλίσης.



Σχήμα 3.12: Κύταρο LSTM

Ταυτόχρονα το κύτταρο LSTM έχει την δυνατότητα να διατηρεί την εξάρτηση απομακρυσμένων εισόδων τις ακολουθίας δημιουργώντας την ικανότητα να επιλύουμε προβλήματα με συμφραζόμενα (context). Το κύτταρο LSTM αποτελείται τις πύλες εισόδου (input gate, i_t), εξόδου (output gate, o_t) και λήθης (forget gate , f_t). Αναλυτικά όπως βλέπουμε και από το σχήμα 3.12 η έξοδος του κυττάρου υπολογίζεται ως εξής για την χρονική στιγμή t:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{3.20}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{3.21}$$

$$\tilde{C}_t = tanh(W_c[h_{t-1}, x_t] + b_c) \tag{3.22}$$

$$\boxed{C_t = f_t * C_{t-1} + i_t * \tilde{C}_t} \tag{3.23}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{3.24}$$

$$\boxed{h_t = o_t * tanh(C_t)} \tag{3.25}$$

με $\sigma(.)$ την σιγμοειδή συνάρτηση και β τις πολώσεις των πυλών. Μπορούμε να ερμηνεύσουμε την λειτουργία του κυττάρου με τα παρακάτω χαρακτηριστικά:

- Η θύρα εισόδου ελέγχει την ροή των διανυσμάτων εισόδου $[x_t, h_{t-1}]$ στο κύτταρο.
- Η θύρα λήθης ελέγχει την μετάβαση ή όχι του C_t στο επόμενο κύτταρο του δικτύου.

- Η κρυφή κατάσταση ενσωματώνει ή όχι την παρελθοντική πληροφορία του δικτύου στην τρέχουσα είσοδο.

Οι παραπάνω εξισώσεις αποτελούν μια από τις εκδοχές, την πιο γνωστή, των LSTM που χρησιμοποιούνται, όμως είναι σύνηθες να χρησιμοποιούνται διάφορες τροποποιημένες μορφές όπως αυτό των Schmidhuber & Gers[33].

3.3.5 Μάθηση Ensemble

Η δυνατότητα ένα μοντέλο να βελτιώσει τις δυνατότητες ταξινόμησης του θα μπορούσε να γίνει αν συνδυαστούν δύο ή περισσότερα μοντέλα ταυτόχρονα. Αυτό εκφράζει το Ensemble Learning, τον συνδυασμό ευφυών τεχνικών ώστε να αυξηθεί η απόδοση της ταξινόμησης. Η ιδέα προήλθε από το σκεπτικό πως δύο διαφορετικοί αλγόριθμοι μπορούν να ταξινομούν ένα δείγμα των δεδομένων. Διαθέτουν ιδιαίτερα θετικά χαρακτηριστικά που τους κάνουν δημοφιλείς. Ένα χαρακτηριστικό τους είναι η δυνατότητα να εκτελούνται παράλληλα οι αλγόριθμοι μηχανικής μάθησης και να εξετάζονται στα δοκιμαστικά δεδομένα ταυτόχρονα. Επίσης σπάνια αποδίδουν χαμηλότερα από τους μεμονωμένους αλγορίθμους που τους αποτελούν, χωρίς αυτό βέβαια να αποτελεί γενικό κανόνα. Οι τεχνικές Ensemble χωρίζονται, ανάλογα με τον τρόπο που λαμβάνει υπόψιν τους επιμέρους αλγόριθμους. Οι γενικές κατηγορίες είναι η **Σκληρή Ταξινόμηση (Hard Classification)** και η **Ανεκτική Ταξινόμηση (Soft Classification)**. Η πρώτη λειτουργεί με βάση την πλειοψηφία των ταξινομητών και ταξινομεί το κάθε δείγμα στην κλάση που έλαβε τις περισσότερες ψήφους (voting). Συνήθως αυτός ο συνδυασμός γίνεται με περιττό αριθμό αλγορίθμων ώστε να μην τεθεί θέμα ισοψηφίας και να χρειαστεί να επιλεγεί με άλλον τρόπο η κλάση ταξινόμησης. Για παράδειγμα αν ένα Ensemble μοντέλο περιλαμβάνει έστω τρείς ταξινομητές Α, Β, Γ που ταξινομούν το δείγμα s στις κλάσεις C_0, C_1, C_1 αντίστοιχα το δείγμα στο Ensemble μοντέλο θα ταξινομηθεί στην κλάση C_1 που έλαβε δύο ψήφους σε αντίθεση με την C_0 που έλαβε μια ψήφο. Σε περίπτωση ισοψηφίας υπάρχουν διάφοροι τρόποι που επιλέγεται η κλάση ταξινόμησης είτε τυχαία, δηλαδή με την επιλογή μιας κλάσης από τις ισοψηφείς στην τύχη, είτε στρατηγικά δηλαδή με την επιλογή ενός ταξινομητή που θα καθορίζει το αποτέλεσμα σε αυτή την περίπτωση. Αντίθετα με την χρήση Soft Classification μπορούμε να επιλέξουμε το ποσοστό επιρροής του κάθε ταξινομητή στην τελική ταξινόμηση. Για παράδειγμα μπορεί να ορισθεί πως η ψήφος του ταξινομητή Α είναι πιο ισχυρή από τις ψήφους των δύο άλλων ταξινομητών, σε συνέχεια του προηγούμενου παραδείγματος, με έναν συντελεστή $W_A = 3$ (θεωρώντας τους άλλους συντελεστές W_B και W_Γ μοναδιαίους) επομένως σε αυτή την περίπτωση το δείγμα s θα ταξινομηθεί στην κλάση C_0 που έλαβε τρείς ψήφους από τον ταξινομητή Α. Μερικές φορές όμως είναι πολύ πιο χρήσιμο να δίνουμε σημασία και στην απόδοση του ταξινομητή στο κάθε δείγμα. Για την εσωτερική αξιολόγηση του ο ταξινομητής χρησιμοποιεί ένα διάνυσμα $\hat{p}$ μήκους D, όπου D ο αριθμός των πιθανών κλάσεων του προβλήματος. Το διάνυσμά αυτό αποτελείται από τις D-πιθανότητες $p(s|C_i), i = 0, 2, ..., D-1$, δηλαδή τις πιθανότητες το δείγμα s να ανήκει σε κάθε μια από τις D κλάσεις. Το διάνυσμα αυτό ονομάζεται διάνυσμα με πιθανότητες εμπιστοσύνης (Confidence Score Vector). Ο ταξινομητής επιλέγει να τα ταξινομήσει το κάθε στοιχείο στην κλάση

που έχει την μεγαλύτερη πιθανότητα στο διάνυσμα. Όταν όμως χρησιμοποιούμε μια Ensemble μέθοδο, συνδυάζοντας αρκετούς ταξινομητές είναι χρήσιμο συνδυάζουμε, με κατάλληλους συντελεστές, τα διανύσματα εμπιστοσύνης ώστε να έχουμε ένα αθροιστικό διάνυσμα εμπιστοσύνης με το δείγμα να ταξινομείται στην κλάση με την μεγαλύτερη πιθανότητα. Η παραπάνω διαδικασία μπορεί να γίνει κατανοητή με την χρήση ενός παραδείγματος. Έστω και πάλι οι τρείς ταξινομητές Α, Β, Γ και τα διανύσματα εμπιστοσύνης τους:

- P_A = [0.2, 0.6, 0.2]
- P_B = [0.4, 0.3, 0.3]
- P_Γ = [0.3, 0.4, 0.3]

Αν θεωρήσουμε πως ο ταξινομητής Α έχει βάρος $W_A = 0.3$, ο Β $W_B = 0.4$ και ο Γ $W_\Gamma = 0.3$ τότε το συνολικό διάνυσμα εμπιστοσύνης θα είναι:

$$R_{ens} = \frac{1}{3}[0.3*0.2+0.4*0.4+0.3*0.3, 0.3*0.6+0.4*0.3+0.3*0.3, 0.3*0.2+0.4*0.3+0.3*0.3] = \frac{1}{3}[0.22, 0.42, 0.27]$$

και το δείγμα θα ταξινομηθεί στην κλάση C_1 αφού αυτή λαμβάνει το μεγαλύτερο σκορ εμπιστοσύνης.

3.4 Επεξεργασία Φυσικής Γλώσσας

Η Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing) αποτελεί πεδίο της Επιστήμης των Υπολογιστών (Computer Science) αλλά και της Τεχνητής Νοημοσύνης με βασικό ερευνητικό ενδιαφέρον, την κατανόηση, την παραγωγή αλλά και την επεξεργασίας της φυσικής γλώσσας. Το 1950 ο Turing πρωτοασχολήθηκε [95] με την δημιουργία υπολογιστών κατάλληλων να επεξεργαστούν δεδομένα κειμένου, ενώ το 1954 ξεκίνησε η προσπάθεια δημιουργίας μεταφραστικών μηχανών, χωρίς όμως να έχουν ιδιαίτερη επιτυχία. Το 1964 έμελλε να είναι μια ημερομηνία σταθμός για την επεξεργασία φυσικής γλώσσας αφού δημιουργήθηκε μια μηχανή, η ELIZA [99], που μπορούσε να αλληλοεπιδρά με τον άνθρωπο, κάνοντας λογικές ερωτήσεις ανάλογα με τις απαντήσεις του ανθρώπου. Η ELIZA αποτέλεσε δημιούργημα του εργαστήριο τεχνητής νοημοσύνης του MIT σε συνεργασία με ψυχολόγους που θεώρησαν την δημιουργία μιας μηχανής ικανής να επικοινωνήσει με τον άνθρωπο ιδιαίτερα σημαντική για ανθρώπους που έπασχαν από κατάθλιψη. Η μεγαλύτερη ανάπτυξη του κλάδου της Επεξεργασίας Φυσική Γλώσσας ήρθε την δεκαετία του 1980 κατά την οποία υπήρξε μεγάλη ανάπτυξη σε διάφορους τομείς όπως η Κατάταξη Σημείων του Λόγου (Part-of-Speech Tagging), η Δημιουργία Κειμένου (Text Generation), η Μετάφραση Κειμένου (Machine Translation) αλλά και η δημιουργία Οντολογιών (Ontology). Τα βασικότερα ερευνητικά ενδιαφέροντα της Επεξεργασίας Φυσικής Γλώσσας συνοψίζονται παρακάτω:

- Παραγωγή Κειμένου (Text Generation): για την δημιουργία ερωτήσεων/απαντήσεων σε φόρμες διαλογικών συστημάτων.

- Περίληψη Κειμένου (Text Summarization): για την δημιουργία σύντομων περιλήψεων δεδομένων κειμένου, όπως τα βιβλία.

- Μηχανική Μετάφραση (Machine Translation): για την δημιουργία αυτόματων μεταφραστών από μια γλώσσα σε μια άλλη.

- Εξόρυξη Κειμένου και Ανάκτηση Πληροφορίας (Text Mining and Information Retrieval: για την δημιουργία μηχανών που επεξεργάζονται συλλογές κειμένων με σκοπό την εξαγωγή χρήσιμων πληροφοριών από αυτές (text mining).

- Γλωσσολογική και Συντακτική Ανάλυση (Language and Syntactic Analysis): για την γραμματική αλλά και συντακτική ανάλυση μιας πρότασης ή την δημιουργία συντακτικών δέντρων για την απόδοση της πληροφορίας μιας πρότασης.

- Σημασιολογική και Πραγματολογική Ανάλυση (Semantics and Pragmatics Analysis): για την ανάλυση των λέξεων της πρότασης που περιέχουν την μεγαλύτερη πληροφορία και για την ανάλυσης του περιεχομένου και του νοήματος της πρότασης αντίστοιχα.

Σαφώς τα παραπάνω μπορούν να συνδυαστούν και δεν αποτελούν πλήρως ανεξάρτητους τομείς έρευνας.

3.4.1 Τεχνικές Επεξεργασίας Κειμένου

Τα δεδομένα κειμένου συνήθως περιέχουν πληροφορίες η οποίες για να μπορέσουν να χρησιμοποιηθούν από αλγορίθμους Μηχανικής Μάθησης πρέπει πρώτα να επεξεργαστούν. Η επεξεργασία, ή προεπεξεργασία όπως συχνά αναφέρεται, των δεδομένων οφείλει να διατηρεί το νοηματικό περιεχόμενο του κειμένου και να απαλείφει τυχόν θορύβους που εμπεριέχονται σε αυτό. Υπάρχουν ορισμένες τεχνικές για την επεξεργασία κειμένου οι οποίες αναπτύσσονται συνοπτικά παρακάτω:

- Έλεγχος ορθογραφίας (Spell Checker): Συνήθως πριν πραγματοποιήσουμε οποιαδήποτε άλλη επεξεργασία στο κείμενο ελέγχουμε την ορθότητα στην ορθογραφία των λέξεων που χρησιμοποιεί. Ο έλεγχος ορθογραφία γίνεται με την χρήση μεγάλων λεξικών με τα οποία συγκρίνονται οι λέξεις του κειμένου. Συνήθως χρησιμοποιείται η απόσταση Levenshtein [54], ή αλλιώς Ελάχιστη Απόσταση Επεξεργασίας (Minimum Edit Distance) η οποία αντικαθιστά την λέξη του κειμένου με την λέξη του λεξικού που απέχει την μικρότερη απόσταση. Η απόσταση αυτή λαμβάνει υπόψιν της, με κατάλληλες μετρικές-ποινής (penalty metrics), την πιθανή διαγραφή, εισαγωγή και αντιστροφή χαρακτήρων για την διόρθωση της λέξης και υπολογίζει με βάση τις ποινές που υποβλήθηκαν την απόσταση των δύο λέξεων.

- Διαγραφή Stop-Words: Έχει αποδειχθεί πως συνδετικές λέξεις όπως τα the, is, which, at κλπ δεν προσδίδουν πληροφορίες στο κείμενο και μπορούν να θεωρηθούν ως θόρυβος για ένα σύστημα Μηχανικής Μάθησης. Το γεγονός αυτό διατυπώθηκε από τον Hans

Luhn [60] ενώ ο Savoy [28] παρέθεσε μια εκτενή λίστα με Stop-Words. Διαγράφοντας τις λέξεις αυτές μειώνουμε τις διαστάσεις του κειμένου ώστε να γίνει πιο λειτουργικό.

- Stemming: Όπως ορίζει και η ονομασία του η διαδικασία του Stemming κατά την οποία η λέξεις μετατρέπονται στις αντίστοιχες ρίζες τους (stems). Η διαδικασία αυτή μπορεί να γίνει με διάφορους τρόπους, είτε με την διαγραφή της κλητικής κατάληξης είτε με την απομάκρυνση των προσφυμάτων (suffix). Για παράδειγμα η λέξη closing θα μετατραπεί σε close ενώ το amusment σε amus. Το 1968 δημιουργείτε μια πρώτη προσπάθεια για την δημιουργία αλγορίθμου που κάνει stemming από τον Julie Beth Lovins [59]. Το 1980 ο Μαρτιν Πορτερ δημιουργεί τον Porter Stemmer[76] που αποτελεί τον πιο διαδεδομένο αλγόριθμο Stemming μέχρι και σήμερα. Είναι ένας αλγόριθμος για την αγγλική γλώσσα και αποτελείται από τα παρακάτω βήματα:

Αλγόριθμος Porter Stemmer	
1:	Διαγραφή κατάληξης -ing, -ed
2:	Μετατροπή τελικού -y σε -i αν υπάρχει άλλο φωνήεν
3:	Μετατροπή διπλών καταλήξεων σε μονές
4:	Διαγραφή των -ic-, -full, -ness αν υπάρχουν
5:	Διαγραφή των -ant, -ence αν υπάρχουν
6:	Διαγραφή τελικού -e

- Lemmatizing: Αντίστοιχα με το stemming ο αλγόριθμος Lemmatizing μετατρέπει τις λέξεις στα λήμμα τα τους. Σε αντίθεση με την διαγραφή καταλήξεων που κάνει ο stemmer, ο Lemmatizer αντικαθιστά της λέξεις με την βάση τους. Για παράδειγμα το am που δεν θα υποστεί καμία αλλαγή με τον stemmer ο Lemmatizer θα το μετατρέψει σε be. Για τον λόγο αυτό απαιτούνται ειδικά λεξικά τα οποία αντιστοιχούν τις λέξεις στις βάσεις τους. Ένα ιδιαίτερα γνωστό είναι το WordNet Lemmatizer [1] το οποίο περιέχει ένα τεράστιο πλήθος αντιστοιχίσεων.

- Part-of-Speech (POS) Tagger: Ένας POS tagger είναι ιδιαίτερα χρήσιμος σε προβλήματα λεξιλογικού περιεχομένου αφού προσθέτει ετικέτες στις λέξεις του κειμένου ανάλογα με το μέρος του λόγου στο οποίο ανήκουν. Συχνά αποκαλείτε και ως grammatical tagger αφού στην ουσία κάνει γραμματική ανάλυση των προτάσεων για να καταλήξει στη μορφολογική κατηγορία της λέξης. Οι αλγόριθμοι αυτοί μπορούν να χωριστούν σε δύο μεγάλες κατηγορίες ανάλογα με τον τρόπο που επισημαίνουν τις λέξεις: τους Στοχαστικούς και τους βασισμένους σε Κανόνες (Rule-Based). Οι POS taggers αποτέλεσαν ενδιαφέρον έρευνας στον τομέα της υπολογιστικής γλωσσολογίας (computation linguistics) από το 1960 αφού στην ουσία θα εκτελούσαν, μέσω αλγορίθμων, γραμματική ανάλυση του κείμενου κάτι που μέχρι τότε εκτελούσαν γλωσσολόγοι με το χέρι. Το 1960 δημιουργείται το γνωστό Brown Corpus, από το Brown University, το οποίο χρησιμοποιείται μέχρι και σήμερα και αποτελείται από 1.000.000 λέξεις. Το 1980

[1] https://wordnet.princeton.edu/

3.5 Μηχανική Μάθηση σε Κείμενο

Οι αλγόριθμοι Μηχανικής μάθησης είναι δομημένοι ώστε να κατατάσσουν και να επεξεργάζονται αριθμητικά δεδομένα. Για να εφαρμόσουμε τους αλγορίθμους Μηχανικής Μάθησης σε δεδομένα που έχουν την μορφή κειμένου πρέπει με κάποιον τρόπο αυτά να μετασχηματιστούν σε αριθμητικά δεδομένα, ως διάνυσμα χαρακτηριστικών (Feature Vector), τα οποία θα τεθούν ως είσοδος στον εκάστοτε αλγόριθμο. Δημιουργήθηκε έτσι η ανάγκη για έναν συμπαγή τρόπο αναπαράστασης των λέξεων και των χαρακτήρων σε αριθμούς, είτε με την μορφή αραιών πινάκων (Sparse Vectors) είτε με την χρήση Clustering. Η δημιουργία του διανύσματος χαρακτηριστικών γίνεται κυρίως με την χρήση δύο μοντέλων, τα distributed και distributional. Τα distributed μοντέλα χρησιμοποιούν ANN για την πρόβλεψη και την δημιουργία του διανύσματος χαρακτηριστικών. Αντίθετα, τα distributional μοντέλα αποτελούν 'Count' μοντέλα αφού το βασικό τους χαρακτηριστικό είναι η καταμέτρηση του πλήθους συνεμφανίσεων των λέξεων. Στην πράξη τα distributed μοντέλα τείνουν να αποδίδουν αρκετά καλύτερα [6]. Παρακάτω αναλύονται μερικές από τις θεωρίες που αναπτύχθηκαν στον τομέα της Φυσική Επεξεργασία Γλώσσας για την αναπαράσταση δεδομένων κειμένου.

3.5.1 Σύνολο από Λέξεις (Bag of Words)

Η πιο απλή τεχνική για την δημιουργία διανύσματος χαρακτηριστικών μέσω του Συνόλου από Λέξεις (BoW). Το μοντέλο BoW βασίζεται σε ένα αραιό διάνυσμα (sparse vector) αναπαράστασης της κάθε λέξης, με το μήκος του διανύσματος να ορίζεται όσο και το μέγεθος του λεξικού. Με αυτή την προσέγγιση κάθε λέξη του λεξικού λαμβάνει έναν προσδιοριστικό αύξοντα αριθμό id, το αντίστοιχο sparse vector της λέξης να λάμβανε τιμή 1 στην θέση id και τιμή 0 σε οποιαδήποτε άλλη. Η κωδικοποίηση αυτή ονομάζεται One-hot encoding και η ονομασία προέρχεται από την τιμή 1 που λαμβάνει το διάνυσμα της κάθε λέξης. Οι λέξεις του κειμένου αναπαρίστανται με ένα μοναδικό αριθμό, χωρίς να λαμβάνεται υπόψιν η σειρά τους ή η εξάρτηση τους από τις γειτονικές τους λέξεις. Το γεγονός ότι δεν επηρεάζεται από συμφραζόμενα κάνει την δημιουργία χαρακτηριστικών ιδιαίτερα απλή με ελάχιστο υπολογιστικό κόστος, που τον κάνει και ιδιαίτερα δημοφιλή. Η πρώτη αναφορά σε ένα μοντέλο που δημιουργίας χαρακτηριστικών από δεδομένα κειμένου έγινε από τον Harris το 1954 [40] και αποτελεί την βάση λειτουργίας του BoW. Σε αυτό το σημείο είναι χρήσιμο να ορίσουμε το Uni-gram μοντέλο κατά το οποίο μια λέξη μέσα σε ένα κείμενο αποτελεί αυθαίρετο στοιχείο με μηδενική εξάρτηση από την προηγούμενη και την επόμενη. Έτσι αν w_i η λέξη i σε μια πρόταση και w_{i-1}, w_{i+1} η προηγούμενη και η επόμενη αντίστοιχα τότε $P(w_i|w_{i-1}) = P(w_i|w_{i+1}) = 0$, δηλαδή η δεσμευμένη πιθανότητα εμφάνισης της λέξης w_i δεδομένης της w_{i-1} ή της w_{i+1} είναι μηδενική. Οι πιθανότητες αυτές ονομάζονται Uni-gram πιθανότητες και είναι αυτές οι οποίες χρησιμοποιούνται από το BoW. Σαφώς το BoW αντιμετωπίζει ιδιαίτερες δυσκολίες όταν για την ταξινόμηση απαιτείται η κατανόηση του κειμένου [66]. Χαρακτηριστικό παράδειγμα αποτελούν οι προτάσεις:

- This is not true! I do like this.

- This is true! I do not like this.

που έχουν ακριβώς τα αντίθετα νοήματα όμως το ίδιο διάνυσμα χαρακτηριστικών κατά το μοντέλο BoW αφού δεν λαμβάνει υπόψιν του την σειρά με την οποία εμφανίζονται η λέξεις στο κείμενο. Η προεπεξεργασία του κειμένου, με τεχνικές που αναπτύχθηκαν στο 5.2, πριν την χρήση του μοντέλου BoW είναι ιδιαίτερα σημαντική ώστε αυτό να αποδίδει το μέγιστο. Συνήθως πριν την χρήση του BoW αφαιρούνται λέξεις οι οποίες έχουν ιδιαίτερα μεγάλη συχνότητα εμφάνισης αλλά δεν προσδίδουν κάποια πληροφορία στην ταξινόμηση, οι οποίες ονομάζονται Stopwords.

3.5.2 Term Frequency-Inverse Document Frequency (TF-IDF)

Πολλές φορές το μοντέλο TF-IDF θεωρείται κοινό με το BoW αν και στην ουσία αποτελεί επέκταση του, παρόλα αυτά επειδή η χρήση του ενός απαιτεί την χρήση του άλλου συγχέονται ως ένα μοντέλο. Όπως και το μοντέλο BoW έτσι και το TF-IDF βασίζεται στην δημιουργία διανύσματος χαρακτηριστικών σε αραιούς πίνακες που όμως λαμβάνουν μια πραγματική τιμή στις σημαντικές θέσης της πρότασης σε αντίθεση με τα 1' που τοποθετεί ο BoW. Ένα πρόβλημα το οποίο παρατηρήθηκε στο Bow μοντέλο και ώθησε τους ερευνητές στην δημιουργία του TF-IDF είναι πως μερικές φορές λέξεις οι οποίες χρησιμοποιούνται ιδιαίτερα συχνά σε πολλά κείμενα δεν φέρουν την πληροφορία των προτάσεων σε αντίθεση με λέξεις που εμφανίζονται συχνά σε κάποιο μεμονωμένο κείμενο. Το γεγονός αυτό διατύπωσε η Sparks[2] [47] δημιουργώντας την έννοια του Inverse Document Frequency σε μια προσπάθεια να ενισχυθούν οι λέξεις που εμφανίζονται συχνά σε ένα κείμενο και να επιβληθεί μια ποινή στις πιο συχνές λέξεις σε ένα σύνολο δεδομένων κατά την διαδικασία προεπεξεργασίας του. Το IDF υπολογίζει το πόσο σπάνια εμφανίζεται μια λέξη μέσα σε ένα σύνολο κείμενων, οπότε αν συνδυαστεί με την Term Frequency - TF δημιουργείται ένα μοντέλο που μπορεί να ανιχνεύσει, σε ένα βαθμό, λέξεις που περιέχουν την μεγαλύτερη πληροφορία σε ένα κείμενο. Υπάρχουν διάφορες επεκτάσεις του μοντέλου αυτού που χρησιμοποιούν διάφορες μετρικές για τον υπολογισμό της συχνότητας TF όπως η λογαριθμική κλίμακα συχνότητας $tf(word) = log(1 + f_{word})$ ή μια ιδιαίτερα δημοφιλής κανονικοποίηση της για την αποφυγή πολώσεων σε μεγάλα κείμενα $tf(word) = 0.5 + 0.5\frac{f_{word}}{maxf}$. Όμοια μπορεί να υπολογιστεί σε λογαριθμική κλίμακα ο όρος $idf(word) = log\frac{n_{word}}{N}$ όπου N το πλήθος των κειμένων και n_{word} το πλήθος των κειμένων που η λέξη έκανε την εμφάνιση της. Ο τελικός δείκτης του μοντέλου υπολογίζεται ως το γινόμενο των δεικτών Tf , Idf ως εξής:

$$Tf - Idf(word) = tf(word) * idf(word) \tag{3.26}$$

Παρόλα αυτά και αυτό το μοντέλο παρουσιάζει αρκετές αδυναμίες. Όπως και στο BoW το μοντέλο δεν μπορεί να αντλήσει από τα συμφραζόμενα την πληροφορίες που προέρχονται από την σειρά τον λέξεων ενώ επίσης φράσεις που είναι παρόμοιες σημασιολογικά αναγνωρίζονται ως τελείως ξένες, όπως για παράδειγμα οι φράσεις 'Used car , Όλδ μοτορςαρ''. Ταυτόχρονα η πολυπλοκότητα του μοντέλου είναι σαφώς μεγαλύτερη σε σύγκριση με την απλοϊκή προσέγγιση

[2] https://en.wikipedia.org/wiki/Karen_Sp%C3%A4rck_Jones

του BoW και σε πολύ μεγάλα δεδομένα κειμένων η αραιή μορφή πινάκων να καταλαμβάνει μεγάλο χώρο μνήμης και μην είναι λειτουργική.

3.5.3 Μοντέλα n-gram

Σε συνέχεια του μοντέλου BoW το οποίο χρησιμοποιεί αποκλειστικά τις Uni-gram πιθανότητες αναπτύχθηκαν νέα μοντέλα που υπολογίζουν την σχέση της κάθε λέξεις με τις n γειτονικές της. Έτσι οι πιθανότητες που στο uni-gram μοντέλο ήταν μηδενικές πλέον αντικαθίστανται από την πιθανότητα εμφάνισης της κάθε λέξης στα συμφραζόμενα του κειμένου. Συνήθως χρησιμοποιούνται τα $bi-gram$ και $tri-gram$ μοντέλα που δίνουν αρκετά ικανοποιητικά αποτελέσματα και πιο σπάνια υπολογίζονται μοντέλα $n-gram$ τα οποία απαιτούν μεγαλύτερη υπολογιστική ισχύ. Τα μοντέλα αυτά χρησιμοποιούνται στα Κρυφά Μαρκοβιανά Μοντέλα (Hidden Markov Models) [8] αλλά παρουσιάζουν προβλήματα όταν εμφανίζονται λέξεις που δεν έχουν ξαναπαρουσιαστεί στο λεξικό εκπαίδευσης (Out of Vocabulary - OoV).

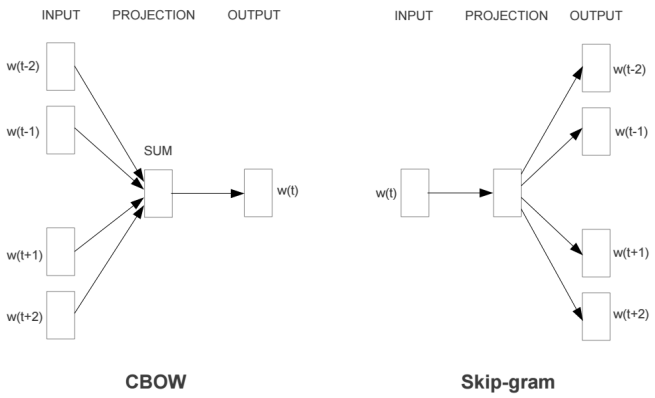
3.5.4 Διανύσματα Λέξεων (Word Embeddings)

Η δημιουργία των (Word Embeddings) ήταν η πρώτη που επήλθε χρονικά στην δημιουργία αναπαραστάσεων για δεδομένα κειμένου από τον Bengio το 2003 [11]. Αποτελούν ANN τα οποία εκπαιδεύονται με αλγορίθμους Stochastic gradient descent-SGD σε ένα μεγάλο σε έκταση κείμενο (corpus), της τάξης των δισεκατομμυρίων λέξεων, χρησιμοποιώντας τεχνικές μη επιβλεπόμενης μάθησης. Το δίκτυό του Bengio αποτελείται από ένα επίπεδό και χρησιμοποιεί την συνάρτηση Softmax στην έξοδο του ώστε να υπολογίζει τις n-gram πιθανότητες. Όμως οι υπολογιστικές δυνατότητες της εποχής δεν έδωσαν την δυνατότητα στον Bengio να εκπαιδεύσει το δίκτυο του με ένα μεγάλο λεξικό. Το 2008 οι Collobert και Weston [23] ανέδειξαν πόσο χρήσιμο θα ήταν για τα προβλήματα επεξεργασίας φυσικής γλώσσας η χρήση προ-εκπαιδευμένων πινάκων (Word Embeddings) εξοικονομώντας μεγάλο υπολογιστικό κόστος.

3.5.4.1 Word2Vec

Το 2013 ο Mikolov[64, 63] εισάγει την έννοια του αλγορίθμου Word2Vec που συμβάλει στην δημιουργία μεγάλης ερευνητικής περιοχής γύρω από την ομοιότητα κειμένων (Text Similarity). Το Word2Vec αποτελεί ANN δύο κρυφών στρωμάτων και επεξεργάζεται ένα μεγάλο κείμενο για να εξάγει διανύσματα χαρακτηριστικών για τις λέξεις του κειμένου. Μια βασική διαφορά του Word2Vec σε σχέση με τους προγενέστερους αλγορίθμους αποτελεί η χρήση γραμμικών συναρτήσεων ενεργοποίησης στους νευρώνες του δικτύου. Λαμβάνοντας υπόψιν τα συμφραζόμενα και με μηδενική ανθρώπινη παρέμβαση ο αλγόριθμος Word2Vec εκπαιδεύεται από τις γειτονικές σχέσεις λέξεων, υπολογίζοντας την συχνότητα συν-εμφάνισης(Co-occurance) τους, και υπολογίζει αριθμητικές αναπαραστάσεις για την δημιουργία ενός συμπαγούς λεξιλογικού διανύσματος. Το διάνυσμα αυτό περιέχει αριθμητικές τιμές με την μορφή πιθανοτήτων, με την χρήση της συνάρτησης $softmax$, για την ομοιότητα και την συσχέτιση μεταξύ των λέξεων. Το βασικό πλεονέκτημα του Word2Vec έναντι των άλλων αλγορίθμων για

την δημιουργία Word Embeddings αποτελεί η ανάλυση σχέσεων μεταξύ των λέξεων (Latent Semantics Analysis) που αποτελεί μια ειδική περιοχή της Επεξεργασίας Φυσική Γλώσσας. Οι μετρικές που κυρίως χρησιμοποιούνται είναι η ομοιότητα συνημίτονού (Cosine Similarity) αλλά και η Ευκλείδεια Απόσταση. Ο αλγόριθμος χρησιμοποιεί δύο βασικές τεχνικές της Επεξεργασίας Φυσικής Γλώσσας για την δημιουργία κατανεμημένων αναπαραστάσεων των λέξεων του κειμένου (Distributed Representation of Words), έναν για την εύρεση της λέξης των συμφραζομένων (Continuous Bag Of Words (CBoW)) και έναν για την εύρεση των συμφραζομένων με βάση την λέξη (Skip-Gram) όπως φαίνεται στο Σχήμα 3.13. Αναλυτικότερα, οι



Σχήμα 3.13: Λειτουργία Continuous Bag Of Words(αριστερά) και Skip-Gram (δεξιά)

δύο τεχνικές που προαναφέρθηκαν μπορούν να μεταφραστούν μέσω της συνάρτησης κόστους στα παρακάτω προβλήματα ελαχιστοποίησης:

Συνάρτηση Κόστους CBoW:

$$J(\theta) = \frac{1}{T}\sum_{t=1}^{T} p(w_t|w_{t-n}, w_{t-n-1}, ..., w_{t-1}, w_{t+1}, ..., w_{t+n-1}, w_{t+n}) \tag{3.27}$$

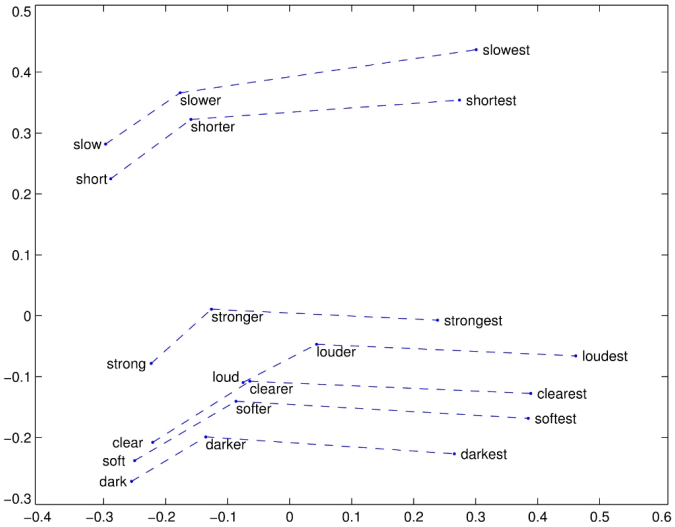
Συνάρτηση Κόστους Skip-gram:

$$J(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=-n}^{n} p(w_{t+j}|w_t) \tag{3.28}$$

Όπως γίνεται αντιληπτό, η συνάρτηση κόστους για το μοντέλο CBoW αποσκοπεί στην πρόβλεψη της λέξης w_t χρησιμοποιώντας την γνώση των n προηγούμενων και n επόμενων λέξεων ενώ αντίθετα το μοντέλο Skip-gram αποσκοπεί στην πρόβλεψη των $2n$ γειτονικών λέξεων που συνοδεύουν την λέξη w_t. Συνηθίζεται να χρησιμοποιείται ένα παράθυρο 10 λέξεων ως συμφραζόμενα(Context) σε Skip-Gram μοντέλο και 5 λέξεις για Continuous Bag Of Words μοντέλο. Ο αλγόριθμος Word2Vec αποτέλεσε πολύ σημαντικό εργαλείο στην κατηγοριοποίηση προτάσεων και στα συστήματα συστάσεων (Recommendation Systems).

3.5.4.2 Global Vectors (GloVe)

Η δημιουργία του GloVe από τον Pennington το 2014 αποτελεί την πλέον χρησιμοποιούμενη μορφή Word Embeddings και αποτελεί τη βασική τεχνική της παρούσας διπλωματικής εργασίας. Συγκεκριμένα, όπως αναφέρει ο Πεννινγτον, η πληροφορία σε ένα κείμενο είναι η πιθανότητα εμφάνισης δυο γειτονικών λέξεων. Ο αλγόριθμος GloVe σε αντίθεση με τον Word2Vec, που είναι βασίζεται σε προβλέψεις, είναι ένας Count-based αλγόριθμος. Έτσι ο αλγόριθμος GloVe κωδικοποιεί τις λέξεις χρησιμοποιώντας διανυσματικές διαφορές. Όπως φαίνεται στο Σχήμα 3.14 οι ευθείες που ενώνουν λέξεις όπως το short με το shorter και το slow με το slower είναι παράλληλες αφού τα αντίστοιχα Word Embeddings έχουν δημιουργηθεί με γνώμονα τις λεξιλογικές διαφορές.



Σχήμα 3.14: Λεξιλογικές Αποστάσεις με την χρήση GloVe

3.5.5 Μετρικές Αξιολόγησης

Μετά την εκπαίδευση του αλγορίθμου Μηχανικής Μάθησης είναι απαραίτητο να αξιολογήσουμε την απόδοση του στην ικανότητα να ταξινομεί τα δείγματα των δεδομένων αξιολόγησης και για το σκοπό αυτό χρησιμοποιούνται διάφορες μετρικές.

Μετρική Accuracy: αποτελεί την πιο διαδεδομένη μετρική αξιολόγησης της απόδοσης ταξινομητών. Είναι ιδιαίτερα απλοϊκός και υπολογίζει το ποσοστό των στοιχείων του συνόλου δοκιμής που ταξινομήθηκαν στην σωστή κατηγορία. Ορίζεται ως:

$$Accuracy = \frac{\text{Πλήθος σωστά ταξινομημένων δειγμάτων του Test Set}}{\text{Σύνολο δειγμάτων Test Set}} \tag{3.29}$$

Παρόλα αυτά η μετρική accuracy παρότι μας δείχνει το ποσοστό του συνόλου που ταξινομείται σωστά, δεν μας προσδίδει πληροφορίες όσο αφορά την ταξινόμηση στοιχείων σε κάθε κατηγορία. Για να αποκτήσουμε αυτές τις πληροφορίες μπορεί να χρησιμοποιηθεί ένα σύνολο πρόσθετων μετρικών που παρουσιάζονται παρακάτω. Αν θεωρήσουμε πως έχουμε ένα δυαδικό

πρόβλημα ταξινόμησης και ονομάσουμε τις δύο κατηγορίες αυτές ως Positive και Negative μπορούμε να ορίσουμε τις παρακάτω έννοιες:

- True Positive-TP: δείγματα που ταξινομούνται σωστά στην κατηγορία Positive.
- True Negative-TN: δείγματα που ταξινομούνται σωστά στην κατηγορία Negative.
- False Positive-FP: δείγματα που ταξινομούνται λανθασμένα στην κατηγορία Positive.
- False Negative-FN:δείγματα που ταξινομούνται λανθασμένα στην κατηγορία Negative.

Με βάση τα μεγέθη που ορίσθηκαν παραπάνω μπορούμε να ορίσουμε τις μετρικές Ακρίβειας (Precission) και Ανάκλησης (Recall).

- Μετρική Ανάκλησης (Recall): Η μετρική αυτή εκφράζει το ποσοστό των δειγμάτων που ταξινομούνται σωστά μεταξύ όλων των δειγμάτων που ανήκουν στην κατηγορία Positive και ορίζεται ως:
$$Recall = \frac{TP}{TP + FN} \tag{3.30}$$
- Μετρική Ακρίβειας (Precission): Η μετρική αυτή, σε αντίθεση με την προηγούμενη, εκφράζει το ποσοστό των δειγμάτων που ταξινομήθηκαν σωστά στην κατήγορία Positive και ορίζεται ως:
$$Precission = \frac{TP}{TP + FP} \tag{3.31}$$

Όπως μπορεί εύκολα να γίνει αντιληπτό οι δύο αυτές μετρικές έρχονται σε αντίθεση μεταξύ τους και η αύξηση της μιας οδηγεί στην μείωση της άλλης. Για τον λόγο αυτό είναι χρήσιμο να δημιουργηθεί μια μετρική που να συνδυάζει της παραπάνω. Η μετρική αυτή ονομάζεται $\mathrm{F}_1 Score$ και αποτελεί τον αρμονικό μέσο όρο των παραπάνω μετρικών:

$$F_1 = \frac{2}{\frac{1}{Precission} + \frac{1}{Recall}} = 2\frac{Precission \times Recall}{Precission + Recall} \tag{3.32}$$

- Receiver Operating Characteristic-ROC: αποτελεί την καμπύλη που συνδέει το ποσοστό TP, που χαρακτηρίζεται αλλιώς και ως ευαισθησία (sensitivity), με το FP, που χαρακτηρίζεται αλλιώς ως 1-ειδικότητα (specificity). Η καμπύλη ROC συμβάλει στην εύρεση του βέλτιστου ορίου απόφασης (decision threshold) που ελαχιστοποιεί το ποσοστό σφάλματος, σχεδιάζοντας τα ποσοστά ευαισθησίας-ειδικότητας για διαφορετικά όρια ταξινόμησης. Ταυτόχρονα με την καμπύλη ROC μπορούμε να αποφανθούμε για τις περιοχές που κάποιος ταξινομητής υπερτερεί έναντι άλλου. Το πιο διαδεδομένο στατιστικό στοιχείο που συνδέεται με την καμπύλη ROC είναι το εμβαδόν κάτω από την καμπύλη (Area Under Curve-AUC) και παρέχει ένα συνολικό μέτρο αξιολόγησης για όλα τα πιθανά όρια ταξινόμησης. Η μετρική AUC εκφράζει πιθανότητα η εμπιστοσύνη του ταξινομητή πως ένα δείγμα της κατηγορίας Positive ανήκει πράγματι στην κατηγορία αυτή να είναι μεγαλύτερη από την εμπιστοσύνη του ταξινομητή πως ένα δείγμα της κατηγορίας Negative ανήκει πράγματι στην κατηγορία Negative.

- Μετρική Ομοιότητας Συνημίτονου (Cosine Similarity): αποτελεί κανονικοποιημένη μορφή του Ευκλείδειου εσωτερικού γινομένου μεταξύ δύο διανυσμάτων. Γεωμετρικά η ομοιότητα συνημιτόνου αποτελεί το συνημίτονο της γωνία που σχηματίζουν τα δύο διανύσματα μεταξύ τους. Έτσι όταν τα δύο διανύσματα είναι παράλληλα η μετρική έχει αποτέλεσμα 1 ενώ όταν τα δύο διανύσματα είναι κάθετα έχει αποτέλεσμα 0. Η μετρική αυτή χρησιμοποιείται σε προβλήματα παλινδρόμησης όπου δεν υπάρχουν διακριτές τιμές ταξινόμησης και πρέπει να ελεγχθεί η αξιοπιστία του μοντέλου. Έτσι, συγκρίνονται τα αποτελέσματα της ταξινόμησης με τις ορθές τιμές των δεδομένων δοκιμής μέσω της ομοιότητας συνημιτόνου τους. Η μετρική ομοιότητας αυτή δίνεται από την παρακάτω εξίσωση:

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{3.33}$$

με A, B τα δύο διανύσματα και θ την γωνία μεταξύ τους. Έτσι, αν το μοντέλο, έστω προβλέψει μια αλληλουχία τιμών A=[1,1,3] ενώ οι πραγματικές τιμές της αλληλουχίας είναι οι B=[1,2,2] η μετρική ομοιότητας συνημιτόνου θα έχει ως αποτέλεσμα 0.904. Αυτό αποτελεί και το βασικό της πλεονέκτημα σε προβλήματα τα οποία αποτελούνται από πολλές κατηγορίες ή που η ταξινόμηση πρέπει να γίνει σε συνεχείς τιμές που δεν μπορούν να διακρητοποιηθούν. Στο αντίστοιχο παράδειγμα η μετρική ακρίβειας θα είχε σαν αποτέλεσμα 0.333, το οποίο απέχει πολύ από το παραπάνω, χωρίς όμως να λαμβάνει υπόψιν τις σχετικά 'κοντινές' προβλέψεις του μοντέλου.

Κεφάλαιο 4

Αρχιτεκτονικές Μοντέλων Ανίχνευσης ΜΓΦ

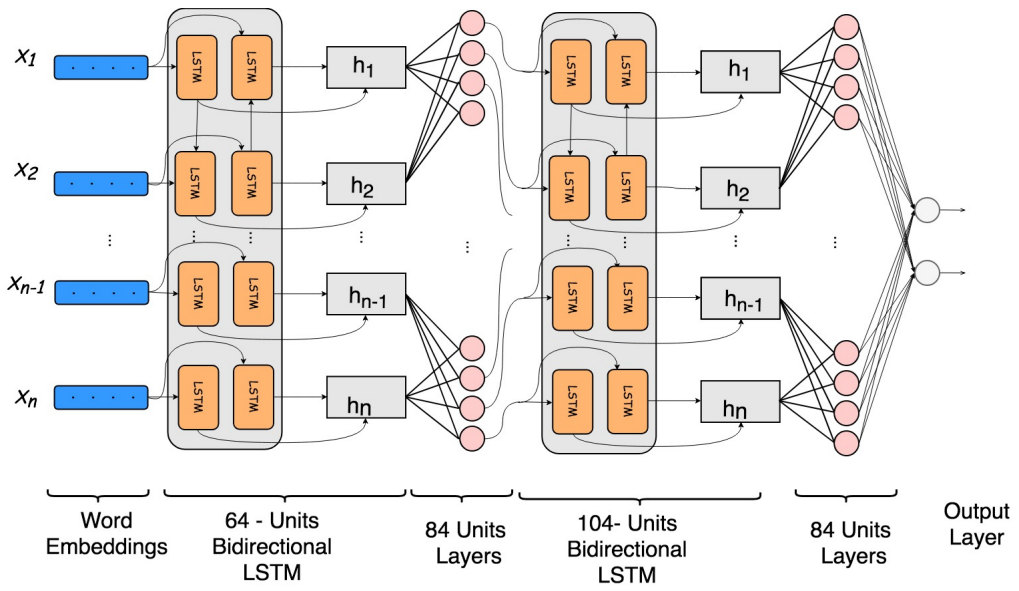
Στο παρόν κεφάλαιο γίνεται μια ανάλυση ορισμένων από τα μοντέλα που δημιουργήθηκαν για την αντιμετώπιση του προβλήματος της χρήσης μεταφορικής γλώσσας και την αντιμετώπιση προβλημάτων ΜΦΓ. Τα μοντέλα που παρουσιάζονται αποτελούν Βαθιά Νευρωνικά Δίκτυα με εισόδους που παρουσιάζονται και αναλύονται στο επόμενο κεφάλαιο.

4.1 Αμφίδρομο LSTM

Όπως αναπτύχθηκε στην ενότητα 3.3.4.1 τα αμφίδρομα LSTM αποτελούν ένα σημαντικό εργαλείο σε ακολουθιακά δεδομένα που πιθανόν η δομή τους να απαιτεί την μελέτη τόσο προγενέστερων όσο και μελλοντικών εισόδων. Στα μεταφορικά φαινόμενα είναι πιθανό μια λέξη σε μελλοντική είσοδο να ανατρέπει το νοηματικό περιεχόμενο του κειμένου, επομένως κρίνεται εύλογη η διαχείριση του κειμένου και από τις δύο πλευρές ανάγνωσης του με χρήση αμφίδρομων δικτύων. Για το μοντέλο αυτό χρησιμοποιούμε στην είσοδο προεκπαιδευμένα Word Embeddings τα οποία τροφοδοτούν δύο αμφίδρομα στρώματα LSTM. Αναλυτικότερα, στο στάδιο εισόδου τοποθετείται ένα αμφίδρομο LSTM 64-μονάδων η έξοδος του οποίο συνδέεται με ένα στρώμα πλήρως συνδεδεμένου ANN 84-νευρώνων πριν συνδεθεί με το επόμενο στρώμα αμφίδρομου LSTM 104-μονάδων. Η έξοδος του δικτύου αποτελείται από ένα νευρωνικό δίκτυο δύο στρωμάτων με 84 και 2-νευρώνες αντίστοιχα. Το μοντέλο εκπαιδεύεται χρησιμοποιώντας αλγόριθμο βελτιστοποίησης Adam και Cross-Entropy σαν συνάρτηση κόστους. Χάριν συντομίας το μοντέλο αυτό θα αναφέρεται ως **BiLSTM**.

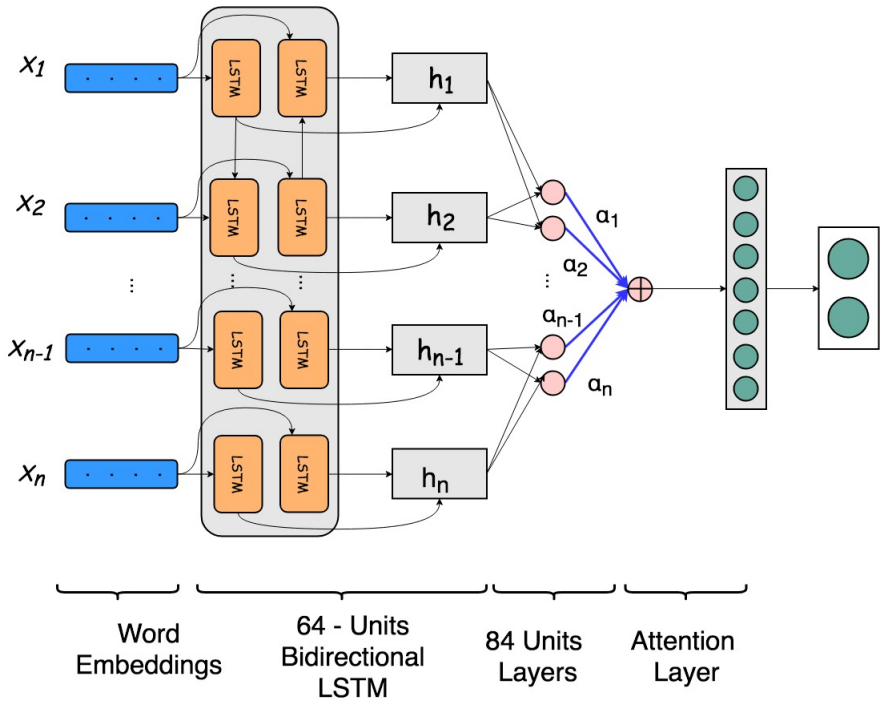
4.2 Αμφίδρομο LSTM με Μηχανισμό Προσοχής

Στην ενότητα 3.3.4.2 αναλύθηκε λεπτομερώς η λειτουργία ενός επιπέδου που προστίθεται στην κορυφή ενός νευρωνικού δικτύου και βοηθά το δίκτυο να εστιάσει την 'προσοχή' του σε κάποια συγκεκριμένη είσοδο της ακολουθίας. Σε ένα πρόβλημα όπως η ανίχνευση μεταφορικών γλωσσικών φαινομένων η δυνατότητα αυτή μπορεί να αποτελέσει κλειδί αφού τις περισσότερες



Σχήμα 4.1: Αμφίδρομο LSTM

φορές η αντίθεση μερικών λέξεων στο κείμενο είναι αυτές που το κρίνουν σαν μεταφορικό.



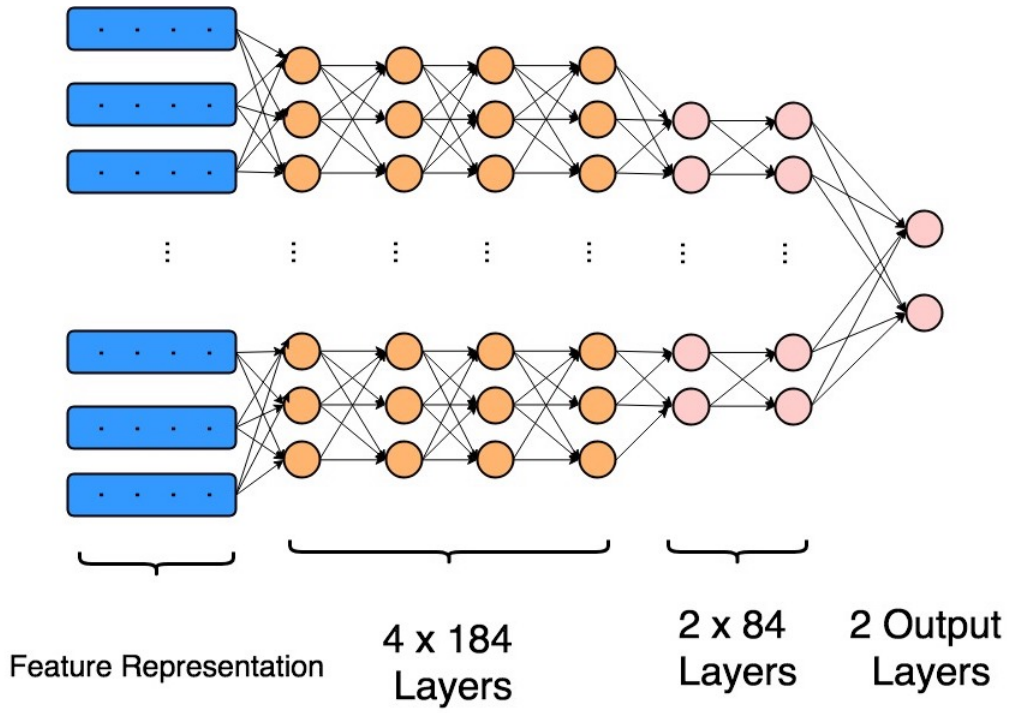
Σχήμα 4.2: Αμφίδρομο LSTM με μηχανισμό προσοχής

Επιπλέον, χρησιμοποιούμε την αμφίδρομη μορφή της μονάδας LSTM ώστε να ενισχύσουμε το μηχανισμό προσοχής στην ανεύρεση αντιθέσεων στα δεδομένα. Σύμφωνα με τα παραπάνω λοιπόν δημιουργούμε ένα μοντέλο που χρησιμοποιεί Word Embeddings για χαρακτηριστικά στην κορυφή του οποίου τοποθετείται ένα αμφίδρομο 64-μονάδων LSTM με μηχανισμό προ-

σοχής. Στο στάδιο εξόδου τοποθετείται ένα στρώμα πλήρως συνδεδεμένου ANN. Το μοντέλο χρησιμοποιεί βελτιστοποίηση adadelta και Cross-Entropy σαν συνάρτηση κόστους ενώ εκπαιδεύεται ανάλογα το μέγεθος του συνόλου δεδομένων ώστε να αποφευχθούν φαινόμενα overfiting. Στο επόμενο κεφάλαιο θα αναφερόμαστε σε αυτό το μοντέλο με την σύντομη ονομασία **AttentionLSTM**.

4.3 Βαθύ Νευρωνικό Δίκτυο - DNN

Μια διαφορετική προσέγγιση είναι η χρήση βαθιών νευρωνικών δικτύων ώστε να χρησιμοποιήσουμε ένα σύνολο χαρακτηριστικών που εξάγουμε από κάθε tweet σε αντίθεση με τα Word Vectors που χρησιμοποιήθηκαν για τα μοντέλα LSTM. Σαν χαρακτηριστικά χρησιμοποιήθηκαν στοιχεία του κειμένου αλλά και n-gram μοντέλα με αποτέλεσμα τα διανύσματα χαρακτηριστικών να έχουν αρκετά μεγάλες διαστάσεις. Έτσι δημιουργήθηκε ένα βαθύ νευρωνικό δίκτυο με μεγάλο πλήθος νευρώνων σε κάθε επίπεδο του ώστε να μπορεί να μοντελοποιεί επακριβώς τα δεδομένα αυτά.



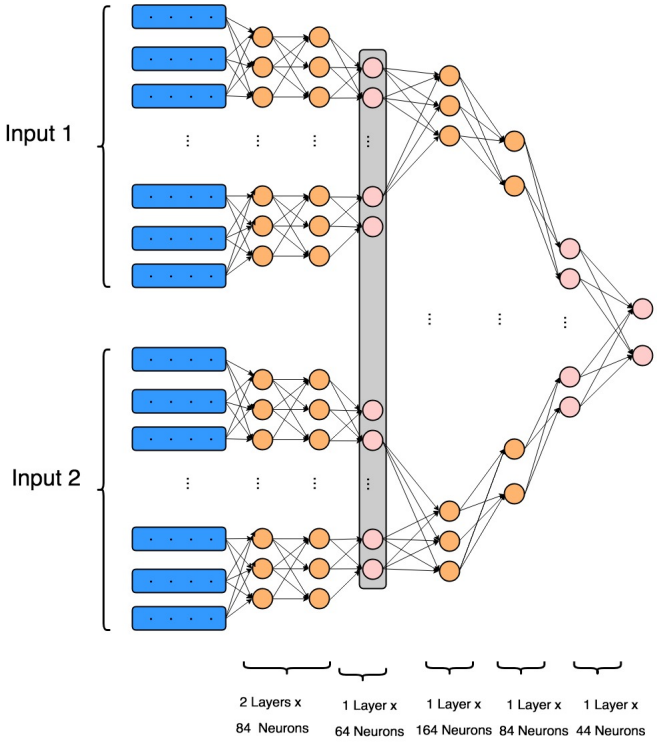
Σχήμα 4.3: Βαθύ Νευρωνικό Δίκτυο

Το μοντέλο αυτό αποτελείται από έξι επίπεδα κρυφών, πλήρως συνδεδεμένων, νευρώνων με τα τέσσερα πρώτα να απαρτίζονται από 184 νευρώνες και τα επόμενα δύο από 84. Στο στάδιο εξόδου τοποθετείται ένα πυκνό στρώμα δύο νευρώνων με softmax συνάρτηση ενεργοποίησης. Στα κρυφά στάδια η συνάρτηση ενεργοποίησης είναι η ReLu ενώ χρησιμοποιείται και στρώμα Dropout για την αποφυγή overfitting. Όπως και προηγουμένως χρησιμοποιείται ο αλγόριθμος

βελτιστοποίησης Adam και η επιλογή εποχών εκπαίδευσης καθορίζεται από το μέγεθος και την πολυπλοκότητα των δεδομένων, παρόλα αυτά η εκπαίδευση του δικτύου σε μεγάλες διαστάσεις διανύσματος χαρακτηριστικών απαιτούν περισσότερες εποχές εκπαίδευσης σε σχέση με τα δύο προηγούμενα μοντέλα. Στο μοντέλο αυτό δοκιμάσαμε να εισάγουμε διαφορετικά είδη χαρακτηριστικών όπως uni-gram, συχνότητες Tf-idf των uni-gram και bi-gram μοντέλων, χαρακτηριστικά που εξάγονται από το κείμενο καθώς και συνδυασμούς αυτών. Στο επόμενο κεφάλαιο θα αναφερόμαστε σε αυτά τα μοντέλα με τις αντίστοιχες ονομασίες **DNN-uni, DNN-Tfidf και DNN-all**, ανάλογα με τα χαρακτηριστικά που το τροφοδοτούμε.

4.4 Βαθύ Νευρωνικό Δίκτυο Δύο Εισόδων

Μια επέκταση του προηγούμενου μοντέλου θα μπορούσε να αποτελέσει μια συνένωση δύο διαφορετικών νευρωνικών δικτύων με ανεξάρτητες εισόδους. Αναλυτικότερα, δύο διαφορετικών διαστάσεων χαρακτηριστικά για τα δεδομένα τροφοδοτούνται σε δύο συμμετρικά βαθιά νευρωνικά δίκτυα δύο κρυφών καταστάσεων με 84-νευρώνες.



Σχήμα 4.4: Βαθύ Νευρωνικό Δίκτυο Δύο Εισόδων

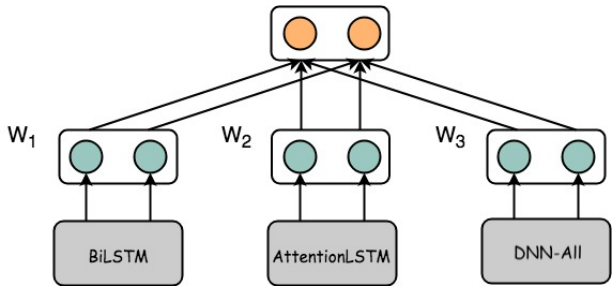
Οι έξοδοι των συμμετρικών δικτύων, αποτελούνται από 64-νευρώνες και συνενώνονται ώστε να τροφοδοτήσουν ένα νέο βαθύ νευρωνικό δίκτυο. Το δίκτυο αυτό αποτελείται από τρία κρυφά επίπεδα με κλιμακούμενο πλήθος νευρώνων, 164-84-44 αντίστοιχα και ReLu συναρτήσεις ενεργοποίησης. Στο στάδιο εξόδου χρησιμοποιούνται και πάλι δύο νευρώνες με

Softmax ενεργοποιήσεις.

Στα πειράματα και τα αποτελέσματα που παρατίθενται στο Κεφάλαιο 5, χρησιμοποιείται σαν πρώτη είσοδος του μοντέλου ένα διάνυσμα με συχνότητες Tf-idf για uni-gram και bi-gram μοντέλα, ενώ για δεύτερη είσοδος επιλέγονται χαρακτηριστικά που εξάγονται από το κείμενο όπως περιγράφονται στην ενότητα 5.3. Το μοντέλο αυτό θα αναφέρεται σαν **2inpDNN** στους πίνακες της αξιολόγησης/σύγκρισης αποτελεσμάτων στο Κεφάλαιο 5.

4.5 Συνδυασμός Μοντέλων-Ensemble Model

Στην ενότητα 3.3.5 παρουσιάστηκε η δυνατότητα συνδυασμού μοντέλων και αλγοριθμικών τεχνικών για την επίτευξη καλύτερων αποτελεσμάτων. Είναι ιδιαίτερα σύνηθες τα διαφορετικά μοντέλα να κατανοούν διαφορετικά χαρακτηριστικά καλύτερα από άλλα με αποτέλεσμα ο συνδυασμός τους να επιφέρει καλύτερη κατανόηση των δεδομένων. Έτσι και στο πρόβλημα αναγνώρισης μεταφορικών γλωσσικών φαινομένων χρησιμοποιούμε ένα συνδυασμό τριών ταξινομητών με την μέθοδο Ανεκτικής Ταξινόμησης. Συγκεκριμένα, επιλέγουμε τα αποτελέσματα από το Αμφίδρομο LSTM, το Αμφίδρομο LSTM με Μηχανισμό και το Βαθύ Νευρωνικό Δίκτυο και κλιμακώνουμε τα αποτελέσματα τους με κατάλληλους συντελεστές W_1, W_2, W_3 όπως φαίνεται στο Σχήμα 4.5. Οι συντελεστές αυτοί αποτελούν τις υπερπαραμέτρους (hyper-parametres) του συστήματος και προσδιορίζονται δυναμικά κατά την διάρκεια της επικύρωσης (validation) των δεδομένων. Στο συνδυασμό χρησιμοποιούμε ένα πλήρες σύνολο χαρακτηριστικών που εξάγουμε από το κείμενο, όπως περιγράφεται στο επόμενο κεφάλαιο, ως είσοδο στο Βαθύ Νευρωνικό Δίκτυο. Ο συνδυασμός αυτός στην πορεία της εργασίας θα καλείτε ως **Deep Ensemble Soft Classifier - DESC** λόγω των αλγορίθμων βαθιών νευρωνικών δικτύων που συνδυάζονται και των τεχνικών που χρησιμοποιούνται.



Σχήμα 4.5: Deep Ensemble Soft Classifier - DESC Μοντέλο

4.6 Παρατηρήσεις

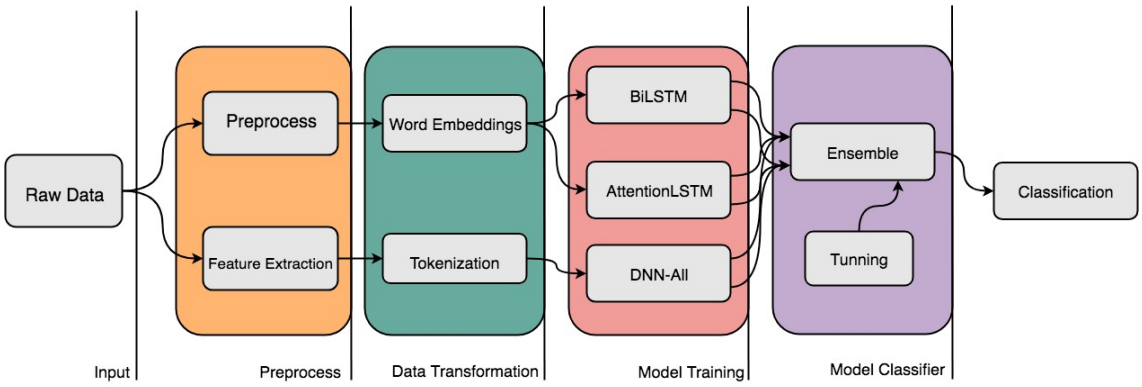
Τα μοντέλα που περιγράφηκαν στις παραπάνω ενότητες θα αποτελέσουν τους βασικούς αλγορίθμους για την αντιμετώπιση του προβλήματος αναγνώρισης ΜΓΦ. Η απλοποιημένη σχηματική αναπαράσταση τους καθώς και η σύντομη περιγραφή της αρχιτεκτονικής τους δεν διατυπώνει πλήρως κάποια χαρακτηριστικά τους που δεν αναπτύχθηκαν εκτενώς. Παρακάτω γίνεται μια προσπάθεια αποσαφήνισης ορισμένων από αυτά τα στοιχεία:

- Σε όλες τις μονάδες LSTM χρησιμοποιείται η συνάρτηση ReLu.
- Σε όλα τα μοντέλα στο στάδιο εξόδου χρησιμοποιείται η συνάρτηση ενεργοποίησης Softmax.
- Τα παραπάνω μοντέλα αφορούν δεδομένα δυαδικής ταξινόμησης. Σε δεδομένα με παραπάνω κατηγορίες το στάδιο εξόδου τροποποιείται ώστε να διαθέτει τόσους νευρώνες όσες και οι πιθανές κλάσεις.

Κεφάλαιο 5

Ανίχνευση Μεταφορικών Γλωσσικών Φαινομένων στο Twitter

Στο παρόν κεφάλαιο παρουσιάζονται αναλυτικά τα πειράματα και τα αποτελέσματα ανίχνευσης ΜΦΓ, καθώς και τα δεδομένα τα οποία χρησιμοποιήθηκαν. Επιπρόσθετα, παρουσιάζονται και συγκρίνονται τα αποτελέσματα διάφορων αλγορίθμων, με διαφορετικές μετρικές απόδοσης όπως αναπτύχθηκαν στην παράγραφο 3.5.5. Τέλος, συγκρίνονται τα μοντέλα που παρουσιάστηκαν στο κεφάλαιο 4 με αποτελέσματα από συγγενικές και ενδεικτικές εργασίες αναφοράς. Η διαδικασία ταξινόμησης ακολουθεί μια ροή εργασιών (pipeline) τεσσάρων σταδίων. Η διαδικασία αυτή περιγράφεται εκτενώς στο κεφάλαιο αυτό και αναπαρίσταται σχηματικά στο παρακάτω Σχήμα 5.1



Σχήμα 5.1: Ροή εργασιών για ανίχνευση ΜΓΦ

5.1 Δεδομένα

Το πρόβλημα της ανίχνευσης ΜΓΦ, πέραν της ιδιομορφίας του λόγου, αποτελεί ένα ιδιαίτερα δύσκολο πρόβλημα αφού δεν υπάρχουν πολλά διαθέσιμα δεδομένα για επεξεργασία. Για

την αξιολόγηση του πειράματος χρησιμοποιήθηκαν τρία διαφορετικά σύνολα δεδομένων που επικεντρώνονται σε διαφορετικά μεταφορικά γλωσσικά φαινόμενα.

5.1.1 Ανίχνευση Ειρωνείας στο Βρετανικό Twitter

Στο διεθνές Workshop Semantic Evaluation-SemEval το 2018 προτάθηκε το πρόβλημα εντοπισμού ειρωνικών και σαρκαστικών tweet στα πλαίσια της δημιουργικής γλώσσας που μπορούν να προσφέρουν τα Social Media[42]. Η συλλογή των δεδομένων του προβλήματος έγινε με την χρήση hashtag που παραπέμπουν σε ειρωνικά tweet, όπως τα #sarcasm, #irony, #not και στην συνέχεια ελέγχθηκαν χειροκίνητα. Μετέπειτα τα hastag αφαιρέθηκαν από τα tweet για να μην προδίδουν την ειρωνεία τους. Τα δεδομένα αποτελούνται από 3,834 tweets για εκπαίδευση και 784 tweets για έλεγχο και χαρακτηρίζονται από την κατηγορία στην οποία ανήκουν, ειρωνικό ή μη ειρωνικό. Τα tweet συλλέχθηκαν την περίοδο μεταξύ 01/12/2014 και 04/01/2015 από 2,676 διαφορετικούς χρήστες. Το σύνολο των δεδομένων είναι πλήρως ισορροπημένο με 2,396 tweet σε κάθε κατηγορία. Τα δεδομένα αξιολόγησης περιέχουν μια αναλογία 40% ειρωνικά και 60% μη ειρωνικά tweet ενώ τα δεδομένα εκπαίδευσης είναι μοιρασμένα με 1911 ειρωνικά και 1923 μη ειρωνικά tweet. Παρότι τα δεδομένα χαρακτηρίζουν ένα δυαδικό πρόβλημα ταξινόμησης, παρουσιάζουν πολλά διαφορετικά είδη ειρωνείας.

Πλήθος Tweet	Σύνολο Εκπαίδευσης	Σύνολο αξιολόγησης
Ειρωνικά	1911	311
Μη ειρωνικά	1923	473
Σύνολο	3834	784

Πίνακας 5.1: Κατανομή ειρωνικών δεδομένων

Για παράδειγμα περιέχονται ειρωνικά σχόλια που παρουσιάζουν αντίθεση συναισθημάτων '*I feel so blessed to get ocular migraines.*' με την λέξη blessed να έρχεται σε αντίθεση με την λέξη migraines, ειρωνικά σχόλια που δεν παρουσιάζουν καμία συναισθηματική αντίθεση '*Human brains disappear every day. Some of them have never even appeared.*' αλλά και ειρωνικά σχόλια που οφείλονται σε περιστασιακά γεγονότα '*Event technology session is having Internet problems. #irony #HSC2024*'.

Μερικά ακόμα παραδείγματα των δεδομένων:

- I really love this year's summer; weeks and weeks of awful weather *(Ειρωνικό)*
- Go ahead drop me hate, I'm looking forward to it. *(Ειρωνικό)*
- I just love when you test my patience!! #not. *(Ειρωνικό)*
- Had no sleep and have got school now #not happy *(Μη ειρωνικό)*

Τύπος ΜΓΦ	# Tweet Εκπαίδευσης	Μέσο Συναίσθημα Εκπαίδευσης	# Tweet Αξιολόγησης	Μέσο Συναίσθημα Αξιολόγησης
Σαρκασμός	5000	-2.25	746	-1.94
Ειρωνεία	1000	-1.70	81	-1.35
Μεταφορά	2000	-0.54	198	-0.34
Σύνολο	8000	-1.99	1025	-1.78

Πίνακας 5.2: Κατανομή δεδομένων Μεταφορικής Γλώσσας

5.1.2 Ανίχνευση Μεταφορικών Γλωσσικών Φαινομένων στο Twitter

Ένα δεύτερο σύνολο δεδομένων που χρησιμοποιήθηκε προέρχεται επίσης από το Workshop SemEval 2015. Σκοπός του προβλήματος ήταν η καταμέτρηση του συναισθήματος που περιέχουν τα μεταφορικά δεδομένα. Τα δεδομένα προέρχονται επίσης από το Twitter και περιέχουν τρία είδη μεταφορικής γλώσσας: ειρωνεία, σαρκασμό και μεταφορά. Οι τιμές συναισθήματος λαμβάνουν τιμές σε μια κλίμακα 11 διακριτών τιμών, από το -5 έως το 5, που έχουν επισημειωθεί από ανεξάρτητους αναγνώστες χειροκίνητα μέσω της πλατφόρμας CrowdFlower[1]. Τιμές κοντά στο -5 υποδηλώνουν αρνητικότητα και κριτική, αντίθετα τιμές κοντά στο +5 υποδηλώνουν θετικά μηνύματα ενώ με 0 υποσημειώνονται συναισθηματικά ουδέτερα tweet. Λόγω της μεταφορική χρήσης της γλώσσα το μεγαλύτερο πλήθος των tweet περιέχουν αρνητικές τιμές συναισθήματος. Έτσι, σκοπός του προβλήματος είναι η αναγνώριση του συναισθήματος που εκφράζει ένα κείμενο μικρού μήκους όταν σε αυτό υπάρχουν μεταφορικά γλωσσικά φαινόμενα. Συνολικά συλλέχθηκαν 9000 tweet με σαρκαστικό, ειρωνικό και μεταφορικό περιεχόμενο που μερικές φορές δεν μπορεί να γίνει διακριτό. Σαν δεδομένα εκπαίδευσης χρησιμοποιούνται 8000 tweet ενώ για δοκιμή τα υπόλοιπα 1000.

Μερικά παραδείγματα από tweet των δεδομένων μαζί με τα αντίστοιχα συναισθηματικά σκορ που έχουν επισημειωθεί:

- A paperless office has about as much chance as a paperless bathroom (-3)
- Today will be about as close as you'll ever get to a "PERFECT 10" in the weather world! Happy Mother's Day! Sunny and pleasant! High 80. (3)
- I love when I'm ready to go to sleep but can't because I can clearly hear my neighbors music from inside my house! NOT (-3)
- This makes me feel soooo good about myself not (-2)

5.1.3 Αναγνώριση Σαρκασμού στο Twitter

Για το γλωσσικό φαινόμενο του σαρκασμού χρησιμοποιήθηκαν δύο διαφορετικά σύνολα από δεδομένα, και τα δύο έχουν προέλθει από το μέσο κοινωνικής δικτύωσης Twitter. Το πρώτο αποτελεί ένα σύνολο 39.780 tweet εκπαίδευσης και 1.975 tweet αξιολόγησης που δημιουργήθηκε από τον Ghosh[34] με βάση τα hashtag #sarcasm , #sarcastic, #not. Τα

[1] https://www.figure-eight.com/

δεδομένα, τόσο εκπαίδευσης όσο και αξιολόγησης είναι σχεδόν ισορροπημένα με 18.488 σαρκαστικά δεδομένα, ποσοστό δηλαδή 46.5% των δεδομένων.

Μερικά παραδείγματα από tweet των δεδομένων μαζί με τις αντίστοιχες κλάσης που έχουν επισημειωθεί:

- I love it when people start rumours about me not (Σαρκαστικό)
- Thank you alarm clock for never going off . hatebeinglate (Σαρκαστικό)
- Don't you just love Mondays ? ! ? ! ? (Σαρκαστικό)
- I get so nervous to speak to people now , I can't even get my words out (Μη Σαρκαστικό)

Το δεύτερο σετ δεδομένων προέρχεται επίσης από το witter, και δημιουργήθηκε από τον Riloff[82] για την διερεύνηση του σαρκασμού ως μια αντίθεση θετικών και αρνητικών συναισθημάτων. Τα δεδομένα αποτελούνται από 2278 tweet εκ των οποίων τα 506 είναι σαρκαστικά. Το συγκεκριμένο σετ δεδομένων είναι αρκετά μη ισορροπημένο αλλά έχει υψηλή ακρίβεια αφού έχει υποσημειωθεί χειροκίνητα από ερευνητές με το μεγαλύτερο Kappa-σκορ[2]. Όπως και τα παραπάνω δεδομένα, έτσι και αυτό έχουν επιλεχθεί με βάσει τα hashtag που δηλώνουν σαρκαστικό περιεχόμενο. Μερικά παραδείγματα από tweet των δεδομένων μαζί με τις αντίστοιχες κλάσης που έχουν επισημειωθεί:

- I love it whenever my brother gets up and turns on all the lights... (Σαρκαστικό)
- Love working on my last day of summer... (Σαρκαστικό)
- Thank god I invested in an umbrella this year. (Μη Σαρκαστικό)

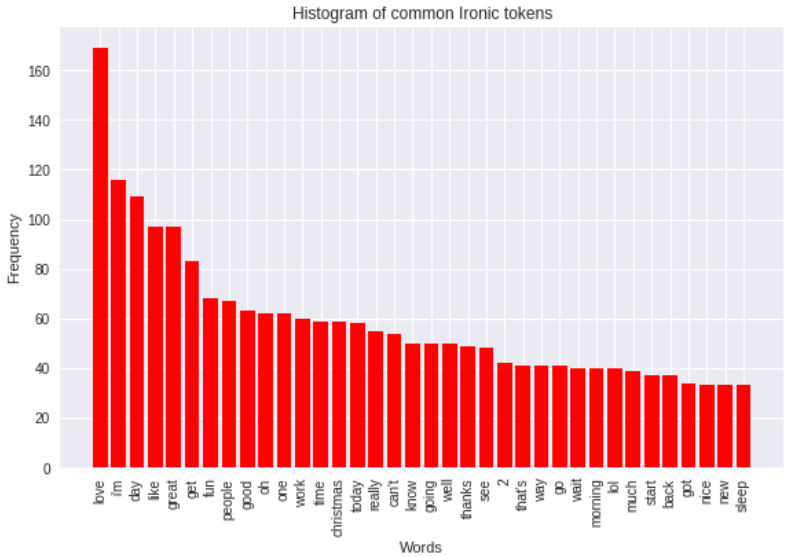
5.1.4 Χαρακτηριστικά Μεταφορικής Γλώσσας

Με βάση την συλλογή δεδομένων που χρησιμοποιούμε είναι χρήσιμο να διερευνήσουμε μερικές από τις πιο συχνές λέξεις που χρησιμοποιούνται στα ειρωνικά και τα σαρκαστικά δεδομένα ώστε να κατανοήσουμε τον τρόπο με τον οποίο διατυπώνεται στα Social Media.

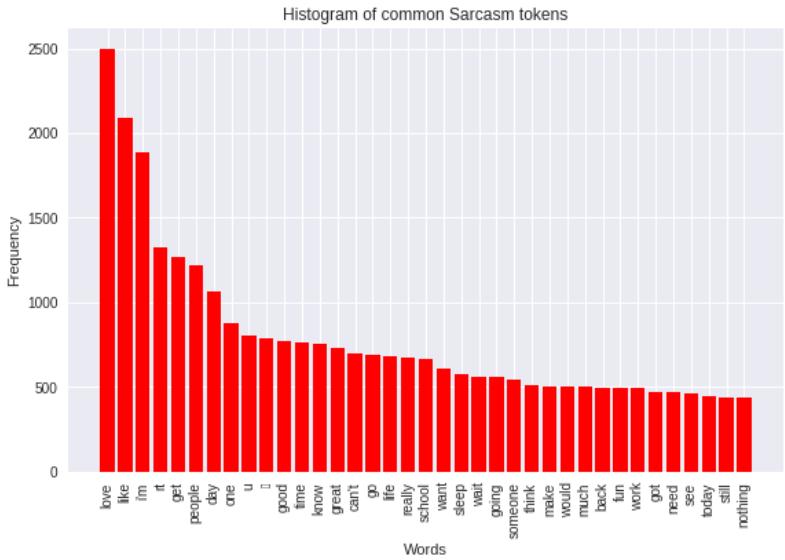
Στα Σχήματα 5.2-5.4 μπορούμε να παρατηρήσουμε την αντίθεση συναισθημάτων που κυριαρχούν στα ειρωνικά και στα σαρκαστικά σχόλια. Λέξεις όπως το love, like, good, fine, great κλπ που αποτυπώνουν θετικά συναισθήματα χρησιμοποιούνται σε ΜΓΦ, δηλώνοντας όμως το ακριβώς αντίθετο συναίσθημα. Μέσω των σχετικών ιστογραμμάτων μπορεί να γίνει αντιληπτή και η δυσκολία του διαχωρισμού των ΜΓΦ από υψηλά συναισθηματικά φορτισμένες εκφράσεις που περιέχουν αντίστοιχο λεξιλόγιο.

Ταυτόχρονα, όπως μπορεί να παρατηρηθεί στο Σχήμα 5.4 το λεξιλόγιο είναι παρόμοιο και στα τρία είδη της μεταφορικής γλώσσας με τις υψηλά συναισθηματικά λέξεις να είναι αυτές που κυριαρχούν. Ένα ακόμα χαρακτηριστικό των μεταφορικών δεδομένων είναι η πιο σύντομη έκφραση που χρησιμοποιούν. Τα περισσότερα από τα tweet χρησιμοποιούν 50 με 100

[2] https://en.wikipedia.org/wiki/Cohen%27s_kappa



Σχήμα 5.2: Συχνότητα λέξεων στα ειρωνικά δεδομένα



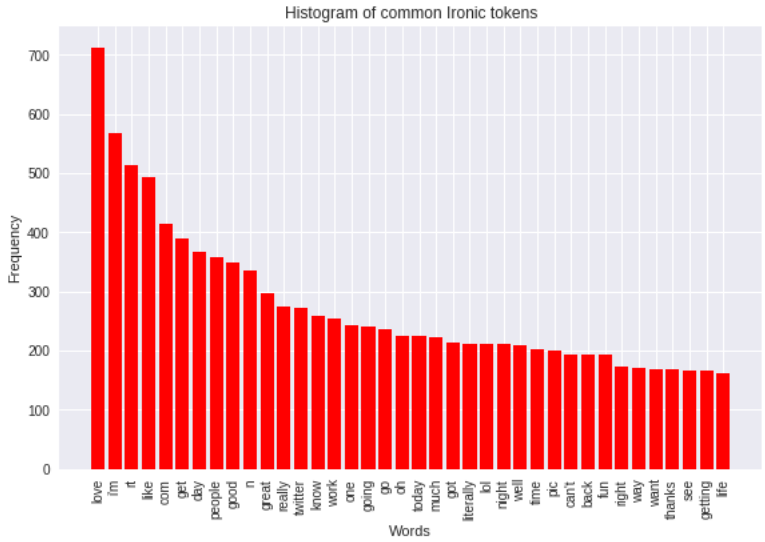
Σχήμα 5.3: Συχνότητα λέξεων στα σαρκαστικά δεδομένα

χαρακτήρες σε αντίθεση με τα κυριολεκτικά που συνήθως εξαντλούν τους 120 χαρακτήρες του tweet, όπως φαίνεται στο Σχήμα 5.5.

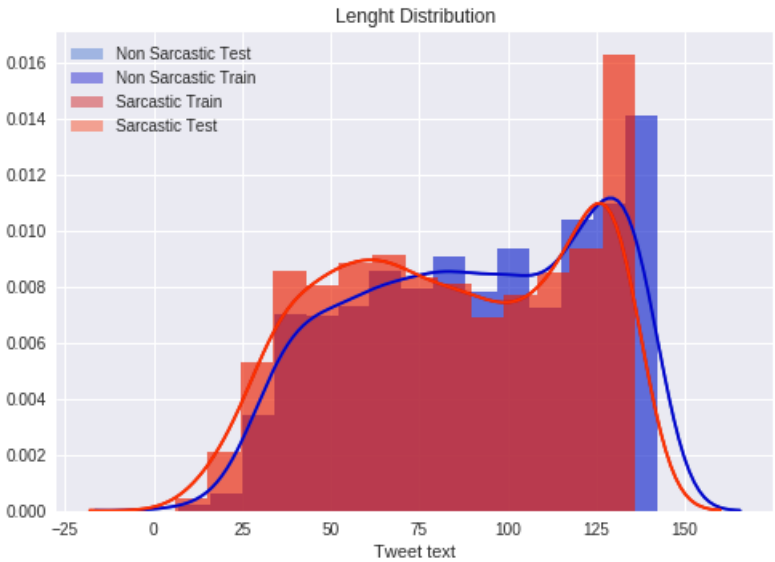
Τα χαρακτηριστικά αυτά είναι απαραίτητο να ληφθούν υπόψιν ώστε να δώσουν την δυνατότητα στους χρησιμοποιούμενους αλγορίθμους να κατανοήσουν τις διαφορές των μεταφορικών και των κυριολεκτικών φαινομένων.

5.2 Προεπεξεργασία δεδομένων

Όπως περιγράφηκε στην προηγούμενη ενότητα τα δεδομένα αποτελούν μικρού μήκους κείμενα, το πολύ 120 χαρακτήρων. Αν παρατηρήσουμε τον τρόπο έκφρασης των χρηστών



Σχήμα 5.4: Συχνότητα λέξεων στα μεταφορικά δεδομένα



Σχήμα 5.5: Μέσο μήκος tweet μεταφορικών δεδομένων

στο Twitter είναι εύκολο να διακρίνουμε ορισμένα χαρακτηριστικά που κυριαρχούν. Μερικά από αυτά η χρήση hashtag στα οποία πιθανόν να υπάρχουν σύνολα από συζευγμένες λέξεις, όπως για παράδειγμα τα #yeahright, #thankgod, #noway και άλλα, οι απαντήσεις σε άλλους χρήστες με την χρήση του @user, η χρήση ανορθόγραφων λέξεων και πιθανόν αρκτικόλεξων όπως το lol, για το laugh out loud καθώς και η χρήση εικονιδίων emoji. Λαμβάνοντας υπόψιν τα χαρακτηριστικά αυτά στο στάδιο προεπεξεργασίας των δεδομένων ακολουθήθηκε η παρακάτω διαδικασία::

- Τα κεφαλαία μετατρέπονται σε μικρά για πιο εύρωστη επεξεργασία του κειμένου
- Αφαιρούνται τα # από το tweet και γίνεται έλεγχος για τυχόν συζευγμένες λέξεις σε

αυτό ώστε να διαχωριστούν και να αντιμετωπιστούν ως λέξεις του κειμένου[9].

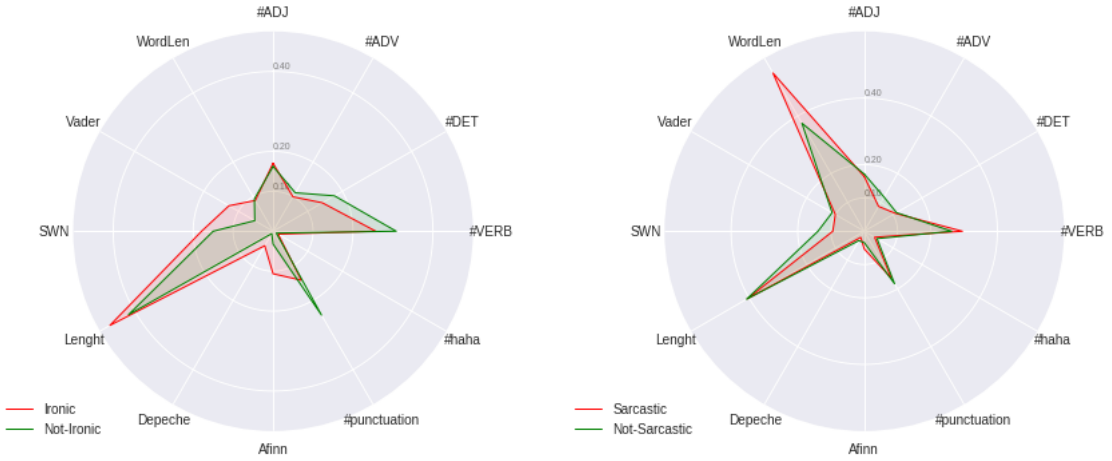
- Αφαιρούνται οι σύνδεσμοι που μπορεί να περιέχονται στο tweet αφού δεν προσδίδουν κάποιο χαρακτηριστικό με βάση το σαρκαστικό περιεχόμενο του tweet.
- Μετατροπή των emoji στην περιγραφή τους ώστε να μπορούν να αντιμετωπιστούν ως λέξεις του κειμένου.
- Διόρθωση ανορθόγραφων λέξεων με βάση την κοντινότερη τους λέξη σε ένα μεγάλο αγγλικό λεξικό.
- Stemming των λέξεων με την χρήση του Porter Stemmer.

5.3 Δημιουργία Χαρακτηριστικών-Feature Engineering

Για κάθε Tweet δημιουργείται ένα διάνυσμα χαρακτηριστικών το οποίο περιέχει 39 στοιχεία που συνδυάζονται με n-gram χαρακτηριστικά των δεδομένων. Τα χαρακτηριστικά αυτά μπορούν να καταταχθούν σε τέσσερις ευρύτερες κατηγορίες.

- 12 συντακτικά χαρακτηριστικά: που καταγράφουν την χρήση ρημάτων, επιρρημάτων, αντωνυμιών, επιθέτων κλπ στο tweet.
- 7 εκφραστικά χαρακτηριστικά: που υποδηλώνουν τον τρόπο έκφρασης του χρήστη και περιλαμβάνουν το πλήθος λέξεων του tweet, το πλήθος emoji που χρησιμοποιήθηκαν, το λόγο emoji ανά λέξη του tweet, το μέσο μήκος κάθε λέξης, το πλήθος σημείων στίξης καθώς και τα επαναλαμβανόμενα γράμματα που μπορεί να δηλώνουν έμφαση όπως το Wooooow.
- 13 συναισθηματικά χαρακτηριστικά: όπως μπορεί να γίνει εύκολα αντιληπτό από τα παραπάνω παραδείγματα η ειρωνεία και ο σαρκασμός αποτελούν μια αντίθεση συναισθημάτων μεταξύ της κυριολεκτικής και της μεταφορικής ερμηνείας του κειμένου. Για τον λόγο αυτό χρησιμοποιήθηκαν διάφορα εργαλεία ανάλυσης συναισθήματος. Συγκεκριμένα υπολογίστηκαν με την χρήση του SentiWordNet[3] ο μέσος όρος αρνητικών και θετικών πολώσεων των λέξεων ανά tweet καθώς και η διαφορά τους. Επίσης χρησιμοποιήθηκε το VADER (Valence Aware Dictionary and sEntiment Reasoner), που έχει εκπαιδευτεί από ένα μεγάλο corpus για την ανάδειξη θετικών και αρνητικών λέξεων στο κείμενο, το οποίο εισάγει την χρήση άρνησης σε λεξικό συναισθηματικής ανάλυσης, για εκφράσεις της μορφής 'not good¨, όπου υπολογίστηκαν οι μέσοι όροι αρνητικών και θετικών συναισθημάτων στο tweet καθώς και η αντίθεση τους. Ταυτόχρονα χρησιμοποιείται άλλος έναν αναλυτής συναισθήματος από την γνωστή βιβλιοθήκη ανάλυσης κειμένου στην Python TextBlob [3], που καταγράφει την φόρτιση του κειμένου μαζί με ένα συντελεστή υποκειμενικότητας, αλλά και το λεξικό Afin[1] που καταγράφει το συναίσθημα του tweet

[3] https://textblob.readthedocs.io/en/dev/api_reference.html#module-textblob.en.sentiments



(α′) Διάγραμμα Χαρακτηριστικών για Ειρωνικά Δεδομένα

(β′) Διάγραμμα Χαρακτηριστικών για Σαρκαστικά Δεδομένα

Σχήμα 5.6: Χαρακτηριστικά Ειρωνικών-Σαρκαστικών Δεδομένων

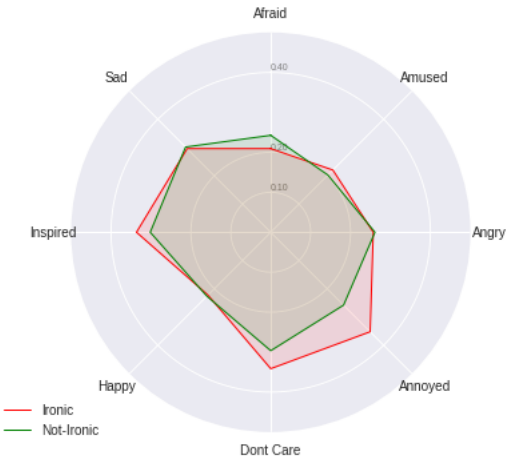
με ένα μοντέλο βασισμένο στις λεξιλογικές αλληλουχίες. Τέλος, καταγράφονται λέξεις που αναδεικνύουν γέλιο στο περιεχόμενο του tweet, όπως τα hahaha, ahaha και τα παράγωγα τους αλλά και εκφράσεις όπως το lol.

Τα παραπάνω δεδομένα προσεγγίζουν την σύνταξη και την μορφολογία του κειμένου, τον τρόπο έκφρασης και την συναισθηματική φόρτιση του χρήστη, αλλά και τις συναισθηματικές αντιθέσεις που περιέχονται σε κάθε σχόλιο. Υπολογίζοντας τις μέσες τιμές των παραπάνω δεδομένων σχεδιάζουμε τα διάγραμματα του Σχήματος 5.6 για να διακρίνουμε τις διαφορές ειρωνείας και σαρκασμού με τα κυριολεκτικά δεδομένα, σε συγκεκριμένα χαρακτηριστικά (το πλήθος επιρρημάτων, αντικειμένων, ρημάτων, προσδιορισμών, το πλήθος εκφράσεων γέλιου, το πλήθος σημείων στίξης, το μέσο μήκος λέξης, το μέσο μέγεθος tweet, καθώς και συναισθηματικές αντιθέσεις με την χρήση των Affin, SentiWordNet, Vader, Depeche). Αντίθετα με όσα αναμέναμε παρατηρούμε μεγαλύτερη έκταση σχολίων στα ειρωνικά δεδομένα, ενώ ταυτόχρονα μεγαλύτερες συναισθηματικές διαφορές με βάση τα εργαλεία Vader, Afinn, SWN. Μια ακόμα αξιόλογη παρατήρηση είναι η χρήση σημαντικά περισσότερων σημείων στίξης, ρημάτων και προσδιορισμών στα κυριολεκτικά δεδομένα. Αντίθετα, στα σαρκαστικά δεδομένα διακρίνουμε την χρήση λέξεων μεγαλύτερης έκτασης και μικρές διαφορές στην χρήση επιρρήματων.

- 8 χαρακτηριστικά διάθεσης: Τα χαρακτηριστικά αυτά εξάγονται με την χρήση του λεξικού DepecheMood[91], που έχει δημιουργηθεί από μια συλλογή κειμένων του Rappler[4] και αναπαριστά κάθε λέξη με ένα διάνυσμα διάθεσης. Συγκεκριμένα, καταγράφει 8 χαρακτηριστικά προδιάθεσης του κειμένου: την χαρά, την λύπη, τον φόβο, τον θυμό, την αδιαφορία, την τέρψη, την ενόχληση και έμπνευση, παρουσιάζοντας state-of-the-art αποτελέσματα. Στο σχήμα 5.7 παρουσιάζονται οι μέσες τιμές των διανυσμάτων διάθε-

[4] www.rappler.com

σης για ειρωνικά και σαρκαστικά δεδομένα. Το πολικό αυτό διάγραμμα αναδυκνύει την επιθετικότητα που χαρακτηρίζει τα ειρωνικά σχόλια σε σύγκριση με τα κυριολεκτικά.



Σχήμα 5.7: Στοιχεία Διάθεσης στα Ειρωνικά Δεδομένα

Τα 39 χαρακτηριστικά που αναφέρθηκαν παραπάνω συνδυάζονται με uni-gram και bi-gram μοντέλα και αποτελούν τον πίνακα χαρακτηριστικών που χρησιμοποιείται στα πειράματα με Νευρωνικά Δίκτυα Εμπρόσθιας τροφοδότησης. Αντίθετα, στα Επαναλαμβανόμενα Νευρωνικά Δίκτυα(RNN) χρησιμοποιούμε προ-εκπαιδευμένα Word Embeddings τα οποία παρουσιάζουν τα καλύτερα αποτελέσματα. Αναλυτικότερα, χρησιμοποιούνται τα GloVe: Global Vectors[74] 100-διαστάσεων που δημιουργήθηκαν από το Standford, τα οποία έχουν εκπαιδευτεί πάνω σε δυο δισεκατομμύρια tweet.

5.4 Αξιολόγηση Αλγορίθμων

Μετά την δημιουργία των απαραίτητων χαρακτηριστικών και την προεπεξεργασία των συνόλων δεδομένων θα χρησιμοποιήσουμε για κάθε ένα από αυτά διάφορους αλγόριθμους ταξινόμησης, όπου και θα αξιολογήσουμε τα αποτελέσματα τους αλλά και θα τα συγκρίνουμε με παρεμφερείς εργασίες. Για την αξιολόγηση θα χρησιμοποιήσουμε τους αλγορίθμους που αναπτύξαμε λεπτομερώς στο Κεφάλαιο 4, ενώ ως μέτρο εσωτερικής σύγκρισης θα χρησιμοποιηθεί ο ταξινομητής SVM που αποτελεί ένα ιδιαίτερα σύνηθες εργαλείο στον ερευνητικό τομέα της αναγνώρισης συναισθήματος(Sentiment Analysis).

5.4.1 Αναγνώριση Ειρωνείας

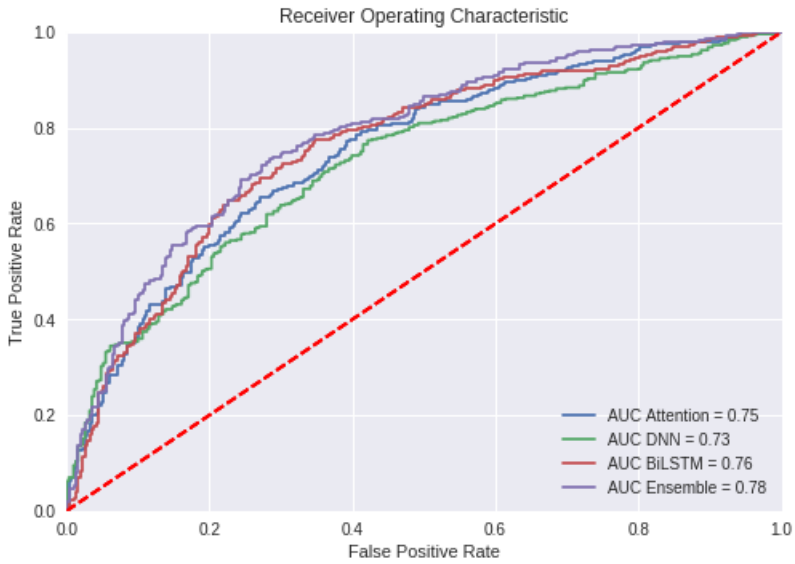
Για το σύνολο ειρωνικών δεδομένων από το SemEval-2018 χρησιμοποιούμε πέντε μετρικές αξιολόγησης, τις: Accuracy, Precision, Recall, $F_1 - Score$ και AUC. Τα αποτελέσματα συνοψίζονται στο Πίνακα 5.3 .

Είναι φανερό πως το μοντέλο συνδυασμού μοντέλων, DESC, που περιγράφηκε στην προηγούμενη ενότητα 3.3.5 παρουσιάζει τα καλύτερα αποτελέσματα όσο αφορά τις μετρικές αξιο-

Αλγόριθμος	Accuracy	Precision	Recall	F_1	AUC
DNN-2inp	0.63	0.64	0.62	0.63	0.65
DNN-Tfidf	0.65	0.65	0.63	0.64	0.70
DNN-uni	0.65	0.68	0.65	0.65	0.72
DNN-All	0.66	0.69	0.66	0.67	0.75
SVM-Tfidf	0.65	0.68	0.65	0.66	0.70
SVM-FeatVec	0.59	0.59	0.59	0.59	0.60
SVM-All	0.66	0.69	0.66	0.67	0.75
AttentionLSTM	0.71	0.70	0.71	0.70	0.75
BiLSTM	0.71	0.71	0.71	0.70	0.76
DESC	**0.74**	**0.73**	**0.73**	**0.73**	**0.78**

Πίνακας 5.3: Σύγκριση αλγορίθμων για ειρωνικά δεδομένα

λόγησης. Όμως, ένα εξίσου σημαντικό γεγονός είναι πως αυξάνεται αισθητά και η περιοχή κάτω από την καμπύλη πρόβλεψης, AUC, όπως φαίνεται στο Σχήμα 5.8.



Σχήμα 5.8: Καμπύλες ROC για τα μοντέλα του συνδυασμού στα Ειρωνικά Δεδομένα

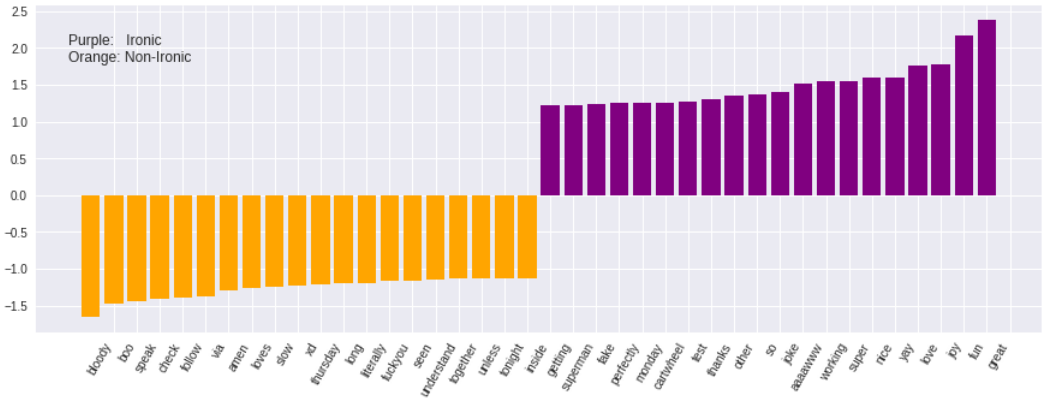
Για να μπορέσουμε να εμβαθύνουμε την ανάλυση μας περαιτέρω, χρησιμοποιούμε μια μέθοδο αναπαράστασης των λέξεων με βάση την σημαντικότητα(feature importance) τους, ως χαρακτηριστικά, στην ταξινόμηση. Στο Σχήμα 5.9 παρουσιάζονται 7 παραδείγματα ειρωνικών δεδομένων τα οποία ταξινομήθηκαν ορθά. Με την χρήση της βιβλιοθήκης eli5[5] υπογραμμίζονται οι λέξεις με την μεγαλύτερη θετική (πράσινο) και αρνητική (κόκκινο) πόλωση, με βάση την συμβολή της Tf-Idf αναπαράστασης τους στην διαδικασία ταξινόμηση. Όμοια, στο Σχήμα 5.10 παρουσιάζεται ένα ιστόγραμμα σχεδιασμένο, επίσης, με βάση τις συχνότητες Tf-idf στο

[5] https://eli5.readthedocs.io/en/latest/index.html

Tweet	Polarity
can u help more conservatives needed on tsu get paid 4 posting stuff like this you can go to	Not-Ironic
just: walked in to starbucks and asked for a tall blonde hahahaha irony	Ironic
not gonna win	Not-Ironic
he is exactly that sort of person weirdo	Ironic
so much sarcasm at work mate 1010 boring 100 dead mate full on shit absolutely sleeping mate cant handle the sarcasm	Ironic
corny jokes are my absolute favorite	Not-Ironic
people complain about my backround pic and all i feel is like hey dont blame me albert e might have spoken those words sarcasm life	Ironic

Σχήμα 5.9: Επιρροή Λέξεων στην εκπαίδευση των αλγορίθμων με την χρήση του εργαλείου eli5

οποίο καταγράφονται οι 20 πιο σημαντικές λέξεις για τα ειρωνικά δεδομένα. Λέξεις με μεγάλη συναισθηματική φόρτιση όπως οι great, fun, joy ή εκφράσεις χαράς όπως οι yay, awwww είναι αυτές που παρουσιάζουν την μεγαλύτερη σημαντικότητα σε ειρωνικά σχόλια, χαρακτηριστικό της δυσκολίας στην ερμηνεία του συναισθήματος σε ειρωνικά δεδομένα.



Σχήμα 5.10: Σημαντικότητα Λέξεων στα Ειρωνικά Δεδομένα

Το συγκεκριμένο σύνολο δεδομένων αποτέλεσε ανοιχτό πρόβλημα στο SemEval-2018 επομένως μπορούμε πολύ εύκολα να συγκρίνουμε τα αποτελέσματα του συνδυαστικού μοντέλου που προτείνεται με μερικά από τα μοντέλα των ερευνητών που αντιμετώπισαν το πρόβλημα, όπως αυτά αναπτύχθηκαν στο Κεφάλαιο 2. Τα αποτελέσματα συνοψίζονται στο πίνακα ;;. Σε όλες τις σχετικές συγκρίσεις αναφερόμαστε σε μετρικές αξιολόγησης οι οποίες χρησιμοποιήθηκαν στις αντίστοιχες εργασίες.

Από τα παραπάνω γίνεται αντιληπτό πως η συνδυαστική μέθοδος αποδίδει αρκετά καλά αποτελέσματα σε ειρωνικά δεδομένα του Αγγλικού Twitter. Η υψηλή Ανάκληση της ομάδας WLV αποδίδει την ικανότητα του μοντέλου να ταξινομεί σωστά το μεγαλύτερο ποσοστό μιας κλάσης επιφέρει όμως λανθασμένες ταξινομήσεις στην άλλη κλάση και συνεπώς χαμηλές μετρικές ακρίβειας και $F_1 - score$, όπου το μοντέλο DESC που προτείνεται αποδίδει καλύτερα.

5.4.2 Αναγνώριση Μεταφορικών Γλωσσικών Φαινομένων

Στο σύνολο αυτό λόγω του μεγάλου πλήθους κλάσεων χρησιμοποιείται σαν μετρική αξιολόγησης η ομοιότητα συνημίτονου αλλά και η ελάχιστη διαφορά τετραγώνων ώστε οι αλγόριθ-

Μέθοδος	Accuracy	Precision	Recall	F_1
THU-NGN	0.73	0.63	0.80	0.71
NTUA-SLP	0.73	0.65	0.69	0.67
WLV	0.64	0.53	**0.84**	0.65
NIHRIO, NCL	0.70	0.61	0.69	0.65
SIRIUS-LC	0.68	0.60	0.59	0.60
DESC	**0.74**	**0.73**	0.73	**0.73**

Πίνακας 5.4: Σύγκριση μοντέλου με συγγενείς εργασίες για ειρωνικά δεδομένα από το SemEval-2018

μοι να έχουν την δυνατότητα να συγκριθούν με άλλους διαγωνιζόμενους της συγκεκριμένης εργασίας.

Αλγόριθμος	Cosine	MSE
DNN-2inp	0.602	4.23
DNN-Tfidf	0.71	3.17
DNN-uni	0.69	8.43
DNN-FeatVec	0.68	3.23
DNN-All	0.789	2.79
SVM-Tfidf	0.72	2.89
SVM-FeatVec	0.70	3.39
SVM-All	0.723	2.81
AttentionLSTM	0.749	2.86
BiLSTM	0.704	3.22
DESC	**0.801**	**2.68**

Πίνακας 5.5: Σύγκριση αλγορίθμων για μεταφορικά δεδομένα από το SemEval-2015

Όπως και το προηγούμενο σύνολο δεδομένων έτσι και αυτό αποτέλεσε ανοιχτό πρόβλημα του SemEval-2015, δίνοντας την δυνατότητα να συγκριθούν τα αποτελέσματα της συνδυαστικής μεθόδου που προτείνεται με τα αποτελέσματα των ερευνητών που ασχολήθηκαν με το συγκεκριμένο πρόβλημα. Τα σχετικά αποτελέσματα συνοψίζονται, κατά αντιστοιχία με το προηγούμενο πρόβλημα αναγνώρισης ειρωνείας, στους Πίνακες 5.5 και 5.6.

5.4.3 Αναγνώριση Σαρκασμού

Όπως αναφέρθηκε στο κεφάλαιο 5.1.3 για την ανίχνευση σαρκασμού χρησιμοποιήθηκαν δυο σύνολα δεδομένων. Η αξιολόγηση των αλγορίθμων έγινε σε κάθε ένα ξεχωριστά χρησιμοποιώντας τις ίδιες μετρικές αξιολόγησης με τα δεδομένα Ειρωνείας. Λόγω των διαθέσιμων περιορισμένων υπολογιστικών δυνατοτήτων χρησιμοποιήθηκαν μόνο 8.000 tweet από τα 39.000

Αλγόριθμος	Cosine	MSE
ClaC	0.758	**2.12**
UPF	0.710	2.46
LLT-PolyU	0.678	2.60
LT3	0.658	3.40
elirf	0.658	3.10
DESC	**0.801**	2.68

Πίνακας 5.6: Σύγκριση μοντέλου με συγγενείς εργασίες για μεταφορικά δεδομένα

που περιέχουν τα δεδομένα εκπαίδευσης των Ghosh&Veale σε συνδυασμό με 2.000 tweet για επικύρωση και 2.000 tweet αξιολόγησης, όπως ακριβώς στην εργασία [34].

Αλγόριθμος	Accuracy	Precision	Recall	F_1	AUC
DNN-2inp	0.79	0.78	0.81	0.79	0.83
DNN-Tfidf	0.844	0.85	0.84	0.84	0.86
DNN-uni	0.76	0.81	0.76	0.70	0.74
DNN-FeatVec	0.68	0.68	0.67	0.67	0.69
DNN-All	0.84	0.85	0.84	0.84	0.87
SVM-Tfidf	0.53	0.76	0.55	0.45	0.51
SVM-FeatVec	0.62	0.62	0.62	0.62	0.68
SVM-All	0.74	0.75	0.74	0.74	0.81
AttentionLSTM	0.80	0.86	0.80	0.80	0.88
BiLSTM	0.82	0.84	0.81	0.81	0.88
DESC	**0.85**	**0.85**	**0.84**	**0.85**	**0.89**

Πίνακας 5.7: Σύγκριση αλγορίθμων για σαρκαστικά δεδομένα των Ghosh&Veale[34]

Τα αποτελέσματα και σε αυτή την περίπτωση δείχνουν μια μικρή ανωτερότητα της τεχνικής συνδυασμού DESC σε σχέση με την απόδοση μεμονωμένων αλγορίθμων (βλέπε Πίνακα 5.7). Παρόλα αυτά, στα δεδομένα αυτά η τεχνική συνδυασμού DESC παρουσιάζει μια σημαντική απόκλιση από τα αποτελέσματα του state-of-the-art μοντέλο που παρουσιάστηκε στο [34]. Σαφώς, στα αποτελέσματα αυτά θα πρέπει να ληφθεί υπόψιν οτι χρησιμοποιήθηκε μόνο το 20% των διαθέσιμων δεδομένων εκπαίδευσης του μοντέλου των Ghosh&Veale(βλέπε Πίνακα 5.8).

Στο δεύτερο μέρος της ανίχνευσης σαρκασμού στο Twitter χρησιμοποιούμε ένα πιο μικρό αλλά ακριβές σύνολο δεδομένων. Τα δεδομένα αυτά, του Riloff[82], αξίζει να σημειωθεί πως δεν είναι ισορροπημένα όπως τα προηγούμενα δεδομένα που χρησιμοποιήθηκαν κατά την πειραματική αξιολόγηση. Η μη ισορροπημένη κατανομή των δεδομένων του Riloff μας δίνει την δυνατότητα να εξετάσουμε την ευρωστία των μοντέλων που χρησιμοποιούμε, γνωρίζοντας

Αλγόριθμος	Precision	Recall	F_1
Ghosh&Veale	0.92	0.92	0.92
DESC	0.85	0.84	0.85

Πίνακας 5.8: Σύγκριση μεθόδου συνδυασμού με το μοντέλο των Ghosh&Veale[34]

πως το πλήθος των μη σαρκαστικών tweet είναι ρεαλιστικά πάντοτε μεγαλύτερο από αυτό των σαρκαστικών. Επιπρόσθετα, για να υπάρχει η δυνατότητα σύγκρισης με συγγενείς εργασίες που χρησιμοποίησαν τα δεδομένα του Riloff δεν αντιμετωπίστηκε η μη-ισορροπημένη δομή των δεδομένων. Τα αποτελέσματα παρατίθενται στον Πίνακα 5.9.

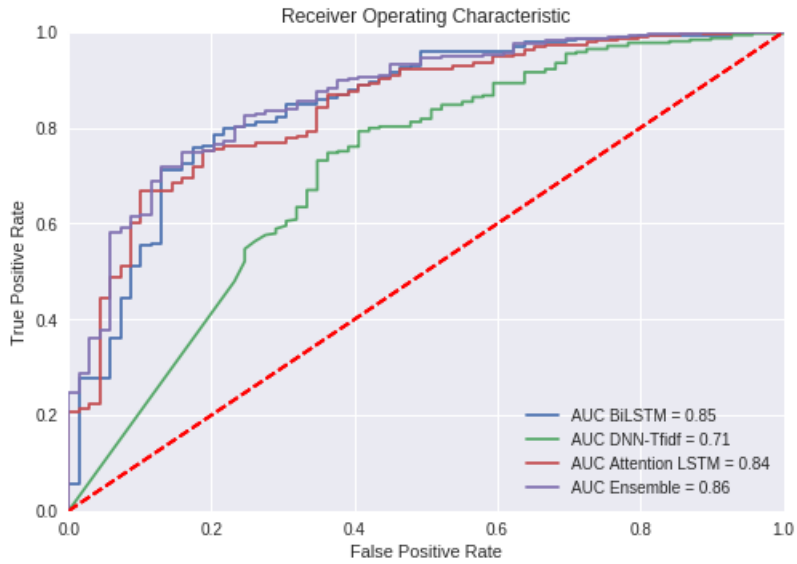
Αλγόριθμος	Accuracy	Precision	Recall	F_1	AUC
DNN-2inp	0.82	0.78	0.81	0.79	0.84
DNN-Tfidf	0.83	0.81	0.83	0.8	0.72
DNN-uni	0.79	0.78	0.81	0.79	0.74
DNN-FeatVec	0.81	0.81	0.83	0.80	0.72
DNN-All	0.83	0.81	0.83	0.81	0.82
SVM-Tfidf	0.82	0.80	0.82	0.80	0.80
SVM-FeatVec	0.82	0.73	0.81	0.75	0.76
SVM-All	0.83	0.81	0.83	0.81	0.81
AttentionLSTM	0.85	0.83	0.85	0.83	0.84
BiLSTM	0.85	0.85	0.85	0.85	0.85
DESC	**0.87**	**0.87**	**0.86**	**0.87**	**0.86**

Πίνακας 5.9: Σύγκριση αλγορίθμων για σαρκαστικά δεδομένα του Riloff[82]

Από το Πίνακα 5.10, μπορούμε και πάλι να παρατηρήσουμε πως και στο σύνολο των συγκεκριμένων δεδομένων η τεχνική συνδυασμού επιδεικνύει την καλύτερη απόδοση για την ανίχνευση μεταφορικών γλωσσικών φαινομένων, έχοντας μεγάλη διαφορά από την απόδοση του Riloff. Χρησιμοποιούμε παράλληλα σαν μέτρο σύγκρισης την αξιολόγηση του αλγορίθμου των Ghosh & Veale πάνω στα δεδομένα του Riloff.

Αλγόριθμος	Precision	Recall	F_1
Riloff[82]	0.44	0.62	0.51
Ghosh[34]	0.88	0.88	0.88
DESC	0.87	0.86	0.87

Πίνακας 5.10: Σύγκριση αλγορίθμων για σαρκαστικά δεδομένα του Riloff[82]



Σχήμα 5.11: Καμπύλες ROC για τα μοντέλα του συνδυασμού στα Σαρκαστικά Δεδομένα

5.4.3.1 Σύνοψη Σαρκαστικών Αποτελεσμάτων

Έχοντας προσπαθήσει να επιλύσουμε το φαινόμενο του σαρκασμού σε δύο διαφορετικά ως προς τον τύπο και την κατανομή τους δεδομένα θα δοκιμάσουμε να συνοψίσουμε τα αποτελέσματα αυτά με συγγενείς εργασίες που χρησιμοποίησαν ίδια μορφολογικά δεδομένα από το Twitter. Είναι σαφές πως το πλήθος, ο τύπος και η κατανομή των δεδομένων επηρεάζουν σημαντικά την εκπαίδευση και μελλοντικά την ικανότητα του μοντέλου να ανιχνεύει σαρκαστικά δεδομένα. Στον Πίνακα 5.11 καταγράφουμε συνοπτικά την εξέλιξη της έρευνας στα σαρκαστικά δεδομένα όσον αφορα τις χρησιμοποιούμενες, στις αντίστοιχες εργασίες, μετρικές αξιολόγησης και σχετικά αποτελέσματα.

Αλγόριθμος	Accuracy	Precision	Recall	F_1	AUC
Davidov[26]-2010	-	0.79	0.86	0.82	-
Tsur[94]-2010	-	0.76	0.81	0.80	-
Ibanez[38]-2011	0.76	-	-	-	-
Riloff[82]-2013	0.78	0.44	0.62	0.51	-
Liebrecht[56]-2013	-	-	-	-	0.76
Rajadesingan[77]-2015	0.83	-	-	-	0.83
Ghosh&Veale[34]-2015	-	**0.92**	**0.92**	**0.92**	-
DESC-2018	**0.85**	0.85	0.84	0.85	**0.89**

Πίνακας 5.11: Σύγκριση αλγορίθμων για σαρκαστικά δεδομένα στο Twitter

Σημείωση: Το '-' χρησιμοποιείται για τις μετρικές αξιολόγησης για τις οποίες δεν αναφέρονται αποτελέσματα στις αντίστοιχες δημοσιεύσεις.

Παρατηρούμε πως στο σύνολο των μεταφορικών φαινομένων η τεχνική συνδυασμού DESC

που δημιουργήσαμε υπερσκελίζει state-of-the-art αποτελέσματα και επιδεικνύει καλύτερη συμπεριφορά για το σύνολο των μετρικών αξιολόγησης (βλέπε Πίνακα 5.11), πέραν τις περίπτωσης των Ghosh&Veale[34] .

Κεφάλαιο 6

Συμπεράσματα

Στην παρούσα διπλωματική εργασία αντιμετωπίζουμε το πρόβλημα ανίχνευσης και αναγνώρισης *μεταφορικών γλωσσικών φαινομένων* (ΜΓΦ) στο μέσο κοινωνική δικτύωσης Twitter και της *ταξινόμησης* (classifcation) των σχετικών tweet. Αναλυτικότερα, συλλέχθηκαν δεδομένα από τέσσερεις διαφορετικές πηγές [42, 35, 34, 82] τα οποία αναφέρονται σε τρία διαφορετικά ΜΦΓ τα οποία εξετάζονται και αντιμετωπίζονται στη παρούσα εργασία, συγκεκριμένα: αναγνώριση ειρωνείας, σαρκασμού και μεταφοράς και ταξινόμηση των αντίστοιχων κειμένων.

Τα δεδομένα ποικίλουν και όσον αφορά στο γλωσσικό φαινόμενο που τα χαρακτηρίζει αλλά και ως προς την κατανομή και ισορροπίας (balance) τους (όσον αφορά στο πλήθος των αντίστοιχων κατηγοριών, π.χ., 'σαρκαστικό' vs. 'μη-σαρκαστικό'). Δύο από τα σύνολα δεδομένων, ειρωνείας[42] και σαρκασμού[34] χαρακτηρίζονται σαν πλήρως ισορροπημένα (balanced) δυαδικής ταξινόμησης προβλήματα, με την ευρωστία των αλγορίθμων που αναπτύχθηκαν και εξετάστηκαν να αξιολογείτε με ένα μη-ισορροπημένα (unbalanced) σύνολο δεδομένων [82]. Ταυτόχρονα, χρησιμοποιείται ένα σύνολο δεδομένων που αποτελείται από *σαρκαστικά*, *ειρωνικά* και *μεταφορικά* tweet τα οποία ταξινομούνται βάση του συναισθήματος που εκφράζουν σε μια διακριτή κλίμακα 11-κλάσεων από το -5 (απόλυτα αρνητικό συναίσθημα) μέχρι το 5 (απόλυτα θετικό συναίσθημα). Με αυτόν τον τρόπο εξετάζουμε και αντιμετωπίζουμε διαφορετικά προβλήματα αναγνώρισης ΜΦΓ όσον αφορά στο τύπο της κατηγοριοποίησης (δυαδική και διακριτή πολλαπλών κατηγοριών), αλλά και όσον αφορά στη κατανομή των δεδομένων (ανισοροπία στο πλήθος των κατηγοριών).

Στην συνέχεια τα δεδομένα προεπεξεργάζονται με τεχνικές που αναλύονται στην ενότητα 5.2 ώστε να δημιουργηθούν τα κατάλληλα Embeddings για την αναπαράσταση των αντίστοιχων συνόλων δεδομένων εκπαίδευσης. Επιπρόσθετα, εξάγονται χαρακτηριστικά που καταδεικνύουν εκφραστικά, συντακτικά, συναισθηματικά και ψυχολογικά γνωρίσματα των σχετικών κειμένων. Τα χαρακτηριστικά που εξάγονται αλλά και τα Word Embeddings τροφοδοτούν ένα σύνολο από διαφορετικές μεθοδολογίες και αλγορίθμους *βαθιάς μηχανικής μάθησης*. Παρατηρούμε πως τα μοντέλα βαθιάς μάθησης που αποδίδουν καλύτερα, συγκρινόμενα μεταξύ τους αλλά και με Μηχανές Διανυσματικής Υποστήριξης (SVM), είναι αρχιτεκτονικές Αμφίδρομου LSTM (βλ. 4.1), LSTM με μηχανισμό προσοχής (βλ. 4.2) και βαθιού νευρωνικού δικτύου

(βλ. 4.3).

Τα προαναφερθέντα μοντέλα συνδυάζονται με την τεχνική Ensemble Ανεκτικής Ταξινόμησης (soft classification) και τη δημιουργία ενός πρωτότυπου ταξινομητή, Deep Ensemble Soft Classifier-DESC. Ο ταξινομητής DESC επιδεικνύει πολύ καλή συμπεριφορά στα περισσότερα σύνολα δεδομένων, όσον αφορά σε state-of-the-art δημοσιευμένα αποτελέσματα. Ενδεικτικά, τα αποτελέσματα του ταξινομητή DESC αποδίδουν F_1=0.73 στα ειρωνικά δεδομένα σε σύγκριση με το F_1=0.71 που επιτυγχάνει ο καλύτερος ταξινομητής της ομάδας THU-NGN[101] στο διεθνή διαγωνισμό SemEval-2018, ενώ στα μεταφορικά δεδομένα παρουσιάζουν ομοιότητα συνημιτόνου 0.801, αρκετά μεγαλύτερη από την ομάδα CLaC[70] που επέδειξε τα καλύτερα αποτελέσματα, 0.758, στο διεθνη διαγωνισμό SemEval-2015. Αναφορικά με τα σαρκαστικά δεδομένα, χρησιμοποιώντας το 20% των δεδομένων εκπαίδευσης των Ghosh&Veale, το μοντέλο DESC, με F_1=0.85 δεν προσεγγίζει τα αποτελέσματα των Ghosh&Veale[34] (F_1=0.92) αλλά διαφέρει σημαντικά από τον ταξινομητή SVM, που στα αντίστοιχα δεδομένα αποδίδει F_1=0.74. Αντίθετα με τα δεδομένα των Ghosh&Veale, στα μη-ισορροπημένα δεδομένα του Riloff, το μοντέλο DESC προσεγγίζει την state-of-the-art απόδοση F_1=0.88 των Ghosh&Veale, αποδίδοντας F_1=0.87. Το γεγονός αυτό μας δίνει την δυνατότητα να αποδώσουμε ένα μέρος της απόκλισης στο σύνολο δεδομένων [34] στην ελλιπή εκπαίδευση του μοντέλου σε σχέση με το state-of-the-art μοντέλο.

Ένα ακόμα ιδιαίτερα θετικό χαρακτηριστικό του μοντέλου DESC είναι η αύξηση του εμβαδού κάτω από την καμπύλη, ROC(AUC metric), συγκρινόμενο με τα αντίστοιχα ποσοστά μεμονωμένων αλγορίθμων. Όλα τα παραπάνω μπορούν να συγκλίνουν στη διαπίστωση πως το συνδυαστικό μοντέλο DESC παρουσιάζει βέλτιστα αποτελέσματα στη πλειοψηφία του συνόλου δεδομένων, χωρίς αυτό να σημαίνει πως δεν επιδέχεται βελτιώσεις.

6.1 Μελλοντικά Σχέδια και Δουλειά

Η ανίχνευση ΜΓΦ αποτελεί ένα ανοιχτό ζήτημα στο πεδίο της επεξεργασίας φυσικής γλώσσας (natural language processing) και επιφέρει αρκετές βελτιώσεις ώστε να επιτευχθούν βέλτιστα και εύρωστα αποτελέσματα. Το DESC περιλαμβάνει πληροφορίες από τα uni-gram και bi-gram μοντέλα, τα οποία βασίζονται σε μετρικές συχνότητας όρων (term frequency), Tf-idf, σε συνδυασμό με στοιχεία που αποτυπώνουν βασικά χαρακτηριστικά του συναισθήματος και της διάθεσης των κειμένων αναφοράς και των αντίστοιχων χρηστών κοινωνικών δικτύων.

Παρόλα αυτά, όπως φαίνεται στο [77], τα ΜΓΦ χαρακτηρίζονται από νοηματικές αντιθέσεις, κάτι το οποίο μας οδηγεί στη δυνατότητα αξιοποίησης ως χαρακτηριστικών εκείνων των στοιχείων που αποτυπώνουν τις αντιθέσεις συναισθημάτων, αλλά και χρονικών αντιθέσεων. Για παράδειγμα, είναι γνωστό ότι η χρήση ενεστώτα και αορίστου χρόνου στο ίδιο κείμενο αποτελεί σημαντική ένδειξη σαρκασμού και ειρωνείας, και μπορεί να αξιοποιηθεί στο χαρακτηρισμό και τη τελική κατάταξη του κειμένου. Με βάση την έρευνα του Rajadesingan αναδεικνύεται ότι η ανάλυση συμπεριφοράς του χρήστη θα μπορούσε να αξιοποιηθεί χωρίς απαραίτητα να απαιτείται το ιστορικό του, αξιοποιώντας στοιχεία τα οποία μπορούν να χαρακτηρίσουν το 'χαρακτήρα' του, τα οποία μπορούν να εξαχθούν από τα αντίστοιχα tweet του. Ένα ακόμα ενδιαφέρον

χαρακτηριστικό που θα μπορούσε να ενσωματωθεί στο DESC είναι η καταμέτρηση λέξεων στο tweet που χρησιμοποιούνται κυρίως στον προφορικό λόγο, όπως προτείνει ο Barbieri[5] με την χρήση του ANC λεξικού. Τα αντιφατικά επιρρήματα είναι επίσης ένα χαρακτηριστικό που καταδεικνύει το στοιχείο του 'απρόσμενου', που εμφανίζεται ιδιαίτερα σε ΜΓΦ, όπως καταγράφει και αναδεικνύει ο Reyes[81]. Τέλος, η χρήση του εργαλείου Linguistic Inquiry and Word Count (LIWC) για την εξαγωγή λεκτικών χαρακτηριστικών από το κείμενο με την δημιουργία διανυσμάτων 90-διαστάσεων , όπως προτείνει ο Ibanez[38], θα μπορούσε να αποτελέσει πρόσθετα χαρακτηριστικά στην ανάλυση που παρουσιάστηκε στην ενότητα 5.3.

Τέλος, η αξιοποίηση των μεθοδολογιών, τεχνικών, αλγορίθμων και αποτελεσμάτων της παρούσας εργασίας σε ένα *ενιαίο και διαλειτουργικό πακέτο*, μέσω και της κατάλληλης προσαρμογής και ανάπτυξης σχετικών προγραμματικών διαμέσων (APIs), μπορεί να οδηγήσει στην ανάπτυξη *αυτοματοποιημένων υπηρεσιών αναγνώρισης και ταυτοποίησης*, αλλά και *εξαγωγής τάσεων* (trend analytics) και χαρακτηρισμού χρονικών ή/και τοπικών καταστάσεων και φαινομένων (π.χ., επιδημιών, φυσικών καταστροφών κλπ.)

Βιβλιογραφία

[1] F. AArup Nielsen. "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs". In: *ArXiv e-prints* (Mar. 2011).

[2] Salvatore Attardo. "Irony as relevant inappropriateness". In: *Journal of Pragmatics* 32.6 (2000), pp. 793–826. ISSN: 0378-2166. DOI: https://doi.org/10.1016/S0378-2166(99)00070-3. URL: http://www.sciencedirect.com/science/article/pii/S0378216699000703.

[3] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. *SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining.* Vol. 10. Jan. 2010.

[4] Francesco Barbieri, Francesco Ronzano, and Horacio Saggion. "UPF-taln: SemEval 2015 Tasks 10 and 11. Sentiment Analysis of Literal and Figurative Language in Twitter". In: *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015).* Denver, Colorado: Association for Computational Linguistics, June 2015, pp. 704–708. URL: http://www.aclweb.org/anthology/S15-2119.

[5] Francesco Barbieri and Horacio Saggion. "Modelling Irony in Twitter". In: *EACL.* 2014.

[6] Marco Baroni, Georgiana Dinu, and Germán Kruszewski. "Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors". In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).* Baltimore, Maryland: Association for Computational Linguistics, 2014, pp. 238–247. DOI: 10.3115/v1/P14-1023. URL: http://www.aclweb.org/anthology/P14-1023.

[7] M. Basavanna. *Dictionary of Psychology.* Allied Publishers, 2000. ISBN: 978-81-7764-030-4. URL: https://books.google.gr/books?id=gjNAaP-JU0EC.

[8] Leonard E. Baum and Ted Petrie. "Statistical Inference for Probabilistic Functions of Finite State Markov Chains". In: *The Annals of Mathematical Statistics* 37.6 (1966), pp. 1554–1563. DOI: 10.1214/aoms/1177699147. URL: https://doi.org/10.1214/aoms/1177699147.

[9] Christos Baziotis, Nikos Pelekis, and Christos Doulkeridis. "DataStories at SemEval-2017 Task 4: Deep LSTM with Attention for Message-level and Topic-based Sentiment Analysis". In: *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. Vancouver, Canada: Association for Computational Linguistics, Aug. 2017, pp. 747–754.

[10] Christos Baziotis et al. "NTUA-SLP at SemEval-2018 Task 3: Tracking Ironic Tweets using Ensembles of Word and Character Level Attentive RNNs". In: *Proceedings of The 12th International Workshop on Semantic Evaluation*. New Orleans, Louisiana: Association for Computational Linguistics, 2018, pp. 613–621. URL: http://aclweb.org/anthology/S18-1100.

[11] Yoshua Bengio et al. "A Neural Probabilistic Language Model". In: *J. Mach. Learn. Res.* 3 (Mar. 2003), pp. 1137–1155. ISSN: 1532-4435. URL: http://dl.acm.org/citation.cfm?id=944919.944966.

[12] Julia Birke and Anoop Sarkar. *A Clustering Approach for Nearly Unsupervised Recognition of Nonliteral Language*. Jan. 2006.

[13] Daria Bogdanova. "A Framework for Figurative Language Detection Based on Sense Differentiation". In: *ACL*. 2010.

[14] Mondher Bouazizi and Tomoaki Ohtsuki. "A Pattern-Based Approach for Sarcasm Detection on Twitter". In: *IEEE Access* 4 (2016), pp. 1–1.

[15] Andrea Bowes and Albert Katz. "When Sarcasm Stings". In: *Discourse Processes* 48 (2011), pp. 215–236.

[16] *Building semantic concordances. In C. Fellbaum (Ed.), WordNet: An electronic lexical database*. (pp. 199–216): Cambridge, MA: MIT Press., 1998.

[17] Konstantin Buschmeier, Philipp Cimiano, and Roman Klinger. *An Impact Analysis of Features in a Classification Approach to Irony Detection in Product Reviews*. Jan. 2014. DOI: 10.3115/v1/W14-2608.

[18] Q Butler. *The istituto oratoria of quintilian*. 1953.

[19] Paula Carvalho et al. "Clues for Detecting Irony in User-Generated Contents: Oh... It's "so easy" ;-)". In: *International Conference on Information and Knowledge Management, Proceedings* (2009).

[20] Paula Carvalho et al. "Clues for Detecting Irony in User-generated Contents: Oh... It's "So Easy" ;-)". In: *Proceedings of the 1st International CIKM Workshop on Topic-sentiment Analysis for Mass Opinion*. TSA '09. New York, NY, USA: ACM, 2009, pp. 53–56. ISBN: 978-1-60558-805-6. DOI: 10.1145/1651461.1651471. URL: http://doi.acm.org/10.1145/1651461.1651471.

[21] Danqi Chen and Christoper Manning. *A Fast and Accurate Dependency Parser using Neural Networks*. Jan. 2014. DOI: 10.3115/v1/D14-1082.

[22] Richard Chin. "The Science of Sarcasm? Yeah, Right". In: *https://www.smithsonianmag.com/science-nature/the-science-of-sarcasm-yeah-right-25038/* ().

[23] Ronan Collobert and Jason Weston. "A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning". In: *Proceedings of the 25th International Conference on Machine Learning*. ICML '08. New York, NY, USA: ACM, 2008, pp. 160–167. ISBN: 978-1-60558-205-4. DOI: 10.1145/1390156.1390177. URL: http://doi.acm.org/10.1145/1390156.1390177.

[24] H Colston and R Gibbs. *A brief history of irony*. Jan. 2007.

[25] A.P. Cowie, R. Mackin, and I.R. McCaig. *Oxford dictionary of current idiomatic English: Phrase, clause and sentence idioms*. . 2. Oxford University Press, 1985. URL: https://books.google.gr/books?id=wYS_tAEACAAJ.

[26] Dmitry Davidov, Oren Tsur, and Ari Rappoport. "Semi-supervised Recognition of Sarcastic Sentences in Twitter and Amazon". In: *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*. CoNLL '10. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 107–116. ISBN: 978-1-932432-83-1. URL: http://dl.acm.org/citation.cfm?id=1870568.1870582.

[27] Tom De Smedt and Walter Daelemans. *Pattern for Python*. Vol. 13. June 2012.

[28] Ljiljana Dolamic and Jacques Savoy. "When Stopword Lists Make the Difference". In: *Journal of the American Society for Information Science and Technology* 61 (2010).

[29] Dynel Marta. "Linguistic approaches to (non)humorous irony". In: *HUMOR* 27.4 (2014), p. 537. ISSN: 16133722. DOI: 10.1515/humor-2014-0097. URL: https://www.degruyter.com/view/j/humr.2014.27.issue-4/humor-2014-0097/humor-2014-0097.xml (visited on 07/31/2018).

[30] Rong-En Fan et al. "LIBLINEAR: A Library for Large Linear Classification". In: *J. Mach. Learn. Res.* 9 (June 2008), pp. 1871–1874. ISSN: 1532-4435. URL: http://dl.acm.org/citation.cfm?id=1390681.1442794.

[31] anna Feldman and J. Peng. "An approach to automatic figurative language detection : A pilot study". In: 2009.

[32] Robert M. French and Christophe Labiouse. *Four Problems with Extracting Human Semantics from Large Text Corpora*.

[33] F. A. Gers and J. Schmidhuber. "Recurrent nets that time and count". In: *Proceedings of the IEEE-INNS-ENNS*. Vol. 3. 2000, 189–194 vol.3. DOI: 10.1109/IJCNN.2000.861302.

[34] Aniruddha Ghosh and Tony Veale. *Fracking Sarcasm using Neural Network*. June 2016. DOI: 10.13140/RG.2.2.16560.15363.

[35] Aniruddha Ghosh et al. *SemEval-2015 Task 11: Sentiment Analysis of Figurative Language in Twitter*. June 2015. DOI: 10.18653/v1/S15-2080.

[36] Debanjan Ghosh, Weiwei Guo, and Smaranda Muresan. "Sarcastic or Not: Word Embeddings to Predict the Literal or Sarcastic Meaning of Words". In: *EMNLP*. 2015.

[37] Mayte Giménez, Ferran Pla, and Lluís-F. Hurtado. "ELiRF: A SVM Approach for SA tasks in Twitter at SemEval-2015". In: *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. Denver, Colorado: Association for Computational Linguistics, June 2015, pp. 574–581. URL: http://www.aclweb.org/anthology/S15-2096.

[38] Roberto I. González-Ibáñez, Smaranda Muresan, and Nina Wacholder. "Identifying Sarcasm in Twitter: A Closer Look". In: *ACL*. 2011.

[39] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[40] Zellig S. Harris. "Distributional Structure". In: *WORD* 10.2-3 (1954), pp. 146–162. DOI: 10.1080/00437956.1954.11659520. URL: https://doi.org/10.1080/00437956.1954.11659520.

[41] D. Hazarika et al. "CASCADE: Contextual Sarcasm Detection in Online Discussion Forums". In: *ArXiv e-prints* (May 2018).

[42] Cynthia Van Hee, Els Lefever, and Véronique Hoste. "SemEval-2018 Task 3: Irony Detection in English Tweets". In: *SemEval@NAACL-HLT*. 2018.

[43] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: *Neural Comput.* 9.8 (Nov. 1997), pp. 1735–1780. ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735. URL: http://dx.doi.org/10.1162/neco.1997.9.8.1735.

[44] Yu-Hsiang Huang, Hen-Hsen Huang, and Hsin-Hsi Chen. "Irony Detection with Attentive Recurrent Neural Networks". In: *ECIR*. 2017.

[45] Clayton J. Hutto and Eric Gilbert. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text". In: *ICWSM*. 2014.

[46] O. İrsoy and C. Cardie. "Bidirectional Recursive Neural Networks for Token-Level Labeling with Structure". In: *ArXiv e-prints* (Dec. 2013).

[47] KAREN SPARCK JONES. "A STATISTICAL INTERPRETATION OF TERM SPECIFICITY AND ITS APPLICATION IN RETRIEVAL". In: *Journal of Documentation* 28.1 (1972), pp. 11–21. DOI: 10.1108/eb026526. URL: https://doi.org/10.1108/eb026526.

[48] A. Joshi, P. Bhattacharyya, and M. J. Carman. "Automatic Sarcasm Detection: A Survey". In: *ArXiv e-prints* (Feb. 2016).

[49] M. Khodak, N. Saunshi, and K. Vodrahalli. "A Large Self-Annotated Corpus for Sarcasm". In: *ArXiv e-prints* (Apr. 2017).

[50] D. P. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization". In: *ArXiv e-prints* (Dec. 2014).

[51] Ioannis Korkontzelos et al. *SemEval-2013 task 5: Evaluating Phrasal Semantics*. June 2013.

[52] Roger J. Kreuz and Sam Glucksberg. "How to Be Sarcastic: The Echoic Reminder Theory of Verbal Irony". In: *Journal of Experimental Psychology: General* 118.4 (1989), pp. 374–386.

[53] L. Kumar, A. Somani, and P. Bhattacharyya. ""Having 2 hours to write a paper is fun!": Detecting Sarcasm in Numerical Portions of Text". In: *ArXiv e-prints* (Sept. 2017).

[54] Vladimir Iosifovich Levenshtein. "Binary codes capable of correcting deletions, insertions and reversals." In: *Soviet Physics Doklady* 10.8 (Feb. 1966), pp. 707–710.

[55] Linlin Li and Caroline Sporleder. *Linguistic Cues for Distinguishing Literal and Non-Literal Usages*. Vol. 2. Aug. 2010.

[56] Christine Liebrecht, Florian Kunneman, and Antal van den Bosch. "The perfect solution for detecting sarcasm in tweets #not". In: *WASSA@NAACL-HLT*. 2013.

[57] Seppo Linnainmaa. *The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors*. Finnish. 1970.

[58] Nick Littlestone. "Learning Quickly When Irrelevant Attributes Abound: A New Linear-Threshold Algorithm". In: *Machine Learning* 2.4 (Apr. 1988), pp. 285–318. ISSN: 1573-0565. DOI: 10.1023/A:1022869011914. URL: https://doi.org/10.1023/A:1022869011914.

[59] Julie B. Lovins. "Development of a stemming algorithm". In: *Mechanical Translation and Computational Linguistics* 11 (1968), pp. 22–31.

[60] H. P. Luhn. "The Automatic Creation of Literature Abstracts". In: *IBM J. Res. Dev.* 2.2 (Apr. 1958), pp. 159–165. ISSN: 0018-8646. DOI: 10.1147/rd.22.0159. URL: http://dx.doi.org/10.1147/rd.22.0159.

[61] Warren S. McCulloch and Walter Pitts. "A logical calculus of the ideas immanent in nervous activity". In: *The bulletin of mathematical biophysics* 5.4 (Dec. 1943), pp. 115–133. ISSN: 1522-9602. DOI: 10.1007/BF02478259. URL: https://doi.org/10.1007/BF02478259.

[62] T. Mikolov, Q. V. Le, and I. Sutskever. "Exploiting Similarities among Languages for Machine Translation". In: *ArXiv e-prints* (Sept. 2013).

[63] T. Mikolov et al. "Distributed Representations of Words and Phrases and their Compositionality". In: *ArXiv e-prints* (Oct. 2013).

[64] T. Mikolov et al. "Efficient Estimation of Word Representations in Vector Space". In: *ArXiv e-prints* (Jan. 2013).

[65] Marvin Minsky and Seymour Papert. *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA, USA: MIT Press, 1969.

[66] Jeff Mitchell and Mirella Lapata. "Composition in Distributional Models of Semantics". In: *Cognitive Science* 34.8 (2010), pp. 1388–1429.

[67] S. M. Mohammad and F. Bravo-Marquez. "Emotion Intensities in Tweets". In: *ArXiv e-prints* (Aug. 2017).

[68] Andrew Nicholson, Juanita Whalen, and Penny Pexman. "Children's processing of emotion in ironic language". In: *Frontiers in Psychology* 4 (2013), p. 691. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2013.00691. URL: https://www.frontiersin.org/article/10.3389/fpsyg.2013.00691.

[69] Kamal Nigam and Matthew Hurst. "Towards a Robust Metric of Polarity". In: *Computing Attitude and Affect in Text: Theory and Applications*. Ed. by James G. Shanahan, Yan Qu, and Janyce Wiebe. Dordrecht: Springer Netherlands, 2006, pp. 265–279. ISBN: 978-1-4020-4102-0. DOI: 10.1007/1-4020-4102-0_20. URL: https://doi.org/10.1007/1-4020-4102-0_20.

[70] Canberk Özdemir and Sabine Bergler. "CLaC-SentiPipe: SemEval2015 Subtasks 10 B,E, and Task 11". In: *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. Denver, Colorado: Association for Computational Linguistics, June 2015, pp. 479–485. URL: http://www.aclweb.org/anthology/S15-2081.

[71] Sebastian Padó and Mirella Lapata. "Dependency-Based Construction of Semantic Space Models". In: *Comput. Linguist.* 33.2 (June 2007), pp. 161–199. ISSN: 0891-2017. DOI: 10.1162/coli.2007.33.2.161. URL: http://dx.doi.org/10.1162/coli.2007.33.2.161.

[72] Ted Pedersen, Siddharth Patwardhan, and Jason Michelizzi. "WordNet::Similarity: Measuring the Relatedness of Concepts". In: *Demonstration Papers at HLT-NAACL 2004*. HLT-NAACL–Demonstrations '04. Stroudsburg, PA, USA: Association for Computational Linguistics, 2004, pp. 38–41. URL: http://dl.acm.org/citation.cfm?id=1614025.1614037.

[73] James Pennebaker, Martha E. Francis, and Roger J. Booth. "Linguistic inquiry and word count (LIWC)". In: (1999).

[74] Jeffrey Pennington, Richard Socher, and Christopher D Manning. "Glove: Global Vectors for Word Representation." In: *EMNLP*. Vol. 14. 2014, pp. 1532–1543.

[75] Wim Peters and Yorick Wilks. "Data-Driven Detection of Figurative Language Use in Electronic Language Resources". In: *Metaphor and Symbol* 18.3 (July 2003), pp. 161–173. ISSN: 1092-6488. DOI: 10.1207/S15327868MS1803_03. URL: https://doi.org/10.1207/S15327868MS1803_03.

[76] M. F. Porter. "An algorithm for suffix stripping". In: *Program* 14.3 (1980), pp. 130–137. DOI: 10.1108/eb046814. URL: https://doi.org/10.1108/eb046814.

[77] Ashwin Rajadesingan, Reza Zafarani, and Huan Liu. "Sarcasm Detection on Twitter: A Behavioral Modeling Approach". In: *WSDM*. 2015.

[78] Harsh Rangwani, Devang Kulshreshtha, and Anil Kumar Singh. *NLPRL-IITBHU at SemEval-2018 Task 3: Combining Linguistic Features and Emoji pre-trained CNN for Irony Detection in Tweets*. Jan. 2018. DOI: 10.18653/v1/S18-1104.

[79] Kumar Ravi and Ravi Vadlamani. *A novel automatic satire and irony detection using ensembled feature selection and data mining*. Dec. 2016. DOI: 10.1016/j.knosys.2016.12.018.

[80] Antonio Reyes, Paolo Rosso, and Davide Buscaldi. "From humor recognition to irony detection: The figurative language of social media". In: *Data & Knowledge Engineering* 74 (2012), pp. 1–12. ISSN: 0169-023X. DOI: https://doi.org/10.1016/j.datak.2012.02.005. URL: http://www.sciencedirect.com/science/article/pii/S0169023X12000237.

[81] Antonio Reyes, Paolo Rosso, and Tony Veale. "A Multidimensional Approach for Detecting Irony in Twitter". In: *Lang. Resour. Eval.* 47.1 (Mar. 2013), pp. 239–268. ISSN: 1574-020X. DOI: 10.1007/s10579-012-9196-x. URL: http://dx.doi.org/10.1007/s10579-012-9196-x.

[82] Ellen Riloff et al. "Sarcasm as contrast between a positive sentiment and negative situation". English (US). In: *EMNLP 2013 - 2013 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*. Association for Computational Linguistics (ACL), 2013, pp. 704–714. ISBN: 978-1-937284-97-8.

[83] Omid Rohanian et al. "WLV at SemEval-2018 Task 3: Dissecting Tweets in Search of Irony". In: *SemEval@NAACL-HLT*. 2018.

[84] F. Rosenblatt. "The Perceptron: A Probabilistic Model for Information Storage and Organization in The Brain". In: *Psychological Review* (1958), pp. 65–386.

[85] Sara Rosenthal et al. *SemEval-2014 Task 9: Sentiment Analysis in Twitter*. Jan. 2014. DOI: 10.3115/v1/S14-2009.

[86] Sara Rosenthal et al. *SemEval-2014 Task 9: Sentiment Analysis in Twitter*. Jan. 2014. DOI: 10.3115/v1/S14-2009.

[87] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors". In: *Nature* 323 (Oct. 1986), p. 533. URL: http://dx.doi.org/10.1038/323533a0.

[88] Dan Sperber and Deirdre Wilson. "Irony and the use-mention distinction". In: (1981).

[89] Caroline Sporleder and Linlin Li. "Unsupervised Recognition of Literal and Non-literal Use of Idiomatic Expressions". In: *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*. EACL '09. Stroudsburg, PA, USA: Association for Computational Linguistics, 2009, pp. 754–762. URL: http://dl.acm.org/citation.cfm?id=1609067.1609151.

[90] Nitish Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *J. Mach. Learn. Res.* 15.1 (Jan. 2014), pp. 1929–1958. ISSN: 1532-4435. URL: http://dl.acm.org/citation.cfm?id=2627435.2670313.

[91] J. Staiano and M. Guerini. "DepecheMood: a Lexicon for Emotion Analysis from Crowd-Annotated News". In: *ArXiv e-prints* (May 2014).

[92] Carlo Strapparava and Alessandro Valitutti. "WordNet Affect: an Affective Extension of WordNet". In: *LREC*. 2004.

[93] Mike Thelwall et al. "Sentiment strength detection in short informal text". In: *Journal of the American Society for Information Science and Technology* 61.12 (), pp. 2544–2558. DOI: 10.1002/asi.21416. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21416.

[94] Oren Tsur, Dmitry Davidov, and Ari Rappoport. *ICWSM - A Great Catchy Name: Semi-Supervised Recognition of Sarcastic Sentences in Online Product Reviews*. Jan. 2010.

[95] A. M. TURING. "COMPUTING MACHINERY AND INLIGENCE". In: *Mind* LIX.236 (1950), pp. 433–460. DOI: 10.1093/mind/LIX.236.433. URL: http://dx.doi.org/10.1093/mind/LIX.236.433.

[96] Cynthia Van Hee, Els Lefever, and Veronique Hoste. "LT3: Sentiment Analysis of Figurative Tweets: piece of cake #NotReally". In: *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. Denver, Colorado: Association for Computational Linguistics, June 2015, pp. 684–688. URL: http://www.aclweb.org/anthology/S15-2115.

[97] Raymond W. Gibbs. *On the psycholinguistics of sarcasm*. Vol. 115. Mar. 1986. DOI: 10.1037/0096-3445.115.1.3.

[98] Byron C. Wallace, Do Kook Choe, and Eugene Charniak. "Sparse, contextually informed models for irony detection: Exploiting user communities, entities and sentiment". English (US). In: *ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference*. Vol. 1. Association for Computational Linguistics (ACL), 2015, pp. 1035–1044.

[99] Joseph Weizenbaum. "ELIZA Computer Program for the Study of Natural Language Communication Between Man and Machine". In: *Commun. ACM* 9.1 (Jan. 1966), pp. 36–45. ISSN: 0001-0782. DOI: 10.1145/365153.365168. URL: http://doi.acm.org/10.1145/365153.365168.

[100] Michael Wilson. "MRC psycholinguistic database: Machine-usable dictionary, version 2.00". In: *Behavior Research Methods, Instruments, & Computers* 20.1 (Jan. 1988), pp. 6–10. ISSN: 1532-5970. DOI: 10.3758/BF03202594. URL: https://doi.org/10.3758/BF03202594.

[101] Chuhan Wu et al. "THU_NGN at SemEval-2018 Task 3: Tweet Irony Detection with Densely connected LSTM and Multi-task Learning". In: *SemEval@NAACL-HLT*. 2018.

[102] Hongzhi Xu et al. "LLT-PolyU: Identifying Sentiment Intensity in Ironic Tweets". In: *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. Denver, Colorado: Association for Computational Linguistics, June 2015, pp. 673–678. URL: http://www.aclweb.org/anthology/S15-2113.

[103] M. D. Zeiler. "ADADELTA: An Adaptive Learning Rate Method". In: *ArXiv e-prints* (Dec. 2012).

Γλωσσάριο

Ελληνικός όρος	Αγγλικός όρος
Αναπαραστάσεις Λέξεων	Word Embeddings
Ανεκτικής Ταξινόμησης	Soft Classification
Δεδομένα Αξιολόγησης	Test Set
Δεδομένα Επικύρωσης	Validation Set
Δεδομένα Εκπαίδευσης	Train Set
Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης	Long Short-Term Memory -LSTM
Διατροπικής Εντροπία	Crossentropy
Ευρωστία	Robustness
κ-Πλησιέστεροι Γειτόνες	k-Nearest Neighbours
Μεταφορικά Γλωσσικά Φαινόμενα	Figurative Language
Συνδυασμός Μεθόδων	Ensemble