



Εθνικό Μετσόβιο Πολυτεχνείο
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΥΠΟΛΟΓΙΣΤΩΝ

Ανάλυση Συναισθήματος και Γνώμης

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

της

ΑΓΓΕΛΙΚΗΣ ΣΠΥΡΙΔΟΥΛΑΣ ΒΛΑΧΟΣΤΕΡΓΙΟΥ

Αθήνα, Νοέμβριος 2018



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
& ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΥΠΟΛΟΓΙΣΤΩΝ

ΑΝΑΛΥΣΗ ΣΥΝΑΙΣΘΗΜΑΤΟΣ ΚΑΙ ΓΝΩΜΗΣ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

της

ΑΓΓΕΛΙΚΗΣ ΣΠΥΡΙΔΟΥΛΑΣ ΒΛΑΧΟΣΤΕΡΓΙΟΥ

Συμβουλευτική Επιτροπή: Ανδρέας - Γεώργιος Σταφυλοπάτης
Γεώργιος Στάμου
Κωνσταντίνος Καρπούζης

Εγκρίθηκε από την επταμελή επιτροπή την 26^η Νοεμβρίου 2018.

...
Ανδρέας - Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

...
Κωνσταντίνος Καρπούζης Παναγιώτης Τσανάκας
Ερευνητής Α ΕΠΙΣΕΥ Καθηγητής Ε.Μ.Π.

...
Γεώργιος Καρυδάκης
Επ/ρος Καθ/τής Παν/μίου Αιγαίου

...
Γεώργιος Στάμου
Αν. Καθ/τής Ε.Μ.Π.

...
Φοίβος Μυλωνάς
Αν. Καθ/τής Ιονίου Παν/μίου

...
Ευάγγελος Σπύρου
Επ/ρος Καθ/τής Τ.Ε.Ι. Στερεάς Ελλάδας

Αθήνα, Νοέμβριος 2018

...

ΑΓΓΕΛΙΚΗ ΣΠΥΡΙΔΟΥΛΑ ΒΛΑΧΟΣΤΕΡΓΙΟΥ

© 2018 - Με επιφύλαξη παντός δικαιώματος - All rights reserved

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς το συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευτεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περιεχόμενα

1	Εισαγωγή	1
1.1	Κίνητρο και Ορισμός Προβλήματος	1
1.2	Ενσωμάτωση του Πλαισίου στην Αναγνώριση Συναισθημάτων	7
1.3	Μονοτροπικότητα και Πολυτροπικότητα	8
1.4	Σύνοψη	9
2	Εννοιολογικό πλαίσιο	11
2.1	Ορισμός του πλαισίου	11
2.2	Το Πλαίσιο στην Επικοινωνία Ανθρώπου-Μηχανής (HCI)	11
2.3	Μοντελοποίηση του Πλαισίου στη Συναισθηματική Υπολογιστική	12
2.4	Αναπαράσταση Συναισθήματος σύμφωνα με το Πλαίσιο	13
2.5	Εντοπισμός του Πλαισίου στα Υπάρχοντα Σώματα Κειμένων	14
2.6	Ενσωμάτωση του Πλαισίου στη Συναισθηματική Υπολογιστική	15
2.7	Προκλήσεις και Μελλοντικές Επεκτάσεις	15
3	Αναπαραστάσεις Συναισθήματος	17
3.1	Αποσαφήνιση Όρων	17
3.2	Αναπαράσταση Συναισθήματος στη Συναισθηματική Υπολογιστική	21
3.2.1	Αναπαράσταση σε Κατηγορίες	21
3.2.2	Διαστατικές Αναπαραστάσεις	22
3.2.3	Θεωρίες Γνωστικής Αξιολόγησης	24
3.3	Αναπαράσταση Συναισθήματος στον τομέα της Φυσικής Γλώσσας	27
4	Πολυτροπική Αναγνώριση Συναισθηματικών Καταστάσεων	31
4.1	Αναγνώριση Συναισθηματικών Καταστάσεων από σήματα Ομιλίας και Φυσιολογίας	31
4.1.1	Εισαγωγή	31
4.1.2	Σχετική Βιβλιογραφία	32
4.1.3	Πειραματική Μελέτη	33
4.1.4	Εξαγωγή χαρακτηριστικών Ομιλίας	34
4.1.5	Εξαγωγή χαρακτηριστικών Φυσιολογίας	35
4.1.6	Πειραματικά Αποτελέσματα	36

4.1.7	Συμπεράσματα	39
5	Μονοτροπική Αναγνώριση Συναισθηματικών Καταστάσεων	41
5.1	Αναγνώριση Συναισθηματικών Καταστάσεων από Λεκτικό Περιεχόμενο ..	41
5.1.1	Εισαγωγή	41
5.1.2	Σχετική Βιβλιογραφία	43
5.1.3	Εποπτική παρουσίαση της προτεινόμενης αρχιτεκτονικής της φυσικής εξέλιξης των σχετικών ερωτημάτων με το πλαίσιο	45
5.1.4	Ερευνητικά Αποτελέσματα	47
5.1.5	Προκλήσεις και Μελλοντικές Επεκτάσεις	50
6	Η συνεισφορά του πλαισίου στον τομέα της Φυσικής Γλώσσας	53
6.1	Εισαγωγή	53
6.2	Σχετική Βιβλιογραφία	55
6.3	Μέθοδοι Ενσωμάτωσης	56
6.3.1	Υπολογισμός Παραμέτρων	58
6.3.2	Συμπέρασμα	59
6.4	Εξαγόμενα χαρακτηριστικά: Σημασιολογική και Συντακτική Δομή των προτάσεων	60
6.4.1	Σημασιολογικά χαρακτηριστικά	60
6.4.2	Συντακτικά χαρακτηριστικά	61
6.5	Εξεταζόμενα Σύνολα Δεδομένων	64
6.6	Πειραματική Διαδικασία	65
6.6.1	Πρόβλημα ταξινόμησης και Πειραματικές ρυθμίσεις	65
6.7	Πειραματικά Αποτελέσματα	67
6.7.1	Θεωρείται αυτή η κατηγορία κοινωνικών μέσων από την άποψη του μεγέθους και του είδους της, κατάλληλη για την ενσωμάτωση του πλαισίου;	68
6.7.2	Είναι απαραίτητο να διευρύνουμε την ομάδα των επιλεγμένων χαρακτηριστικών για τη διευκόλυνση της ενσωμάτωσης του πλαισίου σε συστήματα SA;	68
6.7.3	Ποιός είναι ο συνδυασμός χαρακτηριστικών με τις καλύτερες επιδόσεις για τον εντοπισμό και την ενσωμάτωση πληροφοριών πλαισίου;	69
6.8	Συμπεράσματα	69
7	Εκμάθηση Αναπαραστάσεων της Φυσικής Γλώσσας με Ανταγωνιστικά Παραγωγικά Δίκτυα σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών	71
7.1	Εισαγωγή	71
7.2	Σχετική Βιβλιογραφία	74
7.2.1	Τεχνικές Μηχανικής Μάθησης για NLP διεργασίες	75

7.2.2	Μέθοδοι Βαθιών Νευρωνικών Δικτύων σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών	78
7.2.3	Ανταγωνιστικά Παραγωγικά Δίκτυα για διεργασίες Φυσικής Γλώσσας	80
7.3	Προτεινόμενη Προσέγγιση	82
7.3.1	Διατύπωση Προβλήματος	82
7.3.2	Ανταγωνιστική αρχιτεκτονική νευρωνικού δικτύου σε επίπεδο κειμένου	85
7.4	Πειραματική Επικύρωση	88
7.4.1	Σύνολα Δεδομένων	88
7.4.2	Συστήματα Αναφοράς	90
7.4.3	Υλοποίηση	90
7.4.4	Πειραματικά Αποτελέσματα	92
7.5	Συζήτηση	94
8	Σύνοψη - Συμπεράσματα	99
9	Κατάλογος δημοσιεύσεων	103
9.1	Περιοδικά	103
9.2	Κεφάλαια σε βιβλία	103
9.3	Συνέδρια	103
	Βιβλιογραφία	107

Κατάλογος Σχημάτων

3.1	Η κλίμακα SAM χρησιμοποιήθηκε για τη βαθμολόγηση των συναισθηματικών διαστάσεων του σθένους (άνω σειρά), την ενεργοποίηση (μεσαία γραμμή) και την κυριαρχία (κάτω σειρά) [19].	23
3.2	Παράδειγμα του προγράμματος επισημείωσης Feeltrace κατά τη διάρκεια μιας περιόδου παρακολούθησης. Ο οριζόντιος άξονας αποδίδει τη διάσταση σθένους και ο κάθετος άξονας τη διάσταση της ενεργοποίησης. Ο σχολιαστής μπορεί να μετακινήσει το δρομέα (υποδεικνύεται από έναν κύκλο) σε αυτήν τη διασύνδεση, για να αξιολογήσει μία συναισθηματική εκδήλωση που παρουσιάζεται παράλληλα σε ένα ξεχωριστό παράθυρο.	24
4.1	Οι συχνότητες (άξονας y) για δέκα (10) εκπαιδευόμενους σε συνθήκες IB και CT αντίστοιχα και οι κατανομές των απαντήσεών τους (άξονας x) σε δευτερόλεπτα αναφορικά με την διακοπή ομιλίας του εκπαιδευτή από τον εκπαιδευόμενο. Η κατακόρυφη διακεκομμένη μπλε γραμμή ορίζει το κατώφλι σύμφωνα με το οποίο διακρίνεται η απάντηση μικρότερης και μεγαλύτερης διάρκειας αντίστοιχα.	35
5.1	Επισκόπηση του προτεινόμενου συστήματος: (a) Εντοπισμός και εξαγωγή του σημασιολογικών εννοιών πλαισίου από το σώμα κειμένου SEMAINE. (b) Αναπαράσταση κάθε έννοιας με εκφράσεις πολλαπλών λέξεων ενισχυμένες με τιμές που αντιστοιχούν σε συναισθήματα. (c) Μία σειρά Ψυχολογικών Βάσεων (Psychological Foundations) εμπλουτίζεται. (d) Δείχνουμε ότι η προτεινόμενη προσέγγιση μπορεί να συσχετίσει όρους σημασιολογίας, γνωστικής και συναισθηματικής πληροφορίας.	44
5.2	Οι τέσσερις χαρακτήρες πρακτόρων: η Poppy: χαρούμενη και εξωστρεφής, η Prudence: λογική και συνετή, ο Spike: θυμωμένος και ο Obadiah: καταθλιπτικός και ζοφερός.	45
6.1	Γραφική αναπαράσταση μιας γραμμικής αλυσίδας ενός απλού μοντέλου CRF [121].	63

6.2	Ένα αντιπροσωπευτικό παράδειγμα μιας φράσης από το σύνολο δεδομένων Movies Review Dataset “the movie begins in the past where a young boy named sam attempts to save celebi from a hunter.” στη μορφή CoNLL μετά τη διαδικασία της προεπεξεργασίας. Η διαδικασία αυτή περιλαμβάνει τμηματοποίηση (tokenization) και διαχωρισμό των προτάσεων, επισημείωση των μερών του λόγου (part-of-speech (POS) tagging), λημματοποίηση, σύστημα αναγνώρισης ονομαστικής καταχώρησης (Name Entity Recognition (NER)), χρήση αναλυτή (parsing), και coreference resolution κάνοντας χρήση του Stanford CoreNLP επισημειωτή [139].	64
7.1	Η αρχιτεκτονική του μοντέλου GAN, το οποίο αποτελείται από ένα διευκρινιστή D και μία γεννήτρια G , κατά τη διάρκεια εκπαίδευσης του μοντέλου [241].	84
7.2	Η αρχιτεκτονική του προτεινόμενου δικτύου GAN για την εκμάθηση της αναπαράστασης σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών αντίστοιχα. Το δίκτυο αποτελείται από τα ακόλουθα τμήματα: η γεννήτρια G , ο DAE κωδικοποιητής Enc και ο αποκωδικοποιητής Dec αντίστοιχα, η διαδικασία της αλλοίωσης (corruption process) C and το δίκτυο του διευκρινιστή D [241].	87
7.3	Οι καμπύλες ανάκλησης-ακρίβειας του προτεινόμενου μοντέλου μας στις συλλογές δεδομένων 20Newsgroups, Movie Reviews και Finegrained.	95
7.4	t-SNE οπτικοποίηση για την αναπαράσταση σε επίπεδο κειμένου με τη χρήση του μοντέλου GAN.	96
7.5	t-SNE οπτικοποίηση για την αναπαράσταση σε επίπεδο πρότασης με τη χρήση του μοντέλου GAN.	96
7.6	t-SNE οπτικοποίηση για την αναπαράσταση σε επίπεδο χαρακτηριστικών με τη χρήση του μοντέλου GAN.	96

Κατάλογος Πινάκων

4.1	Επιλεχθέντα χαρακτηριστικά Φυσιολογίας των δέκα (10) εκπαιδευόμενων, πέντε (5) ατόμων της κάθε συνθήκης (IB και CT) αντίστοιχα κατά τη διάρκεια ενός επιχειρησιακού έργου αφόπλισης μίας προσομοιωμένης βόμβας. Τα χαρακτηριστικά επισημαίνονται ως εξής: καρδιακός ρυθμός (heart rate (hr)), αριστερός κοιλιακός (συσταλτικός) χρόνος (Left Ventral (systolic) Ejection Time (LVET)) και προσυσταλτική περίοδος (pre-ejection period (PEP)).	36
4.2	Οι βασικοί ταξινομητές (base learners) και το σύνολο μάθησης (ensemble learning) που χρησιμοποιήθηκαν για τους δέκα (10) εκπαιδευόμενους, πέντε (5) για κάθε κατάσταση (IB και CT) αντίστοιχα. Η καλύτερη μέθοδος για κάθε εκπαιδευόμενο υπογραμμίζεται με έντονο χρώμα.	38
5.1	Λεξιλόγιο που χρησιμοποιείται πιο συχνά από τους χειριστές και τους χρήστες αντίστοιχα στο σώμα κειμένου SEMAINE.	49
6.1	Τα κύρια χαρακτηριστικά μίας σειράς σωμάτων κειμένων. Ο Πίνακας συγκεντρώνει χαρακτηριστικά όπως το είδος των δεδομένων, το επίπεδο ανάλυσης, το μέγεθος του κειμένου, τον τύπο κατηγοριοποίησης και τα εξεταζόμενα χαρακτηριστικά. Η επισημείωση των χαρακτηριστικών είναι ως εξής: λεξιλόγιο (vocabulary (voc)): μονογράμματα (unigrams) και διγράμματα (bigrams), μέρη του λόγου (Part-Of-Speech (pos)), λέξεις με συναισθηματικό περιεχόμενο (sentiment words (sw)), μοτίβα συντακτικά (syntactic patterns (sp)), θέση των λέξεων (position (p)) και ανάλυση λέξεων (discourse (d)).[248]	55
6.2	Οι συλλογές κειμένων που χρησιμοποιήθηκαν για την κατηγοριοποίηση σε δύο ομάδες (two-class). Ο Πίνακας περιλαμβάνει επίσης τον αριθμό των μονογραμμάτων (unigrams) and διγραμμάτων (bigrams) μετά τη διαδικασία της προεπεξεργασίας.	65
6.3	Χαρακτηριστικά που χρησιμοποιήθηκαν για το πρόβλημα της επισημείωσης των ακολουθιών[248].	67

6.4	Ταξινόμηση των συλλογών δεδομένων Movie Review (MR) [163] και Finegrained Sentiment (FSD) [228] ως προς τα μέτρα αξιολόγησης: ακρίβεια (Precision (Pr.)), ανάκληση (Recall (R.)) και F_1 για κάθε λεξιλογική αναπαράσταση (π.χ. μονογράμματα (unigrams), ή μονογράμματα (unigrams) και διγράμματα (bigrams)). Η καλύτερη απόδοση σημειώνεται με έντονα γράμματα.[248]	70
7.1	Βασικά χαρακτηριστικά ορισμένων από τις πρόσφατα δημοσιευμένες ερευνητικές δραστηριότητες αναφορικά με τα Βαθιά Νευρωνικά Δίκτυα για τον τομέα της Ανάλυσης Γνώμης. Ο Πίνακας αναφέρεται στο επίπεδο ανάλυσης (Level of Analysis), στον τύπο του νευρωνικού Δικτύου που εφαρμόσαμε (Model), τα σύνολα δεδομένων που χρησιμοποιήσαμε (Datasets), το είδος της εφαρμογής (Task) και τις μετρικές αξιολόγησης (Evaluation Metrics). Οι μετρικές αξιολόγησης επισημειώνονται ως εξής: Accuracy (Acc.), Macro-F1 measure, Root Mean Square Error (RMSE), Mean Absolute Error (MAE) και Pearson correlation coefficient (r)[241].	77
7.2	Συλλογές δεδομένων δοκιμών για τη διερεύνηση της χρήσης των Ανταγωνιστικά Παραγωγικών Δικτύων αναφορικά με την εκμάθηση καταναμημένων αναπαραστάσεων της Φυσικής Γλώσσας σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών αντίστοιχα. [241]	90
7.3	Η απόδοση του προτεινόμενου μοντέλου GAN ως ένα Ανταγωνιστικά Παραγωγικό μοντέλο όταν χρησιμοποιούμε διαφορετικά είδη δεδομένων σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών. Το μοντέλο μας συγκρίνεται με ένα απλό GAN το οποίο χρησιμεύει ως σύστημα αναφοράς [241].	94

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να εκφράσω τις ειλικρινείς ευχαριστίες μου σε όλους όσους με βοήθησαν κατά τη διάρκεια της διδακτορικής μου διατριβής, τα τελευταία έξι χρόνια. Καταρχάς θα ήθελα να ευχαριστήσω τον επιβλέποντα Καθηγητή, κ. Στέφανο Κόλλια, για την εμπιστοσύνη που μου έδειξε δίνοντάς μου την ευκαιρία να ασχοληθώ εκτενέστερα με το κομμάτι της Συναισθηματικής Υπολογιστικής. Τον ευχαριστώ θερμά για την πολύτιμη βοήθεια και υποστήριξή του, μα πάνω απ' όλα για την ελευθερία που μου διέθεσε κατά την πορεία της διατριβής. Επίσης, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή, κ. Σταφυλοπάτη που ανέλαβε την επίβλεψη του τελευταίου τμήματος της διατριβής μου από το Μάρτιο του 2018 έως σήμερα. Τον ευχαριστώ ιδιαίτερος για τη στήριξη και την εξαιρετική συνεργασία που είχαμε. Ευχαριστίες οφείλω και στα άλλα δύο μέλη της συμβουλευτικής επιτροπής της διδακτορικής μου διατριβής, στον Αναπληρωτή Καθηγητή Γεώργιο Στάμου και στον ερευνητή Α' του Ε.Π.Ι.Σ.Ε.Υ. Κωνσταντίνο Καρπούζη.

Αξιομνημόνευτη είναι η συνεργασία μου και ερευνητική μου συμβίωση με όλα τα μέλη του Εργαστηρίου Ψηφιακής Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων και του Εργαστηρίου Ευφυών Συστημάτων και κυρίως με τους Δρ. Αμαρυλλίς Ραουζαίου, Γεώργιο Τσατήρη, Γεώργιο Μαρανδιανό, Νικόλαο Παπανελόπουλο, Μιχάλη Γιαζιτζόγλου, Enrique Alfonso, Γεώργιο Στρατογιάννη, Ελένη Βάθη και Αθανάσιο Τάγαρη.

Οφείλω να εκφράσω τις θερμές ευχαριστίες μου σε όλους τους φίλους μου εκτός Ε.Μ.Π. για τη συνεχή ενθάρρυνση που μου παρείχαν στη διάρκεια της ερευνητικής μου δραστηριότητας καθώς και για την υπομονή που επέδειξαν όλα αυτά τα χρόνια. Τέλος, είμαι ευγνώμων προς την οικογένειά μου για την ανεκτίμητη συμπαράστασή της σε όλη τη διάρκεια των σπουδών μου. Την ευχαριστώ γιατί με την αγάπη της, ήταν αυτή που στήριξε από μικρή τα όνειρά μου και φρόντιζε να έχω ΠΑΝΤΑ τις απαραίτητες προϋποθέσεις για να τα πραγματοποιώ.

ΠΕΡΙΛΗΨΗ

Τα τελευταία χρόνια έχει παρατηρηθεί μία αύξηση του αριθμού των προσπαθειών για την αυτόματη αναγνώριση και κατηγοριοποίηση του ανθρωπίνου συναισθήματος χρησιμοποιώντας σήματα φυσιολογίας, σήματα από το πρόσωπο, τη φωνή, καθώς επίσης και προσωπικές ερμηνείες από κείμενα μεγάλων κοινωνικών δεδομένων. Αρκετοί είναι οι τομείς της έρευνας που θα μπορούσαν να επωφεληθούν από αυτά τα συστήματα: διαδραστικά συστήματα διδασκαλίας, τα οποία να επιτρέπουν στους εκπαιδευτικούς να γνωρίζουν το άγχος των φοιτητών, πρόληψη των ατυχημάτων (π.χ. εντοπισμός της κόπωσης του οδηγού), στρατιωτικά ομαδικά καθήκοντα που χαρακτηρίζονται από μεγάλης διάρκειας περιόδους άγχους και πίεσης και εφαρμογές στον τομέα της Υγείας για την έγκαιρη διάγνωση νευροεκφυλιστικών νόσων (π.χ. νόσος του Πάρκινσον), όπου η εκδήλωση των συμπτωμάτων συμβαίνει πολλά χρόνια μετά την έναρξη του νευροεκφυλισμού.

Ωστόσο, παρά τις μέχρι τώρα ερευνητικές προσπάθειες, δεν έχει επιτευχθεί ο μακροπρόθεσμος στόχος της δημιουργίας ενός ισχυρού πλαισίου αναγνώρισης του εξεταζόμενου τομέα έρευνας που να βασίζεται στην ανάλυση και στην ερμηνεία του. Δεν υπάρχει καμία αμφιβολία ότι η “δημιουργία του συναισθήματος” (affect production) επηρεάζεται από το εκάστοτε πλαίσιο που λαμβάνει χώρα τη δεδομένη στιγμή, όπως το έργο στο οποίο υποβάλλεται ο χρήστης, τα άτομα που αλληλεπιδρούν με το χρήστη, η ταυτότητα αλλά και η εκφραστικότητα τους. Η οποιαδήποτε λοιπόν συμπληρωματική μορφή πληροφορίας πλαισίου αναφορικά με τον εξεταζόμενο τομέα έρευνας μας βοηθά ώστε να απαντήσουμε στο ερώτημα: τί είναι πιθανότερο να συμβεί, εκτρέποντας έτσι τον ταξινομητή από τις πιθανότερες/σχετικές κατηγορίες. Χωρίς το πλαίσιο, ακόμη και οι άνθρωποι μπορεί να παρερμηνεύουν τις παρατηρούμενες εκφράσεις του. Έτσι, με την αντιμετώπιση των προκλήσεων υπό το πρίσμα της αναγνώρισης του συναισθήματος υπό συγκεκριμένο πλαίσιο (context-aware affect analysis), δηλαδή με την καλύτερη μελέτη των πληροφοριών πλαισίου, με την ερμηνεία του σε συγκεκριμένους τομείς εφαρμογών, την αναπαράστασή του, τη μοντελοποίησή του, μπορούμε να προσεγγίσουμε καλύτερα την αναγνώριση του συναισθήματος σε πραγματικό χρόνο.

Αντίστοιχα, στον τομέα των προσωπικών ερμηνειών από το κείμενο (Sentiment Analysis) αλλά και γενικότερα στον τομέα της Φυσικής Γλώσσας (Natural Language Processing (NLP)) η συνεισφορά του πλαισίου έγκειται στην καλύτερη αναγνώριση, ερμηνεία και επεξεργασία των απόψεων (opinions) και συναισθημάτων (sentiments) σε κείμενα, τα οποία εξετάζονται σε επίπεδο κειμένου (document-level), προτάσεων (sentence-level) και χαρακτηριστικών (aspect-level) αντίστοιχα. Στην περίπτωση αυτή, λαμβάνονται υπόψιν η σημασιολογία, οι γνωστικές και οι συναισθηματικές πληροφορίες των υποκειμενικών απαντήσεων των ατόμων. Ειδικότερα, στον τομέα αυτό, η συνεισφορά μας έγκειται στην εκπαίδευση ισχυρών αναπαραστάσεων χαρακτηριστικών από μη επισημειωμένα δεδομένα με τη χρήση Νευρωνικών Δικτύων και συγκεκριμένα με τη χρήση των Ανταγωνιστικά Παραγωγικών Μοντέλων (GANs), η χρήση των οποίων έχει επιδείξει εντυπωσιακά

αποτελέσματα στον τομέα της Όρασης Υπολογιστών. Η πρωτοτυπία της συγκεκριμένης μεθόδου έγκειται στον τρόπο υλοποίησης του μοντέλου, στην επιλογή των υπερπαραμέτρων, στη χρήση μη επιβλεπόμενης μάθησης και στην πειραματική επικύρωση του προτεινόμενου μοντέλου σε σώματα κειμένου που προέρχονται από διαφορετικές πηγές αναφορικά με το είδος τους και την έκτασή τους.

ABSTRACT

Recent years have witnessed a number of important efforts aimed at technologies for automatic detection and analysis of human emotion using physiological, facial and voice signals, as well as personal interpretations from social media texts. There are a number of research areas that could benefit from such systems: interactive teaching systems which allow teachers to identify student stress, accident prevention (e.g. driver's fatigue detection), military tasks which are characterized by long stress and pressure periods and health care applications for timely diagnosis of neurodegenerative diseases (e.g., Parkinson's disease), where the manifestation of symptoms occurs many years after the onset of neurodegeneration.

However, despite the research efforts so far, the long-term objective of **creating a strong framework** concerning the examined research field that should be based on its analysis and interpretation has not been achieved so far. There is no doubt that "affect production" is affected from the context that is taking place at the moment, such as the resources that are nearby, What (the human is doing), the situation that is occurring, the identities of the people around the user as well as the degree of expressiveness. Any other complementary contextual information examined assists us to respond to the question: what is more likely to happen, thus diverting the classifier from more likely/relevant categories. Without the framework, even people can misinterpret their observable expressions. Thus, by addressing the above described challenges under the context of analyzing emotion in a specific context (context-aware affect analysis), i.e. studying the contextual parameters of the framework through its interpretation in specific areas of application, its representation, its modeling, we may better approach context analysis in real time.

In the **field of personal interpretations from the text (Sentiment Analysis)** and more generally in the field of Natural Language Processing (NLP), the contribution of the framework lies in a more effective analysis, interpretation and processing of opinions and sentiments in text which are examined at document, sentence and aspect level respectively. In this case, the semantic, cognitive and emotional information of the subjective responses of individuals are taken into account. To be more precise, in this area, our contribution is based on the idea of training strong feature representations from unlabelled data using Neural Networks and particularly using the Generative Adversarial Networks (GANs), which have proved very popular in the Computer Vision domain. The originality of our proposed method lies in the model's implementation, in the hyper-parameters' choice, in

the use of unsupervised learning and finally in its experimental validation on text corpora which come from different sources depending on their genre and length.

ΕΚΤΕΤΑΜΕΝΗ ΠΕΡΙΛΗΨΗ

Τα τελευταία 20 χρόνια έχει πραγματοποιηθεί μία σειρά σημαντικών προσπαθειών που στοχεύουν σε τεχνολογίες αναφορικά με τη μοντελοποίηση, την ανάλυση και τη σύνθεση αλληλεπιδράσεων ανθρώπου-ανθρώπου και ανθρώπου-υπολογιστή (HCI). Δεδομένου ότι η ανθρώπινη ύπαρξη και η ύπαρξη ηλεκτρονικών υπολογιστών έχουν γίνει εξαιρετικά αλληλένδετες, οι πρόσφατες εξελίξεις σε αυτόν τον τομέα έρευνας έχουν ήδη μετατοπιστεί από τον τομέα του HCI στον τομέα του “έξυπνου” HCI (iHCI), μεταφέροντάς μας από το παραδοσιακό πληκτρολόγιο και το ποντίκι σε φυσικές λειτουργίες αλληλεπίδρασης όπως η κατανόηση μιας σειράς ανθρώπινων συμπεριφορών μέσω συναισθηματικών και κοινωνικών σημάτων. Με άλλα λόγια, προσεγγίζοντας το HCI αναφορικά με τις ρυθμίσεις HCI που βασίζονται στην αφή, τη χειρονομία και την κίνηση (π.χ. Microsoft Kinect), έχουμε τη δυνατότητα να ανιχνεύουμε τις διαφορές και τις αλλαγές στην επικοινωνιακή συμπεριφορά του ανθρώπου και έτσι να ξεκινούμε αλληλεπιδράσεις σύμφωνα με αυτές τις πληροφορίες. Ως εκ τούτου, το αποτέλεσμα είναι μία όλο και αυξανόμενη τεχνολογία αιχμής σε τομείς όπως η Συναισθηματική Υπολογιστική (AC), η κοινωνική επεξεργασία σήματος, η HCI [58], η ψυχική υγεία, οι τεχνολογίες εκμάθησης κ.λ.π. Ωστόσο, έχει δοθεί ελάχιστη προσοχή στις πιθανές επιπτώσεις που αυτές οι τεχνολογίες μπορούν να επιφέρουν όταν ενσωματωθεί η έννοια του πλαισίου στα συστήματα ανάλυσης της συναισθηματικής συμπεριφοράς των ανθρώπων.

Η ενσωμάτωση του πλαισίου στην ανάλυση της συναισθηματικής ανθρώπινης συμπεριφοράς βρίσκεται στη διασταύρωση των συναισθηματικών υπολογιστικών συστημάτων και των συναισθηματικά συνειδητών συστημάτων αλληλεπίδρασης των ηλεκτρονικών υπολογιστών, δεδομένου ότι οι πληροφορίες περιβάλλοντος/πλαισίου δεν μπορούν να αποφευχθούν στην αυτόματη ανάλυση της συναισθηματικής συμπεριφοράς του ανθρώπου. Οι συνεισφορές των συναισθηματικά υπολογιστικών συστημάτων που έχουν επίγνωση του πλαισίου επιβεβαιώθηκαν κατά τη διάρκεια των δύο πρόσφατα οργανωμένων συνεδρίων σχετικά με την Αναγνώριση Συναισθήματος όταν λαμβάνουμε υπόψη το πλαίσιο (CBAR 2012 και CBAR 2013 [94]), που διεξήχθησαν σε συνεργασία με το SocialCom2012 και το ACII 2013 αντίστοιχα. Πιο συγκεκριμένα, το συνέδριο CBAR 2013 ήταν ένα από τα συντομότερα workshops του ACII 2013 με τις πιο ενδιαφέρουσες ερευνητικές φιγούρες (Schuller, Gratch), οι οποίες είναι ενδεικτικές της αναδυόμενης αυτής ερευνητικής περιοχής. Αυτές οι εξέχουσες προσωπικότητες της AC ερευνητικής κοινότητας παρείχαν μια πολύτιμη πηγή πληροφόρησης σχετικά με το πλαίσιο μάθησης στην

αναγνώριση των επιπτώσεων και πώς διαμορφώνεται η έννοια του πλαισίου και πώς οι άνθρωποι ερμηνεύουν τις εκφράσεις των ανθρώπων και των μηχανών. Έτσι, με την αντιμετώπιση των προβλημάτων που σχετίζονται με την ανάλυση των επιπτώσεων, δηλαδή την προσεκτική μελέτη των σχετικών πληροφοριών και της συνάφειάς τους με τα συστήματα που έχουν επίγνωση του περιβάλλοντος/πλαισίου, της εκπροσώπησής τους, της μοντελοποίησης και της ενσωμάτωσής τους, συμπεριλαμβανομένης της επίπτωσής τους στην απόδοση των υπάρχουσών μεθόδων ανάλυσης επιπτώσεων, ο “ερευνητικός χάρτης” που διαμορφώνεται δείχνει τα πιο σημαντικά βήματα που πρέπει να υλοποιηθούν ακόμη για την ανάλυση των επιπτώσεων αυτών στον πραγματικό κόσμο.

Κατά την προσπάθεια εντοπισμού ενός τυπικού ορισμού για συστήματα ανάλυσης του συναισθήματος (affect analysis systems) και εφαρμογές επίγνωσης πλαισίου (context-aware), ένα σημείο εκκίνησης θα ήταν η διερεύνηση του τρόπου με τον οποίο έχει οριστεί μέχρι στιγμής ο όρος “πλαίσιο”. Ο Schilit και οι συνεργάτες του [207] εισήγαγαν τον όρο context awareness στην πληροφορική, δίνοντας έμφαση στις σημαντικές πτυχές του πλαισίου: με ποιόν είσαι μαζί, πού βρίσκεσαι, ποιοί πόροι βρίσκονται σε κοντινή απόσταση. Έτσι, οι εφαρμογές επίγνωσης πλαισίου εξετάζουν το ποιός, πού, πότε και τί (ο χρήστης τί κάνει) και χρησιμοποιούν αυτές τις πληροφορίες για να προσδιορίσουν για ποιό λόγο συμβαίνει μία κατάσταση. Σύμφωνα με ένα παρόμοιο ορισμό, ο Brown και οι συνεργάτες του [24] ορίζουν το πλαίσιο ως το πού βρίσκονται οι άνθρωποι, τις ταυτότητές των ανθρώπων που περιβάλλουν το χρήστη, την ώρα της ημέρας, την εποχή, τη θερμοκρασία, κλπ. Άλλες προσεγγίσεις [194] αναφέρονται στο πλαίσιο ως η τοποθεσία του χρήστη, το περιβάλλον, η ταυτότητα και ο χρόνος, ενώ άλλοι χρησιμοποιούν συνώνυμα του όρου “πλαίσιο” αναφερόμενοι στο πλαίσιο είτε ως το περιβάλλον [24] είτε ως την κατάσταση που λαμβάνει χώρα [73].

Με βάση την ευρύτερη προσέγγιση του Abowd [2], το πλαίσιο μπορεί να οριστεί ως ένας συνδυασμός των τεσσάρων τύπων πλαισίου (contextual types): ταυτότητα, ώρα, τοποθεσία και δραστηριότητα. Αυτό το μοντέλο πλαισίου εμπλουτίστηκε αργότερα [271] προκειμένου να συμπεριληφθούν πρόσθετοι τύποι πλαισίου που ονομάζονται “σχέσεις”. Στην ανθρώπινη επικοινωνία, η βιβλιογραφία υποδεικνύει ότι οι άνθρωποι αξιολογούν τις καταστάσεις με βάση πληροφορίες που προσδιορίζουν το πλαίσιο, όπως οπτικές πληροφορίες παρελθόντος χρόνου [62], τη γενικότερη κατανόηση της κατάστασης [30], τις λεκτικές πληροφορίες παρελθόντος χρόνου [119], το πολιτιστικό υπόβαθρο [140], το φύλο των συμμετεχόντων, τις γνώσεις σχετικά με τη γενική ρύθμιση της αλληλεπίδρασης κατά την οποία ένα συναισθηματικό φαινόμενο λαμβάνει χώρα [24] αλλά και την κοινωνική κατάσταση [16].

Χωρίς την ύπαρξη του πλαισίου, ακόμη και οι άνθρωποι μπορούν να παρερμηνεύσουν τα παρατηρηθέντα συναισθηματικά συνθήματα όπως αυτά του προσώπου, της φωνής καθώς και των χειρονομιών αντίστοιχα. Όσον αφορά στον πραγματικό κόσμο, για εφαρμογές πλαισίου επίγνωσης (context-aware) στη Συναισθηματική Υπολογιστική, το πλαίσιο ορίζεται ως κάθε πληροφορία που μπορεί να χρησιμοποιηθεί για το χαρακτηρισμό της

κατάστασης αναφορικά με την αλληλεπίδραση μεταξύ των χρηστών και του συστήματος. Έτσι, ο ορισμός που προτείνεται στην εργασία [207] προσεγγίζει καλύτερα την κατανόηση των ανθρωπίνων σημάτων. Ένας ακόμα καταλληλότερος ορισμός [60], συνοψίζει τις βασικές πτυχές του πλαισίου σε σχέση με την ανθρώπινη επικοινωνιακή συμπεριφορά από την άποψη της λεγόμενης γενικευμένης **W5+** διαμόρφωσης: ποιός (**Who**) εμπλέκεται (π.χ. δυαδική/τριαδική αλληλεπίδραση μεταξύ των ατόμων ή αλληλεπίδραση μεταξύ ενός ατόμου και ενός εικονικού χαρακτήρα), τί (**What**) ανακοινώνεται (π.χ. γλωσσικό μήνυμα, μη-γλωσσικό σήμα επικοινωνιακής μορφής, το συναίσθημα), πώς (**How**) μεταφέρονται οι πληροφορίες (έκφραση του προσώπου του ατόμου, η κίνηση του κεφαλιού, ο τόνος της φωνής, οι χειρονομίες των χεριών είτε του σώματος), γιατί (**Why**), δηλαδή, σε ποιά (**Which**) πλαίσιο οι πληροφορίες μεταφέρονται, πού (**Where**) είναι ο χρήστης, ποιά (**What**) είναι η τρέχουσα απασχόλησή του και πώς (**How**) αυτός/αυτή αισθάνεται καθώς και ποιές δράσεις πρέπει να ληφθούν για την ικανοποίηση των αναγκών και των απαιτήσεων των χρηστών.

Δυστυχώς, μέχρι στιγμής οι προσπάθειες για την κατανόηση της ανθρώπινης συμπεριφοράς (human behavior understanding) είναι συνήθως ανεξάρτητες του πλαισίου (context-insensitive) [169], ενώ η προτεινόμενη μεθοδολογία αντιμετωπίζει μία ή περισσότερες **W5+** ερωτήσεις ξεχωριστά κάθε φορά, όπως το πού, τί [9], πότε, ποιός [229], γιατί και πώς [89] (**W5+**). Συνολικά, απαιτείται περαιτέρω έρευνα για την ταυτόχρονη απάντηση των **W5+** ερωτημάτων αλλά και για τη δημιουργία ενός ενοποιημένου πλαισίου διαμόρφωσης.

Στη βιβλιογραφία αναφορικά με την ανάλυση του συναισθήματος, σχετικά λίγες ερευνητικές προσπάθειες κάνουν χρήση των πληροφοριών πλαισίου και σε γενικές γραμμές χρησιμοποιούνται διαφορετικές προσεγγίσεις για τη διαμόρφωση του πλαισίου. Στο [125], ένα μονοτροπικό πλαίσιο για τη βραχυπρόθεσμη μοντελοποίηση του πλαισίου σε δυαδική αλληλεπίδραση προτάθηκε, όπου το πλαίσιο ορίζεται ως το σύνολο των συνθημάτων ομιλίας του ομιλητή από το παρελθόν, ενώ σύμφωνα με το [86], στα πολυτροπικά πλαίσια αναγνώρισης συναισθήματος η έννοια του πλαισίου ορίζεται ως το σύνολο των οπτικοακουστικών συναισθηματικών εκφράσεων που προηγούνται και που ακολουθούν, χρησιμοποιώντας αμφίδρομη μεγάλη και μικρή διάρκειας Νευρωνικά Δίκτυα (BLSTM). Συνολικά, από την άποψη των προσεγγίσεων της αναγνώρισης των συναισθημάτων, οι εξαρτώμενες από το πλαίσιο αμφίδρομη κατεύθυνσης προσεγγίσεις έχει αποδειχτεί ότι μπορούν να ξεπεράσουν τις μεθόδους που δεν εξετάζουν το πλαίσιο όπως τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNNs) και τα κρυφά μαρκοβιανά μοντέλα (HMM). Ωστόσο, τα οριστικά συμπεράσματα εξαρτώνται από τον σχεδιασμό των σωμάτων κειμένου και ενδεχομένως να μην καλύπτουν όλο το φάσμα της ανθρώπινης συναισθηματικά ενισχυμένης επικοινωνίας από την άποψη της μοντελοποίησης της αλληλεπίδρασης. Επιπλέον, όσον αφορά στην ποσοτικοποίηση της επίδρασης του πλαισίου η Kaliouby και οι συνεργάτες της [62] περιγράφουν την έκφραση του συναισθήματος ως μία ακολουθία των ταυτόχρονων διαδοχικών δράσεων του προσώπου, είτε ως μικροεκφρά-

σεις, όπου εδώ το πλαίσιο ορίζεται ως η σχέση μεταξύ των διαδοχικών μικροεκφράσεων. Τέλος, στο [93] περιγράφεται ένα σύστημα επίγνωσης πλαισίου που σχετίζεται με την αναγνώριση του πόνου, όπου οι τύποι πλαισίου (contextual types) ορίζονται ως ο τόπος, η εργασία, κλπ., ενώ για εφαρμογές όπως οι προσωπικές συλλογές φωτογραφιών υποβοηθούμενης επισημείωσης [128] οι συναφείς πληροφορίες παρέχονται μέσω των σχέσεων μεταξύ των ανθρώπων, των γεγονότων και των τοποθεσιών. Και στις δύο εφαρμογές, που βασίζονται σε Transferable Belief Model (TBM) μοντέλα και σε πιθανοτικά μοντέλα αντίστοιχα, σημαντικές βελτιώσεις στην απόδοση έχουν παρατηρηθεί.

Η πρώτη ενότητα της διατριβής ολοκληρώνεται με τη διαπίστωση ότι η ενσωμάτωση του πλαισίου στη συναισθηματική υπολογιστική βρίσκεται ακόμη σε αρχικό στάδιο. Παρά το αναπτυσσόμενο ενδιαφέρον για τα ενδιάμεσα στάδια της επικοινωνίας ανθρώπου-μηχανής (HCI), δεν έχει ακόμη εξεταστεί η καταλληλότητά τους ως προς την ενσωμάτωση του πλαισίου στα στάδια της εξαγωγής (elicitation), της ερμηνείας (interpretation) καθώς και της αναγνώρισης (recognition) συναισθήματος. Τα στάδια αυτά περιλαμβάνουν μία σειρά προκλήσεων όπως η μοντελοποίηση των γνωστικών θεωριών του πλαισίου με βάση την συναισθηματική αλληλεπίδραση, η άντληση πληροφοριών από το πλαίσιο, η ενσωμάτωση των πληροφοριών πλαισίου στα σώματα κειμένου ενσυναίσθησης, ο εντοπισμός των δομών πλαισίου (contextual types) καθώς και η ενσωμάτωσή τους για τη βελτίωση των επιδόσεων των πολυτροπικών εφαρμογών. Κατά συνέπεια, τα προαναφερθέντα ζητήματα είναι στενά συσχετιζόμενα με αυτές τις καταστάσεις και ως εκ τούτου είναι επιβεβλημένος ο συνυπολογισμός τους στην διαδικασία εξαγωγής συναισθηματικών παραμέτρων και αναγνώρισης συναισθήματος. Για τους λόγους αυτούς όπως αναφέρθηκε και προηγουμένως απαιτείται περαιτέρω έρευνα για την απάντηση των ερωτημάτων της **W5+** διαμόρφωσης του πλαισίου. Κάτω από αυτές τις συνθήκες, θα μπορούσαμε να έχουμε μία πρώτη εικόνα για το κατά πόσο είναι αναγκαίο η επισημείωση των κειμένων να πραγματοποιείται σύμφωνα με το εξαρτώμενο πλαίσιο καθώς και κατά πόσο τα υπάρχοντα σώματα κειμένων πρέπει να επαναεπισημειωθούν είτε να ηχογραφηθούν από την αρχή.

Αναφορικά με τη διαδικασία επισημείωσης, οι προκλήσεις της επίτευξης αυτού του στόχου ξεκινούν από την εφαρμογή ενός γενικού πλαισίου για τη συλλογή δεδομένων αυθόρμητης συμπεριφοράς. Το ερευνητικό πεδίο είναι ακόμη στα αρχικά του στάδια και δεν έχουν παρατηρηθεί σημαντικές προσπάθειες για τη συλλογή αυθόρμητων δεδομένων εξαρτώμενα από το πλαίσιο ειδικά σχεδιασμένα για την ανάλυση τύπων πλαισίου (contextual types). Οι περισσότερες ερευνητικές δουλειές μέχρι στιγμής βασίζονται στη χρήση σωμάτων κειμένου που είναι ήδη επισημειωμένα, είτε επισημειώνονται κάθε φορά για την ικανοποίηση των αναγκών του εκάστοτε πειράματος. Παράλληλα, κατά τη διάρκεια αλληλεπίδρασης μίας ομάδας περιλαμβάνεται μια πληθώρα από διαφορετικά “συνθήματα” και μέχρι στιγμής δεν υπάρχει ένα πρότυπο σχολιασμού ή ένα πρωτόκολλο συλλογής δεδομένων που να είναι εύκολα εφαρμόσιμο. Επιπλέον, ο συνδυασμός της προοπτικής της αξιολόγησης γεγονότων θα μπορούσε να παρέχει μία σειρά από οφέλη, όπως η βελτιωμένη αναγνώριση των μικτών συναισθημάτων, η ικανότητα ενσωμάτωσης συ-

ναφών πληροφοριών σε αυτόματες αρχιτεκτονικές συναισθηματικής ανάλυσης είτε πολυτροπικές είτε μονοτροπικές καθώς και ένα καλύτερο επίπεδο ερμηνείας του συναισθήματος. Ως εκ τούτου, χωρίς σημαντική προσπάθεια για τη συλλογή και την επισημείωση δεδομένων θα είναι δύσκολο αν όχι αδύνατο να αντιμετωπιστούν τα τρέχοντα προβλήματα της έρευνας στον τομέα αυτό.

Πιο συγκεκριμένα, στον τομέα των αυτόματων πολυτροπικών αρχιτεκτονικών συναισθηματικής ανάλυσης, μελετάμε την κατανόηση της σχέσης μεταξύ της εξωτερικής συμπεριφοράς και της εσωτερικής συναισθηματικής κατάστασης των δύο μελών της ομάδας (“εκπαιδευτή” “εκπαιδευόμενου”) κατά τη διάρκεια ενός απαιτητικού επιχειρησιακού έργου (δηλαδή, την απενεργοποίηση μιας βόμβας). Σε αυτή τη μελέτη αξιολογήσαμε τις λεκτικές απαντήσεις του μέλους της ομάδας (δηλαδή τη διάρκεια της απάντησης) σε σχέση με τις εξωτερικές και τις εσωτερικές συναισθηματικές τους ενδείξεις. Τα εξωτερικά συναισθήματα συμπεριφοράς περιλαμβάνουν τις λεκτικές εκφράσεις του εκπαιδευόμενου, ενώ οι εσωτερικές ενδείξεις βασίζονται σε φυσιολογικά σήματα. Ειδικότερα, διαφοροποιούμε τα φυσιολογικά πρότυπα του εκπαιδευόμενου που συμβαίνουν μετά την έναρξη ομιλίας του εκπαιδευτή ανάλογα με το εάν ανήκουν σε ένα σύντομο ή ένα μεγαλύτερο χρονικό διάστημα ανταπόκρισης. Με βάση την παραδοχή ότι οι μακροπρόθεσμες συμπεριφορές είναι πιθανό να προκληθούν από απαιτητικά γεγονότα γνωστικών εργασιών ή/και αγχωτικές αλληλεπιδράσεις, υποθέτουμε ότι οι εσωτερικοί μηχανισμοί που παράγονται σε αυτά τα διαστήματα έντονης συναισθηματικής δραστηριότητας θα αντικατοπτρίζονται στη φυσιολογία του εκπαιδευόμενου. Ένα σώμα κειμένου μίας ομάδας δύο ατόμων χρησιμοποιήθηκε για να εξετάσουμε τη συσχέτιση μεταξύ των φυσιολογικών σημάτων των “εκπαιδευομένων”, ακολουθώντας τις ερωτήσεις του εκπαιδευτή, για να προβλέψουμε εάν συνέβησαν σε σύντομο ή μεγαλύτερης διάρκειας χρονικό διάστημα. Τα αποτελέσματα υποδεικνύουν ότι υπάρχει συσχέτιση μεταξύ της ανάληψης και της εσωτερικής συναισθηματικής κατάστασης. Επιπρόσθετα, στόχος μας ήταν η περαιτέρω μελέτη αυτής της σύνδεσης δημιουργώντας μία σειρά σύνθετων ταξινομητών. Ως εκ τούτου, μελετήσαμε διάφορους βασικούς ταξινομητές και διαφορετικά μεγέθη του συνόλου σύνθετων ταξινομητών για να προσδιορίσουμε την καλύτερη προσέγγιση για την οικοδόμηση ενός σταθερού σύνθετου συνόλου που γενικεύει καλά τόσο στις εξωτερικές όσο και στις εσωτερικές ενδείξεις των ατόμων.

Η έλλειψη της συναισθηματικής εκφραστικότητας είναι ένα από τα κύρια χαρακτηριστικά περιόδων άγχους όταν τα μέλη μίας ομάδας καλούνται να εκτελέσουν καθήκοντα συνεργασίας βασιζόμενοι κυρίως σε γνωστικά (cognitive) χαρακτηριστικά. Στις περιπτώσεις αυτές, τα μέλη της ομάδας δυσκολεύονται να εκφράσουν συνειδητά τα συναισθήματά τους και παρουσιάζουν διαφορετικά μοτίβα αντιλήψεων αναφορικά με τη μεταφορά συναισθηματικών πληροφοριών κατά τη διάρκεια της ομαδικής εργασίας για την επίτευξη του κοινού σκοπού [105, 178]. Διαπιστώνουμε λοιπόν, ότι η δυνατότητα παρακολούθησης της εσωτερικής κατάστασης των μελών που αλληλεπιδρούν μεταξύ τους μέσα σε τέτοια περιβάλλοντα μπορεί να μας δώσει νέες ιδέες σχετικά με τους μηχανισμούς της μεταξύ

τους αλληλεπίδρασης, της ενσυναίσθησής τους και της γνωστικής τους ικανότητας.

Είναι γεγονός ότι κατά τη διάρκεια υψηλού φόρτου εργασίας, πίεσης και μεγάλης διάρκειας περιόδων άγχους, το συμπαθητικό νευρικό σύστημα είναι υπεύθυνο για την ενεργοποίηση όλων των αδένων και των οργάνων που σχετίζονται με την υπεράσπιση του ανθρωπίνου σώματος από τις αντιληπτές απειλές. Αυτή η ενεργοποίηση συνδέεται με αλλαγές στη διέγερση (arousal) που επηρεάζονται περαιτέρω από το συναίσθημα (emotion), τη γνώση (cognition) ή την προσοχή (attention). Οι υψηλής διάρκειας περίοδοι άγχους οδηγούν σε αυξημένη δραστηριότητα του συμπαθητικού νευρικού συστήματος και εκδηλώνονται μέσω μίας σειράς σωματικών αντιδράσεων, όπως η αύξηση του καρδιακού παλμού, η αύξηση της ροής του αίματος στα άκρα, η αύξηση του ρυθμού αναπνοής, κλπ. Έτσι, ο συνδυασμός περισσότερων του ενός σημάτων φυσιολογίας μπορεί να θεωρηθεί ως μία αξιόπιστη μέτρηση των μεταβολών του άγχους, αφού παρέχει αξιόπιστες εκτιμήσεις αναφορικά με τα συναισθήματα (emotion), τη διέγερση (arousal) και τη γενική γνώση (general cognition)[142].

Επιχειρούμε λοιπόν να εξετάσουμε το συσχετισμό μεταξύ των σημάτων φυσιολογίας των μελών της ομάδας (εκπαιδευτής-εκπαιδευόμενος) και των απαντήσεων τους κατά τη διάρκεια της συνομιλίας τους (turn-taking responses), η οποία ορίζεται ως το χρονικό διάστημα της έναρξης της ομιλίας του εκπαιδευόμενου από το τέλος της ομιλίας του εκπαιδευτή. Έχει παρατηρηθεί ότι τα μέλη μίας ομάδας όταν συμμετέχουν σε εξαιρετικά απαιτητικές εργασίες (operational tasks) δεν παρατηρούν συχνά τα ερεθίσματα που τους προκαλούν άγχος και πίεση [153, 222]. Έτσι, αν και επικοινωνούν με τα άλλα μέλη της ομάδας, τα συναισθήματα που εκφράζουν συχνά δεν είναι εμφανή μέσω της οπτικοακουστικής παρατήρησης. Για παράδειγμα, μια αναμενόμενη αντίδραση από ένα άτομο που του ζητείται να αποπλύνει μια προσομοιωμένη βόμβα, είναι τα υψηλά επίπεδα (εσωτερικού/μη ορατού) άγχους. Αυτή η έμφυτη διαφορά μεταξύ της εξωτερικής παρατηρήσιμης συμπεριφοράς των μελών μιας ομάδας και της εσωτερικής συναισθηματικής τους κατάστασης δεν είναι γίνεται εύκολα αντιληπτή και μπορεί να γεφυρωθεί μόνο μέσω της παρακολούθησης χαρακτηριστικών των σημάτων φυσιολογίας τους.

Η διάρκεια των απαντήσεων τους θεωρείται μία εξίσου σημαντική ένδειξη, καθώς σχετίζεται με τις πνευματικές διαδικασίες υψηλών ποσοστών άγχους [183]. Λαμβάνοντας υπόψιν μας ότι οι δείκτες φυσιολογίας αντικατοπτρίζουν τις πτυχές των ψυχικής/πνευματικής κατάστασης και συχνά το μέγεθος της δυσφορίας [63], διερευνούμε κατά πόσο τα σήματα φυσιολογίας σύντομης αλλά και μεγαλύτερης διάρκειας απαντήσεων σχετίζονται με διαφορετικά μοτίβα φυσιολογίας. Η περαιτέρω καταγραφή και ερμηνεία αυτής της συνεχιζόμενης και εξελισσόμενης αλληλεπίδρασης, βασίζεται στη χρήση δύο στρατηγικών μάθησης του συνόλου (ensemble learning). Πιστεύουμε ότι η διερεύνηση των νευροφυσιολογικών αλλαγών κατά τη διάρκεια των απαντήσεων των συμπαικτών παρέχει μία καλύτερη κατανόηση της δυναμικής της ομάδας.

Σύμφωνα με την τρέχουσα βιβλιογραφία, μία σειρά ερευνών έχει διεξαχθεί αναφορικά με τη δυναμική του διαλόγου και με το συσχετισμό του έντονου άγχους και της αυξημέ-

νης πίεσης με παραμέτρους φυσιολογίας. Παρόλα αυτά, τα προαναφερθέντα πεδία έχουν μελετηθεί σε μεγάλο βαθμό μεμονωμένα και όχι συνδυαστικά. Για παράδειγμα, οι εργασίες αναφορικά με τη μετάβαση της έναρξης της ομιλίας του εκπαιδευόμενου από το τέλος της ομιλίας του εκπαιδευτή [183] θα μπορούσαν να επωφεληθούν από τις προσωπικές σχέσεις των ατόμων που αλληλεπιδρούν την παρακολούθηση και την καταγραφή των αντιδράσεων φυσιολογίας τους όπως τα υψηλά επίπεδα άγχους κατά τη διάρκεια ομαδικών επιχειρησιακών καθηκόντων [53]. Επιπλέον, ο συσχετισμός των απαντήσεων με παραμέτρους φυσιολογίας μελετήθηκε και για την αξιολόγηση του τρόπου με τον οποίο τα επίπεδα θυμού των ενηλίκων μπορούν να επηρεάσουν τα παιδιά [63]. Τέλος, προηγούμενες μελέτες έχουν αποδείξει τα πλεονεκτήματα της χρήσης της μάθησης του συνόλου τόσο στη μονοτροπική [206, 202] όσο και στην πολυτροπική αναγνώριση των συναισθηματικών καταστάσεων [81, 201].

Από όσο γνωρίζουμε, δεν υπάρχει πειραματική επιβεβαίωση της εφαρμογής μεθόδων μάθησης του συνόλου (ensemble learning) για τη μελέτη της συσχέτισης μεταξύ της ομιλίας των μελών μιας ομάδας και των εσωτερικών συναισθηματικών καταστάσεων που συνάγονται από παραμέτρους φυσιολογίας δύο συμπαικτών (εκπαιδευτή-εκπαιδευόμενος) οι οποίοι προσπαθούν να αφοπλίσουν μια προσομοιωμένη βόμβα (πλαίσιο) [157].

Στόχος μας λοιπόν είναι να ξεδιπλώσουμε αυτή τη συσχέτιση. Τα ερευνητικά μας αποτελέσματα ενισχύουν την αρχική μας υπόθεση, ότι δηλαδή τα πρότυπα φυσιολογίας μεταδίδουν πληροφορίες της εσωτερικής κατάστασης του εκπαιδευόμενου, καθώς διαφοροποιούνται ανάλογα με τη διάρκεια των απαντήσεων του εκπαιδευτή. Τέλος, τα αποτελέσματά μας ενισχύονται επιπρόσθετα μέσω των μεθόδων μάθησης του συνόλου, που ξεπερνούν τους περισσότερους βασικούς ταξινομητές (base learners).

Μία από τις δύο κύριες κατευθυντήριες αρχές που διέπουν την ανθρώπινη συνομιλία είναι ότι μόνο ένα άτομο πρέπει να μιλάει κάθε φορά και ότι δεν μπορούμε να έχουμε σιωπή. Στα πλαίσια λοιπόν της οργάνωσης των συνομιλιών, όταν ακολουθούνται αυτοί οι κανόνες μπορούμε να διαπιστώσουμε ποιός μιλάει και ποιός θα μιλήσει στη συνέχεια. Αυτή η διαδικασία αναφέρεται συχνά στη βιβλιογραφία ως “στροφή” και ορίζεται ως η χρονική διάρκεια από το τέλος της ομιλίας ενός ατόμου έως την αρχή της πρώτης λέξης του άλλου συνομιλητή (turn-taking).

Οι “στροφές” συχνά κυμαίνονται από πολύ μικρές έως αρκετά μεγάλες, υποδεικνύοντας μικρότερης ή μεγαλύτερης διάρκειας αντίστοιχα συναισθηματικά και αγχωτικά επεισόδια, όπου παρατηρούνται φαινόμενα όπως οι επικαλύψεις ομιλίας. Αντίστοιχα, στην περίπτωση του εξεταζόμενου σώματος κειμένου, οι μεγαλύτερης διάρκειας απαντήσεις (στροφές) αποτελούν πολύτιμη πληροφορία σχετικά με την αντιληπτική, γνωστική και συναισθηματική κατάσταση του εκπαιδευόμενου, αντανακλώντας τόσο την εξωτερική του κατάσταση, που μπορεί εύκολα να παρατηρηθεί, όσο και τη μη παρατηρήσιμη εσωτερική συναισθηματική του κατάσταση. Επιλέγουμε λοιπόν να διερευνήσουμε αυτό το είδος αλληλεπιδραστικού πλαισίου μεταξύ των δύο μελών και των δύο ομάδων, βασιζόμενοι στην υπόθεση ότι η συμπεριφορά του εκπαιδευτή είναι περισσότερο ελεγχόμενη,

ελαχιστοποιώντας έτσι την επίδραση των διακυμάνσεων της συναισθηματικής του κατάστασης στη συναισθηματική κατάσταση του εκπαιδευόμενου.

Στα πλαίσια της προσπάθειάς μας να διαχωρίσουμε τις απαντήσεις (στροφές) ως προς τη διάρκειά τους σχεδιάζουμε ένα κατώφλι στο 70ο εκατοστημόριο των τιμών απόκρισης. Η τιμή αυτή υπολογίζεται εμπειρικά μετά από τη σχεδίαση των ιστογραμμάτων της στροφής των απαντήσεων του κάθε εκπαιδευόμενου ξεχωριστά. Οι αρνητικές τιμές αυτού του μέτρου αναφέρονται στις περιπτώσεις όπου ο εκπαιδευόμενος διακόπτει την ομιλία του εκπαιδευτή. Ευρήματα όπως οι αλληλοεπικαλύψεις ομιλίας και οι πολύ σύντομες διάρκειας προτάσεις ευθυγραμμίζονται με τα υψηλά επίπεδα του άγχους σε ιδιαίτερα απαιτητικά επιχειρησιακά καθήκοντα [97].

Μετά από προσεκτική επισκόπηση των διαφόρων περιπτώσεων της συναισθηματικής κατάστασης των μελών της ομάδας, παραθέτουμε μία σειρά ενδιαφερόντων παρατηρήσεων. Εντοπίζουμε μία σειρά παραδειγμάτων κατά τη διάρκεια της αλληλεπίδρασης στα οποία ο εκπαιδευτής εξηγεί τον τρόπο σύνδεσης των μπλε και των κόκκινων καλωδίων. Σε αυτή την περίπτωση, η απάντηση του εκπαιδευόμενου είναι σύντομη (δηλ., ο εκπαιδευόμενος χρησιμοποιεί λέξεις όπως *ok/ναι/όχι*). Αντίστοιχα, υπάρχουν και παραδείγματα όπου ο εκπαιδευτής εξηγεί με περισσότερες λεπτομέρειες τη διαδικασία απόπλισης της βόμβας για να επιβεβαιώσει κατά πόσο ο εκπαιδευόμενος ακολουθεί πραγματικά τις οδηγίες του.

Καταλήγουμε λοιπόν στη διαπίστωση ότι στην πρώτη περίπτωση, όπου η δήλωση είναι απλή, απαιτείται χαμηλή γνωστική προσπάθεια, δηλαδή μία σύντομη απάντηση, ενώ μία μεγαλύτερης διάρκειας απάντηση ακολουθεί στη δεύτερη περίπτωση, όπου ο εκπαιδευόμενος επαναλαμβάνει τις οδηγίες του εκπαιδευτή για να αποδείξει ότι τις κατάλαβε. Σε αυτή την περίπτωση, αναμένουμε ότι ο εκπαιδευόμενος θα βρίσκεται σε πνευματική εγρήγορση.

Ένα BIOPAC MP150, με μία τυποποιημένη διαμόρφωση ηλεκτροδίων μολύβδου II χρησιμοποιήθηκε για την καταγραφή των ηλεκτροκαρδιογραφικών αλλαγών (ECG), των μεταβολών της πίεσης του αίματος σε πραγματικό χρόνο και της καρδιογραφικής αντίστασης (impedance cardiography (ZKG)). Τα συνεχή δεδομένα καταγράφηκαν για κάθε μέλος και των δύο ομάδων που συμμετέχουν σε κάθε επιχειρησιακό καθήκον και αναλύονται εκτός σύνδεσης. Οι πρώτες χρονικές σειρές για κάθε επιχειρησιακό καθήκον (task) χωρίζονται σε διαστήματα των τριάντα δευτερολέπτων (30 secs) ξεκινώντας από το τέλος κάθε συνεδρίας (session). Η επιλογή των τριάντα δευτερολέπτων σχετίζεται με το γεγονός ότι αυτή είναι η απαιτούμενη διάρκεια για την ανάλυση καρδιαγγειακών δεδομένων για περαιτέρω ανάλυση. Η εξαγωγή των χαρακτηριστικών από τα δεδομένα πραγματοποιήθηκε με τη χρήση του προγράμματος Moving Ensemble Average Program (MEAP) [39].

Συγκεκριμένα, εξήχθησαν εικοσιτέσσερα (24) καρδιαγγειακά χαρακτηριστικά. Μεταξύ αυτών των χαρακτηριστικών συμπεριλαμβάνονται: ο καρδιακός ρυθμός (HR), ο αριθμός κολιακός (συσταλτικός) χρόνος (Left Ventral (systolic) Ejection Time (LVET)),

ο χρόνος_p, ο χρόνος_s, ο χρόνος_t, η χρονική διάρκεια συστολής της καρδιάς, η προσυσταλτική περίοδος (pre-ejection period (PEP)), η κοιλιακή συσταλτικότητα (Ventricular Contractility (VC)), η καρδιακή παροχή (Cardiac Output (CO)) και η ολική περιφερική αντίσταση (Total Peripheral Resistance (TPR)). Η προσυσταλτική περίοδος (PEP) είναι ο χρόνος από την έναρξη της καρδιακής αποβολής του μυός έως το άνοιγμα της αορτικής βαλβίδας. Όταν μειώνεται η προσυσταλτική περίοδος (PEP), αυξάνεται η κοιλιακή συσταλτικότητα (VC). Το VC σύμφωνα με τη βιβλιογραφία έχει αποδειχθεί ότι σχετίζεται με τα υψηλά επίπεδα άγχους κατά τη διάρκεια απαιτητικών επιχειρησιακών εργασιών [158, 188, 221, 210]. Αντίστοιχα, η καρδιακή παροχή (CO) είναι η ποσότητα του άντλησης του αίματος σε λίτρα ανά λεπτό. Το TPR απεικονίζει την αγγειοδιαστολή (δηλ., την περισσότερη ροή αίματος) και τη αγγειοσυστολή (δηλ., τη λιγότερη ροή), έννοιες οι οποίες σχετίζονται με την παρασυμπαθητική και τη συμπαθητική δραστηριότητα, αντίστοιχα.

Ερευνητικές προσπάθειες στον τομέα της φυσιολογίας έχουν αποδείξει ότι η TPR αυξάνεται αναμφισβήτητα όταν ένα άτομο βρίσκεται σε κατάσταση απειλής και μειώνεται αντίστοιχα σε μία κατάσταση πρόκλησης, ενώ η καρδιακή παροχή (CO) είτε παραμένει αμετάβλητη είτε μειώνεται σε κατάσταση απειλής και αυξάνεται σε συνθήκες πρόκλησης [233].

Διαπιστώνουμε λοιπόν, ότι υπάρχει ένα ευρύ φάσμα μεταβλητότητας μεταξύ των εκπαιδευόμενων αναφορικά με την εργασία που καλούνται να πραγματοποιήσουν, δηλαδή την επιτυχή απόπλιση της βόμβας. Αυτή η παρατήρηση υποδεικνύει ότι ενδεχομένως υπάρχουν μηχανισμοί που ενεργοποιούνται στους εκπαιδευόμενους μέσω των σημάτων φυσιολογίας τους, τα οποία δεν εντοπίζονται παρά μόνο όταν μελετηθούν οι επιδόσεις της μάθησης του συνόλου (ensemble learning), που όμως δεν φαίνεται να είναι αποδοτικές για τους εκπαιδευόμενους P101, P103, P201.

Αξίζει επίσης να σημειωθεί ότι σε αυτούς τους τρεις εκπαιδευόμενους, παρατηρείται αύξηση του καρδιακού τους ρυθμού (HR). Για περαιτέρω επεξεργασία αναφορικά με αυτή την ένδειξη, εξετάζουμε τις οπτικοακουστικές εγγραφές και τα σήματα του καρδιακού τους παλμού (HR). Παρατηρούμε ότι υπάρχει μία διαφορά στα επίπεδα διέγερσης (δηλ. άγχους) σε σχέση με το είδος της διάρκειας της απάντησης (μικρότερης/μεγαλύτερης) και ότι η διέγερση αυτή εμπεριέχεται τόσο σε μικρότερης όσο και σε μεγαλύτερης διάρκειας απαντήσεις, ανάλογα με τον εκάστοτε εκπαιδευόμενο.

Συγκεκριμένα, συναντάμε παραδείγματα εκπαιδευόμενων οι οποίοι χρειάστηκαν αρκετό χρόνο για να απαντήσουν αφού πρώτα έδωσαν λανθασμένη απάντηση και κατά συνέπεια τους ζητήθηκε να προσπαθήσουν ξανά για να επιβεβαιώσουν ότι κατανόησαν τα βήματα της απόπλισης της βόμβας. Ως εκ τούτου, φαίνεται ότι η εργασία την οποία κλήθηκαν να υλοποιήσουν αποτέλεσε ένα αρκετά αγχωτικό ερέθισμα για αυτούς και συνεπώς τα επίπεδα τόσο της γνωστικής δραστηριότητάς τους όσο και του άγχους τους αντίστοιχα ήταν υψηλά.

Την ίδια στιγμή, παρατηρήσαμε υψηλά επίπεδα διέγερσης και σε σύντομης διάρκειας παραδείγματα, κατά τα οποία, ο εκπαιδευόμενος χρησιμοποιούσε λέξεις όπως ok/ναί/όχι.

Μία τέτοια συναισθηματική κατάσταση δεν ευθυγραμμίζεται με το έργο “αφόπλισης της βόμβας”, δεδομένου ότι περιμέναμε ότι οι μικρότερης διάρκειας απαντήσεις θα σχετίζονται με χαμηλά επίπεδα διέγερσης. Αντιθέτως, τα πειραματικά μας αποτελέσματα υποδηλώνουν ότι παρόλο που ενδέχεται να μην υπάρχουν προφανή (ακουστικά/ορατά) σήματα διέγερσης, τα σήματα φυσιολογίας είναι ικανά να παρέχουν μία συμπληρωματική αλλά μη επικαλυπτόμενη πληροφορία αναφορικά με την κατάσταση ενός ατόμου γενικότερα.

Αυτό το εύρημα είναι μείζονος σημασίας, ιδιαίτερα σε απαιτητικά γνωστικά επιχειρησιακά καθήκοντα στα οποία ένα από τα δύο μέλη της ομάδας ελέγχει και κατευθύνει τη συζήτηση. Η συγκεκριμένη διαπίστωση ευθυγραμμίζεται και με προηγούμενες παλαιότερες αλλά και σύγχρονες ερευνητικές μελέτες [233, 79, 27].

Τέλος, καταλήγουμε ότι σε αυτό το τμήμα της διατριβής, εξετάσαμε τη συσχέτιση των σημάτων φυσιολογίας (αντιδράσεις που προκλήθηκαν στο σώμα (αύξηση καρδιακών παλμών κοκ)) των δύο μελών μίας ομάδας που συμμετείχαν σε ένα σενάριο “αφόπλισης μίας βόμβας” σε συνδυασμό με τις εκφραστικές τους απαντήσεις. Τα πειραματικά αποτελέσματα αναδεικνύουν ότι το λεκτικό κομμάτι της αλληλεπίδρασης του εκπαιδευτή και του εκπαιδευόμενου κατά τη διάρκεια αυτού του αγχωτικού καθήκοντος, μεταφέρει πληροφορίες σχετικά με την εσωτερική κατάσταση του εκπαιδευόμενου. Αντικατοπτρίζουν επίσης την ποσότητα των απαντήσεων μέσω της ομιλίας τους αναφορικά με ένα ερέθισμα, απαντήσεις οι οποίες μπορεί να είναι επιπρόσθετα συνδεδεμένες με το ποσοστό των μη ορατών κοινωνικο-γνωστικών δραστηριοτήτων και οι οποίες δεν είναι πάντα προφανείς μέσω των παραδοσιακών μεθόδων παρατήρησης. Τέλος, προτείνουμε δύο υπάρχοντα σύνολα μαθησιακών μεθόδων που είναι νέα στον τομέα της συλλογής των εξωτερικών και εσωτερικών ενδείξεων των ομιλητών, αποδεικνύοντας ότι αυτές οι μέθοδοι μπορούν να αποφέρουν βελτιώσεις σε σχέση με τις παραδοσιακές μεθόδους ανάλυσης. Επιβεβαιώνοντας λοιπόν τη μοναδικότητα των προσωπικών χαρακτηριστικών κάθε συμμετέχοντα διαπιστώσαμε ότι είναι ακόμη πιο έκδηλη η ανάγκη για την ύπαρξη υπολογιστικών μοντέλων που να εστιάζουν στην καλύτερη αναγνώριση, ερμηνεία και επεξεργασία των απόψεων (opinions) και συναισθημάτων (sentiments) των συμμετεχόντων κατά τη διάρκεια τέτοιων αλληλεπιδράσεων όπου ο προσδιορισμός του πλαισίου είναι μείζονος σημασίας.

Η ικανότητα των ανθρώπων να αλληλεπιδρούν αποτελεσματικά σε κοινωνικό επίπεδο εξαρτάται σε μεγάλο βαθμό από την ευαισθητοποίησή (awareness) τους στο πλαίσιο της αλληλεπίδρασης που λαμβάνει χώρα, όπου είναι σημαντικό να γνωρίζουμε τις πτυχές ενός φορμαλισμού του πλαισίου **W5+**, δηλαδή τις απαντήσεις στα ερωτήματα: **Who**, **What**, **Why**, **Where**, **What** και **How**. Ο στόχος αυτός επιτυγχάνεται γεφυρώνοντας το εννοιολογικό κενό και λαμβάνοντας υπόψιν τη σημασιολογία, τις γνωστικές και τις συναισθηματικές πληροφορίες της μη λεκτικής συμπεριφοράς. Διερευνούμε λοιπόν, το εάν και πώς αυτές οι πληροφορίες θα μπορούσαν να ενσωματωθούν στην αυτόματη ανάλυση της συναισθηματικής συμπεριφοράς. Έτσι, προτείνουμε μία μεθοδολογία εξαγωγής σημασιολογικής έννοιας, η εφαρμογή της οποίας παρουσιάζεται σε ενδεικτικά παραδείγματα από το σώμα κειμένου SEMAINE επικυρώνοντας την προτεινόμενη προσέγγιση.

Μία από τις κύριες προκλήσεις των τελευταίων ετών είναι η δημιουργία περισσότερο φυσικών, ευαίσθητων και ευφυών μηχανών, που δε θα είναι σε θέση να επικοινωνούν μόνο, αλλά θα μπορούν να κατανοήσουν τα σήματα καθώς και το πλαίσιο μέσα στο οποίο θα λαμβάνει χώρα ένα γεγονός [240]. Έτσι, εκτός από τη μονοτροπική ή πολυτροπική επικοινωνία μέσω ενός είτε περισσότερων καναλιών και μέσω του λεκτικού περιεχομένου (σημασιολογία), οι μηχανές υπολογιστών χρειάζεται να αναγνωρίζουν, να ερμηνεύουν και να επεξεργάζονται τις συναισθηματικές πληροφορίες όπως οι άνθρωποι. Στην ανθρώπινη γνώση, η σκέψη και το αίσθημα είναι άρρηκτα συνδεδεμένες έννοιες: τα συναισθήματά μας είναι συχνά προϊόν των δικών μας σκέψεων, καθώς και οι αντιδράσεις μας είναι συχνά το προϊόν αποτελέσματος της συναισθηματικής μας κατάστασης. Αλλά, τί σημαίνει για ένα μηχάνημα να είναι κοινωνικά ευφύε μετά την ενσωμάτωση του πλαισίου; Μέχρι στιγμής, κατά τη διάρκεια συνομιλιών, το εννοιολογικό πλαίσιο έχει οριστεί ως η προηγούμενη οπτική πληροφορία [62], η γενική κατανόηση της κατάστασης [30], η παρελθοντική λεκτική πληροφορία [119], το πολιτιστικό υπόβαθρο των συνομιλητών [140], το φύλο τους, η γνώση του γενικότερου πλαισίου αλληλεπίδρασης στο οποίο πραγματοποιείται ένα συναισθηματικό φαινόμενο [24] και οι κοινωνικές καταστάσεις [16]. Κατά συνέπεια, μελέτες σχετικά με έξυπνες διεπαφές για την επικοινωνία Ανθρώπου-Υπολογιστή (HCI), οι οποίες ενσωματώνουν το εννοιολογικό πλαίσιο, αντιστοιχίζονται με τα ακόλουθα ερωτήματα τα οποία συνθέτουν το **W5+ φορμαλισμό**: Με **Ποιόν** είσαι μαζί; (π.χ., αλληλεπιδράσεις μεταξύ δύο ή περισσότερων ατόμων [242]), **Τί** μεταδίδεται; (π.χ., (μη) γλωσσικό μήνυμα/λεκτικό σήμα και συναίσθημα), **Πώς** μεταδίδονται οι πληροφορίες; (μέσω των ανθρωπίνων συναισθηματικών συνθημάτων), **Γιατί**, δηλαδή, σε ποιο πλαίσιο οι πληροφορίες μεταβιβάζονται;, **Πού** είναι ο χρήστης; **Ποιά** είναι η τρέχουσα εργασία του; **Πώς** αισθάνεται (η διάθεσή του να είναι μεταβαλλόμενη από αρνητική σε θετική;); **Ποιά** είναι τα μέτρα που πρέπει να ληφθούν για την ικανοποίηση των αναγκών, των στόχων και των καθηκόντων του ανθρώπου; [60].

Δυστυχώς, μέχρι στιγμής οι προσπάθειες της κατανόησης των ανθρωπίνων συναισθημάτων είναι συνήθως ανεξάρτητες από το περιβάλλον-πλαίσιο στο οποίο πραγματοποιούνται [91]. Υπό το πρίσμα αυτών των παρατηρήσεων, η κατανόηση της διαδικασίας της φυσικής εξέλιξης των σχετικών με το πλαίσιο ερωτημάτων (context-related), όταν οι άνθρωποι αλληλεπιδρούν σε ένα κοινωνικό περιβάλλον, θα μπορούσε να προσφέρει νέες ιδέες για το μηχανισμό του πλαισίου αλληλεπίδρασης και της αποτελεσματικότητάς του.

Σύμφωνα με τη βιβλιογραφία τα ερωτήματα “Who”, “What”, “Where” έχουν απαντηθεί κυρίως ξεχωριστά ή σε ομάδες δύο ή τριών ερωτημάτων χρησιμοποιώντας τις πληροφορίες που εξάγονται από τα πολυτροπικά σήματα [266]. Έτσι, μέχρι σήμερα, δεν υπάρχει ένας συγκεκριμένος ορισμός του φορμαλισμού **W5+**, δεδομένου ότι τα συστήματα που απαντούν στις περισσότερες ερωτήσεις “W” βασίζονται σε διαφορετικές ψυχολογικές θεωρίες του συναισθήματος και αναπαραστάσεις μοντέλων ενσυναίσθησης εξυπηρετώντας κάθε φορά τους συγκεκριμένους σκοπούς και στόχους της εκάστοτε έρευνας.

Πρόσφατη έρευνα σχετικά με την πρόοδο στις ερωτήσεις του “Why” και “How” οδή-

γησε στον αναδυόμενο ερευνητικό τομέα της ανάλυση του συναισθήματος (Sentiment Analysis (SA)) από σώματα κειμένων Φυσικής Γλώσσας [165, 133, 29], μέσω του εντοπισμού και της εξόρυξης απόψεων (Opinion Mining (OM)) και συναισθημάτων (sentiments), όπου στόχος είναι η βαθιά κατανόηση των σημασιολογικών κανόνων της γλώσσας. Επιπλέον, η ερμηνεία των γνωστικών και συναισθηματικών πληροφοριών που σχετίζεται με τη Φυσική Γλώσσα και, ως εκ τούτου, βοηθούν στην απόκτηση νέων γνώσεων και στη λήψη αποφάσεων, σε σχέση με τις κοινωνικό-συναισθηματικές αξίες και τα ιδανικά του ατόμου, είναι ζωτικής σημασίας. Το πρόβλημα κατά την προσπάθεια μίμησης αυτών των γνωστικών και λειτουργικών διαδικασιών, είναι ότι ενώ οι γνωστικές πληροφορίες είναι συνήθως αντικειμενικές και αμερόληπτες, ενώ η απάντηση των ερωτημάτων “Γιατί” είναι περισσότερο υποκειμενική.

Σύμφωνα με αυτή την άποψη, ο μακροπρόθεσμος στόχος μας είναι να κατανοήσουμε εάν και με ποιό τρόπο το εννοιολογικό πλαίσιο μπορεί να ενσωματωθεί στην αυτόματη ανάλυση της ανθρώπινης συμπεριφοράς και να προτείνουμε ένα νέο πλαίσιο ενσωμάτωσης το οποίο να **(1)** περιλαμβάνει τον εντοπισμό και την εξαγωγή του σημασιολογικού πλαισίου, **(2)** να εμπλουτίζει περισσότερο μία σειρά Ψυχολογικών Βάσεων (Psychological Foundations) με τιμές αισθήσεων και **(3)** να ενισχύει τα συναισθηματικά μοντέλα με πληροφορίες του πλαισίου και με αναπαραστάσεις αναφορικά με την εκτίμηση αξιολόγησης, χρησιμοποιώντας διαθέσιμες στο κοινό πηγές γνώσης (OKS) στην επεξεργασία της Φυσικής Γλώσσας [244]. Το πρώτο βήμα προς αυτή την κατεύθυνση, είναι η διερεύνηση της σημασιολογικής γνωστικής και συναισθηματικής πληροφορίας, που σχετίζεται με το λεκτικό περιεχόμενο (σημασιολογία), το οποίο για τις ανάγκες της έρευνάς μας είναι ο διάλογος μεταξύ του χρήστη και του χειριστή του σώματος κειμένου SEMAINE [143], διατηρώντας σταθερό το σχετικό ερώτημα αναφορικά με το Πού πραγματοποιείται αυτή η συνομιλία.

Αυτή η μέθοδος επισημείωσης λαμβάνοντας υπόψιν το εννοιολογικό πλαίσιο, δεν περιορίζεται μόνο σε μια απλή συντακτική ανάλυση της συνομιλίας ¹. Στις παραδοσιακές μεθόδους επισημείωσης, τα συναισθήματα και οι πληροφορίες από τα συμφραζόμενα δεν συνάγονται πάντα από τις γνωσιακές εκτιμήσεις (appraisals) και συνεπώς οι πληροφορίες που σχετίζονται με το πλαίσιο αναφορικά με τις αιτίες της δημιουργίας αυτών των συναισθημάτων δεν λαμβάνονται υπόψιν [31].

Τα συναισθήματα είναι περίπλοκες καταστάσεις των ανθρωπίνων αισθημάτων (feelings) και οδηγούν σε σωματικές και ψυχολογικές αντιδράσεις επηρεάζοντας τόσο τη σκέψη όσο και τη συμπεριφορά μας. Η μελέτη των συναισθημάτων παραμένει ακόμα ένα ουσιαστικό και ανοικτό προς μελέτη κομμάτι της Ψυχολογίας. Αναφορικά με τον τομέα της Φυσικής Γλώσσας (Natural Language Processing (NLP)) το ερευνητικό ενδιαφέρον εστιάζεται στον εντοπισμό του συναισθήματος μέσα από το κείμενο. Κυρίως, η έρευνα σχετικά με την ανίχνευση των συναισθημάτων από το κείμενο έχει επικεντρωθεί καταγράφοντας λέξεις συναισθηματικού περιεχομένου με βάση τρία μοντέλα συναισθημάτων, δηλ. τις

¹Το μέγεθος του εξεταζόμενου παραθύρου της συνομιλίας αντιστοιχεί σε 16 στροφές.

αναπαραστάσεις σε κατηγορίες, τις διαστατικές αναπαραστάσεις και τις κατηγορίες γνωστικής αξιολόγησης, όπως αναπτύχθηκε στην Ενότητα 3.2.3.1 [28, 90].

Για ένα περισσότερο εφαρμόσιμο σημασιολογικό μοντέλο ενσωμάτωσης πλαισίου, έναντι ενός θεωρητικού, όπως το μοντέλο συνιστωσών (Componential Model), η έρευνα έχει επικεντρωθεί στη συγκέντρωση γνώμων και συναισθημάτων από τη Φυσική Γλώσσα. Το εγχείρημα αυτό θεωρείται μία ερευνητική πρόκληση, καθώς προϋποθέτει μία βαθιά κατανόηση των έμμεσων και σημασιολογικών γλωσσικών κανόνων, που σχετίζονται με μία σειρά ανεπίλυτων προβλημάτων της Φυσικής Γλώσσας, όπως ο χειρισμός άρνησης, η αναγνώριση ονομάτων-οντοτήτων, η άρση της αμφισημίας λέξεων κλπ. Οι προσεγγίσεις με βάση την έννοια [234] στοχεύουν στην κατανόηση της εννοιολογικής και της συναισθηματικής πληροφορίας που σχετίζεται με τις γλωσσικές απόψεις. Επιπροσθέτως, οι προσεγγίσεις που βασίζονται σε έννοιες μπορούν να αναλύσουν εκφράσεις πολλών λέξεων που δε μεταφέρουν απαραίτητα συγκίνηση, αλλά σχετίζονται με έννοιες που το κάνουν. Για παράδειγμα, αντί της συγκέντρωσης μεμονωμένων απόψεων για ένα “συμβάν” (π.χ. πάρτυ γενεθλίων), οι χρήστες γενικά ενδιαφέρονται περισσότερο για τη σύγκριση των διαφορετικών γεγονότων σύμφωνα με ένα συγκεκριμένο σύνολο σημασιολογικά συναφών εννοιών, π.χ. “κεικ”, “ένας έκπληκτος φίλος”, ή “το δώρο” (το οποίο μπορεί να θεωρηθεί ως επιπρόσθετη πληροφορία του εννοιολογικού πλαισίου για τη βελτίωση της αναζήτησης σχετικών αποτελεσμάτων). Αυτή η πληροφορία αναφέρεται σε γνώση η οποία είναι ήδη γνωστή και συνήθως δεν απαιτεί επιπλέον επεξηγήσεις, είναι όμως απαραίτητη για την σωστή αποδόμηση του κειμένου της Φυσικής Γλώσσας σε συναισθήματα. Για παράδειγμα, η έννοια “μικρό δωμάτιο” θα πρέπει να εκτιμηθεί ως αρνητικό σχόλιο για την αξιολόγηση ενός ξενοδοχείου ενώ η έννοια “μικρή ουρά” ως θετική για την αξιολόγηση ενός ταχυδρομείου, είτε η έννοια “διάβασε το βιβλίο” ως θετική για μια κριτική βιβλίου, αλλά αρνητική για μια κριτική ταινίας.

Διαπιστώνουμε λοιπόν ότι σταδιακά, η νέα διεπιστημονική περιοχή που βρίσκεται στο σταυροδρόμι μεταξύ της Συναισθηματικής Υπολογιστικής, της αλληλεπίδραση ανθρώπου-υπολογιστή (HCI), των κοινωνικών επιστημών, της γλωσσολογίας, της Ψυχολογίας, και των μεθόδων ενσωμάτωσης του εννοιολογικού πλαισίου διακρίνεται ως ένα ξεχωριστό ερευνητικό πεδίο. Είναι επομένως δυνατή η καλύτερη αναγνώριση, ερμηνεία και επεξεργασία των απόψεων (opinions) και των συναισθημάτων (sentiments), μέσω της ενσωμάτωσης πληροφοριών εννοιολογικού πλαισίου και συμφραζομένων για την καλύτερη κατανόηση των δεοντολογικών ζητημάτων σχετικά με τη φύση του νου και τη δημιουργία συναισθηματικών μηχανών. Για εφαρμογές σε ερευνητικά πεδία όπως HCI σε πραγματικό χρόνο και ανάλυση μεγάλων κοινωνικών δεδομένων [3], η βαθύτερη κατανόηση της Φυσικής Γλώσσας δεν είναι αυστηρή προϋπόθεση: μία αίσθηση της σημασιολογίας που σχετίζεται με το κείμενο και κάποιες επιπλέον συναισθηματικές πληροφορίες, είναι συχνά ικανοποιητικές για την εκτέλεση γρήγορων εργασιών όπως π.χ. η αναγνώριση συναισθημάτων και η ανίχνευση γνωστικής και συναισθηματικής πληροφορίας.

Όσον αφορά στην **επιλογή του σώματος**, η νέα τάση είναι η συλλογή δεδομένων

σε πραγματικό χρόνο μέσα από νέες πηγές ανάλυσης γνώσης και ανάλυσης αισθήσεων που αφθονούν. Οι κάμερες τοποθετημένες σε smartphones, touchpads ή άλλες συσκευές επιτρέπουν στους χρήστες να δημοσιεύουν γνωμοδοτήσεις σε μορφή ήχου ή οπτικοακουστικού περιεχομένου και όχι μόνο σε κείμενο. Εκτός από την ομιλούμενη γλώσσα σε γραπτό κείμενο για ανάλυση, η οπτικοακουστική μορφή παρέχει εξίσου μία ευκαιρία να εξουσιάσουμε τις απόψεις και το συναίσθημα.

Επιπλέον, όσον αφορά στην **παρουσία και στη θέση πολλαπλών λέξεων-εννοιών (multi-word concepts)** στο τμήμα κειμένου, απαιτείται περαιτέρω εξέταση, καθώς τα διγράμματα (bi-grams) και τριγράμματα (tri-grams), συχνά λαμβάνονται υπόψιν ως χρήσιμα χαρακτηριστικά. Ορισμένες μέθοδοι βασίζονται επίσης στην απόσταση μεταξύ των όρων. Πληροφορίες για τα μέρη του λόγου (POS) (ουσιαστικά, επίθετα, επιρρήματα, ρήματα, κλπ.) χρησιμοποιούνται επίσης ευρέως στην ανάλυση κειμένου ως βασικός τρόπος αποσαφήνισης λέξεων-εννοιών. Πιο συγκεκριμένα, ορισμένα επίθετα, έχει αποδειχθεί ότι είναι καλοί δείκτες του συναισθήματος και μερικές φορές χρησιμοποιούνται για την επιλογή χαρακτηριστικών αναφορικά με την ταξινόμηση του συναισθήματος. Με άλλα λόγια, στην περίπτωση αυτή, η ανίχνευση των συναισθημάτων πραγματοποιείται μέσω επιλεγμένων φράσεων και ενός αριθμού προκαθορισμένων μοντέλων POS, τα περισσότερα από τα οποία αποτελούνται από ένα επίθετο ή ένα επίρρημα. Ωστόσο, τέτοιες προσεγγίσεις καθώς και οι επιδόσεις τους είναι άρρηκτα συνδεδεμένες με τον εξεταζόμενο τομέα εφαρμογής.

Μετακινούμενοι από την πολυτροπική και τη μονοτροπική ανάλυση συναισθήματος στο πεδίο των προσωπικών ερμηνειών από το κείμενο (Sentiment Analysis) και γενικότερο στο πεδίο της Φυσικής Γλώσσας (NLP), αναφορικά με περισσότερο υπολογιστικές προσεγγίσεις, η συμβολή του προτεινόμενου πλαισίου και κατ' επέκταση της διατριβής, έγκειται σε μια πιο αποτελεσματική ανάλυση, ερμηνεία και επεξεργασία απόψεων και συναισθημάτων από δεδομένα κειμένου που εξετάζονται σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών αντίστοιχα. Σε αυτή την περίπτωση, λαμβάνονται υπόψιν οι σημασιολογικές, γνωστικές και συναισθηματικές πληροφορίες των υποκειμενικών απαντήσεων των ατόμων. Συγκεκριμένα, ερευνήσαμε μία μέθοδο βασισμένη στη μεθοδολογία CRF για να ενσωματώσουμε τη δομή των προτάσεων (σύνταξη και σημασιολογία) και τις πληροφορίες πλαισίου για τον εντοπισμό των συναισθημάτων των προτάσεων. Επιβεβαιώσαμε επίσης τη χρησιμότητα της θεωρίας της ρητορικής δομής (RST), λαμβάνοντας υπόψιν τον ρόλο του λόγου των τμημάτων κειμένου. Έτσι, η συμβολή μας στοχεύει στην επανεξέταση της αποτελεσματικότητας των μεθόδων CRF και RST αναφορικά με την ενσωμάτωση των πληροφοριών πλαισίου στα συστήματα ανάλυσης συναισθημάτων. Και οι δύο μέθοδοι αξιολογούνται σε δύο διαφορετικού μεγέθους και είδους πηγές πληροφοριών, στο σύνολο δεδομένων αναθεώρησης ταινιών και στο σύνολο δεδομένων Fine-grained Sentiment Data (FSD). Τέλος, συζητούμε τα διδάγματα που αντλήσαμε από αυτά τα πειραματικά περιβάλλοντα όσον αφορά στην αντιμετώπιση των ακόλουθων βασικών ερευνητικών ερωτημάτων, όπως το κατά πόσο υπάρχει ένας κατάλληλος τύπος αποθετηρίου

κοινωνικών μέσων για την ενσωμάτωση πληροφοριών σχετικών με τα συμφραζόμενα, κατά πόσο η επέκταση της ομάδας των επιλεγμένων χαρακτηριστικών θα μπορούσε να βελτιώσει την ενσωμάτωση περιεχομένου SA και ποιός είναι ο αποδοτικότερος συνδυασμός χαρακτηριστικών αναφορικά με την επίτευξη τέτοιων βελτιωμένων επιδόσεων.

Αναφερόμενοι στα πλεονεκτήματα αυτής της αναδυόμενης ερευνητικής περιοχής (context-aware SA), τα τελευταία σχετίζονται με την ανίχνευση σημαντικών γεγονότων στις ειδήσεις [100], με τη συλλογή απόψεων καταναλωτών για προϊόντα [113], με την εξαγωγή του προσανατολισμού (δηλαδή θετικής ή αρνητικής γνώμης) για το κείμενο [166], κλπ. Πρόσφατα, έχει αποδειχθεί ότι μπορεί να αποφέρει ανταγωνιστικά πλεονεκτήματα για επιχειρήσεις, καθώς εξάγει το συναίσθημα σε επίπεδο προτάσεων ή φράσεων (sentence- ή phrase-level) [165]. Μπορούμε επομένως να συμπεράνουμε ότι, η απόδοση των συστημάτων SA με γνώμονα το πλαίσιο στο οποίο λαμβάνουν χώρα εξαρτάται σε μεγάλο βαθμό από μία σειρά προκαθορισμένων παραμέτρων κατά τη διάρκεια της πειραματικής διαδικασίας [68, 165].

Ιδιαίτερο ενδιαφέρον περιστρέφεται γύρω από τον τομέα της SA όταν η επιλογή του σώματος κειμένου προέρχεται από μεγάλα κοινωνικά δεδομένα (π.χ., ταινίες, αξιολογήσεις προϊόντων) [166]. Προηγούμενες μελέτες έχουν προτείνει διάφορες προσεγγίσεις για την αύξηση της ευρωστίας τέτοιων συστημάτων που έχουν επίγνωση του πλαισίου. Η αυτοματοποιημένη ενσωμάτωση του συναισθήματος έχει χαρακτηριστεί ως πολλά υποσχόμενη προσέγγιση [165]. Από τη μια πλευρά, διαφορετικές τεχνικές Μηχανικής Μάθησης (Machine Learning (ML)), εκμεταλλεύονται μεθόδους διανυσματικών αναπαραστάσεων των κειμένων είτε μεθόδους που βασίζονται στη χρήση λεξικών [227] αποδίδοντας έτσι σημασιολογικό προσανατολισμό (polarity) σε μεμονωμένες λέξεις.

Άλλες ερευνητικές προσπάθειες προτείνουν τη χρήση κανόνων (rule-based) και μη επιβλεπόμενων μεθόδων μάθησης για τη βελτίωση της ανάλυσης συναισθημάτων από πηγές κειμένου. Οι μέχρι τώρα ενδείξεις αποδεικνύουν ότι όλοι οι αλγόριθμοι τελευταίας τεχνολογίας λειτουργούν καλά σε μεμονωμένες προτάσεις χωρίς όμως να λαμβάνουν υπόψη οποιαδήποτε πληροφορία πλαισίου. Παρόλα αυτά, η ακρίβειά τους είναι δραματικά χαμηλότερη σε επίπεδο κειμένου, λόγω του ότι δε λαμβάνουν υπόψη τα συμφραζόμενα.

Ως εκ τούτου, πρόσφατες μελέτες σχετικά με τα context-aware SA συστήματα έχουν αρχίσει να εξετάζουν μεθόδους που παρέχουν περισσότερο έλεγχο. Για παράδειγμα, οι ερευνητές του [36] χρησιμοποιούν τη μεθοδολογία Conditional Random Field (CRF) για να εντοπίσουν το υποκείμενο της εξεταζόμενης πρότασης, δηλαδή το άτομο που εκφέρει την άποψή του, προσεγγίζοντάς το ως μία ακολουθιακή διεργασία επισημείωσης (sequential tagging task) και στη συνέχεια ενσωματώνουν την πληροφορία πλαισίου στο σύστημα. Μία αντίστοιχη έρευνα [103] προτείνει τον εντοπισμό του υποκειμένου της γνώσης με τη χρήση CRF για την περαιτέρω ενσωμάτωση των πληροφοριών πλαισίου στα συστήματα SA. Εκτός από αυτή την κατεύθυνση της έρευνας, έχει χρησιμοποιηθεί επίσης και η ανάλυση του λόγου (discourse analysis) για την προσαρμογή της προηγούμενης πολικότητας των όρων. Περαιτέρω βελτιώσεις αναφέρθηκαν όταν προτάθηκε μία περισσότερο εξε-

λιγμένη προσέγγιση ανάλυσης του λόγου (discourse analysis) που βασίζεται στη θεωρία Rhetorical Structure Theory (RST) [138].

Επανεξετάζοντας λοιπόν την αποτελεσματικότητα αυτών των μεθόδων για την ενσωμάτωση πληροφοριών πλαισίου στην SA, η προτεινόμενη προσέγγισή μας συνδυάζει τις μεθόδους CRF και RST με την SA για την αποτελεσματικότερη ανάλυση του πλαισίου. Συγκεκριμένα, συγκριτικά με τους υπάρχοντες αλγορίθμους ανάλυσης γνώμης και συναισθήματος τόσο σε επίπεδο προτάσεων όσο και σε επίπεδο κειμένων, η προτεινόμενη προσέγγιση προσπαθεί να βελτιώσει την SA αξιοποιώντας πλήρως τη δομή των προτάσεων. Συγκεκριμένα χρησιμοποιεί α) τις πληροφορίες πλαισίου για τον εντοπισμό της σχέσης μεταξύ των προτάσεων και για τη βελτίωση της SA του κειμένου, β) λαμβάνει υπόψιν το σύνολο των λέξεων της γλώσσας του Διαδικτύου και των συναισθημάτων και τέλος γ) επεκτείνει το σύνολο των χαρακτηριστικών συμφραζομένων (context) που κάνουν χρήση σημασιολογικών, συντακτικών, δομικών και RST χαρακτηριστικών του πλαισίου.

Μέχρι στιγμής, ένα ευρύ φάσμα χαρακτηριστικών έχει ελεγχθεί μεμονωμένα από ένα μεγάλο αριθμό ερευνητικών ομάδων, κυρίως όμως με ελεγχόμενες συνθήκες και καθορισμένες παραδοχές.

Μία σημαντική ερευνητική κατεύθυνση στον context-aware SA τομέα είναι η βελτίωση της ευρωστίας αυτών των συστημάτων SA μετά την ενσωμάτωση του πλαισίου [165, 5]. Οι υπάρχουσες μέθοδοι βασίζονται σε λεξικά συναισθημάτων, τα οποία είναι ενιαίοι κατάλογοι συναισθηματικών όρων που υποδηλώνουν τις αλλαγές αισθήματος. Δημοφιλή παραδείγματα λεξικών αποτελούν το General Inquirer [223], το Subjectivity Lexicon και Subjectivity Annotations και το SentiWordNet.

Επιπλέον, η γνώση του κάθε τομέα (domain) διαδραματίζει σημαντικό ρόλο, καθώς το γλωσσικό πλαίσιο ενός όρου επηρεάζει συχνά και τη συσχέτισή του με τον εντοπισμό του συναισθήματος. Οι πρώιμες εργασίες για τη SA χρησιμοποίησαν συντακτικές σχέσεις για να προσδιορίσουν νέους όρους συναισθημάτων, οι οποίοι μπορούν να θεωρηθούν ως μία πρώιμη μορφή εκμετάλλευσης της έννοιας του πλαισίου [95]. Το συναίσθημα εκφράζεται συχνά με τρόπο που δυσχεραίνει τον εντοπισμό του κατά την επεξεργασία προτάσεων ή παραγράφων μεμονωμένα. Συνεπώς, το πλαίσιο παραμένει βασικό συστατικό για την περαιτέρω βελτίωση των συστημάτων SA.

Συνολικά, η έρευνα στην SA περιοχί οργανώνεται σε δύο κατηγορίες: σε μεθόδους Μηχανικής Μάθησης (Machine Learning) και σε μεθόδους που βασίζονται στη χρήση λεξικών (Lexicon-Based). Σύμφωνα με την πρώτη κατηγορία, ο αλγόριθμος εκπαιδεύεται με δεδομένα τα οποία συνοδεύονται με ετικέτες συναισθηματικού περιεχομένου και στη συνέχεια το μοντέλο αυτο χρησιμοποιείται για να κατηγοριοποιήσει νέα έγγραφα. Αυτή η μέθοδος όμως είναι συνήθως χρονοβόρα λόγω του μεγάλου όγκου δεδομένων.

Όσον αφορά στη μέθοδο που βασίζεται στη χρήση λεξικών, αυτή αφορά στην εξαγωγή και στην ομαδοποίηση των βαθμολογιών των όρων που σχετίζονται με τα συναισθήματα (δηλ. τις προηγούμενες πολικότητες) για την πρόβλεψη του συναισθήματος. Παρόλα αυτά, δεδομένου ότι τα λεξικά αυτά είναι προκαθορισμένα, δεν μπορούν να προσαρμο-

στούν σε καινούργιες ή σε συγκεκριμένες μορφές έκφρασης. Επιπλέον, οι προσεγγίσεις που βασίζονται στη χρήση λεξικών δεν παράγουν τόσο υψηλά επίπεδα εμπιστοσύνης κατά την ανάλυση, κάτι το οποίο παρέχεται αυτόματα μέσω των προσεγγίσεων της Μηχανικής Μάθησης.

Η κατανόηση του συναισθήματος των προτάσεων, μας επιτρέπει να συνοψίσουμε τις απόψεις που θα μπορούσαν να βοηθήσουν τους ανθρώπους να λαμβάνουν τεκμηριωμένες αποφάσεις. Όλοι οι υπερσύγχρονοι αλγόριθμοι λειτουργούν ικανοποιητικά σε μεμονωμένες προτάσεις χωρίς να λαμβάνουν υπόψιν οποιαδήποτε πληροφορία πλαισίου. Κατά συνέπεια, η ακρίβειά τους είναι δραματικά χαμηλότερη σε επίπεδο κειμένου. Υπάρχουν πολλές δυσκολίες εξαιτίας κυρίως των ιδιαίτερων χαρακτηριστικών και της διαφορετικότητας στη διάρθρωση της φράσης, στον τρόπο με τον οποίο οι άνθρωποι εκφράζουν τις απόψεις τους (π.χ. απόψεις που εκφράζονται εμμέσως διαμέσου της σύγκρισης κ.λπ.). Επιπλέον, η περίπλοκη δομή της φράσης και τα συναισθήματα κάνουν την ανάλυση της τελευταίας ακόμη πιο δύσκολη.

Για το λόγο αυτό, δεν εξετάζουμε μόνο τη σύνταξη που μπορεί να επηρεάσει το συναίσθημα, αλλά και τα emoticons, τις θετικές και αρνητικές λέξεις, τις λέξεις άρνησης (negation words), την ενσωμάτωση πληροφοριών σχετικά με τη δομή της φράσης, τις λέξεις που εμπεριέχουν έννοιες σύγκρισης, τη θέση των θετικών και των αρνητικών λέξεων μέσα στην πρόταση και τα χαρακτηριστικά πλαισίου-RST. Ως εκ τούτου, χρησιμοποιούμε ένα μοντέλο CRF [121] για να εντοπίσουμε και να συλλέξουμε τα συντακτικά και σημασιολογικά χαρακτηριστικά των προτάσεων και στη συνέχεια χρησιμοποιούμε το RST μοντέλο για να ξεδιπλώσουμε τις ρητορικές σχέσεις για την ανάλυση συναισθήματος.

Conditional Random Field (CRF): Το CRF είναι γνωστό για τις διαδικασίες επισημείωσης ακολουθιών [121]. Τα CRF αποτελούν μία κατηγορία διακριτού μη προσανατολισμένου πιθανοτικού γραφικού μοντέλου που εφαρμόζεται γενικά στην Αναγνώριση Προτύπων και στη Μηχανική Μάθηση και έχουν σχεδιαστεί ειδικά για τη βελτιστοποίηση της πρόβλεψης της δομής. Ένας “γενικός” ταξινομητής προβλέπει μία ετικέτα για ένα μόνο δείγμα χωρίς να αναφέρεται σε δείγματα “γειτονικά/συνδεδεμένα”, ωστόσο μία γραμμική αλυσίδα CRF μπορεί να λάβει υπόψιν το πλαίσιο.

Συνεπώς, στην περίπτωση που επιθυμούμε να καταγράψουμε τις πληροφορίες πλαισίου (π.χ. γειτονικές προτάσεις ή προτάσεις που συνδέονται με λέξεις μετάβασης), η διαδικασία αναγνώρισης του συναισθήματος αντιμετωπίζεται ως ένα είδος επισημείωσης αλληλουχιών. Συγκεκριμένα, τα CRF παρέχουν ένα πιθανοτικό πλαίσιο για τον υπολογισμό της πιθανότητας Y που είναι εξαρτημένη από το X , όπου το X είναι μία τυχαία μεταβλητή/διάνυσμα μίας αλληλουχίας δεδομένων προς επισημείωση και Y είναι μια τυχαία μεταβλητή/διάνυσμα ακολουθιών ετικετών. Αξίζει να αναφέρουμε στο σημείο αυτό ότι τα X και Y θα μπορούσαν να ακολουθούν και μία πολυπλοκότερη δομή γραφημάτων.

Μέχρι στιγμής, το μοντέλο CRF, στην απλούστερη μορφή του, έχει χρησιμοποιηθεί ευρέως στον τομέα επισημείωσης κειμένου [141]. Παραδείγματα CRF αποτελούν οι ερευνητικές προσπάθειες [121] και [141], στις οποίες γίνεται χρήση της γραμμικής δομής της

αλυσίδας. Μία άλλη παρατήρηση αποκαλύπτει ότι υπάρχει αντιστοιχία ένα-προς-ένα μεταξύ των καταστάσεων του μοντέλου και των ετικετών.

Rhetorical Structure Theory (RST): Σε ένα κείμενο Φυσικής Γλώσσας, οι σχέσεις μεταξύ των τμημάτων του κειμένου αρχικά “ξεδιπλώνονται” και εν συνεχεία χρησιμοποιούνται για τη διάκριση σημαντικών τμημάτων του κειμένου από λιγότερο σημαντικά τμήματα αναφορικά με τη συμβολή τους στο γενικότερο “συναισθηματικό αποτύπωμα” του κειμένου. Σε μία διάρθρωση του λόγου (discourse structure), η προαναφερθείσα διαδικασία είναι ένα σημαντικό μέρος του τί καθιστά ένα κείμενο συνεκτικό. Έτσι, η ανάλυση της δομής του λόγου χωρίζεται σε δύο διαδικασίες: στον κατακερματισμό του λόγου και στην ανάλυσή του. Η τμηματοποίηση του λόγου (Discourse Segmentation) σχετίζεται με τον προσδιορισμό ορίων σε μία ακολουθία λέξεων, προσδιορίζοντας το σημείο έναρξης των επόμενων μονάδων λόγου (units). Η ανάλυση του λόγου (Discourse Parsing) σχετίζεται με τον προσδιορισμό των σχέσεων μεταξύ των ακολουθιών των μονάδων λόγου (π.χ. η αιτιότητα, η αντίθεση κλπ.). Στην περίπτωση μας, το σύνολο αυτών των σχέσεων σχηματίζει ένα δέντρο. Τα φύλλα της αναπαράστασης του δέντρου του λόγου (Discourse Tree (DT)) αντιστοιχούν σε συνεχόμενα κείμενα, τα οποία ονομάζονται επίσης μονάδες στοιχειώδους λόγου (Elementary Discourse Units (EDUs)). Τέλος, τα EDUs συνδέονται μεταξύ τους επίσης με την ίδια μορφή σχέσεων (π.χ. αντίθεση) και τα μεγάλα πεδία/ζευγάρια κειμένου που προκύπτουν υποβάλλονται και αυτά σε μία αναδρομική σύνδεση. Μια κλίμακα που συνδέεται με μια ρητορική σχέση μπορεί να είναι είτε ένας πυρήνας (nucleus) είτε ένας δορυφόρος (satellite).

Προηγούμενες μελέτες σχετικά με την ανάλυση του λόγου (Discourse Analysis) έχουν αποδειχτεί αρκετά επιτυχείς για τον εντοπισμό των προσεγγίσεων της Μηχανικής Μάθησης και των χαρακτηριστικών που είναι περισσότερο χρήσιμα για την αυτόματη κατάτμηση (Discourse Segmentation) και ανάλυση (Discourse Parsing) του λόγου [220]. Δυστυχώς ένα από τα μειονεκτήματα αυτής της μεθόδου είναι ο υψηλός χρόνος επεξεργασίας που απαιτείται για την ανάλυση του λόγου στα κείμενα Φυσικής Γλώσσας [96], γεγονός που παρεμποδίζει την εφαρμογή αυτής της μεθόδου σε σενάρια μεγάλης κλίμακας.

Μελετήσαμε λοιπόν διεξοδικά τη χρησιμότητα του μοντέλου CRF και της θεωρίας του λόγου RST για να ενσωματώσουμε τα χαρακτηριστικά της δομής της πρότασης και τις πληροφορίες πλαισίου σε συστήματα που έχουν επίγνωση του συναισθηματικού περιβάλλοντος. Για την πειραματική αξιολόγηση των μοντέλων μας χρησιμοποιήσαμε δύο πηγές πληροφοριών που έχουν διαφορετικό μέγεθος και ανήκουν σε διαφορετικό είδος. Αρχικά, δείξαμε πώς να βελτιώσουμε την SA εκμεταλλευόμενοι πλήρως τη δομή των προτάσεων, χρησιμοποιώντας πληροφορίες πλαισίου για να καταγράψουμε τη σχέση μεταξύ των προτάσεων, βελτιώνοντας έτσι και την SA του εγγράφου. Ο λόγος αυτής της επιτυχίας έγκειται στην επέκταση του συνόλου των χαρακτηριστικών πλαισίου που χρησιμοποιούνται συνδυαστικά με σημασιολογικά, συντακτικά και συναισθηματικά χαρακτηριστικά RST έναντι των παραδοσιακών προσεγγίσεων.

Τα πειραματικά μας αποτελέσματα υποστηρίζουν ότι για να ταξινομήσουμε τη συλ-

λογία FSD, ο συνδυασμός λέξεων που αναφέρονται στα συναισθήματα σε συνδυασμό με τα μονογράμματα/διγράμματα (unigrams/bigrams) παρέχει επίσης αρκετά ακριβή αποτελέσματα. Επιπλέον, μόνο όταν συνδυάζουμε χαρακτηριστικά πλαισίου-RST με μονογράμματα/διγράμματα (unigrams/bigrams) μπορούμε να επιτύχουμε βελτιώσεις απόδοσης με σαφή πολικότητα σε επίπεδο προτάσεων. Τα σημαντικότερα χαρακτηριστικά των ταξινομητών πολικότητας καταγράφουν ουσιαστικά τον τρόπο με τον οποίο οι όροι πολικότητας (polarity terms) χρησιμοποιούνται σε μια πρόταση. Μία πιθανή επέκταση αυτής της δουλειάς θα μπορούσε να είναι η χρήση επιπρόσθετων χαρακτηριστικών συμβατών με τις μεθόδους μας αναφορικά με το εξεταζόμενο είδος δεδομένων καθώς και τη χρήση επιπρόσθετων μέτρων αξιολόγησης.

Η τελευταία συμβολή αυτής της διατριβής σχετίζεται με τον τομέα της ανάλυσης των συναισθημάτων και της εξόρυξης γνώμης και βασίζεται στην προτεινόμενη ιδέα μας να εκπαιδεύουμε ισχυρές αναπαραστάσεις χαρακτηριστικών από δεδομένα χωρίς ετικέτες χρησιμοποιώντας Νευρωνικά Δίκτυα και ιδιαίτερα χρησιμοποιώντας τα Παραγωγικά Ανταγωνιστικά Δίκτυα (GAN), η χρήση των οποίων έχει αποδειχτεί ιδιαίτερος δημοφιλής στον τομέα της Όρασης Υπολογιστών. Η πρωτοτυπία της προτεινόμενης μεθόδου έγκειται στην υλοποίηση του μοντέλου, στην επιλογή των υπερπαραμέτρων, στη χρήση της μη επιβλεπόμενης μάθησης και τέλος στην πειραματική επικύρωση των σωμάτων κειμένων που προέρχονται από διαφορετικές πηγές ανάλογα με το είδος και το μήκος τους.

Δεν υπάρχει αμφίβολία ότι η ικανότητα των μοντέλων Βαθιάς Μάθησης να μαθαίνουν ισχυρές αναπαραστάσεις χαρακτηριστικών από μη επισημειωμένα δεδομένα έχει αρκετές εφαρμογές σε μία μεγάλη ποικιλία εργασιών εκμάθησης μηχανών. Ένας τρόπος για να δημιουργήσουμε τέτοιες αναπαραστάσεις είναι να εκπαιδεύσουμε βαθιά παραγωγικά μοντέλα που να μπορούν να μάθουν να συλλάβουν τη σύνθετη κατανομή των πραγματικών δεδομένων. Οι προσεγγίσεις του ανταγωνιστικά παραγωγικού δικτύου (GAN) έδειξαν εντυπωσιακά αποτελέσματα στην παραγωγή μοντέλων εικόνων. Παρόλα αυτά, έχει γίνει σχετικά περιορισμένη δουλειά αναφορικά με την αξιολόγηση της απόδοσης αυτών των μεθόδων για την εκμάθηση της Φυσικής Γλώσσας, τόσο σε επιβλεπόμενες όσο και σε μη επιβλεπόμενες ρυθμίσεις σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών. Εκτεταμένη επικύρωση της έρευνάς μας βασίζεται σε μία σειρά πειραμάτων που πραγματοποιήθηκαν στα σύνολα δεδομένων: 20Newsgroups, το Movie Review (MR) και το σύνολο δεδομένων Fine-grained Sentiment (FSD). Η πειραματική ανάλυσή μας επιβεβαιώνει ότι τα GANs μπορούν να μάθουν με επιτυχή τρόπο αναπαραστάσεις κειμένων Φυσικής Γλώσσας και στα τρία προαναφερθέντα επίπεδα.

Η απόδοση των μεθόδων Μηχανικής Μάθησης εξαρτάται σε μεγάλο βαθμό από την επιλογή των δεδομένων ή την αναπαράσταση των χαρακτηριστικών στα οποία εφαρμόζονται. Για το λόγο αυτό, μεγάλο μέρος της ουσιαστικής προσπάθειας στην ανάπτυξη αλγορίθμων Μηχανικής Μάθησης σχετίζεται με το σχεδιασμό της διαδικασίας της προεπεξεργασίας και του μετασχηματισμού των δεδομένων και της αναπαράστασης των δεδομένων που μπορούν να υποστηρίξουν τη Μηχανική Μάθηση με αποτελεσματικό τρόπο.

Τέτοιου είδους μηχανική χαρακτηριστικών είναι σημαντική αλλά χρονοβόρα εργασία, ενώ συγχρόνως υπογραμμίζει και την αδυναμία των τωρινών αλγορίθμων Μηχανικής Μάθησης. Παρόλο που υπάρχει μια μεγάλη ποικιλία προσεγγίσεων στην αναπαράσταση της μάθησης (Representation Learning) γενικά, η βασική έννοια περιστρέφεται γύρω από την εκμάθηση συγκεκριμένων χαρακτηριστικών από τα δεδομένα και στη συνέχεια τη χρήση των τελευταίων για την επίλυση μιας σειράς λειτουργιών, για παράδειγμα, την επίλυση μιας ξεχωριστής (πιθανώς μη σχετιζόμενης) εργασίας για την οποία έχουμε ένα μεγάλο αριθμό επισημειωμένων παραδειγμάτων/δειγμάτων. Ως αποτέλεσμα, η εμφάνιση μιας σειράς δεδομένων μεγάλης κλίμακας, όπως το ImageNet [189], το οποίο περιέχει 14.197.122 εικόνες και έχει επισημειωθεί χειροκίνητα, επέτρεψε την ευρύτερη χρήση και τη δημοτικότητα των συνελκτικών νευρωνικών δικτύων (CNNs) ακόμη και σε λιγότερο σχετικούς τομείς, όπως αυτός της ιατρικής απεικόνισης. Επί του παρόντος, η πλειονότητα των ταξινομητών δεν είναι αποτελεσματική όπως αναμένεται όταν το μέγεθος του συνόλου δεδομένων εκπαίδευσης είναι μικρό. Κατασκευάζοντας ένα μεγάλο και επισημειωμένο σύνολο δεδομένων, ωστόσο, είναι μια χρονοβόρα διαδικασία και συνήθως απαιτεί γνώσεις του εκάστοτε εξεταζόμενου τομέα, καθιστώντας την διαδικασία αυτή ακόμη πιο δαπανηρή. Ως εκ τούτου, υπάρχει ένα χάσμα μεταξύ των οφελών ενός μεγάλου συνόλου δεδομένων και της δυσκολίας της απόκτησης επισημειωμένων δεδομένων. Προκειμένου να επεκταθεί το εύρος και η ευκολία εφαρμογής της Μηχανικής Μάθησης, θα ήταν ιδιαίτερα επιθυμητό οι αλγόριθμοι αναπαράστασης να είναι λιγότερο εξαρτημένοι από τη μηχανική των χαρακτηριστικών, έτσι ώστε οι νέες εφαρμογές να μπορούν να κατασκευαστούν γρηγορότερα. Αυτό θα μπορούσε να είναι δυνατό μέσω της εκμάθησης αναπαραστάσεων (Learning Representations) των δεδομένων που διευκολύνουν την εξαγωγή χρήσιμων πληροφοριών αναφορικά με την κατασκευή ταξινομητών. Μία καλή αναπαράσταση είναι αυτή που είναι επίσης χρήσιμη ως είσοδος σε έναν εποπτευόμενο (supervised) ταξινομητή. Μεταξύ των διαφόρων μεθόδων αναπαράστασης εκμάθησης (LR), το τρέχον κεφάλαιο επικεντρώνεται σε μεθόδους Βαθιάς Μάθησης: εκείνες δηλαδή που διαμορφώνονται από τη σύνθεση πολλαπλών μη γραμμικών μετασχηματισμών, με στόχο την απόδοση περισσότερο αφηρημένων- και, τελικά, περισσότερο χρήσιμων αναπαραστάσεων.

Μεταξύ των αποτελεσματικών προσεγγίσεων που έχουν προκύψει για την εκπαίδευση των βαθιά παραγωγικών μοντέλων, αυτό που βασίζεται στον παραλλαγμένο αυτόματο κωδικοποιητή (VAE) [115, 187] και αυτό που βασίζεται στην προσέγγιση που χρησιμοποιεί ανταγωνιστικά παραγωγικά (GAN) δίκτυα κυριαρχούν τα τελευταία χρόνια. Στο πρώτο, τα παρατηρούμενα δεδομένα x θεωρούμε ότι παράγονται από ένα σύνολο στοχαστικών λανθανόντων μεταβλητών z . Το VAE εισάγει ένα δίκτυο συμπερασμάτων (υλοποιημένο με τη χρήση ενός βαθιού νευρωνικού δικτύου) για να προσεγγίσουμε τις καταχρηστικές κατανομές πάνω από το z , και στη συνέχεια μεγιστοποιεί ένα κατώτερο όριο στην πιθανότητα καταγραφής του $p(x)$. Η τελευταία προσέγγιση χρησιμοποιεί τα GANs [83]. Στην αρχική διατύπωση του GAN, μια γεννήτρια βαθιού νευρωνικού δικτύου μαθαίνει να χαρτογραφεί δείγματα από μία αυθαίρετη κατανομή της παρατηρούμενης κατανομή

δεδομένων. Ένα δεύτερο βαθύ νευρωνικό δίκτυο που ονομάζεται διευκρινιστής είναι εκπαιδευμένο στο να διακρίνει μεταξύ των δειγμάτων της εμπειρικής κατανομής και των δειγμάτων που παράγονται από τη γεννήτρια. Η γεννήτρια εκπαιδεύεται στη δημιουργία δειγμάτων που θα ξεγελάσουν το δίκτυο του διευκρινιστή και έτσι ένα αντίπαλο παιχνίδι παίζεται μεταξύ των δύο δικτύων, συγκλίνοντας σε ένα σημείο που είναι ένα τοπικό ελάχιστο για το διευκρινιστή και ένα τοπικό μέγιστο για τη γεννήτρια αντίστοιχα. Τόσο η προσέγγιση VAE όσο και η GAN έχουν δείξει εντυπωσιακά αποτελέσματα στην παραγωγή μοντέλων εικόνων [181, 88]. Παρόλα αυτά, σχετικά λίγη δουλειά έχει πραγματοποιηθεί αναφορικά με την αξιολόγηση της απόδοσης αυτών των μοντέλων μάθησης για την αναπαράσταση της Φυσικής Γλώσσας. Τα GANs αναφορικά με την επεξεργασία της Φυσικής Γλώσσας (NLP) χαρακτηρίζονται ως ισχυρές μέθοδοι, καθώς ασχολούνται με τη δημιουργία προτάσεων: συγκεκριμένα, παράγουν προτάσεις με συγκεκριμένα χαρακτηριστικά (συναισθήματα και ερωτήσεις) και επωφελούνται από τη μη επιβλεπόμενη φύση αυτών των βαθιών νευρωνικών μοντέλων (DNN). Ένας λόγος για τον οποίο τα GANs δεν μπορούν να εφαρμοστούν απευθείας στον τομέα της Φυσικής Γλώσσας είναι το γεγονός ότι οι προτάσεις δεν αναπαρίστανται με συνεχή τρόπο και συνεπώς δεν μπορεί να υπάρξουν διαφοροποιήσεις. Αντίθετα, το κείμενο αναπαρίσταται από διακριτές λέξεις (tokens) (όπως “Άνθρωπος”, “κορίτσι”, κ.λπ.). Έτσι, όταν θέλουμε να ενημερώσουμε ελαφρώς την παραγόμενη πρόταση σύμφωνα με τη συμπεριφορά του διευκρινιστή (discriminator), δεν μπορεί να παραχθεί μία πρόταση. Στην περιοχή της Όρασης Υπολογιστών (CV), η έξοδος της γεννήτριας είναι μια εικόνα (ένας πίνακας που αποτελείται από πραγματικούς αριθμούς) που μπορεί να υποστεί μικρές ενημερώσεις ώστε να δυσκολευτεί το δίκτυο του διευκρινιστή να διακρίνει την πραγματική από την ψεύτικη εικόνα. Πρόσφατα, ωστόσο, τα VAE χρησιμοποιήθηκαν με επιτυχία για τη δημιουργία μοντέλων γλώσσας [18], για τη μοντελοποίηση εγγράφων και να απαντήσει σε ερωτήσεις [146]. Αυτό το έγγραφο προσπαθεί να διελευκάνει το κατά πόσο τα GANs μπορούν να χρησιμοποιηθούν για την εκμάθηση αναπαραστάσεων της Φυσικής Γλώσσας σε μία μη επιβλεπόμενη κατάσταση.

Ιδιαίτερα, στο έργο μας, διατυπώνουμε το πρόβλημα της Μηχανικής Μάθησης ως εξής. Προτείνουμε μια νέα επέκταση των GANs που αντικαθιστά τον παραδοσιακό δυαδικό διευκρινιστή ταξινομητή με αυτόν που αποδίδει μία κλιμακούμενη ενέργεια σε κάθε σημείο της περιοχής εξόδου της γεννήτριας. Ο διευκρινιστής ελαχιστοποιεί την απώλεια που χρησιμοποιείται για την εκπαίδευση των ταξινομητών “μεγίστου περιθωρίου”, ενώ η γεννήτρια προσπαθεί να δημιουργήσει δείγματα με χαμηλή ενέργεια κάτω από το διευκρινιστή. Δείχνουμε ότι μια ισορροπία Nash [196] κάτω από αυτές τις συνθήκες συνθέτει μια γεννήτρια που ταιριάζει με την κατανομή δεδομένων (υποθέτοντας άπειρη χωρητικότητα). Πραγματοποιούμε πειράματα με το διευκρινιστή, με τη μορφή ενός autoencoder denoising (DAE), που προαιρετικά συμπεριλαμβάνει ένα ρυθμιστή που επιβάλλει τα παραγόμενα δείγματα που έχουν υψηλή ομοιότητα συνημίτονου προς άλλα δείγματα στη μικρές ομάδες δεδομένων (mini-batch). Η προτεινόμενη αρχιτεκτονική του νευρωνικού δικτύου βασίζεται σε μια παραλλαγή του προσφάτως προτεινόμενου energy-based GAN

[270] που έχει αποδειχθεί ότι είναι κατάλληλο για την παραγωγή υψηλής ανάλυσης ψηφιακών εικόνων MNIST [52], παρέχοντας μια ποιοτική αξιολόγηση των αναπαραστάσεων εκμάθησης. Επιπλέον, αποφασίσαμε να αντικαταστήσουμε το τυποποιημένο πιθανοτικό GAN που χρησιμοποιήσαμε στο μοντέλο ενέργειας (χρησιμοποιώντας κατανομές του Gibbs [80]), όπως έχει προταθεί από το έργο των Kim και Bengio [114] και αποφασίσαμε να παρουσιάσουμε την ίδια ισορροπία Nash ως ένα βασικό δίκτυο GAN, αλλά μέσω μιας διαφορετικής και πιο γενικευμένης κατηγορίας loss functionals, όπως το hinge loss. Αυτή η επιλογή του πειραματικού σχεδιασμού βασίζεται στην προσπάθειά μας να επιλέξουμε το ζευγάρι των μοντέλων που συγκλίνει [263] και που επιδεικνύει μεγαλύτερη σταθερότητα από τα κανονικά GANs κατά τη διάρκεια της εκπαίδευσης [197].

Η Μηχανική Μάθηση για τον τομέα της Φυσικής Γλώσσας μπορεί να εφαρμοστεί σε ένα ευρύ φάσμα μορφών κειμένου, από κριτικές πολλών προτάσεων και σχόλια, σε εκφράσεις γνώμης μέσω λέξεων. Η πιο συχνή προσέγγιση είναι η ταξινόμηση σε επίπεδο κειμένου. Μια άλλη προσέγγιση, η ταξινόμηση σε επίπεδο προτάσεων, περιορίζει την ανάλυση σε απλές προτάσεις αντί για ολόκληρα κείμενα. Συνήθως είναι πιο δύσκολο να επιλυθεί το πρόβλημα, αφού δεν υπάρχουν πολλές πληροφορίες που μπορούν να χρησιμοποιηθούν από τον ταξινομητή, καθώς οι προτάσεις είναι συνήθως πολύ μικρότερες από τα κείμενα. Σε σύγκριση με την ανάλυση κειμένων και προτάσεων, η ανάλυση σε επίπεδο χαρακτηριστικών (aspect level) είναι περισσότερο fine-grained. Σκοπός του είναι να εξάγει και να συνοψίσει τις απόψεις των ανθρώπων που εκφράζονται στις πτυχές των οντοτήτων, που ονομάζονται επίσης και στόχοι (targets).

Αναφορικά με τις προσεγγίσεις Μηχανικής Μάθησης για τις NLP διεργασίες, η ταξινόμηση σε επίπεδο κειμένου ή φράσης υλοποιείται συχνά με τη χρήση **επιβλεπόμενων μεθόδων Μηχανικής Μάθησης**. Αυτό περιλαμβάνει την εκπαίδευση ενός μοντέλου χρησιμοποιώντας ένα μεγάλο σύνολο σχολιασμένων δεδομένων που είναι σχετικό με το θέμα (topic-specific). Οποιαδήποτε υπάρχουσα επιβλεπόμενη μέθοδος Μηχανικής Μάθησης μπορεί να εφαρμοστεί στην ταξινόμηση σε επίπεδο κειμένου, όπως μηχανές φορέα υποστήριξης (SVM) ή τα κρυφά Μαρκοβιανά μοντέλα (HMMs). Συγκεκριμένα, οι Pang κ.ά. [166] σύγκριναν πολλές μεθόδους Μηχανικής Μάθησης χρησιμοποιώντας έναν ταξινομητή αξιολόγησης ταινιών, καταλήγοντας στο συμπέρασμα ότι οι SVM και οι Naive Bayes ταξινομητές έχουν την καλύτερη απόδοση συνολικά. Η παλινδρόμηση (regression) σε επίπεδο κειμένου ή φράσης είναι κρίσιμης σημασίας δεδομένου ότι πολλά προβλήματα δεν μπορούν να λυθούν αποτελεσματικά με θετική-αρνητική ταξινόμηση, η οποία χρησιμοποιείται συχνά για κριτικές προϊόντων, όπου επικρατεί βαθμολογία 1-5 αστερών. Οι Pang et al. [164] σύγκριναν διάφορες μεθόδους παλινδρόμησης, όπως η παλινδρόμηση SVM και η ταξινόμηση των πολυάριθμων SVM (SVM multiclass). Οι Qu et al. [179] επέκτειναν την αναπαράσταση bag-of-words εκμεταλλευόμενοι τους τροποποιητές της άρνησης και του συναισθήματος, οι οποίοι έχουν μεγαλύτερη επιρροή στην παλινδρόμηση απ' ό,τι σε προβλήματα ταξινόμησης. Αντίστοιχα, οι Mejnava et al. [145] συγκέντρωσαν στρατηγικές επιλογής χαρακτηριστικών, όπως το stemming, η συχνότητα εμφάνισης των όρων,

τα n-grams, το σημείο ομιλίας και χαρακτηριστικά εμπλουτισμένα με άρνηση. Κατέληξαν λοιπόν στο συμπέρασμα ότι ένα μικρότερο σύνολο χαρακτηριστικών είναι περισσότερο αποδοτικό συγκριτικά με ένα μεγαλύτερο σύνολο μεγάλων βάσεων δεδομένων.

Από την άλλη πλευρά, η **μη επιβλεπόμενη Μηχανική Μάθηση** για διεργασίες της Φυσικής Γλώσσας βασίζεται στην κυρίαρχη επίδραση των λέξεων και φράσεων συναισθηματικού περιεχομένου για την εκτέλεση της ταξινόμησης χωρίς τη χρήση δαπανηρών επισημειωμένων δεδομένων. Αυτό επιτυγχάνεται είτε με την εξαγωγή των συντακτικών προτύπων των προτάσεων είτε με τη χρήση λεξικών συναισθηματικών όρων. Και οι δύο προσεγγίσεις βασίζονται στη μέτρηση του προσανατολισμού του συναισθήματος (SO) των φράσεων και κατ' επέκταση ολόκληρου του κειμένου. Η ανάλυση που βασίζεται στη χρήση λεξικών με όρους συναισθηματικού περιεχομένου υπολογίζει τις τιμές του συναισθηματικού προσανατολισμού (SO) των λέξεων και φράσεων, που συνοψίζονται στην πολικότητα ολόκληρου του κειμένου. Γενικά, οι ταξινομητές συναισθημάτων που βασίζονται στα λεξικά συναισθηματικών όρων, δείχνουν μια θετική μεροληψία/προτίμηση, η οποία μπορεί να καθοριστεί προσαρμόζοντας την τιμή των περισσότερο σπάνιων αρνητικών εκφράσεων [226]. Μία από τις ελλείψεις τους είναι ότι οι μέθοδοι που βασίζονται σε λεξικά συναισθηματικών όρων δεν λειτουργούν καλά σε δεδομένα που εξαρτώνται από τον εξεταζόμενο τομέα, καθιστώντάς τα λιγότερο αποτελεσματικά όταν χρησιμοποιούνται για τομείς που παρουσιάζουν περισσότερο ενδιαφέρον για την ανάλυση των αισθήσεων (SA), όπως η πολιτική. Το ζήτημα αυτό εξετάζεται εν μέρει στο [268]; παρόλα αυτά οι επιβλεπόμενες μέθοδοι μάθησης εξακολουθούν να υπερβαίνουν τις μεθόδους που βασίζονται στη χρήση συγκεκριμένων λεξικών, με εξειδικευμένη ορολογία αναφορικά με τον εξεταζόμενο τομέα.

Πρόσφατα, τα νευρωνικά δίκτυα (NNs) άρχισαν να επεκτείνονται στο πεδίο της Φυσικής Γλώσσας με τη μορφή τόσο των επιβλεπόμενων όσο και των μη επιβλεπόμενων μεθόδων αναπαράστασης μάθησης. Από την πλευρά της **μη επιβλεπόμενης αναπαράστασης της μάθησης** [160], αναπτύχθηκε και επικυρώθηκε ένα μεγάλο μέρος της πρώιμης έρευνας αναφορικά με τη σύγχρονη βαθιά μάθηση [98, 102, 238, 40]. Παρόλο που οι επιβλεπόμενες προσεγγίσεις έχουν σαφείς στόχους που μπορούν να βελτιστοποιηθούν άμεσα, οι μη επιβλεπόμενες μάθησης προσεγγίσεις βασίζονται σε εξουσιοδοτημένες εργασίες, όπως η ανακατασκευή, η εκτίμηση της πυκνότητας ή η παραγωγή/δημιουργία, εργασίες οι οποίες δεν ενθαρρύνουν άμεσα χρήσιμες αναπαραστάσεις συγκεκριμένων διεργασιών. Ως αποτέλεσμα, μεγάλος όγκος δουλειάς έχει πραγματοποιηθεί αναφορικά με το σχεδιασμό στόχων και αρχιτεκτονικών που προορίζονται να ενθαρρύνουν την εκμάθηση χρήσιμων αναπαραστάσεων.

Παρά τις δυσκολίες αυτές, υπάρχουν αξιόλογες εφαρμογές μη επιβλεπόμενης μάθησης. Pretrained διανύσματα λέξεων είναι ζωτικής σημασίας για πολλά σύγχρονα συστήματα NLP [41]. Αυτές οι αναπαραστάσεις, η εκμάθηση των οποίων έχει πραγματοποιηθεί μέσω της μοντελοποίησης λέξεων που εμφανίζουν συσχέτιση (word co-occurrences), αυξάνουν την αποδοτικότητα των δεδομένων και την ικανότητα της γενίκευσης των συστη-

μάτων Φυσικής Γλώσσας [171, 34]. Η εκμάθηση αναπαράστασης φράσεων, προτάσεων και εγγράφων είναι ένας ανοικτός χώρος της έρευνας. Εμπνευσμένοι από την επιτυχία των διανυσμάτων από λέξεις, η δουλειά των Kiros et al. [117] πρότεινε skip-thought διανύσματα, μία μέθοδο εκπαίδευσης μίας πρότασης κωδικοποιητή προβλέποντας την προηγούμενη και την επόμενη πρόταση. Η αναπαράσταση που προέκυψε από την εκμάθηση αυτού του μοντέλου λειτουργεί ανταγωνιστικά σε μία ευρεία σειρά διεργασιών που έχουν αξιολογηθεί. Οι περισσότερες προηγμένες τεχνικές εκπαίδευσης, όπως η εξομάλυνση του στρώματος [8], βελτίωσαν περαιτέρω τα αποτελέσματα. Ωστόσο, η απόδοση των skip-thought διανυσμάτων εξακολουθεί να ξεπερνιέται από τα μοντέλα επιβλεπόμενης μάθησης, τα οποία βελτιστοποιούν άμεσα την επιθυμητή μετρική της απόδοσης για ένα συγκεκριμένο σύνολο δεδομένων. Αυτό συμβαίνει και για τις δύο εργασίες ταξινόμησης κειμένου, οι οποίες μετρούν κατά πόσο μία συγκεκριμένη έννοια είναι σωστά κωδικοποιημένη σε μια αναπαράσταση. Αυτό συμβαίνει ακόμη και όταν τα σύνολα δεδομένων χαρακτηρίζονται ως μικρά σύμφωνα με τα σύγχρονα πρότυπα, τα οποία συχνά αποτελούνται μόνο από μερικές χιλιάδες επισημειωμένα παραδείγματα.

Αναφορικά με τις προσεγγίσεις Βαθιών Νευρωνικών Δικτύων σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών, συγκεντρώνουμε και παραθέτουμε τις τελευταίες ερευνητικές προσπάθειες, αναφορικά με την εφαρμογή των Νευρωνικών Δικτύων (ΝΔ) για διεργασίες της Φυσικής Γλώσσας, εστιάζοντας στην ανάλυση κειμένων, προτάσεων και χαρακτηριστικών.

Οι Rahman et al. [182] εισήγαγαν μια παραλλαγή της μακράς βραχυπρόθεσμης μνήμης (LSTM) και τη δοκίμασαν στη Large Movie Review Dataset των κειμένων των 50K. Επιπλέον, σύμφωνα με την έρευνα των τελευταίων αποδείχτηκε ότι παρόλο που η ακρίβεια του συστήματος δεν αυξήθηκε σε σύγκριση με τα κανονικά μοντέλα LSTM, η σταθερότητα και η ακρίβειά του παρουσίασαν βελτίωση. Επιπλέον, σύμφωνα με την εργασία [122], η τελευταία εισήγαγε ένα επαναλαμβανόμενο συνελκτικό νευρωνικό δίκτυο (RCNN) που εφάρμοσε μια αμφίδρομη επαναλαμβανόμενη δομή για να εντοπίσει τις πληροφορίες πλαισίου του κειμένου και στη συνέχεια χρησιμοποίησε ένα στρώμα μέγιστης συγκέντρωσης για να καταγράψει τα βασικά τμήματα-μέρη του κειμένου. Οι ερευνητές του [213] πραγματοποίησαν δοκιμές σύγκρισης με ένα CNN, ένα LSTM και ένα συνδυασμό CNN και LSTM στη βάση δεδομένων Large Movie Review για την ταξινόμηση του συναισθήματος, αποδεικνύοντας ότι το δίκτυο CNN + LSTM είναι αποτελεσματικότερο σε σχέση με τα αντίστοιχα προαναφερθέντα Νευρωνικά Δίκτυα μεμονομένα. Μια παρόμοια ερευνητική προσπάθεια διεξήχθη από τους συγγραφείς της [261] συνδυάζοντας τα δίκτυα CNN και LSTM για τη δημιουργία πολλαπλών παραλλαγών, οι οποίες επιτυγχάνουν αποδόσεις της τελευταίας τεχνολογίας. Για άλλη μια φορά, χρησιμοποιήθηκε η βάση δεδομένων Large Movie Review, αλλά οι συγγραφείς εξέφρασαν την πεποίθησή τους ότι το προτεινόμενο μοντέλο μπορεί να λειτουργήσει αποδοτικά τόσο σε εφαρμογές ήχου όσο και σε βίντεο. Τέλος, οι Liu et al. [136] πειραματίστηκαν με RNNs, εισάγοντας την έννοια της πολυδιάστατης εκμάθησης (multitask learning), η οποία συνδέει όλα τα σχετικά κα-

θήκοντα σε ένα ενιαίο σύστημα, το οποίο εκπαιδεύεται από κοινού. Οι συγγραφείς πραγματοποίησαν ταξινόμηση τόσο σε επίπεδο κειμένου στη βάση δεδομένων Large Movie Review αλλά και σε επίπεδο προτάσεων στη Stanford Sentiment Treebank (SST) βάση δεδομένων, επιτυγχάνοντας κορυφαία αποτελέσματα απόδοσης μετά τον εντοπισμό του συνδυασμού των περισσότερων αποδοτικών παραμέτρων (fine-tuning).

Επιπλέον, τα Νευρωνικά Δίκτυα έχουν χρησιμοποιηθεί ευρέως για την ταξινόμηση **σε επίπεδο προτάσεων**. Ιδιαίτερα, οι Chen et al. [35] πρότειναν ένα μοντέλο CNN-LSTM το οποίο ταξινομεί πρώτα τις προτάσεις σε non-target, one-target και σε προτάσεις πολλαπλών στόχων (multi-target). Το μοντέλο τους δοκιμάστηκε σε πολλά σύνολα δεδομένων (Stanford Sentiment Treebank, σε κριτικές ταινιών και σε σχόλια προϊόντων του Amazon) πετυχαίνοντας κορυφαία απόδοση σε μερικά από αυτά. Σύμφωνα με το ότι ο συνδυασμός CNN και RNN θεωρείται πολύ δημοφιλής προσέγγιση, οι συγγραφείς των [251] πρότειναν μία παραλλαγή του, η οποία πέτυχε υψηλή ακρίβεια στις βάσεις δεδομένων Stanford Sentiment Treebank και Movie Reviews. Επιπλέον, οι Conneau et al. [43] πρότειναν ένα πολύ βαθύ συνελκτικό δίκτυο που αποτελείται από 29 στρώματα, με αποτέλεσμα τη βελτιωμένη απόδοση και την απόδειξη της χρήσης πολλαπλών επιπέδων για τον τομέα της Φυσικής Γλώσσας επίσης. Ειδικότερα, οι συγγραφείς αξιολόγησαν το μοντέλο τους σε οκτώ διαφορετικά σύνολα δεδομένων που περιλαμβάνουν κριτικές ταινιών και ειδήσεις, δοκιμάζοντας διάφορα μοντέλα Φυσικής Γλώσσας όπως SA, την κατηγοριοποίηση ειδήσεων και την ταξινόμηση θέματος. Επιπλέον, οι Wang J et al. [251] πρότειναν ένα περιφερειακό μοντέλο CNN-LSTM το οποίο χρησιμοποίησε μία ξεχωριστή πρόταση ως την εξεταζόμενη περιοχή για την εξαγωγή των συναισθηματικών πληροφοριών. Το μοντέλο αυτό εκπαιδεύτηκε με το σύνολο δεδομένων Stanford Sentiment Treebank και με Κινέζικα κείμενα χρησιμοποιώντας 2K προτάσεις από κοινωνικά φόρουμ, παρουσιάζοντας καλύτερη απόδοση από τα CNNs, RNNs ή LSTMs αναφορικά με την πρόβλεψη του σθένους και της διέγερσης, αντίστοιχα.

Τέλος, διερευνήθηκε και η ανάλυση της Φυσικής Γλώσσας **σε επίπεδο χαρακτηριστικών (aspect level)**. Οι Du et al. [59] μοντελοποίησαν τα συναισθήματα και το συντακτικό πλαίσιο κάτω από συγκεκριμένες πτυχές για να αποκτήσουν καλύτερη ενσωμάτωση λέξεων, οι οποίες δόθηκαν ως εισροή σε ένα CNN για την κατάταξη συναισθημάτων των αξιολογήσεων προϊόντων του Amazon. Τα αποτελέσματά τους έδειξαν μία βελτίωση σε σύγκριση με τις παραδοσιακές μεθόδους ενσωμάτωσης λέξεων. Αντίστοιχα, οι Wang et al. [252] εφάρμοσαν τα δίκτυα LSTMs για την ταξινόμηση συναισθημάτων σε επίπεδο χαρακτηριστικών, επιτυγχάνοντας μια κορυφαία απόδοση του 89.9%. Το μοντέλο αξιολογήθηκε χρησιμοποιώντας το σύνολο δεδομένων SemEval 2014 ενώ οι αρχικές ενσωματώσεις λέξεων έγιναν από το Glove², καταγράφοντας τα σημαντικά μέρη της φράσης όταν δίνονται διάφορα χαρακτηριστικά. Οι Poria et al. [177] εξέτασαν ένα βαθύ CNN για την εξαγωγή χαρακτηριστικών χρησιμοποιώντας 7 επίπεδα NN. Για την αξιολόγηση χρησιμοποιήθηκε το σύνολο δεδομένων SemEval 2014 και ένα σύνολο δεδομένων SA και

²<https://nlp.stanford.edu/projects/glove/>

παρατηρήθηκε βελτίωση στα ποσοστά της ακρίβειας και της ανάκλησης. Έχουν εφαρμοστεί επίσης δίκτυα βαθιάς μνήμης σε ταξινόμηση κειμένου ανά επίπεδο και έχουν παρουσιάσει συγκρίσιμα αποτελέσματα με τα LSTM ενώ είναι 15 φορές ταχύτερα την ίδια στιγμή [230].

Αφού συνοψίσαμε τις τρέχουσες και τις προηγούμενες ερευνητικές προσπάθειες που έχουν διεξαχθεί στην περιοχή των βαθιών νευρωνικών δικτύων σε επίπεδο των κειμένων, της φράσης και των χαρακτηριστικών της ανάλυσης, παρατηρήσαμε ότι, στις περισσότερες περιπτώσεις, έχουν χρησιμοποιηθεί δίκτυα όπως το CNN ή τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNN) και συγκεκριμένα τα δίκτυα αμφίδρομης μακράς βραχυπρόθεσμης μνήμης (BLSTMs). Επομένως, πολύ λίγες ερευνητικές προσπάθειες έχουν διεξαχθεί για την ανάλυση κειμένων χρησιμοποιώντας παραγωγικά μοντέλα και συγκεκριμένα GAN. Θεωρώντας ότι ένας από τους στόχους της εργασίας μας είναι να εξετάσουμε κατά πόσο τα GANs μπορούν να χρησιμοποιηθούν για την εκμάθηση των αναπαραστάσεων των κειμένων της Φυσικής Γλώσσας με ένα ανεξέλεγκτο τρόπο, στην συνέχεια, συνοψίζουμε τα περισσότερα από τα πλέον σύγχρονα ερευνητικά έργα που έχουν χρησιμοποιήσει την προαναφερθείσα αρχιτεκτονική. Επιπλέον, κινητοποιημένοι από το γεγονός ότι η δομή του GAN δικτύου μπορεί να ενσωματώσει διάφορες λειτουργίες απώλειας (loss functions), το προτεινόμενο μοντέλο μας παρέχει ένα μεγαλύτερο βαθμό ελευθερίας. Επιπλέον, υποστηρίζουμε ότι είναι ικανό να προσφέρει ελπιδοφόρες και δημιουργικές λύσεις που να παράγουν χρήσιμα δεδομένα.

Αναφορικά με τα Παραγωγικά Ανταγωνιστικά Δίκτυα για τις NLP διεργασίες, εξετάζουμε την πρόσφατη έρευνα αναφορικά με τη διερεύνηση της δομής της Φυσικής Γλώσσας με τη χρήση δικτύων (VAE)s [116] και GANs [83]. Η αξιολόγηση των βαθιών παραγωγικών μοντέλων θεωρείται ερευνητική πρόκληση μέχρι στιγμής. Από όσο γνωρίζουμε, υπάρχουν ελάχιστες ερευνητικές προσπάθειες για την ανάλυση του κειμένου με τη χρήση των GANs. Οι Zhang et al. [269] πρότειναν ένα πλαίσιο για τη χρήση του LSTM και του CNN για την αμοιβαία ανταγωνιστική εκπαίδευση για την παραγωγή ρεαλιστικού κειμένου. Ο latent code z τροφοδοτείται στη γεννήτρια LSTM σε κάθε βήμα, ενώ το CNN λειτουργεί ως ταξινομητής δυαδικών προτάσεων ο οποίος διακρίνει μεταξύ πραγματικών δεδομένων και δειγμάτων που παράγονται. Τα LSTM δίκτυα ή οι περιοδικές επαναλαμβανόμενες μονάδες (GRU) που διατηρούν τη μακροπρόθεσμη μνήμη των tokens, έχουν αποδείξει ότι λειτουργούν καλά στην πράξη χρησιμοποιώντας την maximum likelihood estimation.

Ωστόσο, η εκπαίδευση με τη μέγιστη πιθανότητα έχει τα μειονεκτήματά της, όπως η μεροληψία έκθεσης (exposure bias). Αυτό αναφέρεται στην κατάσταση κατά την οποία, κατά τη διάρκεια της εκπαίδευσης, η πρόβλεψη της επόμενης λέξης η οποία εξαρτάται από την προηγούμενη είναι μη εφικτή, καθώς η προηγούμενη λέξη μπορεί να μην έχει εμφανιστεί στα δεδομένα εκπαίδευσης. Εάν η γεννήτρια κάνει ένα λάθος νωρίς στη διαδικασία παραγωγής, η παραγόμενη πρόταση θα παραμείνει αποκλίνουσα καθώς παράγονται περισσότερες λέξεις. Για την επίλυση αυτού του ζητήματος, μία σειρά μεθοδολογιών έχει

εφαρμοστεί στο παρελθόν, όπως η προγραμματισμένη δειγματοληψία [13] παρέχοντας βαθμολογίες αλληλουχίας που αφορούν συγκεκριμένες εργασίες. Ένα πρόβλημα με την εφαρμογή του GAN δικτύου στο κείμενο είναι ότι τα gradients του δικτύου της γεννήτριας (generator) δεν μπορούν να προωθήσουν σωστά τις διακριτές μεταβλητές. Στο [269], το πρόβλημα αυτό επιλύεται κάνοντας την πρόβλεψη της λέξης ομαλοποιώντας κάθε φορά τις λέξεις στο χώρο ενσωμάτωσης λέξεων. Επιπλέον, οι Yu et al. [262] πρότειναν την παράκαμψη αυτού του προβλήματος μοντελοποιώντας τη γεννήτρια ως stochastic policy. Στην περίπτωση αυτή το σήμα ανταμοιβής προέρχεται από τον διευκρινιστή GAN. Για την ανάλυση του κειμένου, είναι δυνατό να δημιουργηθούν δεδομένα εκπαίδευσης από ένα καθορισμένο σύνολο γραμματικών και έπειτα να αξιολογηθούν τα παραγωγικά μοντέλα με βάση το εάν (ή πόσο καλά) τα παραγόμενα δείγματα συμφωνούν με την προκαθορισμένη γραμματική [225]. Θεωρώντας ότι είναι δύσκολο να γίνει μια καλή αξιολόγηση για τη δημιουργία κειμένου δεδομένου ότι δεν υπάρχει αντικειμενικός τρόπος να εκτιμηθεί κατά πόσο μία “τεχνητή” πρόταση είναι περισσότερο plausible ή ρεαλιστική από ό,τι μία άλλη, μερικές ερευνητικές προσπάθειες έχουν χρησιμοποιήσει τα BLEU ποσοστά δειγμάτων σε μεγάλο αριθμό αθέατων δεδομένων δοκιμών. Η ικανότητα παραγωγής παρόμοιων προτάσεων σε μη ορατά πραγματικά δεδομένα θεωρείται ένδειξη για τη μέτρηση ποιότητας [262].

Συγκεκριμένα, όσον αφορά στη δημιουργία διαδοχικών δεδομένων με GANs δίκτυα, μια εναλλακτική προσέγγιση που βασίζεται στη μάθηση ενίσχυσης χρησιμοποιήθηκε για την εκπαίδευση του δικτύου GAN. Γνωρίζουμε μόνο μία εργασία που χρησιμοποιεί GAN για τη δημιουργία συνεχόμενων ακολουθιών, στόχος της οποίας ήταν η παραγωγή πολυφωνικής μουσικής χρησιμοποιώντας ένα GAN με έναν LSTM διευκρινιστή [150]. Μία ερευνητική δουλειά αναφορικά με την παραγωγή ελεγχόμενου κειμένου [101] χρησιμοποίησε τον μεταβλητό αυτόματο κωδικοποιητή (VAE) μαζί με τις ελεγχόμενες πληροφορίες δημιουργώντας προτάσεις κατηγορίας. Τέλος, οι Zhang et al. [269] και οι Semeniuta et al. [211] χρησιμοποίησαν το δίκτυο GAN για την παραγωγή κειμένου επιτυγχάνοντας state-of-the-art αποτελέσματα. Τέλος, μια άλλη κατηγορία προσέγγισης είναι το δίκτυο conditional GAN [149], το οποίο μας επιτρέπει να κατευθύνουμε τα δεδομένα κατά τη διαδικασία παραγωγής. Αυτή η προσέγγιση χρησιμοποιήθηκε κυρίως για τη δημιουργία εικόνων [181, 4]. Πρόσφατα, οι αρχιτεκτονικές conditional GAN έχουν επίσης χρησιμοποιηθεί στην επεξεργασία κειμένου, συμπεριλαμβανομένης της μετάφρασης [262] και της παραγωγής διαλόγου [129].

Έχοντας συνοψίσει την εργασία που διεξήχθη στον τομέα των GAN για τις εφαρμογές Φυσικής Γλώσσας, παρατηρούμε ότι παρόλο που η ερευνητική προσπάθεια έχει μετατοπιστεί στις προσεγγίσεις GAN για εφαρμογές της Φυσικής Γλώσσας, των οποίων τα μοντέλα του διευκρινιστή και γεννήτριας είναι κυρίως CNN και/ή RNN (BLSTMs), παρόλα αυτά, δεν έχει επιλυθεί μία σειρά περιορισμών.

Έχοντας αυτό υπόψιν μας, προσπαθούμε να ερευνήσουμε κατά πόσο τα GANs θεωρούνται μια κατάλληλη επιλογή μοντέλου για να μάθουμε αναπαραστάσεις της Φυσι-

κής Γλώσσας σε μία μη επιβλεπόμενη ρύθμιση σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών. Επιπλέον, επανεξετάζουμε το παραδοσιακό πλαίσιο GAN από μία εναλλακτική προσέγγιση, αυτή της energy-based. Συγκεκριμένα, προτείνουμε μία αρχιτεκτονική νευρωνικού δικτύου που βασίζεται σε μια παραλλαγή της energy-based GAN [270] για την ανταγωνιστικά παραγωγική εκπαίδευση του δικτύου. Η συμβολή μας βασίζεται στη χρήση μιας απλής hinge loss, στο σημείο όπου το σύστημα συγκλίνει, έτσι ώστε η γεννήτρια του energy-based GAN δικτύου να παράγει σημεία που να ακολουθούν την υποκειμένη κατανομή δεδομένων. Προτείνουμε λοιπόν να χρησιμοποιήσουμε μία αρχιτεκτονική αυτόματου κωδικοποιητή ως διευκρινιστή στην οποία η ενέργεια είναι το σφάλμα ανοικοδόμησης. Η πειραματική επιλογή του σχεδιασμού μας βασίζεται στην προσπάθειά μας να αποκτήσουμε το ζευγάρι των μοντέλων το οποίο να συγκλίνει [263] και το οποίο θα επιδεικνύει μία περισσότερο σταθερή συμπεριφορά από τα κανονικά GAN δίκτυα κατά τη διάρκεια της εκπαίδευσής του [197].

Συγκεκριμένα, προσπαθήσαμε να παρουσιάσουμε την προτεινόμενη προσέγγισή μας επισημαίνοντας τους περιορισμούς των σημερινών energy-based μοντέλων ενώ συγχρόνως παραθέτουμε τα πλεονεκτήματα των δικτύων GANs. Τα GANs έχουν μεγάλη σημασία για την ανάπτυξη των γενετικών μοντέλων. Δεδομένου ότι αποτελούν μία ισχυρή κατηγορία τους, τα GANs επιλύουν το πρόβλημα της δημιουργίας δεδομένων τα οποία μπορούν να ερμηνευτούν με φυσικό τρόπο. Ειδικότερα, για τη δημιουργία μεγάλης διάστασης δεδομένων, η δομή του νευρωνικού δικτύου που υιοθετείται δεν περιορίζει τη διάσταση της παραγωγής, χάρη στην οποία διευρύνεται σημαντικά το πεδίο των παραγόμενων δειγμάτων. Επιπρόσθετα, η δομή του νευρωνικού δικτύου μπορεί να ενσωματώσει διάφορες loss functions, αυξάνοντας έτσι τον βαθμό ελευθερίας του σχεδιασμού του μοντέλου. Όσον αφορά στα energy-based μοντέλα, έχουν χρησιμοποιηθεί για να απομονώσουν τις εξαρτήσεις έναντι των μεταβλητών καθορίζοντας μία energy-based συνάρτηση. Η συνάρτηση της ενέργειας συνδέει κάθε διαμόρφωση των μεταβλητών με μία κλίμακα ενεργειακής τιμής. Έτσι, οι χαμηλότερες τιμές ενέργειας ανατίθενται σε περισσότερα πιθανές διαμορφώσεις. Η λειτουργία αυτή έχει χρησιμοποιηθεί, για παράδειγμα, για την εκτίμηση της κατανομής πιθανότητας σύμφωνα με την κατανομή Boltzmann η οποία ορίζεται από μία ενεργειακή συνάρτηση και έναν κατάλληλο συντελεστή κανονικοποίησης. Στην περίπτωση αυτή, η ενεργειακή συνάρτηση ορίζεται για να εκχωρήσει μία τιμή πιθανότητας που δεν είναι κανονικοποιημένη. Ο συντελεστής κανονικοποίησης διαδραματίζει ένα σημαντικό ρόλο περιορίζοντας την ενεργειακή συνάρτηση για τη σωστή εκτίμηση της κατανομής πιθανότητας. Παρόλα αυτά, εισάγει δυσκολίες κατά τη διάρκεια της διαδικασίας εκπαίδευσης, η οποία απαιτεί κατάλληλο αριθμό δειγμάτων και επιβραδύνει την πρόοδο της εκπαίδευσης είτε απαιτεί ορισμένες δομές μοντέλου για να πάρει τα δείγματα. Για να ξεπεραστούν αυτοί οι περιορισμοί, παρουσιάζουμε στην ακόλουθη ενότητα την προτεινόμενη αρχιτεκτονική του νευρωνικού μας δικτύου.

Προτείνουμε ένα μοντέλο του οποίου ο διευκρινιστής θεωρείται ως μια συνάρτηση ενέργειας που αποδίδει χαμηλά ποσοστά ενέργειας προς τις περιοχές κοντά στα πολλα-

πλά δεδομένα και τις υψηλότερες ενέργειες σε άλλες περιοχές, ενώ η γεννήτρια θεωρούμε ότι έχει εκπαιδευτεί να παράγει αντίθετα (contrastive) δείγματα με ελάχιστες ενέργειες. Με τον όρο αντίθετα (contrastive) δείγματα, αναφερόμαστε σε ένα σημείο δεδομένων το οποίο προκαλεί μια έλξη ενέργειας, όπως για παράδειγμα μία λανθασμένη επισημείωση στην επιβλεπόμενη μάθηση και το οποίο υποδεικνύει από περιοχές χαμηλής πυκνότητας δεδομένων σε μία μη επιβλεπόμενη μάθηση. Στόχος μας είναι να εκπαιδεύσουμε τον διευκρινιστή να εκχωρήσει σε υψηλές ενέργειες σε αυτά τα παραγόμενα δείγματα. Βλέποντας τον διευκρινιστή ως μία συνάρτηση ενέργειας, το προτεινόμενο μας σύστημα, μας επιτρέπει να χρησιμοποιήσουμε μία μεγάλη ποικιλία αρχιτεκτονικών και συναρτήσεων απώλειας-εκτός από τον συνηθισμένο δυαδικό ταξινομητή με λογικής απόδοσης (logistic output) - που έχουν εισαχθεί στο προσφάτως προταθέν μοντέλο “Energy-based ανταγωνιστικό παραγωγικό δίκτυο” (EBGAN) [270]. Ωστόσο, το μοντέλο που προτείνουμε είναι μια τροποποιημένη έκδοση του EBGAN μοντέλου καθώς προτείνουμε τη χρήση μιας αρχιτεκτονικής autoencoder, με την ενέργεια να είναι το σφάλμα ανοικοδόμησης, στη θέση του διευκρινιστή. Συγκεκριμένα, προτείνουμε αντί να χρησιμοποιούμε ένα μόνο κομμάτι πληροφοριών στόχου για να εκπαιδεύσουμε το μοντέλο, να χρησιμοποιήσουμε τους διαφορετικούς στόχους που προσφέρονται από το αποτέλεσμα της διαδικασίας της ανακατασκευής (reconstruction-based output) για το διευκρινιστή. Με τη δυαδική logistic loss, μόνο δύο στόχοι είναι δυνατοί, έτσι, μέσα σε μια μίνι-παρτίδα (mini-batch), οι κλίσεις (gradients) που αντιστοιχούν σε διαφορετικά δείγματα είναι πιθανότατα πολύ μακριά από ορθογώνια. Αυτό οδηγεί σε ανεπαρκή εκπαίδευση και μείωση των μεγεθών της μίνι-παρτίδας (mini-batch), η οποία συχνά δεν αποτελεί επιλογή για το hardware. Από την άλλη πλευρά, η ανασυγκρότηση της απώλειας (reconstruction loss) πιθανόν να παράγει πολύ διαφορετικές κατευθύνσεις κλίσης μέσα στη μίνι-παρτίδα, επιτρέποντας μεγαλύτερο μέγεθος μίνι-παρτίδας χωρίς απώλεια απόδοσης. Επιπλέον, αποφασίσαμε να χρησιμοποιήσουμε τους αυτόματους κωδικοποιητές δεδομένου ότι παραδοσιακά έχουν χρησιμοποιηθεί για να αντιπροσωπεύουν μοντέλα με βάση την ενέργεια. Όταν εκπαιδεύονται με regularization terms, οι αυτόματοι κωδικοποιητές έχουν τη δυνατότητα να μάθουν μια πολλαπλή ενέργεια χωρίς επίβλεψη ή αρνητικά παραδείγματα. Αυτό σημαίνει ότι ακόμα και όταν πρόκειται για energy-based μοντέλο αυτόματης κωδικοποίησης εκπαιδευμένο για να ανοικοδομήσει ένα πραγματικό δείγμα, ο διευκρινιστής συμβάλλει στην ανακάλυψη πολλαπλών δεδομένων από το ίδιο το δίκτυο. Αντίθετα, χωρίς την παρουσία αρνητικών παραδειγμάτων από τη γεννήτρια, η λειτουργία του διευκρινιστή ο οποίος εκπαιδεύεται με δυαδική λογική απώλειας κρίνεται άσκοπη.

Το προτεινόμενο μοντέλο μας εμπνέεται επίσης και από το έργο των Kim και του Bengio [114]. Ωστόσο, διαφέρει με τον ακόλουθο τρόπο. Η προσέγγισή τους χρησιμοποιεί ένα τυποποιημένο πιθανοτικό GAN σε ένα μοντέλο ενέργειας (χρησιμοποιώντας κατανομές Gibbs), επιτρέποντας την εκπαίδευση του διευκρινιστή να μοντελοποιήσει την κατανομή όταν η ισορροπία Nash επιτυγχάνεται, αλλά εξακολουθεί να έχει προβλήματα ανάμειξης (mixing problems), ειδικά με βαθιά μοντέλα που επιβραδύνουν τη εκπαίδευση.

Αντίθετα, η προσέγγισή μας είναι απαλλαγμένη από το πιθανοτικό πλαίσιο, ενώ παρουσιάζει την ίδια ισορροπία Nash ως ένα παραδοσιακό μοντέλο GAN, αλλά μέσω μιας διαφορετικής και περισσότερο γενικευμένης κατηγορίας των συναρτήσεων απώλειας (loss functions).

Τα Ανταγωνιστικά δίκτυα είναι ιδιαίτερος δημοφιλή στον τομέα της Όρασης υπολογιστών [83, 69], παρόλα αυτά όμως, δεν έχουν συγκεντρώσει επιτυχή αποτελέσματα όταν αυτά εφαρμόζονται σε διεργασίες SA. Η παρούσα διατριβή προτείνει ένα τρόπο επίλυσης και προσέγγισης του προβλήματος της εκμάθησης αναπαράστασης σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών με τη χρήση παραγωγικά ανταγωνιστικών δικτύων με μη επιβλεπόμενο τρόπο.

Ιδιαίτερα, η διατριβή μας προτείνει μια νέα επέκταση των γενετικά ανταγωνιστικών δικτύων που αντικαθιστούν τον παραδοσιακό δυαδικό διευκρινιστή ταξινομητή με αυτόν που αποδίδει μια κλιμακωτή ενέργεια σε κάθε σημείο στην περιοχή εξόδου της γεννήτριας. Ο διευκρινιστής ελαχιστοποιεί την απώλεια hinge loss, ενώ η γεννήτρια επιχειρεί να παράγει δείγματα με χαμηλή ενέργεια κάτω από το διευκρινιστή. Δείχνουμε ότι μια ισορροπία Nash κάτω από αυτές τις συνθήκες οδηγεί σε μια γεννήτρια που ταιριάζει με την κατανομή των δεδομένων (υποθέτοντας άπειρη χωρητικότητα). Τα πειράματα που διεξήχθησαν με το διευκρινιστή που λαμβάνει τη μορφή ενός autoencoder και περιλαμβάνουν προαιρετικά έναν ρυθμιστή ο οποίος τιμωρεί τα παραγόμενα δείγματα που έχουν υψηλή ομοιότητα συνημίτονου με άλλα δείγματα στην παρτίδα mini-batch. Τέλος, οπτικοποίησαμε την αναπαράσταση δεδομένων και των τριών αρχιτεκτονικών, προβάλλοντας τα χαρακτηριστικά στα επίπεδα των προτεινόμενων αρχιτεκτονικών νευρωνικών δικτύων κειμένων, προτάσεων και χαρακτηριστικών για την Ανάλυση Συναισθήματος αντιστοίχως.

Υπάρχουν αρκετά ελπιδοφόρα μονοπάτια για μελλοντική επέκταση αυτής της ερευνητικής προσπάθειας που υπογραμμίζονται από τις προτεινόμενες αρχιτεκτονικές νευρωνικών δικτύων σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών για τον τομέα της Ανάλυσης Συναισθήματος από δεδομένα κειμένου. Λαμβάνοντας υπόψιν ότι, γενικά, η δομή του νευρικού δικτύου των μοντέλων GANs μπορεί να ενσωματώσει διάφορες λειτουργίες απώλειας, σκοπεύουμε να αυξήσουμε τον βαθμό ελευθερίας του σχεδιασμού του προτεινόμενου μοντέλου μας. Ιδιαίτερα, επειδή η προτεινόμενη προσέγγισή μας είναι μια εναλλακτική ενεργειακή οπτική γωνία του πλαισίου GAN, θα ήταν ενδιαφέρον να ενσωματωθεί η οικογένεια ενεργειακών λειτουργικών απωλειών που παρουσιάζονται στο [124] αναφορικά με τα ενεργειακά μοντέλα GAN. Έτσι, η υπό όρους ρύθμιση που παρουσιάζεται στο [54] είναι μία πολλά υποσχόμενη προοπτική για να εξερευνήσουμε αναμένοντας ότι θα προσελκύσει μεγαλύτερη προσοχή σε μια ευρύτερη ομάδα των GANs από ενεργειακής άποψης. Επιπλέον, λαμβάνοντας υπόψιν ότι τα GANs είναι επίσης σημαντικά και χρήσιμα για ημι-επιβλεπόμενη μάθηση, σκοπεύουμε να χρησιμοποιήσουμε τη διαδικασία εκπαίδευσης των GANs για την απόκτηση pretrained δεδομένων χρησιμοποιώντας μη επισημειωμένα δεδομένα. Πιο συγκεκριμένα, στοχεύουμε πρώτα στη χρήση μεγάλου

αριθμού μη επισημειωμένων δεδομένων για την εκπαίδευση των προτεινόμενων προτύπων GANs και με βάση την κατανόηση των εκπαιδευμένων GANs σε μη επισημειωμένα δεδομένα, σκοπεύουμε στη συνέχεια να χρησιμοποιήσουμε ένα μικρό σύνολο δεδομένων με ετικέτα ώστε να εκπαιδύσουμε το μοντέλο διάκρισης τόσο για εργασίες ταξινόμησης όσο και παλινδρόμησης. Τέλος, τα μελλοντικά μας σχέδια περιλαμβάνουν την κατανόηση του γιατί ο αυτόνομος κωδικοποιητής (autoencoder) στο μοντέλο του διευκρινιστή στο δίκτυο GAN φαίνεται να παράγει σημαντικά καλύτερες αναπαραστάσεις από έναν τυπικό αυτόματο κωδικοποιητή; αποτέλεσμα το οποίο αποδεικνύεται από την πειραματική αξιολόγηση.

EXTENDED ABSTRACT

The last 20 years has witnessed a number of important efforts aimed at technologies for modeling, analysis and synthesis of human-human and human-computer interactions (HCI). Due to the fact that human and computer existence have become extremely interwoven, recent developments in that research area have already shifted from HCI to intelligent HCI (iHCI), moving from traditional keyboard and mouse to natural humanlike interactive functions including understanding certain human behaviors such as affective and social signals. In other words, by approaching HCI towards the HCI settings that are based on touch, gesture and movement (e.g. Microsoft Kinect), we have the ability to detect subtleties and changes in the human's communicative behavior and thus to initiate interactions based on this information. Hence, the result is an increasingly growing state-of-the-art in domains such as Affective Computing (AC), social signal processing, HCI [58], gaming [199], mental health [174], learning technologies [131] etc. However, relatively little attention was paid to the potential impact these technologies can achieve when the contextual aspect is incorporated in human affective behavior analysis systems.

Incorporating context in affective human behavior analysis lies at the intersection of context aware affective computing systems and affective aware intelligent human computer interaction systems since contextual information cannot be discounted in doing automatic analysis of human affective behavior. The contributions of context aware affective computing systems were demonstrated during the two recently organized workshops on Context-based Affect Recognition (CBAR 2012 and CBAR 2013 [94]), held in conjunction with SocialCom2012 and ACII 2013 respectively. More specifically, the CBAR 2013 workshop was one of the shortest workshops in ACII 2013 with the most interesting keynotes (Schuller, Gratch) which is indicative of the research area's status. These prominent figures in AC research community provided a valuable source of information discussing about learning context in affect recognition and how context shapes how people interpret the expressions of people and machines. Thus, by tackling the issues of context based affect analysis, i.e. careful study of contextual information and its relevance in context aware systems, its representation, its modeling and incorporation including its effect on the performance of existing affect analysis methods, it is possible to outline a roadmap showing the most important steps still to be made towards real-world affect analysis.

According to the first work which introduced the term context awareness in CS, [207] the important aspects of context are: Who you are with, Where you are, When, What

resources are nearby. Thus, context aware systems look at the Who, Where, When and What (the human is doing) entities and use this information to determine Why the situation is occurring. In a similar definition, in [24] authors define context as location, identities of the people around the user, the time of day, season, temperature, etc. Other approaches [194] include context as the user's location, environment, identity and time while others [23, 73] have simply provided synonyms for context. For a more extended overview on context-awareness the reader is referred to [2].

To fulfill the need of incorporating context in multiparty human communicative behavior, the context model which was formalized as the combination of the "Identity", "Time", "Location" and "Activity" contextual types presented in [2] was enriched [271] to include an additional contextual type called "Relations", referring to information about any possible relation that a person may establish with others during an interaction (multiparty). The "Relations" type expresses a dependency among the interlocutors that emerges and also acts as indices to other sources of contextual information. For example, given a person's affiliation, related information such as social associations, connections, information about friends, enemies, neighbors, co-workers, relatives, etc. can be acquired. Recently the term Relations has been used to refer to the relation between the individual and the social context in terms of perceived involvement [17] and to the changes detected in a group's involvement in a multiparty interaction [16]. Thus, it is clear that understanding the multiparty human communicative behavior implies understanding the modifications of the social structure and dynamics of small [78] and large groups (friends, colleagues, families, students, etc.) [38] and the changes in individuals' behaviors and attitudes that occur because of their membership in social and situational settings [258].

As far as real-world, context-aware affective computing frameworks is concerned, context is defined as any information that can be used to characterize the situation that is relevant to the interaction between users and the system, with the definition of [60] approaching better the understanding of human affect signals in AC. More specifically, this work [60] summarizes appropriately the key aspects of context with respect to the human communicative behavior in terms of the context-aware formalization: the human-human multiparty interactions, the non-linguistic conversational signal or emotion that is communicated, how the information is communicated (the person's facial expression, head movement, tone of voice, and hand and body gestures), the context under which the information is passed on: where the user is, what his current task is, and how he/she feels and the (re)action that should be taken to satisfy user's needs and requirements. Thus, so far the efforts on human multiparty behavior understanding are usually context independent [169] while the above presented context-aware methodology answers one or more questions separately each time [9, 229]. Overall, further research is needed in answering the above context-aware aspects simultaneously and propose a context formalization framework.

Following context integration on affect production and context incorporation in emotion

corpora, this section reviews context awareness in automatic affect analysis systems, both single-modal and multimodal. In [125] a unimodal context aware affect analysis system for short-term context modeling in dyadic interactions is proposed, where context is defined as the speech cues from the past utterance of the speaker, while in [86] a multimodal phoneme recognition system used Bidirectional Long Short Term Memory (BLSTM) networks to incorporate arbitrarily large amount contextual data from past and future contextual information. Overall, in terms of emotion analysis approaches, Bidirectional context-sensitive approaches proved to outperform methods that do not consider context from both past and future observations such as Recurrent Neural Networks (RNNs) and Hidden Markov Models (HMMs). However, such conclusions depend on the design of the corpus and may not cover the entire spectrum of human affectively enhanced communication in terms of interaction modeling.

Moreover, in terms of visual signals, the interpretation of the observed facial expression [108] has been also attempted in the past. Facial expressions are accordingly displayed in a particular context, such as location (outdoor, indoor), situation (driving a car or being treated in a hospital), the undergoing task, other people involved, the personality of the expresser [167, 266]. However, to the best of our knowledge, no vision-based model takes into account the context of the application for spontaneous expression analysis. A recent exemption is discussed in the work of [92] where the Transferable Belief Model (TBM) is used to easily add one or more context variables in the model of facial expressions classification for the pain analysis application. In this work, several contextual variables can be defined (the place, the task, the answer to a writing question, etc.) to reduce the set of the expected facial expressions. At the same time only “the place” as a contextual variable is used and takes two values depending on whether the expresser is in the hospital or not in order to identify if the videotaped expression is painful or not. Furthermore, the former work consists of an extension of [93], where a context aware clinical system related to pain analysis is described, where the contextual variables are defined as the place, the task, etc.

Incorporating context into Affective Computing is still at an early stage. Despite the growing interest in the intermediate stages of human-machine communication (HCI), their relevance with respect to its integration into the elicitation, interpretation and recognition stages has not yet been examined. These stages include a series of challenges such as the modeling of context aware cognitive theories based on emotional interaction, the extraction of information, the integration of the contextual information into context aware corpora, the identification of contextual types and the their integration to improve the performance of multimodal applications.

Consequently, the aforementioned issues are closely related to these situations and therefore it is imperative that they will taken into account in the process of extracting contextual parameters and during emotion analysis. For these reasons as we have mentioned above, further research is needed to answer the context related questions of the **W5+**

formalization. Under these circumstances, we could have a first insight of whether it is necessary to annotate the corpora in accordance with the dependent context, and whether existing corpora have to be re-annotated or recorded from scratch.

With respect to the annotation process, the challenges of achieving this goal start with the implementation of a general framework for collecting spontaneous behavioral data. The research field is still in its early stages, and no significant efforts have been made so far to collect spontaneous framework-specific data specifically designed for contextual types. Most research papers so far are based on the use of already-labeled corpora, or they are labeled every time to fulfill the needs of the current experiment.

At the same time, during a group interaction there is a plethora of different “cues” and so far there is no an agreed annotation template or a data collection protocol that is easily applicable. In addition, the combination of the perspective of the event evaluation could provide a range of benefits such as improved recognition of mixed emotions, the ability to incorporate relevant information into automated emotional analysis architectures, and a better level of interpretation of affect. Therefore, without much effort to collect and label data, it will be difficult, if not impossible, to address the current research problems in the Affective Computing research field.

Thus, our contributions provide comprehensive and thorough answers to the currently unaddressed questions posed above, and fully explore the terrain of context in Affective Computing (both multimodal and unimodal) shedding light in each of its subcomponents. More specifically, exploring the terrain of the modeling of the cognitive theories of context based on affective interaction, the extraction of context information, the incorporation of contextual information in affective corpora focusing on how it should be represented, what contextual information is relevant, as well as how the integration of contextual information, would improve the performance of multimodal frameworks. Such knowledge is expected to enable several exciting directions for further investigation. It could also enable technologies such as context based and affect-aware intelligent tutors, human-embodied conversational agent interactions, independent living and personal wellness technologies, broadcast video news technologies, face recognition in personal photos in the wild, face recognition systems that utilize body and clothing, and educational tools. Moreover, in the area of arts various applications, using the term context, range from analysis of aesthetics, to arts installations. Similarly, in the area of clinical applications context aware technologies such as depression severity detection, stress/pain monitoring, emotion related disorders such as autism etc. could be developed. Such technologies would have large impact in domains such as arts installations and entertainment, education (e.g. gaming applications) and learning styles and healthcare applications.

To be more precise, in the area of multimodal frameworks, we study the understanding of the relationship between the external behavior and inner affective state of two team members (“instructor”-“defuser”) during a demanding operational task (i.e., bomb defusion). In this study we assessed team member’s verbal responses (i.e., length of duration) in

relation to their external as well as internal affective cues. External behavioral cues include defuser's verbal expressions while inner cues are based on physiological signals. More specifically, we differentiate between "defusers'" physiological patterns occurring after the "instructor's" turns according to whether they belong to a short or a long turn-taking response interval. Based on the assumption that longer turn-taking behaviors are likely to be caused by demanding cognitive task events and/or stressful interactions, we hypothesize that inner mechanisms produced in these intense affective activity intervals will be reflected on defuser's physiology. A dyadic team corpus was used to examine the association between the "defusers'" physiological signals following the "instructor's" questions to predict whether they occurred in a short or long turn-taking period of time. The results suggest that an association does exist between turn taking and inner affective state. Additionally, it was our goal to further unpack this association by creating diverse ensembles. As such, we studied various base learners and different ensemble sizes to determine the best approach towards building a stable diverse ensemble that generalizes well on the external and inner cues of individuals.

Lack of emotional expressivity is one of the main deficits that characterizes periods of stress when team members perform highly cognitive cooperative tasks. In cases such as this, teammates may find it more difficult to express their conscious feelings and show different patterns in perceiving and conveying emotional information when working together to meet a common goal [105, 178]. In light of these observations, having a way to monitor the internal state of teammates within such contexts might provide us new insights with respect to the mechanisms of their interaction and affectivity.

During high workload and high stress tasks, the sympathetic nervous system is accountable for activating glands and organs that are responsible for defending the body from perceived threats. This activation is associated with changes in arousal that are further influenced by emotion, cognition or attention. Stress results in increased sympathetic activity and can be tracked for example through bodily reactions, such as an increase in heart rate, greater blood flow to extremities and an increase in the respiration rate etc. Thus, a combination of more than one physiological indicator would be considered a more sensitive measure of changes in stress and can be used to provide estimations of emotion, arousal and general cognition [142].

We attempt to shed light on the association between two team members' physiological states and their speech, measured via their conversational turn-taking duration. Team members in highly-demanding operational tasks do not often notice triggers that cause them to be emotionally and mentally stressed [153, 222]. Thus, they might communicate with their teammates and express their emotions in ways that may not be noticed in observable audio-visual cues. For instance, one would expect that asking an individual to disarm a simulated bomb would result in high levels of (internal) stress. This inherent gap between teammates' external observable behavior and their inner affective state is not well understood and can be potentially bridged by monitoring their physiology. The duration of response utterances

is also reported to be very important, as it can be indicative of conflicting mental and stress procedures [183]. Because physiological indicators reflect aspects of underlying mental states and specifically the amount of distress [63], we explore whether physiological signals of long and short response utterance duration exhibit different physiological patterns. To further capture and interpret this ongoing and evolving interplay, we examine the use of two ensemble learning strategies. We believe that the investigation of physiological changes during the response periods can provide a better understanding of a team dynamics.

There has been a lot of research on dialogue dynamics as well as the relationship of stress to underlying physiology, but these fields have largely been separate in the literature. For example, work on turn-taking behavior in dialogue systems [183] could benefit from an understand of interpersonal dynamics of physiological stress response during a cooperative task [53]. Links between turn-taking behavioral responses and physiology have been studied for assessing how adults' anger levels affected children [63]. Moreover, previous studies have shown the advantages of using ensemble learning in both unimodal [208, 206, 202] and multimodal behavioral analysis[81, 201].

To the best of our knowledge, there is no experimental evidence of applying ensemble learning to study the link between external behavior of turn-taking responses and inner affective states inferred from physiological signal indicators of two teammates (instructor-defuser) trying to disarm a simulated bomb [157]. This section of the presented Ph.D. thesis is an effort to unfold this association based on the experimental evidence from the Dyadic Team Corpus. Our results indicate that physiological patterns convey information about the defuser's inner state, because they differ according to the duration of turn-taking behavioral replies with respect to the instructor's turns. Finally, our results are further enhanced through ensemble learning methods, which outperform the individual base learners in most cases and interesting observations are discussed.

One of the main indicators of an interactional speech episode is often called a "turn-taking" and is defined as the time duration between the end of someone's turn and the beginning of the other interlocutor's corresponding turn. Turn-taking responses may span from very short to very long, which may indicate shorter or longer emotional and stressful episodes. In a similar way, in our teammate corpus, longer turn taking behavioral responses provided valuable information about the defuser's perceived cognition and affective state, reflected their external observable as well as implicit inner affective states. We choose to investigate that type of interactional context between the two teammates, motivated by the fact that the instructor's behavior is more controllable, thus minimizing the effect of the instructor's variability on the defuser's behavior.

To further distinguish between short and long turn-taking behavioral responses we drew a threshold at the 70th percentile of response values. This threshold was computed empirically after plotting the histograms of turn-taking behavioral response instances from the data of each defuser separately. Negative values of this measure mean that the defuser started talking before the instructor had finished the current turn. Phenomena such as

overlapped speech and very short utterances are aligned with high levels of stress in highly-demanding operational tasks [97].

After carefully inspecting various turn-taking behavioral instances in our corpus, we came across a number of interesting tendencies. There were examples during the interactional context in which the instructor explained how the blue and red wires are connected. In that case, the defuser's reply is short (i.e., the defuser uses words such as ok/yes/no). Then, as a follow up, the instructor explained with more detail the process of bomb defusion to check whether the defuser is really following his instructions. In the first case, where the statement is simple and elicits low cognitive effort, a short reply occurred, while a long one occurred in the second case, where the defuser repeated the instructor's guidelines to confirm that he correctly understood the task. In this case, it was expected that the defuser was much more mentally alert.

As far as the extracted physiological features, a BIOPAC MP150, with a standard lead II electrode configuration was used to record electrocardiography (ECG), real-time changes in blood pressure, and impedance cardiography (ZKG). Continuous data were recorded for each participant throughout each task and analyzed offline. The raw time series for each task segmented into thirty second intervals relative to the end of each session, such that a few seconds from the beginning of the "bomb defusion" task were cut out. This was done because a minimum of 30 seconds of cardiovascular data are necessary for further analysis. The Moving Ensemble Average Program (MEAP)[39] was used to extract features from the data by computing an ensembled average over each epoch.

Particularly, we extracted 24 cardiovascular features. Some of these features included heart rate (HR), LVET (left ventral (systolic) ejection time), p_time, s_time, t_time, x_time, systole_time, pre-ejection period (PEP), ventricular contractility (VC), cardiac output (CO) and total peripheral resistance (TPR). PEP is the time from the onset of the heart muscle depolarization to the opening of the aortic valve. When PEP decreases, VC increases. VC has been shown to be related to task engagement [158, 188, 221, 210]. CO is the amount of blood pumped in liters per minute. TPR reflects vasodilation (more blood flow) and vasoconstriction (less blow flow), which are related to parasympathetic and sympathetic activity, respectively. Prior work has shown that TPR unambiguously increases when an individual is in a threat state and decreases in a challenge state, whereas CO either remains unchanged or decreases in a threat state and increases in a challenge state [233].

We further observed that there is a wide variability across defusers with respect to the given task. This observation indicates that there might be mechanisms triggered in defusers with high learning accuracy, reflected their physiological signals, which are not present in defusers with low learning performance (i.e. P101, P103, P201). It is also noteworthy that for these three defusers, the selected physiological feature is HR. To further elaborate on this tendency, we go through the audiovisual recordings and the HR signals. We notice that there is a difference in the arousal levels (i.e., stress) with respect to the type of behavioral replies (short/long) and that arousal affectivity is present both in short and long turn-taking

responses, depending on the defuser.

More specifically, we came across examples of defusers who took a long time to respond after having given a wrong answer once and were asked to try again to confirm the bomb defusion steps. Hence, it appears that the task was a sufficiently stressful stimulus for them. In these long turn-taking examples, it is also reasonable to assume that high cognitive activity or stressor events occurred. At the same time, high levels of arousal are noticed in short turn-taking examples, in which for example the defuser uses words such ok/yes/no. This tendency is not aligned with the “bomb defusion” task, considering that we were expecting that short turn-taking examples would reflect low levels of arousal. On the contrary, our observation suggests that even though there may be no obvious (audible/visible) signals of arousal, physiological signals may provide a complementary, not overlaid though, view of a person’s state. This finding is of particular importance, especially in cognitively demanding tasks in which one of the teammates manipulates the discussion and is also aligned with previous research studies [233, 79, 27].

Finally, we concluded that our study provided an analysis of physiological signals in a dyadic team “bomb defusion” scenario in association with their expressive behavioral cues. The results suggested that physiological responses convey information about the defuser’s inner state. They also reflected the amount of the defuser’s verbal responses with respect to a stimuli and can be further linked with the amount of underlying socio-cognitive activity, which is not always obvious through traditional observational methods. Last, we proposed two existing ensemble learning methods which are new to the field of generalizing external and inner cues of speakers, showing that these methods can yield improvements over traditional analysis methods.

The ability of humans to effectively interact socially relies heavily on their awareness of the context the interaction takes place, i.e. the “bomb defusion” task that we have presented above. In order for computer systems to accordingly possess the same ability, it is crucial they are also context-aware in terms of a formalization of context based on the W5+ framework aspects of Who, What, Why, Where, What and How. Research work presented in this paper contributes towards this goal by bridging the conceptual gap and exploiting semantics and cognitive and affective information of non verbal behavior and investigating whether and how this information could be incorporated in automatic analysis of affective behavior. A semantic concept extraction methodology is proposed and its application to indicative examples from the SEMAINE corpus is presented that validates the proposed approach.

One of the main challenges of recent years is to create more natural, sensitive and socially intelligent machines, that are not able only to communicate but also to understand social signals and make sense of the various social contextual settings [240]. Thus, besides communication through various channels and through verbal content (semantics), machines also need to be able to recognize, interpret, and process emotional information as humans. In human cognition, thinking and feeling are mutually present: emotions are often the

product of our thoughts, as well as our reflections are often the product of our affective states. But, what does it mean to be socially intelligent when incorporating interaction context? So far, in natural conversations context awareness is defined as past visual information [62], general situational understanding [30], past verbal information [119], cultural background [140], gender of the participants, knowledge of the general interaction setting in which an emotional phenomenon is taking place [24], discourse and social situations [16]. Accordingly, studies in intelligent Human-Computer interfaces (iHCI), which incorporate context, correspond to the following contextual aspects, known as W5+ formalization: Who you are with (e.g. dyadic/multiparty interactions [243]), What is communicated (e.g., (non)-linguistic message/conversational signal, and emotion), How the information is communicated (the person's affective cues), Why, i.e., in which context the information is passed on, Where the user is, What his current task is, How he/she feels (has his mood been polarized changing from negative to positive) and which (re)action should be taken to satisfy human's needs, goals and tasks [60].

Unfortunately, so far the efforts on human affective behavior understanding are usually context independent [91]. In light of these observations, understanding the process of a natural progression of context-related questions when people interact in a social environment could provide new insights into the mechanisms of their interaction context and affectivity. The "Who", "What", "Where" context-related questions have been mainly answered either separately or in groups of two or three using the information extracted from multimodal input streams [266]. Thus, as of date, no general W5+ formalization exists, like the systems that answer to most of the "W" questions are founded on different psychological theories of emotion and they all fit specific purposes according to the goals of a particular research in various fields.

Recent research on progressing to the questions of "Why" and "How" has led to the emerging field of sentiment analysis [165, 133, 29], through mining opinions and sentiments from natural language, which involves a deep understanding of semantic rules proper of a language. Furthermore, the interpretation of cognitive and affective information associated with natural language and, hence, further inferring new knowledge and making decisions, in connection with one's social and emotional values and ideals, is of crucial importance. The problem when trying to emulate such cognitive and affective processes, is that while cognitive information is usually objective and unbiased, answering the "Why" context-related question through affective information is rather subjective and argumentative.

Under this view, our long-term goal is to understand whether and how context is incorporated in automatic analysis of human affective behavior and to propose a novel context-aware incorporation framework (Fig. 1) which (1): includes detection and extraction of semantic context concepts, (2): enriches better a number of Psychological Foundations with sentiment values and (3): enhances emotional models with context information and context concept representation in appraisal estimation, using publicly available on-line knowledge sources (OKS) in natural language processing [?]. As a first step in this work,

we focus on bridging the gap at concept level by exploiting semantics cognitive and affective information, associated with the image verbal content (semantics), which for the needs of our research is the contextual interactional information between the user and the operator of the SEMAINE database [143], keeping fixed the “Where” context-related question. This context concept-based annotation method, that we are examining, allows the system to go beyond a mere syntactic analysis of the semantics associated with fixed window sizes³. In most of traditional annotation methods, emotions and contextual information are not always inferred by appraisals and thus contextual information about the causes of that emotion is not taken into account [31].

Emotions are complex states of feeling, resulting in physical and psychological reactions influencing both our thought and behavior. The study of emotions still remains an essential and open part of psychology. Of interest to Natural Language Processing (NLP) is being able to tell which emotion is expressed in the text. Predominantly, research on detecting emotions from text has focused on capturing emotion words based on three **emotion models**, i.e. categories of basic emotions, emotion dimensions and cognitive-appraisal categories, particularly the componential model [28, 90].

Unlike the categorical and dimensional approaches, recently, increasingly attention has been dedicated to another set of psychological models, referred to as componential models of emotion, which are based on the appraisal theory and might be more appropriate for developing context-aware frameworks [151]. However, how to use the appraisal approach for automatic analysis of affect is an open research problem. In the componential models of emotion, various ways of linking automatic emotion analysis and appraisal models of emotion are proposed. This link aims to enable the addition of contextual information into automatic emotion analyzers, and enrich their interpretation capability in terms of a more sensitive and richer representation.

However, these emotional models have some limitations. Categorical approaches usually fail to describe the complex range of emotions that can occur in daily communication. Furthermore, the dimensional space neither allows to compare affect words according to their reciprocal distance, nor models the fact that two or more emotions might be experienced at the same time.

Particularly, a number of 2D-dimensional approaches are mainly used to **visualize Psychological Foundations**. An early example is Russell’s circumplex model [190], which uses the dimensions of arousal and valence to plot 150 affective labels. Similarly, Whissell considers emotions as a continuous 2D space with evaluation and activation as dimensions [253]. Another bi-dimensional model is Plutchik’s wheel of emotions [176], according to which emotions are adaptive as they are based on evolutionary principles, even though we conceive emotions as feeling states. These feeling states are part of a process involving both cognition and behavior and containing several feedback loops. Eventually, all such

³The window length corresponds to 16 conversational turns and is displayed on figures for future visualization purposes.

approaches work at word level, so they are unable to grasp the affective valence of multiple-word concepts.

However, since the above models currently focus on the objective inference of affective information when associated with natural language opinions, appraisal-based emotions are not taken into account. Nevertheless, in view of their suitability to context modeling, emphasis should be given on emotional models based on cognitive appraisal, which characterize emotional states in terms of detailed evaluation of emotions acquisition and especially implicit methods. For an extended overview on modeling affect, the reader is referred to [28, 90].

For a more applicable semantic context concept model, rather than a theoretical one such as the componential model, research has been focused on mining opinions and sentiments from natural language. This is challenging, as it requires a deep understanding of the explicit and implicit and semantic language rules, struggling with NLP's unresolved problems such as negation handling, named-entity recognition, word-sense disambiguation, etc. Concept-based approaches [234] aim to grasp the conceptual and affective information associated with natural language opinions. Additionally, concept-based approaches can analyze multi-word expressions that don't explicitly convey emotion, but are related to concepts doing so. For example, instead of gathering isolated opinions about a whole "event" (e.g. birthday party), users are generally more interested in comparing different events according to their specific set of semantically related concepts, e.g. "cake", "surprised friend", or "gift" (which can be considered as contextual information for improving search results), associated with a set of affectively related concepts, e.g. "celebration" or "special occasion". This taken-for-granted information referring to obvious things people normally know and usually leave uncommented, is necessary to properly deconstruct natural language text into sentiments. For example, the concept "small room" should be appraised as negative for a hotel review and "small queue" as positive for a post office, or the concept "go read the book" as positive for a book review but negative for a movie review.

Gradually, the new multi-disciplinary area that lies at the crossroads between Affective Computing, Human-Computer Interaction (HCI), social sciences, linguistics, psychology and context awareness is distinguishing itself as a separate field. It is thus possible to better recognize, interpret and process opinions and sentiments, incorporate contextual information and finally to understand the related ethical issues about the nature of mind and the creation of emotional machines. For applications in fields such as real-time HCI and big social data analysis [3], deep natural language understanding is not strictly required: a sense of the semantics associated with text and some extra information (affect) associated with such semantics are often sufficient to quickly perform tasks such as emotion recognition and cognitive and affective information detection.

We have illustrated a method for extracting context concept aspects from SEMAINE corpus interaction. The proposed framework only leverages on any taken-for-granted information. By allowing sentiments to flow from multi-word concept to multi-word concept, we could

possibly achieve a better understanding of the contextual role of each concept within the sentence.

As far as the selection of the corpus is concerned, on which the experiments will be performed every time, the new trend is the collection of data in real time through new sources of opinion mining and sentiment analysis which abound. Webcams installed in smartphones, touchpads, or other devices let users post opinions in an audio or audiovisual format rather than in text. Aside from converting spoken language to written text for analysis, the audiovisual format provides an opportunity to mine opinions and sentiment. Many new areas might be useful in opinion mining, such as facial expression, body movement. Affect analysis, a related field, addresses the use of linguistic, acoustic and (potentially) video information. This field focuses on a broader set of emotions or the estimation of continuous emotion primitives; for example, valence can be related to sentiment.

Furthermore, as far as the presence and the position of the multi-word concepts in the text unit, further examination is necessary, as typically bi-grams and tri-grams, are often taken into account as useful features. Some methods also rely on the distance between terms. Part-of-speech (POS) information (nouns, adjectives, adverbs, verbs, etc.) is also commonly exploited in general textual analysis as a basic form of word-sense disambiguation. Certain adjectives, in particular, have been proved to be good indicators of sentiment and sometimes have been used to guide feature selection for sentiment classification. In other works, the detection of sentiments was performed through selected phrases, which were chosen via a number of pre-specified POS patterns, most including an adjective or an adverb. However, such approaches and their performance are strictly bound to the considered domain of application and to the related topics.

Finally, most of the literature on sentiment analysis has focused on text written in English and, consequently, most resources developed, e.g., sentiment lexicons, are in English. Adapting such resources to other languages should be seriously considered as the choice of words and their intended meaning are personally, contextually, culturally and socially dependent and differ on the level of the different expertise and purposes of tagging users, resulting many times in tags that use various levels of abstraction to describe a resource.

Moving from multimodal to unimodal analysis and more precisely in the field of personal interpretations from the text (Sentiment Analysis) and more generally in the field of Natural Language Processing (NLP), with respect to more computational approaches, the contribution of our proposed framework lies in a more effective analysis, interpretation and processing of opinions and sentiments in text which are examined at document, sentence and aspect level respectively. In this case, the semantic, cognitive and emotional information of the subjective responses of individuals are taken into account. Particularly, we investigated a method based on Conditional Random Fields (CRFs) to incorporate sentence structure (syntax and semantics) and context information to identify sentiments of sentences. It also demonstrated the usefulness of the Rhetorical Structure Theory (RST) taking into

consideration the discourse role of text segments. Thus, our contribution aims to reconsider the effectiveness of CRF and RST methods in incorporating the contextual information into Sentiment Analysis systems. Both methods are evaluated on two, different in size and genre of information sources, the Movie Review Dataset and the Finegrained Sentiment Dataset (FSD). Finally, we discuss the lessons learned from these experimental settings with respect to addressing the following key research questions such as whether there is an appropriate type of social media repository to incorporate contextual information, whether extending the pool of the selected features could improve context incorporation into SA systems and which is the best performing feature combination to achieve such improved performance.

Incorporating context information to improve Sentiment Analysis (SA) is an emerging research area due to the clear benefits of context-aware applications, including: the detection of important events in news [100], consumers' opinions on products [113], extraction of the sentiment orientation (i.e. positive or negative) of opinionated text [166], etc. Recently, it has been shown that it can yield competitive advantages for businesses [10], as extracting sentiment at the fine-grained level [66, 185, 67, 186, 6, 184, 77] (e.g. at the sentence- or phrase-level) has received increasing attention due to its challenging nature [165]. It can be thus inferred that, the performance of context-aware SA systems strongly depends on a number of predefined parameters during the experimental setting [68, 165]. Particular interest revolves round the SA when it is performed in the domain of movies and general product reviews [166].

Previous studies have proposed various approaches to increase the robustness of context-aware SA systems: Understanding the sentiment of sentences allows us to summarize online opinions which could make informed decisions. Automated sentiment incorporation has seen great research efforts for many years and has achieved some promising results [165]. On one hand, different machine learning techniques, exploit patterns in vector representations of text and lexicon-based methods [227] account for semantic orientation in individual words, while on the other hand, some researchers have proposed rule-based (and unsupervised) methods to improve SA. Still, even though the current insight is that all of the state-of-the-art algorithms perform well on individual sentences without considering any context information, their accuracy is dramatically lower on the document level, due to the fact that they fail to consider context.

Hence, recent studies on context-aware SA have started to consider methods that provide more control, e.g. [36] uses CRFs to tackle opinion source identification as a sequential tagging task, whereas [103] identifies the target of the opinion with CRF to further incorporate the context information into SA systems. In addition to this research direction, discourse analysis has also been employed to adjust prior polarity of terms. Further improvements were reported when a more sophisticated weighting approach based on RST [138] was introduced.

Therefore, it is time to reconsider the effectiveness of these methods for incorporating

context information into SA. The contribution of this paper is to present an approach of combining CRF and RST methods with SA to effectively analyze context. Particularly, compared to existing algorithms of sentiment analysis both on sentence and document levels, our proposed approach attempts to improve SA by taking full advantage of the sentence structure, by using context information to capture the relationship among sentences and to improve document-level SA, by taking into consideration the Internet language word set and emoticons and finally by extending the pool of the contextual features used with semantic, syntactic, structural and context-aware RST features. So far, a wide range of features has been independently tested by a large number of research teams, mostly in constrained settings.

An important research direction in context-aware SA is improving the robustness of SA systems after incorporating context. The state-of-the-art in automated SA has been reviewed extensively [165, 5]. Existing methods rely on sentiment lexicons, which are enumerative lists of sentiment terms that indicate the sentiment changes. Popular examples include General Inquirer[223], Subjectivity Lexicon and Subjectivity Sense Annotations and SentiWordNet. Additionally, domain knowledge plays a key role, since a sentiment term's linguistic context often impacts its sentiment charge. Early work on SA used syntactic relations to identify new sentiment terms, which can be considered as an early form of context exploitation [95]. Sentiment is often expressed in a subtle manner, which makes it difficult to identify when processing sentences or paragraphs in isolation. Thus, context remains an essential ingredient to further improve SA.

On the whole, research in SA can be organized into two broad methods: Machine Learning and Lexicon-Based. With Machine learning approaches, an algorithm is trained with sentiment labeled data and the learned model is used to classify new documents. This method requires the initial labeled data, which is typically generated through labor-intensive human annotation. As far as the Lexicon-based method concerns, the former involves the extraction and aggregation of terms' sentiment scores offered by a lexicon (i.e. prior polarities) to make sentiment prediction. Nevertheless, it is observed that the main strength of the lexicon-based approaches is at the same time also their weakness. Considering that the lexicons are predefined, they are unable to adapt to novel or domain specific forms of expressions. In addition, lexicon-based approaches do not naturally produce the level of confidence during the analysis, which is automatically provided by machine-learning approaches.

Understanding the sentiment of sentences allows us to summarize opinions which could help people make informed decisions. All of the state-of-the-art algorithms perform well on individual sentences without considering any context information, but their accuracy is dramatically lower on the document level because they fail to consider context. There are many difficulties due to the special characteristics and diversity in sentence structure in the way people express their opinions (e.g. opinions expressed indirectly through comparison etc.). In addition, complicated sentence structure and emoticons make sentiment analysis

even more challenging.

Thus, based on our proposed work, we do not only consider syntax that may influence the sentiment, including newly emerged Internet language, emoticons, positive and negative words and negation words, but also incorporate information about sentence structure, like conjunction words and comparisons, the position of positive and negative words and the context-aware RST features. Therefore, we employ a CRF [121] model to capture syntactic, semantic and contextual features of sentences and the RST to unfold the rhetorical relations for SA.

Conditional Random Fields(CRF) method: CRF is well known for sequence labeling tasks [121]. CRFs are a class of discriminative undirected probabilistic graphical model generally applied in pattern recognition and machine learning, where they are specifically designed to optimize structure prediction. A “generic” classifier predicts a label for a single sample without regarding to “neighboring/connected” samples, however a (linear-chain) CRF can take context into account. Hence, in the case we want to capture the context information (e.g. neighboring sentences or sentences connected by transition words), the procedure of sentiment identification becomes a kind of sequence labeling. Particularly, CRFs provide a probabilistic framework for calculating the probability of Y globally conditioned on X , where X is a random variable/vector over sequence data to be labeled, and Y is a random variable/vector over corresponding label sequences. X and Y could have a natural and/or complicated graph structure. So far, the CRF model, in its simplest form, has been widely used in the text labeling domain [141]. CRF examples given in seminal works such as [121] and [141] assumed linear chain structure as well. A further observation reveals that there is a one-to-one correspondence between states and labels.

Rhetorical Structure method in SA: Within a natural language text, rhetorical relations that hold between parts of the text, unfold and are typically used to distinguish important text segments from less important ones in terms of their contribution to a text’s overall sentiment. Within a discourse structure, the former is an important part of what makes a text coherent. Thus, the analysis of the discourse structure is divided into two tasks: discourse segmentation and discourse parsing. Discourse segmentation is the task of taking a sequence of word and punctuation tokens as input and identifying boundaries where new discourse units begin. Discourse parsing is the task of taking a sequence of discourse units and identifying relationships, such as causality, contrast and specification between them. In our case, the set of these relationships form a tree. The leaves of the Discourse Tree (DT) representation correspond to contiguous atomic text spans, also called Elementary Discourse Units (EDUs). The adjacent EDUs are connected by a rhetorical relation (e.g. elaboration), and the resulting larger text spans are recursively also subject to this relation linking. A span linked by a rhetorical relation can be either a nucleus or a satellite depending on how central the message is to the author.

Previous studies on discourse analysis have been quite successful in identifying what machine learning approaches and what features are more useful for automatic discourse

segmentation and parsing[220]. However, one of the reported downsides of SA guided w.r.t. RST is the high processing time required for analyzing discourse in natural language text [96]. This problem seems to obstruct the applicability of such methods in large-scale scenarios.

We have thoroughly studied the usefulness of the CRF model and the RST discourse theory to incorporate sentence structure and context information into context-aware Sentiment Analysis systems, in two information sources which are different in size and genre. First, we have shown how to improve SA by taking full advantage of the sentences structure, using context information to capture the relationship among sentences and to improve document-level SA. The reason of this success lies in extending the pool of the contextual features used with semantic, syntactic and context-aware RST features rather than repeating the already existing approaches.

Our experimental results show that to classify the FSD collection, the combination of sentiment-carrying words combined with unigrams/bigrams provides also quite accurate results. Moreover, it is indeed only when we combine context-aware RST features with unigrams/bigrams, we can obtain clear polarity classification performance improvements at the sentence level. The most valuable features of the polarity classifiers essentially capture the way in which polar terms are used in a sentence. With this in mind, one possible way of further exploiting would be the sentence's discourse units and their rhetorical roles within the sentence (inter-sentence analysis). For future work, we plan to develop additional features for our methods according to the type of the data examined, to consider additional evaluation measures and finally to compare our proposed context incorporation method with deep neural network approaches.

Our final contribution in the area of Sentiment Analysis and Opinion Mining is based on our proposed idea of training strong feature representations from unlabelled data using Neural Networks and particularly using the Generative Adversarial Neural Networks (GANs), which have proved very popular in the Computer Vision domain. The originality of our proposed method lies in the model's implementation, in the hyper-parameters' choice, in the use of unsupervised learning and finally in its experimental validation on text bodies which come from different sources depending on their genre and length.

There is no doubt that the ability to learn robust, resizable feature representations from unlabeled data has potential applications in a wide variety of machine learning tasks. One way to create such representations is to train deep generative models that can learn to capture the complex distribution of real-world data. Generative adversarial network (GAN) approaches have shown impressive results in producing generative models of images, but relatively little work has been done on evaluating the performance of these methods for the learning representation of natural language, both in supervised and unsupervised settings at the document, sentence, and aspect level. Extensive research validation experiments were performed by leveraging the 20 Newsgroups corpus, the Movie Review (MR) Dataset, and the Finegrained Sentiment Dataset (FSD). Our experimental analysis suggests that

GANs can successfully learn representations of natural language texts at all three aforementioned levels.

The performance of machine learning (ML) methods is heavily dependent on the choice of data or feature representation to which they are applied. For that reason, much of the actual effort in deploying ML algorithms goes into the design of preprocessing the pipelines and data transformations that result in a representation of the data that can support effective ML. Such feature engineering is important but labor-intensive, which highlights the weakness of current learning algorithms. Even though there are a large variety of approaches to representation learning in general, the underlying concept is to learn some set of features from data, and then use these features to solve, for example, a separate (possibly unrelated) task for which we have a large number of labeled examples. As a result, the emergence of large-scale datasets, such as ImageNet [189], which contains 14.197.122 manually labeled images, has allowed the wider-spread use and popularity of convolutional neural networks (CNNs) even in the unrelated task of medical imaging. Currently, the majority of existing classifiers cannot perform as expected when the size of the training dataset is small. Constructing a large labeled dataset, however, is time-consuming and usually requires domain knowledge, making it even more costly. Therefore, there is a gap between the potential benefits of having a large dataset and the difficulty in obtaining labeled data. In order to expand the scope and ease of applicability of ML, it would be highly desirable to make learning algorithms less dependent on feature engineering, so that novel applications can be constructed faster. This could be possible by learning representations (LRs) of the data that make it easier to extract useful information when building classifiers or other predictors. A good representation is one that is also useful as input to a supervised predictor. Among the various methods of LR, this paper focuses on deep learning methods: those that are formed by the composition of multiple nonlinear transformations, with the goal of yielding more abstract—and, ultimately, more useful—representations.

Among the effective approaches that have emerged to train deep generative models, the one that is based on the variational autoencoder (VAE) [115, 187] and the approach that uses generative adversarial networks (GANs) have dominated in recent years. In the former one, the observed data x is assumed to be generated from a set of stochastic latent variables z . The VAE introduces an inference network (implemented using a deep neural network) to approximate the intractable distributions over z , and then maximizes a lower bound on the log-likelihood of $p(x)$. The latter approach uses GANs [83]. In the original GAN formulation, a generator deep neural network learns to map samples from an arbitrary distribution to the observed data distribution. A second deep neural network called the discriminator is trained to distinguish between samples from the empirical distribution and samples that are produced by the generator. The generator is trained to create samples that will fool the discriminator, and so an adversarial game is played between the two networks, converging on a saddle point that is a local minimum for the discriminator and a

local maximum for the generator. Both VAE and GAN approaches have shown impressive results in producing generative models of images [181, 88], but relatively little work has been done on evaluating the performance of these models for learning representations of natural language. GANs for natural language processing (NLP) are considered powerful methods, as they deal with generating sentences: specifically, they produce sentences with certain characteristics (sentiment and questions) and take advantage of the unsupervised nature of these deep neural network (DNN) models. One reason that GANs cannot be directly applied to natural language is the fact that the space in which sentences are present is not continuous and therefore not differentiable. On the contrary, text is represented atomically in terms of discrete tokens (like “man”, “girl”, etc). So, when we want to update the generated sentence slightly according to the discriminator’s behavior, we may not get a sentence. In the computer vision (CV) domain, the output of the generator is an image (a matrix consisting of real valued numbers) which can undergo small updates to make it more difficult for the discriminator to differentiate between the real and fake image. Recently, however, VAEs have been used successfully to create language models [18], to model documents, and to perform question answering [146]. This paper attempts to shed light on whether GANs can be used to learn representations of natural language in an unsupervised setting.

Particularly, in our work, we formulate the ML problem as follows. We propose a novel extension of GANs that replaces the traditional binary classifier discriminator with one that assigns a scalar energy to each point in the generator’s output domain. The discriminator minimizes the hinge loss function used for training “maximum-margin” classifiers, while the generator attempts to generate samples with low energy under the discriminator. We show that a Nash equilibrium [196] under these conditions yields a generator that matches the data distribution (assuming infinite capacity). We conducted experiments with the discriminator in the form of a denoising autoencoder (DAE), optionally including a regularizer that penalizes generated samples having a high cosine similarity to other samples in the mini-batch. Our proposed neural network architecture is based on a variation of the recently proposed energy-based GAN [270] that has been proven to be suitable for the task of generating high-resolution MNIST digit images [52], providing a qualitative evaluation of the learned representations.

Additionally, we decided to replace the standard probabilistic GAN that we cast into the energy model (using Gibbs distributions [80]), as has been proposed by the work of Kim and Bengio [114], and we decided to present the same Nash equilibrium as a standard GAN, but through a different and more generalized class of loss functionals, such as hinge loss. This experimental design selection is based on our attempt in get the pair of models to converge [263] and to exhibit more stable behavior than regular GANs during training [197].

ML for NLP can be performed with a wide range of text formats, from multi-sentence reviews and comments, to single-word expressions of opinion. The most frequent approach

is document-level classification. Another approach, sentence-level classification, limits the analysis to single sentences instead of whole documents. It is typically a harder problem to solve since there is not much information that can be used by the classifier, with sentences being usually much shorter than documents. Compared with document- and sentence-level analysis, aspect-level analysis is more finegrained. Its task is to extract and summarize people's opinions expressed on the aspects of entities, which are also called targets.

With respect to the Machine Learning Approaches for NLP Tasks, Document- or sentence-level classification is often implemented by using **supervised ML methods**. This involves the training of a model using a large body of annotated data which is topic-specific. Any existing supervised ML method can be applied to document-level classification, such as support vector machines (SVMs) or hidden Markov models. Particularly, [166] compared many ML methods on a movie review classifier, concluding that SVMs and Naive Bayes had the best performance overall. Document- or sentence-level regression has also received much attention since many problems cannot be solved with a positive-negative classification, which is frequently used for product reviews, where a 1-5-star rating is prevalent. Pang et al. [164] compared various regression methods, such as SVM regression, SVM multiclass classification, and one-versus-all. Qu et al. [179] extended the bag-of-words representation by exploiting negation and sentiment modifiers, which are more influential in regression than in classification problems. Mejova et al. [145] reviewed feature selection strategies, such as stemming, term frequency, n-grams, point-of-speech, and negation-enriched features. They concluded that a smaller set of features outperformed a larger set for big datasets.

On the other hand, **unsupervised ML** for NLP relies on the dominating influence of sentiment words and phrases to perform the classification without the use of costly annotated data. This has been achieved either by extracting the syntactic patterns of the sentences or by using sentiment lexicons. Both approaches rely on the measurement of the sentiment orientation (SO) of phrases and eventually of the whole document. Lexicon-based analysis calculates the SO values of words and phrases, summing up to the polarity of the whole document. Such classifiers can incorporate negation and intensification since such operations can be easily identified by lexicons. In general, lexicon-based sentiment classifiers show a positive bias, which can be fixed by adjusting the value of the rarer negative expressions [226]. One of their deficits is that lexicon-based methods do not perform well on domain-dependent data, making them less efficient when used for domains that are more challenging for sentiment analysis (SA), such as politics. This issue was partially addressed in [268], but supervised methods still outperform lexicon-based methods for domain-specific problems.

Recently, neural networks (NNs) have started expanding to the field of NLP in the form of both supervised and unsupervised representation learning methods. In terms of **unsupervised representation learning** [160], much of the early research into modern deep learning was developed and validated via this approach [98, 102, 238, 40]. Unsupervised learning is promising due to its ability to scale beyond only the subsets and domains

of data that can be cleaned and labeled given resources, privacy, or other constraints. This advantage is also its difficulty. While supervised approaches have clear objectives that can be directly optimized, unsupervised approaches rely on proxy tasks, such as reconstruction, density estimation, or generation, which do not directly encourage useful representations for specific tasks. As a result, much work has gone into designing objectives, priors, and architectures meant to encourage the learning of useful representations.

Despite these difficulties, there are notable applications of unsupervised learning. Pretrained word vectors are a vital part of many modern NLP systems [41]. These representations, learned by modeling word co-occurrences, increase the data efficiency and generalization capability of NLP systems [171, 34]. How to learn representations of phrases, sentences, and documents is an open area of research. Inspired by the success of word vectors, Kiros et al. [117] proposed skip-thought vectors, a method of training a sentence encoder by predicting the preceding and the following sentence. The representation learned by this objective performs competitively on a broad suite of evaluated tasks. More advanced training techniques, such as layer normalization [8], further improved results. However, skip-thought vectors are still outperformed by supervised models, which directly optimize the desired performance metric on a specific dataset. This is the case for both text classification tasks, which measure whether a specific concept is well encoded in a representation, and more general semantic similarity tasks. This occurs even when the datasets are relatively small by modern standards, often consisting of only a few thousand labeled examples.

With respect to the Deep Neural Network Approaches at Document, Sentence, and Aspect Level, we reviewed initially the latest research efforts, with respect to the application of NNs for NLP tasks, focusing on document-, sentence-, and aspect-level analysis. Rahman et al. [182] introduced a variation of long short-term memory (LSTM) and tested it on the Large Movie Review Dataset of 50K **documents**. It showed that even though accuracy was not increased compared to normal LSTM models, the stability and consistency were improved. Moreover, [122] introduced a recurrent convolutional neural network (RCNN) which applied a bidirectional recurrent structure to capture the contextual information of the document and then employed a max-pooling layer to capture the key components of the text. The researchers in [213] performed comparison tests with single CNN, single LSTM, and a combination of CNN and LSTM on the Large Movie Review Dataset for sentiment classification, showing that the CNN + LSTM network performed better than the respective standalone NNs. A similar research effort was conducted by [261] by combining CNN and LSTM to produce multiple variations which achieved state-of-art performance. Once again, the Large Movie Review Dataset was used, but the authors expressed their belief that the proposed model has potential in both audio and video. Finally, Liu et al. [136] experimented with RNNs, introducing the concept of multitask learning, which connects all the related tasks into a single system, trained jointly. They performed both document-level classification on the Large Movie Review Dataset and sentence-level classification on the Stanford Sentiment Treebank (SST), achieving state-of-art performance results after

fine-tuning.

Moreover, NNs have been widely used for **sentence-level** classification as well. Particularly, Chen et al. [35] proposed a CNN–LSTM model which first classifies the sentences into non-target, one-target, and multi-target sentences. Their model was tested on many datasets (Stanford Sentiment Treebank, Movie Reviews, and Amazon product reviews), and it achieved state-of-the-art performance on some of them. Based on the fact that the combination of CNN and RNN is considered a very popular approach, the authors of [251] proposed their variant, which achieved high accuracy on the typical Stanford Sentiment Treebank and Movie Reviews datasets. Moreover, Conneau et al. [43] used a very deep convolutional network that consisted of 29 layers, resulting in improved performance and proof of the “benefit of depth” for NLP tasks as well. To be more precise, the authors evaluated their model on eight different datasets comprising Movie Reviews and news, testing various NLP tasks such as SA, news categorization, and topic classification. Moreover, Wang J et al. [249] proposed a regional CNN–LSTM model which used an individual sentence as the region for the extraction of affective information. The model was trained with the Stanford Sentiment Treebank dataset and Chinese Valence Arousal (VA) texts using 2K sentences from social forums, and it performed better than CNNs, RNNs, or LSTMs for valence and arousal prediction, respectively.

Finally, **aspect-level analysis** has been also explored. Du et al. [59] modeled both sentiment and syntactic context under specific aspects to acquire better word embeddings, which were given as input to a CNN for sentiment classification of Amazon product reviews. Their results showed an improvement compared to traditional word-embedding methods. Wang et al. [252] applied LSTMs for aspect-based sentiment classification, achieving a state-of-the-art performance of 89.9%. The model was evaluated on the SemEval 2014 dataset while the word embeddings were initialized by Glove⁴, capturing the important parts of a sentence when different aspects were given. Poria et al. [177] tested a deep CNN for aspect extraction using seven levels of NNs. The SemEval 2014 dataset and an aspect-based SA dataset were used for the evaluation and showed improvement in precision and recall. Deep memory networks have also been applied to aspect-level text classification and have shown comparable results to LSTMs while being 15 times faster at the same time [230].

After summarizing the current and previous research efforts that have been conducted within the area of deep neural networks at the document, sentence, and aspect levels of analysis, we observed that, in most cases, networks such as CNNs or recurrent neural networks (RNNs) and particular bidirectional long short-term memory (BLSTMs) networks have been applied; thus, very little work has been conducted on text analysis using generative models and particular GANs. Considering that one of our aims is to examine whether GANs can be used to learn representations of NL texts in an unsupervised manner, we summarize most of the state-of-the-art research works that have used the GAN architecture.

⁴<https://nlp.stanford.edu/projects/glove/>

Additionally, motivated by the fact that the GAN neural network structure can integrate various loss functions, our proposed model was designed to have a better degree of freedom. Moreover, it is expected to provide promising solutions for creatively producing data that are meaningful to humans.

With respect to the Generative Adversarial Networks for NLP Tasks, we review recent research on discovering rich structure in natural language with variational autoencoders (VAEs) [116] and GANs [83]. Evaluating deep generative models has been challenging so far. To the best of our knowledge, there are very few works on text analysis using GANs. Zhang et al. [269] proposed a framework for employing LSTM and CNN for adversarial training to generate realistic text. The latent code z was fed to the LSTM generator at every time step, while CNN acted as binary sentence classifier which discriminated between real data and generated samples. RNN extensions, such as LSTMs or gated recurrent units (GRUs) that retain long-term memory of tokens, have been shown to work well in practice using maximum likelihood estimation.

However, training using maximum likelihood has its downsides, like exposure bias. This refers to the situation in which, during training, the prediction of the next word conditioned on the previous word becomes infeasible since the previous word may not have been seen in the training data. If the generator makes an error early on in the generation process, the generated sentence will keep diverging further away as more words are generated. To solve this issue, a few things have been tried in the past, like scheduled sampling [13] and having task-specific sequence scores. One problem with applying GAN to text is that the gradients from the discriminator cannot properly back-propagate through discrete variables. In [269], this problem was solved by making the word prediction at every time “soft” at the word-embedding space. Moreover, Yu et al. [262] proposed bypassing this problem by modeling the generator as a stochastic policy. The reward signal came from the GAN discriminator, judged on a complete sequence, and was passed back to the intermediate state-action steps using Monte-Carlo search. For text, it is possible to create oracle training data from a fixed set of grammars and then evaluate generative models based on whether (or how well) the generated samples agree with the predefined grammar [225]. Considering that it is hard to make a good evaluation for generating text since there is no objective way to assess whether an artificial sentence is more plausible or realistic than another, some works have used BLEU scores of samples on a large amount of unseen test data. The ability to generate similar sentences to unseen real data is considered a measurement of quality [262].

Notably, in terms of sequential data generation with GANs, an alternative approach based on reinforcement learning was used to train the GAN. We are aware of only one preliminary work using GANs to generate continuous-valued sequences, and it aimed to produce polyphonic music using a GAN with an LSTM generator and discriminator [150]. A work with respect to controllable text generation [101] applied the variable autoencoder (VAE) together with controllable information to generate category sentences. Finally,

Zhang et al. [269] and Semeniuta et al. [211] used GANs for text generation and achieved state-of-the-art results. Finally, another category of approach is the conditional GANs [149] that condition the model on additional information and, therefore, allow us to direct the data generation process. This approach has been mainly used for image generation tasks [181, 4]. Recently, conditional GAN architectures have been also used in NLP, including translation [262] and dialogue generation [129].

After summarizing the work conducted in the area of GANs for NLP tasks, we can observe that even though the research effort has been shifted to GAN approaches for NLP tasks, whose discriminator and generator models are mainly CNNs and/or RNNs (BLSTMs), still a number of limitations and development trends remain unsolved.

With this in mind, our attempt is to investigate whether GANs are a suitable model selection to learn representations of natural language in an unsupervised setting at the document, sentence, and aspect level. Additionally, we revisit the traditional GAN framework from an alternative energy-based perspective. Particularly, we propose a neural network architecture that is based on a variation of the energy-based GAN formulation [270] for generative adversarial training. Our contribution is based on the use of a simple hinge loss, at the point when the system reaches convergence, so that the generator of the energy-based GAN produces points that follow the underlying data distribution. We propose to use a denoising autoencoder architecture as a discriminator in which the energy is a reconstruction error. Our experimental design selection is based on our attempt to get the pair of models to converge [263] and to exhibit more stable behavior than regular GANs during training [197].

Particularly, we attempted to set the scene of our proposed approach by highlighting the limitations of the current energy-based models and discuss the advantages of GANs. GANs have great significance to the development of generative models. As a powerful class of generative methods, GANs solve the problem of generating data that can be naturally interpreted. Especially for the generation of high-dimensional data, the adopted neural network structure does not limit the generation dimension, which greatly broadens the scope of the generated samples. Moreover, the neural network structure can integrate various loss functions, thereby increasing the degree of freedom of the model design. As far as the energy-based models are concerned, they have been used to capture dependencies over variables by defining an energy function. The energy function associates each configuration of the variables with a scalar energy value. Lower energy values should be assigned to more likely or plausible configurations and, conversely, higher energy values should go to others. This has been used, for example, to estimate the probability distribution based on a Boltzmann distribution defined by an energy function and an appropriate normalization factor. In this case, the energy function is defined to assign a probability value that is not normalized. The normalization factor plays an important role by constraining the energy function to properly estimate the probability distribution. However, it introduces difficulties during the learning procedure, which requires an appropriate number of samples

and makes the learning progress slow and noisy or requires certain model structures to get the samples. To overcome these limitations, we present in the following subsection our proposed neural network architecture.

To further formulate our model, we propose to view the model’s discriminator as an energy function that attributes low energies to the regions near the data manifold and higher energies to other regions. At the same time the model’s generator is seen as being trained to produce contrastive samples with minimal energies. With the term contrastive samples, we refer to a data point that causes an energy pull-up, such as an incorrect label in supervised learning and points from low data density regions in unsupervised learning. Our aim is to train the discriminator to assign high energies to these generated samples. Viewing the discriminator as an energy function, our proposed system allows us to use a wide variety of architectures and loss functionals—in addition to the usual binary classifier with logistic output—that have been introduced in the recently proposed “Energy-Based Generative Adversarial Network” model (EBGAN) [270]. However, our proposed model is a modified version of the EBGAN framework since we propose the use of an autoencoder architecture, with the energy being the reconstruction error, in place of the discriminator. Particularly, we suggest that rather than using a single bit of target information to train the model, the reconstruction-based output offers diverse targets for the discriminator. With the binary logistic loss, only two targets are possible, so, within a mini-batch, the gradients corresponding to different samples are most likely far from orthogonal. This leads to inefficient training, and reducing the mini-batch sizes is often not an option on current hardware. On the other hand, the reconstruction loss will likely produce very different gradient directions within the mini-batch, allowing for a larger mini-batch size without a loss of efficiency. Moreover, we decided to use the autoencoders since they have traditionally been used to represent energy-based models and arise naturally. When trained with some regularization terms, autoencoders have the ability to learn an energy manifold without supervision or negative examples. This means that even when an energy-based autoencoding model is trained to reconstruct a real sample, the discriminator contributes to discovering the data manifold by itself. To the contrary, without the presence of negative examples from the generator, a discriminator trained with binary logistic loss becomes pointless.

Furthermore, our proposed work is also inspired by the work of Kim and Bengio [114]. However, it differs in the following way. Their approach uses a standard probabilistic GAN cast into an energy model (using Gibbs distributions), allowing them to learn a discriminator that models the distribution when Nash equilibrium is reached, but it still has mixing problems, especially with deep models that slow the learning. On the contrary, our approach gets rid of the probabilistic setting, while presenting the same Nash equilibrium as a standard GAN, but through a different and more generalized class of loss functionals.

Adversarial networks have enjoyed much success in the CV domain [83, 69], but to our best knowledge, they have not yet achieved comparably successful results when

applied to SA tasks. This study proposes an elegant solution to the problem of learning representations at the document, sentence, and aspect level based on generative adversarial training in an unsupervised way.

Particularly, our work proposes a novel extension of generative adversarial networks that replaces the traditional binary classifier discriminator with one that assigns a scalar energy to each point in the generator's output domain. The discriminator minimizes a hinge loss, while the generator attempts to generate samples with low energy under the discriminator. We show that a Nash equilibrium under these conditions yields a generator that matches the data distribution (assuming infinite capacity). Experiments were conducted with the discriminator taking the form of an autoencoder, optionally including a regularizer that penalizes generated samples having a high cosine similarity to other samples in the mini-batch. Finally, we visualized the data representation of all three architectures by projecting the features into the layers of the proposed adversarial document-, sentence-, and aspect-level neural network architectures for SA, respectively.

There are several promising directions for future work to further extend the presented research work based on our proposed adversarial document-, sentence-, and aspect-level neural network architectures for SA. Considering that, in general, the neural network structure of our GAN models can integrate various loss functions, we do intend to increase the degree of freedom of the model design. Particularly, since our proposed approach is an alternative energy-based perspective of the GAN framework, it would be interesting to incorporate the family of energy-based loss functionals presented in [124] into the energy-based GAN models. Thus, the conditional setting presented in [54] is a promising setup to explore, and we expect that will attract more attention to a broader view of GANs from the energy-based perspective. Additionally, taking into account that GANs are also meaningful and instructive for semi-supervised learning, we do intend to use the training process of GANs to achieve pretrained data using unlabeled data. To be more precise, we aim first to use a large amount of unlabeled data to train our proposed GAN models, and, based on the understanding of the trained GANs over the unlabeled data, we intend then to use a small amount of labeled data to train the discriminative model for classification and regression tasks. Finally, our future plans include gaining an understanding of why the denoising autoencoder in the GAN discriminator appears to produce significantly better representations than a standard autoencoder, a result demonstrated by the experimental evaluation.

Γλωσσάριο Τεχνικών Όρων

Acceptance	αποδοχή
Activation	ενεργοποίηση
Affect	συναίσθημα
Affect analysis	ανάλυση συναισθήματος
Affective Computing	Συναισθηματική Υπολογιστική
Affect production	δημιουργία συναισθήματος
Affect synthesis	σύνθεση συναισθήματος
Appraisal	γνωσιακή εκτίμηση
Arousal	διέγερση
Attention	προσοχή
Attitude	συμπεριφορά
Awareness	ευαισθητοποίηση
Base learners	βασικοί ταξινομητές
Baselines	συστήματα αναφοράς
BLSTM	αμφίδρομης μεγάλης και μικρής διάρκειας Νευρωνικά Δίκτυα
Bi-grams	διγράμματα

Cardiac Output	καρδιακή παροχή
Claim	απαίτηση
Clause	έκφραση
Cognitive Neuroscience	γνωστική Νευροεπιστήμη
Cognitive task	γνωστική εργασία
Componential model	Συνθετικό Μοντέλο Συναισθημάτων
Context-aware	επίγνωση πλαισίου
Contextual type	τύπος πλαισίου
Deep Generative Models	βαθιά παραγωγικά μοντέλα
Discourse analysis	ανάλυση του λόγου
Discourse segmentation	τμηματοποίηση του λόγου
Disgust	αηδία
Disposition	συναισθηματική διαταγή
Document-level analysis	ανάλυση σε επίπεδο κειμένου
Domain	τομέας
Dominance	κυριαρχία
Emotional elicitation	εξαγωγή συναισθήματος
Emotional classes	συναισθηματικές κατηγορίες
Emotions	συναισθήματα
Ensemble learners	σύνολο σύνθετων κατηγοριοποιητών
Evaluations	εκτιμήσεις
Fear	φόβος
Feelings	αισθήματα

Happiness	ευτυχία
Heart Rate	καρδιακός ρυθμός
Hidden Markov Models	κρυφά μαρκοβιανά μοντέλα
Human behavior understanding	κατανόηση της ανθρώπινης συμπεριφοράς
Intensity	ένταση
Interpretation	ερμηνεία
Iterative scaling algorithms	αλγόριθμοι επαναλαμβανόμενης κλίμακας
GANs	Ανταγωνιστικά Παραγωγικά Μοντέλα
Left Ventral (systolic) Ejection Time	αριστερός κοιλιακός (συσταλτικός) χρόνος
Multi-word concepts	πολλαπές λέξεις-έννοιες
Name Entity Recognition	σύστημα αναγνώρισης ονομαστικής καταχώρησης
Naturalistic data	δεδομένα αυθόρμητης μορφής
Natural Language Processing	επεξεργασία Φυσικής Γλώσσας
Negation words	λέξεις άρνησης
Normalization constant	σταθερά κανονικοποίησης
Operational tasks	απαιτητικές εργασίες
Opinion Mining	εξόρυξη Γνώμης
Part-Of-Speech	μέρη του λόγου
Polarity	πολικότητα
Precision	ακρίβεια

Pre-ejection period	προσυσταλτική περίοδος
Psychological Foundations	Ψυχολογικές Βάσεων
Ratings	βαθμολογήσεις
Recall	ανάκληση
Recurrent Neural Networks	επαναλαμβανόμενα νευρωνικά δίκτυα
Recommendation systems	συστήματα συστάσεων
Repositories	αποθήκες
Reviews	διαδικασίες αξιολόγησης/κριτικής
Rhetorical Structure Theory	Θεωρία του λόγου
Sadness	θλίψη
Self Assessment Manikins	κλίμακα Αυτοαξιολόγησης
Semantic parser	σημασιολογικός αναλυτής
Sentiment Analysis	ανάλυση προσωπικών ερμηνειών από το κείμενο
Sentiment words	λέξεις με συναισθηματικό περιεχόμενο
Sequential tagging task	ακολουθιακή διεργασία επισημείωσης
Session	συνεδρία
Shallow parsers	ρηχοί αναλυτές
Skin conductivity	αγωγιμότητα του δέρματος
Support Vector Machines	μηχανές διανυσματικής υποστήριξης
Surprise	έκπληξη
Syntactic patterns	μοτίβα συντακτικά
Task	καθήκον

Textual cues	συνθήματα κειμένου
Tokenization	τμηματοποίηση
Total Peripheral Resistance	ολική περιφερική αντίσταση
Tri-grams	τριγράμματα
Turn-taking	η στροφή στην ανάλυση των συνομιλιών
Unigrams	μονογράμματα
Unit	μονάδα λόγου
Valence	σθένος
Ventricular Contractility	κοιλιακή συσταλτικότητα

Κεφάλαιο 1

Εισαγωγή

1.1 Κίνητρο και Ορισμός Προβλήματος

Τα συναισθήματα εμφανίζονται καθημερινά στη ζωή μας και συνήθως συμβαίνουν στο πλαίσιο μίας κατάστασης ή μίας αλληλεπίδρασης, διαμορφώνοντας την ερμηνεία μας για τα γεγονότα που βιώνουμε. Τα συναισθήματα εκφράζονται συγκαλυμμένα ή μη, με ένταση και σαφήνεια και εξελίσσονται με το πέρασμα του χρόνου. Οι ορατές και μη συναισθηματικές μας καταστάσεις τείνουν να διαμορφώνουν την μονοτροπική είτε πολυτροπική επικοινωνία μας μέσω των εκδηλώσεων του προσώπου, της φωνής ή της φυσιολογίας που προκαλούνται στο σώμα (εφίδρωση, αύξηση καρδιακών παλμών κοκ) συνδυαστικά είτε μεμονωμένα. Δημιουργείται λοιπόν η ανάγκη για την ύπαρξη υπολογιστικών μοντέλων συναισθημάτων, ικανά να ερμηνεύουν τέτοιες σύνθετες πτυχές της συναισθηματικής έκφρασης λαμβάνοντας υπόψιν το πλαίσιο στο οποίο εμφανίζεται μία συναισθηματική εκδήλωση, οδηγώντας έτσι σε μία αποτελεσματική αλληλεπίδραση.

Η ανάπτυξη υπολογιστικών μεθόδων που να κατανοούν τα ανθρώπινα συναισθήματα, θα μπορούσε επίσης να μας προσφέρει τα εργαλεία ώστε να εξερευνήσουμε τη μη ορατή συναισθηματική ανθρώπινη έκφραση, όπως για παράδειγμα τη συμπεριφορά των ατόμων με ψυχικές ή ψυχολογικές διαταραχές, είτε να βοηθήσει στην έγκαιρη διάγνωση νευροεκφυλιστικών νόσων, όπου η εκδήλωση των συμπτωμάτων συμβαίνει αρκετά χρόνια μετά την έναρξη του νευροεκφυλισμού. Η σχέση μεταξύ των υπολογιστικών προσεγγίσεων συνδυαστικά με τις εφαρμογές στο τομέα της Υγείας αποτελεί ένα συναρπαστικό μονοπάτι για εξερεύνηση.

Όσον αφορά σε παραδείγματα πραγματικών εμπειριών, εξετάζουμε την περίπτωση ενός επιβάτη αεροπλάνου ο οποίος ενημερώνεται από έναν υπάλληλο του αερολιμένα ότι οι αποσκευές του χάθηκαν και ότι δε θα λάβει οικονομική αποζημίωση για αυτή την τάλαιπωρία. Αυτή η κατάσταση θέτει ήδη ένα πλαίσιο για τη συναισθηματική απάντηση που θα περίμενε κάποιος, ενδεχομένως αρνητικά συναισθήματα, όπως η δυσαρέσκεια. Επιπρόσθετα, δεδομένου ότι ο επιβάτης και ο εργαζόμενος του αεροδρομίου έχουν εκδηλώσει αρνητικά συναισθήματα μέχρι εκείνο το σημείο της συζήτησης, ίσως είναι πιθανό να ακολουθήσουν αρνητικά συναισθήματα. Αν δηλαδή, η τυπική εξέλιξη αυτών των συ-

ναισθημάτων (συναισθηματική αλληλουχία) τείνει να ακολουθεί συγκεκριμένα μοτίβα, τότε και στη συνέχεια αναμένεται να ισχυροποιηθεί η ικανότητά μας να ερμηνεύουμε τις συναισθηματικές εκδηλώσεις.

Στο ίδιο σενάριο των αποσκευών που χάθηκαν, ο επιβάτης μπορεί να εκφράσει την οργή του μέσω διαφόρων επικοινωνιακών καναλιών. Αυξάνοντας την ένταση της ομιλίας του, χρησιμοποιώντας υβριστικά λόγια, ανασηκώνοντας τα φρύδια του, πραγματοποιώντας γρήγορες κινήσεις με τα χέρια του μεταφέρει δυσφορία και συνδυαστικά είτε μεμονωμένα με το αντίστοιχο λεκτικό περιεχόμενο μπορεί να μεταδώσει θυμό. Είναι πιθανό επίσης να καταφύγει σε σαρκασμό συνδυάζοντας συγκρουόμενα πολυτροπικά κανάλια επικοινωνίας για παράδειγμα, χαμογελαστό πρόσωπο με θυμωμένη φωνή. Αυτή η πολυτροπική φύση του συναισθήματος καθιστά σημαντική τη σωστή κωδικοποίηση και μοντελοποίηση των πληροφοριών που μεταδίδονται αλλά και τον εντοπισμό και την ενσωμάτωση του πλαισίου προκειμένου να αποκτηθεί μία πιο ολοκληρωμένη εικόνα του εκφραζόμενου συναισθήματος. Έτσι, είναι πιθανό για παράδειγμα κάποιος να μην καταφέρει να αναγνωρίσει την ευτυχία ή τη χαρά μέσω της ομιλίας εάν αυτή η συγκίνηση μεταφέρεται αποκλειστικά και μόνο μέσω ενός λαμπερού χαμόγελου ή να μην μπορεί να ερμηνεύσει ένα χαμόγελο ως έκφραση σαρκαστικού θυμού εκτός και αν ληφθεί υπόψιν η αντίστοιχη θυμωμένη ομιλία και η κατάσταση που διαδραματίζεται (πλαίσιο).

Επιστρέφοντας λοιπόν στη συναισθηματική κατάσταση του επιβάτη του αεροδρομίου, η σαφήνεια, η ένταση και ο βαθμός της δυσαρέσκειας και του θυμού του δεν είναι σταθερά, αλλά εξελίσσονται κατά τη διάρκεια της αλληλεπίδρασης. Για παράδειγμα, ο “υπόγειος” θυμός του επιβάτη στην αρχή της συζήτησης μπορεί να κλιμακωθεί σε μανία όταν έρχεται αντιμέτωπος με την αναφορά του υπαλλήλου του αεροδρομίου. Αυτό δείχνει ότι τα συναισθήματα δεν αντιπροσωπεύονται πάντα από απλές κατηγορικές περιγραφές, αλλά μπορούν να εξελίσσονται μέσα στο χρόνο. Ως εκ τούτου, μία συναισθηματική κατάσταση μπορεί να εκπροσωπείται με όρους συνεχών χαρακτηριστικών. Τέτοιες συνεχείς αναπαραστάσεις θα μπορούσαν επίσης να θεωρηθούν ως ένας πιο γενικός τρόπος για να περιγράψουμε τα συναισθήματα, και συγκεκριμένα συναισθηματικές εκδηλώσεις που είναι πολύπλοκες και ασαφείς.

Η Ανάλυση Συναισθήματος (Sentiment Analysis) ή η Εξόρυξη Γνώμης (Opinion Mining) είναι η υπολογιστική μελέτη των απόψεων (opinions), των συναισθημάτων (sentiments), των εκτιμήσεων (appraisals) και των στάσεων (attitudes) των ανθρώπων σε οντότητες όπως προϊόντα, υπηρεσίες, οργανώσεις, άτομα, θέματα, γεγονότα κ.ο.κ [134]. Η ταχεία ανάπτυξη του πεδίου συμπίπτει με την ανάπτυξη των κοινωνικών μέσων στο διαδίκτυο, όπως για παράδειγμα, σχόλια, συζητήσεις σε φόρουμ, Twitter και κοινωνικά δίκτυα, καθώς για πρώτη φορά έχουμε ένα τεράστιο όγκο δεδομένων καταγεγραμμένο σε ψηφιακή μορφή. Τα τελευταία 20 χρόνια, η ανάλυση συναισθημάτων έχει εξελιχθεί σε έναν από τους πιο δραστήριους τομείς της έρευνας, στον τομέα της επεξεργασίας της Φυσικής Γλώσσας. Μελετάται επίσης συνδυαστικά με τον τομέα της Εξόρυξης Δεδομένων, της Εξόρυξης Ιστού, της Εξόρυξης Κειμένου και της Ανάκτησης Πληροφοριών με μία σειρά

εφαρμογών όπως οι επιστήμες της διαχείρισης, η πολιτική επιστήμη, οι επικοινωνίες, η επιστήμη της Υγείας και ακόμη και η ιστορία, λόγω της σημασίας της για τις επιχειρήσεις και την κοινωνία στο σύνολό της. Η εξάπλωση αυτή οφείλεται στο γεγονός ότι οι απόψεις κατέχουν κεντρικό ρόλο σε όλες σχεδόν τις ανθρώπινες δραστηριότητες και αποτελούν βασικό παράγοντα που επηρεάζει την ανθρώπινη συμπεριφορά. Οι πεπειθήσεις, οι αντιλήψεις μας, οι επιλογές που κάνουμε εξαρτώνται, σε σημαντικό βαθμό και από το πώς οι άλλοι βλέπουν και αξιολογούν τον κόσμο. Για το λόγο αυτό, όταν καλούμαστε να λάβουμε μία απόφαση, συχνά ανατρέχουμε στις απόψεις των άλλων.

Στις μέρες μας, εάν κάποιος επιθυμεί για παράδειγμα να αγοράσει ένα καταναλωτικό προϊόν, δεν περιορίζεται πλέον απλώς στο να ρωτήσει τη γνώμη των φίλων του και της οικογένειάς του. Αντίθετα μπορεί να ανατρέξει σε πολλές κριτικές χρηστών και συζητήσεις σχετικά με το προϊόν σε δημόσια φόρουμ στο διαδίκτυο. Επίσης, για έναν οργανισμό, ενδέχεται να μην είναι πλέον απαραίτητο να διεξάγονται έρευνες και δημοσκοπήσεις για τη συγκέντρωση των απόψεων των πολιτών, καθώς υπάρχει αφθονία τέτοιων πληροφοριών διαθέσιμων στο κοινό. Τα τελευταία χρόνια έχουμε διαπιστώσει ότι οι καταγεγραμμένες τοποθετήσεις στα κοινωνικά μέσα ενημέρωσης έχουν βοηθήσει στην αναμόρφωση των επιχειρήσεων και επηρεάζουν σημαντικά τη δημόσια γνώμη.

Ωστόσο, η εύρεση και παρακολούθηση ιστότοπων γνώσης στον Παγκόσμιο Ιστό και η απομόνωση των πληροφοριών που περιέχονται σε αυτό εξακολουθεί να αποτελεί ένα τεράστιο έργο λόγω της διάδοσης ποικίλων σελίδων στο διαδίκτυο. Κάθε ιστότοπος περιέχει συνήθως ένα τεράστιο όγκο κειμένων γνώμης που δεν είναι πάντα εύκολο να αποκρυπτογραφηθεί σε εκτενή blogs και σε τις καταχωρίσεις σε φόρουμ. Κάτω από αυτές τις συνθήκες ο μέσος αναγνώστης δυσκολεύεται να εντοπίσει σχετικούς ιστότοπους, να εξάγει και να συνοψίσει τις απόψεις τους. Επομένως, η ανάγκη για αυτοματοποιημένα συστήματα ανάλυσης γνώμης είναι έκδηλη. Εξαιτίας αυτού, υπάρχουν πολλές νεοσύστατες εταιρείες που εστιάζουν στην παροχή υπηρεσιών ανάλυσης γνώμης. Αυτές οι πρακτικές εφαρμογές παρέχουν ισχυρά κίνητρα για έρευνα στην ανάλυση του συναισθήματος και γνώμης.

Η τρέχουσα έρευνα έχει δημιουργήσει πολυάριθμες τεχνικές για διάφορα καθήκοντα (tasks) ανάλυσης συναισθημάτων, τα οποία περιλαμβάνουν τόσο επιβλεπόμενες όσο και μη επιβλεπόμενες μεθόδους. Στην πρώτη περίπτωση, οι πρώτες ερευνητικές προσπάθειες βασίστηκαν στη χρήση μηχανών διανυσματικής υποστήριξης (SVM), μέγιστης εντροπίας, Naive Bayes κ.λπ. Αντίστοιχα, η χρήση της μη επιβλεπόμενης μάθησης περιελάμβανε διάφορες μεθόδους που εκμεταλλεύονταν τα λεξικά συναισθημάτων, τη γραμματική ανάλυση και τα συντακτικά επαναλαμβανόμενα μοτίβα [134, 133, 165].

Την τελευταία δεκαετία περίπου, οι μέθοδοι της Βαθιάς Μάθησης έχουν αναδειχθεί ως μια ισχυρή τεχνική Μηχανικής Μάθησης [82] και παράγουν αποτελέσματα τελευταίας τεχνολογίας σε πολλούς τομείς εφαρμογής, που εκτείνονται από τον τομέα της Όρασης Υπολογιστών και την αναγνώριση ομιλίας έως τον τομέα της Φυσικής Γλώσσας. Συγκεκριμένα, η εφαρμογή της Βαθιάς Μάθησης στην ανάλυση συναισθημάτων και γνώμης σε

επίπεδο κειμένου, πρότασης και χαρακτηριστικών έχει γίνει πολύ δημοφιλής πρόσφατα.

Αυτή η εργασία προτείνει λοιπόν τόσο θεωρητικές όσο και υπολογιστικές μεθόδους που στοχεύουν στη μοντελοποίηση με μεγαλύτερη ακρίβεια της συναισθηματικής έκφρασης λαμβάνοντας υπόψη το πλαίσιο εντός του οποίου λαμβάνει χώρα και τις συνεχείς και εξελισσόμενες ιδιότητες του συναισθήματος κατά τη διάρκεια μιας συναισθηματικής εκδήλωσης ή μιας αποτελεσματικής αλληλεπίδρασης. Στόχος μας είναι να περιγράψουμε και να επικυρώσουμε πειραματικά τις μεθοδολογίες που οδηγούν σε πιο αποτελεσματικά συστήματα αναγνώρισης συναισθημάτων, με περισσότερες δυνατότητες. Ταυτόχρονα, ελπίζουμε να ρίξουμε φως στον τρόπο με τον οποίο οι πολυτροπικές εκφράσεις διαμορφώνονται υπό το πρίσμα της μη-ορατής συναισθηματικής κατάστασης, ενώ συγχρόνως επανεξετάζουμε μία σειρά ποιοτικών ψυχολογικών παρατηρήσεων και μοντέλων που να είναι περισσότερο συμβατά με την ενσωμάτωση του πλαισίου. Το τελευταίο μέρος αυτής της εργασίας προτείνει τη χρήση τεχνικών Βαθιας Μάθησης με τη χρήση Νευρωνικών Δικτύων (Deep Learning) για την ανάλυση και την ερμηνεία απόψεων και συναισθημάτων σε επίπεδο κειμένου, προτάσεων και χαρακτηριστικών.

Στόχος μας είναι να χρησιμοποιήσουμε τις υπολογιστικές μας μεθόδους για να διερευνήσουμε και να ποσοτικοποιήσουμε τις αναλύσεις μεταξύ των ορατών και μη ορατών συναισθηματικών εκφράσεων, λαμβάνοντας κάθε φορά υπόψη το πλαίσιο στο οποίο εκτυλίσσεται το εξεταζόμενο γεγονός, προκειμένου να προωθηθεί η κατανόηση αυτής της περίπλοκης ψυχολογικής κατάστασης, προσεγγίζοντάς την από υπολογιστική άποψη.

Υπάρχουν πέντε βασικοί πυλώνες σε αυτό το έργο. Αρχικά, η έρευνά μας τοποθετημένη **στην τομή των ερευνητικών περιοχών της Αλληλεπίδρασης του Πλαισίου και της Συναισθηματικής Υπολογιστικής, διερευνώντας εάν και πώς το πλαίσιο μπορεί να ενσωματωθεί στην αυτόματη ανάλυση της ανθρώπινης συναισθηματικής συμπεριφοράς**. Αρκετές σχετικές πτυχές συζητούνται, μεταξύ των οποίων εντάσσονται η μοντελοποίηση, η απόκτηση και η επισημείωση σωμάτων κειμένου ενισχυμένων συναισθηματικά καθώς και θέματα που σχετίζονται με την ενσωμάτωση των πληροφοριών πλαισίου σε ένα πολυμορφικό πλαίσιο σύντηξης της συναισθηματικής ανάλυσης. Οι πτυχές αυτές εξετάζονται κριτικά όσον αφορά στις προκλήσεις που περιλαμβάνουν, ενώ, σε ένα ευρύτερο πλαίσιο συζήτησης, εντοπίζονται οι μελλοντικές κατευθύνσεις αυτής της πρόσφατα ενεργούς, αλλά κυρίως ανεξερευνήτης ερευνητικής περιοχής. Συνολικά, η εργασία στοχεύει τόσο στην τεκμηρίωση της σημερινής κατάστασης, καθώς και στο σχολιασμό της εξέλιξης του επερχόμενου θέματος του Πλαισίου στο χώρο της Συναισθηματικής Υπολογιστικής. Το έργο αυτό περιγράφεται λεπτομερώς στο Κεφάλαιο 2 και στο Κεφάλαιο 3. Το περιεχόμενο αυτών των κεφαλαίων βασίζεται κατά κύριο λόγο στο κείμενο των δημοσιεύσεων [242, 244].

Στο δεύτερο μέρος της δουλειάς μας θέτουμε **το πρόβλημα της πολυτροπικής αναγνώρισης των συναισθημάτων από δύο κανάλια επικοινωνίας**. Στο Κεφάλαιο 4 η έννοια του πλαισίου ορίζεται ως **η σχέση μεταξύ της εξωτερικής συμπεριφοράς και της εσωτερικής συναισθηματικής κατάστασης δύο μελών της ίδιας ομάδας** (“εκπαιδευ-

τής” - “εκπαιδευόμενος”) κατά τη διάρκεια ενός απαιτητικού επιχειρησιακού έργου όπως η απενεργοποίηση μιας βόμβας. Τα μέλη της ομάδας αξιολογούνται σύμφωνα με τις προφορικές τους απαντήσεις (δηλ. τη διάρκεια των απαντήσεών τους) σε σχέση με τις εξωτερικές και εσωτερικές συναισθηματικές τους ενδείξεις. Στα εξωτερικά συμπεριφοριστικά συναισθήματα περιλαμβάνονται οι φωνητικές εκφράσεις του εκπαιδευόμενου, ενώ οι εσωτερικές γνώσεις βασίζονται σε σήματα φυσιολογίας. Συγκεκριμένα, κατηγοριοποιούμε τις ενδείξεις των σημάτων φυσιολογίας των “εκπαιδευμένων” που λαμβάνουν χώρα μετά την ολοκλήρωση της ομιλίας του “εκπαιδευτή” ανάλογα με το εάν ανήκουν σε μία μικρότερης ή μία μεγαλύτερης διάρκειας απάντηση. Με βάση την υπόθεση ότι οι μακροπρόθεσμες αντιδράσεις είναι πιθανό να προκληθούν από απαιτητικά γεγονότα γνωστικών εργασιών (cognitive task) και/ή αγχωτικές αλληλεπιδράσεις, υποθέτουμε ότι οι εσωτερικοί μηχανισμοί που δημιουργούνται σε αυτά τα έντονα συναισθηματικά διαστήματα δραστηριότητας θα αντικατοπτρίζονται στα χαρακτηριστικά φυσιολογίας του εκπαιδευόμενου.

Για την συσχέτιση των σημάτων φυσιολογίας του “εκπαιδευόμενου” τα οποία έπονται των ερωτήσεων του εκπαιδευτή αναφορικά με την κατηγοριοποίηση εάν συμβαίνουν σε σύντομο ή μεγαλύτερο χρονικό διάστημα, χρησιμοποιήθηκε ένα σώμα κειμένου που περιλαμβάνει τις συνομιλίες δύο μελών μίας ομάδας. Τα αποτελέσματα επιβεβαιώνουν τη συσχέτιση που υπάρχει ανάμεσα στο πλαίσιο της συνομιλίας των δύο μελών της ομάδας και την εσωτερική συναισθηματική κατάσταση του εκπαιδευόμενου. Επιπρόσθετα, κατηγοριοποιούμε αυτή τη σύνδεση εξετάζοντας ποικίλους βασικούς κατηγοριοποιητές (base learners) και ένα σύνολο σύνθετων κατηγοριοποιητών (ensemble learners) για την καλύτερη δημιουργία ενός σταθερού σύνθετου κατηγοριοποιητή που γενικεύει καλά στις εξωτερικές και εσωτερικές ενδείξεις των ατόμων. Το περιεχόμενο αυτού του κεφαλαίου βασίζεται κατά κύριο λόγο στο κείμενο της δημοσίευσης [246].

Διαπιστώνουμε λοιπόν, ότι η ικανότητα των ανθρώπων να αλληλεπιδρούν αποτελεσματικά εξαρτάται σε μεγάλο βαθμό από την ευαισθητοποίησή τους για το πλαίσιο της αλληλεπίδρασης που λαμβάνει χώρα. Θέτουμε λοιπόν την έννοια του πλαισίου σε ένα περισσότερο θεωρητικό μοντέλο στόχος του οποίου είναι η γεφύρωση του εννοιολογικού κενού. Αξιοποιώντας λοιπόν τη σημασιολογία, τις γνωστικές πληροφορίες και τις πληροφορίες συναισθήματος σχετικά με τη λεκτική συμπεριφορά, εξετάζουμε κατά πόσο η πληροφορία αυτή μπορεί να ενσωματωθεί σε συστήματα αυτόματης ανάλυσης της ανθρώπινης συμπεριφοράς. Προτείνουμε λοιπόν μια **μεθοδολογία σημασιολογικής εξαγωγής πολλαπλών εννοιών**. Η εφαρμογή της στα ενδεικτικά παραδείγματα του σώματος κειμένου SEMAINE επικυρώνει την προτεινόμενη προσέγγιση (Κεφάλαιο 5). Τα θεμελιώδη ζητήματα λοιπόν που προκύπτουν είναι κυρίως η μετατόπιση του ενδιαφέροντος στην καλύτερη αναγνώριση, ερμηνεία και επεξεργασία των απόψεων (opinions) και συναισθημάτων (sentiments) μέσα στην πρόταση δεδομένου ότι μπορούμε να εστιάζουμε σε απαντήσεις ερωτημάτων όπως το “Γιατί” που είναι περισσότερο υποκειμενικές. Το περιεχόμενο αυτού του κεφαλαίου βασίζεται στο κείμενο της δημοσίευσης [245].

Έτσι, προσπαθώντας να αναδείξουμε **την συνεισφορά του εννοιολογικού πλαισίου αναφορικά με την ερμηνεία των απόψεων (opinions) και των συναισθημάτων (sentiments)** στο τέταρτο μέρος αυτής της εργασίας μελετούμε διεξοδικά τη χρησιμότητα δύο υπολογιστικών μοντέλων, του Conditional Random Field (CRF) και της θεωρίας του λόγου Rhetorical Structure Theory (RST) για την ενσωμάτωση των δομικών χαρακτηριστικών της πρότασης και των πληροφοριών πλαισίου σε συστήματα που έχουν επίγνωση του συναισθηματικού τους περιβάλλοντος. Εφαρμόζουμε τα δύο προτεινόμενα μοντέλα μας σε δύο πηγές πληροφοριών διαφορετικού μεγέθους (size) και είδους (genre), επιβεβαιώνοντας ότι ο τομέας της ανάλυσης Συναισθήματος από πόρους της Φυσικής Γλώσσας (Sentiment Analysis (SA)) μπορεί να βελτιωθεί λαμβάνοντας υπόψη τη δομή των προτάσεων και χρησιμοποιώντας πληροφορίες πλαισίου. Η συνεισφορά αυτής της ερευνητικής προσπάθειας έγκειται στην επέκταση του συνόλου των χαρακτηριστικών πλαισίου που προτείνουμε και στη συνδυαστική τους χρήση με σημασιολογικά, συντακτικά χαρακτηριστικά και με χαρακτηριστικά πλαισίου RST. Αυτό το έργο περιγράφεται λεπτομερώς στο Κεφάλαιο 6 και παρουσιάζεται στις δημοσιευμένες εργασίες [248, 247] και ως κεφάλαιο βιβλίου [15], ενώ έχει τιμηθεί από την επιτροπή του 3ου International Workshop: Emotions, Modality, Sentiment Analysis and the Semantic Web του ESWC (ESWC2017) συνεδρίου το Μάιο του 2017 **με το βραβείο καλύτερης δημοσίευσης (best paper award)**.

Τέλος, το τελευταίο μέρος αυτής της εργασίας προτείνει τη διερεύνηση σύγχρονων **υπολογιστικών μεθοδολογιών στον τομέα της Ανάλυσης Γνώμης και Συναισθήματος**. Συγκεκριμένα, η ικανότητα εκμάθησης ισχυρών αναπαραστάσεων χαρακτηριστικών από μη επιβλεπόμενα δεδομένα, έχει πιθανές εφαρμογές σε μια μεγάλη ποικιλία διεργασιών Μηχανικής Μάθησης. Ένας τρόπος για να δημιουργήσουμε τέτοιες αναπαραστάσεις είναι να εκπαιδύσουμε βαθιά παραγωγικά μοντέλα (Deep Generative Models) τα οποία μπορούν να αποτυπώνουν τη σύνθετη κατανομή δεδομένων σε πραγματικές συνθήκες. Οι προσεγγίσεις των Ανταγωνιστικά Παραγωγικών Δικτύων (Generative Adversarial Networks (GANs)) έχουν πρόσφατα παρουσιάσει εντυπωσιακά αποτελέσματα στην παραγωγή μοντέλων (generative models) εικόνων. Ειδικότερα πρόκειται για μία νέα προσέγγιση που έχει μελετηθεί ελάχιστα και παρουσιάζει τεράστιο και διαρκώς αυξανόμενο ερευνητικό ενδιαφέρον. Ειδικότερα, στον τομέα της Φυσικής Γλώσσας, η συνεισφορά μας έγκειται στην εκπαίδευση ισχυρών αναπαραστάσεων χαρακτηριστικών από μη επισημειωμένα δεδομένα με τη χρήση των Ανταγωνιστικά Παραγωγικών Μοντέλων (GANs), τα οποία παρουσιάζουν εντυπωσιακά αποτελέσματα στον τομέα της Όρασης Υπολογιστών. Η πρωτοτυπία της συγκεκριμένης μεθόδου έγκειται στον τρόπο υλοποίησης του μοντέλου, στην επιλογή των υπερπαραμέτρων, στη χρήση της μη επιβλεπόμενης μάθησης και στην πειραματική επικύρωση του προτεινόμενου μοντέλου σε σώματα κειμένου που προέρχονται από διαφορετικές πηγές αναφορικά με το είδος τους και την έκτασή τους, συσχετισμοί που παρουσιάζουν πολύ μικρή έως καθόλου ερευνητική δραστηριότητα. Αυτή η κατεύθυνση εισάγεται στο Κεφάλαιο 7 και παρουσιάζεται στη δημοσιευμένη εργασία

[241].

Τοποθετώντας λοιπόν την παρούσα ερευνητική προσπάθεια στο ευρύτερο πλαίσιο της Συναισθηματικής Υπολογιστικής, θα μπορούσε κανείς να τη δει ως μία προσπάθεια για να κατανοήσουμε καλύτερα τις προκλήσεις που εμπλέκονται αναγνωρίζοντας πολύπλοκες συναισθηματικές εκφράσεις και να προτείνουμε υπολογιστικές λύσεις για την ανάλυση γνώμης και συναισθήματος [44, 209]. Η προσέγγισή μας στοχεύει στο να εμπλουτίσει τα συστήματα αναγνώρισης συναισθήματος, εξετάζοντας όχι μόνο την τρέχουσα παρατήρηση, αλλά επίσης και τις πληροφορίες γύρω από αυτό λαμβάνοντας υπόψιν το πλαίσιο στο οποίο εκτυλίσσεται η συναισθηματική έκφραση.

Επιπλέον, μέσω της ενσωμάτωσης του πλαισίου κινούμαστε προς τη χρήση πληροφοριών υψηλότερου επιπέδου (higher level), υπό την έννοια της επισήμανσης συναισθηματικά σημαντικών γεγονότων μίας αλληλεπίδρασης ή εξετάζοντας το πλαίσιο και τον καθορισμό μιας συναισθηματικής συνομιλίας και αλληλεπίδρασης.

Παρά τις αρκετές προκλήσεις, οι συνέπειες του πεδίου της Συναισθηματικής Υπολογιστικής είναι πολλαπλές τόσο για την επιστήμη όσο και για την τεχνολογία [209, 173]. Πράγματι, από μία τεχνολογική σκοπιά, ο τομέας αυτός έχει τη δυνατότητα να αλλάξει την αντίληψή μας για το τί είναι μία μηχανή και τον τρόπο με τον οποίο αλληλεπιδρούμε μαζί της. Θα μπορούσαμε να σκεφτούμε πραγματικά εξατομικευμένους υπολογιστές που θα μπορούσαν να αναγνωρίσουν τη διάθεσή μας, π.χ. άγχος εναντίον χαράς από οπτικο-ακουστικά σήματα και από το περιβάλλον (πλαίσιο), όπως για παράδειγμα το χρόνο και την ημέρα της εβδομάδας ή τον αριθμό των ραντεβού στο διαδίκτυο, το ημερολόγιο και να μας προτείνουν μουσική, ταινίες ή ακόμη και να κατατάσσουν τα νέα της εισερχόμενης ειδησεογραφίας. Ακόμη και αν η τρέχουσα απόδοση της αναγνώρισης συναισθήματος και γνώμης είναι αρκετά μακριά από την άριστη κατανόηση της συναισθηματικής κατάστασης του χρήστη, π.χ. θετική έναντι αρνητικής και του βαθμού έντασής της, θα μπορούσε να παρέχει μία κατάλληλη απάντηση από τον υπολογιστή που θα έφερνε την ανθρώπινη αλληλεπίδραση με τον υπολογιστή (HCI) πιο κοντά στην παραγωγή μίας ανθρώπινης εμπειρίας. Από τη σκοπιά της σύνθεσης, μία ανάλυση του τρόπου με τον οποίο οι εκφράσεις του προσώπου, οι σωματικές και φωνητικές εκφράσεις διαμορφώνονται από το μη-ορατό συναίσθημα, θα μπορούσε να επιτρέψει τη δημιουργία ευαίσθητων εικονικών πρακτόρων που να εκφράζουν συναισθήματα, οι οποίοι θα μπορούσαν να έχουν πολλές εφαρμογές στην ψυχαγωγία, την εκπαίδευση και την έρευνα.

1.2 Ενσωμάτωση του Πλαισίου στην Αναγνώριση Συναισθημάτων

Η ερμηνεία μιας συναισθηματικής εκδήλωσης συχνά απαιτεί πληροφορίες πέραν αυτών που περιλαμβάνονται στις τρέχουσες εμφανιζόμενες λεπτομέρειες, μια πληροφορία δηλαδή που να τονίζει την σημασία της ενσωμάτωσης του πλαισίου. Πράγματι, η έρευνα

στον τομέα της Ψυχολογίας υποδηλώνει ότι η κατανόηση της συναισθηματικής κατάστασης του ανθρώπου επηρεάζεται από το πλαίσιο, το οποίο με μία ευρεία έννοια μπορεί να αναφέρεται σε γλωσσικές και σημασιολογικές πληροφορίες, στο πολιτιστικό υπόβαθρο και το φύλο των συμμετεχόντων, στη γνώση του γενικού πλαισίου στο οποίο λαμβάνει χώρα μία συναισθηματική αλληλεπίδραση, στον τόπο κλπ. Για παράδειγμα, η βιβλιογραφία στον τομέα της Ψυχολογίας υποδηλώνει ότι οι πληροφορίες του προσώπου όταν εξετάζονται μεμονομένα μπορεί να μην είναι ικανές για να αποσαφηνίσουν το συναίσθημα που εκφράζεται και έτσι οι άνθρωποι τείνουν να χρησιμοποιούν ένα πλαίσιο, όπως προηγούμενες οπτικές πληροφορίες [62], τη γενική κατανόηση της κατάστασης [30], παρελθούσες λεκτικές πληροφορίες [119] είτε το πολιτισμικό τους υπόβαθρο [140] ώστε να πάρουν μια συναισθηματική απόφαση. Επιπλέον, τα συναισθήματα είναι συνήθως αργές μεταβαλλόμενες καταστάσεις και συνήθως διαρκούν λιγότερο από ένα λεπτό [159]. Επομένως, ένα συναίσθημα μπορεί να εκτείνεται σε πολλές διαδοχικές προτάσεις μιας συνομιλίας και οι συναισθηματικές μεταβάσεις να τείνουν να είναι ομαλές.

Στον τομέα της αναγνώρισης συναισθημάτων, είναι σχετικά μικρή η ερευνητική δραστηριότητα αναφορικά με την ενσωμάτωση του πλαισίου στα υπάρχοντα συστήματα. Συγκεκριμένα, στο [42] ο συγγραφέας περιγράφει ένα πλαίσιο για την δημιουργία ενός εικονικού πράκτορα διδασκαλίας, που θα εξετάζει μια ποικιλία μεταβλητών πλαισίου όπως την προσωπικότητα του μαθητή και το στόχο του εκπαιδευτή. Οι συγγραφείς του [125] προτείνουν ένα μονοτροπικό πλαίσιο για τη βραχυπρόθεσμη μοντελοποίηση των χρονικών αλληλουχιών στις δυαδικές αλληλεπιδράσεις, όπου η προηγούμενη πρόταση της ομιλίας του ομιλητή και του συνομιλητή του λαμβάνονται υπόψιν κατά την αναγνώριση των συναισθημάτων. Στην εργασία [126] ο διάλογος και μία σειρά λεκτικών χαρακτηριστικών χρησιμοποιούνται ως συμπληρωματική πληροφορία πλαισίου εκτός από τα ακουστικά (χρήση προσωδίας) χαρακτηριστικά. Στο [130] οι συγγραφείς κάνουν χρήση λεκτικών χαρακτηριστικών και χαρακτηριστικών προσωδίας και διαλόγου από τις τελευταίες δύο στροφές¹ αναγνωρίζοντας την τρέχουσα συναισθηματική κατάσταση του ομιλητή.

1.3 Μονοτροπικότητα και Πολυτροπικότητα

Τα συναισθήματα είναι σύνθετες δυναμικές διεργασίες που εκφράζονται μέσω πολλών καναλιών έκφρασης οι οποίες ενδέχεται να φέρουν συμπληρωματικές ή ακόμη και αντικρουόμενες πληροφορίες [144]. Η ανθρώπινη ομιλία μπορεί να εκφράσει πλούσια

¹ Στην ανάλυση των συνομιλιών, η στροφή (turn-taking) είναι ένας όρος για τον τρόπο με τον οποίο λαμβάνει χώρα μία συζήτηση. Μόλις επιλεγεί το θέμα και ξεκινήσει η συζήτηση, τότε πρέπει να ακολουθηθούν δύο κατευθυντήριες αρχές. Το να γνωρίζουμε πότε είναι αποδεκτό ή υποχρεωτικό να κάνει κάποιος από τους συνομιλητές μια στροφή στη συζήτηση. Προϋπόθεση αυτού είναι η γνώση του τρόπου αναγνώρισης των κατάλληλων σημείων ανταλλαγής και η γνώση της διάρκειας των παύσεων μεταξύ των στροφών. Είναι επίσης σημαντικό να γνωρίζουμε υπό ποιές συνθήκες επιτρέπεται η επικάλυψη ομιλίας. Δεδομένου ότι οι συνομιλίες πρέπει να ακολουθούν μία μορφή οργάνωσης, οι προαναφερθέντες κανόνες πρέπει να ακολουθούνται.

συναισθηματική πληροφορία: το συναίσθημα φαίνεται να επηρεάζει τον τόνο της φωνής, την ένταση και τα φασματικά χαρακτηριστικά του σήματος ομιλίας καθώς και την άρθρωση, την ποιότητα της φωνής και τον ρυθμό λόγου [75, 155, 203].

Ακολουθώντας αυτές τις κατευθύνσεις, ο τομέας της Συναισθηματικής Υπολογιστικής έχει υπογραμμίσει τη σημασία των πολυτροπικών πληροφοριών, προχωρώντας προς πολυτροπικές προσεγγίσεις της αναγνώρισης του συναισθήματος [209]. Αρκετές ερευνητικές προσπάθειες συναντώνται στη βιβλιογραφία που συνδυάζουν τα στοιχεία του προσώπου και της ομιλίας για την αναγνώριση των συναισθημάτων, χρησιμοποιώντας ποικίλες προσεγγίσεις συμπεριλαμβανομένων συστημάτων βασισμένα σε Hidden Markov Models (HMM) [50], multi-stream HMMs [265], Support Vector Machines (SVM) [26], αλγόριθμους Ada-Boost [265], προσαρμοστικά Νευρωνικά Δίκτυα (Neural Networks (NNs)) [168], Support Vector Regression (SVR) [107]. Οι πληροφορίες σχετικά με το λεκτικό περιεχόμενο μπορούν επίσης να συνδυαστούν με οπτικοακουστικό υλικό [72] όπου εφαρμόζεται μία αρχιτεκτονική NN χρησιμοποιώντας χαρακτηριστικά προσώπου, φωνής αλλά και λεκτικά χαρακτηριστικά. Τέλος, οι ερευνητές έχουν μελετήσει επίσης τη χρήση σημάτων φυσιολογίας για την αναγνώριση των συναισθημάτων. Οι συγγραφείς του [156] χρησιμοποίησαν μετρήσεις της θερμοκρασίας του σώματος του ατόμου, τη γαλβανική απόκριση του δέρματος και τον καρδιακό ρυθμό, ενώ οι συγγραφείς του [110] συνδύασαν σήματα φυσιολογίας (ηλεκτροκαρδιογράφημα, αγωγιμότητα του δέρματος (skin conductivity), θερμοκρασία κλπ.) με πληροφορίες ομιλίας. Επίσης, οι εργασίες [156, 110] χρησιμοποιούν είτε τον k-Nearest Neighbor (kNN) είτε τον Linear Discriminant (LDA) για την ταξινόμηση που επιθυμούν να πραγματοποιήσουν. Τέλος, μία αναλυτική έρευνα της πολυτροπικής αναγνώρισης συναισθημάτων αποτελεί η εργασία [46].

1.4 Σύνοψη

Η παρούσα εργασία οργανώνεται ως εξής: Το Κεφάλαιο 2 ασχολείται με τον ορισμό του πλαισίου, τη μοντελοποίησή του, τον εντοπισμό του στα υπάρχοντα σώματα κειμένου, εστιάζοντας στις προκλήσεις και στις μελλοντικές επεκτάσεις. Το Κεφάλαιο 3 παρουσιάζει μία επισκόπηση των αναπαραστάσεων του συναισθήματος στη Συναισθηματική Υπολογιστική και στον τομέα της Φυσικής Γλώσσας. Τα Κεφάλαια 4 και 5 αναφέρονται στην ανάλυση συναισθημάτων από πολυτροπικά χαρακτηριστικά και από το λεκτικό περιεχόμενο των ομιλητών αντίστοιχα. Συγχρόνως, το Κεφάλαιο 6 εισάγει το συνδυασμό δύο υπολογιστικών μεθόδων που υπογραμμίζουν τη συνεισφορά του εννοιολογικού πλαισίου αναφορικά με την ερμηνεία των απόψεων (opinions) και των συναισθημάτων (sentiments). Το Κεφάλαιο 7 αυτής της ερευνητικής προσπάθειας προτείνει τη χρήση τεχνικών Βαθιάς Μάθησης με τη χρήση Νευρωνικών Δικτύων (Deep Learning) για την ανάλυση και την ερμηνεία απόψεων και συναισθημάτων σε επίπεδο κειμένου, προτάσεων και χαρακτηριστικών. Τέλος, η διατριβή καταλήγει με το Κεφάλαιο 8 στο οποίο παρουσιάζονται τα συμπεράσματα της προτεινόμενης έρευνας.

Κεφάλαιο 2

Εννοιολογικό πλαίσιο

2.1 Ορισμός του πλαισίου

Κατά την προσπάθεια εντοπισμού ενός τυπικού ορισμού για συστήματα ανάλυσης του συναισθήματος (affect analysis systems) και εφαρμογές επίγνωσης πλαισίου (context-aware), ένα σημείο εκκίνησης θα ήταν η διερεύνηση του τρόπου με τον οποίο έχει οριστεί μέχρι στιγμής ο όρος “πλαίσιο”. Ο Schilit και οι συνεργάτες του [207] εισήγαγαν τον όρο context awareness στην πληροφορική, δίνοντας έμφαση στις σημαντικές πτυχές του πλαισίου: με ποιόν είσαι μαζί, πού βρίσκεσαι, ποιοί πόροι βρίσκονται σε κοντινή απόσταση. Έτσι, οι εφαρμογές επίγνωσης πλαισίου εξετάζουν το ποιός, πού, πότε και τί (ο χρήστης τί κάνει) και χρησιμοποιούν αυτές τις πληροφορίες για να προσδιορίσουν για ποιό λόγο συμβαίνει μία κατάσταση. Σύμφωνα με ένα παρόμοιο ορισμό, ο Brown και οι συνεργάτες του [24] ορίζουν το πλαίσιο ως το πού βρίσκονται οι άνθρωποι, τις ταυτότητές των ανθρώπων που περιβάλλουν το χρήστη, την ώρα της ημέρας, την εποχή, τη θερμοκρασία, κλπ. Άλλες προσεγγίσεις [194] αναφέρονται στο πλαίσιο ως η τοποθεσία του χρήστη, το περιβάλλον, η ταυτότητα και ο χρόνος, ενώ άλλοι χρησιμοποιούν συνώνυμα του όρου “πλαίσιο” αναφερόμενοι στο πλαίσιο είτε ως το περιβάλλον [24] είτε ως την κατάσταση που λαμβάνει χώρα [73].

2.2 Το Πλαίσιο στην Επικοινωνία Ανθρώπου-Μηχανής (HCI)

Με βάση την ευρύτερη προσέγγιση του Abowd [2], το πλαίσιο μπορεί να οριστεί ως ένας συνδυασμός των τεσσάρων τύπων πλαισίου (contextual types): ταυτότητα, ώρα, τοποθεσία και δραστηριότητα. Αυτό το μοντέλο πλαισίου εμπλουτίστηκε αργότερα [271] προκειμένου να συμπεριληφθούν πρόσθετοι τύποι πλαισίου που ονομάζονται “σχέσεις”. Στην ανθρώπινη επικοινωνία, η βιβλιογραφία υποδεικνύει ότι οι άνθρωποι αξιολογούν τις καταστάσεις με βάση πληροφορίες που προσδιορίζουν το πλαίσιο, όπως οπτικές πληροφορίες παρελθόντος χρόνου [62], τη γενικότερη κατανόηση της κατάστασης [30], τις λεκτικές πληροφορίες παρελθόντος χρόνου [119], το πολιτιστικό υπόβαθρο [140], το φύλο

των συμμετεχόντων, τις γνώσεις σχετικά με τη γενική ρύθμιση της αλληλεπίδρασης κατά την οποία ένα συναισθηματικό φαινόμενο λαμβάνει χώρα [24] αλλά και την κοινωνική κατάσταση [16]. Χωρίς την ύπαρξη του πλαισίου, ακόμη και οι άνθρωποι μπορούν να παρερμηνεύσουν τα παρατηρηθέντα συναισθηματικά συνθήματα όπως αυτά του προσώπου, της φωνής καθώς και των χειρονομιών αντίστοιχα. Όσον αφορά στον πραγματικό κόσμο, για εφαρμογές πλαισίου επίγνωσης (context-aware) στη Συναισθηματική Υπολογιστική, το πλαίσιο ορίζεται ως κάθε πληροφορία που μπορεί να χρησιμοποιηθεί για το χαρακτηρισμό της κατάστασης αναφορικά με την αλληλεπίδραση μεταξύ των χρηστών και του συστήματος. Έτσι, ο ορισμός που προτείνεται στην εργασία [207] προσεγγίζει καλύτερα την κατανόηση των ανθρωπίνων σημάτων. Ένας ακόμα καταλληλότερος ορισμός [60], συνοψίζει τις βασικές πτυχές του πλαισίου σε σχέση με την ανθρώπινη επικοινωνιακή συμπεριφορά από την άποψη της λεγόμενης γενικευμένης **W5+** διαμόρφωσης: ποιός (**Who**) εμπλέκεται (π.χ. δυαδική/τριαδική αλληλεπίδραση μεταξύ των ατόμων ή αλληλεπίδραση μεταξύ ενός ατόμου και ενός εικονικού χαρακτήρα), τί (**What**) ανακοινώνεται (π.χ. γλωσσικό μήνυμα, μη-γλωσσικό σήμα επικοινωνιακής μορφής, το συναίσθημα), πώς (**How**) μεταφέρονται οι πληροφορίες (έκφραση του προσώπου του ατόμου, η κίνηση του κεφαλιού, ο τόνος της φωνής, οι χειρονομίες των χεριών είτε του σώματος), γιατί (**Why**), δηλαδή, σε ποιό (**Which**) πλαίσιο οι πληροφορίες μεταφέρονται, πού (**Where**) είναι ο χρήστης, ποιά (**What**) είναι η τρέχουσα απασχόλησή του και πώς (**How**) αυτός/αυτή αισθάνεται καθώς και ποιές δράσεις πρέπει να ληφθούν για την ικανοποίηση των αναγκών και των απαιτήσεων των χρηστών. Δυστυχώς, μέχρι στιγμής οι προσπάθειες για την κατανόηση της ανθρώπινης συμπεριφοράς (human behavior understanding) είναι συνήθως ανεξάρτητες του πλαισίου (context-insensitive) [169], ενώ η προτεινόμενη μεθοδολογία αντιμετωπίζει μία ή περισσότερες **W5+** ερωτήσεις ξεχωριστά κάθε φορά, όπως το πού, τί [9], πότε, ποιός [229], γιατί και πώς [89] (**W5+**). Συνολικά, απαιτείται περαιτέρω έρευνα για την ταυτόχρονη απάντηση των **W5+** ερωτημάτων αλλά και για τη δημιουργία ενός ενοποιημένου πλαισίου διαμόρφωσης.

2.3 Μοντελοποίηση του Πλαισίου στη Συναισθηματική Υπολογιστική

Στη βιβλιογραφία αναφορικά με την ανάλυση του συναισθήματος, σχετικά λίγες ερευνητικές προσπάθειες κάνουν χρήση των πληροφοριών πλαισίου και σε γενικές γραμμές χρησιμοποιούνται διαφορετικές προσεγγίσεις για τη διαμόρφωση του πλαισίου. Στο [125], ένα μονοτροπικό πλαίσιο για τη βραχυπρόθεσμη μοντελοποίηση του πλαισίου σε δυαδική αλληλεπίδραση προτάθηκε, όπου το πλαίσιο ορίζεται ως το σύνολο των συνθημάτων ομιλίας του ομιλητή από το παρελθόν, ενώ σύμφωνα με το [86], στα πολυτροπικά πλαίσια αναγνώρισης συναισθήματος η έννοια του πλαισίου ορίζεται ως το σύνολο των οπτικοακουστικών συναισθηματικών εκφράσεων που προηγούνται και που ακο-

λουθούν, χρησιμοποιώντας αμφίδρομης μεγάλης και μικρής διάρκειας Νευρωνικά Δίκτυα (BLSTM). Συνολικά, από την άποψη των προσεγγίσεων της αναγνώρισης των συναισθημάτων, οι εξαρτώμενες από το πλαίσιο αμφίδρομης κατεύθυνσης προσεγγίσεις έχει αποδειχτεί ότι μπορούν να ξεπεράσουν τις μεθόδους που δεν εξετάζουν το πλαίσιο όπως τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNNs) και τα κρυφά μαρκοβιανά μοντέλα (HMM). Ωστόσο, τα οριστικά συμπεράσματα εξαρτώνται από τον σχεδιασμό των σωμάτων κειμένου και ενδεχομένως να μην καλύπτουν όλο το φάσμα της ανθρώπινης συναισθηματικά ενισχυμένης επικοινωνίας από την άποψη της μοντελοποίησης της αλληλεπίδρασης. Επιπλέον, όσον αφορά στην ποσοτικοποίηση της επίδρασης του πλαισίου η Kalliouby και οι συνεργάτες της [62] περιγράφουν την έκφραση του συναισθήματος ως μία ακολουθία των ταυτόχρονων διαδοχικών δράσεων του προσώπου, είτε ως μικροεκφράσεις, όπου εδώ το πλαίσιο ορίζεται ως η σχέση μεταξύ των διαδοχικών μικροεκφράσεων. Τέλος, στο [93] περιγράφεται ένα σύστημα επίγνωσης πλαισίου που σχετίζεται με την αναγνώριση του πόνου, όπου οι τύποι πλαισίου (contextual types) ορίζονται ως ο τόπος, η εργασία, κλπ., ενώ για εφαρμογές όπως οι προσωπικές συλλογές φωτογραφιών υποβοηθούμενης επισημείωσης [128] οι συναφείς πληροφορίες παρέχονται μέσω των σχέσεων μεταξύ των ανθρώπων, των γεγονότων και των τοποθεσιών. Και στις δύο εφαρμογές, που βασίζονται σε Transferable Belief Model (TBM) μοντέλα και σε πιθανοτικά μοντέλα αντίστοιχα, σημαντικές βελτιώσεις στην απόδοση έχουν παρατηρηθεί.

2.4 Αναπαράσταση Συναισθήματος σύμφωνα με το Πλαίσιο

Υπάρχει ένα αυξανόμενο ενδιαφέρον για την κατανόηση των οφελών αλλά και των περιορισμών των εναλλακτικών θεωρητικών μοντέλων που περιγράφουν τα συναισθήματα και τις σχετικές αναπαραστάσεις τους. Τρία είναι τα κυρίαρχα θεωρητικά μοντέλα συναισθήματος στον τομέα της Συναισθηματικής Υπολογιστικής: αναπαράσταση σε κατηγορίες, οι θεωρίες διαστατικών συναισθημάτων και οι θεωρίες αξιολόγησης [85, 266]. Ωστόσο, αναφορικά με την καταλληλότητά τους ως προς τη μοντελοποίηση του πλαισίου, δίνεται έμφαση στα μοντέλα ενσυναίσθησης που βασίζονται στη γνωστική εκτίμηση, που χαρακτηρίζουν τις συναισθηματικές καταστάσεις σύμφωνα με τις συνθήκες εξαγωγής των δεδομένων. Αυτές οι συνθήκες περιλαμβάνουν την εξοικείωση, την ικανοποίηση, τη συνάφεια με τους στόχους κάποιου να χαρακτηρίζει την αιτία ή το αντικείμενο ενός συναισθήματος, όπως προκύπτει από το πλαίσιο, καθώς και την πρόβλεψη των συναισθημάτων στα συστήματα Τεχνητής Νοημοσύνης. Η πλειοψηφία αυτών των βαθύτερων προσεγγίσεων υιοθετούν τη θεωρία των Orthony, Clore, και Collins (OCC) [161] είτε το μοντέλο του Scherer [205], ωστόσο, το OCC μοντέλο κρίνεται περισσότερο κατάλληλο μεταξύ των δύο, ενώ έχει χρησιμοποιηθεί και στη σύνθεση συναισθήματος (affect synthesis) για το σχεδιασμό “χαρακτήρων συνομιλητή με σωματική υπόσταση” [266].

Περισσότερο πρόσφατες ερευνητικές απόπειρες υποστηρίζουν ότι ένα άλλο σύνολο ψυχολογικών μοντέλων, που αναφέρεται ως “comproential” μοντέλο του συναισθήματος και βασίζεται στη θεωρία αξιολόγησης, μπορεί να είναι καταλληλότερο για την ανάπτυξη συστημάτων επίγνωσης πλαισίου (context-aware) [151]. Σύμφωνα με τη θεωρία αξιολόγησης, τα συναισθήματα δημιουργούνται μέσα από τη συνεχή, αναδρομική, υποκειμενική εκτίμηση τόσο της δική μας εσωτερικής κατάστασης αλλά και της κατάστασης του έξω κόσμου (ανησυχίες/ανάγκες) [85]. Παρά τις πρωτοποριακές προσπάθειες του Scherer και των συνεργατών του (π.χ., [198]), η χρήση της μεθόδου εκτίμησης που βασίζεται στην αυτόματη μέτρηση του συναισθήματος παραμένει ένα ανοικτό ερώτημα της έρευνας. Στην πρόσφατη δουλειά του [151], παρουσιάζεται μία προσπάθεια να συνδεθεί η αυτόματη αναγνώριση των συναισθημάτων με τα μοντέλα εκτίμησης του συναισθήματος σύμφωνα με την οποία μια σειρά από οφέλη αναμένεται να προκύψει αναφορικά με την αυτόματη αναγνώριση συναισθημάτων.

2.5 Εντοπισμός του Πλαισίου στα Υπάρχοντα Σώματα Κειμένων

Έχοντας σχεδιασμένα και επισημειωμένα σώματα κειμένων από την άποψη του πλαισίου θα διευκόλυνε τη διαδικασία διαμόρφωσης πλαισίου και την ενσωμάτωσή του σε αρχιτεκτονικές συναισθηματικής υπολογιστικής. Η διαδικασία επισημείωσης επιβεβαιώνει ότι το πλαίσιο είναι πράγματι πολύ δύσκολο να αποσπαστεί από δεδομένα αυθόρμητης μορφής (naturalistic data), καθώς σε αυτές τις ρυθμίσεις αλληλεπίδρασης κάθε ορισμός θα μπορούσε να αντικατοπτρίζει διαφορετικές υποθέσεις σχετικά με την επισημείωση σχετικών πληροφοριών. Επιπλέον, λόγω του γεγονότος ότι δεν υπάρχει ένα συμφωνημένο πρωτόκολλο απόκτησης δεδομένων, η εφαρμογή του οποίου θα παρείχε βελτιωμένα αποτελέσματα, τόσο ο εντοπισμός όσο και η εξόρυξη σχετικής πληροφορίας με το πλαίσιο γίνεται ακόμα πιο δύσκολη. Ακόμη και για έναν ειδικό, είναι δύσκολο να ορίσουμε τι συνιστά το “πλαίσιο” σε συνδυασμό με το συναίσθημα. Μέχρι στιγμής, δεν υπάρχουν αποθήκες (repositories) σωμάτων κειμένου που να εξαρτώνται από το πλαίσιο και που θα μπορούσαν να βοηθήσουν στη διαμόρφωση της κατανόησης του ίδιου του προβλήματος ρίχνοντας φως σε βασικά προβλήματα της επισημείωσης όσον αφορά στην ενσωμάτωση του πλαισίου. Οι λίγες εξαιρέσεις, που ικανοποιούν τις παραπάνω ανάγκες σχετίζονται με την αλληλεπίδραση μεταξύ του ανθρώπου και ενός εικονικού πράκτορα. Για παράδειγμα, στο σώμα κειμένου SEMAINE [143], η όλη διαδικασία απόκτησης δεδομένων καθοδηγείται από το ακόλουθο πλαίσιο: ο χρήστης κάθεται, κοιτάζοντας τον πράκτορα, έχει ένα συγκεκριμένο στυλ και διάθεση, ο χρήστης δεν είναι επιτρέπεται να υποβάλει ερωτήσεις, ο πράκτορας κάνει τις ερωτήσεις, κλπ.

2.6 Ενσωμάτωση του Πλαισίου στην Συναισθηματική Υπολογιστική

Μία σειρά από τεχνικές Μηχανικής Μάθησης έχουν χρησιμοποιηθεί για την αναγνώριση συναισθήματος. Ειδικά, όσον αφορά στις εφαρμοζόμενες στρατηγικές ταξινόμησης, ο κύριος περιορισμός των μέχρι τώρα μοντέλων είναι ότι η ταξινόμησή τους γίνεται μη λαμβάνοντας υπόψιν το πλαίσιο μέσα στο οποίο πραγματοποιούνται. Από την άποψη των φωνητικών και οπτικών σημάτων και συνθημάτων κειμένου (textual cues), λίγες προσπάθειες έχουν γίνει για την ενσωμάτωση των ηχητικών και οπτικών χαρακτηριστικών που να εξαρτώνται από το πλαίσιο και να χρησιμοποιούν Dynamic Bayesian Networks (DBN) για τη μοντελοποίηση της συναισθηματικής κατάστασης δύο συνομιλητών [125]. Επιπλέον, στο [86], τα δίκτυα BLSTM χρησιμοποιήθηκαν για την ενσωμάτωση μεγάλου όγκου πληροφοριών πλαισίου. Επιπλέον, από την άποψη των οπτικών σημάτων, η ερμηνεία της παρατηρούμενης έκφρασης του προσώπου [108], έχει ήδη επιχειρηθεί στο παρελθόν. Οι εκφράσεις του προσώπου εμφανίζονται κατά συνέπεια σε ένα συγκεκριμένο πλαίσιο, όπως η θέση (εξωτερική, εσωτερική), η κατάσταση (όταν νοσηλεύονται σε νοσοκομείο), το υποβαλλόμενο έργο, οι άλλοι άνθρωποι, η ταυτότητα και η προσωπικότητα [167, 266] του ομιλητή. Ωστόσο, μέχρι στιγμής, δεν υπάρχει κάποιο μοντέλο όρασης που να συμπεριλαμβάνει το πλαίσιο κατά τη διαδικασία αναγνώρισης της αυθόρμητης έκφρασης. Μια πρόσφατη εξαίρεση είναι η εργασία [92], όπου το μοντέλο TBM χρησιμοποιείται για την εύκολη προσθήκη της μεταβλητής πλαισίου (contextual type) “θέση” στο μοντέλο ταξινόμησης ανθρωπίνων εκφράσεων κατά την εφαρμογή της αναγνώρισης του πόνου.

2.7 Προκλήσεις και Μελλοντικές Επεκτάσεις

Η ενσωμάτωση του πλαισίου στη συναισθηματική υπολογιστική βρίσκεται ακόμη σε αρχικό στάδιο. Παρά το αναπτυσσόμενο ενδιαφέρον για τα ενδιάμεσα στάδια της επικοινωνίας ανθρώπου-μηχανής (HCI), δεν έχει ακόμη εξεταστεί η καταλληλότητά τους ως προς την ενσωμάτωση του πλαισίου στα στάδια της εξαγωγής (elicitation), της ερμηνείας (interpretation) καθώς και της αναγνώρισης (recognition) συναισθήματος. Τα στάδια αυτά περιλαμβάνουν μία σειρά προκλήσεων όπως η μοντελοποίηση των γνωστικών θεωριών του πλαισίου με βάση την συναισθηματική αλληλεπίδραση, η άντληση πληροφοριών από το πλαίσιο, η ενσωμάτωση των πληροφοριών πλαισίου στα σώματα κειμένου ενσυναίσθησης, ο εντοπισμός των δομών πλαισίου (contextual types) καθώς και η ενσωμάτωσή τους για τη βελτίωση των επιδόσεων των πολυτροπικών εφαρμογών. Κατά συνέπεια, τα προαναφερθέντα ζητήματα είναι στενά συσχετιζόμενα με αυτές τις καταστάσεις και ως εκ τούτου είναι επιβεβλημένος ο συνυπολογισμός τους στην διαδικασία εξαγωγής συναισθηματικών παραμέτρων και αναγνώρισης συναισθήματος. Για τους λόγους αυτούς όπως αναφέρθηκε και προηγουμένως απαιτείται περαιτέρω έρευνα για την απάντηση των

ερωτημάτων της **W5+** διαμόρφωσης του πλαισίου. Κάτω από αυτές τις συνθήκες, θα μπορούσαμε να έχουμε μία πρώτη εικόνα για το κατά πόσο είναι αναγκαίο η επισημείωση των κειμένων να πραγματοποιείται σύμφωνα με το εξαρτώμενο πλαίσιο καθώς και κατά πόσο τα υπάρχοντα σώματα κειμένων πρέπει να επαναεπισημειωθούν είτε να ηχογραφηθούν από την αρχή. Αναφορικά με τη διαδικασία επισημείωσης, οι προκλήσεις της επίτευξης αυτού του στόχου ξεκινούν από την εφαρμογή ενός γενικού πλαισίου για τη συλλογή δεδομένων αυθόρμητης συμπεριφοράς. Το ερευνητικό πεδίο είναι ακόμη στα αρχικά του στάδια και δεν έχουν παρατηρηθεί σημαντικές προσπάθειες για τη συλλογή αυθόρμητων δεδομένων εξαρτώμενα από το πλαίσιο ειδικά σχεδιασμένα για την ανάλυση τύπων πλαισίου (contextual types). Οι περισσότερες ερευνητικές δουλειές μέχρι στιγμής βασίζονται στη χρήση σωμάτων κειμένου που είναι ήδη επισημειωμένα, είτε επισημειώνονται κάθε φορά για την ικανοποίηση των αναγκών του εκάστοτε πειράματος. Παράλληλα, κατά τη διάρκεια αλληλεπίδρασης μίας ομάδας περιλαμβάνεται μια πληθώρα από διαφορετικά “συνθήματα” και μέχρι στιγμής δεν υπάρχει ένα πρότυπο σχολιασμού ή ένα πρωτόκολλο συλλογής δεδομένων που να είναι εύκολα εφαρμόσιμο. Επιπλέον, ο συνδυασμός της προοπτικής της αξιολόγησης γεγονότων θα μπορούσε να παρέχει μία σειρά από οφέλη, όπως η βελτιωμένη αναγνώριση των μικτών συναισθημάτων, η ικανότητα ενσωμάτωσης συναφών πληροφοριών σε αυτόματες αρχιτεκτονικές συναισθηματικής ανάλυσης καθώς και ένα καλύτερο επίπεδο ερμηνείας του συναισθήματος. Ως εκ τούτου, χωρίς σημαντική προσπάθεια για τη συλλογή και την επισημείωση δεδομένων θα είναι δύσκολο αν όχι αδύνατο να αντιμετωπιστούν τα τρέχοντα προβλήματα της έρευνας στον τομέα αυτό.

Κεφάλαιο 3

Αναπαραστάσεις Συναισθήματος

3.1 Αποσαφήνιση Όρων

Λαμβάνοντας υπόψιν μας το διεπιστημονικό χαρακτήρα της Συναισθηματικής Υπολογιστικής (Affective Computing), κρίνεται απαραίτητο στο σημείο αυτό μία αναλυτική αποσαφήνιση των όρων που θα χρησιμοποιηθούν στη συνέχεια. Είναι συχνό φαινόμενο να παρατηρούνται αμφισημίες στη χρήση βασικών όρων ανάλογα με την επιστημονική σκοπιά του εκάστοτε κειμένου. Σε αυτή τη φάση, η αποσαφήνιση που προτείνεται δεν έχει σκοπό να διακρίνει σωστές από λανθασμένες χρήσεις όρων, αλλά να καθορίσει με σαφήνεια την ερμηνεία κάθε όρου καθώς και το εννοιολογικό πλαίσιο όπου αυτό χρησιμοποιείται. Στο ερευνητικό πεδίο της Ψυχολογίας, είναι πολύ διαδεδομένη η χρήση του αγγλικού όρου **affect**, ο οποίος συνήθως καλύπτει μία πληθώρα από έννοιες όπως είναι τα συναισθήματα, οι διαθέσεις και οι προτιμήσεις, ενώ είναι συχνά συμπληρωματικός με αγγλικούς όρους όπως **emotion, feeling, sentiment, opinion** [71]. Η ανάγκη για την αποσαφήνιση είναι προφανής από το γεγονός ότι η ερώτηση “Τί είναι συναίσθημα;” δεν έχει μοναδική απάντηση.

Affect: Υπάρχει η άποψη ότι το συναίσθημα ορίζεται από τις φυσιολογικές αντιδράσεις που προκαλούνται στο σώμα (εφίδρωση, αύξηση καρδιακών παλμών κοκ), άλλοι διατείνονται ότι πρόκειται για μία καθαρά νοητική διαδικασία, ενώ ιδιαίτερα διαδεδομένες είναι οι υβριδικές θεωρίες, όπου καθεμία ορίζει τον τρόπο και το βαθμό με τον οποίο συμμετέχει η κάθε ανθρώπινη λειτουργία στη συναισθηματική εμπειρία.

Από την οπτική της Γνωστικής Νευροεπιστήμης (Cognitive Neuroscience), ο Damasio διαχωρίζει τα συναισθήματα, δηλαδή τις σωματικές καταστάσεις που μπορούν να παρατηρηθούν δημοσίως, από τα αισθήματα (feelings) δηλαδή τα νοητικά γεγονότα που γίνονται αντιληπτά μόνο από το άτομο που τα βιώνει [48]. Σχετικές έρευνες ισχυρίζονται ότι ένα συναισθηματικό επεισόδιο ξεκινάει από ένα συναισθηματικά ικανό ερέθισμα (όπως ένα τρομακτικό γεγονός) το οποίο ο οργανισμός αυτόματα αξιολογεί είτε σαν γεγονός που συντελεί στην επιβίωση/ευεξία είτε σαν γεγονός που δεν συνδράμει στην επίτευξη των επιθυμητών στόχων. Η αξιολόγηση λαμβάνει χώρα ως μια σύνθετη αλληλουχία από

σωματικές νευροφυσιολογικής φύσεως αντιδράσεις (π.χ. αύξηση παλμών, ένταση στους μύες του προσώπου), η οποία καταγράφεται στον εγκέφαλο. Από την καταγραφή αυτή αναδύεται η αίσθηση του σώματος όταν βιώνει ένα συναίσθημα.

Σύμφωνα με τους συγγραφείς του [11], βρέθηκε ότι το συναίσθημα είναι ο γενικότερος ορισμός μεταξύ όλων. Συγκεκριμένα, ο Shouse [215] ορίζει το συναίσθημα ως μία μη συνειδητή εμπειρία έντασης, δηλαδή ως μία στιγμή μη δομημένης δυναμικής προσπάθειας. Ταυτόχρονα, προσθέτει ότι το συναίσθημα παίζει ένα σημαντικό ρόλο στον προσδιορισμό της σχέσης μεταξύ των ατόμων, του περιβάλλοντός μας και μεταξύ άλλων παραγόντων. Έτσι, το συναίσθημα ορίζεται ως η θετικής και αρνητικής κλίμακας αξιολόγηση ενός αντικειμένου είτε μιας συμπεριφοράς [232]. Ο Zajonc [264] επιπρόσθετα συμφώνησε ότι το συναίσθημα αποκαλύπτει τις προσωπικές προτιμήσεις του ατόμου, ενώ και η ένταση με την οποία βιώνει τη συγκεκριμένη κατάσταση είναι ενδεικτική της προτίμησής του [11].

Feeling: Το αίσθημα (Feeling) ορίζεται ως η αίσθηση που έχει ήδη ελεγχθεί από προηγούμενες εμπειρίες. Λαμβάνοντας υπόψιν ότι κάθε άτομο έχει ένα ξεχωριστό σύνολο προηγούμενων αισθήσεων, είναι προφανές ότι η ερμηνεία και η επισημείωση των συναισθημάτων τους θα είναι υποκειμενική. Επιπλέον, ο Thoits [232] δήλωσε ότι τα συναισθήματα είναι η απόρροια των ανθρώπινων εμπειριών των φυσικών καταστάσεων όπως η πείνα, ο πόνος, η κόπωση και ότι οι συναισθηματικές καταστάσεις είναι δηλαδή μέρος των προσωπικών μας “Ψυχολογικών αποσκευών” [71]. Αντίστοιχα, οι Friedenberg και Silverman [76] πρόσθεσαν ότι τα αισθήματα αντιστοιχούν στην υποκειμενική εμπειρία ενός συναισθήματος. Η άποψη αυτή επιβεβαιώθηκε αργότερα και από την Wierzbicka [256]. Η τελευταία υποστήριξε ότι τα αισθήματα είναι υποκειμενικά και εμφανίζονται ως σκέψεις του τί συμβαίνει, αλλά και ως σκέψεις που βασίζονται σε ορισμένες επαναλαμβανόμενες σκέψεις-γνωστικά σενάρια που διαμορφώνονται από την κουλτούρα του κάθε ανθρώπου. Την ίδια στιγμή οι συγγραφείς του [256] επεσήμαναν επιπλέον ότι η έννοια των αισθημάτων είναι διεθνής (universal) καθώς όλες οι γλώσσες φαίνεται να έχουν μια λέξη για την έννοια “αίσθημα”.

Emotions: Στον τομέα της Φυσικής Γλώσσας οι Kleinginna και Kleinginna [118] πρότειναν έναν επίσημο ορισμό του συναισθήματος ως ένα πολύπλοκο σύνολο αλληλεπιδράσεων μεταξύ υποκειμενικών και αντικειμενικών παραγόντων, με τη μεσολάβηση του νευρικού συστήματος, που μπορεί α) να προκαλέσει συναισθηματικές εμπειρίες όπως συναισθήματα εγρήγορσης, ευχαρίστησης και δυσαρέσκειας και β) να παράγει γνωστικές διεργασίες όπως οι εκτιμήσεις (evaluations) και οι διαδικασίες αξιολόγησης (reviews).

Στον τομέα της Ψυχολογίας ο όρος συναίσθημα τείνει να χρησιμοποιείται για το χαρακτηρισμό σχετικά σύντομων και έντονων εμπειριών, ενώ διαφοροποιείται από τις διαθέσεις και τις προτιμήσεις που αναφέρονται σε χαμηλότερης έντασης αλλά μεγαλύτερης διάρκειας εμπειρίες [193]. Η διάθεση ενυπάρχει στο συναίσθημα και το χρωματίζει, αλλά σε αντίθεση με το συναίσθημα που αποτελεί συγκεκριμένη αντίδραση, η διάθεση είναι πιο

γενική και δεν χρειάζεται να έχει προέλθει από κάποιο συγκεκριμένο ερέθισμα ([191]). Στο έργο του για την Ψυχολογία των συναισθημάτων, ο Scherer [204] απέδωσε στη σύγκριση του ορισμού των συναισθημάτων την αποτυχία της σαφούς διάκρισης μεταξύ των διαφόρων κατηγοριών/ομάδων φαινομένων. Συγκεκριμένα, ο Scherer συνέστησε ότι το γενικό συναισθηματικό σθένος ή η προτίμηση δεν θα πρέπει να αντιμετωπίζεται με τον ίδιο τρόπο όπως τα συναισθηματικά επεισόδια, ούτε θα πρέπει να είναι μεγαλύτερης διάρκειας σε σύγκριση με συναισθηματικές καταστάσεις όπως οι συμπεριφορές (attitudes). Ωστόσο, τα συναισθήματα όπως μία κατηγορία συναισθηματικής κατάστασης έχει σθένος (valence) και ένταση (intensity) [11]. Την ίδια στιγμή, ο Scherer όρισε την έννοια του συναισθήματος ως επεισόδια συντονισμένων αλλαγών σε πολλά σημεία (συμπεριλαμβανομένης τουλάχιστον της νευροφυσιολογικής ενεργοποίησης, των εκφράσεων της κίνησης, τα υποκειμενικά συναισθήματα αλλά ενδεχομένως και των γνωστικών διαδικασιών) ως απάντηση σε εξωγενή ή ενδογενή γεγονότα μείζονος σημασίας για τον οργανισμό.

Αντίστοιχα, οι Thoits [232] καθόρισαν τα συναισθήματα (Emotions) ως πολιτισμικές κατηγορίες των συναισθημάτων. Ο Dolan [56] υποστήριξε ότι από ψυχολογικής απόψεως, τα συναισθήματα έχουν τρία χαρακτηριστικά: αρχικά, αντίθετα από τις σκέψεις, τα συναισθήματα είναι ενσωματωμένα, δηλαδή βιώνουμε συναισθήματα όχι μόνο ως ψυχικά γεγονότα που βρίσκονται στο μυαλό μας αλλά και ως εμπειρίες μέσω της κίνησης του σώματός μας. Δεύτερον, σε αντίθεση με τη γνώση, τα συναισθήματα είναι περισσότερο δύσκολο να ελεγχθούν. Μπορεί να διαπιστώσουμε ότι είναι πιο εύκολο να αλλάξουμε τις σκέψεις μας παρά τα συναισθήματά μας. Τρίτον, τα συναισθήματα φαίνονται λιγότερο εγκλωβισμένα από τις σκέψεις. Για το λόγο αυτό φαίνεται να έχουν και μεγαλύτερο αντίκτυπο στη συμπεριφορά μας.

Sentiment: Τα συναισθήματα (Sentiments) ορίζονται από την Cattell [32] ως μία αποκτηθείσα και σχετικά μόνιμη και σημαντική νευροψυχική διάθεση στο να αντιδρούμε συναισθηματικά, διανοητικά και αντιφατικά προς ένα συγκεκριμένο αντικείμενο ή κατάσταση με ορισμένο και σταθερό τρόπο, συνειδητοποιώντας τον τρόπο αντίδρασης μας ως προς το συγκεκριμένο αντικείμενο. Παρόμοιος ορισμός έχει δοθεί και από τον Gordon [84], σύμφωνα με την έρευνα του οποίου τα συναισθήματα είναι κοινωνικά κατασκευασμένα πρότυπα αισθήσεων, εκφραστικές χειρονομίες και πολιτιστικές έννοιες που οργανώνονται γύρω από μία σχέση με ένα κοινωνικό αντικείμενο, είτε με ένα άλλο άτομο.

Επιπρόσθετα, η Cattell [32] επεσήμανε περαιτέρω ότι η λειτουργία του συναισθήματος συνοδεύεται στο άτομο από ενδιαφέρον και μία αίσθηση, από την πλευρά του για τις “αξίες” του αντικειμένου. Συνεπώς, το αντικείμενο κατέχει περισσότερο ή λιγότερο μία διαρκή δύναμη να προκαλεί σχετικά έντονες και συχνές αντιδράσεις, θετικές ή αρνητικές σε ένα άτομο [154]. Τέλος, ο Broad [22] προσπάθησε να εξηγήσει ότι τα συναισθήματα δημιουργούνται όταν ένα συγκεκριμένο αντικείμενο γίνεται συνεχώς αντιληπτό από ένα άτομο και με την πάροδο του χρόνου, το άτομο αυτό δημιουργεί μια προδιάθεση (θετικά/αρνητικά προσκείμενος) προς το αντικείμενο αυτό.

Opinions: Οι γνώμες σύμφωνα με την έρευνα είναι συνώνυμες με τις σκέψεις που πραγματοποιεί ένα άτομο. Μια άποψη/γνώμη είναι η ιδέα ή η γνώση που έχει ένα άτομο για κάτι. Σύμφωνα με την εργασία των Pang και Lee [165], μία γνώμη οδηγεί σε ένα συμπέρασμα που έχει μελετηθεί ακόμη και εάν αυτή ενδέχεται να αλλάξει υπό το πρίσμα νέων στοιχείων. Έτσι, οι απόψεις αφήνουν περιθώρια για σφάλματα στη σκέψη, αφού θεωρούνται προσωρινές αποφάσεις, πεποιθήσεις, είτε πιθανές απόψεις. Οι πεποιθήσεις, όπως περιγράφονται στην εργασία των Fishbein και Ajzen [70], αντιστοιχούν στις υποκειμενικές αντιλήψεις ότι ένα αντικείμενο ή ένα πρόσωπο έχει συγκεκριμένα χαρακτηριστικά ή ιδιότητες. Σύμφωνα με τους Kim και Hony [111], μία γνώμη αποτελείται από τα ακόλουθα τέσσερα μέρη: θέμα, κάτοχος γνώμης, αξίωση και συναίσθημα. Δηλαδή, για κάθε γνώμη, υπάρχει ο κάτοχός της που πιστεύει μια αξίωση για ένα θέμα και στη συνέχεια συνδέει ένα θετικό, αρνητικό ή ουδέτερο ¹ συναίσθημα με την αξίωση αυτή.

Τέλος, στα πλαίσια της προσπάθειας του Liu [132] να αναπαραστήσει τον όρο “γνώμη” με μαθηματικό τρόπο, χρησιμοποίησε το ακόλουθο μοντέλο: $\langle o, f, so, h, t \rangle$, όπου o είναι ένα αντικείμενο; f είναι ένα χαρακτηριστικό του αντικειμένου o ; so είναι ο προσανατολισμός ή η πολικότητα της γνώμης για το χαρακτηριστικό f του αντικειμένου o (θετικό, αρνητικό ή ουδέτερο); h είναι κάτοχος γνώμης; και t είναι η χρονική στιγμή της γνώμης που εκφράζεται.

Διαπιστώνουμε λοιπόν, ότι υπάρχει μία σειρά παραγόντων που διαφοροποιούν τις έννοιες που αναπτύχθηκαν προηγουμένως. Συγκεκριμένα:

- Τα συναισθήματα (**Affects**) είναι στην πραγματικότητα, ένας προκάτοχος των αισθημάτων (**Feelings**) και των σχετικά σύντομων και έντονων εμπειριών (**Emotions**).
- Τα Αισθήματα (**Feelings**) είναι επικεντρωμένα στο άτομο και είναι συνειδητά φαινόμενα.
- Οι σύντομη και έντονη διάρκεια εμπειρίες (**Emotions**) είναι προγνωστικές κοινωνικές εκφράσεις των αισθημάτων (**Feelings**) και επηρεάζονται από την πολιτισμική κουλτούρα.
- Οι εκφράσεις του συναισθήματος στον τομέα της Φυσικής Γλώσσας (**Sentiments**) είναι εν μέρει κοινωνικές κατασκευές των συναισθηματικών εμπειριών (**Emotions**) που αναπτύσσονται με την πάροδο του χρόνου και διαρκούν.
- Τέλος, οι απόψεις (**Opinions**) είναι προσωπικές ερμηνείες των πληροφοριών και μπορεί να είναι συναισθηματικά ή μη φορτισμένες.

¹ουδέτερο εδώ δε σημαίνει απουσία, π.χ. “Ο χειμώνας έφτασε. Δεν είναι καλό ή κακό.”

3.2 Αναπαράσταση Συναισθήματος στη Συναισθηματική Υπολογιστική

3.2.1 Αναπαράσταση σε Κατηγορίες

Μετά από δεκαετίες έρευνας, εξακολουθεί να υπάρχει έντονη συζήτηση στην ερευνητική κοινότητα της Ψυχολογίας σχετικά με τον ορισμό και τις ιδιότητες του συναισθήματος. Ορισμένοι ερευνητές υποστηρίζουν ότι υπάρχει ένα μικρό σύνολο βασικών (basic) συναισθημάτων που έχουν αναπτυχθεί μέσω της ανθρώπινης εξέλιξης και αναγνωρίζονται παγκοσμίως. Ο Plutchik και οι συνεργάτες του [176] περιγράφουν μια λίστα με 8 βασικά συναισθήματα: φόβος (fear), θυμός (anger), ευτυχία (happiness), θλίψη (sadness), αποδοχή (acceptance), αηδία (disgust) και έκπληξη (surprise), ο συνδυασμός των οποίων είναι δυνατόν να οδηγήσει στη δημιουργία περισσότερο πολύπλοκων συναισθημάτων. Κάθε ένα από τα προαναφερθέντα συναισθήματα προέρχονται από μία βασική εξελικτική ανάγκη, π.χ. ο φόβος ώστε να ξεφύγουμε από μία απειλή, η ευτυχία μετά την απόκτηση ενός επιθυμητού αντικειμένου, η οργή κατά την αντιμετώπιση ενός εμποδίου κλπ. Οι Ekman και Friesen [61] καθόρισαν ένα σύνολο 6 βασικών συναισθημάτων. Μεταξύ αυτών εντάσσονται ο θυμός, η αηδία, ο φόβος, η χαρά, η θλίψη και η έκπληξη, τα οποία μπορούν να εντοπιστούν και να αναγνωριστούν παγκοσμίως ανεξαρτήτως της πολιτισμικής ταυτότητας και της κουλτούρας του κάθε λαού. Αργότερα, η εργασία αυτή επεκτάθηκε από τον Ekman συμπεριλαμβάνοντας αυτή τη φορά συναισθηματικές καταστάσεις όπως η διασκέδαση, η αμηχανία, η ενοχή, κλπ. [47]. Άλλοι ερευνητές έχουν προτείνει εναλλακτικά σύνολα βασικών συναισθημάτων, που κυμαίνονται από δύο τουλάχιστον συναισθήματα (π.χ. πόνο και ευχαρίστηση όπως προτάθηκε από τον Mowrer [152]) σε πολύ περισσότερα [162]. Συνήθως τα σύνολα αυτά περιλαμβάνουν συναισθήματα όπως ο φόβος, θυμός, θλίψη και χαρά. Ωστόσο, η θεωρία των βασικών συναισθημάτων αμφισβητείται έντονα από άλλους ερευνητές, π.χ., Ortony και Turner [162], οι οποίοι υποστηρίζουν ότι τέτοια συναισθήματα τείνουν να χαρακτηρίζονται ως βασικά επειδή απλώς είναι περισσότερο συνηθισμένα.

Ακολουθώντας αυτές τις κατευθύνσεις τα συστήματα αναγνώρισης συναισθήματος στη βιβλιογραφία είναι συχνά σχεδιασμένα με τέτοιο τρόπο ώστε να ταξινομούν τις συναισθηματικές εκδηλώσεις στις αντίστοιχες συναισθηματικές κατηγορίες, π.χ. [265]. Ωστόσο, από υπολογιστικής άποψης, τα συναισθήματα που επιλέγουμε να εξετάσουμε κατά το σχεδιασμό ενός συστήματος αναγνώρισης συναισθήματος εξαρτώνται συχνά από την εφαρμογή και το σκοπό της, οπότε στην περίπτωση αυτή θα πρέπει να λαμβάνεται υπόψιν και το πλαίσιο μέσα στο οποίο θα πραγματοποιηθεί ο στόχος της εφαρμογής. Για παράδειγμα, η αναγνώριση ενός θετικού έναντι ενός αρνητικού συναισθήματος σε ένα τηλεφωνικό κέντρο (πλαίσιο) [126] είναι εύκολο να πραγματοποιηθεί. Στην περίπτωση όμως ενός εκπαιδευτικού συστήματος, όπως αυτό που προτάθηκε στο [172], στόχος του οποίου είναι η ταξινόμηση μιας παρατήρησης σε μία από τις ακόλουθες τρεις συναισθηματικές κατηγο-

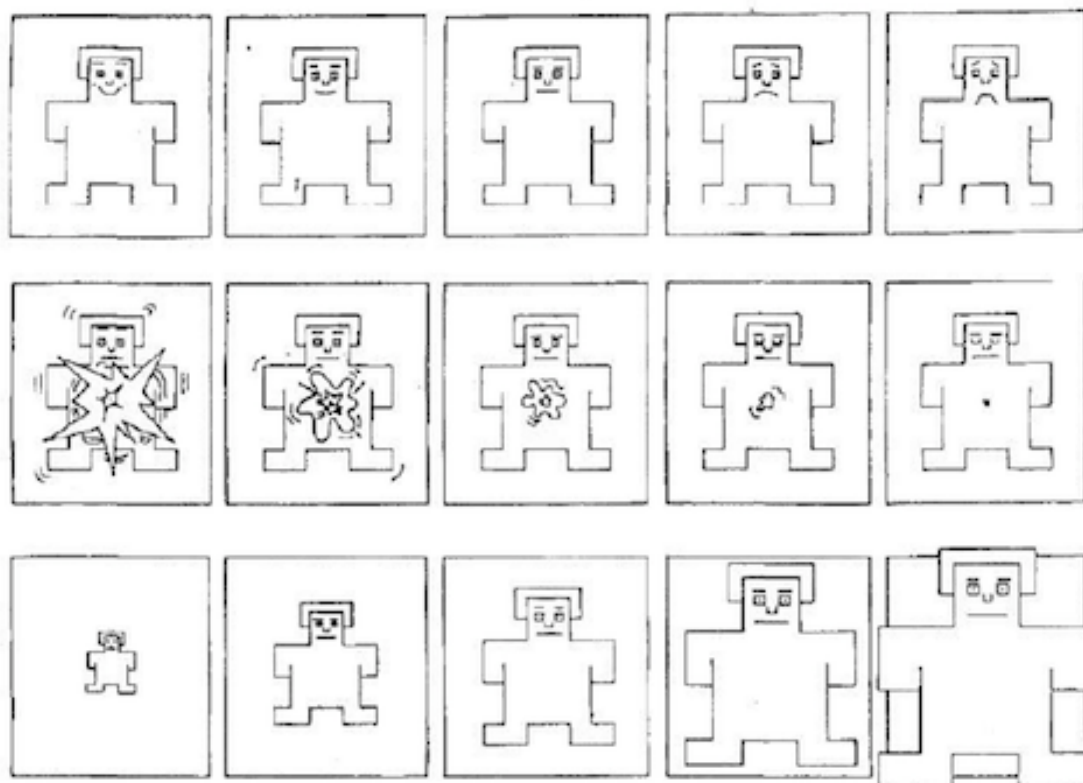
ρίες: ενδιαφέρον, ευχαρίστηση ή απογοήτευση, τότε τόσο η ερμηνεία όσο και η επίγνωση του πλαισίου είναι δύο καθοριστικής σημασίας στοιχεία για την επιτυχή επιλογή της σωστής αναπαράστασης σε κατηγορίες.

3.2.2 Διαστατικές Αναπαραστάσεις

Μια εναλλακτική προσέγγιση για την περιγραφή των συναισθημάτων είναι η αναπαράσταση τους σύμφωνα με χαρακτηριστικά συνεχούς μορφής, που ονομάζονται διαστάσεις. Οι πιο διαδεδομένες τέτοιες διαστάσεις είναι η ενεργοποίηση (activation), το σθένος (valence) και η κυριαρχία (dominance). Η ενεργοποίηση περιγράφει πόσο δραστήρια ή έντονη είναι η συναισθηματική εμπειρία, το σθένος περιγράφει το επίπεδο ευχαρίστησης που σχετίζεται με ένα συναίσθημα και εκτείνεται από θετικές έως αρνητικές τιμές για ευχάριστα και δυσάρεστα συναισθήματα αντίστοιχα και τέλος η κυριαρχία περιγράφει το επίπεδο της κυριαρχίας (ελέγχου) ενός ατόμου κατά τη διάρκεια μιας συναισθηματικής εμπειρίας. Αυτές οι συναισθηματικές διαστάσεις προτάθηκαν στον τομέα της Ψυχολογίας βάσει πειραματικών στοιχείων, σύμφωνα με τα οποία οι άνθρωποι ασυναίσθητα χρησιμοποιούν τέτοιες αναπαραστάσεις για την αξιολόγηση των συναισθημάτων [192, 87]. Αυτή η αναπαράσταση δεν υπονοεί την προεπιλογή μιας σειράς συναισθηματικών κατηγοριών για την ταξινόμηση σύνθετων συναισθηματικών εκδηλώσεων. Πράγματι, διάφορα συστήματα αναγνώρισης στον τομέα της Συναισθηματικής Υπολογιστικής χρησιμοποιούν αναπαραστάσεις διαστάσεων και συχνά εξετάζουν και αναφέρονται μόνο στις διαστάσεις της ενεργοποίησης και του σθένους αντίστοιχα ([257, 267]).

Τα συστήματα που χρησιμοποιούν αναπαραστάσεις διαστάσεων μπορεί να μην απαιτούν τον προκαθορισμό ενός συνόλου συναισθηματικών κατηγοριών (emotional classes), ωστόσο απαιτούν συνήθως την προεπιλογή του αριθμού των επιπέδων της ενεργοποίησης και του σθένους που θα χρησιμοποιηθούν κατά τη διαδικασία της ταξινόμησης. Για παράδειγμα, τα συστήματα μπορεί να χρησιμοποιούν τρία επίπεδα, υποδεικνύοντας χαμηλές, μεσαίες και υψηλές τιμές ή απαιτούν μεγαλύτερη λεπτομέρεια, όπως 5 ή 9 σημεία κλίμακας (point scales). Για να διευκολυνθεί η επισημείωση των συναισθηματικών εκδηλώσεων σε κλίμακες ενεργοποίησης, σθένους και κυριαρχίας αντίστοιχα, οι ερευνητές του τομέα της Ψυχολογίας εισήγαγαν την κλίμακα Αυτοαξιολόγησης (Self Assessment Manikins (SAM)) [19]. Το μοντέλο SAM αποτελείται από διαισθητικές εικονογραφικές παραστάσεις που περιγράφουν τα χαρακτηριστικά σε τρεις διαστάσεις. Ένα παράδειγμα του SAM για μία κλίμακα 5 σημείων αναφορικά με την ενεργοποίηση της αξιολόγησης, του σθένους και της κυριαρχίας απεικονίζεται στο Σχήμα 3.1 ([19]). Στον τομέα της Συναισθηματικής Υπολογιστικής, το μοντέλο SAM έχει χρησιμοποιηθεί για τη βαθμολόγηση σημειακής κλίμακας (rating) συναισθηματικών σωμάτων κειμένου [25].

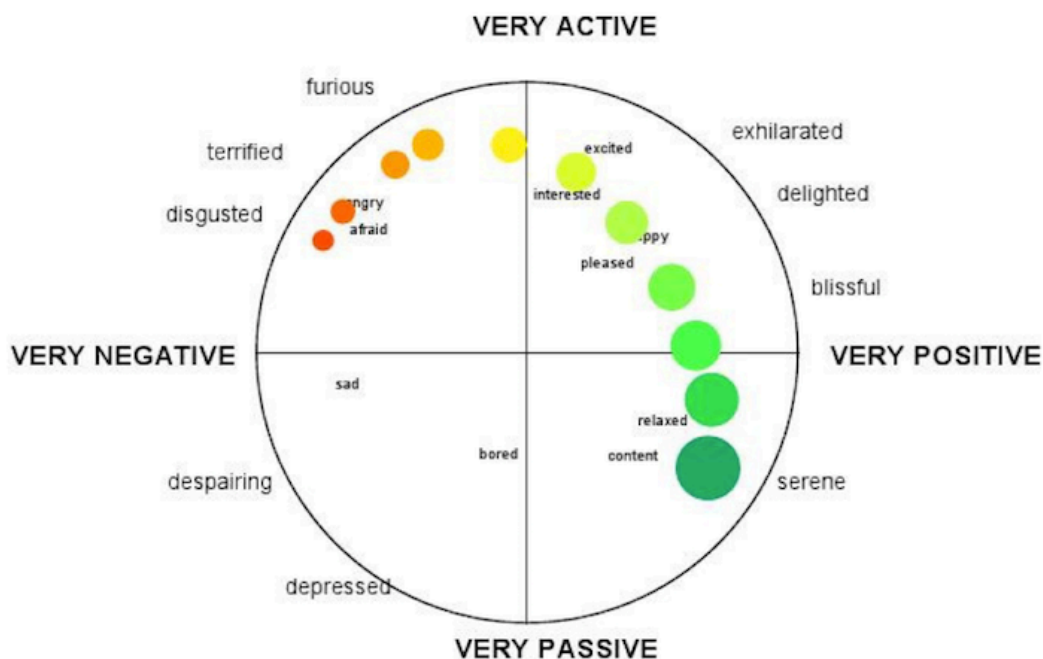
Εναλλακτικά, οι ερευνητές προσπάθησαν να αξιοποιήσουν πλήρως τη συνεχή φύση των εν λόγω αναπαραστάσεων, συλλέγοντας συνεχείς βαθμολογήσεις (ratings) συναισθηματικών διαστάσεων και αναπτύσσοντας συστήματα που να υπολογίζουν τις συνεχείς συναισθηματικές ιδιότητες, π.χ. [260]. Ένα ακόμη βήμα αποτελεί η συγκέντρωση βαθ-



Σχήμα 3.1: Η κλίμακα SAM χρησιμοποιήθηκε για τη βαθμολόγηση των συναισθηματικών διαστάσεων του σθένους (άνω σειρά), την ενεργοποίηση (μεσαία γραμμή) και την κυριαρχία (κάτω σειρά) [19].

μολογιών συνεχούς μορφής των ιδιοτήτων των διαστάσεων κατά τη διάρκεια του χρόνου και η δημιουργία συστημάτων που να αντιπροσωπεύουν συναισθηματικά χαρακτηριστικά, όχι ως σημεία, αλλά ως καμπύλες που λαμβάνουν συνεχείς τιμές και εξελίσσονται στο χρόνο. Η συγκέντρωση τέτοιων βαθμολογιών συνεχούς μορφής προτάθηκε από τους συγγραφείς του [45], οι οποίοι παρουσίασαν το λογισμικό επισημείωσης Feeltrace, που απεικονίζεται στο Σχήμα 3.2. Το Feeltrace δίνει τη δυνατότητα στο χρήστη να παρέχει επισημείωση συναισθηματικού περιεχομένου σε πραγματικό χρόνο μετακινώντας το δρομέα (cursor) σε μια διεπαφή που αντιπροσωπεύει το διδιάστατο χώρο της ενεργοποίησης και του σθένους, τη στιγμή που οι συναισθηματικές εκδηλώσεις εμφανίζονται σε ξεχωριστό πρόγραμμα αναπαραγωγής βίντεο. Το αποτέλεσμα αυτής της διαδικασίας είναι μια συναισθηματική καμπύλη για κάθε συναισθηματική ιδιότητα.

Ανεξάρτητα από την εκάστοτε αναπαράσταση, μία πρόκληση αναφορικά με την βαθμολόγηση των συναισθημάτων είναι η εγγενής υποκειμενικότητα του έργου. Τα άτομα συχνά αντιλαμβάνονται τα συναισθήματα με έναν υποκειμενικό τρόπο, και ως εκ τούτου μπορεί να υπάρχει διαφωνία ως προς την πραγματική συναισθηματική επισημείωση μιας συναισθηματικής εκδήλωσης, αναφορικά με διαφορούμενες συναισθηματικές εκδηλώσεις. Φυσικά, όσο μεγαλύτερο είναι το επίπεδο λεπτομέρειας των συναισθηματικών σχολιασμών, π.χ., περισσότερες ετικέτες που αναπαριστούν το συναίσθημα σε κατηγο-



Σχήμα 3.2: Παράδειγμα του προγράμματος επισημείωσης *Feeltrace* κατά τη διάρκεια μιας περιόδου παρακολούθησης. Ο οριζόντιος άξονας αποδίδει τη διάσταση σθένους και ο κάθετος άξονας τη διάσταση της ενεργοποίησης. Ο σχολιαστής μπορεί να μετακινήσει το δρομέα (υποδεικνύεται από έναν κύκλο) σε αυτήν τη διασύνδεση, για να αξιολογήσει μία συναισθηματική εκδήλωση που παρουσιάζεται παράλληλα σε ένα ξεχωριστό παράθυρο.

ρίες ή περισσότερα επίπεδα συναισθηματικών χαρακτηριστικών, τόσο λιγότεροι θα είναι οι σχολιαστές που αναμένεται να συμφωνήσουν. Το γεγονός αυτό υπογραμμίζει μια θεμελιώδη πρόκληση του προβλήματος της αναγνώρισης των συναισθημάτων, δηλαδή το γεγονός ότι υπάρχει αβεβαιότητα όσον αφορά στην επισημείωση που σχετίζεται με ένα παράδειγμα, σε αντίθεση με το παραδοσιακά επιβλεπόμενο μοτίβο αναγνώρισης όπου υπάρχει σαφής συσχέτιση μεταξύ ενός παραδείγματος και της αντίστοιχης επισημείωσής του.

3.2.3 Θεωρίες Γνωστικής Αξιολόγησης

Μέχρι σήμερα έχουν διατυπωθεί ποικίλες θεωρίες συναισθημάτων που στηρίζονται στη γνωστική αξιολόγηση. Κάθε μία ορίζει ένα διαφορετικό σύνολο συνιστωσών που απαρτίζουν αυτή τη διαδικασία αξιολόγησης. Υπάρχει όμως μια κεντρική ιδέα που τις διέπει, ο ισχυρισμός ότι μπορούμε να διακρίνουμε τις συναισθηματικές καταστάσεις, ανάλογα με το ποιές συνιστώσες αξιολόγησης συμμετέχουν στην εκάστοτε συναισθηματική εμπειρία.

Ένα σημαντικό ερώτημα που πρέπει να απαντηθεί είναι το εάν υπάρχουν “βασικές” ή “θεμελιώδεις” διαστάσεις αξιολόγησης. Όσοι απαντάνε θετικά τείνουν να πιστεύουν και σε μικρά σύνολα από θεμελιώδη συναισθήματα, έτσι ώστε συγκεκριμένοι συνδυα-

σμοί από αξιολογήσεις να προκαλούν (και άρα να επιτρέπουν αντίστροφα και την πρόβλεψη) μία συναισθηματική εμπειρία. Ερευνητές όπως ο Lazarus, ο Scherer, ο Roseman και ο Ellsworth υποδεικνύουν ένα πεπερασμένο σύνολο συνιστωσών που τις θεωρούν ανεξάρτητες της εκάστοτε κατάστασης και των εκάστοτε συνθηκών όπου εξελίσσεται ένα συναισθηματικό επεισόδιο. Δεν αρνούνται την ύπαρξη σημαντικών κοινωνικών και πολιτιστικών επιρροών, τόσο στη συμπεριφορά όσο και στην εννοιολογική τους σύλληψη. Συγκεκριμένα, αυτό που προτείνουν είναι ότι η σχέση ανάμεσα στις αξιολογήσεις και στα συναισθήματα είναι γενική ανά τις κουλτούρες και ίσως και παγκόσμια. Αν άνθρωποι από διαφορετικές κουλτούρες αξιολογούν μία κατάσταση με τον ίδιο τρόπο, τότε θα βιώνουν το ίδιο συναίσθημα. Αν βιώσουν διαφορετικό συναίσθημα είναι γιατί έχουν αξιολογήσει την κατάσταση διαφορετικά [64].

Ο Smith και ο Lazarus [216] υποστήριξαν ότι η γνωστική αξιολόγηση αποτελείται από τις ακόλουθες συνιστώσες, οι οποίες αναφέρονται ως appraisal components ή appraisal dimensions:

- Έλεγχος συνάφειας με προσωπικά κίνητρα (πρωτεύουσα αξιολόγηση)
- Έλεγχος σχέσης με προσωπικούς στόχους/επιδιώξεις (πρωτεύουσα αξιολόγηση)
- Έλεγχος υπευθυνότητας - σε ποιόν αποδίδονται τα εύσημα ή η ευθύνη; (δευτερεύουσα αξιολόγηση)
- Έλεγχος της δυνατότητας του ατόμου να ανταπεξέλθει στη λύση μιας προβληματικής κατάστασης (δευτερεύουσα αξιολόγηση)
- Έλεγχος της δυνατότητας του ατόμου να ανταπεξέλθει συναισθηματικά στην κατάσταση (δευτερεύουσα αξιολόγηση)
- Πρόβλεψη για το μέλλον - πόσο πιθανό είναι να αλλάξει η κατάσταση (δευτερεύουσα αξιολόγηση)

Σε αντίθεση με τους υπέρμαχους των θεωριών αξιολόγησης με πεπερασμένες συνιστώσες υπάρχουν και ερευνητές που πιστεύουν ότι το πλήθος των δυνατών διαστάσεων αξιολόγησης και ο αριθμός των διαφορετικών δυνατών συναισθηματικών εμπειριών είναι ουσιαστικά απεριόριστα. Είναι χαρακτηριστική η περίπτωση όπου διαφορετικές απόψεις αναδύονται όταν μελετώνται τα κοινά και οι διαφορές ανάμεσα σε διαφορετικά άτομα και σε διαφορετικές κουλτούρες.

3.2.3.1 Συνθετικό Μοντέλο Συναισθημάτων

Σε αντίθεση με την αναπαράσταση του συναισθήματος σε κατηγορίες και σε διαστατικές προσεγγίσεις, πρόσφατα, το ενδιαφέρον της ερευνητικής κοινότητας στράφηκε σε ένα άλλο σύνολο Ψυχολογικών μοντέλων, το οποίο αναφέρεται ως συνθετικό μοντέλο συναισθημάτων. Το μοντέλο αυτό βασίζεται στη θεωρία της αξιολόγησης και αναμένεται

ότι είναι καταλληλότερο για την ανάπτυξη περιβαλλόντων που έχουν επίγνωση του πλαισίου (context-aware) [151]. Ωστόσο, η σωστή χρήση της προσέγγισης της αξιολόγησης για την αυτόματη ανάλυση των συναισθημάτων παραμένει ένα ανοιχτό ερευνητικό πρόβλημα. Στο συνθετικό μοντέλο συναισθήματος, διάφοροι τρόποι σύνδεσης της αυτόματης συναισθηματικής ανάλυσης και των μοντέλων αξιολόγησης του συναισθήματος έχουν προταθεί. Αυτός ο σύνδεσμος στοχεύει στο να επιτρέψει την ενσωμάτωση του πλαισίου στα συστήματα αυτόματης ανάλυσης των συναισθημάτων.

Ωστόσο, αυτά τα συναισθηματικά μοντέλα εμπεριέχουν μια σειρά περιορισμών. Η αναπαράσταση συναισθημάτων σε κατηγορίες αποτυγχάνει να περιγράψει το περίπλοκο φάσμα των συναισθημάτων που μπορούν να εμφανιστούν στην καθημερινή επικοινωνία. Επιπλέον, ο χώρος των διαστάσεων δεν επιτρέπει ούτε τη σύγκριση των λέξεων συναισθηματικού περιεχομένου ανάλογα με την αμοιβαία απόστασή τους (reciprocal distance), ούτε τη μοντελοποίηση του γεγονότος ότι δύο ή περισσότερα συναισθήματα μπορεί να συμβαίνουν την ίδια στιγμή.

Για το λόγο αυτό, μια σειρά από προσεγγίσεις 2D διαστάσεων χρησιμοποιούνται κυρίως για την οπτικοποίηση αυτών των Ψυχολογικών Θεμελιών (Psychological Foundations). Ένα πρώιμο παράδειγμα είναι το μοντέλο που προτάθηκε από τον Russell (Russell's Circumplex Model) [190], το οποίο χρησιμοποιεί τις διαστάσεις της διέγερσης (arousal) και του σθένους (valence) για να δημιουργήσει ένα διάγραμμα 150 ετικετών συναισθηματικού περιεχομένου. Ομοίως, ο Whissell αντιμετωπίζει τα συναισθήματα ως 2D διαστάσεις σε συνεχή χώρο χρησιμοποιώντας την αξιολόγηση και την ενεργοποίηση ως διαστάσεις [253]. Ένα άλλο δισδιάστατο μοντέλο είναι ο τροχός των συναισθημάτων του Plutchik [176], σύμφωνα με το οποίο τα συναισθήματα είναι προσαρμοστικά καθώς βασίζονται σε εξελικτικές αρχές, παρόλο που αξιολογούμε τα συναισθήματα (emotions) ως αισθήματα (feelings). Αυτές οι καταστάσεις αισθήματος είναι μέρος μιας διαδικασίας που περιλαμβάνει τόσο γνώση όσο και συμπεριφορά και περιέχει αρκετούς βρόχους ανατροφοδότησης. Δυστυχώς όμως, όλες αυτές οι προσπάθειες είναι αποδοτικές σε επίπεδο λέξεων αλλά όχι πολλαπλών λέξεων εννοιών.

Ωστόσο, δεδομένου ότι τα παραπάνω μοντέλα επικεντρώνονται σήμερα στην αντικειμενικότητα μιας πληροφορίας συναισθηματικού περιεχομένου, όταν αυτά συνδέονται με γλωσσικές απόψεις, τα συναισθήματα που αξιολογούνται σύμφωνα με γνωστικές συνιστώσες (appraisal-based) δεν λαμβάνονται υπόψιν. Εντούτοις, λόγω της καταλληλότητάς τους αναφορικά με την ενσωμάτωση του εννοιολογικού πλαισίου, περισσότερη έμφαση πρέπει να δοθεί στα συναισθηματικά μοντέλα που βασίζονται στην γνωστική αξιολόγηση.

3.3 Αναπαράσταση Συναισθήματος στον τομέα της Φυσικής Γλώσσας

Οι εκφράσεις του συναισθήματος, εξ ορισμού, ορίζονται στον τομέα της Φυσικής Γλώσσας ως θετικές ή αρνητικές. Τα συναισθήματα περιλαμβάνουν συναισθηματικές διαταγές (dispositions) που διαμορφώνονται προς ένα αντικείμενο και σε αντίθεση με τα συναισθήματα (emotions) που είναι σύντομης διάρκειας, τα συναισθήματα (sentiments) σχετικά με ένα αντικείμενο είναι μεγαλύτερης διάρκειας αντίστοιχα.

Η προσέγγιση των συναισθημάτων υπό το πρίσμα της διάρκειας [74] είναι ιδιαιτέρως χρήσιμη στους τομείς της ψυχαγωγίας και των συστημάτων συστάσεων (recommendation systems), καθώς μπορεί να προσφέρει ακόμη περισσότερο ενημερωμένες κριτικές και αξιολογήσεις ή ακόμη και να οδηγήσει σε αναθεωρήσεις μετά από μία σειρά αξιολογήσεων. Για παράδειγμα, μία κακή κριτική (review) της ταινίας Sky Fall, από ένα πρόσωπο που απολαμβάνει γενικά και του αρέσουν οι ταινίες δράσης, θα είχε πιθανώς μεγαλύτερη συνάφεια για συστήματα συστάσεων (recommendation systems) συγκριτικά με μια κακή κριτική από κάποιον που δεν έχει προτίμηση στις ταινίες δράσης. Ωστόσο, το ερευνητικό πεδίο που είναι αφιερωμένο στην ανίχνευση συναισθημάτων στον τομέα της Φυσικής Γλώσσας (Sentiment Analysis (SA)), δεν έχει αντιμετωπίσει τα συναισθήματα ως διαρκή. Ρεαλιστικά, σε αυτό τον ερευνητικό τομέα, η συλλογή των συναισθημάτων μπορεί να πραγματοποιηθεί μόνο με βάση την τακτική τους εμφάνιση μέσω των συναισθηματικών απαντήσεων των χρηστών ως προς ένα αντικείμενο [74]. Μία αντιπροσωπευτική προσπάθεια αυτού του τομέα είναι η έρευνα που παρουσιάζεται στο [175]. Συγκεκριμένα, περιγράφεται η συλλογή ελεύθερα διαθέσιμων δεδομένων χρηστών από το Twitter για μία περίοδο που κυμαίνεται από τρεις μήνες έως ένα έτος, ανάλογα με τον χρήστη, τα tweets του οποίου συγκεντρώθηκαν και φιλτραρίστηκαν σύμφωνα με ένα αντικείμενο (π.χ., πρόσωπο, τόπος, και κατάσταση). Τέλος, τα φιλτραρισμένα tweets, αναλύθηκαν σύμφωνα με τα οκτώ (8) βασικά συναισθήματα του Plutchik: εμπιστοσύνη, χαρά, αναμονή, έκπληξη, θυμό, φόβο, αηδία και θλίψη.

Στην Υπολογιστική Γλωσσολογία και στην επεξεργασία της Φυσικής Γλώσσας, το πεδίο που είναι αφιερωμένο στην έρευνα και στην ανίχνευση απόψεων από το κείμενο είναι γνωστό με την ονομασία “εξόρυξη γνώμης” (Opinion Mining (OM)). Όπως προαναφέρθηκε, οι όροι OM και SA είναι συμπληρωματικοί και επικαλυπτόμενοι, παρόλο που οι απόψεις (opinions) μπορούν να βασιστούν στην προσωπική ερμηνεία της πληροφορίας αλλά δεν μπορεί να φέρουν συναισθηματικό μεγαλύτερης (sentiment) ή μικρότερης (emotion) διάρκειας αντίστοιχα². Προσπαθώντας να ορίσουμε τις απόψεις (opinions), καταλήγουμε

²Στο σημείο αυτό να διευκρινίσουμε ότι εδώ οι όροι sentiment και emotion είναι αλληλένδετοι. Για παράδειγμα, κάποιος εκφέρει τη γνώμη του αναφορικά με μία σύντομη συναισθηματική εμπειρία που του συνέβη (“Μου άρεσε αυτή η ταινία”, καθώς επέστρεφε από τον κινηματογράφο) είτε σύμφωνα με τη γενικότερη άποψη που έχει διαμορφώσει (“Οι ταινίες δράσης είναι οι καλύτερες”, σύμφωνα με τις προτιμήσεις του αναφορικά με τις ταινίες δράσης). Δυστυχώς, σε πολλές εφαρμογές όπου η εξόρυξη απόψεων απαιτείται, για παράδειγμα, στην αξιολόγηση μιας ταινίας ή ενός προϊόντος ο διαχωρισμός μεταξύ της μικρής ή

στο ότι μία άποψη περιλαμβάνει τουλάχιστον τρία μέρη: το θέμα, τον κάτοχο της άποψης και την απαίτηση (claim). Επιπλέον, μία άποψη μπορεί να περιλαμβάνει τα εξής: (1) συναισθηματική έκφραση μικρής (emotional) και μεγάλης διάρκειας (sentimental), (2) χαρακτηριστικό (-α) του θέματος και (3) χρόνος (η ώρα της γνωμοδότησης).

Ιδιαίτερο ενδιαφέρον για την εξόρυξη γνώμης (OM) παρουσιάζει ο προσδιορισμός του κατά πόσο οι διατυπωμένες απόψεις είναι θετικές, αρνητικές ή ουδέτερες. Επίσης, αξίζει να σημειωθεί ότι μία γνώμη μπορεί να κατηγοριοποιηθεί ως ουδέτερη όταν δηλώνεται χωρίς καμία θετική ή αρνητική αξιολόγηση. Για παράδειγμα, “ο Μπιν Λάντεν κρύβεται στο Πακιστάν”³.

Επιπλέον, εάν μια γνώμη περιλαμβάνει ένα συναίσθημα (emotion) είτε μία συναισθηματική έκφραση (sentiment), τότε μπορεί να χρησιμοποιηθούν τα συναισθήματα αυτά για τον χαρακτηρισμό της γνώμης είτε ως αρνητική ή είτε ως θετική. Στην περίπτωση όπου η γνώμη δε σχετίζεται με κανένα συναίσθημα είτε συναισθηματική έκφραση, τότε η πολικότητα (polarity) της εξεταζόμενης γνώμης καθορίζεται ως εξής:

- Αναλύοντας εάν υπάρχει συναισθηματική αντίδραση στο θέμα ή στο χαρακτηρισμό (“Βρήκα αυτό το βιβλίο βαρετό” και “αυτή η ταινία σκηνοθετήθηκε όμορφα”), χρησιμοποιώντας συναισθηματικές λέξεις-κλειδιά από τις υπάρχουσες βάσεων γνώσεων (WordNet [148]) ή από λεξικά τα οποία βασίζονται σε σώματα κειμένων ανάλογα με τον εξεταζόμενο τομέα (domain-dependent corpora)([235, 106, 112]), είτε
- Αναλύοντας τις λέξεις που οι άνθρωποι επιλέγουν να περιγράψουν τις εμπειρίες τους. Υπάρχουν πολλές περιπτώσεις όπου ένας συγγραφέας εκφράζει τη γνώμη του χωρίς να τη συνδέει με οποιεσδήποτε θετικές ή αρνητικές λέξεις (π.χ. “η μπαταρία του υπολογιστή διαρκεί μόνο τρεις ώρες”). Αυτές οι περιπτώσεις θεωρούνται περισσότερο δύσκολες ως προς την κατηγοριοποίησή τους ως θετικές ή αρνητικές, καθώς απαιτούν ένα μεγάλο σώμα της γνώσης. Παρόλα αυτά η κατηγοριοποίησή τους δεν θεωρείται αδύνατη.

Λαμβάνοντας την παραπάνω δήλωση ως παράδειγμα, “Η μπαταρία του υπολογιστή διαρκεί μόνο τρεις ώρες”, η λέξη “μόνο” χαρακτηρίζεται από την Sokolova και Lapalme [219] ως περιγραφική λέξη. Διαπιστώνουμε λοιπόν ότι σε τέτοιες περιπτώσεις κρίνεται απαραίτητη η ενσωμάτωση του πλαισίου, καθώς η λέξη “μόνο” μπορεί να χρησιμεύσει ως υπαινιγμός τόσο για μία θετική όσο και για μία αρνητική γνώμη. Στο παράδειγμά μας, είναι επίσης εμφανές ότι ο χρήστης του υπολογιστή αναμένεται ότι η μπαταρία θα διαρκέσει περισσότερο. Έτσι, η άποψη αυτή μπορεί να κατηγοριοποιηθεί ως αρνητική γνώμη.

της μεγαλύτερης διάρκειας ενός συναισθήματος είναι δύσκολο να κατανοηθεί.

³Ο Sayeed [200] υποστήριξε ότι συχνά η κατηγοριοποίηση μιας γνώμης σε θετική, αρνητική είτε ουδέτερη είναι εξαρτώμενη από την οπτική γωνία του αναγνώστη του κειμένου. Παραδείγματος χάριν, σύμφωνα με έναν αναγνώστη, η πρόταση “Ο Μπιν Λάντεν κρύβεται στο Πακιστάν” μπορεί να χαρακτηριστεί ως μία ουδέτερη πρόταση. Η ίδια πρόταση όμως μπορεί να χαρακτηριστεί είτε ως θετική είτε ως αρνητική λαμβάνοντας υπόψιν το κατά πόσο ο αναγνώστης ενδιαφέρεται για τη συγκεκριμένη πληροφορία και εάν γνωρίζει κατά πόσο η επιλογή της τοποθεσίας από τον Μπιν Λάντεν είναι σωστή.

Σε αντίθεση με τις λίστες των συναισθηματικά φορτισμένων λέξεων σε επίπεδο-λέξης (word-level), η ανάλυση των περιγραφικών λέξεων απαιτεί ένα ευρύτερο πλαίσιο, όπως μια ολόκληρη πρόταση (sentence-level), ολόκληρη παράγραφο (paragraph-level) ή ένα ολόκληρο κείμενο (document-level) [219].

Κεφάλαιο 4

Πολυτροπική Αναγνώριση Συναισθηματικών Καταστάσεων

4.1 Αναγνώριση Συναισθηματικών Καταστάσεων από σήματα Ομιλίας και Φυσιολογίας

4.1.1 Εισαγωγή

Η έλλειψη της συναισθηματικής εκφραστικότητας είναι ένα από τα κύρια χαρακτηριστικά περιόδων άγχους όταν τα μέλη μίας ομάδας καλούνται να εκτελέσουν καθήκοντα συνεργασίας βασιζόμενοι κυρίως σε γνωστικά (cognitive) χαρακτηριστικά. Στις περιπτώσεις αυτές, τα μέλη της ομάδας δυσκολεύονται να εκφράσουν συνειδητά τα συναισθήματά τους και παρουσιάζουν διαφορετικά μοτίβα αντιλήψεων αναφορικά με τη μεταφορά συναισθηματικών πληροφοριών κατά τη διάρκεια της ομαδικής εργασίας για την επίτευξη του κοινού σκοπού [105, 178]. Διαπιστώνουμε λοιπόν, ότι η δυνατότητα παρακολούθησης της εσωτερικής κατάστασης των μελών που αλληλεπιδρούν μεταξύ τους μέσα σε τέτοια περιβάλλοντα μπορεί να μας δώσει νέες ιδέες σχετικά με τους μηχανισμούς της μεταξύ τους αλληλεπίδρασης, της ενσυναίσθησής τους και της γνωστικής τους ικανότητας.

Κατά τη διάρκεια υψηλού φόρτου εργασίας, πίεσης και μεγάλης διάρκειας περιόδων άγχους, το συμπαθητικό νευρικό σύστημα είναι υπεύθυνο για την ενεργοποίηση όλων των αδένων και των οργάνων που σχετίζονται με την υπεράσπιση του ανθρωπίνου σώματος από τις αντιληπτές απειλές. Αυτή η ενεργοποίηση συνδέεται με αλλαγές στη διεγερση (arousal) που επηρεάζονται περαιτέρω από το συναίσθημα (emotion), τη γνώση (cognition) ή την προσοχή (attention). Οι υψηλής διάρκειας περίοδοι άγχους οδηγούν σε αυξημένη δραστηριότητα του συμπαθητικού νευρικού συστήματος και εκδηλώνονται μέσω μίας σειράς σωματικών αντιδράσεων, όπως η αύξηση του καρδιακού παλμού, η αύξηση της ροής του αίματος στα άκρα, η αύξηση του ρυθμού αναπνοής, κλπ. Έτσι, ο συνδυασμός περισσότερων του ενός σημάτων φυσιολογίας μπορεί να θεωρηθεί ως μία αξιόπιστη μέτρηση των μεταβολών του άγχους, αφού παρέχει αξιόπιστες εκτιμήσεις ανα-

φορικά με τα συναισθήματα (emotion), τη διέγερση (arousal) και τη γενική γνώση (general cognition)[142].

Στο τρέχον κεφάλαιο, εξετάζουμε το συσχετισμό μεταξύ των σημάτων φυσιολογίας των μελών της ομάδας (εκπαιδευτής-εκπαιδευόμενος) και των απαντήσεων τους κατά τη διάρκεια της συνομιλίας τους (turn-taking responses), η οποία ορίζεται ως το χρονικό διάστημα της έναρξης της ομιλίας του εκπαιδευόμενου από το τέλος της ομιλίας του εκπαιδευτή. Έχει παρατηρηθεί ότι τα μέλη μίας ομάδας όταν συμμετέχουν σε εξαιρετικά απαιτητικές εργασίες (operational tasks) δεν παρατηρούν συχνά τα ερεθίσματα που τους προκαλούν άγχος και πίεση [153, 222]. Έτσι, αν και επικοινωνούν με τα άλλα μέλη της ομάδας, τα συναισθήματα που εκφράζουν συχνά δεν είναι εμφανή μέσω της οπτικοακουστικής παρατήρησης. Για παράδειγμα, μια αναμενόμενη αντίδραση από ένα άτομο που του ζητείται να αφοπλίσει μια προσομοιωμένη βόμβα, είναι τα υψηλά επίπεδα (εσωτερικού/μη ορατού) άγχους. Αυτή η έμφυτη διαφορά μεταξύ της εξωτερικής παρατηρήσιμης συμπεριφοράς των μελών μιας ομάδας και της εσωτερικής συναισθηματικής τους κατάστασης δεν είναι γίνεται εύκολα αντιληπτή και μπορεί να γεφυρωθεί μόνο μέσω της παρακολούθησης χαρακτηριστικών των σημάτων φυσιολογίας τους.

Η διάρκεια των απαντήσεων τους θεωρείται μία εξίσου σημαντική ένδειξη, καθώς σχετίζεται με τις πνευματικές διαδικασίες υψηλών ποσοστών άγχους [183]. Λαμβάνοντας υπόψιν μας ότι οι δείκτες φυσιολογίας αντικατοπτρίζουν τις πτυχές των ψυχικής/πνευματικής κατάστασης και συχνά το μέγεθος της δυσφορίας [63], διερευνούμε κατά πόσο τα σήματα φυσιολογίας σύντομης αλλά και μεγαλύτερης διάρκειας απαντήσεων σχετίζονται με διαφορετικά μοτίβα φυσιολογίας. Η περαιτέρω καταγραφή και ερμηνεία αυτής της συνεχιζόμενης και εξελισσόμενης αλληλεπίδρασης, βασίζεται στη χρήση δύο στρατηγικών μάθησης του συνόλου (ensemble learning). Πιστεύουμε ότι η διερεύνηση των νευροφυσιολογικών αλλαγών κατά τη διάρκεια των απαντήσεων των συμπαικτών παρέχει μία καλύτερη κατανόηση της δυναμικής της ομάδας.

4.1.2 Σχετική Βιβλιογραφία

Σύμφωνα με την τρέχουσα βιβλιογραφία, μία σειρά ερευνών έχει διεξαχθεί αναφορικά με τη δυναμική του διαλόγου και με το συσχετισμό του έντονου άγχους και της αυξημένης πίεσης με παραμέτρους φυσιολογίας. Παρόλα αυτά, τα προαναφερθέντα πεδία έχουν μελετηθεί σε μεγάλο βαθμό μεμονωμένα και όχι συνδυαστικά. Για παράδειγμα, οι εργασίες αναφορικά με τη μετάβαση της έναρξης της ομιλίας του εκπαιδευόμενου από το τέλος της ομιλίας του εκπαιδευτή [183] θα μπορούσαν να επωφεληθούν από τις προσωπικές σχέσεις των ατόμων που αλληλεπιδρούν την παρακολούθηση και την καταγραφή των αντιδράσεων φυσιολογίας τους όπως τα υψηλά επίπεδα άγχους κατά τη διάρκεια ομαδικών επιχειρησιακών καθηκόντων [53]. Επιπλέον, ο συσχετισμός των απαντήσεων με παραμέτρους φυσιολογίας μελετήθηκε και για την αξιολόγηση του τρόπου με τον οποίο τα επίπεδα θυμού των ενηλίκων μπορούν να επηρεάσουν τα παιδιά [63]. Τέλος, προηγούμενες μελέτες έχουν αποδείξει τα πλεονεκτήματα της χρήσης της μάθησης του συνόλου

τόσο στη μονοτροπική [206, 202] όσο και στην πολυτροπική αναγνώριση των συναισθηματικών καταστάσεων [81, 201].

Από όσο γνωρίζουμε, δεν υπάρχει πειραματική επιβεβαίωση της εφαρμογής μεθόδων μάθησης του συνόλου (ensemble learning) για τη μελέτη της συσχέτισης μεταξύ της ομιλίας των μελών μιας ομάδας (Ενότητα 4.1.4) και των εσωτερικών συναισθηματικών καταστάσεων που συνάγονται από παραμέτρους φυσιολογίας (Ενότητα 4.1.5) δύο συμπαικτών (εκπαιδευτής-εκπαιδευόμενος) οι οποίοι προσπαθούν να αφοπλίσουν μια προσομοιωμένη βόμβα (πλαίσιο) (Ενότητα 4.1.3.1) [157].

Στόχος μας λοιπόν είναι να ξεδιπλώσουμε αυτή τη συσχέτιση. Τα ερευνητικά μας αποτελέσματα ενισχύουν την αρχική μας υπόθεση, ότι δηλαδή τα πρότυπα φυσιολογίας μεταδίδουν πληροφορίες της εσωτερικής κατάστασης του εκπαιδευόμενου, καθώς διαφοροποιούνται ανάλογα με τη διάρκεια των απαντήσεων του εκπαιδευτή (Ενότητα 4.1.6.2). Τέλος, τα αποτελέσματά μας ενισχύονται επιπρόσθετα μέσω των μεθόδων μάθησης του συνόλου, που ξεπερνούν τους περισσότερους βασικούς ταξινομητές (base learners).

4.1.3 Πειραματική Μελέτη

4.1.3.1 Περιγραφή του Σώματος Κειμένου

Χρησιμοποιούμε ένα σώμα κειμένου [157] το οποίο περιγράφει την αλληλεπίδραση μεταξύ ατόμων μέλη μιας ομάδας, στόχος των οποίων είναι η επιτυχής αφοπλισή μιας προσομοιωμένης βόμβας. Ειδικότερα, το σενάριο που εκτυλίσσεται ακολουθεί τις παρακάτω παραδοχές: αποτελείται από δύο (2) πειραματικές συνθήκες μεταξύ μιας ομάδας δύο (2) ατόμων, είκοσι (20) ατόμων συνολικά. Σύμφωνα με την πρώτη παραδοχή (**The IB Breaker Conversation (IB)**) επιτρέπεται στους συμπαίκτες να γνωριστούν μεταξύ τους πριν την έναρξη της πειραματικής διαδικασίας μέσω μιας σειράς προσωπικών ερωτημάτων. Αντίστοιχα, σύμφωνα με τη δεύτερη συνθήκη, η οποία ονομάζεται κατάσταση ελέγχου (**Control Task (CT)**), οι συμπαίκτες προχωρούν στην αφοπλισή της βόμβας χωρίς να υπάρχει προηγούμενη εξοικείωση μεταξύ τους. Το συγκεκριμένο σώμα κειμένου αποτελείται από μία σειρά προσομοιωμένων σεναρίων “αφοπλισής της βόμβας”. Σε κάθε σενάριο ένα μέλος της ομάδας λειτουργεί ως “εκπαιδευτής” και το δεύτερο μέλος της ομάδας λειτουργεί ως “εκπαιδευόμενος”. Ο “εκπαιδευτής” ακολουθεί ένα εγχειρίδιο με οδηγίες για το πώς να αφοπλίσει τη βόμβα. Σύμφωνα με τις οδηγίες που δόθηκαν στον “εκπαιδευτή”, είναι ευθύνη των “εκπαιδευόμενων” να ολοκληρώσουν επιτυχώς την εργασία που τους ανατέθηκε, δηλαδή την αφοπλισή της βόμβας. Μετά από κάθε ξεχωριστό καθήκον (task) τα μέλη της ομάδας αλλάζουν ρόλους (δηλαδή, σε κάθε ομάδα και τα δύο μέλη έχουν και το ρόλο του εκπαιδευτή και του εκπαιδευόμενου), οδηγώντας έτσι σε ένα σύνολο τεσσάρων (4) κύριων καθηκόντων, εκ των οποίων κάθε ένα διαρκεί κατά μέσο όρο 5 λεπτά (5 mins). Να αναφέρουμε στο σημείο αυτό ότι για τα πειραματικά αποτελέσματα που παρουσιάζονται στην Ενότητα 4.1.6.2, λαμβάνουμε υπόψιν μας μόνο δέκα (10) ομάδες συνολικά (πέντε (5) από κάθε κατάσταση) και εξετάζουμε μόνο την περι-

πτωση κατά την οποία ο συμμετέχων Α είναι ο εκπαιδευτής και ο συμμετέχων Β είναι ο εκπαιδευόμενος, δηλαδή δεν εξετάζουμε την περίπτωση της αντιστροφής των ρόλων.

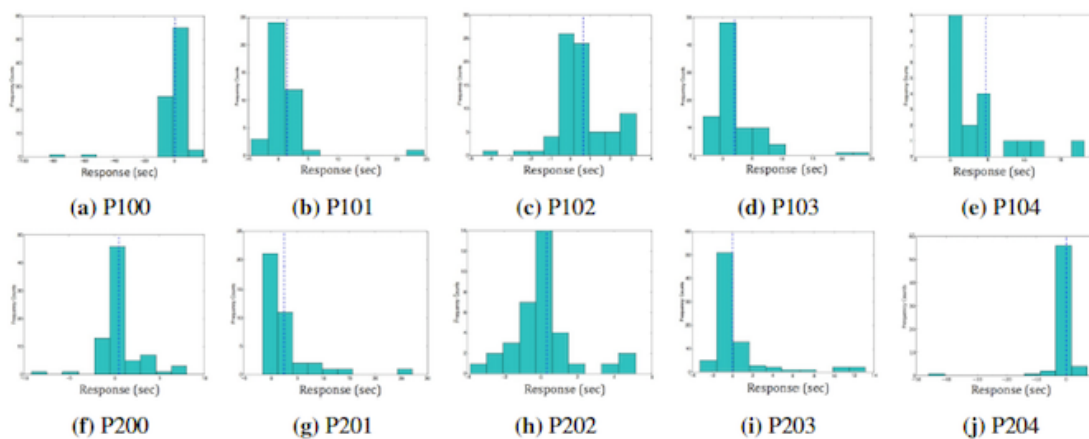
4.1.4 Εξαγωγή χαρακτηριστικών Ομιλίας

Μία από τις δύο κύριες κατευθυντήριες αρχές που διέπουν την ανθρώπινη συνομιλία είναι ότι μόνο ένα άτομο πρέπει να μιλάει κάθε φορά και ότι δεν μπορούμε να έχουμε σιωπή. Στα πλαίσια λοιπόν της οργάνωσης των συνομιλιών, όταν ακολουθούνται αυτοί οι κανόνες μπορούμε να διαπιστώσουμε ποιός μιλάει και ποιός θα μιλήσει στη συνέχεια. Αυτή η διαδικασία αναφέρεται συχνά στη βιβλιογραφία ως “στροφή” και ορίζεται ως η χρονική διάρκεια από το τέλος της ομιλίας ενός ατόμου έως την αρχή της πρώτης λέξης του άλλου συνομιλητή (turn-taking).

Οι “στροφές” συχνά κυμαίνονται από πολύ μικρές έως αρκετά μεγάλες, υποδεικνύοντας μικρότερης ή μεγαλύτερης διάρκειας αντίστοιχα συναισθηματικά και αγχωτικά επεισόδια, όπου παρατηρούνται φαινόμενα όπως οι επικαλύψεις ομιλίας. Αντίστοιχα, στην περίπτωση του εξεταζόμενου σώματος κειμένου (Ενότητα 4.1.3.1), οι μεγαλύτερης διάρκειας απαντήσεις (στροφές) αποτελούν πολύτιμη πληροφορία σχετικά με την αντιληπτική, γνωστική και συναισθηματική κατάσταση του εκπαιδευόμενου, αντανακλώντας τόσο την εξωτερική του κατάσταση, που μπορεί εύκολα να παρατηρηθεί, όσο και τη μη παρατηρήσιμη εσωτερική συναισθηματική του κατάσταση. Επιλέγουμε λοιπόν να διερευνήσουμε αυτό το είδος αλληλεπιδραστικού πλαισίου μεταξύ των δύο μελών και των δύο ομάδων, βασιζόμενοι στην υπόθεση ότι η συμπεριφορά του εκπαιδευτή είναι περισσότερο ελεγχόμενη, ελαχιστοποιώντας έτσι την επίδραση των διακυμάνσεων της συναισθηματικής του κατάστασης στη συναισθηματική κατάσταση του εκπαιδευόμενου.

Στα πλαίσια της προσπάθειάς μας να διαχωρίσουμε τις απαντήσεις (στροφές) ως προς τη διάρκειά τους (Σχήμα 4.1) σχεδιάζουμε ένα κατώφλι στο 70ο εκατοστημόριο των τιμών απόκρισης. Η τιμή αυτή υπολογίζεται εμπειρικά μετά από τη σχεδίαση των ιστογραμμάτων της στροφής των απαντήσεων του κάθε εκπαιδευόμενου ξεχωριστά. Οι αρνητικές τιμές αυτού του μέτρου αναφέρονται στις περιπτώσεις όπου ο εκπαιδευόμενος διακόπτει την ομιλία του εκπαιδευτή. Ευρήματα όπως οι αλληλοεπικαλύψεις ομιλίας και οι πολύ σύντομες διάρκειες προτάσεις ευθυγραμμίζονται με τα υψηλά επίπεδα του άγχους σε ιδιαίτερα απαιτητικά επιχειρησιακά καθήκοντα [97].

Μετά από προσεκτική επισκόπηση των διαφόρων περιπτώσεων της συναισθηματικής κατάστασης των μελών της ομάδας, παραθέτουμε μία σειρά ενδιαφέροντων παρατηρήσεων. Εντοπίζουμε μία σειρά παραδειγμάτων κατά τη διάρκεια της αλληλεπίδρασης στα οποία ο εκπαιδευτής εξηγεί τον τρόπο σύνδεσης των μπλε και των κόκκινων καλωδίων. Σε αυτή την περίπτωση, η απάντηση του εκπαιδευόμενου είναι σύντομη (δηλ., ο εκπαιδευόμενος χρησιμοποιεί λέξεις όπως *ok/ναι/όχι*). Αντίστοιχα, υπάρχουν και παραδείγματα όπου ο εκπαιδευτής εξηγεί με περισσότερες λεπτομέρειες τη διαδικασία απόπλισης της βόμβας για να επιβεβαιώσει κατά πόσο ο εκπαιδευόμενος ακολουθεί πραγματικά τις οδηγίες του.



Σχήμα 4.1: Οι συχνότητες (άξονας y) για δέκα (10) εκπαιδευόμενους σε συνθήκες IB και CT αντίστοιχα και οι κατανομές των απαντήσεών τους (άξονας x) σε δευτερόλεπτα αναφορικά με την διακοπή ομιλίας του εκπαιδευτή από τον εκπαιδευόμενο. Η κατακόρυφη διακεκομμένη μπλε γραμμή ορίζει το κατώφλι σύμφωνα με το οποίο διακρίνεται η απάντηση μικρότερης και μεγαλύτερης διάρκειας αντίστοιχα.

Καταλήγουμε λοιπόν στη διαπίστωση ότι στην πρώτη περίπτωση, όπου η δήλωση είναι απλή, απαιτείται χαμηλή γνωστική προσπάθεια, δηλαδή μία σύντομη απάντηση, ενώ μία μεγαλύτερης διάρκειας απάντηση ακολουθεί στη δεύτερη περίπτωση, όπου ο εκπαιδευόμενος επαναλαμβάνει τις οδηγίες του εκπαιδευτή για να αποδείξει ότι τις κατάλαβε. Σε αυτή την περίπτωση, αναμένουμε ότι ο εκπαιδευόμενος θα βρίσκεται σε πνευματική εγρήγορση.

4.1.5 Εξαγωγή χαρακτηριστικών Φυσιολογίας

Ένα BIOPAC MP150, με μία τυποποιημένη διαμόρφωση ηλεκτροδίων μολύβδου II χρησιμοποιήθηκε για την καταγραφή των ηλεκτροκαρδιογραφικών αλλαγών (ECG), των μεταβολών της πίεσης του αίματος σε πραγματικό χρόνο και της καρδιογραφικής αντίστασης (impedance cardiography (ZKG)). Τα συνεχή δεδομένα καταγράφηκαν για κάθε μέλος και των δύο ομάδων που συμμετέχουν σε κάθε επιχειρησιακό καθήκον και αναλύονται εκτός σύνδεσης. Οι πρώτες χρονικές σειρές για κάθε επιχειρησιακό καθήκον (task) χωρίζονται σε διαστήματα των τριάντα δευτερολέπτων (30 secs) ξεκινώντας από το τέλος κάθε συνεδρίας (session). Η επιλογή των τριάντα δευτερολέπτων σχετίζεται με το γεγονός ότι αυτή είναι η απαιτούμενη διάρκεια για την ανάλυση καρδιαγγειακών δεδομένων για περαιτέρω ανάλυση. Η εξαγωγή των χαρακτηριστικών από τα δεδομένα πραγματοποιήθηκε με τη χρήση του προγράμματος Moving Ensemble Average Program (MEAP) [39].

Συγκεκριμένα, εξήχθησαν εικοσιτέσσερα (24) καρδιαγγειακά χαρακτηριστικά. Μεταξύ αυτών των χαρακτηριστικών συμπεριλαμβάνονται: ο καρδιακός ρυθμός (HR), ο αριθμός κολιακός (συσταλτικός) χρόνος (Left Ventral (systolic) Ejection Time (LVET)),

Μέλη Ομάδας	Σενάριο Συνθήκης	Επιλεχθέντα χαρακτηριστικά Φυσιολογίας
P100	IB	s_time, systole_time
P101	IB	hr
P102	IB	lvet,p_time, x_time
P103	IB	hr
P104	IB	t_time
P200	CT	diastole_time
P201	CT	hr
P202	CT	pep
P203	CT	hr
P204	CT	hr

Πίνακας 4.1: Επιλεχθέντα χαρακτηριστικά Φυσιολογίας των δέκα (10) εκπαιδευόμενων, πέντε (5) ατόμων της κάθε συνθήκης (IB και CT) αντίστοιχα κατά τη διάρκεια ενός επιχειρησιακού έργου απόπλισης μίας προσομοιωμένης βόμβας. Τα χαρακτηριστικά επισημαίνονται ως εξής: καρδιακός ρυθμός (heart rate (hr)), αριστερός κοιλιακός (συσταλτικός) χρόνος (Left Ventral (systolic) Ejection Time (LVET)) και προσυσταλτική περίοδος (pre-ejection period (PEP)).

ο χρόνος_p, ο χρόνος_s, ο χρόνος_t, η χρονική διάρκεια συστολής της καρδιάς, η προσυσταλτική περίοδος (pre-ejection period (PEP)), η κοιλιακή συσταλτικότητα (Ventricular Contractility (VC)), η καρδιακή παροχή (Cardiac Output (CO)) και η ολική περιφερική αντίσταση (Total Peripheral Resistance (TPR)). Η προσυσταλτική περίοδος (PEP) είναι ο χρόνος από την έναρξη της καρδιακής αποβολής του μυός έως το άνοιγμα της αορτικής βαλβίδας. Όταν μειώνεται η προσυσταλτική περίοδος (PEP), αυξάνεται η κοιλιακή συσταλτικότητα (VC). Το VC σύμφωνα με τη βιβλιογραφία έχει αποδειχθεί ότι σχετίζεται με τα υψηλά επίπεδα άγχους κατά τη διάρκεια απαιτητικών επιχειρησιακών εργασιών [158, 188, 221, 210]. Αντίστοιχα, η καρδιακή παροχή (CO) είναι η ποσότητα του άντλησης του αίματος σε λίτρα ανά λεπτό. Το TPR απεικονίζει την αγγειοδιαστολή (δηλ., την περισσότερη ροή αίματος) και τη αγγειοσυστολή (δηλ., τη λιγότερη ροή), έννοιες οι οποίες σχετίζονται με την παρασυμπαθητική και τη συμπαθητική δραστηριότητα, αντίστοιχα.

Ερευνητικές προσπάθειες στον τομέα της φυσιολογίας έχουν αποδείξει ότι η TPR αυξάνεται αναμφισβήτητα όταν ένα άτομο βρίσκεται σε κατάσταση απειλής και μειώνεται αντίστοιχα σε μία κατάσταση πρόκλησης, ενώ η καρδιακή παροχή (CO) είτε παραμένει αμετάβλητη είτε μειώνεται σε κατάσταση απειλής και αυξάνεται σε συνθήκες πρόκλησης [233].

4.1.6 Πειραματικά Αποτελέσματα

Στην ενότητα αυτή, στόχος μας είναι να ξεδιπλώσουμε τον άμεσο συσχετισμό μεταξύ των επιπέδων της εξωτερικής κοινωνικό-γνωστικής κατάστασης και των εσωτερικών συναισθηματικών μηχανισμών. Χρησιμοποιώντας ένα σύνολο μάθησης (ensemble learning)

προσπαθούμε να εκμεταλλευτούμε διαφορετικές συμπεριφορές των επιλεγμένων βασικών ταξινομητών (base learners) για να βελτιώσουμε την ακρίβεια των ταξινομητών του συνολικού συστήματος εκμάθησης. Σκοπός μας είναι να δείξουμε ότι τα μοτίβα φυσιολογίας των εκπαιδευόμενων διαφέρουν μεταξύ περιόδων μεγαλύτερων και μικρότερων απαντήσεων και ότι υπάρχουν επίσης διαφοροποιήσεις ως προς τα επίπεδο άγχους μεταξύ των εκπαιδευομένων.

4.1.6.1 Μεθοδολογία

Εξαγωγή χαρακτηριστικών: Λαμβάνοντας υπόψιν ότι ορισμένα από τα χαρακτηριστικά σχετίζονται μεταξύ τους σε μεγάλο βαθμό, μειώνουμε το σύνολο αυτό επιλέγοντας τα χαρακτηριστικά με τη μέθοδο Correlation Feature Selection (CFS). Συγκεκριμένα, το CFS επιλέγει τα χαρακτηριστικά που συσχετίζονται με την ετικέτα κλάσης, αλλά δεν σχετίζονται με τα χαρακτηριστικά που έχουν επιλεγεί προηγουμένως.

Σύνολα Μάθησης (Ensemble Learning): Επιθυμώντας να διατηρήσουμε την ανομοιογένεια των βασικών εκπαιδευτών [55] χρησιμοποιούμε δύο ετερογενή συστήματα εκπαίδευσης: Voting [214] και Meta-learning [259] και σχολιάζουμε το μέγεθος του συνόλου των ταξινομητών. Δημιουργούμε λοιπόν τα δύο ακόλουθα σύνολα συνδυάζοντας τους ακόλουθους βασικούς ταξινομητές: K-Nearest Neighbor με $K=5$ (KNN)¹, Naive Bayes (NB), Decision Tree (DT), Random Tree (RT), Support Vector Machines με RBF kernel (SVM-RBF), Multi-layer Perceptron (MLP) και Random Forest (RF)[21]². Τα πειράματα εκτελέστηκαν χρησιμοποιώντας τη διαδικασία leave-one-instance-out ελέγχου εγκυρότητας. Η προσέγγιση αυτή εφαρμόστηκε για κάθε εκπαιδευόμενο ξεχωριστά και για τις δύο ομάδες και των δύο συνθηκών IB και CT αντίστοιχα, καθώς θέλαμε να διερευνήσουμε τις μοναδικές ιδιαιτερότητες του κάθε συμμετέχοντος αναφορικά με τη συναισθηματική του κατάσταση, τη φυσιολογία του και με το κατά πόσο η εξοικείωσή του με το άλλο μέλος της ομάδας τον επηρεάζει.

4.1.6.2 Αποτελέσματα

Οι βασικοί ταξινομητές (base learners) και οι μέθοδοι μάθησης του συνόλου (ensemble learning) που επιλέχθηκαν για τη μελέτη μας παρατίθενται στον Πίνακα 4.2. Τα πειραματικά αποτελέσματά μας κυμαίνονται από 43,75% έως 88,89% γεγονός που υποδηλώνει ότι τα σήματα φυσιολογίας περιέχουν πληροφορίες που σχετίζονται με την ποσότητα και την ποιότητα των απαντήσεων και είναι ενδεικτικές της συναισθηματικής κατάστασης

¹Το $K=5$ επιλέχθηκε εμπειρικά καθώς οδηγούσε σε καλύτερη απόδοση του συστήματος, λαμβάνοντας υπόψιν τον περιορισμό του μικρού μεγέθους των δειγμάτων δύο εκπαιδευόμενων (οι P104 και P202 αποτελούνται από 9 και 10 δείγματα αντίστοιχα).

²Ο ταξινομητής αυτός είναι ισχυρός και δουλεύει σχετικά ικανοποιητικά χωρίς την ανάγκη τροποποίησης των παραμέτρων του.

Participants	Condition	Base Learners							Ensemble Learners	
		KNN(5)	NB	DT	RT	SVM-RBF	MLP	RF	Voting	Meta-learning
P100	IB	68.97	75.86	62.07	75.86	72.41	65.52	75.86	72.41	72.41
P101	IB	43.75	56.25	50.00	50.00	43.75	50.00	50.00	43.75	62.50
P102	IB	80.00	73.33	73.33	73.33	40.00	66.67	73.33	73.33	73.33
P103	IB	34.38	46.88	53.13	43.75	59.38	53.13	43.75	59.38	62.50
P104	IB	66.67	44.44	88.89	66.67	55.56	66.67	66.67	55.56	55.56
P200	CT	77.78	66.67	88.89	88.89	88.89	77.78	88.89	88.89	83.33
P201	CT	43.75	43.75	56.25	18.75	31.25	50.00	25.00	31.25	62.50
P202	CT	70.00	80.00	90.00	60.00	60.00	80.00	80.00	60.00	70.0
P203	CT	88.00	88.00	88.00	72.00	88.00	84.00	72.00	88.00	88.00
P204	CT	77.78	66.67	77.78	77.78	77.78	66.67	66.67	77.78	66.67

Πίνακας 4.2: Οι βασικοί ταξινομητές (base learners) και το σύνολο μάθησης (ensemble learning) που χρησιμοποιήθηκαν για τους δέκα (10) εκπαιδευόμενους, πέντε (5) για κάθε κατάσταση (IB και CT) αντίστοιχα. Η καλύτερη μέθοδος για κάθε εκπαιδευόμενο υπογραμμίζεται με έντονο χρώμα.

των εκπαιδευόμενων. Επιπροσθέτως, παρατηρούμε μία μεγάλη διαφορά στην απόδοση των εκπαιδευόμενων, υπογραμμίζοντας και πάλι τη σημασία των μοναδικών χαρακτηριστικών του κάθε εκπαιδευόμενου. Ιδιαίτερα, τα επιλεγθέντα χαρακτηριστικά φυσιολογίας των εκπαιδευόμενων P100, P102, P202, P203 και P204 είναι εμφανώς περισσότερο συνδεδεμένα με τον τύπο της απάντησης ως προς τη διάρκειάς της (μικρή/μεγάλη) σε σύγκριση με τα αντίστοιχα μοτίβα των εκπαιδευόμενων P101, P103, P104 και P201.

Ανάλυση επιδόσεων της μάθησης του συνόλου (ensemble learning) και του μεγέθους της: Παρατηρούμε ότι τα σύνολα για τους δύο διαφορετικούς συνδυασμούς που υλοποιήθηκαν είτε ξεπερνούν τον καλύτερο βασικό ταξινομητή (P101, P103, P201) είτε πλησιάζουν τα ίδια ποσοστά απόδοσης με τους εκπαιδευόμενους P200, P203, P204. Παρόλο που πειραματιστήκαμε με ένα μέγεθος που κυμαίνεται από 5 έως 10 βασικούς ταξινομητές, παρουσιάζουμε μόνο τους περισσότερο ακριβείς. Επιπλέον, έχουμε εξετάσει με έναν άρτιο και περιττό αριθμό μεγέθους του συνόλου. Πειραματικά, λοιπόν διαπιστώνουμε ότι η χρήση περιττού αριθμού αναφορικά με το μέγεθος του συνόλου παρέχει υψηλότερη απόδοση μάθησης.

Επιλεγθέντα χαρακτηριστικά φυσιολογίας: Επεξεργαζόμενοι τα χαρακτηριστικά που παρουσιάζονται στον Πίνακα 4.1 αναφορικά με την απόπλιση της βόμβας, διαπιστώνουμε ότι για το 50% των εκπαιδευόμενων το χαρακτηριστικό φυσιολογίας που ενεργοποιείται περισσότερο συχνά είναι ο καρδιακός ρυθμός (HR). Βάσει αυτού, υποθέτουμε ότι το HR σχετίζεται άμεσα με τα επίπεδα διέγερσης (arousal) και έχει μεγάλη σημασία για το εξεταζόμενο έργο. Όσον αφορά στα υπόλοιπα χαρακτηριστικά φυσιολογίας, παρατηρούμε ότι αυτά ποικίλουν μεταξύ των εκπαιδευόμενων. Αυτό το εύρημα ενισχύει την αρχική παραδοχή της μοναδικότητας των χαρακτηριστικών μεταξύ των συμμετεχόντων. Τέλος, η πρώτη διαπίστωση είναι επίσης ευθυγραμμισμένη με τα πειραματικά αποτελέσματα και

τις παρατηρήσεις μας, υποδηλώνοντας ότι η παράμετρος αναφορικά με την εξοικείωση των μελών της ομάδας πριν την έναρξη του έργου τουλάχιστον στο εξεταζόμενο τμήμα του σώματος κειμένου, δεν επηρεάζει τα αποτελέσματά μας.

4.1.7 Συμπεράσματα

Διαπιστώνουμε λοιπόν, όπως αναφέρθηκε και στην Ενότητα 4.1.6.2, ότι υπάρχει ένα ευρύ φάσμα μεταβλητότητας μεταξύ των εκπαιδευόμενων αναφορικά με την εργασία που καλούνται να πραγματοποιήσουν, δηλαδή την επιτυχή απόπλιση της βόμβας. Αυτή η παρατήρηση υποδεικνύει ότι ενδεχομένως υπάρχουν μηχανισμοί που ενεργοποιούνται στους εκπαιδευόμενους μέσω των σημάτων φυσιολογίας τους, τα οποία δεν εντοπίζονται παρά μόνο όταν μελετηθούν οι επιδόσεις της μάθησης του συνόλου (ensemble learning), που όμως δεν φαίνεται να είναι αποδοτικές για τους εκπαιδευόμενους P101, P103, P201.

Αξίζει επίσης να σημειωθεί ότι σε αυτούς τους τρεις εκπαιδευόμενους, παρατηρείται αύξηση του καρδιακού τους ρυθμού (HR). Για περαιτέρω επεξεργασία αναφορικά με αυτή την ένδειξη, εξετάζουμε τις οπτικοακουστικές εγγραφές και τα σήματα του καρδιακού τους παλμού (HR). Παρατηρούμε ότι υπάρχει μία διαφορά στα επίπεδα διέγερσης (δηλ. άγχους) σε σχέση με το είδος της διάρκειας της απάντησης (μικρότερης/μεγαλύτερης) και ότι η διέγερση αυτή εμπεριέχεται τόσο σε μικρότερη όσο και σε μεγαλύτερη διάρκεια απαντήσεων, ανάλογα με τον εκάστοτε εκπαιδευόμενο.

Συγκεκριμένα, συναντάμε παραδείγματα εκπαιδευόμενων οι οποίοι χρειάστηκαν αρκετό χρόνο για να απαντήσουν αφού πρώτα έδωσαν λανθασμένη απάντηση και κατά συνέπεια τους ζητήθηκε να προσπαθήσουν ξανά για να επιβεβαιώσουν ότι κατανόησαν τα βήματα της απόπλισης της βόμβας. Ως εκ τούτου, φαίνεται ότι η εργασία την οποία κλήθηκαν να υλοποιήσουν αποτέλεσε ένα αρκετά αγχωτικό ερέθισμα για αυτούς και συνεπώς τα επίπεδα τόσο της γνωστικής δραστηριότητάς τους όσο και του άγχους τους αντίστοιχα ήταν υψηλά.

Την ίδια στιγμή, παρατηρήσαμε υψηλά επίπεδα διέγερσης και σε σύντομης διάρκειας παραδείγματα, κατά τα οποία, ο εκπαιδευόμενος χρησιμοποιούσε λέξεις όπως *ok/ναι/όχι*. Μία τέτοια συναισθηματική κατάσταση δεν ευθυγραμμίζεται με το έργο “απόπλισης της βόμβας”, δεδομένου ότι περιμέναμε ότι οι μικρότερης διάρκειας απαντήσεις θα σχετίζονται με χαμηλά επίπεδα διέγερσης. Αντιθέτως, τα πειραματικά μας αποτελέσματα υποδηλώνουν ότι παρόλο που ενδέχεται να μην υπάρχουν προφανή (ακουστικά/ορατά) σήματα διέγερσης, τα σήματα φυσιολογίας είναι ικανά να παρέχουν μία συμπληρωματική αλλά μη επικαλυπτόμενη πληροφορία αναφορικά με την κατάσταση ενός ατόμου γενικότερα.

Αυτό το εύρημα είναι μείζονος σημασίας, ιδιαίτερα σε απαιτητικά γνωστικά επιχειρησιακά καθήκοντα στα οποία ένα από τα δύο μέλη της ομάδας ελέγχει και κατευθύνει τη συζήτηση. Η συγκεκριμένη διαπίστωση ευθυγραμμίζεται και με προηγούμενες παλαιότερες αλλά και σύγχρονες ερευνητικές μελέτες [233, 79, 27].

Κεφάλαιο 5

Μονοτροπική Αναγνώριση Συναισθηματικών Καταστάσεων

5.1 Αναγνώριση Συναισθηματικών Καταστάσεων από Λεκτικό Περιεχόμενο

5.1.1 Εισαγωγή

Η ικανότητα των ανθρώπων να αλληλεπιδρούν αποτελεσματικά σε κοινωνικό επίπεδο εξαρτάται σε μεγάλο βαθμό από την ευαισθητοποίησή (awareness) τους στο πλαίσιο της αλληλεπίδρασης που λαμβάνει χώρα, όπου είναι σημαντικό να γνωρίζουμε τις πτυχές ενός φορμαλισμού του πλαισίου **W5+**, δηλαδή τις απαντήσεις στα ερωτήματα: **Who, What, Why, Where, What** και **How**. Ο στόχος αυτός επιτυγχάνεται γεφυρώνοντας το εννοιολογικό κενό και λαμβάνοντας υπόψιν τη σημασιολογία, τις γνωστικές και τις συναισθηματικές πληροφορίες της μη λεκτικής συμπεριφοράς. Διερευνούμε λοιπόν, το εάν και πώς αυτές οι πληροφορίες θα μπορούσαν να ενσωματωθούν στην αυτόματη ανάλυση της συναισθηματικής συμπεριφοράς. Έτσι, προτείνουμε μία μεθοδολογία εξαγωγής σημασιολογικής έννοιας, η εφαρμογή της οποίας παρουσιάζεται σε ενδεικτικά παραδείγματα από το σώμα κειμένου SEMAINE επικυρώνοντας την προτεινόμενη προσέγγιση.

Μία από τις κύριες προκλήσεις των τελευταίων ετών είναι η δημιουργία περισσότερο φυσικών, ευαίσθητων και ευφών μηχανών, που δε θα είναι σε θέση να επικοινωνούν μόνο, αλλά θα μπορούν να κατανοήσουν τα σήματα καθώς και το πλαίσιο μέσα στο οποίο θα λαμβάνει χώρα ένα γεγονός [240]. Έτσι, εκτός από τη μονοτροπική ή πολυτροπική επικοινωνία μέσω ενός είτε περισσοτέρων καναλιών και μέσω του λεκτικού περιεχομένου (σημασιολογία), οι μηχανές υπολογιστών χρειάζεται να αναγνωρίζουν, να ερμηνεύουν και να επεξεργάζονται τις συναισθηματικές πληροφορίες όπως οι άνθρωποι. Στην ανθρώπινη γνώση, η σκέψη και το αίσθημα είναι άρρηκτα συνδεδεμένες έννοιες: τα συναισθήματά μας είναι συχνά προϊόν των δικών μας σκέψεων, καθώς και οι αντιδράσεις μας είναι συχνά το προϊόν αποτελέσματος της συναισθηματικής μας κατάστασης. Αλλά,

τί σημαίνει για ένα μηχάνημα να είναι κοινωνικά ευφύες μετά την ενσωμάτωση του πλαισίου; Μέχρι στιγμής, όπως συζητήθηκε και στην Ενότητα 1.2 και στο Κεφάλαιο 2 κατά τη διάρκεια συνομιλιών, το εννοιολογικό πλαίσιο έχει οριστεί ως η προηγούμενη οπτική πληροφορία [62], η γενική κατανόηση της κατάστασης [30], η παρελθοντική λεκτική πληροφορία [119], το πολιτιστικό υπόβαθρο των συνομιλητών [140], το φύλο τους, η γνώση του γενικότερου πλαισίου αλληλεπίδρασης στο οποίο πραγματοποιείται ένα συναισθηματικό φαινόμενο [24] και οι κοινωνικές καταστάσεις [16]. Κατά συνέπεια, μελέτες σχετικά με έξυπνες διεπαφές για την επικοινωνία Ανθρώπου-Υπολογιστή (HCI), οι οποίες ενσωματώνουν το εννοιολογικό πλαίσιο, αντιστοιχίζονται με τα ακόλουθα ερωτήματα τα οποία συνθέτουν το **W5+ φορμαλισμό**:

- Με **Ποιόν** είσαι μαζί; (π.χ., αλληλεπιδράσεις μεταξύ δύο ή περισσότερων ατόμων [242]),
- **Τί** μεταδίδεται; (π.χ., (μη) γλωσσικό μήνυμα/λεκτικό σήμα και συναίσθημα),
- **Πώς** μεταδίδονται οι πληροφορίες; (μέσω των ανθρωπίνων συναισθηματικών συνθημάτων),
- **Γιατί**, δηλαδή, σε ποιο πλαίσιο οι πληροφορίες μεταβιβάζονται;
- **Πού** είναι ο χρήστης;
- **Ποιά** είναι η τρέχουσα εργασία του;
- **Πώς** αισθάνεται (η διάθεσή του να είναι μεταβαλλόμενη από αρνητική σε θετική;)
- **Ποιά** είναι τα μέτρα που πρέπει να ληφθούν για την ικανοποίηση των αναγκών, των στόχων και των καθηκόντων του ανθρώπου; [60].

Δυστυχώς, μέχρι στιγμής οι προσπάθειες της κατανόησης των ανθρωπίνων συναισθημάτων είναι συνήθως ανεξάρτητες από το περιβάλλον-πλαίσιο στο οποίο πραγματοποιούνται [91]. Υπό το πρίσμα αυτών των παρατηρήσεων, η κατανόηση της διαδικασίας της φυσικής εξέλιξης των σχετικών με το πλαίσιο ερωτημάτων (context-related), όταν οι άνθρωποι αλληλεπιδρούν σε ένα κοινωνικό περιβάλλον, θα μπορούσε να προσφέρει νέες ιδέες για το μηχανισμό του πλαισίου αλληλεπίδρασης και της αποτελεσματικότητάς του.

Σύμφωνα με τη βιβλιογραφία τα ερωτήματα “Who”, “What”, “Where” έχουν απαντηθεί κυρίως ξεχωριστά ή σε ομάδες δύο ή τριών ερωτημάτων χρησιμοποιώντας τις πληροφορίες που εξάγονται από τα πολυτροπικά σήματα [266]. Έτσι, μέχρι σήμερα, δεν υπάρχει ένας συγκεκριμένος ορισμός του φορμαλισμού **W5+**, δεδομένου ότι τα συστήματα που απαντούν στις περισσότερες ερωτήσεις “W” βασίζονται σε διαφορετικές ψυχολογικές θεωρίες του συναισθήματος και αναπαραστάσεις μοντέλων ενσυναίσθησης εξυπηρετώντας κάθε φορά τους συγκεκριμένους σκοπούς και στόχους της εκάστοτε έρευνας.

Πρόσφατη έρευνα σχετικά με την πρόοδο στις ερωτήσεις του “Why” και “How” οδήγησε στον αναδυόμενο ερευνητικό τομέα της ανάλυση του συναισθήματος (Sentiment

Analysis (SA)) από σώματα κειμένων Φυσικής Γλώσσας [165, 133, 29], μέσω του εντοπισμού και της εξόρυξης απόψεων (Opinion Mining (OM)) και συναισθημάτων (sentiments), όπου στόχος είναι η βαθιά κατανόηση των σημασιολογικών κανόνων της γλώσσας. Επιπλέον, η ερμηνεία των γνωστικών και συναισθηματικών πληροφοριών που σχετίζεται με τη Φυσική Γλώσσα και, ως εκ τούτου, βοηθούν στην απόκτηση νέων γνώσεων και στη λήψη αποφάσεων, σε σχέση με τις κοινωνικό-συναισθηματικές αξίες και τα ιδανικά του ατόμου, είναι ζωτικής σημασίας. Το πρόβλημα κατά την προσπάθεια μίμησης αυτών των γνωστικών και λειτουργικών διαδικασιών, είναι ότι ενώ οι γνωστικές πληροφορίες είναι συνήθως αντικειμενικές και αμερόληπτες, ενώ η απάντηση των ερωτημάτων “Γιατί” είναι περισσότερο υποκειμενική.

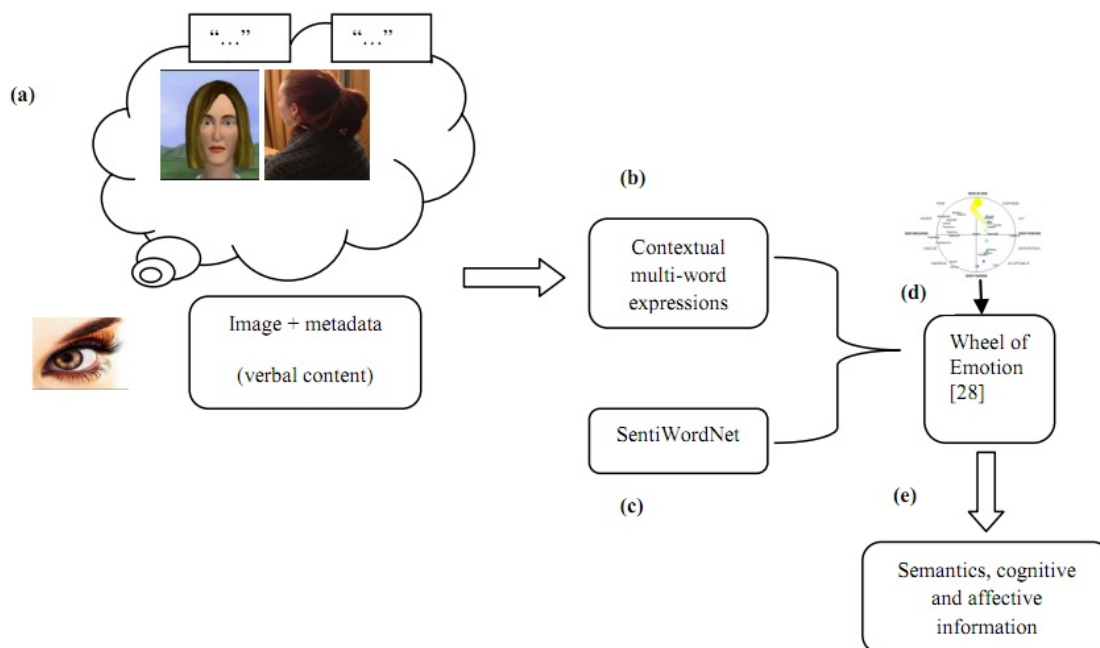
Σύμφωνα με αυτή την άποψη, ο μακροπρόθεσμος στόχος μας είναι να κατανοήσουμε εάν και με ποιό τρόπο το εννοιολογικό πλαίσιο μπορεί να ενσωματωθεί στην αυτόματη ανάλυση της ανθρώπινης συμπεριφοράς και να προτείνουμε ένα νέο πλαίσιο ενσωμάτωσης (Εικόνα 5.1) το οποίο να (1) περιλαμβάνει τον εντοπισμό και την εξαγωγή του σημασιολογικού πλαισίου, (2) να εμπλουτίζει περισσότερο μία σειρά Ψυχολογικών Βάσεων (Psychological Foundations) με τιμές αισθήσεων και (3) να ενισχύει τα συναισθηματικά μοντέλα με πληροφορίες του πλαισίου και με αναπαραστάσεις αναφορικά με την εκτίμηση αξιολόγησης, χρησιμοποιώντας διαθέσιμες στο κοινό πηγές γνώσης (OKS) στην επεξεργασία της Φυσικής Γλώσσας [244]. Το πρώτο βήμα προς αυτή την κατεύθυνση, είναι η διερεύνηση της σημασιολογικής γνωστικής και συναισθηματικής πληροφορίας, που σχετίζεται με το λεκτικό περιεχόμενο (σημασιολογία), το οποίο για τις ανάγκες της έρευνάς μας είναι ο διάλογος μεταξύ του χρήστη και του χειριστή του σώματος κειμένου SEMAINE [143], διατηρώντας σταθερό το σχετικό ερώτημα αναφορικά με το Πού πραγματοποιείται αυτή η συνομιλία.

Αυτή η μέθοδος επισημείωσης λαμβάνοντας υπόψιν το εννοιολογικό πλαίσιο, δεν περιορίζεται μόνο σε μια απλή συντακτική ανάλυση της συνομιλίας¹. Στις παραδοσιακές μεθόδους επισημείωσης, τα συναισθήματα και οι πληροφορίες από τα συμφραζόμενα δεν συνάγονται πάντα από τις γνωσιακές εκτιμήσεις (appraisals) και συνεπώς οι πληροφορίες που σχετίζονται με το πλαίσιο αναφορικά με τις αιτίες της δημιουργίας αυτών των συναισθημάτων δεν λαμβάνονται υπόψιν [31].

5.1.2 Σχετική Βιβλιογραφία

Τα συναισθήματα είναι περίπλοκες καταστάσεις των ανθρωπίνων αισθημάτων (feelings) και οδηγούν σε σωματικές και ψυχολογικές αντιδράσεις επηρεάζοντας τόσο τη σκέψη όσο και τη συμπεριφορά μας. Η μελέτη των συναισθημάτων παραμένει ακόμα ένα ουσιαστικό και ανοικτό προς μελέτη κομμάτι της Ψυχολογίας. Αναφορικά με τον τομέα της Φυσικής Γλώσσας (Natural Language Processing (NLP)) το ερευνητικό ενδιαφέρον εστιάζεται στον εντοπισμό του συναισθήματος μέσα από το κείμενο. Κυρίως, η έρευνα σχετικά

¹Το μέγεθος του εξεταζόμενου παραθύρου της συνομιλίας αντιστοιχεί σε 16 στροφές.



Σχήμα 5.1: Επισκόπηση του προτεινόμενου συστήματος: (a) Εντοπισμός και εξαγωγή του σημασιολογικών εννοιών πλαισίου από το σώμα κειμένου SEMAINE. (b) Αναπαράσταση κάθε έννοιας με εκφράσεις πολλαπλών λέξεων ενισχυμένες με τιμές που αντιστοιχούν σε συναισθήματα. (c) Μία σειρά Ψυχολογικών Βάσεων (Psychological Foundations) εμπλουτίζεται. (d) Δείχνουμε ότι η προτεινόμενη προσέγγιση μπορεί να συσχετίσει όρους σημασιολογίας, γνωστικής και συναισθηματικής πληροφορίας.

με την ανίχνευση των συναισθημάτων από το κείμενο έχει επικεντρωθεί καταγράφοντας λέξεις συναισθηματικού περιεχομένου με βάση τρία μοντέλα συναισθημάτων, δηλ. τις αναπαραστάσεις σε κατηγορίες, τις διαστατικές αναπαραστάσεις και τις κατηγορίες γνωστικής αξιολόγησης, όπως αναπτύχθηκε στην Ενότητα 3.2.3.1 [28, 90].

Σημασιολογικές έννοιες πλαισίου. Για ένα περισσότερο εφαρμόσιμο σημασιολογικό μοντέλο ενσωμάτωσης πλαισίου, έναντι ενός θεωρητικού, όπως το μοντέλο συνιστωσών (Componential Model), η έρευνα έχει επικεντρωθεί στη συγκέντρωση γνώμων και συναισθημάτων από τη Φυσική Γλώσσα. Το εγχείρημα αυτό θεωρείται μία ερευνητική πρόκληση, καθώς προϋποθέτει μία βαθιά κατανόηση των έμμεσων και σημασιολογικών γλωσσικών κανόνων, που σχετίζονται με μία σειρά ανεπίλυτων προβλημάτων της Φυσικής Γλώσσας, όπως ο χειρισμός άρνησης, η αναγνώριση ονομάτων-οντοτήτων, η άρση της αμφισημίας λέξεων κλπ. Οι προσεγγίσεις με βάση την έννοια [234] στοχεύουν στην κατανόηση της εννοιολογικής και της συναισθηματικής πληροφορίας που σχετίζεται με τις γλωσσικές απόψεις. Επιπροσθέτως, οι προσεγγίσεις που βασίζονται σε έννοιες μπορούν να αναλύσουν εκφράσεις πολλών λέξεων που δε μεταφέρουν απαραίτητα συγκίνηση, αλλά σχετίζονται με έννοιες που το κάνουν. Για παράδειγμα, αντί της συγκέντρωσης μεμονωμένων απόψεων για ένα “συμβάν” (π.χ. πάρτυ γενεθλίων), οι χρήστες γενικά ενδιαφέρονται περισσότερο για τη σύγκριση των διαφορετικών γεγονότων σύμφωνα με

ένα συγκεκριμένο σύνολο σημασιολογικά συναφών εννοιών, π.χ. “κεικ”, “ένας έκπληκτος φίλος”, ή “το δώρο” (το οποίο μπορεί να θεωρηθεί ως επιπρόσθετη πληροφορία του εννοιολογικού πλαισίου για τη βελτίωση της αναζήτησης σχετικών αποτελεσμάτων). Αυτή η πληροφορία αναφέρεται σε γνώση η οποία είναι ήδη γνωστή και συνήθως δεν απαιτεί επιπλέον επεξηγήσεις, είναι όμως απαραίτητη για την σωστή αποδόμηση του κειμένου της Φυσικής Γλώσσας σε συναισθήματα. Για παράδειγμα, η έννοια “μικρό δωμάτιο” θα πρέπει να εκτιμηθεί ως αρνητικό σχόλιο για την αξιολόγηση ενός ξενοδοχείου ενώ η έννοια “μικρή ουρά” ως θετική για την αξιολόγηση ενός ταχυδρομείου, είτε η έννοια “διάβασε το βιβλίο” ως θετική για μια κριτική βιβλίου, αλλά αρνητική για μια κριτική ταινίας.

5.1.3 Εποπτική παρουσίαση της προτεινόμενης αρχιτεκτονικής της φυσικής εξέλιξης των σχετικών ερωτημάτων με το πλαίσιο

5.1.3.1 Σώμα Κειμένου

Το προτεινόμενο μοντέλο βασίζεται στο σώμα κειμένου SEMAINE [143]. Το συγκεκριμένο σώμα κειμένου περιλαμβάνει συνεδρίες οι οποίες έχουν επισημειωθεί χειροκίνητα, κατά τις οποίες ένας άνθρωπος-χρήστης αλληλεπιδρά με έναν άνθρωπο-φορέα που ενεργεί ως εικονικός πράκτορας. Αυτές οι αλληλεπιδράσεις βασίζονται σε ένα σενάριο που περιλαμβάνει τέσσερις (4) χαρακτήρες πρακτόρων (Σχήμα 5.2). Οι δηλώσεις/προτάσεις του εικονικού πράκτορα καθορίζονται από ένα σενάριο, ωστόσο, ορισμένες αποκλίσεις στο σενάριο εμφανίζονται στο σώμα κειμένου.



Σχήμα 5.2: Οι τέσσερις χαρακτήρες πρακτόρων: η Porry: χαρούμενη και εξωστρεφής, η Prudence: λογική και συνετή, ο Spike: θυμωμένος και ο Obadiah: καταθλιπτικός και ζοφερός.

5.1.3.2 Προεπεξεργασία

Μέρος της προεπεξεργασίας αποτελεί αρχικά η ερμηνεία όλων των συναισθηματικών ενδείξεων σθένους (affective valence indicators) που περιέχονται συνήθως στο λεκτικό

περιεχόμενο των μεταγραφών (transcriptions), όπως είναι τα ειδικά σημεία στίξης, οι λέξεις που ξεκινούν με κεφαλαίο γράμμα (upper-case words), οι λέξεις που δηλώνουν θαυμασμό και οι αρνήσεις. Ο χειρισμός των αρνήσεων είναι ένας σημαντικός προβληματισμός σε ένα τέτοιο σενάριο, καθώς μπορεί να αντιστρέψει την ερμηνεία της εξεταζόμενης πρότασης. Δεύτερον, μετά τη μετατροπή των κεφαλαίων γραμμάτων σε μικρά, εφαρμόζουμε τη διαδικασία της λημματοποίησης και χωρίζουμε τις προτάσεις σε μεμονωμένες εκφράσεις.

Στη συνέχεια, τα n-grams που παράγονται δεν χρησιμοποιούνται άκριτα ως μοτίβα συγκεκριμένων λέξεων, αλλά χρησιμοποιούνται ως σημεία αναφοράς για τη διαδικασία της προεπεξεργασίας, προκειμένου να εξάγουμε έννοιες πολλαπλών λέξεων από προτάσεις πλούσιες σε πληροφορία. Έτσι λοιπόν, το παρόν τμήμα της προτεινόμενης αρχιτεκτονικής, σε αντίθεση με άλλους ρηχούς αναλυτές (shallow parsers), μπορεί να αναγνωρίζει σύνθετες έννοιες ακόμα και όταν χρησιμοποιούνται ακανόνιστα ρήματα (irregular verbs) είτε όταν βρίσκονται “διάσπαρτα” ενδιάμεσα από επίθετα και επιρρήματα, όπως για παράδειγμα, η έννοια “αγοράζω ένα πασχαλινό δώρο” στην πρόταση “αγόρασα πολλά πολύ ωραία Πασχαλινά δώρα”.

5.1.3.3 Αναλυτής Εννοιών πολλαπλών λέξεων Σημασιολογικού Περιεχομένου

Ο σκοπός του σημασιολογικού αναλυτή (semantic parser) είναι να κατατμήσει τις προτάσεις σε εκφράσεις (clauses) και, ως εκ τούτου, να ανακατασκευάσει αυτές τις εκφράσεις σε έννοιες. Αυτή η αποδόμηση χρησιμοποιεί λεξικά που βασίζονται σε ακολουθίες (lexemes) που αντιπροσωπεύουν τις έννοιες πολλαπλών λέξεων που εξάγονται από το ConceptNet, το WordNet [148] και άλλους γλωσσικούς πόρους. Επίσης, σύμφωνα με τον Stanford Parser ² η λειτουργία του οποίου βασίζεται στο Python NLTK ³, μία γενική παραδοχή κατά τη διάρκεια διαχωρισμού εκφράσεων (clauses) είναι ότι, αν ένα κομμάτι του κειμένου περιέχει μία πρόθεση ή μία υποδεέστερη συσχέτιση, τότε οι λέξεις που προηγούνται αυτών δεν ερμηνεύονται ως γεγονότα αλλά ως αντικείμενα. Δεύτερον, η εξάρτηση της στοιχείων της δομής επεξεργάζονται με τη βοήθεια του Stanford Lemmatizer για κάθε μία πρόταση. Κάθε πιθανό τμήμα του ουσιαστικού που σχετίζεται με ατομικά κομμάτια ρήματος είναι ζευγαρωμένο με το ρήμα από το οποίο προήλθε και χρησιμοποιείται για την ανίχνευση εκφράσεων πολλαπλών λέξεων της μορφής “verb plus object”.

5.1.3.4 Λεξικό Απόψεων και Συναισθημάτων

Οι τρέχουσες προσεγγίσεις για την ανάλυση εννοιών (concept-level) στον τομέα της ανάλυσης συναισθήματος από κείμενο βασίζονται κυρίως στην ύπαρξη βάσεων γνώσεων της ενσυναίσθησης, όπως η ANEW [20], το WordNet-Affect [224] και το SentiWordNet [65]. Ωστόσο, για τις ανάγκες της τρέχουσας εργασίας μας, χρησιμοποιούμε το SentiWordNet,

²<http://nlp.stanford.edu:8080/parser/>

³<http://nltk.org>

το οποίο είναι λεξικό γνώμης σε επίπεδο ιδεών και περιέχει πολλαπλές έννοιες λέξεων και εκφράσεων που χαρακτηρίζονται από τις αντίστοιχες βαθμολογίες πολικότητας.

5.1.4 Ερευνητικά Αποτελέσματα

Σε αυτή την ενότητα, παρουσιάζουμε την ανάλυση και το σχολιασμό μίας σειράς παραδειγμάτων ενδεικτικών της έννοιας του πλαισίου που χρησιμοποιούν το σώμα κειμένου SEMAINE [143]. Επιπλέον, παρέχουμε έναν κατάλογο των κύριων ερευνητικών αποτελεσμάτων και συμπερασμάτων που παρατηρήθηκαν κατά τη διάρκεια της μελέτης μας.

Το πρώτο παράδειγμα προέρχεται από τη **Συνεδρία 70 για τον Prudence**, εστιάζοντας στις φράσεις [16-32] και [48-64]. Κατά τη διάρκεια αυτού του διαδραστικού πλαισίου, η συζήτηση περιστρέφεται γύρω από το ερώτημα: Τί συζητείται; αναφερόμενοι στο θέμα “διακοπές” και στο θέμα “ταξίδι” στις φράσεις [50-64]. Ο χρήστης δεν ενδιαφέρεται για το θέμα της “εργασίας” και κατά συνέπεια λέει ότι το θέμα είναι βαρετό. Κατά τη διάρκεια της αλληλεπίδρασης, τόσο ο χειριστής όσο και ο χρήστης έχουν τις ίδιες υποκειμενικές απόψεις, καθώς και οι δύο επαναλαμβάνουν πολλές φορές τις λέξεις “υπερβολικό” και “παράλογο” αναφερόμενοι στο θέμα “ταξίδι”.

19 - Prudence: “And have you considered where you might go for holidays?”
20 - User: “(Looks around in thought). Ah yeah. I’m thinking about going to Australia.”
...
56 - User: “(Nods smiling). That is absolutely absurd. I concur. (Looks around smiling). But it was... fantastic fun. Eh... four guys doing a road trip from Houston to New Orleans as well. Emm... Obviously myself included. (Licks lips). An... A lot of eating, a lot of drinking, a lot of not talking about research. It was fantastic.”
57 - Prudence: “Well it sounds like it was an... excessive trip.”
58 - User: “Very excessive.”

Το δεύτερο παράδειγμα αναφέρεται στον ρόλο του χειριστή **Spike στη Συνεδρία 73**. Μελετώντας τις φράσεις 56-73, παρατηρούμε ότι ο χρήστης είναι ενθουσιασμένος με το χειριστή. Εκφράσεις πολλαπλών λέξεων όπως “είναι πολύ” και “piss me off” υπογραμμίζονται. Ο χρήστης λέει ότι συχνά τον μπερδεύουν με Αμερικανό, ενώ είναι Καναδός και δείχνει ενοχλημένος από αυτή τη σύγχυση. Λαμβάνοντας υπόψη ότι ο ρόλος του Spike είναι να προκαλέσει το θυμό του χρήστη, αυτές οι εκφράσεις χρησιμοποιούνται για την ενίσχυση της ενόχλησης του τελευταίου.

56 - User: "Hmm... The world doesn't really think highly of Americans."

57 - Spike: "Yeah... But Canadians are just the same aren't they?" . . .

72 - User: "That's a lot. (Smiles, nods)."

73 - Spike: "That would piss me o."

Στη Συνεδρία 72, αναφορικά με το ρόλο του **Obadiah**, εντοπίζουμε έναν αριθμό εκφράσεων που σχετίζονται με τις εκτιμήσεις. Αυτό το εύρημα είναι ευθυγραμμισμένο με το είδος της συναισθηματικής κατάστασης του Obadiah. Σε αυτό το ρόλο, ο χειριστής εκφράζει τις απόψεις του σχετικά με τη ζωή, οι οποίες πυροδοτούν επαναλήψεις μικρής διάρκειας. Αυτό είναι έκδηλο στις φράσεις [37-38] και [41-42], στις οποίες λέξεις συναισθηματικά συνδεδεμένες όπως "ευτυχισμένη", "συναίσθημα", "θλιβερή", "βαρετός" και "ενδιαφέρον" επαναλαμβάνονται διαρκώς.

37 - Obadiah: "Life is hard sometimes."

38 - User: "(Nods). Life can suck sometimes. I agree." . . .

41 - Obadiah: "Yeah. But you can't be cheery all the time."

42 - User: "(Shakes head). Oh God I'm not cheery (laughs).[...]"

Τέλος, **στη Συνεδρία 71, η Poppy** στις φράσεις [24-32] λόγω του χαρακτήρα της φαίνεται να ευθυγραμμίζεται με τις προτάσεις του χρήστη, παρέχοντας ανατροφοδότηση όπως "hmm", ή "yeah", αλλά χωρίς όμως να επαναλαμβάνει τις λέξεις του χρήστη.

24 - User: "Yeah it's... very fast. Very... high contact which I... tend to like. Emm... Haven't done it in a while so (smiling and wide eyed) I guess that makes me a bit sad."

25 - Poppy: "Ah..."

26 - User: "Emm... Yeah."

"Τί συζητείται;" προσδιορισμός του θέματος. Λόγω του γεγονότος ότι στο εξεταζόμενο σώμα κειμένου SEMAINE, μόνο ο χρήστης είναι ο ομιλητής, ο τελευταίος καταλαμβάνει το 65,5% της συνολικής διάρκειας της ομιλίας, συνεπώς, η ομιλία δεν είναι ισοκατανεμημένη μεταξύ του χρήστη και του χειριστή. Επιπλέον, αυτό το φαινόμενο παρατηρείται και κατά τον υπολογισμό του ποσοστού της ομιλίας του χρήστη για όλες τις συνεδρίες που αντιστοιχούν στον ρόλο ενός συγκεκριμένου χειριστή. Διαπιστώνουμε λοιπόν ότι τα ποσοστά δραστηριότητας της ομιλίας κυμαίνονται από 60,6% για τον Obadiah (ελάχιστο ποσοστό) έως 70,4% για Prudence (μέγιστο ποσοστό ομιλίας). Ωστόσο, αυτό θα μπορούσε να εξηγηθεί εν μέρει λαμβάνοντας υπόψιν το ρόλο των προσωπικοτήτων των τεσσάρων χειριστών (που διαδραματίζονται από τους ανθρώπους-χειριστές).

Για παράδειγμα, η λογική και συνετή Prudence παρασύρει το χρήστη ώστε να μιλήσει περισσότερο, ενώ η καταθλιπτική διάθεση του Obadiah μπορεί να οδηγήσει το χρήστη να

μιλήσει λιγότερο. Αντιθέτως, όσον αφορά στο ρόλο της Poppy, στη Συνεδρία 26, ο ευτυχισμένος χειριστής ζητά από το χρήστη, “πού είναι το καλύτερο ζύπνημα που έχεις κάνει ποτέ;” και ο χρήστης απαντά “σε μια σκηνή στο kilimanjaro.” Εδώ, η ερώτηση του πράκτορα πυροδοτεί την έναρξη ενός νέου θέματος χωρίς να το προσδιορίζει εντελώς. Είναι η απάντηση του χρήστη που επιλέγει το νέο θέμα, έχοντας όμως προηγουμένως ακολουθήσει τις ενδείξεις που έχουν δοθεί από τις ερωτήσεις του πράκτορα.

“Γιατί” και “Πώς” αισθάνεται;- σχετικές με το πλαίσιο ερωτήσεις: εντοπισμός του στυλ ενσυναίσθησης και των συναισθημάτων (sentiments) των χειριστών και των χρηστών. Εκτός από το αλληλεπιδραστικό θέμα της συζήτησης, το λεξιλογικό και το συναισθηματικό προφίλ του χρήστη και του χειριστή εξαρτώνται επίσης από το είδος του σώματος κειμένου. Συνολικά, αναμένεται ότι το συγκεκριμένο λεξιλόγιο που χρησιμοποιείται αντιστοιχεί στην αλληλεπίδραση διάρκειας πέντε (5) λεπτών και σε ένα περιορισμένο λεξιλόγιο, εξατομικευμένο σχετικά με το ρόλο του και με το γλωσσικό-λεκτικό ρόλο αντίστοιχα που πρεσβεύει. Εξετάζοντας λοιπόν τις λέξεις που χρησιμοποιούνται περισσότερο από τον καταθλιπτικό Obadiah (“άθλια”, “απογοητευτικά”, “μίζερα”) και από την Poppy (“εξαιρετικά”, “συναρπαστικά”, “ευτυχία”) παρατηρείται ότι είναι δυνατή η εξαγωγή πληροφοριών αναφορικά με το στυλ ενσυναίσθησης του κάθε ρόλου του χειριστή.

Συνολικά, όσον αφορά στην ταυτότητα του χειριστή, οι συναισθηματικές εκφράσεις είναι πολυάριθμες, εκτός από τις συνεδρίες της Prudence. Αυτό πιθανότατα οφείλεται στην προσωπικότητα της Prudence που καλείται να παίξει ο χειριστής, αφού πρόκειται για μία λογική προσωπικότητα που εκφράζει εκτιμήσεις σχετικά με τη συμπεριφορά του χρήστη και ζητά από το χρήστη να εκφράσει τη στάση του σχετικά με συγκεκριμένα πράγματα. Επιπλέον, για το ρόλο του Spike, εντοπίσαμε ότι οι πιο συχνές λέξεις που χρησιμοποιούνται (Πίνακας 5.1) όπως οι λέξεις “άνοητος” και “ενοχλημένος”, ενσωματώνουν ένα περισσότερο προσβλητικό ύφος. Μία πιθανή εξήγηση αποτελεί το γεγονός ότι όταν ο χειριστής που έχει το ρόλο του Spike έχει προσβλητικό ύφος και τόνο απευθυνόμενος στο χρήστη, ο τελευταίος μερικές φορές επαναλαμβάνει τα λόγια του χειριστή.

Λέξεις	Ποσοστό	Χειριστής	Λέξεις	Ποσοστό	Χρήστης
miserable	0.02561	Obadiah	bloody	0.03224	2
excellent	0.02579	Prudence	beautiful	0.09622	9
fool	0.03250	Spike	shipped	0.02506	11
annoyed	0.03123	Spike	hang	0.02506	11
excellent	0.03417	Poppy	language	0.03233	12
aha	0.03379	Poppy	room	0.04196	16

Πίνακας 5.1: Λεξιλόγιο που χρησιμοποιείται πιο συχνά από τους χειριστές και τους χρήστες αντίστοιχα στο σώμα κειμένου SEMAINE.

Από την άλλη πλευρά, για τους **χρήστες** του σώματος κειμένου SEMAINE, οι συνη-

θέστερες λέξεις αντιστοιχούν σε διαφορετικά θέματα συζήτησης και στις απόψεις τους (opinions), ορίζονται δηλαδή είτε από επιρρήματα είτε/και περιλαμβάνουν λέξεις όπως “σαββατοκύριακο”, “διακοπές” που είναι ενδεικτικές του θέματος της συζήτησης. Κατά συνέπεια, τα συναισθήματα του χρήστη είναι ευθυγραμμισμένα με το ρόλο του πράκτορα που υποδύεται ο χειριστής. Όπως αναμενόταν, οι συνεδρίες της Porpy και της Prudence εκφράζουν μία συναισθηματική πληροφορία με αρνητικά συναισθήματα. Τέλος, αναφορικά με τις συνεδρίες του Spike, η ομιλία του είναι περισσότερο κατανεμημένη.

5.1.5 Προκλήσεις και Μελλοντικές Επεκτάσεις

Διαπιστώνουμε ότι σταδιακά, η νέα διεπιστημονική περιοχή που βρίσκεται στο σταυροδρόμι μεταξύ της Συναισθηματικής Υπολογιστικής, της αλληλεπίδραση ανθρώπου-υπολογιστή (HCI), των κοινωνικών επιστημών, της γλωσσολογίας, της Ψυχολογίας, και των μεθόδων ενσωμάτωσης του εννοιολογικού πλαισίου διακρίνεται ως ένα ξεχωριστό ερευνητικό πεδίο. Είναι επομένως δυνατή η καλύτερη αναγνώριση, ερμηνεία και επεξεργασία των απόψεων (opinions) και των συναισθημάτων (sentiments), μέσω της ενσωμάτωσης πληροφοριών εννοιολογικού πλαισίου και συμφραζομένων για την καλύτερη κατανόηση των δεοντολογικών ζητημάτων σχετικά με τη φύση του νου και τη δημιουργία συναισθηματικών μηχανών. Για εφαρμογές σε ερευνητικά πεδία όπως HCI σε πραγματικό χρόνο και ανάλυση μεγάλων κοινωνικών δεδομένων [3], η βαθύτερη κατανόηση της Φυσικής Γλώσσας δεν είναι αυστηρή προϋπόθεση: μία αίσθηση της σημασιολογίας που σχετίζεται με το κείμενο και κάποιες επιπλέον συναισθηματικές πληροφορίες, είναι συχνά ικανοποιητικές για την εκτέλεση γρήγορων εργασιών όπως π.χ. η αναγνώριση συναισθημάτων και η ανίχνευση γνωστικής και συναισθηματικής πληροφορίας.

Όσον αφορά στην **επιλογή του σώματος**, η νέα τάση είναι η συλλογή δεδομένων σε πραγματικό χρόνο μέσα από νέες πηγές ανάλυσης γνώσης και ανάλυσης αισθήσεων που αφθονούν. Οι κάμερες τοποθετημένες σε smartphones, touchpads ή άλλες συσκευές επιτρέπουν στους χρήστες να δημοσιεύουν γνωμοδοτήσεις σε μορφή ήχου ή οπτικοακουστικού περιεχομένου και όχι μόνο σε κείμενο. Εκτός από την ομιλούμενη γλώσσα σε γραπτό κείμενο για ανάλυση, η οπτικοακουστική μορφή παρέχει εξίσου μία ευκαιρία να εξουσιάσουμε τις απόψεις και το συναίσθημα.

Επιπλέον, όσον αφορά στην **παρουσία και στη θέση πολλαπλών λέξεων-εννοιών (multi-word concepts)** στο τμήμα κειμένου, απαιτείται περαιτέρω εξέταση, καθώς τα διγράμματα (bi-grams) και τριγράμματα (tri-grams), συχνά λαμβάνονται υπόψιν ως χρήσιμα χαρακτηριστικά. Ορισμένες μέθοδοι βασίζονται επίσης στην απόσταση μεταξύ των όρων. Πληροφορίες για τα μέρη του λόγου (POS) (ουσιαστικά, επίθετα, επιρρήματα, ρήματα, κλπ.) χρησιμοποιούνται επίσης ευρέως στην ανάλυση κειμένου ως βασικός τρόπος αποσαφήνισης λέξεων-εννοιών. Πιο συγκεκριμένα, ορισμένα επίθετα, έχει αποδειχθεί ότι είναι καλοί δείκτες του συναισθήματος και μερικές φορές χρησιμοποιούνται για την επιλογή χα-

ρακτηριστικών αναφορικά με την ταξινόμηση του συναισθήματος. Με άλλα λόγια, στην περίπτωση αυτή, η ανίχνευση των συναισθημάτων πραγματοποιείται μέσω επιλεγμένων φράσεων και ενός αριθμού προκαθορισμένων μοντέλων POS, τα περισσότερα από τα οποία αποτελούνται από ένα επίθετο ή ένα επίρρημα. Ωστόσο, τέτοιες προσεγγίσεις καθώς και οι επιδόσεις τους είναι άρρηκτα συνδεδεμένες με τον εξεταζόμενο τομέα εφαρμογής.

Επιλογή γλώσσας. Σύμφωνα με την τρέχουσα βιβλιογραφία, το μεγαλύτερο μέρος της αναφορικά με την ανάλυση συναισθημάτων επικεντρώνεται σε κείμενα της Αγγλικής γλώσσας και κατά συνέπεια, οι περισσότεροι γλωσσικοί πόροι έχουν αναπτυχθεί σε αυτή, π.χ., λεξικά συναισθημάτων είναι στα αγγλικά. Η προσαρμογή αυτών των πόρων και σε άλλες γλώσσες πρέπει να ληφθεί σοβαρά υπόψιν καθώς η επιλογή των λέξεων και η επιδιωκόμενη σημασία τους είναι εξαρτώμενη από τα συναισθηματικά, πολιτισμικά και κοινωνικά χαρακτηριστικά και διαφέρει ανάλογα με το επίπεδο της εμπειρογνομosύνης και τους σκοπούς της επισημείωσης των χρηστών.

Κεφάλαιο 6

Η συνεισφορά του πλαισίου στον τομέα της Φυσικής Γλώσσας

6.1 Εισαγωγή

Όπως αναφέρθηκε στην Ενότητα 3.1, κρίνεται απαραίτητη η καλύτερη αναγνώριση, ερμηνεία και επεξεργασία των απόψεων (opinions) και των συναισθημάτων (sentiments) συνδυαστικά με τις πληροφορίες εννοιολογικού πλαισίου για τη δημιουργία συναισθηματικών μηχανών. Συγκεκριμένα, αναφορικά με εφαρμογές σε ερευνητικά πεδία όπως η επικοινωνία ανθρώπου-μηχανής (HCI) σε πραγματικό χρόνο και η ανάλυση μεγάλων κοινωνικών δεδομένων [3], τα τελευταία μπορούν να επωφεληθούν από τη βαθύτερη κατανόηση της Φυσικής Γλώσσας και από την ενσωμάτωση του πλαισίου σε επίπεδο κειμένου.

Τα πλεονεκτήματα αυτής της αναδυόμενης ερευνητικής περιοχής (context-aware SA) σχετίζονται με την ανίχνευση σημαντικών γεγονότων στις ειδήσεις [100], με τη συλλογή απόψεων καταναλωτών για προϊόντα [113], με την εξαγωγή του προσανατολισμού (δηλαδή θετικής ή αρνητικής γνώμης) για το κείμενο [166], κλπ. Πρόσφατα, έχει αποδειχθεί ότι μπορεί να αποφέρει ανταγωνιστικά πλεονεκτήματα για επιχειρήσεις, καθώς εξάγει το συναίσθημα σε επίπεδο προτάσεων ή φράσεων (sentence- ή phrase-level) [165]. Μπορούμε επομένως να συμπεράνουμε ότι, η απόδοση των συστημάτων SA με γνώμονα το πλαίσιο στο οποίο λαμβάνουν χώρα εξαρτάται σε μεγάλο βαθμό από μία σειρά προκαθορισμένων παραμέτρων κατά τη διάρκεια της πειραματικής διαδικασίας [68, 165].

Ιδιαίτερο ενδιαφέρον περιστρέφεται γύρω από τον τομέα της SA όταν η επιλογή του σώματος κειμένου προέρχεται από μεγάλα κοινωνικά δεδομένα (π.χ., ταινίες, αξιολογήσεις προϊόντων) [166]. Προηγούμενες μελέτες έχουν προτείνει διάφορες προσεγγίσεις για την αύξηση της ευρωστίας τέτοιων συστημάτων που έχουν επίγνωση του πλαισίου. Η αυτοματοποιημένη ενσωμάτωση του συναισθήματος έχει χαρακτηριστεί ως πολλά υποσχόμενη προσέγγιση [165]. Από τη μια πλευρά, διαφορετικές τεχνικές Μηχανικής Μάθησης (Machine Learning (ML)), εκμεταλλεύονται μεθόδους διανυσματικών αναπαραστάσεων των κειμένων είτε μεθόδους που βασίζονται στη χρήση λεξικών [227] αποδίδοντας έτσι

σημασιολογικό προσανατολισμό (polarity) σε μεμονωμένες λέξεις.

Άλλες ερευνητικές προσπάθειες προτείνουν τη χρήση κανόνων (rule-based) και μη επιβλεπόμενων μεθόδων μάθησης για τη βελτίωση της ανάλυσης συναισθημάτων από πηγές κειμένου. Οι μέχρι τώρα ενδείξεις αποδεικνύουν ότι όλοι οι αλγόριθμοι τελευταίας τεχνολογίας λειτουργούν καλά σε μεμονωμένες προτάσεις χωρίς όμως να λαμβάνουν υπόψη οποιαδήποτε πληροφορία πλαισίου. Παρόλα αυτά, η ακρίβειά τους είναι δραματικά χαμηλότερη σε επίπεδο κειμένου, λόγω του ότι δε λαμβάνουν υπόψη τα συμφραζόμενα.

Ως εκ τούτου, πρόσφατες μελέτες σχετικά με τα context-aware SA συστήματα έχουν αρχίσει να εξετάζουν μεθόδους που παρέχουν περισσότερο έλεγχο. Για παράδειγμα, οι ερευνητές του [36] χρησιμοποιούν τη μεθοδολογία Conditional Random Field (CRF) για να εντοπίσουν το υποκείμενο της εξεταζόμενης πρότασης, δηλαδή το άτομο που εκφέρει την άποψή του, προσεγγίζοντάς το ως μία ακολουθιακή διεργασία επισημείωσης (sequential tagging task) και στη συνέχεια ενσωματώνουν την πληροφορία πλαισίου στο σύστημα. Μία αντίστοιχη έρευνα [103] προτείνει τον εντοπισμό του υποκειμένου της γνώσης με τη χρήση CRF για την περαιτέρω ενσωμάτωση των πληροφοριών πλαισίου στα συστήματα SA. Εκτός από αυτή την κατεύθυνση της έρευνας, έχει χρησιμοποιηθεί επίσης και η ανάλυση του λόγου (discourse analysis) για την προσαρμογή της προηγούμενης πολικότητας των όρων. Περαιτέρω βελτιώσεις αναφέρθηκαν όταν προτάθηκε μία περισσότερο εξελιγμένη προσέγγιση ανάλυσης του λόγου (discourse analysis) που βασίζεται στη θεωρία Rhetorical Structure Theory (RST) [138].

Επανεξετάζοντας λοιπόν την αποτελεσματικότητα αυτών των μεθόδων για την ενσωμάτωση πληροφοριών πλαισίου στην SA, η προτεινόμενη προσέγγισή μας συνδυάζει τις μεθόδους CRF και RST με την SA για την αποτελεσματικότερη ανάλυση του πλαισίου. Συγκεκριμένα, συγκριτικά με τους υπάρχοντες αλγόριθμους ανάλυσης γνώσης και συναισθήματος τόσο σε επίπεδο προτάσεων όσο και σε επίπεδο κειμένων, η προτεινόμενη προσέγγιση προσπαθεί να βελτιώσει την SA αξιοποιώντας πλήρως τη δομή των προτάσεων. Συγκεκριμένα χρησιμοποιεί α) τις πληροφορίες πλαισίου για τον εντοπισμό της σχέσης μεταξύ των προτάσεων και για τη βελτίωση της SA του κειμένου, β) λαμβάνει υπόψη το σύνολο των λέξεων της γλώσσας του Διαδικτύου και των συναισθημάτων και τέλος γ) επεκτείνει το σύνολο των χαρακτηριστικών συμφραζομένων (context) που κάνουν χρήση σημασιολογικών, συντακτικών, δομικών και RST χαρακτηριστικών του πλαισίου.

Μέχρι στιγμής, ένα ευρύ φάσμα χαρακτηριστικών έχει ελεγχθεί μεμονωμένα από ένα μεγάλο αριθμό ερευνητικών ομάδων, κυρίως όμως με ελεγχόμενες συνθήκες και καθορισμένες παραδοχές. Ο Πίνακας 6.1 συνοψίζει τα κύρια χαρακτηριστικά σημαντικών μελετών που πραγματοποιήθηκαν σε αυτό τον τομέα. Διαπιστώνουμε λοιπόν, ότι δεν υπάρχει σαφής εικόνα της συνεισφοράς του κάθε συνόλου χαρακτηριστικών ενώ συγχρόνως υπάρχουν ελάχιστες ενδείξεις αναφορικά με τον τρόπο συμπεριφοράς και συνεισφοράς της ίδιας ομάδας χαρακτηριστικών σε πηγές πληροφοριών διαφορετικού μεγέθους και είδους. Έτσι, δημιουργείται η ανάγκη συστηματικών μελετών που να συγκρίνουν τα σημαντικότερα χαρακτηριστικά σε ομοιογενείς συνθήκες.

Εργασία	Είδος Δεδομένων	Επίπεδο Ανάλυσης	Μέγεθος	Κατηγοριοποίηση	Χαρακτηριστικά
Pang et al.[166]	Movie Reviews	Docs.	1400	pos/neg	voc, pos, p
Turney[235]	Reviews	Docs.	410	pos/neg	sp
Pang and Lee[163]	Movie Reviews	Docs.	2000	pos/neg	voc, p
Beineke et al. [12]	Movie Reviews	Sents.	2500	Summarization	voc, l
Wiebe and Riloff [254]	Press articles	Sents.	9289	subj/obj	pos, sp
Taboada et al.[227]	Reviews	Docs.	400	pos/neg	pos, sw, d
Heerschoop et al.[96]	Movie Reviews	Docs.	1000	pos/neg	sw, d
Katz et al.[109]	Hotel Reviews	Docs.	30.000	pos/neg	voc, pos, sp, p
Katz et al.[109]	Movie Reviews	Docs.	2000	pos/neg	voc, pos, sp, p

Πίνακας 6.1: Τα κύρια χαρακτηριστικά μίας σειράς σωμάτων κειμένων. Ο Πίνακας συγκεντρώνει χαρακτηριστικά όπως το είδος των δεδομένων, το επίπεδο ανάλυσης, το μέγεθος του κειμένου, τον τύπο κατηγοριοποίησης και τα εξεταζόμενα χαρακτηριστικά. Η επισημείωση των χαρακτηριστικών είναι ως εξής: λεξιλόγιο (*vocabulary (voc)*): μονογράμματα (*unigrams*) και διγράμματα (*bigrams*), μέρη του λόγου (*Part-Of-Speech (pos)*), λέξεις με συναισθηματικό περιεχόμενο (*sentiment words (sw)*), μοτίβα συντακτικά (*syntactic patterns (sp)*), θέση των λέξεων (*position (p)*) και ανάλυση λέξεων (*discourse (d)*).[248]

6.2 Σχετική Βιβλιογραφία

Μία σημαντική ερευνητική κατεύθυνση στον context-aware SA τομέα είναι η βελτίωση της ευρωστίας αυτών των συστημάτων SA μετά την ενσωμάτωση του πλαισίου [165, 5]. Οι υπάρχουσες μέθοδοι βασίζονται σε λεξικά συναισθημάτων, τα οποία είναι ενιαίοι κατάλογοι συναισθηματικών όρων που υποδηλώνουν τις αλλαγές αισθήματος. Δημοφιλή παραδείγματα λεξικών αποτελούν το General Inquirer [223], το Subjectivity Lexicon και Subjectivity Annotations και το SentiWordNet.

Επιπλέον, η γνώση του κάθε τομέα (*domain*) διαδραματίζει σημαντικό ρόλο, καθώς το γλωσσικό πλαίσιο ενός όρου επηρεάζει συχνά και τη συσχέτισή του με τον εντοπισμό του συναισθήματος. Οι πρώιμες εργασίες για τη SA χρησιμοποίησαν συντακτικές σχέσεις για να προσδιορίσουν νέους όρους συναισθημάτων, οι οποίοι μπορούν να θεωρηθούν ως μία πρώιμη μορφή εκμετάλλευσης της έννοιας του πλαισίου [95]. Το συναίσθημα εκφράζεται συχνά με τρόπο που δυσχεραίνει τον εντοπισμό του κατά την επεξεργασία προτάσεων ή παραγράφων μεμονωμένα. Συνεπώς, το πλαίσιο παραμένει βασικό συστατικό για την περαιτέρω βελτίωση των συστημάτων SA.

Συνολικά, η έρευνα στην SA περιοχή οργανώνεται σε δύο κατηγορίες: σε μεθόδους Μηχανικής Μάθησης (*Machine Learning*) και σε μεθόδους που βασίζονται στη χρήση λεξικών (*Lexicon-Based*). Σύμφωνα με την πρώτη κατηγορία, ο αλγόριθμος εκπαιδεύεται με δεδομένα τα οποία συνοδεύονται με ετικέτες συναισθηματικού περιεχομένου και στη συνέχεια το μοντέλο αυτο χρησιμοποιείται για να κατηγοριοποιήσει νέα έγγραφα. Αυτή η μέθοδος όμως είναι συνήθως χρονοβόρα λόγω του μεγάλου όγκου δεδομένων.

Όσον αφορά στη μέθοδο που βασίζεται στη χρήση λεξικών, αυτή αφορά στην εξαγωγή και στην ομαδοποίηση των βαθμολογιών των όρων που σχετίζονται με τα συναισθη-

ματα (δηλ. τις προηγούμενες πολικότητες) για την πρόβλεψη του συναισθήματος. Παρόλα αυτά, δεδομένου ότι τα λεξικά αυτά είναι προκαθορισμένα, δεν μπορούν να προσαρμοστούν σε καινούργιες ή σε συγκεκριμένες μορφές έκφρασης. Επιπλέον, οι προσεγγίσεις που βασίζονται στη χρήση λεξικών δεν παράγουν τόσο υψηλά επίπεδα εμπιστοσύνης κατά την ανάλυση, κάτι το οποίο παρέχεται αυτόματα μέσω των προσεγγίσεων της Μηχανικής Μάθησης.

6.3 Μέθοδοι Ενσωμάτωσης

Η κατανόηση του συναισθήματος των προτάσεων, μας επιτρέπει να συνοψίσουμε τις απόψεις που θα μπορούσαν να βοηθήσουν τους ανθρώπους να λαμβάνουν τεκμηριωμένες αποφάσεις. Όλοι οι υπερσύγχρονοι αλγόριθμοι λειτουργούν ικανοποιητικά σε μεμονωμένες προτάσεις χωρίς να λαμβάνουν υπόψιν οποιαδήποτε πληροφορία πλαισίου. Κατά συνέπεια, η ακρίβειά τους είναι δραματικά χαμηλότερη σε επίπεδο κειμένου. Υπάρχουν πολλές δυσκολίες εξαιτίας κυρίως των ιδιαίτερων χαρακτηριστικών και της διαφορετικότητας στη διάρθρωση της φράσης, στον τρόπο με τον οποίο οι άνθρωποι εκφράζουν τις απόψεις τους (π.χ. απόψεις που εκφράζονται εμμέσως διαμέσου της σύγκρισης κ.λπ.). Επιπλέον, η περίπλοκη δομή της φράσης και τα συναισθήματα κάνουν την ανάλυση της τελευταίας ακόμη πιο δύσκολη.

Για το λόγο αυτό, δεν εξετάζουμε μόνο τη σύνταξη που μπορεί να επηρεάσει το συναίσθημα, αλλά και τα emoticons, τις θετικές και αρνητικές λέξεις, τις λέξεις άρνησης (negation words), την ενσωμάτωση πληροφοριών σχετικά με τη δομή της φράσης, τις λέξεις που εμπεριέχουν έννοιες σύγκρισης, τη θέση των θετικών και των αρνητικών λέξεων μέσα στην πρόταση και τα χαρακτηριστικά πλαισίου-RST. Ως εκ τούτου, χρησιμοποιούμε ένα μοντέλο CRF [121] για να εντοπίσουμε και να συλλέξουμε τα συντακτικά και σημασιολογικά χαρακτηριστικά των προτάσεων και στη συνέχεια χρησιμοποιούμε το RST μοντέλο για να ξεδιπλώσουμε τις ρητορικές σχέσεις για την ανάλυση συναισθήματος.

Conditional Random Field (CRF):

Το CRF είναι γνωστό για τις διαδικασίες επισημείωσης ακολουθιών [121]. Τα CRF αποτελούν μία κατηγορία διακριτού μη προσανατολισμένου πιθανοτικού γραφικού μοντέλου που εφαρμόζεται γενικά στην Αναγνώριση Προτύπων και στη Μηχανική Μάθηση και έχουν σχεδιαστεί ειδικά για τη βελτιστοποίηση της πρόβλεψης της δομής. Ένας “γενικός” ταξινομητής προβλέπει μία ετικέτα για ένα μόνο δείγμα χωρίς να αναφέρεται σε δείγματα “γειτονικά/συνδεδεμένα”, ωστόσο μία γραμμική αλυσίδα CRF μπορεί να λάβει υπόψιν το πλαίσιο.

Συνεπώς, στην περίπτωση που επιθυμούμε να καταγράψουμε τις πληροφορίες πλαισίου (π.χ. γειτονικές προτάσεις ή προτάσεις που συνδέονται με λέξεις μετάβασης), η διαδικασία αναγνώρισης του συναισθήματος αντιμετωπίζεται ως ένα είδος επισημείωσης αλληλουχιών. Συγκεκριμένα, τα CRF παρέχουν ένα πιθανοτικό πλαίσιο για τον υπολογισμό της πιθανότητας Y που είναι εξαρτημένη από το X , όπου το X είναι μία τυχαία μετα-

βλητή/διάνυσμα μίας αλληλουχίας δεδομένων προς επισημείωση και Y είναι μια τυχαία μεταβλητή/διάνυσμα ακολουθιών ετικετών. Αξίζει να αναφέρουμε στο σημείο αυτό ότι τα X και Y θα μπορούσαν να ακολουθούν και μία πολυπλοκότερη δομή γραφημάτων.

Μέχρι στιγμής, το μοντέλο CRF, στην απλούστερη μορφή του, έχει χρησιμοποιηθεί ευρέως στον τομέα επισημείωσης κειμένου [141]. Παραδείγματα CRF αποτελούν οι ερευνητικές προσπάθειες [121] και [141], στις οποίες γίνεται χρήση της γραμμικής δομής της αλυσίδας. Μία άλλη παρατήρηση αποκαλύπτει ότι υπάρχει αντιστοιχία ένα-προς-ένα μεταξύ των καταστάσεων του μοντέλου και των ετικετών. Στο Σχήμα 6.1 παρουσιάζεται μια απλή απεικόνιση του τρόπου εμφάνισης ενός μοντέλου CRF.

Έστω ότι δίνεται μία αλληλουχία (π.χ. ένα κείμενο αποτελούμενο από πολλαπλές προτάσεις) $X := (x_1, x_2, \dots, x_m)$ και η αντίστοιχη ακολουθία ετικετών (κάθε πρόταση επισημειώνεται ως μία ετικέτα) $Y := (y_1, y_2, \dots, y_m)$, η πιθανότητα του Y όταν συμβαίνει το X όπως ορίζεται στο CRF μοντέλο, τότε το $Pr(Y|X)$, εκφράζεται ως εξής:

$$Pr(Y|X) = \frac{1}{Z_X} \exp\left(\sum_{j=1}^{K \times L} F_j(Y, X)\right) = \frac{1}{Z_X} \exp\left(\sum_{i=1, k=1}^{m, K} \lambda_k f_k(y_{i-1}, y_i, X) + \sum_{i=1, l=1}^{m, L} \mu_l g_l(y_i, X)\right) \quad (6.1)$$

όπου Z_X είναι η σταθερά κανονικοποίησης (normalization constant) η οποία αθροίζει την πιθανότητα όλων των καταστάσεων της αλληλουχίας ίση με τη μονάδα. Η εξίσωση 6.1 περιλαμβάνει δύο τύπους συναρτήσεων ενδεικτικών χαρακτηριστικών:

1. $f_k(y_{i-1}, X)$ είναι μία αυθαίρετη συνάρτηση χαρακτηριστικών αναφορικά με τη συνολική αλληλουχία X και τη θέση των καταστάσεων i και $i - 1$;
2. $g_l(y_i, X)$ είναι η συνάρτηση χαρακτηριστικών της κατάστασης στη θέση i και στην αλληλουχία X .

Αντίστοιχα, τα λ_k και μ_l είναι τα θετικά βάρη μετά τη διαδικασία της εκπαίδευσης (από τα δεδομένα εκπαίδευσης) για τις συναρτήσεις χαρακτηριστικών f_k και g_l , αντικατοπτρίζοντας τη σταθερότητα (confidence) του μοντέλου των αντίστοιχων συναρτήσεων χαρακτηριστικών f_k και g_l . Οι τελευταίες μπορούν να περιγράψουν οποιοδήποτε χαρακτηριστικό της μετάβασης από την κατάσταση y_{i-1} στην y_i , καθώς επίσης την y_i και τα γενικά (global) χαρακτηριστικά της κατάστασης X . Για να δώσουμε ένα συγκεκριμένο παράδειγμα, η συνάρτηση χαρακτηριστικών f_k μπορεί να θεωρηθεί ότι αξιολογείται ως εξής:

1. 1 όταν το y_{i-1} έχει ετικέτα P ;
2. 1 όταν το y_i έχει ετικέτα N ;
3. 1 όταν το x_{i-1} περιέχει θετικά emoticons;
4. 1 όταν το x_i περιέχει συνδετικές λέξεις (conjunction words) στην αρχή της πρότασης;

5. 0 οπουδήποτε αλλού.

Από την άλλη πλευρά, η συνάρτηση χαρακτηριστικών g_l μπορεί να θεωρηθεί ότι αξιολογείται ως εξής:

1. 1 όταν το y έχει ετικέτα P ;
2. 1 όταν το x_i περιέχει επίθετα θετικού περιεχομένου και όχι λέξεις που δηλώνουν άρνηση (negation words);
3. 0 οπουδήποτε αλλού.

6.3.1 Υπολογισμός Παραμέτρων

Ο στόχος της εκτίμησης των παραμέτρων είναι η εκμάθηση του συνόλου βαρών/παραμέτρων στο μοντέλο CRF. Υποθέτουμε ότι το $\Theta := \{\lambda_k, \mu_l | 1 \leq k \leq K; 1 \leq l \leq L\}$ είναι η παράμετρος του CRF μοντέλου μας. Είναι γνωστό ότι η παράμετρος Θ υπολογίζεται από την αρχή της μέγιστης πιθανότητας (maximum likelihood), η οποία υπολογίζεται από την μεγιστοποίηση της πιθανής καταγραφής της πιθανότητας της συνάρτησης των επισημειωμένων ακολουθιών στα δεδομένα εκπαίδευσης $D := (\mathbf{X}, \mathbf{Y}) = \{(X^{(1)}, Y^{(1)}), \dots, (X^{(M)}, Y^{(M)})\}$, όπου η παράμετρος M είναι ο αριθμός των δειγμάτων εκπαίδευσης. Έτσι, η log-likelihood συνάρτηση ορίζεται ως εξής:

$$\begin{aligned} \mathcal{L}(\Theta) &= \log Pr(\mathbf{Y}|\mathbf{X}; \Theta) \\ &= \log \prod_{j \in \{1, \dots, M\}} Pr(Y^{(j)}|X^{(j)}; \Theta) \\ &= \sum_{j \in \{1, \dots, M\}} \log Pr(Y^{(j)}|X^{(j)}; \Theta). \end{aligned} \quad (6.2)$$

Για την αποφυγή του over-fitting κατά τη διαδικασία της εκπαίδευσης των δεδομένων, μέθοδοι σταθεροποίησης (regularization methods) [170] συχνά προστίθενται στο $\mathcal{L}(\Theta)$ μέσω της συνάρτησης 6.2. Ένας πολύ συνηθισμένος τρόπος να προσθέσουμε ένα Gaussian prior στις παραμέτρους, ορίζεται ως εξής:

$$\mathcal{L}'(\Theta) = \sum_{j \in \{1, \dots, M\}} \log Pr(Y^{(j)}|X^{(j)}; \Theta) - \sum_k \frac{\lambda_k^2}{2\sigma_k^2} - \sum_l \frac{\mu_l^2}{2\sigma_l^2}, \quad (6.3)$$

όπου το σ_k^2 και το σ_l^2 είναι τα variances των Gaussians priors (τα οποία ισοδυναμούν με 1).

Το τελευταίο βήμα αναφορικά με την εκτίμηση της μέγιστης πιθανότητας (maximum likelihood) είναι η διαφοροποίηση της regularized log-likelihood συνάρτησης ως προς κάθε παράμετρο λ_k είτε μ_l .

$$\frac{\partial \mathcal{L}'(\times)}{\partial \lambda_k} = E_{\tilde{p}(Y,X)}[F_k(Y, X)] - \sum_j E_{Y|X^{(j)};\lambda}[F_k(Y, X^{(j)})], \quad (6.4)$$

όπου η παράμετρος $\tilde{p}(Y, X)$ είναι η εμπειρική κατανομή των δεδομένων εκπαίδευσης. Να επισημάνουμε σε αυτό το σημείο ότι θέτοντας αυτή την παράγωγο ίση με μηδέν αποδίδεται το μέγιστο μοντέλο εντροπίας περιορισμού για την παράμετρο λ_k . Η προσδοκία του κάθε χαρακτηριστικού αναφορικά με την κατανομή του μοντέλου ισούται με την αναμενόμενη τιμή της εμπειρικής κατανομής των δεδομένων εκπαίδευσης. Πολλοί αλγόριθμοι επαναλαμβανόμενης κλίμακας (Iterative Scaling algorithms) (π.χ. ο Generalized Iterative Scaling (GIS) και ο Improved Iterative Scaling (IIS) [121]) μπορούν να χρησιμοποιηθούν για τη βελτιστοποίηση του $\mathcal{L}'(\times)$. Επιπρόσθετα, μερικές stochastic gradient μέθοδοι μπορούν επίσης να χρησιμοποιηθούν για τη βελτιστοποίηση παραμέτρων [212]. Στο κεφάλαιο αυτό, αναφερόμαστε σε μία quasi-Newton μέθοδο που ονομάζεται περιορισμένης μνήμης Broyden-Fletcher-Goldfarb-Shanno αλγόριθμος (L-BFGS) [135] ο οποίος συγκλίνει αρκετά γρηγορότερα από τον αρχικό BFGS αλγόριθμο.

6.3.2 Συμπέρασμα

Έχοντας ορίσει την conditional probability των καταστάσεων της αλληλουχίας από το CRF μοντέλο στη συνάρτηση 6.1 και τις εκτιμώμενες παραμέτρους $\hat{\Theta}$, η ετικέτα πρόβλεψης της ακολουθίας ορίζεται ως εξής:

$$Y^* = \operatorname{argmax}_Y Pr(Y|X; \hat{\Theta}). \quad (6.5)$$

Η εξίσωση 6.5 μπορεί να επιλυθεί ικανοποιητικά με τη χρήση του αλγορίθμου Viterbi [180]. Η οριακή πιθανότητα των καταστάσεων σε κάθε θέση της αλληλουχίας μπορεί να υπολογιστεί από μία διαδικασία δυναμικού προγραμματισμού παρόμοια με τη διαδικασία forward-backward του μοντέλου Hidden Markov Model (HMM) [180].

Υπολογίζουμε τις forward μεταβλητές $\alpha_i(y|X)$ σύμφωνα με τα δύο ακόλουθα βήματα:

1. Θέτουμε το $\alpha_1(y|X)$ ίσο με την πιθανότητα έναρξης της κατάστασης y .
2. Για $i > 1$, χρησιμοποιούμε την συνάρτηση 6.6.

$$\begin{aligned} \alpha_{i+1}(y|X) &= \sum_{y'} \alpha_i(y'|X) \exp(F_{i+1}(y', y, X)) \\ &= \sum_{y'} \alpha_i(y'|X) \sum_k \lambda_k f_k(y_i = y', y_{i+1} = y, X) \\ &\quad + \sum_{y'} \alpha_i(y'|X) \sum_l \mu_l g_l(y_{i+1} = y, X) \end{aligned} \quad (6.6)$$

Έτσι, οι παράγοντες σταθεροποίησης της συνάρτησης 6.1 ορίζονται ως:

$$Z_X = \sum_y \sum_j \alpha_j(y|X).$$

Με τη βοήθεια του Z_X , υπολογίζουμε την οριακή πιθανότητα της κάθε πρότασης, η οποία χαρακτηρίζεται ως **θετική** πρόταση όταν δίνεται η συνολική ακολουθία της πρότασης.

$$Pr(y_i = \text{“}P\text{”}|X) = \frac{a_i(\text{“}P\text{”}|X) \cdot \beta_i(\text{“}P\text{”}|X)}{Z_X}, \quad (6.7)$$

όπου $\beta_i(y|X)$ είναι οι backward τιμές και ορίζονται με παρόμοιο τρόπο όπως οι τιμές $a_i(y|X)$ στη forward αλληλουχία. Η $a_i(y|X)$ κωδικοποιεί την πιθανότητα ύπαρξης στην τρέχουσα κατάσταση y όταν δίνονται όλες οι παρατηρήσεις από την παρατήρηση 1 στην παρατήρηση $i - 1$. Αντίστοιχα, η $\beta_i(y|X)$ κωδικοποιεί την πιθανότητα της ύπαρξης της στην τρέχουσα κατάσταση y όταν είναι γνωστές οι observations από την $i + 1$ κατάσταση έως το τέλος της αλληλουχίας/πρότασης. Τέλος, η $Pr(y_i|X)$ υπολογίζεται από τη συνάρτηση 6.7 και είναι η εξομαλυμένη πιθανότητα των $a_i(y|X)$ και $\beta_i(y|X)$.

6.4 Εξαγώμενα χαρακτηριστικά: Σημασιολογική και Συντακτική Δομή των προτάσεων

6.4.1 Σημασιολογικά χαρακτηριστικά

Αριθμός λέξεων θετικού/αρνητικού περιεχομένου. Διαισθητικά, μία πρόταση η οποία περιλαμβάνει περισσότερες λέξεις θετικού/αρνητικού περιεχομένου είναι περισσότερο πιθανό να χαρακτηριστεί ως θετική ή αρνητική αντίστοιχα. Χρησιμοποιούμε δύο λίστες αγγλικών λέξεων συναισθηματικού περιεχομένου οι οποίες περιέχουν 1948 και 4550 λέξεις θετικού και αρνητικού περιεχομένου αντίστοιχα [255].

Εμπεριέχει οποιοδήποτε θετικό/αρνητικό emoticon. Με τη ραγδαία ανάπτυξη του Διαδικτύου και του Παγκόσμιου Ιστού, ένα μεγάλο πλήθος των λέξεων του Διαδικτύου έχουν γίνει όλο και περισσότερο συνηθισμένες. Για παράδειγμα, τα emoticons: “=)”, “:D” και “v_v” αντιπροσωπεύουν έννοιες όπως το χαμόγελο, το γέλιο και τη λύπη αντίστοιχα. Τα emoticons που χρησιμοποιούμε είναι 52 και 35 θετικά και αρνητικά αντίστοιχα.

Προτάσεις που δηλώνουν σύγκριση. Οι άνθρωποι συχνά εκφράζουν την άποψή τους μέσω της σύγκρισης παρόμοιων προϊόντων είτε αντικειμένων. Για παράδειγμα, όταν ένας χρήστης περιγράφει την εμπειρία του με το iPhone, η διαδικασία αυτή πραγματοποιείται μέσω της σύγκρισης για παράδειγμα με Samsung Android τηλέφωνα, τα οποία ενδεχομένως να ανήκουν σε κάποιο φίλο του. Μία πρόταση που δηλώνει σύγκριση μπορεί να μην χαρακτηρίζεται ως αντικειμενική ακόμα και στην περίπτωση που δεν περιλαμβάνει καμία λέξη συναισθηματικού περιεχομένου είτε emoticon. Για τον εντοπισμό προτάσεων συγκριτικού περιεχομένου, χρησιμοποιούμε μέρη του λόγου (Part-of-Speech (POS)) για να επισημειώσουμε κάθε λέξη της πρότασης. Έτσι, μία πρόταση δηλώνει σύγκριση εάν περιλαμβάνει:

- επίθετα που δηλώνουν σύγκριση,
- επιρρήματα που δηλώνουν σύγκριση,
- επίθετα υπερθετικού βαθμού,
- επιρρήματα υπερθετικού βαθμού,
- ενδεικτικές λέξεις κλειδιά (126 λέξεις συνολικά, που περιλαμβάνουν εκφράσεις όπως “συγκριτικά με/σε σύγκριση”, “σε αντίθεση με”, κλπ.)
- είτε κάποια προκαθορισμένα επαναλαμβανόμενα μοτίβα δομής (“as <επίθετο/επίρρημα>”, “το ίδιο με (the same as)”, “παρόμοια με (similar to)”, κλπ.)

Συνδετικές λέξεις. Οι συνδετικές λέξεις χρησιμοποιούνται συχνά για τη συνένωση διαφορετικών τμημάτων μίας πρότασης και μπορούν να έχουν μία σημαντική επιρροή στο συναίσθημα που τη χαρακτηρίζει. Για παράδειγμα, προτάσεις που περιλαμβάνουν λέξεις όπως: “but”, είτε “however” είναι πιθανό να φέρουν διαφορετικό συναισθηματικό περιεχόμενο από την προηγούμενη πρόταση. Παρόλα αυτά, οι προτάσεις συνδέονται συχνά με λέξεις όπως: “and”, “so that”, “before”, είτε “after” που έχουν συχνά παρόμοιο συναισθηματικό περιεχόμενο. Στην παρούσα εργασία διαχωρίζουμε τις συνδετικές λέξεις σε τρεις κατηγορίες: υποδεέστερες (subordinating), συντονισμού (coordinating) και συσχετισμού (correlative).

6.4.2 Συντακτικά χαρακτηριστικά

Η θέση της πρότασης. Διαισθητικά, οι προτάσεις στην αρχή είτε στο τέλος ενός κειμένου είναι συνήθως προτάσεις που περιλαμβάνουν συμπεράσματα. Συνήθως επηρεάζουν το συναισθηματικό περιεχόμενο των άλλων προτάσεων που βρίσκονται σε άλλες θέσεις μέσα στο εξεταζόμενο κείμενο. Λαμβάνουμε υπόψιν μας τρεις διαφορετικές τιμές αναφορικά με τη θέση της πρότασης: στην αρχή, στη μέση και στο τέλος. Εάν η πρόταση βρίσκεται στο αρχικό 20% των προτάσεων ενός κειμένου, χαρακτηρίζουμε την τελευταία ως αρχική πρόταση, τελική εάν βρίσκεται στο τελευταίο 20% των προτάσεων του κειμένου αντίστοιχα και ενδιάμεση σε όλες τις υπόλοιπες περιπτώσεις.

Απλές προτάσεις. Ως απλή πρόταση χαρακτηρίζεται συχνά εκείνη η οποία περιλαμβάνει περισσότερες από μία γνώμες μέσω συνδετικών λέξεων που συνδέουν διαφορετικά μέρη της πρότασης.

Η θέση λέξεων θετικού/αρνητικού περιεχομένου. Εάν δεν υπάρχουν θετικές ή αρνητικές λέξεις, η τιμή του χαρακτηριστικού ισούται με 0; αντίστοιχα η τιμή του χαρακτηριστικού είναι 1 όταν βρίσκονται αποκλειστικά στο αρχικό τμήμα μιας απλής πρότασης; η τιμή του χαρακτηριστικού ισούται με 2 όταν οι λέξεις αυτές βρίσκονται στο δεύτερο τμήμα της πρότασης; και τέλος η τιμή ισούται με -1 όταν βρίσκονται στο αρχικό καθώς και σε άλλα μέρη της πρότασης. Παραθέτουμε το ακόλουθο παράδειγμα (από την αξιολόγηση ενός πελάτη) το οποίο αντικατοπτρίζει τη σημασία αυτού του χαρακτηριστικού:

Εξαγώμενα χαρακτηριστικά: Σημασιολογική και Συντακτική Δομή των προτάσεων

- “Μου αρέσει το χρώμα, αλλά η ταχύτητα είναι κακή”,-αρνητική;
- “Μου αρέσει το χρώμα, αλλά η ταχύτητα είναι εξίσου καλή”,-περισσότερο θετική;
- “Η ταχύτητα είναι κακή, αλλά μου αρέσει το χρώμα”,-θετική.

Η θέση των λέξεων που δηλώνουν άρνηση. Οι λέξεις που δηλώνουν άρνηση θα μπορούσαν να αντιστρέψουν το συναισθηματικό περιεχόμενο της πρότασης. Για τις ανάγκες του πειράματος που παραθέτουμε σε επόμενη ενότητα συγκεντρώσαμε 32 λέξεις που δηλώνουν άρνηση. Η θέση αυτών των λέξεων είναι ιδιαίτερης σημασίας για το συναισθηματικό περιεχόμενο της πρότασης. Στην περίπτωση που η άρνηση βρίσκεται πολύ κοντά σε μία λέξη συναισθηματικού περιεχομένου, είναι πολύ πιθανό να αντιστρέψει το “συναισθηματικό φορτίο” αυτής της λέξης. Για παράδειγμα:

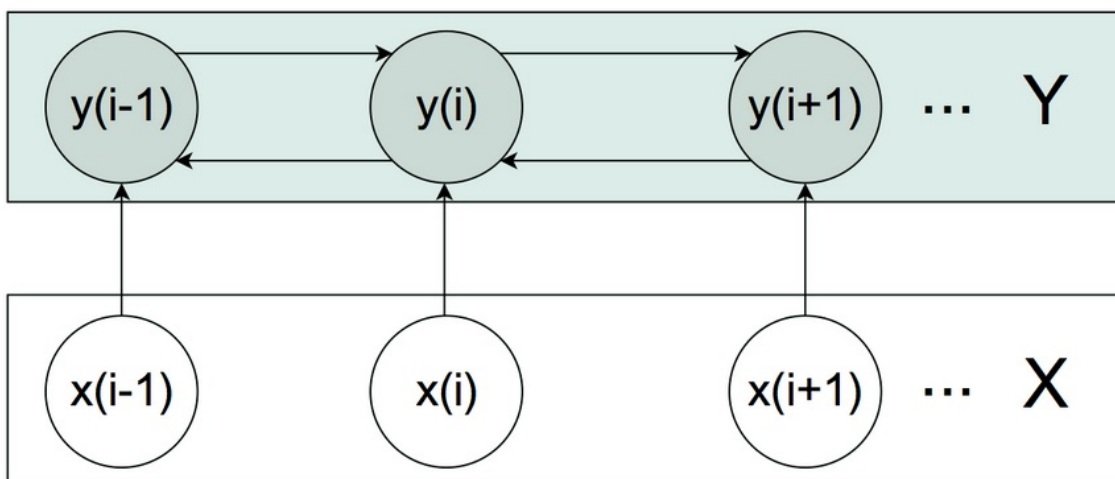
- “No other camera is better than this one!”,-θετική (αρνητική αναφορικά με το “other camera”);
- “This camera is not good!”,-αρνητική.

Υποκείμενο που δηλώνει σύγκριση. Το υποκείμενο σε μία σύγκριση μπορεί να αλλάξει το συναισθηματικό περιεχόμενο μιας πρότασης. Για παράδειγμα, ένας χρήστης ανέρτησε το σχόλιο “I like Pepsi more” στην επίσημη Facebook σελίδα της Coca-cola. Η πρόταση θα μπορούσε να είναι θετικού περιεχομένου εάν το υποκείμενο ήταν η Pepsi, όμως στο συγκεκριμένο πλαίσιο, δεν χαρακτηρίζεται ως θετική πρόταση. Στην παρούσα προσέγγιση, χρησιμοποιούμε τον επισημειωτή Stanford [127] ώστε να πάρουμε τις συσχετίσεις που προκύπτουν. Με τον τρόπο αυτό μπορούμε να γνωρίζουμε ποιο υποκείμενο (συνήθως ουσιαστικό) πραγματοποιεί μία σειρά ενεργειών (συνήθως ρήμα) είτε διαθέτει κάποια χαρακτηριστικά (συνήθως επίθετο). Σύμφωνα με το εγχειρίδιο του Stanford αναφορικά με τις τυπωμένες συσχετίσεις¹, η ετικέτα “nsubj” υποδηλώνει αυτή την εξάρτηση. Για το προαναφερθέν παράδειγμα, το υποκείμενο είναι το “I” αντί για τη λέξη “Pepsi” σύμφωνα με τις ακόλουθες εξαρτήσεις:

- nsubj(“like”-2, “I”-1)
- root(“ROOT”-0, “like”-2)
- dobj(“like”-2, “Pepsi”-3)
- advmod (“like”-2, “more”-4).

Ομοιότητα με τις γειτονικές προτάσεις. Ορίζουμε τα χαρακτηριστικά ώστε να αντιπροσωπεύουν την ομοιότητα μεταξύ μίας πρότασης και των γειτονικών της προτάσεων. Στην παρούσα εργασία, λαμβάνουμε υπόψιν μας μόνο τις προτάσεις που προηγούνται (sim_to_next)

¹http://nlp.stanford.edu/software/dependencies_manual.pdf.



Σχήμα 6.1: Γραφική αναπαράσταση μιας γραμμικής αλυσίδας ενός απλού μοντέλου CRF [121].

και που ακολουθούν αμέσως (sim_to_next) αντίστοιχα και υπολογίζουμε τα αποτελέσματα ομοιότητας, χρησιμοποιώντας τη μέθοδο cosine similarity και latent semantic indexing (LSI) [51]. Η τελευταία αποτελεί μία μέθοδο μείωσης των διαστάσεων που βασίζεται στην αποσύνθεση της μοναδικής τιμής (Singular Value Decomposition (SVD)). Η συγγενής ομοιότητα (cosine similarity) “αιχμαλωτίζει” την ομοιότητα σε επίπεδο λέξης ενώ το LSI χρησιμοποιείται για τον υπολογισμό της σημασιολογικής ομοιότητας.

Rhetorical Structure Theory (RST):

Σε ένα κείμενο Φυσικής Γλώσσας, οι σχέσεις μεταξύ των τμημάτων του κειμένου αρχικά “ξεδιπλώνονται” και εν συνεχεία χρησιμοποιούνται για τη διάκριση σημαντικών τμημάτων του κειμένου από λιγότερο σημαντικά τμήματα αναφορικά με τη συμβολή τους στο γενικότερο “συναισθηματικό αποτύπωμα” του κειμένου. Σε μία διάρθρωση του λόγου (discourse structure), η προαναφερθείσα διαδικασία είναι ένα σημαντικό μέρος του τί καθιστά ένα κείμενο συνεκτικό. Έτσι, η ανάλυση της δομής του λόγου χωρίζεται σε δύο διαδικασίες: στον κατακερματισμό του λόγου και στην ανάλυσή του. Η τμηματοποίηση του λόγου (Discourse Segmentation) σχετίζεται με τον προσδιορισμό ορίων σε μία ακολουθία λέξεων, προσδιορίζοντας το σημείο έναρξης των επόμενων μονάδων λόγου (units). Η ανάλυση του λόγου (Discourse Parsing) σχετίζεται με τον προσδιορισμό των σχέσεων μεταξύ των ακολουθιών των μονάδων λόγου (π.χ. η αιτιότητα, η αντίθεση κλπ.). Στην περίπτωση μας, το σύνολο αυτών των σχέσεων σχηματίζει ένα δέντρο. Τα φύλλα της αναπαράστασης του δέντρου του λόγου (Discourse Tree (DT)) αντιστοιχούν σε συνεχόμενα κείμενα, τα οποία ονομάζονται επίσης μονάδες στοιχειώδους λόγου (Elementary Discourse Units (EDUs)). Τέλος, τα EDUs συνδέονται μεταξύ τους επίσης με την ίδια μορφή σχέσεων (π.χ. αντίθεση).

Προηγούμενες μελέτες σχετικά με την ανάλυση του λόγου (Discourse Analysis) έχουν αποδειχτεί αρκετά επιτυχείς για τον εντοπισμό των προσεγγίσεων της Μηχανικής Μάθησης και των χαρακτηριστικών που είναι περισσότερο χρήσιμα για την αυτόματη κατά-

1	the	the	DT	O	2	det		
2	movie	movie	NN	O	3	nsubj		
3	begins	begin	VBZ	O	0	ROOT		
4	in	in	IN	O	6	case		
5	the	the	DT	DATE	6	det		
6	past	past	NN	DATE	3	nmod		
7	where	where	WRB	O	11	advmod		
8	a	a	DT	O	10	det		
9	young	young	JJ	O	10	amod		
10	boy	boy	NN	O	11	nsubj		
11	named	name	VBN	O	6	acl:relcl		
12	sam	sam	NN	O	13	compound		
13	attempts	attempt	NNS	O	11	dobj		
14	to	to	TO	O	15	mark		
15	save	save	VB	O	11	xcomp		
16	celebi	celebus	NNS	O	15	dobj		
17	from	from	IN	O	19	case		
18	a	a	DT	O	19	det		
19	hunter	hunter	NN	O	15	nmod		
20	.	.	.	O	3	punct		

Σχήμα 6.2: Ένα αντιπροσωπευτικό παράδειγμα μιας φράσης από το σύνολο δεδομένων *Movie Review Dataset* “the movie begins in the past where a young boy named sam attempts to save celebi from a hunter.” στη μορφή CoNLL μετά τη διαδικασία της προεπεξεργασίας. Η διαδικασία αυτή περιλαμβάνει τμηματοποίηση (tokenization) και διαχωρισμό των προτάσεων, επισήμειωση των μερών του λόγου (part-of-speech (POS) tagging), λημματοποίηση, σύστημα αναγνώρισης ονομαστικής καταχώρησης (Name Entity Recognition (NER)), χρήση αναλυτή (parsing), και coreference resolution κάνοντας χρήση του Stanford CoreNLP επισήμειωτή [139].

τμηση (Discourse Segmentation) και ανάλυση (Discourse Parsing) του λόγου [220]. Δυστυχώς ένα από τα μειονεκτήματα αυτής της μεθόδου είναι ο υψηλός χρόνος επεξεργασίας που απαιτείται για την ανάλυση του λόγου στα κείμενα Φυσικής Γλώσσας [96], γεγονός που παρεμποδίζει την εφαρμογή αυτής της μεθόδου σε σενάρια μεγάλης κλίμακας.

6.5 Εξεταζόμενα Σύνολα Δεδομένων

Η αξιολόγηση της προτεινόμενης μεθόδου ενσωμάτωσης πραγματοποιείται σε δύο είδη συνόλων δεδομένων² από διαφορετικούς τομείς και με διαφορετικό μήκος. Επιλέγουμε δηλαδή την αξιολόγηση ταινιών [163] και τη συλλογή FSD που σχετίζεται με κριτικές προϊόντων[228].

Movie Review Dataset: Επιλέγουμε να χρησιμοποιήσουμε το συγκεκριμένο σύνολο δεδομένων (αξιολόγηση ταινιών) που παρέχεται από τους Pang και Lee και είναι ελεύθερα διαθέσιμο³, καθώς πολλές ερευνητικές εργασίες στο τομέα της ανάλυσης συναισθήματος από κείμενο αξιολογούν την αποδοτικότητα των αλγοριθμικών τους προσεγγίσεων σύμφωνα με αυτό το σύνολο δεδομένων γεγονός που το καθιστά ιδανικό μέτρο σύγκρισης των δικών μας πειραματικών αποτελεσμάτων.

²Για την πειραματική μας ρύθμιση, χωρίζουμε τυχαία κάθε συλλογή σε ένα σύνολο δεδομένων εκπαίδευσης και ένα ελέγχου 75% και 25% αντίστοιχα.

³<http://www.cs.cornell.edu/people/pabo/movie-review-data>

Datasets	Subj./Pos. sent.	Obj./Neg. sent.	Unigrams	Bigrams
Movie Reviews [163]	5000	5000	4948	9103
FSD [228]	923	1320	1275	1996

Πίνακας 6.2: Οι συλλογές κειμένων που χρησιμοποιήθηκαν για την κατηγοριοποίηση σε δύο ομάδες (two-class). Ο Πίνακας περιλαμβάνει επίσης τον αριθμό των μονογραμμάτων (unigrams) and διγραμμάτων (bigrams) μετά τη διαδικασία της προεπεξεργασίας.

Finegrained Sentiment Dataset (FSD): Το συγκεκριμένο σύνολο δεδομένων περιέχει 294 κριτικές προϊόντων από διάφορες πηγές στο διαδίκτυο. Τα σχόλια είναι σχετικά ισοκατανεμημένα σε σχέση με τον εξεταζόμενο τομέα (π.χ. βιβλία, DVD, μουσική και βιντεοπαιχνίδια) και η συνολική αξιολόγηση είναι είτε θετική είτε αρνητική. Τα κύρια στατιστικά στοιχεία των συλλογών αναφέρονται στον Πίνακα 6.2.

6.6 Πειραματική Διαδικασία

Σε αυτή την ενότητα περιγράφουμε την προτεινόμενη μέθοδό μας στοχεύοντας στη περαιτέρω διερεύνηση του βέλτιστου μεγέθους και του είδους/μορφής των δεδομένων που κρίνονται ιδανικότερα για την ενσωμάτωση του πλαισίου και για την επιλογή των ιδανικών χαρακτηριστικών σε επίπεδο προτάσεων. Το σύνολο των χαρακτηριστικών αυτών που δημιουργήσαμε, παρουσιάζεται στον Πίνακα 6.6.1.

6.6.1 Πρόβλημα ταξινόμησης και Πειραματικές ρυθμίσεις

Η πειραματική διαδικασία που ακολουθούμε ορίζεται ως εξής:

Υποθέτουμε ότι η είσοδος είναι ένα σύνολο m από έγγραφα (documents): $\{d_1, d_2, \dots, d_m\}$ μαζί με το καθορισμένο θέμα (subject): $\{sub_1, sub_2, \dots, sub_m\}$. Κάθε d_i περιέχει n_i προτάσεις S^i : $\{s_1^i, s_2^i, \dots, s_{n_i}^i\}$. Για όλα τα έγγραφα, ορίζουμε την έξοδο ως εξής: για την j^{th} πρόταση στο i^{th} έγγραφο s_j^i , αποδίδεται ένα συναίσθημα $o_j^i \{P : , N : \}$ και ένα συναίσθημα $o_j^i \in \{S : subjective, O : objective\}$ αντίστοιχα.

Το μοντέλο CRF παρέχει ένα πιθανοτικό πλαίσιο για τον υπολογισμό της πιθανότητας των ακολουθιών των ετικετών Y που υπόκεινται σε ακολουθίες δεδομένων X που πρέπει να επισημειώσουμε.

Μέχρι στιγμής, μία ποικιλία χαρακτηριστικών έχει εξαχθεί από τις προτάσεις για την ανάλυση του συναίσθηματος και τις οποίες αξιοποιούμε μέσω του μοντέλου CRF. Σε αυτή την εργασία, προτείνουμε τη χρήση των χαρακτηριστικών που παρατίθενται στον Πίνακα 6.6.1. Ειδικότερα, χρησιμοποιήσαμε το CRF++⁴, το οποίο είναι ένα εργαλείο ανοιχτού κώδικα για την εφαρμογή του πλαισίου Μηχανικής Μάθησης για τη δημιουργία της

⁴<https://taku910.github.io/crfpp/>

γραμμικής μας αλυσίδας CRF, η οποία ακολουθεί μια αντιστοιχία ένα-προς-ένα μεταξύ των καταστάσεων και των ετικετών. Στόχος μας είναι να εντοπίσουμε και να οριοθετήσουμε τις πληροφορίες πλαισίου (π.χ. γειτονικές προτάσεις ή προτάσεις που συνδέονται με μεταβατικές λέξεις) μεταξύ των προτάσεων σε ένα έγγραφο. Η διαδικασία ενσωμάτωσης του συναισθήματος γίνεται επομένως ένα είδος επισημείωσης της αλληλουχίας. Ο στόχος του μοντέλου μας είναι να αντιστοιχίσει μια ετικέτα σε κάθε ακολουθία προτάσεων που συνθέτει την πρόταση.

Όσον αφορά στη μέθοδο RST, αυτή έχει εισαχθεί με επιτυχία στον τομέα της αξιολόγησης κειμένων [96, 272] ενώ συγχρόνως έχει επιβεβαιωθεί η ικανότητά της να χρησιμοποιεί σχέσεις λόγου (discourse relations) στο κείμενο για τον υπολογισμό των τιμών των συναισθημάτων. Συγκεκριμένα, η προσέγγισή μας χρησιμοποιεί έναν αναλυτή (parser) που υλοποιεί τη θεωρία της ρητορικής δομής (RST) [138] για τον εντοπισμό στοιχείων του λόγου (discourse elements) στο κείμενο. Συγκεκριμένα, μετατρέπουμε τα παραγόμενα αρχεία XML σε μορφή παρόμοια με αυτή της μορφής CoNLL, χρησιμοποιώντας το εργαλείο σχολιασμού Stanford CoreNLP [139]. Μετά το πέρας της ολοκλήρωσης της προεπεξεργασίας των δεδομένων μας, κατατάσσουμε τα τελευταία σε EDUs και χρησιμοποιώντας το εργαλείο ανάλυσης του λόγου (Discourse Parsing from Linear Projection (DPLP)) [104], δημιουργούμε τα δέντρα RST για τις προτάσεις με βάση το προτεινόμενο σύνολο χαρακτηριστικών. Τέλος, με τη χρήση του αναλυτή (parser) RST, παράγουμε περαιτέρω το αντίστοιχο αρχείο για κάθε έγγραφο.

Η φύση των δεδομένων εκπαίδευσης: Είναι επίσης σημαντικό να σημειωθεί ότι τα εκτιμώμενα σύνολα δεδομένων ποικίλουν ανάλογα με το μήκος του κειμένου που αναλύουμε κάθε φορά.

Μέθοδοι προσαρμογής (Adaptation schemes): Πειραματιζόμαστε με τους γραμμικούς ταξινομητές της βιβλιοθήκης LIBLINEAR⁵, η οποία υποστηρίζει τους ταξινομητές SVM και Logistic Regression (LR). Εξετάζουμε εκτενώς αυτούς τους ταξινομητές έναντι των δεδομένων εκπαίδευσης για να επιλέξουμε τον καλύτερο ταξινομητή. Βελτιστοποιούμε τους ταξινομητές χρησιμοποιώντας 5-fold cross-validation σε σχέση με τα δεδομένα εκπαίδευσης. Για κάθε συλλογή, επικυρώνουμε περαιτέρω με το σύνολο δοκιμών τον SVM ταξινομητή που πραγματοποίησε τον καλύτερο χρόνο εκπαίδευσης (αναφορικά με την F_1 μετρική). Με αυτό τον τρόπο, μειώνουμε τη διακύμανση των αποτελεσμάτων απόδοσης και κάνουμε τη σύγκριση λιγότερο εξαρτώμενη από το συγκεκριμένο σύνολο δοκιμών. Επαναλαμβάνουμε αυτή τη διαδικασία δέκα φορές και υπολογίζουμε κατά μέσο όρο την απόδοση σε αυτές τις δέκα υπο-ομάδες (folds). Τέλος, μετράμε τη στατιστική σημασία με τη δοκιμή paired, two-sided micro sign test.

⁵<https://www.csie.ntu.edu.tw/~cjlin/liblinear/>

Set	Features
vocabulary	Unigrams and Bigrams (binary)
num_pos_words	A positive word list of 1948 words [255]
num_neg_words	A negative word list of 4550 words [255]
exist_pos_emotic	A positive emoticon list of 52 emoticons. For example “=)” represents smiling
exist_neg_emotic	A negative emoticon list of 35 emoticons. For example “v_v” represents sadness
exist_comp_sent	Existence of comparative adjectives, adverbs, superlative adjectives and adverbs or phrases (“compare to”, “in contrast”, etc.)
conjunction_words	Subordinating, coordinating, and correlative conjunctions words
sent_post	Sentence position. if the sentence is within first 20% of the sentences, it’s a beginning sentence; an end sentence if within the last 20%, and middle for all others.
post_pos_words	Position of positive words occurring. 0: no positive words occur; 1: only exist in the first part of a sentence; 2: only exist in the second part; -1: exist in both parts (mixed)
post_neg_words	Position of negative words
post_negation_words	Position of negation words
comp_sub	Comparison subject: If the subjectivity is the same as the input subjectivity.
cos_sim_neigh_sent	cosine similarity score to neighboring sentences (previous sentence and next sentence) ^{α□}
LSI_sim_neigh_sent	LSI similarity score to neighboring sentences (previous sentence and next sentence) ^{β□}
context-aware RST	binary feature for each type of RST relationships. Every sentence has only one of these features set to 1

Πίνακας 6.3: Χαρακτηριστικά που χρησιμοποιήθηκαν για το πρόβλημα της επισημείωσης των ακολουθιών[248].

^{α□} We use cosine similarity to capture the word-level similarity.

^{β□} We use the dimension reduction method of Latent Semantic Indexing [51] to measure the semantic similarity.

6.7 Πειραματικά Αποτελέσματα

Στον Πίνακα 6.4 παραθέτουμε τα αποτελέσματα της ταξινόμησής μας και για τα δύο σύνολα δεδομένων. Παρουσιάζουμε τα αποτελέσματά μας ως τη μέση τιμή της βαθμολογίας F_1 για τις θετικές, τις αρνητικές, τις υποκειμενικές και τις αντικειμενικές κατηγορίες αντίστοιχα ώστε να ποσοτικοποιήσουμε την ποιότητα της ταξινόμησης. Αναφέρουμε επίσης μέτρα αξιολόγησης όπως η ακρίβεια (Pr) και η ανάκληση (R) για πληρότητα. Επιπλέον, δίνουμε απαντήσεις σε μία σειρά ερευνητικών ερωτημάτων. Θεωρούμε τα μονογράμματα (unigrams) και τα μονογράμματα σε συνδυασμό με τα διγράμματα (bigrams) ως συστήματα αναφοράς (baselines) και ενσωματώνουμε περαιτέρω το σύνολο χαρακτηριστικών που παρουσιάζεται στον Πίνακα 6.6.1 στον ταξινομητή μας.

6.7.1 Θεωρείται αυτή η κατηγορία κοινωνικών μέσων από την άποψη του μεγέθους και του είδους της, κατάλληλη για την ενσωμάτωση του πλαισίου;

Σύμφωνα με τα αποτελέσματά μας, χαρακτηριστικά όπως το μήκος λέξης/φράσης, οι λέξεις που φέρουν συναίσθημα και η θέση της λέξης/φράσης θα μπορούσαν να είναι ενδεικτικά της υποκειμενικότητας και της αντικειμενικότητας αντίστοιχα. Παρατηρούμε ότι τέτοια χαρακτηριστικά επηρεάζουν την ανάλυση απόδοσης και των δύο συνόλων δεδομένων και σε συνδυασμό με την προτεινόμενη προσέγγιση μας, μπορούμε να ενσωματώσουμε με επιτυχία πληροφορίες πλαισίου.

6.7.2 Είναι απαραίτητο να διευρύνουμε την ομάδα των επιλεγμένων χαρακτηριστικών για τη διευκόλυνση της ενσωμάτωσης του πλαισίου σε συστήματα SA;

Συνολικά, η ταξινόμηση προτάσεων με βάση τα μονογράμματα-διγράμματα (unigrams-bigrams) είναι μία αποτελεσματική και ασφαλής επιλογή. Τα **χαρακτηριστικά μήκους** διαπιστώνουμε ότι τελικά δεν συμβάλλουν στη διάκριση μεταξύ αντικειμενικών και υποκειμενικών προτάσεων. Από την άλλη, τα **χαρακτηριστικά θέσης** φαίνεται να δουλεύουν ιδιαίτερα καλά αναφορικά με τον εντοπισμό του υποκειμενικού περιεχομένου. Ακόμα κι αν δεν έχουμε στη συλλογή δεδομένων της αξιολόγησης ταινιών (MR) χαρακτηριστικά θέσης, υποθέτουμε ότι η ικανότητά τους να ταξινομούν αντικειμενικές προτάσεις είναι περιορισμένη⁶. Επιπλέον, η ενσωμάτωση των **context-aware RST** χαρακτηριστικών δεν λειτούργησε όσο ικανοποιητικά αναμέναμε. Προφανώς, η παρουσία συγκεκριμένων ρητορικών σχέσεων ανά φράση δεν συμβάλλει απαραίτητα στην ευρωστία των ταξινομητών όταν οι τελευταίοι χρησιμοποιούν μονογράμματα (unigrams) και διγράμματα (bigrams). Τέλος, όταν συνδυάσουμε **όλα τα χαρακτηριστικά** σε έναν ενιαίο ταξινομητή, διαπιστώνουμε ότι αποκτούμε έναν καλό ταξινομητή όσον αφορά στην ανάκληση υποκειμενικών προτάσεων. Αυτό οδήγησε σε απόδοση ταξινόμησης που ήταν μερικές φορές χειρότερη από την απόδοση των συστημάτων αναφοράς (baselines).

Με παρόμοιο τρόπο, αναφέρουμε την απόδοση ταξινόμησης στη συλλογή FSD αντίστοιχα. Μία βασική παρατήρηση είναι ότι η παρουσία των χαρακτηριστικών RST (μήκος (length), θέση (positional)) και **context-aware RST** δεν μεταφέρει πολύ περισσότερες πληροφορίες στην απόδοση της συνολικής ταξινόμησης από την άποψη της πολικότητας, ενώ φαίνεται ότι έχει πολύ περισσότερη δύναμη όταν υποδεικνύει υποκειμενικές προτάσεις. Τέλος, ο συνδυασμός του **όλων των χαρακτηριστικών** λειτούργησε καλά, αλλά ήταν κατώτερος από τα χαρακτηριστικά **sentiment-word**.

⁶ Αυτό μπορεί να εξηγηθεί από το γεγονός ότι έχουμε την τάση να χρησιμοποιούμε υποκειμενικές προτάσεις σε συγκεκριμένα μέρη ενός κειμένου π.χ. στην αρχή ή στο τέλος του.

6.7.3 Ποιός είναι ο συνδυασμός χαρακτηριστικών με τις καλύτερες επιδόσεις για τον εντοπισμό και την ενσωμάτωση πληροφοριών πλαισίου;

Ο συνδυασμός με τις καλύτερες επιδόσεις ήταν αυτός που περιελάμβανε τα χαρακτηριστικά **sentiment-word**. Ήταν το μοναδικό χαρακτηριστικό που ήταν σε θέση να βελτιώσει στατιστικά τα συστήματα αναφοράς (baselines) σε όλες τις καταστάσεις σε διαφορετικά σετ δοκιμών. Συνδυάζοντας τα μονογράμματα (unigrams) ή τα διγράμματα (bigrams) με τα χαρακτηριστικά του **sentiment-word** είναι ένας τρόπος να ταξινομούμε τόσο τις γενικές εκφράσεις της κοινής γνώμης όσο και τις εκφράσεις πεποιθήσεων συγκεκριμένων τομέων. Αυτό οδήγησε σε ισχυρούς ταξινομητές υποκειμενικότητας. Με τον ίδιο τρόπο, όσον αφορά στα αποτελέσματα της ταξινόμησης FSD, ένας από τους συνδυασμούς με τις καλύτερες επιδόσεις είναι και πάλι αυτός που περιλαμβάνει τα χαρακτηριστικά **sentiment-carrying words**.

6.8 Συμπεράσματα

Στο συγκεκριμένο κεφάλαιο μελετήσαμε διεξοδικά τη χρησιμότητα του μοντέλου CRF και της θεωρίας του λόγου RST για να ενσωματώσουμε τα χαρακτηριστικά της δομής της πρότασης και τις πληροφορίες πλαισίου σε συστήματα που έχουν επίγνωση του συναισθηματικού περιβάλλοντος. Για την πειραματική αξιολόγηση των μοντέλων μας χρησιμοποιήσαμε δύο πηγές πληροφοριών που έχουν διαφορετικό μέγεθος και ανήκουν σε διαφορετικό είδος. Αρχικά, δείξαμε πώς να βελτιώσουμε την SA εκμεταλλευόμενοι πλήρως τη δομή των προτάσεων, χρησιμοποιώντας πληροφορίες πλαισίου για να καταγράψουμε τη σχέση μεταξύ των προτάσεων, βελτιώνοντας έτσι και την SA του εγγράφου. Ο λόγος αυτής της επιτυχίας έγκειται στην επέκταση του συνόλου των χαρακτηριστικών πλαισίου που χρησιμοποιούνται συνδυαστικά με σημασιολογικά, συντακτικά και συναισθηματικά χαρακτηριστικά RST έναντι των παραδοσιακών προσεγγίσεων.

Τα πειραματικά μας αποτελέσματα υποστηρίζουν ότι για να ταξινομήσουμε τη συλλογή FSD, ο συνδυασμός λέξεων που αναφέρονται στα συναισθήματα σε συνδυασμό με τα μονογράμματα/διγράμματα (unigrams/bigrams) παρέχει επίσης αρκετά ακριβή αποτελέσματα. Επιπλέον, μόνο όταν συνδυάζουμε χαρακτηριστικά πλαισίου-RST με μονογράμματα/διγράμματα (unigrams/bigrams) μπορούμε να επιτύχουμε βελτιώσεις απόδοσης με σαφή πολικότητα σε επίπεδο προτάσεων. Τα σημαντικότερα χαρακτηριστικά των ταξινομητών πολικότητας καταγράφουν ουσιαστικά τον τρόπο με τον οποίο οι όροι πολικότητας (polarity terms) χρησιμοποιούνται σε μια πρόταση. Μία πιθανή επέκταση αυτής της δουλειάς θα μπορούσε να είναι η χρήση επιπρόσθετων χαρακτηριστικών συμβατών με τις μεθόδους μας αναφορικά με το εξεταζόμενο είδος δεδομένων καθώς και τη χρήση επιπρόσθετων μέτρων αξιολόγησης.

Features (MR)	Subjective			Objective			Microavg
	Prec.	Rec.	F_1	Prec.	Rec.	F_1	F_1
Unigrams	0.8939	0.8910	0.8924	0.8916	0.8944	0.8930	0.8927
Length	0.8614	0.8940	0.8774	0.8901	0.8565	0.8730	0.8752
Positional	-	-	-	-	-	-	-
Sentiment-carrying words	0.8926	0.8995	0.8960	0.8989	0.8920	0.8954	0.8958
RST	0.8934	0.8910	0.8922	0.8915	0.8939	0.8927	0.8924
All	0.8876	0.9005	0.8940	0.8993	0.8862	0.8927	0.8934
Uni+Bigrams	0.9043	0.8942	0.8992	0.8956	0.9055	0.9005	0.8999
Length	0.8829	0.8811	0.8820	0.8816	0.8834	0.8825	0.8822
Positional	-	-	-	-	-	-	-
Sentiment-carrying words	0.9016	0.8964	0.899	0.8973	0.9024	0.8998	0.8994
RST	0.9054	0.8888	0.8970	0.8910	0.9073	0.8991	0.8980
All	0.8999	0.9026	0.9012	0.9025	0.8999	0.9012	0.9012
Features (FSD)	Positive			Negative			Microavg
	Prec.	Rec.	F_1	Prec.	Rec.	F_1	F_1
Unigrams	0.6596	0.6175	0.6379	0.7302	0.7647	0.7471	0.7021
Length	0.6451	0.5195	0.5755	0.6897	0.7889	0.7360	0.6745
Positional	0.6720	0.6217	0.6459	0.7352	0.7758	0.7550	0.7104
Sentiment-carrying words	0.6936	0.6117	0.6825	0.7630	0.7808	0.7718	0.7345
RST	0.6690	0.6074	0.6367	0.7285	0.7780	0.7524	0.7055
All	0.6245	0.7348	0.6752	0.7747	0.6737	0.7207	0.6996
Uni+Bigrams	0.6801	0.5872	0.6302	0.7231	0.7960	0.7578	0.7073
Length	0.6618	0.4590	0.5421	0.6742	0.8268	0.7427	0.6705
Positional	0.6958	0.5855	0.6359	0.7260	0.8109	0.7661	0.7152
Sentiment-carrying words	0.7149	0.6578	0.6852	0.7614	0.8063	0.7832	0.7432
RST	0.6878	0.5734	0.6254	0.7194	0.8078	0.7610	0.7082
All	0.6297	0.7385	0.6798	0.7786	0.6793	0.7256	0.7045

Πίνακας 6.4: Ταξινόμηση των συλλογών δεδομένων *Movie Review (MR)* [163] και *Finegrained Sentiment (FSD)* [228] ως προς τα μέτρα αξιολόγησης: ακρίβεια (*Precision (Pr.)*), ανάκληση (*Recall (R.)*) και F_1 για κάθε λεξιλογική αναπαράσταση (π.χ. μονογράμματα (*unigrams*), ή μονογράμματα (*unigrams*) και διγράμματα (*bigrams*). Η καλύτερη απόδοση σημειώνεται με έντονα γράμματα. [248]

Κεφάλαιο 7

Εκμάθηση Αναπαραστάσεων της Φυσικής Γλώσσας με Ανταγωνιστικά Παραγωγικά Δίκτυα σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών

7.1 Εισαγωγή

Η απόδοση των μεθόδων Μηχανικής Μάθησης εξαρτάται σε μεγάλο βαθμό από την επιλογή των δεδομένων ή την αναπαράσταση των χαρακτηριστικών στα οποία εφαρμόζονται. Για το λόγο αυτό, μεγάλο μέρος της ουσιαστικής προσπάθειας στην ανάπτυξη αλγορίθμων Μηχανικής Μάθησης σχετίζεται με το σχεδιασμό της διαδικασίας της προεπεξεργασίας και του μετασχηματισμού των δεδομένων και της αναπαράστασης των δεδομένων που μπορούν να υποστηρίξουν τη Μηχανική Μάθηση με αποτελεσματικό τρόπο. Τέτοιου είδους μηχανική χαρακτηριστικών είναι σημαντική αλλά χρονοβόρα εργασία, ενώ συγχρόνως υπογραμμίζει και την αδυναμία των τωρινών αλγορίθμων Μηχανικής Μάθησης. Παρόλο που υπάρχει μια μεγάλη ποικιλία προσεγγίσεων στην αναπαράσταση της μάθησης (Representation Learning) γενικά, η βασική έννοια περιστρέφεται γύρω από την εκμάθηση συγκεκριμένων χαρακτηριστικών από τα δεδομένα και στη συνέχεια τη χρήση των τελευταίων για την επίλυση μιας σειράς λειτουργιών, για παράδειγμα, την επίλυση μιας ξεχωριστής (πιθανώς μη σχετιζόμενης) εργασίας για την οποία έχουμε ένα μεγάλο αριθμό επισημειωμένων παραδειγμάτων/δειγμάτων. Ως αποτέλεσμα, η εμφάνιση μιας σειράς δεδομένων μεγάλης κλίμακας, όπως το ImageNet [189], το οποίο περιέχει 14.197.122 εικόνες και έχει επισημειωθεί χειροκίνητα, επέτρεψε την ευρύτερη χρήση και τη δημοτικότητα των συνελκτικών νευρωνικών δικτύων (CNNs) ακόμη και σε λιγότερο σχετικούς τομείς, όπως αυτός της ιατρικής απεικόνισης. Επί του παρόντος, η πλειονότητα

των ταξινομητών δεν είναι αποτελεσματική όπως αναμένεται όταν το μέγεθος του συνόλου δεδομένων εκπαίδευσης είναι μικρό. Κατασκευάζοντας ένα μεγάλο και επισημειωμένο σύνολο δεδομένων, ωστόσο, είναι μια χρονοβόρα διαδικασία και συνήθως απαιτεί γνώσεις του εκάστοτε εξεταζόμενου τομέα, καθιστώντας την διαδικασία αυτή ακόμη πιο δαπανηρή. Ως εκ τούτου, υπάρχει ένα χάσμα μεταξύ των οφελών ενός μεγάλου συνόλου δεδομένων και της δυσκολίας της απόκτησης επισημειωμένων δεδομένων. Προκειμένου να επεκταθεί το εύρος και η ευκολία εφαρμογής της Μηχανικής Μάθησης, θα ήταν ιδιαίτερα επιθυμητό οι αλγόριθμοι αναπαράστασης να είναι λιγότερο εξαρτημένοι από τη μηχανική των χαρακτηριστικών, έτσι ώστε οι νέες εφαρμογές να μπορούν να κατασκευαστούν γρηγορότερα. Αυτό θα μπορούσε να είναι δυνατό μέσω της εκμάθησης αναπαράστασεων (Learning Representations) των δεδομένων που διευκολύνουν την εξαγωγή χρήσιμων πληροφοριών αναφορικά με την κατασκευή ταξινομητών. Μία καλή αναπαράσταση είναι αυτή που είναι επίσης χρήσιμη ως είσοδος σε έναν εποπτευόμενο (supervised) ταξινομητή. Μεταξύ των διαφόρων μεθόδων αναπαράστασης εκμάθησης (LR), το τρέχον κεφάλαιο επικεντρώνεται σε μεθόδους Βαθιάς Μάθησης: εκείνες δηλαδή που διαμορφώνονται από τη σύνθεση πολλαπλών μη γραμμικών μετασχηματισμών, με στόχο την απόδοση περισσότερο αφηρημένων- και, τελικά, περισσότερο χρήσιμων αναπαραστάσεων.

Μεταξύ των αποτελεσματικών προσεγγίσεων που έχουν προκύψει για την εκπαίδευση των βαθιά παραγωγικών μοντέλων, αυτό που βασίζεται στον παραλλαγμένο αυτόματο κωδικοποιητή (VAE) [115, 187] και αυτό που βασίζεται στην προσέγγιση που χρησιμοποιεί ανταγωνιστικά παραγωγικά (GAN) δίκτυα κυριαρχούν τα τελευταία χρόνια. Στο πρώτο, τα παρατηρούμενα δεδομένα x θεωρούμε ότι παράγονται από ένα σύνολο στοχαστικών λανθανόντων μεταβλητών z . Το VAE εισάγει ένα δίκτυο συμπερασμάτων (υλοποιημένο με τη χρήση ενός βαθιού νευρωνικού δικτύου) για να προσεγγίσουμε τις καταχρηστικές κατανομές πάνω από το z , και στη συνέχεια μεγιστοποιεί ένα κατώτερο όριο στην πιθανότητα καταγραφής του $p(x)$. Η τελευταία προσέγγιση χρησιμοποιεί τα GANs [83]. Στην αρχική διατύπωση του GAN, μια γεννήτρια βαθιού νευρωνικού δικτύου μαθαίνει να χαρτογραφεί δείγματα από μία αυθαίρετη κατανομή της παρατηρούμενης κατανομής δεδομένων. Ένα δεύτερο βαθύ νευρωνικό δίκτυο που ονομάζεται διευκρινιστής είναι εκπαιδευμένο στο να διακρίνει μεταξύ των δειγμάτων της εμπειρικής κατανομής και των δειγμάτων που παράγονται από τη γεννήτρια. Η γεννήτρια εκπαιδεύεται στη δημιουργία δειγμάτων που θα ξεγελάσουν το δίκτυο του διευκρινιστή και έτσι ένα αντίπαλο παιχνίδι παίζεται μεταξύ των δύο δικτύων, συγκλίνοντας σε ένα σημείο που είναι ένα τοπικό ελάχιστο για το διευκρινιστή και ένα τοπικό μέγιστο για τη γεννήτρια αντίστοιχα. Τόσο η προσέγγιση VAE όσο και η GAN έχουν δείξει εντυπωσιακά αποτελέσματα στην παραγωγή μοντέλων εικόνων [181, 88]. Παρόλα αυτά, σχετικά λίγη δουλειά έχει πραγματοποιηθεί αναφορικά με την αξιολόγηση της απόδοσης αυτών των μοντέλων μάθησης για την αναπαράσταση της Φυσικής Γλώσσας. Τα GANs αναφορικά με την επεξεργασία της Φυσικής Γλώσσας (NLP) χαρακτηρίζονται ως ισχυρές μέθοδοι, καθώς ασχολούνται με τη δημιουργία προτάσεων: συγκεκριμένα, παράγουν προτάσεις με συγκεκριμένα χαρακτη-

ριστικά (συναισθήματα και ερωτήσεις) και επωφελούνται από τη μη επιβλεπόμενη φύση αυτών των βαθιών νευρωνικών μοντέλων (DNN). Ένας λόγος για τον οποίο τα GANs δεν μπορούν να εφαρμοστούν απευθείας στον τομέα της Φυσικής Γλώσσας είναι το γεγονός ότι οι προτάσεις δεν αναπαρίστανται με συνεχή τρόπο και συνεπώς δεν μπορεί να υπάρξουν διαφοροποιήσεις. Αντίθετα, το κείμενο αναπαρίσταται από διακριτές λέξεις (tokens) (όπως “Άνθρωπος”, “κορίτσι”, κ.λπ.). Έτσι, όταν θέλουμε να ενημερώσουμε ελαφρώς την παραγόμενη πρόταση σύμφωνα με τη συμπεριφορά του διευκρινιστή (discriminator), δεν μπορεί να παραχθεί μία πρόταση. Στην περιοχή της Όρασης Υπολογιστών (CV), η έξοδος της γεννήτριας είναι μια εικόνα (ένας πίνακας που αποτελείται από πραγματικούς αριθμούς) που μπορεί να υποστεί μικρές ενημερώσεις ώστε να δυσκολευτεί το δίκτυο του διευκρινιστή να διακρίνει την πραγματική από την ψεύτικη εικόνα. Πρόσφατα, ωστόσο, τα VAE χρησιμοποιήθηκαν με επιτυχία για τη δημιουργία μοντέλων γλώσσας [18], για τη μοντελοποίηση εγγράφων και να απαντήσει σε ερωτήσεις [146]. Αυτό το έγγραφο προσπαθεί να διευκρινίσει το κατά πόσο τα GANs μπορούν να χρησιμοποιηθούν για την εκμάθηση αναπαραστάσεων της Φυσικής Γλώσσας σε μία μη επιβλεπόμενη κατάσταση.

Ιδιαίτερα, στο έργο μας, διατυπώνουμε το πρόβλημα της Μηχανικής Μάθησης ως εξής. Προτείνουμε μια νέα επέκταση των GANs που αντικαθιστά τον παραδοσιακό δυαδικό διευκρινιστή ταξινομητή με αυτόν που αποδίδει μία κλιμακούμενη ενέργεια σε κάθε σημείο της περιοχής εξόδου της γεννήτριας. Ο διευκρινιστής ελαχιστοποιεί την απώλεια που χρησιμοποιείται για την εκπαίδευση των ταξινομητών “μεγίστου περιθωρίου”, ενώ η γεννήτρια προσπαθεί να δημιουργήσει δείγματα με χαμηλή ενέργεια κάτω από το διευκρινιστή. Δείχνουμε ότι μια ισορροπία Nash [196] κάτω από αυτές τις συνθήκες συνθέτει μια γεννήτρια που ταιριάζει με την κατανομή δεδομένων (υποθέτοντας άπειρη χωρητικότητα). Πραγματοποιούμε πειράματα με το διευκρινιστή, με τη μορφή ενός autoencoder denoising (DAE), που προαιρετικά συμπεριλαμβάνει ένα ρυθμιστή που επιβάλλει τα παραγόμενα δείγματα που έχουν υψηλή ομοιότητα συνημίτονου προς άλλα δείγματα στη μικρές ομάδες δεδομένων (mini-batch). Η προτεινόμενη αρχιτεκτονική του νευρωνικού δικτύου βασίζεται σε μια παραλλαγή του προσφάτως προτεινόμενου energy-based GAN [270] που έχει αποδειχθεί ότι είναι κατάλληλο για την παραγωγή υψηλής ανάλυσης ψηφιακών εικόνων MNIST [52], παρέχοντας μια ποιοτική αξιολόγηση των αναπαραστάσεων εκμάθησης. Επιπλέον, αποφασίσαμε να αντικαταστήσουμε το τυποποιημένο πιθανοτικό GAN που χρησιμοποιήσαμε στο μοντέλο ενέργειας (χρησιμοποιώντας κατανομές του Gibbs [80]), όπως έχει προταθεί από το έργο των Kim και Bengio [114] και αποφασίσαμε να παρουσιάσουμε την ίδια ισορροπία Nash ως ένα βασικό δίκτυο GAN, αλλά μέσω μιας διαφορετικής και πιο γενικευμένης κατηγορίας loss functionals, όπως το hinge loss. Αυτή η επιλογή του πειραματικού σχεδιασμού βασίζεται στην προσπάθειά μας να επιλέξουμε το ζευγάρι των μοντέλων που συγκλίνει [263] και που επιδεικνύει μεγαλύτερη σταθερότητα από τα κανονικά GANs κατά τη διάρκεια της εκπαίδευσης [197].

Οι κύριες συνεισφορές μας συνοψίζονται ως εξής:

- Εξετάζουμε κατά πόσο τα GANs μπορούν να χρησιμοποιηθούν για την εκμάθηση

αναπαραστάσεων της Φυσικής Γλώσσας σε ένα μη επιβλεπόμενης μάθησης περιβάλλον σε επίπεδο εγγράφου, πρότασης και χαρακτηριστικών.

- Μεταξύ των διαφόρων μεθόδων των αναπαραστάσεων εκμάθησης, εστιάζουμε σε μεθόδους βαθιάς μάθησης, οι οποίες αντιστοιχούν σε περισσότερο αφηρημένες - και τελικά πιο χρήσιμες - αναπαραστάσεις.
- Γεφυρώνουμε την προσέγγιση της μη επιβλεπόμενης μάθησης με τα GAN και τους αυτόματους κωδικοποιητές.
- Προσεγγίζουμε το παραδοσιακό πλαίσιο GAN από μια εναλλακτική ενεργειακή οπτική (energy-based).
- Προτείνουμε τη χρήση μιας αρχιτεκτονικής νευρωνικών δικτύων που βασίζεται σε μια παραλλαγή του μοντέλου energy-based GAN αναφορικά με μία ανταγωνιστικά παραγωγική εκπαίδευση. Η συμβολή μας βασίζεται στη χρήση ενός απλού hinge loss, στο σημείο σύγκλισης του συστήματος, έτσι ώστε η γεννήτρια του energy-based GAN να παράγει σημεία που να ακολουθούν την υποκείμενη κατανομή των δεδομένων.
- Προτείνουμε επίσης τη χρήση της αρχιτεκτονικής autoencoder ως το δίκτυο του διευκρινιστή, στο οποίο η ενέργεια είναι το σφάλμα ανοικοδόμησης.
- Επικεντρωνόμαστε στο πλεονέκτημα των δικτύων GANs στο να ελέγχουν την επεξεργασία μεγάλου όγκου μη επισημειωμένων δεδομένων και όχι στην ικανότητά τους να παράγουν νέα δεδομένα.
- Τέλος, διεξάγουμε εκτεταμένα πειράματα με τη χρήση δεδομένων διαφορετικών τύπων, μήκους και είδους: το 20Newsgroup σώμα κειμένου, το σύνολο δεδομένων των ταινιών αξιολόγησης (MR) και το Finegrained Sentiment Σύνολο δεδομένων (FSD).

7.2 Σχετική Βιβλιογραφία

Η Μηχανική Μάθηση για τον τομέα της Φυσικής Γλώσσας μπορεί να εφαρμοστεί σε ένα ευρύ φάσμα μορφών κειμένου, από κριτικές πολλών προτάσεων και σχόλια, σε εκφράσεις γνώμης μέσω λέξεων. Η πιο συχνή προσέγγιση είναι η ταξινόμηση σε επίπεδο κειμένου. Μια άλλη προσέγγιση, η ταξινόμηση σε επίπεδο προτάσεων, περιορίζει την ανάλυση σε απλές προτάσεις αντί για ολόκληρα κείμενα. Συνήθως είναι πιο δύσκολο να επιλυθεί το πρόβλημα, αφού δεν υπάρχουν πολλές πληροφορίες που μπορούν να χρησιμοποιηθούν από τον ταξινομητή, καθώς οι προτάσεις είναι συνήθως πολύ μικρότερες από τα κείμενα. Σε σύγκριση με την ανάλυση κειμένων και προτάσεων, η ανάλυση σε επίπεδο χαρακτηριστικών (aspect level) είναι περισσότερο fine-grained. Σκοπός του είναι

να εξάγει και να συνοψίσει τις απόψεις των ανθρώπων που εκφράζονται στις πτυχές των οντοτήτων, που ονομάζονται επίσης και στόχοι (targets).

7.2.1 Τεχνικές Μηχανικής Μάθησης για NLP διεργασίες

Η ταξινόμηση σε επίπεδο κειμένου ή φράσης υλοποιείται συχνά με τη χρήση **επιβλεπόμενων μεθόδων Μηχανικής Μάθησης**. Αυτό περιλαμβάνει την εκπαίδευση ενός μοντέλου χρησιμοποιώντας ένα μεγάλο σύνολο σχολιασμένων δεδομένων που είναι σχετικό με το θέμα (topic-specific). Οποιαδήποτε υπάρχουσα επιβλεπόμενη μέθοδος Μηχανικής Μάθησης μπορεί να εφαρμοστεί στην ταξινόμηση σε επίπεδο κειμένου, όπως μηχανές φορέα υποστήριξης (SVM) ή τα κρυφά Μαρκοβιανά μοντέλα (HMMs). Συγκεκριμένα, οι Pang κ.ά. [166] σύγκριναν πολλές μεθόδους Μηχανικής Μάθησης χρησιμοποιώντας έναν ταξινομητή αξιολόγησης ταινιών, καταλήγοντας στο συμπέρασμα ότι οι SVM και οι Naive Bayes ταξινομητές έχουν την καλύτερη απόδοση συνολικά. Η παλινδρόμηση (regression) σε επίπεδο κειμένου ή φράσης είναι κρίσιμης σημασίας δεδομένου ότι πολλά προβλήματα δεν μπορούν να λυθούν αποτελεσματικά με θετική-αρνητική ταξινόμηση, η οποία χρησιμοποιείται συχνά για κριτικές προϊόντων, όπου επικρατεί βαθμολογία 1-5 αστερών. Οι Pang et al. [164] σύγκριναν διάφορες μεθόδους παλινδρόμησης, όπως η παλινδρόμηση SVM και η ταξινόμηση των πολυάριθμων SVM (SVM multiclass). Οι Qu et al. [179] επέκτειναν την αναπαράσταση bag-of-words εκμεταλλευόμενοι τους τροποποιητές της άρνησης και του συναισθήματος, οι οποίοι έχουν μεγαλύτερη επιρροή στην παλινδρόμηση απ' ό,τι σε προβλήματα ταξινόμησης. Αντίστοιχα, οι Mejova et al. [145] συγκέντρωσαν στρατηγικές επιλογής χαρακτηριστικών, όπως το stemming, η συχνότητα εμφάνισης των όρων, τα n-grams, το σημείο ομιλίας και χαρακτηριστικά εμπλουτισμένα με άρνηση. Κατέληξαν λοιπόν στο συμπέρασμα ότι ένα μικρότερο σύνολο χαρακτηριστικών είναι περισσότερο αποδοτικό συγκριτικά με ένα μεγαλύτερο σύνολο μεγάλων βάσεων δεδομένων.

Από την άλλη πλευρά, η **μη επιβλεπόμενη Μηχανική Μάθηση** για διεργασίες της Φυσικής Γλώσσας βασίζεται στην κυρίαρχη επίδραση των λέξεων και φράσεων συναισθηματικού περιεχομένου για την εκτέλεση της ταξινόμησης χωρίς τη χρήση δαπανηρών επισημειωμένων δεδομένων. Αυτό επιτυγχάνεται είτε με την εξαγωγή των συντακτικών προτύπων των προτάσεων είτε με τη χρήση λεξικών συναισθηματικών όρων. Και οι δύο προσεγγίσεις βασίζονται στη μέτρηση του προσανατολισμού του συναισθήματος (SO) των φράσεων και κατ' επέκταση ολόκληρου του κειμένου. Η ανάλυση που βασίζεται στη χρήση λεξικών με όρους συναισθηματικού περιεχομένου υπολογίζει τις τιμές του συναισθηματικού προσανατολισμού (SO) των λέξεων και φράσεων, που συνοψίζονται στην πολικότητα ολόκληρου του κειμένου. Γενικά, οι ταξινομητές συναισθημάτων που βασίζονται στα λεξικά συναισθηματικών όρων, δείχνουν μια θετική μεροληψία/προτίμηση, η οποία μπορεί να καθοριστεί προσαρμόζοντας την τιμή των περισσότερο σπάνιων αρνητικών εκφράσεων [226]. Μία από τις ελλείψεις τους είναι ότι οι μέθοδοι που βασίζονται σε λεξικά συναισθηματικών όρων δεν λειτουργούν καλά σε δεδομένα που εξαρτώνται από

τον εξεταζόμενο τομέα, καθιστώντάς τα λιγότερο αποτελεσματικά όταν χρησιμοποιούνται για τομείς που παρουσιάζουν περισσότερο ενδιαφέρον για την ανάλυση των αισθήσεων (SA), όπως η πολιτική. Το ζήτημα αυτό εξετάζεται εν μέρει στο [268]; παρόλα αυτά οι επιβλεπόμενες μέθοδοι μάθησης εξακολουθούν να υπερβαίνουν τις μεθόδους που βασίζονται στη χρήση συγκεκριμένων λεξικών, με εξειδικευμένη ορολογία αναφορικά με τον εξεταζόμενο τομέα.

Πρόσφατα, τα νευρωνικά δίκτυα (NNs) άρχισαν να επεκτείνονται στο πεδίο της Φυσικής Γλώσσας με τη μορφή τόσο των επιβλεπόμενων όσο και των μη επιβλεπόμενων μεθόδων αναπαράστασης μάθησης. Από την πλευρά της **μη επιβλεπόμενης αναπαράστασης της μάθησης** [160], αναπτύχθηκε και επικυρώθηκε ένα μεγάλο μέρος της πρώιμης έρευνας αναφορικά με τη σύγχρονη βαθιά μάθηση [98, 102, 238, 40]. Παρόλο που οι επιβλεπόμενες προσεγγίσεις έχουν σαφείς στόχους που μπορούν να βελτιστοποιηθούν άμεσα, οι μη επιβλεπόμενες μάθησης προσεγγίσεις βασίζονται σε εξουσιοδοτημένες εργασίες, όπως η ανακατασκευή, η εκτίμηση της πυκνότητας ή η παραγωγή/δημιουργία, εργασίες οι οποίες δεν ενθαρρύνουν άμεσα χρήσιμες αναπαραστάσεις συγκεκριμένων διεργασιών. Ως αποτέλεσμα, μεγάλος όγκος δουλειάς έχει πραγματοποιηθεί αναφορικά με το σχεδιασμό στόχων και αρχιτεκτονικών που προορίζονται να ενθαρρύνουν την εκμάθηση χρήσιμων αναπαραστάσεων.

Παρά τις δυσκολίες αυτές, υπάρχουν αξιόλογες εφαρμογές μη επιβλεπόμενης μάθησης. Pretrained διανύσματα λέξεων είναι ζωτικής σημασίας για πολλά σύγχρονα συστήματα NLP [41]. Αυτές οι αναπαραστάσεις, η εκμάθηση των οποίων έχει πραγματοποιηθεί μέσω της μοντελοποίησης λέξεων που εμφανίζουν συσχέτιση (word co-occurrences), αυξάνουν την αποδοτικότητα των δεδομένων και την ικανότητα της γενίκευσης των συστημάτων Φυσικής Γλώσσας [171, 34]. Η εκμάθηση αναπαράστασης φράσεων, προτάσεων και εγγράφων είναι ένας ανοικτός χώρος της έρευνας. Εμπνευσμένοι από την επιτυχία των διανυσμάτων από λέξεις, η δουλειά των Kiros et al. [117] πρότεινε skip-thought διανύσματα, μία μέθοδο εκπαίδευσης μίας πρότασης κωδικοποιητή προβλέποντας την προηγούμενη και την επόμενη πρόταση. Η αναπαράσταση που προέκυψε από την εκμάθηση αυτού του μοντέλου λειτουργεί ανταγωνιστικά σε μία ευρεία σειρά διεργασιών που έχουν αξιολογηθεί. Οι περισσότεροι προηγούμενες τεχνικές εκπαίδευσης, όπως η εξομάλυνση του στρώματος [8], βελτίωσαν περαιτέρω τα αποτελέσματα. Ωστόσο, η απόδοση των των skip-thought διανυσμάτων εξακολουθεί να ξεπερνιέται από τα μοντέλα επιβλεπόμενης μάθησης, τα οποία βελτιστοποιούν άμεσα την επιθυμητή μετρική της απόδοσης για ένα συγκεκριμένο σύνολο δεδομένων. Αυτό συμβαίνει και για τις δύο εργασίες ταξινόμησης κειμένου, οι οποίες μετρούν κατά πόσο μία συγκεκριμένη έννοια είναι σωστά κωδικοποιημένη σε μια αναπαράσταση. Αυτό συμβαίνει ακόμη και όταν τα σύνολα δεδομένων χαρακτηρίζονται ως μικρά σύμφωνα με τα σύγχρονα πρότυπα; τα οποία συχνά αποτελούνται μόνο από μερικές χιλιάδες επισημειωμένα παραδείγματα.

Work	Level of Analysis	Model	Dataset	Task	Evaluation metrics
Rahman et al. [182]	Docs.	LSTM	Large Movie Review (50.000 reviews)	classification	Acc.: 80%
Lai et al. [122]	Docs.	RCNN	20Newsgroups; Fudan Set; ACL Anthology Network; SST	classification	Macro-F1: 96.49% Acc.: 95.2% Acc.: 49.19% Acc.: 47.21%
Shen et al. [213]	Docs.	CNN + BLSTM	Large Movie Review (50.000 reviews)	classification	Acc.: 89.7%
Yender & Verna [261]	Docs.	CNN + LSTM	Large Movie Review (50.000 reviews)	classification	Acc.: 89.5%
Liu et al. [136]	Docs. Sents.	RNN	SST1 (Sents.); SST2 (Sents.); Movie Reviews (subj/obj. reviews); Large Movie Review (50.000 reviews)	classification	Acc.: 49.6% Acc.: 87.9% Acc.: 94.1% Acc.: 91.3%
Chen et al. [35]	Sents.	BiLSTM-CRF CNN	MPQA opinion corpus; SST; Movie Reviews (polarity v1.0)	classification; target extraction	- Acc.: up to 88.3% Acc.: 82.3%
Wang X et al. [251]	Sents.	CNN + RNN	SST1; SST2; Movie Reviews	classification	Acc.: 51.50% Acc.: 89.95% Acc.: 82.28%
Conneau et al [43]	Sents.	Very Deep CNN	Product Reviews; News	classification	not reported
Wang J et al. [249]	Sents.	CNN + LSTM	SST; Chinese VA Texts	dimensional regression	RMSE/MAE/r: 1.341/0.987/0.778 RMSE/MAE/r: 0.874/0.689/0.557
Du et al. [59]	Docs.	Deep CNN	Amazon reviews	aspect classification	Acc.: 94.38%
Wang Y et al. [252]	Sents.	attention based LSTM	SemEval 2014 (Task 4)	aspect (binary) classification	Acc.: 89.9%
Poria et al. [177]	Sents.	Deep CNN	SemEval 2014; Aspect-based dataset	aspect extraction; classification	Acc.: up to 87.2%
Tang et al. [230]	Sents.	Deep Memory Network	SemEval 2014	aspect classification	Acc.: 80.95%

Πίνακας 7.1: Βασικά χαρακτηριστικά ορισμένων από τις πρόσφατα δημοσιευμένες ερευνητικές δραστηριότητες αναφορικά με τα Βαθιά Νευρωνικά Δίκτυα για τον τομέα της Ανάλυσης Γνώμης. Ο Πίνακας αναφέρεται στο επίπεδο ανάλυσης (*Level of Analysis*), στον τύπο του νευρωνικού Δικτύου που εφαρμόσαμε (*Model*), τα σύνολα δεδομένων που χρησιμοποιήσαμε (*Datasets*), το είδος της εφαρμογής (*Task*) και τις μετρικές αξιολόγησης (*Evaluation Metrics*). Οι μετρικές αξιολόγησης επισημειώνονται ως εξής: *Accuracy (Acc.)*, *Macro-F1 measure*, *Root Mean Square Error (RMSE)*, *Mean Absolute Error (MAE)* και *Pearson correlation coefficient (r)*[241].

7.2.2 Μέθοδοι Βαθιών Νευρωνικών Δικτύων σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών

Στην τρέχουσα , συγκεντρώνουμε και παραθέτουμε στον Πίνακα 7.1 τις τελευταίες ερευνητικές προσπάθειες, αναφορικά με την εφαρμογή των Νευρωνικών Δικτύων (ΝΔ) για διεργασίες της Φυσικής Γλώσσας, εστιάζοντας στην ανάλυση κειμένων, προτάσεων και χαρακτηριστικών. Στο σημείο αυτό αξίζει να σημειωθεί ότι χωρίζουμε τον Πίνακα 7.1 σε τρία υπομήματα ώστε να έχουμε μία περισσότερο σαφή διάκριση μεταξύ των ερευνητικών προσπαθειών που πραγματοποιήθηκαν αναφορικά με τα τρία προαναφερθέντα επίπεδα ανάλυσης.

Οι Rahman et al. [182] εισήγαγαν μια παραλλαγή της μακράς βραχυπρόθεσμης μνήμης (LSTM) και τη δοκίμασαν στη Large Movie Review Dataset **των κειμένων** των 50K. Επιπλέον, σύμφωνα με την έρευνα των τελευταίων αποδείχτηκε ότι παρόλο που η ακρίβεια του συστήματος δεν αυξήθηκε σε σύγκριση με τα κανονικά μοντέλα LSTM, η σταθερότητα και η ακρίβειά του παρουσίασαν βελτίωση. Επιπλέον, σύμφωνα με την εργασία [122], η τελευταία εισήγαγε ένα επαναλαμβανόμενο συνελκτικό νευρωνικό δίκτυο (RCNN) που εφάρμοσε μια αμφίδρομη επαναλαμβανόμενη δομή για να εντοπίσει τις πληροφορίες πλαισίου του κειμένου και στη συνέχεια χρησιμοποίησε ένα στρώμα μέγιστης συγκέντρωσης για να καταγράψει τα βασικά τμήματα-μέρη του κειμένου. Οι ερευνητές του [213] πραγματοποίησαν δοκιμές σύγκρισης με ένα CNN, ένα LSTM και ένα συνδυασμό CNN και LSTM στη βάση δεδομένων Large Movie Review για την ταξινόμηση του συναισθήματος, αποδεικνύοντας ότι το δίκτυο CNN + LSTM είναι αποτελεσματικότερο σε σχέση με τα αντίστοιχα προαναφερθέντα Νευρωνικά Δίκτυα μεμονομένα. Μια παρόμοια ερευνητική προσπάθεια διεξήχθη από τους συγγραφείς της [261] συνδυάζοντας τα δίκτυα CNN και LSTM για τη δημιουργία πολλαπλών παραλλαγών, οι οποίες επιτυγχάνουν αποδόσεις της τελευταίας τεχνολογίας. Για άλλη μια φορά, χρησιμοποιήθηκε η βάση δεδομένων Large Movie Review, αλλά οι συγγραφείς εξέφρασαν την πεποίθησή τους ότι το προτεινόμενο μοντέλο μπορεί να λειτουργήσει αποδοτικά τόσο σε εφαρμογές ήχου όσο και σε βίντεο. Τέλος, οι Liu et al. [136] πειραματίστηκαν με RNNs, εισάγοντας την έννοια της πολυδιάστατης εκμάθησης (multitask learning), η οποία συνδέει όλα τα σχετικά καθήκοντα σε ένα ενιαίο σύστημα, το οποίο εκπαιδεύεται από κοινού. Οι συγγραφείς πραγματοποίησαν ταξινόμηση τόσο σε επίπεδο κειμένου στη βάση δεδομένων Large Movie Review αλλά και σε επίπεδο προτάσεων στη Stanford Sentiment Treebank (SST) βάση δεδομένων, επιτυγχάνοντας κορυφαία αποτελέσματα απόδοσης μετά τον εντοπισμό του συνδυασμού των περισσότερο αποδοτικών παραμέτρων (fine-tuning).

Επιπλέον, τα Νευρωνικά Δίκτυα έχουν χρησιμοποιηθεί ευρέως για την ταξινόμηση **σε επίπεδο προτάσεων**. Ιδιαίτερα, οι Chen et al. [35] πρότειναν ένα μοντέλο CNN-LSTM το οποίο ταξινομεί πρώτα τις προτάσεις σε non-target, one-target και σε προτάσεις πολλαπλών στόχων (multi-target). Το μοντέλο τους δοκιμάστηκε σε πολλά σύνολα δεδομένων (Stanford Sentiment Treebank, σε κριτικές ταινιών και σε σχόλια προϊόντων του Amazon)

πετυχαίνοντας κορυφαία απόδοση σε μερικά από αυτά. Σύμφωνα με το ότι ο συνδυασμός CNN και RNN θεωρείται πολύ δημοφιλής προσέγγιση, οι συγγραφείς των [251] πρότειναν μία παραλλαγή του, η οποία πέτυχε υψηλή ακρίβεια στις βάσεις δεδομένων Stanford Sentiment Treebank και Movie Reviews. Επιπλέον, οι Conneau et al. [43] πρότειναν ένα πολύ βαθύ συνελκτικό δίκτυο που αποτελείται από 29 στρώματα, με αποτέλεσμα τη βελτιωμένη απόδοση και την απόδειξη της χρήσης πολλαπλών επιπέδων για τον τομέα της Φυσικής Γλώσσας επίσης. Ειδικότερα, οι συγγραφείς αξιολόγησαν το μοντέλο τους σε οκτώ διαφορετικά σύνολα δεδομένων που περιλαμβάνουν κριτικές ταινιών και ειδήσεις, δοκιμάζοντας διάφορα μοντέλα Φυσικής Γλώσσας όπως SA, την κατηγοριοποίηση ειδήσεων και την ταξινόμηση θέματος. Επιπλέον, οι Wang J et al. [251] πρότειναν ένα περιφερειακό μοντέλο CNN-LSTM το οποίο χρησιμοποίησε μία ξεχωριστή πρόταση ως την εξεταζόμενη περιοχή για την εξαγωγή των συναισθηματικών πληροφοριών. Το μοντέλο αυτό εκπαιδεύτηκε με το σύνολο δεδομένων Stanford Sentiment Treebank και με Κινέζικα κείμενα χρησιμοποιώντας 2K προτάσεις από κοινωνικά φόρουμ, παρουσιάζοντας καλύτερη απόδοση από τα CNNs, RNNs ή LSTMs αναφορικά με την πρόβλεψη του σθένους και της διέγερσης, αντίστοιχα.

Τέλος, διερευνήθηκε και η ανάλυση της Φυσικής Γλώσσας **σε επίπεδο χαρακτηριστικών (aspect level)**. Οι Du et al. [59] μοντελοποίησαν τα συναισθήματα και το συντακτικό πλαίσιο κάτω από συγκεκριμένες πτυχές για να αποκτήσουν καλύτερη ενσωμάτωση λέξεων, οι οποίες δόθηκαν ως εισροή σε ένα CNN για την κατάταξη συναισθημάτων των αξιολογήσεων προϊόντων του Amazon. Τα αποτελέσματά τους έδειξαν μία βελτίωση σε σύγκριση με τις παραδοσιακές μεθόδους ενσωμάτωσης λέξεων. Αντίστοιχα, οι Wang et al. [252] εφάρμοσαν τα δίκτυα LSTMs για την ταξινόμηση συναισθημάτων σε επίπεδο χαρακτηριστικών, επιτυγχάνοντας μια κορυφαία απόδοση του 89.9%. Το μοντέλο αξιολογήθηκε χρησιμοποιώντας το σύνολο δεδομένων SemEval 2014 ενώ οι αρχικές ενσωματώσεις λέξεων έγιναν από το Glove¹, καταγράφοντας τα σημαντικά μέρη της φράσης όταν δίνονται διάφορα χαρακτηριστικά. Οι Poria et al. [177] εξέτασαν ένα βαθύ CNN για την εξαγωγή χαρακτηριστικών χρησιμοποιώντας 7 επίπεδα NN. Για την αξιολόγηση χρησιμοποιήθηκε το σύνολο δεδομένων SemEval 2014 και ένα σύνολο δεδομένων SA και παρατηρήθηκε βελτίωση στα ποσοστά της ακρίβειας και της ανάκλησης. Έχουν εφαρμοστεί επίσης δίκτυα βαθιάς μνήμης σε ταξινόμηση κειμένου ανά επίπεδο και έχουν παρουσιάσει συγκρίσιμα αποτελέσματα με τα LSTM ενώ είναι 15 φορές ταχύτερα την ίδια στιγμή [230].

Αφού συνοψίσαμε τις τρέχουσες και τις προηγούμενες ερευνητικές προσπάθειες (στον Πίνακα 7.1) που έχουν διεξαχθεί στην περιοχή των βαθιών νευρωνικών δικτύων σε επίπεδο των κειμένων, της φράσης και των χαρακτηριστικών της ανάλυσης, παρατηρήσαμε ότι, στις περισσότερες περιπτώσεις, έχουν χρησιμοποιηθεί δίκτυα όπως το CNN ή τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNN) και συγκεκριμένα τα δίκτυα αμφίδρομης μακράς βραχυπρόθεσμης μνήμης (BLSTMs). Επομένως, πολύ λίγες ερευνητικές προσπά-

¹<https://nlp.stanford.edu/projects/glove/>

θεις έχουν διεξαχθεί για την ανάλυση κειμένων χρησιμοποιώντας παραγωγικά μοντέλα και συγκεκριμένα GAN. Θεωρώντας ότι ένας από τους στόχους της εργασίας μας είναι να εξετάσουμε κατά πόσο τα GANs μπορούν να χρησιμοποιηθούν για την εκμάθηση των αναπαραστάσεων των κειμένων της Φυσικής Γλώσσας με ένα ανεξέλεγκτο τρόπο, στην επόμενη ενότητα 7.2.3, συνοψίζουμε τα περισσότερα από τα πλέον σύγχρονα ερευνητικά έργα που έχουν χρησιμοποιήσει την προαναφερθείσα αρχιτεκτονική. Επιπλέον, κινητοποιημένοι από το γεγονός ότι η δομή του GAN δικτύου μπορεί να ενσωματώσει διάφορες λειτουργίες απώλειας (loss functions), το προτεινόμενο μοντέλο μας παρέχει ένα μεγαλύτερο βαθμό ελευθερίας. Επιπλέον, υποστηρίζουμε ότι είναι ικανό να προσφέρει ελπιδοφόρες και δημιουργικές λύσεις που να παράγουν χρήσιμα δεδομένα.

7.2.3 Ανταγωνιστικά Παραγωγικά Δίκτυα για διεργασίες Φυσικής Γλώσσας

Σε αυτή την ενότητα, εξετάζουμε την πρόσφατη έρευνα αναφορικά με τη διερεύνηση της δομής της Φυσικής Γλώσσας με τη χρήση δικτύων (VAE)s [116] και GANs [83]. Η αξιολόγηση των βαθιών παραγωγικών μοντέλων θεωρείται ερευνητική πρόκληση μέχρι στιγμής. Από όσο γνωρίζουμε, υπάρχουν ελάχιστες ερευνητικές προσπάθειες για την ανάλυση του κειμένου με τη χρήση των GANs. Οι Zhang et al. [269] πρότειναν ένα πλαίσιο για τη χρήση του LSTM και του CNN για την αμοιβαία ανταγωνιστική εκπαίδευση για την παραγωγή ρεαλιστικού κειμένου. Ο latent code z τροφοδοτείται στη γεννήτρια LSTM σε κάθε βήμα, ενώ το CNN λειτουργεί ως ταξινομητής δυαδικών προτάσεων ο οποίος διακρίνει μεταξύ πραγματικών δεδομένων και δειγμάτων που παράγονται. Τα LSTM δίκτυα ή οι περιοδικές επαναλαμβανόμενες μονάδες (GRU) που διατηρούν τη μακροπρόθεσμη μνήμη των tokens, έχουν αποδείξει ότι λειτουργούν καλά στην πράξη χρησιμοποιώντας την maximum likelihood estimation.

Ωστόσο, η εκπαίδευση με τη μέγιστη πιθανότητα έχει τα μειονεκτήματά της, όπως η μεροληψία έκθεσης (exposure bias). Αυτό αναφέρεται στην κατάσταση κατά την οποία, κατά τη διάρκεια της εκπαίδευσης, η πρόβλεψη της επόμενης λέξης η οποία εξαρτάται από την προηγούμενη είναι μη εφικτή, καθώς η προηγούμενη λέξη μπορεί να μην έχει εμφανιστεί στα δεδομένα εκπαίδευσης. Εάν η γεννήτρια κάνει ένα λάθος νωρίς στη διαδικασία παραγωγής, η παραγόμενη πρόταση θα παραμείνει αποκλίνουσα καθώς παράγονται περισσότερες λέξεις. Για την επίλυση αυτού του ζητήματος, μία σειρά μεθοδολογιών έχει εφαρμοστεί στο παρελθόν, όπως η προγραμματισμένη δειγματοληψία [13] παρέχοντας βαθμολογίες αλληλουχίας που αφορούν συγκεκριμένες εργασίες. Ένα πρόβλημα με την εφαρμογή του GAN δικτύου στο κείμενο είναι ότι τα gradients του δικτύου της γεννήτριας (generator) δεν μπορούν να προωθήσουν σωστά τις διακριτές μεταβλητές. Στο [269], το πρόβλημα αυτό επιλύεται κάνοντας την πρόβλεψη της λέξης ομαλοποιώντας κάθε φορά τις λέξεις στο χώρο ενσωμάτωσης λέξεων. Επιπλέον, οι Yu et al. [262] πρότειναν την παράκαμψη αυτού του προβλήματος μοντελοποιώντας τη γεννήτρια ως stochastic policy.

Στην περίπτωση αυτή το σήμα ανταμοιβής προέρχεται από τον διευκρινιστή GAN. Για την ανάλυση του κειμένου, είναι δυνατό να δημιουργηθούν δεδομένα εκπαίδευσης από ένα καθορισμένο σύνολο γραμματικών και έπειτα να αξιολογηθούν τα παραγωγικά μοντέλα με βάση το εάν (ή πόσο καλά) τα παραγόμενα δείγματα συμφωνούν με την προκαθορισμένη γραμματική [225]. Θεωρώντας ότι είναι δύσκολο να γίνει μια καλή αξιολόγηση για τη δημιουργία κειμένου δεδομένου ότι δεν υπάρχει αντικειμενικός τρόπος να εκτιμηθεί κατά πόσο μία “τεχνητή” πρόταση είναι περισσότερο plausible ή ρεαλιστική από ό,τι μία άλλη, μερικές ερευνητικές προσπάθειες έχουν χρησιμοποιήσει τα BLEU ποσοστά δειγμάτων σε μεγάλο αριθμό αθέατων δεδομένων δοκιμών. Η ικανότητα παραγωγής παρόμοιων προτάσεων σε μη ορατά πραγματικά δεδομένα θεωρείται ένδειξη για τη μέτρηση ποιότητας [262].

Συγκεκριμένα, όσον αφορά στη δημιουργία διαδοχικών δεδομένων με GANs δίκτυα, μια εναλλακτική προσέγγιση που βασίζεται στη μάθηση ενίσχυσης χρησιμοποιήθηκε για την εκπαίδευση του δικτύου GAN. Γνωρίζουμε μόνο μία εργασία που χρησιμοποιεί GAN για τη δημιουργία συνεχόμενων ακολουθιών, στόχος της οποίας ήταν η παραγωγή πολυφωνικής μουσικής χρησιμοποιώντας ένα GAN με έναν LSTM διευκρινιστή [150]. Μία ερευνητική δουλειά αναφορικά με την παραγωγή ελεγχόμενου κειμένου [101] χρησιμοποίησε τον μεταβλητό αυτόματο κωδικοποιητή (VAE) μαζί με τις ελεγχόμενες πληροφορίες δημιουργώντας προτάσεις κατηγορίας. Τέλος, οι Zhang et al. [269] και οι Semeniuta et al. [211] χρησιμοποίησαν το δίκτυο GAN για την παραγωγή κειμένου επιτυγχάνοντας state-of-the-art αποτελέσματα. Τέλος, μια άλλη κατηγορία προσέγγισης είναι το δίκτυο conditional GAN [149], το οποίο μας επιτρέπει να κατευθύνουμε τα δεδομένα κατά τη διαδικασία παραγωγής. Αυτή η προσέγγιση χρησιμοποιήθηκε κυρίως για τη δημιουργία εικόνων [181, 4]. Πρόσφατα, οι αρχιτεκτονικές conditional GAN έχουν επίσης χρησιμοποιηθεί στην επεξεργασία κειμένου, συμπεριλαμβανομένης της μετάφρασης [262] και της παραγωγής διαλόγου [129].

Έχοντας συνοψίσει την εργασία που διεξήχθη στον τομέα των GAN για τις εφαρμογές Φυσικής Γλώσσας, παρατηρούμε ότι παρόλο που η ερευνητική προσπάθεια έχει μετατοπιστεί στις προσεγγίσεις GAN για εφαρμογές της Φυσικής Γλώσσας, των οποίων τα μοντέλα του διευκρινιστή και γεννήτριας είναι κυρίως CNN και/ή RNN (BLSTMs), παρόλα αυτά, δεν έχει επιλυθεί μία σειρά περιορισμών.

Έχοντας αυτό υπόψιν μας, η προσπάθειά μας στο συγκεκριμένο κεφάλαιο αναλύεται στην επόμενη ενότητα 7.3, είναι να ερευνήσουμε κατά πόσο τα GANs θεωρούνται μια κατάλληλη επιλογή μοντέλου για να μάθουμε αναπαραστάσεις της Φυσικής Γλώσσας σε μία μη επιβλεπόμενη ρύθμιση σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών. Επιπλέον, επανεξετάζουμε το παραδοσιακό πλαίσιο GAN από μία εναλλακτική προσέγγιση, αυτή της energy-based. Συγκεκριμένα, προτείνουμε μία αρχιτεκτονική νευρωνικού δικτύου που βασίζεται σε μια παραλλαγή της energy-based GAN [270] για την ανταγωνιστικά παραγωγική εκπαίδευση του δικτύου. Η συμβολή μας βασίζεται στη χρήση μιας απλής hinge loss, στο σημείο όπου το σύστημα συγκλίνει, έτσι ώστε η γεννήτρια

του energy-based GAN δικτύου να παράγει σημεία που να ακολουθούν την υποκείμενη κατανομή δεδομένων. Προτείνουμε λοιπόν να χρησιμοποιήσουμε μία αρχιτεκτονική αυτόματου κωδικοποιητή ως διευκρινιστή στην οποία η ενέργεια είναι το σφάλμα ανοικοδόμησης. Η πειραματική επιλογή του σχεδιασμού μας βασίζεται στην προσπάθειά μας να αποκτήσουμε το ζευγάρι των μοντέλων το οποίο να συγκλίνει [263] και το οποίο θα επιδεικνύει μία περισσότερο σταθερή συμπεριφορά από τα κανονικά GAN δίκτυα κατά τη διάρκεια της εκπαίδευσής του [197].

7.3 Προτεινόμενη Προσέγγιση

Σε αυτή την ενότητα παρουσιάζουμε την προτεινόμενη προσέγγισή μας επισημαίνοντας τους περιορισμούς των σημερινών energy-based μοντέλων ενώ συγχρόνως παραθέτουμε τα πλεονεκτήματα των δικτύων GANs. Τα GANs έχουν μεγάλη σημασία για την ανάπτυξη των γενετικών μοντέλων. Δεδομένου ότι αποτελούν μία ισχυρή κατηγορία τους, τα GANs επιλύουν το πρόβλημα της δημιουργίας δεδομένων τα οποία μπορούν να ερμηνευτούν με φυσικό τρόπο. Ειδικότερα, για τη δημιουργία μεγάλης διάστασης δεδομένων, η δομή του νευρωνικού δικτύου που υιοθετείται δεν περιορίζει τη διάσταση της παραγωγής, χάρη στην οποία διευρύνεται σημαντικά το πεδίο των παραγόμενων δειγμάτων. Επιπρόσθετα, η δομή του νευρωνικού δικτύου μπορεί να ενσωματώσει διάφορες loss functions, αυξάνοντας έτσι τον βαθμό ελευθερίας του σχεδιασμού του μοντέλου. Όσον αφορά στα energy-based μοντέλα, έχουν χρησιμοποιηθεί για να απομονώσουν τις εξαρτήσεις έναντι των μεταβλητών καθορίζοντας μία energy-based συνάρτηση. Η συνάρτηση της ενέργειας συνδέει κάθε διαμόρφωση των μεταβλητών με μία κλίμακα ενεργειακής τιμής. Έτσι, οι χαμηλότερες τιμές ενέργειας ανατίθενται σε περισσότερο πιθανές διαμορφώσεις. Η λειτουργία αυτή έχει χρησιμοποιηθεί, για παράδειγμα, για την εκτίμηση της κατανομής πιθανότητας σύμφωνα με την κατανομή Boltzmann η οποία ορίζεται από μία ενεργειακή συνάρτηση και έναν κατάλληλο συντελεστή κανονικοποίησης. Στην περίπτωση αυτή, η ενεργειακή συνάρτηση ορίζεται για να εκχωρήσει μία τιμή πιθανότητας που δεν είναι κανονικοποιημένη. Ο συντελεστής κανονικοποίησης διαδραματίζει ένα σημαντικό ρόλο περιορίζοντας την ενεργειακή συνάρτηση για τη σωστή εκτίμηση της κατανομής πιθανότητας. Παρόλα αυτά, εισάγει δυσκολίες κατά τη διάρκεια της διαδικασίας εκπαίδευσης, η οποία απαιτεί κατάλληλο αριθμό δειγμάτων και επιβραδύνει την πρόοδο της εκπαίδευσης είτε απαιτεί ορισμένες δομές μοντέλου για να πάρει τα δείγματα. Για να ξεπεραστούν αυτοί οι περιορισμοί, παρουσιάζουμε στην ακόλουθη ενότητα την προτεινόμενη αρχιτεκτονική του νευρωνικού μας δικτύου.

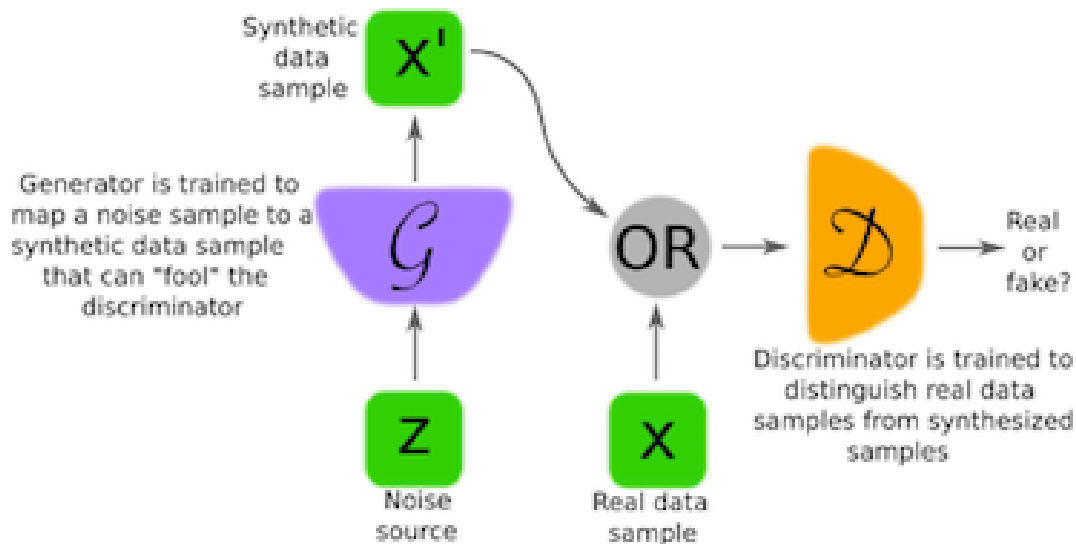
7.3.1 Διατύπωση Προβλήματος

Προτείνουμε ένα μοντέλο του οποίου ο διευκρινιστής θεωρείται ως μια συνάρτηση ενέργειας που αποδίδει χαμηλά ποσοστά ενέργειας προς τις περιοχές κοντά στα πολλα-

πλά δεδομένα και τις υψηλότερες ενέργειες σε άλλες περιοχές, ενώ η γεννήτρια θεωρούμε ότι έχει εκπαιδευτεί να παράγει αντίθετα (contrastive) δείγματα με ελάχιστες ενέργειες. Με τον όρο αντίθετα (contrastive) δείγματα, αναφερόμαστε σε ένα σημείο δεδομένων το οποίο προκαλεί μια έλξη ενέργειας, όπως για παράδειγμα μία λανθασμένη επισημείωση στην επιβλεπόμενη μάθηση και το οποίο υποδεικνύει από περιοχές χαμηλής πυκνότητας δεδομένων σε μία μη επιβλεπόμενη μάθηση. Στόχος μας είναι να εκπαιδεύσουμε τον διευκρινιστή να εκχωρήσει σε υψηλές ενέργειες σε αυτά τα παραγόμενα δείγματα. Βλέποντας τον διευκρινιστή ως μία συνάρτηση ενέργειας, το προτεινόμενο μας σύστημα, μας επιτρέπει να χρησιμοποιήσουμε μία μεγάλη ποικιλία αρχιτεκτονικών και συναρτήσεων απώλειας-εκτός από τον συνηθισμένο δυαδικό ταξινομητή με λογικής απόδοσης (logistic output) - που έχουν εισαχθεί στο προσφάτως προταθέν μοντέλο “Energy-based ανταγωνιστικό παραγωγικό δίκτυο” (EBGAN) [270]. Ωστόσο, το μοντέλο που προτείνουμε είναι μια τροποποιημένη έκδοση του EBGAN μοντέλου καθώς προτείνουμε τη χρήση μιας αρχιτεκτονικής autoencoder, με την ενέργεια να είναι το σφάλμα ανοικοδόμησης, στη θέση του διευκρινιστή. Συγκεκριμένα, προτείνουμε αντί να χρησιμοποιούμε ένα μόνο κομμάτι πληροφοριών στόχου για να εκπαιδεύσουμε το μοντέλο, να χρησιμοποιήσουμε τους διαφορετικούς στόχους που προσφέρονται από το αποτέλεσμα της διαδικασίας της ανακατασκευής (reconstruction-based output) για το διευκρινιστή. Με τη δυαδική logistic loss, μόνο δύο στόχοι είναι δυνατοί, έτσι, μέσα σε μια μίνι-παρτίδα (mini-batch), οι κλίσεις (gradients) που αντιστοιχούν σε διαφορετικά δείγματα είναι πιθανότατα πολύ μακριά από ορθογώνια. Αυτό οδηγεί σε ανεπαρκή εκπαίδευση και μείωση των μεγεθών της μίνι-παρτίδας (mini-batch), η οποία συχνά δεν αποτελεί επιλογή για το hardware. Από την άλλη πλευρά, η ανασυγκρότηση της απώλειας (reconstruction loss) πιθανόν να παράγει πολύ διαφορετικές κατευθύνσεις κλίσης μέσα στη μίνι-παρτίδα, επιτρέποντας μεγαλύτερο μέγεθος μίνι-παρτίδας χωρίς απώλεια απόδοσης. Επιπλέον, αποφασίσαμε να χρησιμοποιήσουμε τους αυτόματους κωδικοποιητές δεδομένου ότι παραδοσιακά έχουν χρησιμοποιηθεί για να αντιπροσωπεύουν μοντέλα με βάση την ενέργεια. Όταν εκπαιδεύονται με regularization terms, οι αυτόματοι κωδικοποιητές έχουν τη δυνατότητα να μάθουν μια πολλαπλή ενέργεια χωρίς επίβλεψη ή αρνητικά παραδείγματα. Αυτό σημαίνει ότι ακόμα και όταν πρόκειται για energy-based μοντέλο αυτόματης κωδικοποίησης εκπαιδευμένο για να ανοικοδομήσει ένα πραγματικό δείγμα, ο διευκρινιστής συμβάλλει στην ανακάλυψη πολλαπλών δεδομένων από το ίδιο το δίκτυο. Αντίθετα, χωρίς την παρουσία αρνητικών παραδειγμάτων από τη γεννήτρια, η λειτουργία του διευκρινιστή ο οποίος εκπαιδεύεται με δυαδική λογική απώλειας κρίνεται άσκοπη.

Το προτεινόμενο μοντέλο μας εμπνέεται επίσης και από το έργο των Kim και του Bengio [114]. Ωστόσο, διαφέρει με τον ακόλουθο τρόπο. Η προσέγγισή τους χρησιμοποιεί ένα τυποποιημένο πιθανοτικό GAN σε ένα μοντέλο ενέργειας (χρησιμοποιώντας κατανομές Gibbs), επιτρέποντας την εκπαίδευση του διευκρινιστή να μοντελοποιήσει την κατανομή όταν η ισορροπία Nash επιτυγχάνεται, αλλά εξακολουθεί να έχει προβλήματα ανάμειξης (mixing problems), ειδικά με βαθιά μοντέλα που επιβραδύνουν τη εκπαίδευση.

Αντίθετα, η προσέγγισή μας είναι απαλλαγμένη από το πιθανοτικό πλαίσιο, ενώ παρουσιάζει την ίδια ισορροπία Nash ως ένα παραδοσιακό μοντέλο GAN, αλλά μέσω μιας διαφορετικής και περισσότερο γενικευμένης κατηγορίας των συναρτήσεων απώλειας (loss functions).



Σχήμα 7.1: Η αρχιτεκτονική του μοντέλου GAN, το οποίο αποτελείται από ένα διευκρινιστή D και μία γεννήτρια G , κατά τη διάρκεια εκπαίδευσης του μοντέλου [241].

Η υπολογιστική διαδικασία και η δομή του προτεινόμενου μοντέλου GAN απεικονίζονται στο Σχήμα 7.1. Ειδικότερα, χρησιμοποιούμε τις συναρτήσεις διαφοροποίησης D και G για να αντιπροσωπεύσουμε τον διευκρινιστή και τη γεννήτρια, αντίστοιχα. Οι εισροές τους είναι πραγματικές τιμές x και τυχαίες μεταβλητές z , αντίστοιχα. Το $G(z)$ αντιστοιχεί στο παραγόμενο δείγμα από το G σύμφωνα με τα δεδομένα κατανομής p_{data} των πραγματικών δεδομένων και το x αντιστοιχεί στα νέα συνθετικά δείγματα. Εάν η είσοδος του διευκρινιστή D προέρχεται από τα πραγματικά δεδομένα x , ο D πρέπει να το ταξινομήσει ως αληθινό. Εάν η είσοδος προέρχεται από το $G(z)$, το D πρέπει να το ταξινομήσει ως ψευδές. Ο σκοπός του D είναι να επιτευχθεί η σωστή ταξινόμηση της πηγής δεδομένων, ενώ ο σκοπός του G είναι να προσεγγίσει τις επιδόσεις του (δηλ. $D(G(z))$) με την απόδοση πραγματικών δεδομένων x επί D (δηλ. $D(x)$). Η διαδικασία βελτιστοποίησης της ανταγωνιστικότητας βελτιώνει σταδιακά την απόδοση των D και G . Τελικά, όταν η ικανότητα διάκρισης της D έχει βελτιωθεί σε υψηλό επίπεδο, αλλά δεν μπορεί να διακρίνει τα δεδομένα της πηγής σωστά, τότε θεωρείται ότι η γεννήτρια G έχει καταγράψει την κατανομή των πραγματικών δεδομένων.

7.3.2 Ανταγωνιστική αρχιτεκτονική νευρωνικού δικτύου σε επίπεδο κειμένου

Το αρχικό GAN μοντέλο [83], που παρουσιάζεται στο Σχήμα 7.1, αποτελείται από ένα min-max ανταγωνιστικό παιχνίδι μεταξύ ενός μοντέλου γεννήτριας G και ενός μοντέλου D διευκρινιστή. Το $G(z)$ είναι ένα νευρωνικό δίκτυο που εκπαιδεύεται για να χαρτογραφήσει τα δείγματα z από μια προηγούμενη κατανομή θορύβου $p(z)$ στο χώρο δεδομένων. Το $D(x)$ είναι ένα δεύτερο νευρωνικό δίκτυο που λαμβάνει ένα δείγμα δεδομένων x ως είσοδο και εξάγει μία μοναδική κλίμακα που αντιπροσωπεύει την πιθανότητα ότι το x προέρχεται από την κατανομή δεδομένων αντί του $G(z)$. Το D είναι εκπαιδευμένο για να μεγιστοποιήσει την πιθανότητα εκχώρησης της σωστής ετικέτας στην είσοδο x , ενώ το μοντέλο G είναι εκπαιδευμένο για να “μπερδεύει” το D , χρησιμοποιώντας το gradient του $D(x)$ αναφορικά με το x για την ενημέρωση των παραμέτρων του. Αυτό το παιχνίδι min-max μπορεί να βελτιστοποιηθεί από τον ακόλουθο κίνδυνο, που δίνεται από την Εξίσωση 7.1, και συνήθως υλοποιείται με μοντέλα νευρωνικών δικτύων; ωστόσο, αυτά τα μοντέλα θα μπορούσαν να υλοποιηθούν με οποιαδήποτε μορφή διαφοροποιήσιμου συστήματος που χαρτογραφεί δεδομένα από ένα χώρο σε άλλο.

$$\phi = \min_G \max_D E_{x \sim p(\text{data})} [\log D(x)] + E_{z \sim p(z)} [\log(1 - D(G(z)))] \quad (7.1)$$

Ένα μειονέκτημα αυτού του μοντέλου είναι ότι δεν υπάρχουν σαφείς τρόποι για συμπεράσματα, και έτσι είναι δεν είναι σαφές πώς θα μπορούσαν να εφαρμοστούν τα μοντέλα GANs σε μη επιβλεπόμενη αναπαράσταση μάθησης. Στην εργασία [83], δύο δυνατές λύσεις προτάθηκαν και διερευνήθηκαν από την ερευνητική κοινότητα στα δύο ακόλουθα έργα. Η πρώτη προσέγγιση είναι να εκπαιδεύσουμε ένα άλλο δίκτυο για να παράγουμε συμπεράσματα, να μάθουμε δηλαδή μια χαρτογράφηση από το x πάλι στο z [57], με μια παραλλαγή αυτής της μεθόδου αντί να χρησιμοποιήσουμε την διαδικασία της ανταγωνιστικής διαδικασίας εκμάθησης για να νομιμοποιήσουμε (regularize) το επίπεδο αναπαράστασης του autoencoder [137]. Η δεύτερη ιδέα είναι να χρησιμοποιήσουμε τις εσωτερικές συνιστώσες του ίδιου του δικτύου ως αναπαράσταση [181]. Παρόλα αυτά, και οι δύο προσεγγίσεις δεν καταλήγουν σε μία αρχιτεκτονική με σταθερή εκπαίδευση με μία σειρά συνόλων δεδομένων αλλά μοντελοποιούν τις υπερπαραμέτρους όταν χρησιμοποιείται ένα πιθανοτικό μοντέλο ως διευκρινιστής. Έτσι, η προσέγγισή μας να βελτιώσουμε την εκπαίδευση του GAN έγκειται στο να αξιολογήσουμε τα εμπειρικά συμπτώματα που παρουσιάζονται κατά τη διάρκεια της εκπαίδευσης [181] μεταβαίνοντας στη χρήση του energy-based μοντέλου GAN, όπου ο διευκρινιστής είναι ένας αυτόματος κωδικοποιητής [270]. Οι αναπαραστάσεις κειμένων μπορούν στη συνέχεια να σχηματιστούν από την κωδικοποιημένη αναπαράσταση του διευκρινιστή.

Συγκεκριμένα, η αρχιτεκτονική του προτεινόμενου μοντέλου μας διατυπώνεται ως εξής: Έστω $x_i \in \{0, 1\}^V$ είναι η δυαδική αναπαράσταση του μοντέλου bag-of-words ενός

κειμένου, όπου V είναι το μέγεθος λεξιλογίου και το x_i είναι η δυαδική τιμή η οποία υποδεικνύει κατά πόσο η λέξη i^{th} είναι παρούσα στο κείμενο ή όχι. Ορίζουμε μια ροή προς τα εμπρός του δικτύου της γεννήτριας $G(z)$ που παίρνει ένα διάνυσμα $z \in R^{h_g}$ ως είσοδο και παράγει ένα διάνυσμα $\hat{x} \in R^V$, με h_g να είναι ο αριθμός των διαστάσεων του διανύσματος θορύβου εισόδου (δειγματοληψία από $N(0, 1)$). Ορίζουμε επίσης ένα μοντέλο διευκρινιστή $D(x)$, το οποίο θεωρούμε ότι είναι η συνάρτηση ενέργειας, που λαμβάνει διανύσματα $x \in R^V$ και παράγει μία εκτίμηση ενέργειας $E \in R$.

Μια κύρια διαφορά αναφορικά με το έργο του [270] είναι ότι χρησιμοποιήσαμε ένα autoencoder denoising (DAE) ως συνάρτηση ενέργειας, καθώς το τελευταίο έχει βρεθεί ότι παράγει ανώτερες αναπαραστάσεις συγκριτικά με τη βασική μορφή ενός αυτόματου κωδικοποιητή [239]. Σε αυτή τη δουλειά, χρησιμοποιούμε επίπεδα κωδικοποίησης και αποκωδικοποίησης. Έτσι, η διαδικασία κωδικοποίησης παρουσιάζεται στην Εξίσωση 7.2:

$$h = f(W^e x^c + b_e) \quad (7.2)$$

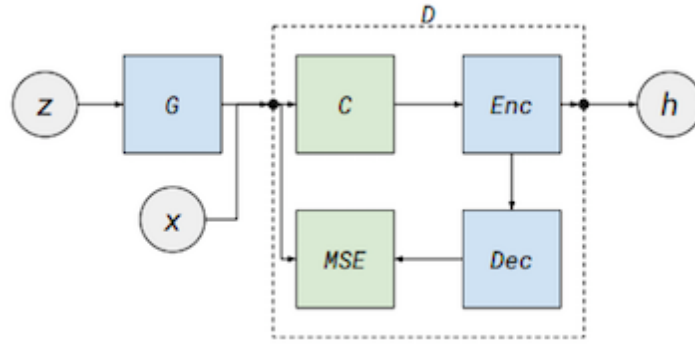
όπου το W^e είναι ένα σύνολο παραμέτρων εκμάθησης που αναφέρονται σε μια ντετερμινιστική χαρτογράφηση/αντιστοίχιση από ένα χώρο δεδομένων στο χώρο αναπαράστασης, b_e είναι ένας μαθησιακός όρος (learned bias term), f είναι μια μη γραμμική συνάρτηση, x^c είναι μια κατεστραμμένη έκδοση του x , και $h \in R^{h_d}$ είναι η κρυφή αναπαράσταση του μεγέθους h_d , με το h_d ισούται με το μέγεθος της κρυφής κατάστασης του autoencoder denoising (DAE). Στη συνέχεια, η διαδικασία αποκωδικοποίησης δίνεται από την ακόλουθη εξίσωση:

$$y = W^d h + b_d \quad (7.3)$$

όπου W^d και b_d είναι ένα άλλο μαθησιακό σύνολο βαρών και όρων μεροληψίας. Η τελική τιμή της ενέργειας είναι ο μέσος όρος τετραγώνου του σφάλματος ανοικοδόμησης:

$$E = \frac{1}{V} \sum_{i=1}^V (x_i - y_i)^2 \quad (7.4)$$

Η αρχιτεκτονική του προτεινόμενου μοντέλου GAN παρουσιάζεται στο Σχήμα 7.2. Η συνάρτηση της ενέργειας είναι εκπαιδευμένη να ωθήσει την ενέργεια των πραγματικών δειγμάτων x και να ωθήσει την ενέργεια των παραγόμενων δειγμάτων \hat{x} [270]. Αυτό δίνεται από την Εξίσωση 7.5, όπου η συνάρτηση f_D είναι η τιμή που πρέπει να ελαχιστοποιηθεί σε κάθε επανάληψη και m είναι ένα περιθώριο μεταξύ των θετικών και αρνητικών ενεργειών. Με άλλα λόγια, η συνάρτηση ενέργειας εξάγει χαμηλές τιμές για τα πολλαπλά δεδομένα και υψηλότερες τιμές οπουδήποτε αλλού.



Σχήμα 7.2: Η αρχιτεκτονική του προτεινόμενου δικτύου GAN για την εκμάθηση της αναπαράστασης σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών αντίστοιχα. Το δίκτυο αποτελείται από τα ακόλουθα τμήματα: η γεννήτρια G , ο DAE κωδικοποιητής Enc και ο αποκωδικοποιητής Dec αντίστοιχα, η διαδικασία της αλλοίωσης (corruption process) C and το δίκτυο του διευκρινιστή D [241].

$$f_D(x, z) = D(x) + \max(0, m - D(G(z))) \quad (7.5)$$

Σε κάθε επανάληψη, η γεννήτρια G εκπαιδεύεται ανταγωνιστικά έναντι του D για να ελαχιστοποιήσει το f_G . Με άλλα λόγια, η γεννήτρια G μαθαίνει να επιλέγει σημεία όπου η ενέργεια πρέπει να αυξηθεί, ενώ ο διευκρινιστής D ορίζεται ως μία αντικειμενική συνάρτηση που έχει εκπαιδευτεί. Ευθυγραμμισμένο με το ρόλο της γεννήτριας, το τελευταίο μοντέλο εκπαιδεύεται για να δημιουργήσει δείγματα που θα ξεγελάσουν το διευκρινιστή, έτσι ώστε το ανταγωνιστικό παιχνίδι που παίζεται μεταξύ των δύο δικτύων να συγκλίνει σε ένα σημείο που είναι τοπικό ελάχιστο για το διευκρινιστή και τοπικό μέγιστο για το μοντέλο της γεννήτριας.

$$f_G(z) = D(G(z)) \quad (7.6)$$

Με παρόμοιο τρόπο, μπορούν να σχηματιστούν και οι αναπαραστάσεις σε επίπεδο προτάσεων και χαρακτηριστικών από την κωδικοποιημένη εκπροσώπηση του διευκρινιστή και η δομή των δικτύων τους είναι κατασκευασμένη κατά τρόπο παρόμοιο με αυτό της δομής του ανταγωνιστικού κειμένου. Συγκεκριμένα, η λειτουργία της ενέργειας αναφορικά με την αναπαράσταση σε επίπεδο πρότασης και χαρακτηριστικών στοχεύει στη μείωση της ενέργειας των πραγματικών δειγμάτων x_{sent} και x_{asp} και στην ώθηση της ενέργειας των παραγόμενων δειγμάτων \hat{x}_{sent} και \hat{x}_{asp} , αντίστοιχα [270]. Αυτό δίνεται από την εξίσωση 7.7 και την εξίσωση 7.8 αντίστοιχα, όπου $f_{D_{sent}}$ και $f_{D_{asp}}$ είναι οι τιμές που πρέπει να ελαχιστοποιούνται σε κάθε επανάληψη και m είναι ένα περιθώριο μεταξύ θετικών και αρνητικών ενεργειών. Έτσι, η ενεργειακή συνάρτηση εξάγει χαμηλές τιμές στα πολλαπλά δεδομένα και υψηλότερες τιμές οπουδήποτε αλλού.

$$f_{D_{sent}}(x_{sent}, z_{sent}) = D(x_{sent}) + \max(0, m - D_{sent}(G(z)_{sent})) \quad (7.7)$$

$$f_{D_{asp}}(x_{asp}, z_{asp}) = D(x_{asp}) + \max(0, m - D_{asp}(G(z)_{asp})) \quad (7.8)$$

Τέλος, όπως εξηγούμε παραπάνω, σε κάθε επανάληψη, όσον αφορά στο ανταγωνιστικό παιχνίδι min-max μεταξύ της γεννήτριας και του διευκρινιστή για τις αναπαραστάσεις σε επίπεδο πρότασης και χαρακτηριστικών, οι γεννήτριες G_{sent} και G_{asp} εκπαιδεύονται ανταγωνιστικά ενάντι του D_{sent} και του D_{asp} για να ελαχιστοποιήσουν τις συναρτήσεις $f_{G_{sent}}$ και $f_{G_{asp}}$, αντίστοιχα. Με άλλα λόγια, οι γεννήτριες G_{sent} και G_{asp} μαθαίνουν να επιλέγουν σημεία όπου η ενέργεια έπρεπε να αυξηθεί, ενώ οι διευκρινιστές D_{sent} και η D_{asp} θεωρούνται ως αντικειμενικές συναρτήσεις που έχουν εκπαιδευτεί και απεικονίζονται στις Εξισώσεις 7.9 και 7.10 αντίστοιχα.

$$f_{G_{sent}(z_{sent})} = D_{sent}(G_{sent}(z_{sent})) \quad (7.9)$$

$$f_{G_{asp}(z_{asp})} = D_{asp}(G_{asp}(z_{asp})) \quad (7.10)$$

7.4 Πειραματική Επικύρωση

Αυτή η ενότητα περιγράφει τα σύνολα δεδομένων σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών (ενότητα 7.4.1), τις αποφάσεις σχεδιασμού για την υλοποίηση του προτεινόμενου μοντέλου και την επιλογή των παραμέτρων της προτεινόμενης GAN αρχιτεκτονικής (ενότητα 7.4.3) αναφορικά με την πειραματική αξιολόγηση και, τέλος, τα πειραματικά αποτελέσματα (ενότητα 7.4.4).

7.4.1 Σύνολα Δεδομένων

Η επιλογή των κατάλληλων συνόλων δεδομένων κρίνεται σημαντική για την αξιολόγηση των μοντέλων Φυσικής Γλώσσας. Υπάρχουν πολλά ελεύθερα διαθέσιμα σύνολα δεδομένων που έχουν χρησιμοποιηθεί εκτενώς στον ακαδημαϊκό χώρο και προέρχονται από τον υποτομέα της Ανάλυσης του Συναισθήματος και Γνώμης (Sentiment Analysis) του τομέα της Φυσικής Γλώσσας. Η χρήση τέτοιων συνόλων δεδομένων τροφοδοτείται από μια άνευ προηγουμένου δραστηριότητα των κοινωνικών δικτύων την τελευταία δεκαετία αλλά και από το ενδιαφέρον για την επεξεργασία των κοινωνικών μέσων που έχει ενισχυθεί με τη χρήση συναισθήματος. Μεταξύ αυτών των συνόλων δεδομένων εντάσσονται αυτά που παρέχονται από το Πανεπιστήμιο του Στάνφορντ, το SST1 και το SST2, το σύνολο δεδομένων Large Movie Review, το MPQA σώμα γνώμης [35], ένα σύνολο δεδομένων αξιολογήσεων από το Amazon [59], το ACL ανθολόγιο και το σώμα κειμένου

20Newsgroups [122]. Παρακάτω, δίνουμε μια λεπτομερή περιγραφή των τριών σωμάτων κειμένου τα οποία χρησιμοποιούμε για την πειραματική μας ανάλυση:

(1) Το σύνολο δεδομένων 20Newsgroups ²: Το σύνολο δεδομένων 20Newsgroups είναι μια συλλογή περίπου 20.000 εγγράφων ομάδας συζήτησης. Πρόκειται για ένα δημοφιλές σύνολο δεδομένων σε πειράματα εφαρμογών κειμένου με χρήση τεχνικών Μηχανικής Μάθησης, όπως π.χ. η ταξινόμηση και η συσταδοποίηση κειμένου. Τα δεδομένα είναι οργανωμένα σε 20 διαφορετικές ομάδες συζήτησης, εκ των οποίων καθεμία αντιστοιχεί σε διαφορετικό θέμα. Ορισμένες από τις ομάδες συζήτησης σχετίζονται πολύ στενά με άλλες ομάδες (π.χ., comp.sys.ibm.pc.hardware/comp.sys.mac.hardware), ενώ άλλες δεν έχουν μεγάλη σχέση μεταξύ τους (π.χ., misc.forsale/soc.religion.christian). Ο διαχωρισμός μεταξύ των δεδομένων εκπαίδευσης (training) (50%), επικύρωσης (validation) (10%) και του σετ δοκιμών (test) (40%) βασίζεται σε μηνύματα που έχουν αναρτηθεί πριν και μετά από μια συγκεκριμένη ημερομηνία, ενώ τα διπλότυπα και οι επικεφαλίδες ταυτοποίησης ομάδων συζήτησης (Xref, ομάδες συζήτησης, διαδρομή, παρακολούθηση, ημερομηνία) έχουν αφαιρεθεί. Η έκδοση 20news-bydate.tar.gz του συνόλου δεδομένων 20Newsgroups αποτελείται από 18.786 κείμενα (αναρτήσεις σχολίων); ωστόσο, μετά από επιπρόσθετη προεπεξεργασία (δηλ. εξαιρώντας έναν αριθμό αντιγράφων και καταργώντας ορισμένες επικεφαλίδες) ο τελικός αριθμός των κειμένων είναι 18.821. Διαχωρίσαμε περαιτέρω τα δεδομένα σε 10.163 ως δεδομένα εκπαίδευσης, 1.130 ως δεδομένα επικύρωσης και 7.528 ως δεδομένα δοκιμών αντίστοιχα.

(2) Σύνολο δεδομένων αναθεώρησης ταινιών (σύνολο δεδομένων υποκειμενικότητας v1.0) ³: Αυτό το σύνολο δεδομένων περιλαμβάνει 5000 υποκειμενικές και 5000 αντικειμενικές επεξεργασμένες προτάσεις. Όσον αφορά στο επίπεδο ανάλυσης, κάθε γραμμή σε αυτά τα δύο αρχεία (υποκειμενικές και αντικειμενικές προτάσεις) αντιστοιχεί σε μία μόνο πρόταση ή απόσπασμα; σε όλες τις προτάσεις (ή τα αποσπάσματα) τα κεφαλαία γράμματα έχουν αντικατασταθεί με μικρά. Επίσης, συμπεριλάβαμε μόνο προτάσεις ή αποσπάσματα που περιείχαν τουλάχιστον 10 tokens. Οι προτάσεις και τα αποσπάσματα έχουν επισημειωθεί αυτόματα. Θεωρούμε ότι, η χρήση του συνόλου δεδομένων Movie Review που παρέχεται ελεύθερα από τον Pang και τον Lee είναι επαρκής, καθώς πολλά άρθρα της ανάλυσης γνώμης χρησιμοποιούν αυτό το σύνολο δεδομένων για την επικύρωση μιας σειράς μεθόδων και προσεγγίσεων, καθιστώντας το ιδανικό υποψήφιο κείμενο για αξιόπιστη συγκριτική αξιολόγηση. Για τις ανάγκες της προτεινόμενης πειραματικής διαδικασίας, χωρίσαμε τυχαία κάθε συλλογή σε δεδομένα εκπαίδευσης και δοκιμών 75% και 25%, αντίστοιχα. Σε αυτό το σύνολο δεδομένων, το σύνολο επικύρωσης αναφέρεται στο 10% του συνόλου εκπαίδευσης.

(3) Σύνολο δεδομένων Finegrained Sentiment ⁴: Το σύνολο δεδομένων Finegrained Sentiment περιέχει 294 κριτικές προϊόντων από διάφορες online πηγές τα οποία επισημαίνονται χειρονακτικά με τα αντίστοιχα συναίσθημα σε επίπεδο προτάσεων. Τα δεδομένα είναι σχεδόν ισοκαταναμημένα αναφορικά με τον τομέα (βιβλία, DVD, ηλεκτρονικά, μουσική, βιντεοπαιχνίδια) και το συνολικό συναίσθημα αξιολόγησης (θετικό, αρ-

Datasets	Training	Validation	Test
20Newsgroup ⁵	10.163	1130	7528
Movie Reviews [163]	7424	76	2500
FSD [228]	2582	287	956

Πίνακας 7.2: Συλλογές δεδομένων δοκιμών για τη διερεύνηση της χρήσης των Ανταγωνιστικά Παραγωγικών Δικτύων αναφορικά με την εκμάθηση καταναμημένων αναπαραστάσεων της Φυσικής Γλώσσας σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών αντίστοιχα. [241]

νητικό, ουδέτερο) και οι ετικέτες συναισθήματος αποδίδονται σε προτάσεις μετά από το σχολιασμό δύο ειδικών. Όσον αφορά στις κριτικές προϊόντων του συνόλου δεδομένων Finegrained, οι ετικέτες γνώμης αναφέρονται σε θετικές και αρνητικές απόψεις σχετικά με τα διάφορα χαρακτηριστικά του προϊόντος, παρόλο που το γενικό συναίσθημα για το προϊόν μπορεί να είναι θετικό ή αρνητικό. Η συλλογή FSD περιλαμβάνει συνολικά 2243 προτάσεις πολικότητας: 923 θετικές και 1320 αρνητικές προτάσεις αντίστοιχα. Για μία ακόμη φορά, χωρίσαμε τυχαία κάθε συλλογή σε δεδομένα εκπαίδευσης και σε δεδομένα δοκιμών με ποσοστά 75% και 25%, αντίστοιχα. Σε αυτό το σύνολο δεδομένων, το σύνολο επικύρωσης αφορά στο 10% του συνόλου εκπαίδευσης. Η πειραματική διάταξη για τις τρεις συλλογές δεδομένων παρουσιάζεται στον Πίνακα 7.2.

7.4.2 Συστήματα Αναφοράς

Δημιουργήσαμε ένα απλό μοντέλο GAN που αποτελείται από μια γεννήτρια τριών-επιπέδων feedforward δικτύων και δίκτυα διευκρινιστών ως το μοντέλο αναφοράς μας. Οι παράμετροι των δικτύων βελτιστοποιήθηκαν χρησιμοποιώντας το σύνολο επικύρωσης. Το μοντέλο baseline υλοποιήθηκε με παρόμοια αρχιτεκτονική και χρησίμευσε ως δίκαιη σύγκριση με την προτεινόμενη μέθοδο (π.χ., αριθμός επιπέδων). Η βασική διαφορά με το μοντέλο που προτείνουμε είναι ότι ο προτεινόμενος τύπος μοντέλου για το δίκτυο του διευκρινιστή είναι ένας denoising autoencoder.

7.4.3 Υλοποίηση

Μετά την προεπεξεργασία και των τριών συνόλων δεδομένων για την αφαίρεση τυχόν δεδομένων που περιέχουν θόρυβο ώστε να μειώσουμε την πολυπλοκότητα των δεδομένων μας. Αποφασίσαμε να καταργήσουμε τα μη σχετικά χαρακτηριστικά, όπως stop words (όπως “εγώ” ή “και”); λημματοποιήσαμε (tokenized) τα σύνολα δεδομένων ώστε να χωρίσουμε τις λέξεις τους σε λέξεις (tokens); και stemmed και καταλήξαμε στο τελευταίο για να μειώσουμε τις λέξεις (tokens) σε ένα μοναδικό τύπο, συνήθως σε μία λέξη-ρίζα. Ως εκ τούτου, η διαδικασία stemming μείωσε τις περιττές λέξεις των συνόλων δεδομένων μας. Ο στόχος μας είναι να δημιουργήσουμε ένα λεξιλόγιο περίπου 40.000 διαφορετικών λέξεων.

Από αυτές, περίπου 2.000 από τις πιο δημοφιλείς διατηρήθηκαν στα σύνολα δεδομένων εκπαίδευσης. Ως αποτέλεσμα, κάθε ανάρτηση αντιπροσωπεύεται από ένα διάνυσμα που περιλαμβάνει 2.000 καταμετρημένες λέξεις.

Παρόλο που η ενσωμάτωση λέξεων (word embedding) είναι σήμερα η κορυφαία τεχνολογία στον τομέα της Φυσικής Γλώσσας αναφορικά με την επίλυση σχετικών με το κείμενο προβλημάτων, εξαιτίας μιας σειράς περιορισμών που παρουσιάσαμε, αποφασίσαμε να χρησιμοποιήσουμε το μοντέλο bag-of-words (BoW). Ειδικότερα, ένας περιορισμός των επιμέρους ενσωματώσεων λέξεων είναι η αδυναμία τους να αναπαριστούν φράσεις όπου ο συνδυασμός δύο ή περισσότερων λέξεων δεν αντιστοιχίζεται με το συνδυασμό των εννοιών των μεμονωμένων λέξεων [147]. Ένας άλλος περιορισμός προέρχεται από την εκμάθηση αυτών των ενσωματώσεων (embeddings) που βασίζονται μόνο σε ένα μικρό παράθυρο (window) γύρω από τις λέξεις: μερικές φορές λέξεις, όπως “καλό” και “κακό”, μοιράζονται σχεδόν το ίδιο σύνολο ενσωματώσεων [217], το οποίο είναι προβληματικό στην περίπτωση που χρησιμοποιείται σε διεργασίες όπως η ανάλυση γνώμης [250]. Επιπλέον, μία γενική προειδοποίηση αναφορικά με τις ενσωματωμένες λέξεις (word embeddings) είναι ότι οι τελευταίες εξαρτώνται σε μεγάλο βαθμό από τις εφαρμογές από τις οποίες χρησιμοποιούνται [120], ως εκ τούτου, το μοντέλο BoW λειτουργεί καλύτερα από τα τρέχοντα μοντέλα ενσωμάτωσης λέξεων (Word2Vec, GloVe) αναφορικά με το εξεταζόμενο σενάριο.

Εκπαιδεύουμε λοιπόν τα δίκτυά μας χρησιμοποιώντας το Keras [37] με το Tensorflow ως back-end [1]. Ακολουθώντας την ερευνητική δραστηριότητα του [99], για μια περισσότερο άμεση σύγκριση, ορίζουμε το μέγεθος αναπαράστασης h_d (το μέγεθος του autoencoder denoising (DAE) του κρυφού επιπέδου/στρώματος) έως 50. Το διάνυσμα εισόδου του θορύβου της γεννήτριας h_g ορίστηκε επίσης να έχει το ίδιο μέγεθος. Η γεννήτρια είναι ένα δίκτυο τροφοδοσίας τριών επιπέδων/στρωμάτων, με ενεργοποιημένες γραμμικές μονάδες (rectified linear unit (ReLU)) στα πρώτα δύο στρώματα και μία μη γραμμική σιγμοειδή μονάδα (sigmoid) στο στρώμα εξόδου. Τα επίπεδα 1 και 2 είναι και τα δύο μεγέθους 300, με το τελικό στρώμα έχει το ίδιο μέγεθος με το λεξιλόγιο. Τα επίπεδα 1 και 2 χρησιμοποιούν batch normalization [197]. Αποφασίσαμε να χρησιμοποιήσουμε το τελευταίο, καθώς η χρήση του προτείνεται και για τα δύο δίκτυα για τη σταθεροποίηση της εκπαίδευσης τους σε βαθύτερα μοντέλα. Ο κωδικοποιητής δηλ. το δίκτυο του διευκρινιστή αποτελείται από ένα μόνο γραμμικό στρώμα που ακολουθείται από μία leaky ReLU μη γραμμική συνάρτηση (με διαρροή 0.02). Αποφασίζουμε να χρησιμοποιήσουμε τη συνάρτηση αυτή, καθώς έχει αποδειχθεί ότι η χρήση της συνάρτησης ReLU με διαρροή μεταξύ των ενδιάμεσων στρωμάτων του διευκρινιστή, παρέχει μεγαλύτερη απόδοση συγκριτικά με τη χρήση τακτικών ReLUs [181]. Ο αποκωδικοποιητής είναι ένας γραμμικός μετασχηματισμός πίσω στο μέγεθος του λεξιλογίου. Βελτιστοποιήσαμε επίσης τόσο το G όσο και το D χρησιμοποιώντας τον Adam βελτιστοποιητή [123] με ένα αρχικό ρυθμό εκμάθησης 0.0001. Η διαφθορά που προκαλεί ο autoencoder (DAE) ρυθμίστηκε τυχαία στο 40% των τιμών εισόδου και χρησιμοποιήσαμε ένα περιθώριο μεγέθους m του 5% του

μεγέθους του λεξιλογίου.

Τέλος, ακολουθούμε την ίδια διαδικασία επικύρωσης με τους Salakhutdinov και Hinton [99] και ορίζουμε το σύνολο της επικύρωσης που έχει οριστεί για την επιλογή μοντέλου των άλλων υπερπαραμέτρων, όπως ο ρυθμός εκμάθησης και ο ο αριθμός των μεταβάσεων που διέρχονται από το σύνολο εκπαίδευσης (με βάση την έγκαιρη διακοπή (early stopping)). Δοκιμάσαμε επίσης τη χρήση ενός κρυμμένου στρώματος μη γραμμικού hyperbolic tangent, που δίνεται στην Εξίσωση 7.11, αντί για τη σιγμοειδή μονάδα, χρησιμοποιώντας πάντα την καλύτερη επιλογή βάσει της απόδοσης που επιλέγεται από την διαδικασία της επικύρωσης. Τέλος, το προτεινόμενο μοντέλο GAN εκπαιδεύτηκε με ρυθμό εκμάθησης 0.01 και χρησιμοποιώντας τη συνάρτηση ενεργοποίησης tanh.

$$\tanh(x) = (\exp(x) - \exp(-x)) / (\exp(x) + \exp(-x)) \quad (7.11)$$

7.4.4 Πειραματικά Αποτελέσματα

Προκειμένου να διεξαχθεί μια συνολική αξιολόγηση της προτεινόμενης μεθόδου μας, χρησιμοποιούμε τρεις τύπους συνόλων δεδομένων από διαφορετικούς τομείς. Είναι επίσης σημαντικό να σημειωθεί ότι τα σύνολα δεδομένων που έχουν ήδη αξιολογηθεί, όπως περιγράφονται στην ενότητα 7.4.1, ποικίλλουν σημαντικά στο μήκος του κειμένου που έχει αναλυθεί και κυμαίνεται από 18.821 και 10.000 έως 3.825 κείμενα και προτάσεις, αντίστοιχα. Αυτή η ποικιλομορφία μας επιτρέπει να αξιολογήσουμε την ευρωστία της προτεινόμενης προσέγγισής μας σε πολλαπλές πειραματικές ρυθμίσεις. Μέχρι σήμερα, δεν υπάρχει συναίνεση όσον αφορά στην επιλογή του καλύτερου μέτρου αξιολόγησης σε σχέση με την απόδοση του μοντέλου GAN [14].

Οι διαφορετικές μετρικές αξιολογούν, τόσο ποιοτικά όσο και ποσοτικά, διάφορες πτυχές της διαδικασίας παραγωγής και είναι απίθανο μια μετρική να καλύπτει όλες τις εξεταζόμενες περιπτώσεις. Επί του παρόντος, δύο ευρέως αποδεκτές μετρικές, το Inception Score [197] και το Frechet Inception Distance, βασίζονται σε pretrained βαθιά δίκτυα για να αντιπροσωπεύουν και να συγκρίνουν στατιστικά τα αρχικά και τα παραγόμενα δείγματα. Ωστόσο, αυτές οι μετρικές έχουν εφαρμοστεί και δοκιμαστεί μόνο στον τομέα της Όρασης Υπολογιστών και όχι στον τομέα της Ανάλυσης Γνώμης. Όσον αφορά στις διεργασίες του τομέα της Ανάλυσης Γνώμης, η ακρίβεια είναι το πιο συχνά χρησιμοποιούμενο μέτρο, σύμφωνα με τα τελευταία δημοσιευμένα έργα που συγκεντρώσαμε στον Πίνακα 7.1, κυρίως όταν εξετάζονται τα CNNs και/ή τα RNNs. Έτσι, υποστηρίζουμε την αναγκαιότητα του να εξετάσουμε και άλλες μετρικές.

Λαμβάνοντας υπόψιν μας ότι η προτεινόμενη αρχιτεκτονική GAN βασίζεται σε μία παραλλαγή του energy-based GAN [270] και για να προχωρήσουμε σε μια άμεση σύγκριση με την προαναφερθείσα εργασία, αποφασίσαμε να χρησιμοποιήσουμε τη μέτρηση των καμπύλων ακριβείας/ανάκλησης για να αξιολογήσουμε την απόδοση του προτεινόμενου μοντέλου που χρησιμοποιούμε στη μελέτη μας. Η ακρίβεια ορίζεται ως το ποσοστό

των σχετικών στοιχείων από τα στοιχεία “top X” που ανακτώνται από τον αλγόριθμό μας, ενώ η ανάκληση είναι το ποσοστό των ανακτηθέντων σχετικών στοιχείων όλων των σχετικών στοιχείων του συνόλου δεδομένων [49].

Για λόγους πληρότητας, αποφασίσαμε να χρησιμοποιήσουμε τη μετρική F-measure για να αξιολογήσουμε τις επιδόσεις του προτεινόμενου συστήματός μας. Η F-measure επιτρέπει τον προσδιορισμό τόσο της ακρίβειας όσο και της ανάκλησης σε μία μόνο τιμή [218]. Η μετρική αυτή υπολογίζεται χρησιμοποιώντας τον ακόλουθο τύπο:

$$F - measure = \frac{(1 + \beta^2) \times recall \times precision}{(\beta^2 \times precision) + recall} \quad (7.12)$$

Στα πειράματά μας χρησιμοποιούμε $b = 1$, δίνοντας έτσι το ίδιο βάρος στα δύο μέτρα της ακρίβειας και της ανάκλησης.

Παρόλο που η ακρίβεια είναι το πιο συνηθισμένο μέτρο για τις εφαρμογές της Ανάλυσης Γνώμης, ένας επιπρόσθετος λόγος που συνηγορεί στη χρήση των παραπάνω μέτρων είναι το ζήτημα της ανισοκατανομής των δεδομένων [237, 195, 33]. Για να είμαστε ακριβέστεροι, σε πολλές περιπτώσεις, τα σύνολα δεδομένων Ανάλυσης Γνώμης υποφέρουν από ανισοκατανομές στις κλάσεις, όπου ο αριθμός των περιπτώσεων που ανήκουν σε μια κλάση ξεπερνά σημαντικά τον αριθμό που ανήκει σε μία άλλη κλάση. Όσον αφορά στα σύνολα δεδομένων που χρησιμοποιήσαμε για να αξιολογήσουμε το σύστημά μας, το σύνολο δεδομένων 20Newsgroups και το Finegrained Sentiment σύνολο δεδομένων περιλαμβάνουν σημαντικές ανισοκατανομές μεταξύ των κλάσεων που περιλαμβάνουν.

Ως εκ τούτου, το πλεονέκτημα του επιλεγμένου μέτρου αξιολόγησης είναι ότι επιτρέπει την αξιολόγηση της απόδοσης των επιδόσεων στις κορυφαίες περιπτώσεις X που είναι σχετικές, ταξινομημένες ανάλογα με την πιθανότητα να ανήκουν σε μία συγκεκριμένη κατηγορία σχετικών περιπτώσεων. Αυτό το μέτρο χρησιμοποιείται συχνά στην ανάκτηση πληροφοριών κατά τη μέτρηση της απόδοσης ενός ερωτήματος: καθώς ο χρήστης δεν είναι πιθανό να εξετάσει χιλιάδες έγγραφα/προτάσεις όσον αφορά ένα χαρακτηριστικό (aspect), τα κορυφαία κείμενα/προτάσεις είναι πιο σημαντικά και συνεπώς λαμβάνουν μεγαλύτερη βαρύτητα. Αυτό το μέτρο μας επιτρέπει να αναλύσουμε την απόδοση του αλγορίθμου που έχουμε αξιολογήσει από πολλαπλές οπτικές. Τέλος, είναι επίσης ιδιαίτερα χρήσιμο για ρεαλιστικές εφαρμογές, όπου συχνά μόνο τα κορυφαία στοιχεία X χρησιμοποιούνται κατά τη διαδικασία αξιολόγησης.

Ο Πίνακας 7.3 παρουσιάζει την απόδοση όσον αφορά στις μετρήσεις ακρίβειας, ανάκλησης και F-measure για τα μοντέλα που εκπαιδεύονται με γεννήτρια τροφοδοσίας τριών επιπέδων και με δίκτυα διευκρινιστών που αποτελούνται από ένα ενιαίο γραμμικό στρώμα (κωδικοποιητής) και ένα γραμμικό μετασχηματισμό πίσω στο μέγεθος του λεξιλογίου (αποκωδικοποιητής) αντίστοιχα. Παρατηρούμε λοιπόν, ότι το σύνολο δεδομένων Movie Reviews σε επίπεδο προτάσεων λειτουργεί καλύτερα, ενώ το σύνολο δεδομένων 20Newsgroups παρέχει αποτελέσματα ελαφρώς καλύτερης ακρίβειας σε σύγκριση με το σύνολο δεδομένων Finegrained.

Dataset	Baseline			Proposed GAN		
	Precision	Recall	F-measure	Precision	Recall	F-measure
20Newsgroups	0.2521	0.0001	$0.996 * 10^{-4}$	0.4188	0.0001	$1.999 * 10^{-4}$
	0.2099	0.0002	$3.996 * 10^{-4}$	0.4012	0.0002	$3.996 * 10^{-4}$
	0.1005	0.0005	$9 * 10^{-4}$	0.3648	0.0005	$9.986 * 10^{-4}$
Movie Reviews	0.3637	0.0001	$1.999 * 10^{-4}$	0.6376	0.0001	$1.999 * 10^{-4}$
	0.3637	0.0002	$3.978 * 10^{-4}$	0.6376	0.0002	$3.998 * 10^{-4}$
	0.3901	0.0005	$9.871 * 10^{-4}$	0.6202	0.0005	$9.991 * 10^{-4}$
Fine-Grained Dataset (FSD)	0.1022	0.0001	$2.101 * 10^{-4}$	0.3522	0.0001	$2 * 10^{-4}$
	0.1152	0.0002	$3.993 * 10^{-4}$	0.3483	0.0002	$3.997 * 10^{-4}$
	0.3185	0.0005	$9 * 10^{-4}$	0.3483	0.0005	$9.985 * 10^{-4}$

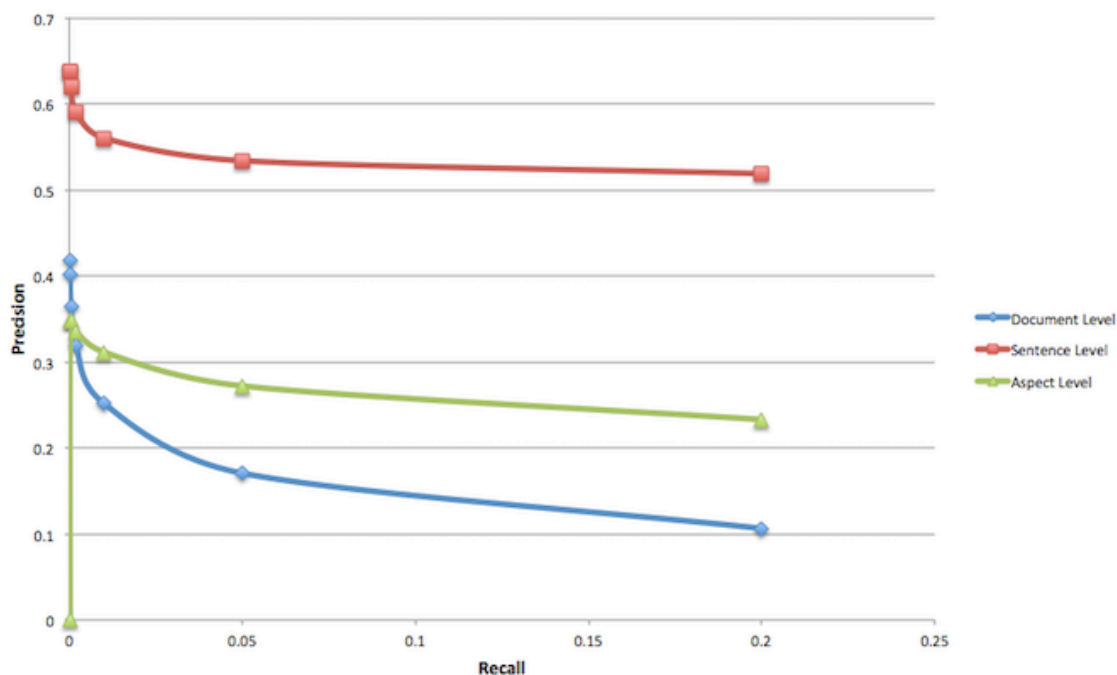
Πίνακας 7.3: Η απόδοση του προτεινόμενου μοντέλου GAN ως ένα Ανταγωνιστικά Παραγωγικό μοντέλο όταν χρησιμοποιούμε διαφορετικά είδη δεδομένων σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών. Το μοντέλο μας συγκρίνεται με ένα απλό GAN το οποίο χρησιμεύει ως σύστημα αναφοράς [241].

Για να οπτικοποιήσουμε τα παραγόμενα αποτελέσματα, δημιουργήσαμε περαιτέρω τις καμπύλες ανάκλισης-ακριβείας και αντιμετωπίσαμε την απόδοση του μοντέλου μας ως μία εργασία ανάκτησης του συνόλου δεδομένων 20Newsgroups, των δεδομένων της Movie Reviews και της FSD. Οι καμπύλες ακριβείας-ανάκλισης για τις τιμές ανάκλισης δίνονται στον Πίνακα 7.3 και στο Σχήμα 7.3.

Στα Σχήματα 7.4–7.6 παρουσιάζουμε τις απεικονίσεις των αναπαραστάσεων που δημιουργήθηκαν χρησιμοποιώντας το εργαλείο t-Distributed Stochastic Neighbor Embedding (t-SNE) [236]. Χρησιμοποιούμε αυτό το εργαλείο και δημιουργούμε τις 2D προβολές της κατανομής των δεδομένων. Συγκεκριμένα, τα σχήματα 7.4–7.6 παρουσιάζει τις κατανομές των δοκιμών σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών (aspect) τα οποία έχουν εκπαιδευτεί από το προτεινόμενο ανταγωνιστικό μοντέλο, αντίστοιχα. Τα κείμενα ανήκουν σε 20 διαφορετικές κατηγορίες θεμάτων, οι προτάσεις του συνόλου δεδομένων της Movie Reviews ανήκουν σε διαφορετικές κατηγορίες, ενώ οι προτάσεις της σειράς Finegrained ανήκουν σε 5 κατηγορίες αντίστοιχα. Για το διαχωρισμό αυτών των ξεχωριστών κατηγοριών χρησιμοποιούμε διαφορετικά έγχρωμα σημεία στα σχήματα 7.4–7.6. Με τη χρήση της ανταγωνιστικής εκπαίδευσης, οι κατανομές χαρακτηριστικών για τα δείγματα δοκιμής και από τα τρία σύνολα δεδομένων είναι σχεδόν δυσδιάκριτες, αποδεικνύοντας ότι η προτεινόμενη προσέγγιση μας μπορεί να αναπαρασταθεί με κοινό τρόπο.

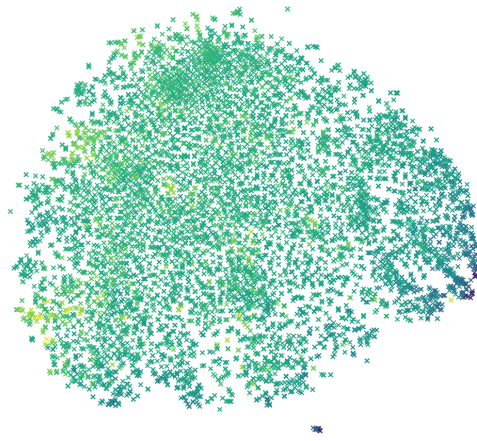
7.5 Συζήτηση

Η παρούσα ενότητα αναφέρεται στα αποτελέσματα της ανάλυσης που πραγματοποιήθηκαν σύμφωνα με τον πειραματικό σχεδιασμό. Όσον αφορά στον **αριθμό των κρυφών επιπέδων** για την αναπαραγωγή χαρακτηριστικών, αρχικά μελετήσαμε το αποτέλεσμα

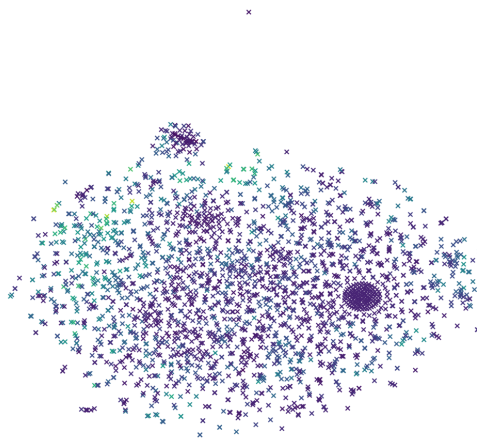


Σχήμα 7.3: Οι καμπύλες ανάκλησης-ακρίβειας του προτεινόμενου μοντέλου μας στις συλλογές δεδομένων 20Newsgroups, Movie Reviews και Finegrained.

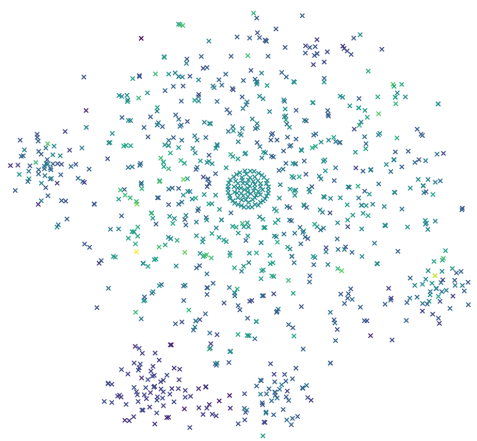
του αριθμού των κρυμμένων επιπέδων τόσο στο δίκτυο της γεννήτριας όσο και στο δίκτυο του διευκρινιστή. Αλλάξαμε τον αριθμό των κρυφών επιπέδων, που κυμαίνεται από ένα έως τέσσερα αρχικά μόνο στο δίκτυο της γεννήτριας, στη συνέχεια στο δίκτυο του διευκρινιστή, και, τέλος, και στα δύο δίκτυα ταυτόχρονα, και παρατηρήσαμε κατά πόσο οι αλλαγές στην αναπαραγωγή χαρακτηριστικών επηρεάζουν την απόδοση ταξινόμησης. Η αξιολόγηση αυτή πραγματοποιήθηκε αποκλειστικά με βάση το σύνολο επικύρωσης και για τις τρεις εξεταζόμενες συλλογές δεδομένων. Στις περισσότερες περιπτώσεις, δύο ή τρία επίπεδα παρέχουν την καλύτερη απόδοση. Παρά τη θεωρητική ύπαρξη μοναδικών λύσεων, η εκπαίδευση του δικτύου GAN είναι δύσκολη και συχνά ασταθής για διάφορους λόγους [181, 197, 7]. Μια προσέγγιση για τη βελτίωση της εκπαίδευσης του μοντέλου GAN είναι η αξιολόγηση των εμπειρικών “συμπτωμάτων” που μπορεί να αντιμετωπιστούν κατά τη διάρκεια της εκπαίδευσης. Αυτά τα συμπτώματα περιλαμβάνουν: δυσκολίες στη σύγκλιση του μοντέλου γεννήτριας-διευκρινιστή [181], την περίπτωση όπου το ανταγωνιστικό παραγωγικό μοντέλο καταρρέει κατά τη δημιουργία παρόμοιων δειγμάτων για διαφορετικές εισροές [197] και, τέλος, την απώλεια του διευκρινιστή (discriminator loss) που συγκλίνει γρήγορα στο μηδέν [7], παρέχοντας μια μη αξιόπιστη διαδρομή για αναβαθμίσεις της κλίσης (gradient updates) στη γεννήτρια. Με βάση την εκτενή πειραματική μας μελέτη, δεν παρατηρήσαμε σημαντική αύξηση της απόδοσης εξαιτίας της αύξησης του αριθμού των κρυφών επιπέδων στο προτεινόμενο δίκτυο (ούτε στη γεννήτρια ούτε στο μοντέλο του διευκρινιστή για τους τρεις διαφορετικούς τύπους συνόλων δεδομένων). Ο αρχικός μας στόχος ήταν να επιλέξουμε για κάθε τύπο συνόλου δεδομένων την ιδανική “Χωρητικότητα” κρυφών επιπέδων και για τα δύο μοντέλα του δικτύου μας. Ωστόσο, τα



Σχήμα 7.4: *t-SNE* οπτικοποίηση για την αναπαράσταση σε επίπεδο κειμένου με τη χρήση του μοντέλου *GAN*.



Σχήμα 7.5: *t-SNE* οπτικοποίηση για την αναπαράσταση σε επίπεδο πρότασης με τη χρήση του μοντέλου *GAN*.



Σχήμα 7.6: *t-SNE* οπτικοποίηση για την αναπαράσταση σε επίπεδο χαρακτηριστικών με τη χρήση του μοντέλου *GAN*.

αποτελέσματά μας είναι ευθυγραμμισμένα με το έργο της εργασίας [231], σύμφωνα με το οποίο ακόμα και αν παρέχουμε επιτυχημένες λύσεις στα παραπάνω προβλήματα κατά την εκπαίδευση των GAN δικτύων, η οικοδόμηση του τελευταίου εξαρτάται ακόμη σε μεγάλο βαθμό από την εξεταζόμενη εφαρμογή.

Για να διερευνήσουμε **τα αποτελέσματα απόδοσης** που επιτεύχθηκαν από το προτεινόμενο μοντέλο μας, χρησιμοποιήσαμε όλο το σετ δοκιμών και των τριών συνόλων δεδομένων ως ερωτήματα και τα συγκρίναμε με το κλάσμα του πλησιέστερου τύπου κειμένου του εξεταζόμενου συνόλου δεδομένων για τα τρία σύνολα εκπαίδευσης, από τα οποία υπολογίσαμε την ομοιότητα βάσει της ομοιότητας του συνημιτόνου μεταξύ των διανυσματικών τους αναπαραστάσεων. Ο μέσος αριθμός των κειμένων και των προτάσεων που επιστράφηκαν με την ίδια ετικέτα όπως το έγγραφο-ερώτημα (ακρίβεια) κατεγράφη. Αναφορικά με την επιλεγμένη μετρική αξιολόγησης, θεωρούμε ότι λόγω της ανισοκατανομής μεταξύ των κλάσεων του συνόλου δεδομένων η καμπύλη ανάκτησης-ακρίβειας αποτελεί αξιόπιστο μέτρο αξιολόγησης. Για την ανάκτηση πληροφοριών, η ακρίβεια είναι ένα μέτρο της σχετικότητας των αποτελεσμάτων, ενώ η ανάκληση είναι ένα μέτρο για το πόσα πραγματικά αποτελέσματα έχουν επιστραφεί. Έτσι, οι καμπύλες απόκρισης ακριβείας που παρουσιάζονται στο Σχήμα 7.3 δείχνουν το συμβιβασμό (trade-off) μεταξύ ακριβείας και ανάκλησης για διαφορετικά όρια-τιμές. Παρατηρήσαμε ότι επιτύχαμε υψηλές βαθμολογίες τόσο για την ακρίβεια όσο και για την ανάκληση αναφορικά με το σύνολο δεδομένων Movie Reviews (σε επίπεδο προτάσεων), που υποδηλώνει ότι η προτεινόμενη αρχιτεκτονική GAN επιστρέφει ακριβή αποτελέσματα (υψηλή ακρίβεια), καθώς και επιστροφή της πλειοψηφίας όλων των θετικών αποτελεσμάτων (υψηλή ανάκληση). Επιπλέον, παρατηρήσαμε ότι το σύστημα GAN σε επίπεδο χαρακτηριστικών για την ίδια τιμή ανάκλησης έχει 0.2% υψηλότερα αποτελέσματα ακριβείας σε σύγκριση με το σύστημα GAN σε επίπεδο εγγράφου. Εκτός από αυτό, τα αποτελέσματά μας δείχνουν ότι λαμβάνοντας υπόψη ότι το προτεινόμενο μοντέλο GAN σε επίπεδο κειμένου επιτυγχάνει υψηλή ακρίβεια αλλά χαμηλή ανάκληση, επιστρέφει δηλαδή πολύ λίγα αποτελέσματα, αλλά τα περισσότερα από τα παραγόμενα κείμενα είναι σωστά όταν συγκριθούν με αυτά της εκπαίδευσης. Τέλος, αφού συγκρίναμε την προτεινόμενη απόδοση μοντέλου με την απόδοση του συστήματος αναφοράς παρατηρήσαμε ότι σε όλες τις περιπτώσεις, και τα τρία διαφορετικά συστήματα αναφοράς για τις ίδιες τιμές ανάκλησης που δίνονται στον Πίνακα 7.3 ξεπερνά το σύστημα αναφοράς GAN. Συνολικά, το πιο ιδανικό σύστημα μεταξύ των τριών, επιτυγχάνοντας τόσο υψηλή ακρίβεια όσο και υψηλή ανάκληση ενώ επιστρέφει πολλά αποτελέσματα που παράγονται σωστά, είναι το μοντέλο GAN σε επίπεδο προτάσεων για το σύνολο δεδομένων Movie Reviews. Επιπλέον, η επιλογή μας για μία μετρική αξιολόγησης αναφορικά με την εξέταση της απόδοσης του μοντέλου μας GAN ευθυγραμμίζεται με το έργο του [231], σύμφωνα με το οποίο η προσπάθεια αξιολόγησης των GANs χρησιμοποιώντας διαφορετικά μέτρα μπορεί να οδηγήσει σε συγκρουόμενα συμπεράσματα σχετικά με την ποιότητα των παραγόμενων δειγμάτων; η απόφαση της επιλογής ενός μέτρου έναντι κάποιου άλλου εξαρτάται από την εξεταζόμενη εφαρμογή.

Διερευνήσαμε επίσης την αναπαράσταση χαρακτηριστικών κατά την εκπαίδευση του μοντέλου GAN χρησιμοποιώντας την t-SNE εργαλειοθήκη. Ο στόχος αυτής της αξιολόγησης είναι να απεικονίσουμε την κατανομή των αραιών πινάκων (sparse matrices) από τα δεδομένα δοκιμής των τριών συνόλων δεδομένων. Η αξιολόγηση αυτή εφαρμόστηκε χρησιμοποιώντας μοντέλα διανυσμάτων χρησιμοποιώντας τα σύνολα δεδομένων 20Newsgroups, το Movie Reviews και το Finegrained. Για τα αραιά δεδομένα μεγάλων διαστάσεως, μειώνουμε αρχικά τις διαστάσεις σε 50 με τη μέθοδο (TruncatedSVD) και στη συνέχεια εκτελούμε το t-SNE. Αυτό συνήθως βελτιώνει την απεικόνιση.

Όσον αφορά στην επιλογή των υπερπαραμέτρων, έχει προταθεί από τους van der Maaten και Hinton [236] ότι οι τιμές πολυπλοκότητας πρέπει να κυμαίνονται μεταξύ 5 και 50. Μετά από πειράματα για τιμές 2-100, παρατηρήσαμε ότι με πολύ μικρές τιμές πολυπλοκότητας, κυριαρχούν οι τοπικές διακυμάνσεις. Έτσι, καταλήγουμε στο συμπέρασμα ότι για να λειτουργήσει σωστά η εκπαίδευση του ανταγωνιστικού μας δικτύου, η τιμή της πολυπλοκότητας θα πρέπει να είναι μικρότερη από τον αριθμό των σημείων. Ως εκ τούτου, για τις ανάγκες των πειραμάτων μας, ορίσαμε την πολυπλοκότητα ως 40 αντί για 30; η τελευταία είναι η προτεινόμενη προεπιλεγμένη τιμή. Επιπλέον, κάθε ένα από τα διαγράμματα που απεικονίζονται στα Σχήματα 7.4–7.6 έγινε με 1000 επαναλήψεις με ρυθμό εκμάθησης (συντά αποκαλούμενο “epsilon”) 10 και έφτασε σε σημείο σταθερότητας με βήμα 1000. Κρατώντας την τιμή αμηχανίας σταθερή στην τιμή 40, πειραματιστήκαμε με τιμές των 10, 20, 60, 120 και 1000 βημάτων, που παράγουν εικόνες για πέντε διαφορετικές διαδρομές σε μια πολυπλοκότητα 40. Παρατηρήσαμε ότι οι πρώτες τέσσερις τερματίστηκαν πριν επιτευχθεί σταθερότητα, ενώ τα παραγόμενα σχέδια φαίνονται μονοδιάστατα αντί για δισδιάστατα με περίεργα σχήματα, υποδεικνύοντας ότι η διαδικασία σταμάτησε πολύ νωρίς. Ως αποτέλεσμα, καταλήγουμε στο συμπέρασμα ότι δεν υπάρχει σταθερός αριθμός βημάτων που να οδηγεί σε σταθερό αποτέλεσμα. Με άλλα λόγια, διαφορετικά σύνολα δεδομένων είναι πιθανό να απαιτούν διαφορετικούς αριθμούς επαναλήψεων για να συγκλίνουν.

Μια πτυχή αναφορικά με τη χρήση της προτεινόμενης προσέγγισης η οποία αποτελεί πρόκληση είναι η δυσκολία εκπαίδευσης ανταγωνιστικών δικτύων. Για παράδειγμα, οι Zhao et al. [270] παρατήρησαν ότι σε προβλήματα Φυσικής Γλώσσας, οι βελτιώσεις των EBGANs ήταν μεγάλες για ορισμένους τύπους θορύβων, αλλά λιγότερο αποτελεσματικές για άλλους. Πρότειναν λοιπόν ότι ο συντονισμός των παραμέτρων θα μπορούσε να οδηγήσει σε καλύτερα αποτελέσματα. Παρατηρήσαμε επίσης ότι το μοντέλο μας απέτυχε να συγκλίνει για ορισμένες παραμέτρους, κάτι που αποτελεί κοινή διαπίστωση σε προβλήματα min-max. Όταν εκπαιδεύονται κατάλληλα, ωστόσο, αυτό το ισχυρό πλαίσιο GAN μπορεί να προσφέρει ικανοποιητικές λύσεις σε ένα από τα πιο σημαντικά προβλήματα στον τομέα της Φυσικής Γλώσσας.

Κεφάλαιο 8

Σύνοψη - Συμπεράσματα

Οι πρόσφατες εξελίξεις στη Συναισθηματική Υπολογιστική (AC) περιλαμβάνουν την έρευνα προς την κατεύθυνση της αυτόματης ανάλυσης της συναισθηματικά βελτιωμένης ανθρώπινης συμπεριφοράς κατά τη διάρκεια της πολυπληθούς αλληλεπίδρασης σε διαφορετικές συνθήκες. Η παρούσα εργασία ερευνά τον τρόπο με τον οποίο το πλαίσιο μπορεί να ενσωματωθεί σε πολυπληθείς μονοτροπικές ή πολυτροπικές αλληλεπιδράσεις στα πλαίσια της Συναισθηματικής Υπολογιστικής.

Ζητήματα σχετικά με την ενσωμάτωση του πλαισίου όπως η σημασία και το κίνητρο για την ενσωμάτωσή του σε συστήματα Συναισθηματικής Υπολογιστικής, τα κατάλληλα μοντέλα ενσυναίσθησης (Κεφάλαιο 3), οι χρήσιμοι πόροι (σώματα κειμένων) για τις πολυπληθείς αλληλεπιδράσεις αναφορικά με την ανάλυση του πλαισίου, η προσέγγιση του πλαισίου ως μία επιπρόσθετη ροή πληροφορίας σε πολυπληθή Συναισθηματικής Υπολογιστικής συστήματα και σε συστήματα Συναισθηματικής Υπολογιστικής που γνωρίζουν το πλαίσιο, προσεγγίστηκαν ως ερευνητικά ερωτήματα στα πλαίσια της αναθεώρησης της τρέχουσας τεχνολογίας στον τομέα της έρευνας. Οι προκλήσεις που προκύπτουν από την ενσωμάτωση του πλαισίου εντοπίζονται και αναλύονται εκτενώς προκειμένου να προβλεφθούν οι μελλοντικές ερευνητικές κατευθύνσεις στον τομέα (Κεφάλαιο 2).

Το Κεφάλαιο 4 αναφέρεται στην πολυτροπική αναγνώριση των συναισθηματικών καταστάσεων των μελών μίας ομάδας που συμμετείχαν σε ένα απαιτητικό επιχειρησιακό έργο. Συγκεκριμένα εξετάσαμε τη συσχέτιση των σημάτων φυσιολογίας (αντιδράσεις που προκλήθηκαν στο σώμα (αύξηση καρδιακών παλμών κοκ)) των δύο μελών μίας ομάδας που συμμετείχαν σε ένα σενάριο “αφόπλισης μίας βόμβας” σε συνδυασμό με τις εκφραστικές τους απαντήσεις. Τα πειραματικά αποτελέσματα αναδεικνύουν ότι το λεκτικό κομμάτι της αλληλεπίδρασης του εκπαιδευτή και του εκπαιδευόμενου κατά τη διάρκεια αυτού του αγχωτικού καθήκοντος, μεταφέρει πληροφορίες σχετικά με την εσωτερική κατάσταση του εκπαιδευόμενου. Αντικατοπτρίζουν επίσης την ποσότητα των απαντήσεων μέσω της ομιλίας τους αναφορικά με ένα ερέθισμα, απαντήσεις οι οποίες μπορεί να είναι επιπρόσθετα συνδεδεμένες με το ποσοστό των μη ορατών κοινωνικο-γνωστικών δραστηριοτήτων και οι οποίες δεν είναι πάντα προφανείς μέσω των παραδοσιακών μεθόδων παρατήρησης. Τέλος, προτείναμε δύο υπάρχοντα σύνολα μαθησιακών μεθόδων που είναι νέα στον τομέα

της συλλογής των εξωτερικών και εσωτερικών ενδείξεων των ομιλητών, αποδεικνύοντας ότι αυτές οι μέθοδοι μπορούν να αποφέρουν βελτιώσεις σε σχέση με τις παραδοσιακές μεθόδους ανάλυσης. Επιβεβαιώνοντας λοιπόν τη μοναδικότητα των προσωπικών χαρακτηριστικών κάθε συμμετέχοντα διαπιστώσαμε ότι είναι ακόμη πιο έκδηλη η ανάγκη για την ύπαρξη υπολογιστικών μοντέλων που να εστιάζουν στην καλύτερη αναγνώριση, ερμηνεία και επεξεργασία των απόψεων (opinions) και συναισθημάτων (sentiments) των συμμετεχόντων κατά τη διάρκεια τέτοιων αλληλεπιδράσεων.

Στο Κεφάλαιο 5, προτείναμε μία μέθοδο για την εξαγωγή πτυχών του εννοιολογικού πλαισίου χρησιμοποιώντας το σώμα κειμένου SEMAINE. Η προτεινόμενη αρχιτεκτονική του συστήματος ενσωματώνει το φορμαλισμό **W5+**, σε συστήματα που έχουν επίγνωση του πλαισίου κατά τη διάρκεια πολυπληθών αλληλεπιδράσεων. Το προτεινόμενο μοντέλο περιλαμβάνει την ανίχνευση και την εξαγωγή σημασιολογικών εννοιών του πλαισίου (context-aware semantic concepts), διανθίζοντας τα μοντέλα ενσυναίσθησης μέσω πληροφοριών του πλαισίου και μέσω της αναπαράστασης εννοιών του πλαισίου κατά την εκτίμηση της αξιολόγησης (appraisal). Ως εκ τούτου, το προτεινόμενο μοντέλο ενσωμάτωσης του πλαισίου W5+ αξιοποιεί μόνο τα οφέλη των ληφθέντων πληροφοριών, επιτρέποντας στα συναισθήματα να μεταφέρονται από πολλαπλές έννοιες σε πολλαπλές έννοιες. Επιπλέον, μετατοπίζει το ενδιαφέρον στην καλύτερη αναγνώριση, ερμηνεία και επεξεργασία των απόψεων (opinions) και συναισθημάτων (sentiments) μέσα στην πρόταση καθώς εκμεταλλεύεται τη σημασιολογία, τις γνωστικές και συναισθηματικές πληροφορίες της μη λεκτικής συμπεριφοράς των ατόμων που αλληλεπιδρούν, εστιάζοντας σε απαντήσεις ερωτημάτων όπως το “Γιατί” οι οποίες είναι περισσότερο υποκειμενικές.

Στα πλαίσια της προσπάθειάς μας να αναδείξουμε την συνεισφορά του εννοιολογικού πλαισίου αναφορικά με την ερμηνεία των απόψεων (opinions) και των συναισθημάτων (sentiments) μελετήσαμε διεξοδικά τη χρησιμότητα δύο υπολογιστικών μοντέλων, του (Conditional Random Field (CRF)) και της θεωρίας του λόγου (Rhetorical Structure Theory (RST)) για την ενσωμάτωση της δομής της πρότασης και των πληροφοριών πλαισίου σε συστήματα που έχουν επίγνωση του συναισθηματικού τους περιβάλλοντος. Τα δύο προτεινόμενα μοντέλα εφαρμόστηκαν σε δύο πηγές πληροφοριών διαφορετικού μεγέθους και είδους. Αρχικά, επιβεβαιώσαμε ότι ο τομέας του SA μπορεί να βελτιωθεί εάν εκμεταλλευτούμε πλήρως τη δομή των προτάσεων και χρησιμοποιώντας πληροφορίες πλαισίου για να καταγράψουμε τη σχέση μεταξύ των προτάσεων, βελτιώνοντας την SA σε επίπεδο κειμένου (document-level). Η συνεισφορά αυτής της ερευνητικής προσπάθειας παρουσιάζεται στο Κεφάλαιο 6 και έγκειται στην επέκταση του συνόλου των χαρακτηριστικών πλαισίου που χρησιμοποιούνται μαζί με σημασιολογικά, συντακτικά και χαρακτηριστικά πλαισίου RST. Πράγματι μόνο όταν συνδυάζουμε χαρακτηριστικά πλαισίου RST με μονογράμματα/διγράμματα (unigrams/bigrams), μπορούμε να επιτύχουμε βελτιώσεις στην απόδοση της κατηγοριοποίησης του συστήματος.

Τέλος, λαμβάνοντας υπόψιν μας ότι ο τομέας της Βαθιάς Μάθησης (Deep Learning) έχει αναδειχθεί ως μια ισχυρή τεχνική Μηχανικής Μάθησης καθώς μαθαίνει πολλαπλά

στρώματα αναπαραστάσεων των δεδομένων παράγοντας αποτελέσματα πρόβλεψης τελευταίας τεχνολογίας, στο τελευταίο Κεφάλαιο της διατριβής (Κεφάλαιο 7) ασχολούμαστε με την εφαρμογή αυτών των τεχνικών στην ανάλυση συναισθημάτων. Ειδικότερα, η ικανότητα των μοντέλων Βαθιάς Μάθησης να μαθαίνουν ισχυρές αναπαραστάσεις χαρακτηριστικών από μη επισημειωμένα δεδομένα έχει αρκετές εφαρμογές σε μία μεγάλη ποικιλία εργασιών εκμάθησης μηχανών. Ένας τρόπος για να δημιουργήσουμε τέτοιες αναπαραστάσεις είναι να εκπαιδύσουμε βαθιά παραγωγικά μοντέλα που να μπορούν να μάθουν να συλλάβουν τη σύνθετη κατανομή των πραγματικών δεδομένων. Οι προσεγγίσεις του ανταγωνιστικού παραγωγικού δικτύου (GAN) έδειξαν εντυπωσιακά αποτελέσματα στην παραγωγή μοντέλων εικόνων. Παρόλα αυτά, έχει γίνει σχετικά περιορισμένη δουλειά αναφορικά με την αξιολόγηση της απόδοσης αυτών των μεθόδων για την εκμάθηση της Φυσικής Γλώσσας, τόσο σε επιβλεπόμενες όσο και σε μη επιβλεπόμενες ρυθμίσεις σε επίπεδο κειμένου, πρότασης και χαρακτηριστικών. Εκτεταμένη επικύρωση της έρευνάς μας βασίζεται σε μία σειρά πειραμάτων που πραγματοποιήθηκαν στα σύνολα δεδομένων: 20Newsgroups, το Movie Review (MR) και το σύνολο δεδομένων Fine-grained Sentiment (FSD). Η πειραματική ανάλυσή μας επιβεβαιώνει ότι τα GANs μπορούν να μάθουν με επιτυχή τρόπο αναπαραστάσεις κειμένων Φυσικής Γλώσσας και στα τρία προαναφερθέντα επίπεδα.

Κεφάλαιο 9

Κατάλογος δημοσιεύσεων

9.1 Περιοδικά

[4] **A. Vlachostergiου**, Ph. Mylonas, K. Karpouzis and A. Stafylopatis, “Domain Adversarial for Natural Language texts”. **(under preparation)**

[3] **A. Vlachostergiου**, K. Karpouzis, Ph. Mylonas, and A. Stafylopatis, “Multitask Learning for Sentiment Analysis”. **(submitted)**

[2] **A. Vlachostergiου**, G. Caridakis, Ph. Mylonas and A. Stafylopatis, “Learning Representations of Natural Language texts with Generative Adversarial Networks at Document, Sentence and Aspect-level”, *Algorithms, Special Issue in Humanistic Data Mining: Tools and Applications*, vol. 11, issue 10, Oct 2018.

[1] **A. Vlachostergiου**, G. Stratogiannis, G. Caridakis, G. Siolas and Ph. Mylonas, “User adaptive and context aware Smart Home using pervasive and semantic technologies”, *Journal of Electrical and Computer Engineering*, vol. 2016, 2016.

9.2 Κεφάλαια σε βιβλία

[1] **A. Vlachostergiου**, G. Marandianos and S. Kollias, “From Conditional Random Field (CRF) to Rhetorical Structure Theory (RST): incorporating Context information in Sentiment Analysis”, in *ESWC (Satellite Events): 283-295*, 2017.

9.3 Συνέδρια

[14] **A. Vlachostergiου** and A. Stafylopatis, “Extracting NLP features corresponding to higher level representations”. **(submitted)**

[13] **A. Vlachostergiου**, A. Tagaris, A. Stafylopatis and S. Kollias, “Multi-Task Learning

for predicting Parkinson's Disease based on Medical Imaging Information", in *the 25th IEEE International Conference on Image Processing (ICIP 2018)*, Athens, Greece, Oct 2018.

[12] **A. Vlachostergiou**, A. Tagaris, A. Stafylopatis and S. Kollias, "Investigating the best performing task conditions of a multi-tasking learning model in Healthcare using Convolutional Neural Networks: Evidence from a Parkinson's disease Database", in *the 25th IEEE International Conference on Image Processing (ICIP 2018)*, Athens, Greece, Oct 2018.

[11] **A. Vlachostergiou**, M. Dennison, C. Neubauer, S. Scherer, P. Khooshabeh and A. Harrison, "Unfolding the External Behavior and Inner Affective State of Teammates through Ensemble Learning: Experimental Evidence from a Dyadic Team Corpus", in *the 11th edition of the Language Resources and Evaluation Conference (LREC 2018)*, Miyazaki, Japan, May 2018. ****This work was conducted during my summer internship at University of Southern California (USC).**

[10] **A. Vlachostergiou**, G. Marandianos and S. Kollias, "Context Incorporation using context - aware language features", in *the 25th European International Signal Processing Conference (EUSIPCO 2017)*, Kos island, Greece, Aug 2017.

[9] **A. Vlachostergiou**, G. Marandianos and S. Kollias, "From HCI and Affective Computing to Sentiment Analysis: extending the pool of context-aware features in Affective-aware systems", in *the 5th International Conference on Affective and Pleasurable Design*, Los Angeles, CA, USA, July 2017.

[8] **A. Vlachostergiou**, G. Marandianos and S. Kollias, "From Conditional Random Field (CRF) to Rhetorical Structure Theory (RST): incorporating Context information in Sentiment Analysis", in *the 3rd International Workshop on Emotions, Modality, Sentiment Analysis and the Semantic Web, (ESWC 2017)*, Portoroz, Slovenia, May 2017.

(Best Workshop Paper Award)

[7] **A. Vlachostergiou**, G. Stratogiannis, G. Siolas, G. Caridakis and Ph. Mylonas, "Pervasive semantic representation and integration of context-aware homes in context sensitive cities", in *the 1st Workshop on Artificial Intelligence and Internet of Things (AI-IoT)*, in conjunction with SETN 2016, Thessaloniki, Greece, May 2016.

[6] **A. Vlachostergiou** and S. Kollias, "A revisit of the incorporation of context-awareness in affective computing systems", in *the 9th International and Interdisciplinary Conference on Modeling and Using Context*, Doctoral Symposium (DS), Larnaca, Cyprus, Nov 2015.

[5] G. Stratogiannis, **A. Vlachostergiou**, G. Siolas, G. Caridakis, Ph. Mylonas, A. Stafy-

Ilopatis and S. Kollias, “User and home appliances pervasive interaction in a sensor driven Smart Home environment: the SandS approach,” in *the 10th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*, Trento, Italy, Nov 2015.

[4] **A. Vlachostergiou**, G. Stratogiannis, G. Caridakis, G. Siolas and Ph. Mylonas, “Smart home context awareness based on Smart and Innovative Cities”, in *the 2nd Workshop on Innovative European Policies and Applied Measures for Developing Smart Cities (IPMSC)*, Island of Rhodes, Greece, Sep 2015.

[3] **A. Vlachostergiou**, G. Caridakis, A. Raouzaïou and S. Kollias, “HCI and Natural Progression of Context-Related Questions”, in *the 17th International Conference on Human-Computer Interaction (HCII)*, Los Angeles, CA, USA, Aug 2015.

[2] **A. Vlachostergiou**, G. Caridakis and S. Kollias, “Investigating Context Awareness of Affective Computing Systems: A Critical Approach”, in *the 6th International Conference on Intelligent Human Computer Interaction (iHCI)*, Evry, France, Dec 2014.

[1] **A. Vlachostergiou**, G. Caridakis and S. Kollias, “Context in Affective Multiparty and Multimodal Interaction: Why, Which, How and Where?”, in *the Workshop on Understanding and Modeling Multiparty, Multimodal Interactions, UM3I@ICMI*, Istanbul, Turkey, Nov 2014.

Βιβλιογραφία

- [1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *OSDI*, volume 16, pages 265–283, 2016.
- [2] Gregory D Abowd, Anind K Dey, Peter J Brown, Nigel Davies, Mark Smith, and Pete Steggles. Towards a better understanding of context and context-awareness. In *Handheld and Ubiquitous Computing*, pages 304–307. Springer, 1999.
- [3] Rajendra Akerkar. *Big data computing*. CRC Press, 2013.
- [4] Grigory Antipov, Moez Baccouche, and Jean-Luc Dugelay. Face aging with conditional generative adversarial networks. In *IEEE International Conference on Image Processing (ICIP)*, pages 2089–2093. IEEE, 2017.
- [5] Orestes Appel, Francisco Chiclana, and Jenny Carter. Main concepts, state of the art and future research questions in sentiment analysis. *Acta Polytechnica Hungarica*, 12(3):87–108, 2015.
- [6] Alessio Palmero Aprosio, Francesco Corcoglioniti, Mauro Dragoni, and Marco Rospocher. Supervised opinion frames detection with raid. In *Semantic Web Evaluation Challenge*, pages 251–263. Springer, 2015.
- [7] Martin Arjovsky and Léon Bottou. Towards principled methods for training generative adversarial networks. *arXiv preprint arXiv:1701.04862*, 2017.
- [8] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.
- [9] Sileye O Ba and Jean-Marc Odobez. Recognizing visual focus of attention from head pose in natural meetings. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(1):16–33, 2009.
- [10] Daniella Bal, Malissa Bal, Arthur Van Bunningen, Alexander Hogenboom, Frederik Hogenboom, and Flavius Frasinca. Sentiment analysis with a multilingual pipeline. In *Web Information System Engineering–WISE*, pages 129–142. Springer, 2011.

- [11] C Daniel Batson, Laura L Shaw, and Kathryn C Oleson. Differentiating affect, mood, and emotion: toward functionally based conceptual distinctions. 1992.
- [12] Philip Beineke, Trevor Hastie, Christopher Manning, and Shivakumar Vaithyanathan. Exploring sentiment summarization. In *Proceedings of the AAAI spring symposium on exploring attitude and affect in text: theories and applications*, volume 39, 2004.
- [13] Samy Bengio, Oriol Vinyals, Navdeep Jaitly, and Noam Shazeer. Scheduled sampling for sequence prediction with recurrent neural networks. In *Advances in Neural Information Processing Systems*, pages 1171–1179, 2015.
- [14] Raffaella Bernardi, Ruket Cakici, Desmond Elliott, Aykut Erdem, Erkut Erdem, Nazli Ikizler-Cinbis, Frank Keller, Adrian Muscat, and Barbara Plank. Automatic description generation from images: A survey of models, datasets, and evaluation measures. *Journal of Artificial Intelligence Research*, 55:409–442, 2016.
- [15] Eva Blomqvist, Katja Hose, Heiko Paulheim, Agnieszka Ławrynowicz, Fabio Ciravegna, and Olaf Hartig. *The Semantic Web: ESWC 2017 Satellite Events: ESWC 2017 Satellite Events, Portorož, Slovenia, May 28–June 1, 2017, Revised Selected Papers*, volume 10577. Springer, 2017.
- [16] Ronald Bock, Andreas Wendemuth, Stefan Gluge, and Ingo Siegert. Annotation and classification of changes of involvement in group conversation. In *Proceedings of the Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 803–808. IEEE, 2013.
- [17] Francesca Bonin, Ronald Bock, and Nick Campbell. How do we react to context? annotation of individual and group engagement in a video corpus. In *Privacy, Security, Risk and Trust (PASSAT), International Conference on Social Computing (SocialCom)*, pages 899–903. IEEE, 2012.
- [18] Samuel R Bowman, Luke Vilnis, Oriol Vinyals, Andrew Dai, Rafal Jozefowicz, and Samy Bengio. Generating sentences from a continuous space. In *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*, pages 10–21, 2016.
- [19] Margaret M Bradley and Peter J Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994.
- [20] Margaret M Bradley and Peter J Lang. Affective norms for english words (anew): Instruction manual and affective ratings. Technical report, Technical Report C-1, The Center for Research in Psychophysiology, University of Florida, 1999.

- [21] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [22] CD Broad. Emotion and sentiment. 1971.
- [23] P. J. Brown. The stick-e document: a framework for creating context-aware applications. In *Proceedings of EP'96, Palo Alto*, pages 182–196, January 1996.
- [24] Peter J Brown, John D Bovey, and Xian Chen. Context-aware applications: from the laboratory to the marketplace. *Personal Communications*, 4(5):58–64, 1997.
- [25] Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N Chang, Sungbok Lee, and Shrikanth S Narayanan. Iemocap: Interactive emotional dyadic motion capture database. *Language resources and evaluation*, 42(4):335, 2008.
- [26] Carlos Busso, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann, and Shrikanth Narayanan. Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Proceedings of the 6th International Conference on Multimodal Interfaces (ICMI)*, pages 205–211. ACM, 2004.
- [27] Susan D Calkins and Nathan A Fox. Self-regulatory processes in early personality development: A multilevel approach to the study of childhood social withdrawal and aggression. *Development and Psychopathology*, 14(3):477–498, 2002.
- [28] Rafael A Calvo and Sidney D’Mello. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1(1):18–37, 2010.
- [29] Erik Cambria, Amir Hussain, Catherine Havasi, and Chris Eckl. SenticSpace: visualizing opinions and sentiments in a multi-dimensional vector space. In *Knowledge-Based and Intelligent Information and Engineering Systems*, pages 385–393. Springer, 2010.
- [30] James M Carroll and James A Russell. Do facial expressions signal specific emotions? judging emotion from the face in context. *Journal of personality and social psychology*, 70(2):205, 1996.
- [31] Ginevra Castellano, George Caridakis, Antonio Camurri, Kostas Karpouzis, Gualtiero Volpe, and Stefanos Kollias. Body gesture and facial expression analysis for automatic affect recognition. *Blueprint for Affective Computing: A sourcebook*, pages 245–255, 2010.
- [32] Raymond B Cattell. Sentiment or attitude? the core of a terminology problem in personality research. *Journal of Personality*, 9(1):6–17, 1940.

- [33] Nitesh V Chawla. Data mining for imbalanced datasets: An overview. In *Data mining and knowledge discovery handbook*, pages 875–886. Springer, 2009.
- [34] Danqi Chen and Christopher Manning. A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 740–750, 2014.
- [35] Tao Chen, Ruifeng Xu, Yulan He, and Xuan Wang. Improving sentiment analysis via sentence type classification using bilstm-crf and cnn. *Expert Systems with Applications*, 72:221–230, 2017.
- [36] Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. Identifying sources of opinions with conditional random fields and extraction patterns. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 355–362. Association for Computational Linguistics, 2005.
- [37] François Chollet et al. Keras: Deep learning library for theano and tensorflow. *URL: <https://keras.io/k>*, 7:8, 2015.
- [38] Tanzeem Choudhury and Alex Pentland. Modeling face-to-face communication using the sociometer. 5:3–8, 2003.
- [39] Matt Cieslak. Moving ensemble averaging program. <https://github.com/mattcieslak/MEAP>, 2017.
- [40] Adam Coates, Andrew Ng, and Honglak Lee. An analysis of single-layer networks in unsupervised feature learning. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, pages 215–223, 2011.
- [41] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12(Aug):2493–2537, 2011.
- [42] Cristina Conati. Probabilistic assessment of user’s emotions in educational games. *Applied artificial intelligence*, 16(7-8):555–575, 2002.
- [43] Alexis Conneau, Holger Schwenk, Loïc Barrault, and Yann Lecun. Very deep convolutional networks for text classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, volume 1, pages 1107–1116, 2017.
- [44] Roddy Cowie. Perceiving emotion: towards a realistic understanding of the task. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3515–3525, 2009.

- [45] Roddy Cowie, Ellen Douglas-Cowie, Susie Savvidou*, Edelle McMahon, Martin Sawey, and Marc Schröder. 'feeltrace': An instrument for recording perceived emotion in real time. In *ISCA tutorial and research workshop (ITRW) on speech and emotion*, 2000.
- [46] Roddy Cowie, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias, Winfried Fellenz, and John G Taylor. Emotion recognition in human-computer interaction. *IEEE Signal processing magazine*, 18(1):32–80, 2001.
- [47] Tim Dalgleish and Mick Power. *Handbook of cognition and emotion*. John Wiley & Sons, 2000.
- [48] Antonio R Damasio. *Looking for Spinoza: Joy, sorrow, and the feeling brain*. Houghton Mifflin Harcourt, 2003.
- [49] Jesse Davis and Mark Goadrich. The relationship between precision-recall and roc curves. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 233–240. ACM, 2006.
- [50] Liyanage C De Silva and Pei Chi Ng. Bimodal emotion recognition. In *Proceedings of the 4th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 332–335. IEEE, 2000.
- [51] Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391, 1990.
- [52] Li Deng. The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE Signal Processing Magazine*, 29(6):141–142, 2012.
- [53] Mark Dennison, Cathy Neubauer, Tony Passaro, Andre Harrison, Stefan Scherer, and Pete Khooshabeh. Using cardiovascular features to classify state changes during cooperation in a simulated bomb diffusal task. In *Proceedings of the Physiologically Aware Virtual Agents Workshop*. IEEE, 2016.
- [54] Emily L Denton, Soumith Chintala, Rob Fergus, et al. Deep generative image models using a laplacian pyramid of adversarial networks. In *Advances in neural information processing systems*, pages 1486–1494, 2015.
- [55] Thomas G Dietterich. Ensemble learning. *The handbook of brain theory and neural networks*, 2:110–125, 2002.
- [56] Raymond J Dolan. Emotion, cognition, and behavior. *science*, 298(5596):1191–1194, 2002.

- [57] Jeff Donahue, Philipp Krähenbühl, and Trevor Darrell. Adversarial feature learning. *arXiv preprint arXiv:1605.09782*, 2016.
- [58] Ellen Douglas-Cowie, Roddy Cowie, Cate Cox, Noam Amier, and Dirk Heylen. The sensitive artificial listener: an induction technique for generating emotionally coloured conversation. In *LREC Workshop on Corpora for Research on Emotion and Affect*, pages 1–4. European Language Resources Association(ELRA), 2008.
- [59] Hui Du, Xueke Xu, Xueqi Cheng, Dayong Wu, Yue Liu, and Zhihua Yu. Aspect-specific sentimental word embedding for sentiment analysis of online reviews. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 29–30. International World Wide Web Conferences Steering Committee, 2016.
- [60] Zoran Duric, Wayne D Gray, Ric Heishman, Fayin Li, Azriel Rosenfeld, Michael J Schoelles, Christian Schunn, and Harry Wechsler. Integrating perceptual and cognitive modeling for adaptive and intelligent human-computer interaction. *Proceedings of the IEEE*, 90(7):1272–1289, 2002.
- [61] Paul Ekman and Wallace V Friesen. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *semiotica*, 1(1):49–98, 1969.
- [62] Rana El Kaliouby, Peter Robinson, and Simeon Keates. Temporal context and the recognition of emotion from facial expression. In *Proceedings of the HCI International Conference*, pages 631–635. American Psychological Association, 2003.
- [63] Mona El-Sheikh, E Mark Cummings, and Virginia L Goetsch. Coping with adults' angry behavior: Behavioral, physiological, and verbal responses in preschoolers. *Developmental Psychology*, 25(4):490, 1989.
- [64] Phoebe C Ellsworth and Klaus R Scherer. Appraisal processes in emotion. *Handbook of affective sciences*, 572:V595, 2003.
- [65] Andrea Esuli and Fabrizio Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, volume 6, pages 417–422. Citeseer, 2006.
- [66] Marco Federici and Mauro Dragoni. Towards unsupervised approaches for aspects extraction.
- [67] Marco Federici and Mauro Dragoni. A knowledge-based approach for aspect-based opinion mining. In *Semantic Web Evaluation Challenge*, pages 141–152. Springer, 2016.

- [68] Ronen Feldman. Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82–89, 2013.
- [69] Basura Fernando, Amaury Habrard, Marc Sebban, and Tinne Tuytelaars. Unsupervised visual domain adaptation using subspace alignment. In *Proceedings of the IEEE international conference on computer vision*, pages 2960–2967, 2013.
- [70] Martin Fishbein and Icek Ajzen. Attitudes and opinions. *Annual review of psychology*, 23(1):487–544, 1972.
- [71] Kristie S Fleckenstein. Defining affect in relation to cognition: A response to susan mcleod. *Journal of Advanced Composition*, pages 447–453, 1991.
- [72] N Fragopanagos and John G Taylor. Emotion recognition in human–computer interaction. *Neural Networks*, 18(4):389–405, 2005.
- [73] David Franklin and Joshua Flascbart. All gadget and no representation makes jack a dull environment. In *Proceedings of the AAAI Spring Symposium on Intelligent Environments*, pages 155–160, 1998.
- [74] Vera V French. The structure of sentiments. *Journal of Personality*, 16(1):78–108, 1947.
- [75] Robert W Frick. Communicating emotion: The role of prosodic features. *Psychological Bulletin*, 97(3):412, 1985.
- [76] Jay Friedenber and Gordon Silverman. *Cognitive science: An introduction to the study of mind*. Sage, 2011.
- [77] Aldo Gangemi, Valentina Presutti, and Diego Reforgiato Recupero. Frame-based detection of opinion holders and topics: a model and a tool. *IEEE Computational Intelligence Magazine*, 9(1):20–30, 2014.
- [78] Daniel Gatica-Perez. Automatic nonverbal analysis of social interaction in small groups: A review. *Image and Vision Computing*, 27(12):1775–1787, 2009.
- [79] Ian R Gellatly and John P Meyer. The effects of goal difficulty on physiological arousal, cognition, and task performance. *Journal of Applied Psychology*, 77(5):694, 1992.
- [80] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741, 1984.
- [81] Michael Glodek, Stephan Tschechne, Georg Layher, Martin Schels, Tobias Brosch, Stefan Scherer, Markus Kächele, Miriam Schmidt, Heiko Neumann, Günther Palm,

- et al. Multiple classifier systems for the classification of audio-visual emotional states. *Affective Computing and Intelligent Interaction (ACII)*, pages 359–368, 2011.
- [82] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. 2016.
- [83] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- [84] Steven L Gordon. *The sociology of sentiments and emotion*, 1981.
- [85] Didier Grandjean, David Sander, and Klaus R Scherer. Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization. *Consciousness and cognition*, 17(2):484–495, 2008.
- [86] Alex Graves, Santiago Fernández, and Jürgen Schmidhuber. Bidirectional lstm networks for improved phoneme classification and recognition. In *Proceedings of the 15th International Conference on Artificial Neural Networks: Formal Models and Their Applications - Volume Part II, ICANN'05*, pages 799–804. Springer-Verlag, 2005.
- [87] Mark K Greenwald, Edwin W Cook, Peter J Lang, et al. Affective judgment and psychophysiological response: Dimensional covariation in the evaluation of pictorial stimuli. *Journal of psychophysiology*, 3(1):51–64, 1989.
- [88] Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Rezende, and Daan Wierstra. Draw: A recurrent neural network for image generation. In *International Conference on Machine Learning*, pages 1462–1471, 2015.
- [89] Hatice Gunes and Massimo Piccardi. Automatic temporal segment detection and affect recognition from face and body display. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(1):64–84, 2009.
- [90] Hatice Gunes and Björn Schuller. Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing*, 31(2):120–136, 2013.
- [91] Hatice Gunes and Björn Schuller. Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing*, 31(2):120 – 136, 2013.
- [92] Zakia Hammal and Miriam Kunz. Pain monitoring: A dynamic and context-sensitive system. *Pattern Recognition*, 45(4):1265–1280, 2012.

- [93] Zakia Hammal, Miriam Kunz, Martin Arguin, and Frédéric Gosselin. Spontaneous pain expression recognition in video sequences. In *Proceedings of the international conference on Visions of Computer Science*, pages 191–210. British Computer Society, 2008.
- [94] Zakia Hammal and Merlin Teodosia Suarez. Towards context based affective computing. In *Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 802–802. IEEE, 2013.
- [95] Vasileios Hatzivassiloglou and Kathleen R McKeown. Predicting the semantic orientation of adjectives. In *35th annual meeting of the association for computational linguistics and 8th conference of the european chapter of the association for computational linguistics*, pages 174–181. ACL, 1997.
- [96] Bas Heerschoop, Frank Goossen, Alexander Hogenboom, Flavius Frasincar, Uzay Kaymak, and Franciska de Jong. Polarity analysis of texts using discourse structure. In *20th International Conference on Information and Knowledge Management*, pages 1061–1070. ACM, 2011.
- [97] Mattias Heldner and Jens Edlund. Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4):555–568, 2010.
- [98] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.
- [99] Geoffrey E Hinton and Ruslan R Salakhutdinov. Replicated softmax: an undirected topic model. In *Advances in neural information processing systems*, pages 1607–1614, 2009.
- [100] Alexander Hogenboom, Frederik Hogenboom, Flavius Frasincar, Kim Schouten, and Otto van der Meer. Semantics-based information extraction for detecting economic events. *Multimedia Tools and Applications*, 64(1):27–52, 2013.
- [101] Zhiting Hu, Zichao Yang, Xiaodan Liang, Ruslan Salakhutdinov, and Eric P Xing. Controllable text generation. *arXiv preprint arXiv:1703.00955*, 2017.
- [102] Fu Jie Huang, Y-Lan Boureau, Yann LeCun, et al. Unsupervised learning of invariant feature hierarchies with applications to object recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2007.
- [103] Niklas Jakob and Iryna Gurevych. Extracting opinion targets in a single-and cross-domain setting with conditional random fields. In *Proceedings of the conference on Empirical Methods in Natural Language Processing*, pages 1035–1045. Association for Computational Linguistics, 2010.

- [104] Yangfeng Ji and Jacob Eisenstein. Representation learning for text-level discourse parsing. In *ACL (1)*, pages 13–24, 2014.
- [105] Gareth R Jones and Jennifer M George. The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of management review*, 23(3):531–546, 1998.
- [106] Jaap Kamps, MJ Marx, Robert J Mokken, and Maarten De Rijke. Using wordnet to measure semantic orientations of adjectives. 2004.
- [107] Ittipan Kanluan, Michael Grimm, and Kristian Kroschel. Audio-visual emotion recognition using an emotion space concept. In *Proceeding of the 16th European Signal Processing Conference*, pages 1–5. IEEE, 2008.
- [108] Ashish Kapoor, Winslow Burleson, and Rosalind W Picard. Automatic prediction of frustration. *International Journal of Human-Computer Studies*, 65(8):724–736, 2007.
- [109] Gilad Katz, Nir Ofek, and Bracha Shapira. Consent: Context-based sentiment analysis. *Knowledge-Based Systems*, 84:162–178, 2015.
- [110] Jonghwa Kim and Elisabeth André. Emotion recognition using physiological and speech signal in short-term observation. In *International Tutorial and Research Workshop on Perception and Interactive Technologies for Speech-Based Systems*, pages 53–64. Springer, 2006.
- [111] Soo-Min Kim and Eduard Hovy. Determining the sentiment of opinions. In *Proceedings of the 20th International Conference on Computational Linguistics*, page 1367. Association for Computational Linguistics (ACL), 2004.
- [112] Soo-Min Kim and Eduard Hovy. Automatic detection of opinion bearing words and sentences. In *Companion Volume to the Proceedings of Conference including Posters/Demos and tutorial abstracts*, 2005.
- [113] Soo-Min Kim and Eduard Hovy. Automatic identification of pro and con reasons in online reviews. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 483–490. Association for Computational Linguistics, 2006.
- [114] Taesup Kim and Yoshua Bengio. Deep directed generative models with energy-based probability estimation. *arXiv preprint arXiv:1606.03439*, 2016.
- [115] Diederik P Kingma and Max Welling. Auto-encoding variational bayes.
- [116] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

- [117] Ryan Kiros, Yukun Zhu, Ruslan R Salakhutdinov, Richard Zemel, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. Skip-thought vectors. In *Advances in neural information processing systems*, pages 3294–3302, 2015.
- [118] Paul R Kleinginna and Anne M Kleinginna. A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and emotion*, 5(4):345–379, 1981.
- [119] Harold R Knudsen and Louis H Muzekari. The effects of verbal statements of context on facial expressions of emotion. *Journal of Nonverbal Behavior*, 7(4):202–212, 1983.
- [120] Igor Labutov and Hod Lipson. Re-embedding words. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 489–493, 2013.
- [121] John Lafferty, Andrew McCallum, and Fernando Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the eighteenth International Conference on Machine Learning, ICML*, volume 1, pages 282–289, 2001.
- [122] Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. Recurrent convolutional neural networks for text classification. In *AAAI*, volume 333, pages 2267–2273, 2015.
- [123] Quoc V Le, Jiquan Ngiam, Adam Coates, Abhik Lahiri, Bobby Prochnow, and Andrew Y Ng. On optimization methods for deep learning. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, pages 265–272. Omnipress, 2011.
- [124] Yann LeCun, Sumit Chopra, Raia Hadsell, M Ranzato, and F Huang. A tutorial on energy-based learning. *Predicting structured data*, 1(0), 2006.
- [125] Chi-Chun Lee, Carlos Busso, Sungbok Lee, and Shrikanth S Narayanan. Modeling mutual influence of interlocutor emotion states in dyadic spoken interactions. In *Proceedings of Interspeech*, pages 1983–1986. ISCA, 2009.
- [126] Chul Min Lee and Shrikanth S Narayanan. Toward detecting emotions in spoken dialogs. *IEEE transactions on Speech and Audio Processing*, 13(2):293–303, 2005.
- [127] Roger Levy and Christopher D Manning. Deep dependencies from context-free statistical parsers: correcting the surface dependency approximation. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, page 327. Association for Computational Linguistics, 2004.

- [128] Dahua Lin, Ashish Kapoor, Gang Hua, and Simon Baker. Joint people, event, and location recognition in personal photo collections using cross-domain context. In *European Conference on Computer Vision*, pages 243–256. Springer, 2010.
- [129] Kevin Lin, Dianqi Li, Xiaodong He, Zhengyou Zhang, and Ming-Ting Sun. Adversarial ranking for language generation. In *Advances in Neural Information Processing Systems*, pages 3155–3165, 2017.
- [130] Jackson Liscombe, Giuseppe Riccardi, and Dilek Hakkani-Tür. Using context to improve emotion detection in spoken dialog systems. In *Ninth European Conference on Speech Communication and Technology*, 2005.
- [131] Gwen C Littlewort, Marian Stewart Bartlett, Linda P Salamanca, and Judy Reilly. Automated measurement of children’s facial expressions during problem solving tasks. In *International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011)*, pages 30–35. IEEE, 2011.
- [132] Bing Liu. Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2:627–666, 2010.
- [133] Bing Liu. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167, 2012.
- [134] Bing Liu. *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press, 2015.
- [135] Dong C Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1-3):503–528, 1989.
- [136] Pengfei Liu, Xipeng Qiu, and Xuanjing Huang. Recurrent neural network for text classification with multi-task learning. *arXiv preprint arXiv:1605.05101*, 2016.
- [137] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. Adversarial autoencoders. *arXiv preprint arXiv:1511.05644*, 2015.
- [138] William C Mann and Sandra A Thompson. Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, 8(3):243–281, 1988.
- [139] Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *ACL (System Demonstrations)*, pages 55–60, 2014.
- [140] Takahiko Masuda, Phoebe C Ellsworth, Batja Mesquita, Janxin Leu, Shigehito Tanida, and Ellen Van de Veerdonk. Placing the face in context: cultural differences

- in the perception of facial emotion. *Journal of personality and social psychology*, 94(3):365, 2008.
- [141] Andrew McCallum. Efficiently inducing features of conditional random fields. In *Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence*, pages 403–410. Morgan Kaufmann Publishers Inc., 2002.
- [142] Bruce S McEwen. Physiology and neurobiology of stress and adaptation: central role of the brain. *Physiological reviews*, 87(3):873–904, 2007.
- [143] Gary McKeown, Michel Valstar, Roddy Cowie, Maja Pantic, and Marc Schroder. The semaine database: annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Transactions on Affective Computing*, 3(1):5–17, 2012.
- [144] Albert Mehrabian. Communication without words. *Communication theory*, pages 193–200, 2008.
- [145] Yelena Mejova and Padmini Srinivasan. Exploring feature definition and selection for sentiment classifiers. In *Proceeding of the International Conference on Web and Social Media (ICWSM)*, 2011.
- [146] Yishu Miao, Lei Yu, and Phil Blunsom. Neural variational inference for text processing. In *International Conference on Machine Learning (ICML)*, pages 1727–1736, 2016.
- [147] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.
- [148] George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.
- [149] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.
- [150] Olof Mogren. C-rnn-gan: Continuous recurrent neural networks with adversarial training. *arXiv preprint arXiv:1611.09904*, 2016.
- [151] Marcello Mortillaro, Ben Meuleman, and Klaus R Scherer. Advocating a componential appraisal model to guide emotion recognition. *International Journal of Synthetic Emotions (IJSE)*, 3(1):18–32, 2012.
- [152] Orval Mowrer. Learning theory and behavior. 1960.
- [153] Lawrence R Murphy. Stress management in work settings: a critical review of the health effects. *American Journal of Health Promotion*, 11(2):112–135, 1996.

- [154] Henry Alexander Murray and Christiana D Morgan. A clinical study of sentiments (i & ii). *Genetic Psychology Monographs*, 1945.
- [155] Iain R Murray and John L Arnott. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, 93(2):1097–1108, 1993.
- [156] Fatma Nasoz, Kaye Alvarez, Christine L Lisetti, and Neal Finkelstein. Emotion recognition from physiological signals using wireless sensors for presence technologies. *Cognition, Technology & Work*, 6(1):4–14, 2004.
- [157] Catherine Neubauer, Joshua Woolley, Peter Khooshabeh, and Stefan Scherer. Getting to know you: a multimodal investigation of team behavior and resilience to stress. In *Proceedings of the 18th International Conference on Multimodal Interaction (ICMI)*, pages 193–200. ACM, 2016.
- [158] David B Newlin and Robert W Levenson. Pre-ejection period: Measuring beta-adrenergic influences upon the heart. *Psychophysiology*, 16(6):546–552, 1979.
- [159] Keith Oatley, Dacher Keltner, and Jennifer M Jenkins. *Understanding emotions*. Blackwell publishing, 2006.
- [160] Bruno A Olshausen and David J Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision research*, 37(23):3311–3325, 1997.
- [161] Andrew Ortony, Gerald L Clore, and Allan Collins. *The cognitive structure of emotions*. Cambridge university press, 1990.
- [162] Andrew Ortony and Terence J Turner. What’s basic about basic emotions? *Psychological review*, 97(3):315, 1990.
- [163] Bo Pang and Lillian Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, page 271. Association for Computational Linguistics, 2004.
- [164] Bo Pang and Lillian Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd annual meeting on association for computational linguistics*, pages 115–124. Association for Computational Linguistics, 2005.
- [165] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135, 2008.

- [166] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics, 2002.
- [167] M. Pantic and M.S. Bartlett. Machine analysis of facial expressions. In *Face Recognition*, pages 377–416. I-Tech Education and Publishing, 2007.
- [168] Maja Pantic, George Caridakis, Elisabeth André, Jonghwa Kim, Kostas Karpouzis, and Stefanos Kollias. Multimodal emotion recognition from low-level cues. In *Emotion-Oriented Systems*, pages 115–132. Springer, 2011.
- [169] Maja Pantic, Anton Nijholt, Alex Pentland, and Thomas S Huanag. Human-centred intelligent human? computer interaction (hci²): how far are we from attaining it? *International Journal of Autonomous and Adaptive Communications Systems*, 1(2):168–187, 2008.
- [170] Fuchun Peng and Andrew McCallum. Information extraction from research papers using conditional random fields. *Information processing & management*, 42(4):963–979, 2006.
- [171] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.
- [172] Rosalind W. Picard. Affective computing. *The MIT Press*, 1997.
- [173] Rosalind W Picard and Roalind Picard. *Affective computing*, volume 252. MIT press Cambridge, 1997.
- [174] RW Picard. Future affective technology for autism and emotion communication. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, 364(1535):3575–3584, 2009.
- [175] R Plutchik. A general psychoevolutionary theory of emotion in r. plutchik & h. kellerman (eds.) emotion: Theory, research, and experience (vol. 1, pp. 189-217), 1980.
- [176] Robert Plutchik. The nature of emotions human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 89(4):344–350, 2001.
- [177] Soujanya Poria, Erik Cambria, and Alexander Gelbukh. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, 108:42–49, 2016.

- [178] L Melita Prati, Douglas Ceasar, Gerald R Ferris, Anthony P Ammeter, and M Ronald Buckley. Emotional intelligence, leadership effectiveness, and team outcomes. *International Journal of Organizational Analysis*, 11(1):21, 2003.
- [179] Lizhen Qu, Georgiana Ifrim, and Gerhard Weikum. The bag-of-opinions method for review rating prediction from sparse text patterns. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 913–921. Association for Computational Linguistics (ACL), 2010.
- [180] Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [181] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
- [182] Lamia Rahman, Nabeel Mohammed, and Abul Kalam Al Azad. A new lstm model by introducing biological cell state. In *the 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, pages 1–6. IEEE, 2016.
- [183] Antoine Raux and Maxine Eskenazi. A finite-state turn-taking model for spoken dialog systems. In *Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 629–637. ACL, 2009.
- [184] Diego Reforgiato Recupero, Valentina Presutti, Sergio Consoli, Aldo Gangemi, and Andrea Giovanni Nuzzolese. Sentilo: frame-based sentiment analysis. *Cognitive Computation*, 7(2):211–225, 2015.
- [185] Andi Rexha, Mark Kröll, Mauro Dragoni, and Roman Kern. Exploiting propositions for opinion mining. In *Semantic Web Evaluation Challenge*, pages 121–125. Springer, 2016.
- [186] Andi Rexha, Mark Kröll, Mauro Dragoni, and Roman Kern. Polarity classification for target phrases in tweets: A word2vec approach. In *International Semantic Web Conference*, pages 217–223. Springer, 2016.
- [187] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *International Conference on Machine Learning (ICML)*, pages 1278–1286, 2014.
- [188] M Richter and GH Gendolla. The heart contracts to reward: monetary incentives and pre-ejection period. *Psychophysiology*, 46(3):451, 2009.

- [189] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.
- [190] James A Russell. Affective space is bipolar. *Journal of personality and social psychology*, 37(3):345, 1979.
- [191] James A Russell. Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145, 2003.
- [192] James A Russell and Albert Mehrabian. Evidence for a three-factor theory of emotions. *Journal of research in Personality*, 11(3):273–294, 1977.
- [193] Cheryl L Rusting. Personality, mood, and cognitive processing of emotional information: three conceptual frameworks. *Psychological bulletin*, 124(2):165, 1998.
- [194] Nick S Ryan, Jason Pascoe, and David R Morse. Enhanced reality fieldwork: the context-aware archaeological assistant. In *Computer Applications in Archaeology*, British Archaeological Reports, pages 182–196. Tempus Reparatum, 1998.
- [195] Takaya Saito and Marc Rehmsmeier. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PLoS one*, 10(3):e0118432, 2015.
- [196] Stephen W Salant, Sheldon Switzer, and Robert J Reynolds. Losses from horizontal merger: The effects of an exogenous change in industry structure on cournot-nash equilibrium. *The Quarterly Journal of Economics*, pages 185–199, 1983.
- [197] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. In *Advances in Neural Information Processing Systems*, pages 2234–2242, 2016.
- [198] David Sander, Didier Grandjean, and Klaus R Scherer. A systems approach to appraisal mechanisms in emotion. *Neural networks*, 18(4):317–352, 2005.
- [199] Jyotirmay Sanghvi, Ginevra Castellano, Iolanda Leite, André Pereira, Peter W McOwan, and Ana Paiva. Automatic analysis of affective postures and body motion to detect engagement with a game companion. In *6th ACM International Conference on Human-Robot Interaction (HRI)*, pages 305–311. IEEE, 2011.
- [200] Asad Sayeed. An opinion about opinions about opinions: subjectivity and the aggregate reader. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 691–696, 2013.

- [201] Martin Schels, Michael Glodek, Sascha Meudt, Miriam Schmidt, David Hrabal, R Böck, Steffen Walter, and Friedhelm Schwenker. Multi-modal classifier-fusion for the classification of emotional states in woz scenarios. In *the 1st International Conference on Affective and Pleasurable Design*, pages 5337–5346, 2012.
- [202] Martin Schels and Friedhelm Schwenker. A multiple classifier system approach for facial expressions in image sequences utilizing gmm supervectors. In *the 20th International Conference on Pattern Recognition (ICPR)*, pages 4251–4254. IEEE, 2010.
- [203] Klaus R Scherer. Vocal affect expression: A review and a model for future research. *Psychological bulletin*, 99(2):143, 1986.
- [204] Klaus R Scherer. Psychological models of emotion. *The neuropsychology of emotion*, 137(3):137–162, 2000.
- [205] Klaus R Scherer. Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion: Theory, methods, research*, 92(120):57, 2001.
- [206] Stefan Scherer, Friedhelm Schwenker, and Günther Palm. Emotion recognition from speech using multi-classifier systems and rbf-ensembles. *Speech, Audio, Image and Biomedical Signal Processing using Neural Networks*, pages 49–70, 2008.
- [207] B Schilit, N Adams, and R Want. Context-aware computing applications. In *Proceedings of the First Workshop on Mobile Computing Systems and Applications*, pages 85–90. IEEE, 1994.
- [208] Björn Schuller, Stephan Reiter, Ronald Muller, Marc Al-Hames, Manfred Lang, and Gerhard Rigoll. Speaker independent speech emotion recognition by ensemble classification. In *IEEE International Conference on Multimedia and Expo*, pages 864–867. IEEE, 2005.
- [209] Nicu Sebe, Ira Cohen, and Thomas S Huang. Multimodal emotion recognition. In *Handbook of Pattern Recognition and Computer Vision*, pages 387–409. World Scientific, 2005.
- [210] Mark D Seery. Challenge or threat? cardiovascular indexes of resilience and vulnerability to potential stress in humans. *Neuroscience & Biobehavioral Reviews*, 35(7):1603–1610, 2011.
- [211] Stanislau Semeniuta, Aliaksei Severyn, and Erhardt Barth. A hybrid convolutional variational autoencoder for text generation. *arXiv preprint arXiv:1702.02390*, 2017.

- [212] Fei Sha and Fernando Pereira. Shallow parsing with conditional random fields. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 134–141. Association for Computational Linguistics, 2003.
- [213] Qianzi Shen, Zijian Wang, and Yaoru Sun. Sentiment analysis of movie reviews based on cnn-blstm. In *International Conference on Intelligence Science*, pages 164–171. Springer, 2017.
- [214] Catherine A Shipp and Ludmila I Kuncheva. Relationships between combination methods and measures of diversity in combining classifiers. *Information fusion*, 3(2):135–148, 2002.
- [215] Eric Shouse. Feeling, emotion, affect. *m/c journal*, 8 (6). Retrieved May, 6:2012, 2005.
- [216] Craig A Smith and Richard S Lazarus. Appraisal components, core relational themes, and the emotions. *Cognition & Emotion*, 7(3-4):233–269, 1993.
- [217] Richard Socher, Cliff Chiung-Yu Lin, Andrew Y Ng, and Christopher D Manning. Parsing natural scenes and natural language with recursive neural networks. 2011.
- [218] Marina Sokolova, Nathalie Japkowicz, and Stan Szpakowicz. Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. In *Australasian joint conference on artificial intelligence*, pages 1015–1021. Springer, 2006.
- [219] Marina Sokolova and Guy Lapalme. Learning opinions in user-generated web content. *Natural Language Engineering*, 17(4):541–567, 2011.
- [220] Radu Soricut and Daniel Marcu. Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 149–156. Association for Computational Linguistics, 2003.
- [221] Derek P Spangler and Bruce H Friedman. Effortful control and resiliency exhibit different patterns of cardiac autonomic control. *International Journal of Psychophysiology*, 96(2):95–103, 2015.
- [222] Franklin Stein. Occupational stress, relaxation therapies, exercise and biofeedback. *Work: A Journal of Prevention, Assessment and Rehabilitation*, 17(3):235–245, 2001.
- [223] Philip J Stone, Dexter C Dunphy, and Marshall S Smith. The general inquirer: A computer approach to content analysis. 1966.

- [224] Carlo Strapparava, Alessandro Valitutti, et al. Wordnet affect: an affective extension of wordnet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC)*, volume 4, pages 1083–1086, 2004.
- [225] Sandeep Subramanian, Sai Rajeswar, Francis Dutil, Chris Pal, and Aaron Courville. Adversarial generation of natural language. In *Proceedings of the 2nd Workshop on Representation Learning for NLP*, pages 241–251, 2017.
- [226] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307, 2011.
- [227] Maite Taboada, Kimberly Voll, and Julian Brooke. Extracting sentiment as a function of discourse structure and topicality. *Simon Fraser Univeristy School of Computing Science Technical Report*, 2008.
- [228] Oscar Täckström and Ryan McDonald. Discovering fine-grained sentiment with latent variable structured prediction models. In *European Conference on Information Retrieval*, pages 368–374. Springer, 2011.
- [229] Fotios Talantzis, Aristodemos Pnevmatikakis, and Anthony G Constantinides. Audio–visual active speaker tracking in cluttered indoors environments. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 38(3):799–807, 2008.
- [230] Duyu Tang, Bing Qin, and Ting Liu. Aspect level sentiment classification with deep memory network. *arXiv preprint arXiv:1605.08900*, 2016.
- [231] Lucas Theis, Aäron van den Oord, and Matthias Bethge. A note on the evaluation of generative models. *arXiv preprint arXiv:1511.01844*, 2015.
- [232] Peggy A Thoits. The sociology of emotions. *Annual review of sociology*, 15(1):317–342, 1989.
- [233] Joe Tomaka, Jim Blascovich, Jeffrey Kibler, and John M Ernst. Cognitive and physiological antecedents of threat and challenge appraisal. *Journal of Personality and Social Psychology*, 73(1):63, 1997.
- [234] Angela Charng-Rurng Tsai, Chi-En Wu, Richard Tzong-Han Tsai, Jane Yung-jen Hsu, et al. Building a concept-level sentiment dictionary based on commonsense knowledge. *Intelligent Systems*, 28(2):22–30, 2013.
- [235] Peter D Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics, 2002.

- [236] Laurens Van Der Maaten. Accelerating t-sne using tree-based algorithms. *The Journal of Machine Learning Research*, 15(1):3221–3245, 2014.
- [237] Jason Van Hulse, Taghi M Khoshgoftaar, and Amri Napolitano. Experimental perspectives on learning from imbalanced data. In *Proceedings of the 24th International Conference on Machine Learning (ICML)*, pages 935–942. ACM, 2007.
- [238] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*, pages 1096–1103. ACM, 2008.
- [239] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11(Dec):3371–3408, 2010.
- [240] Alessandro Vinciarelli, Maja Pantic, and Hervé Bourlard. Social signal processing: Survey of an emerging domain. *Image and Vision Computing*, 27:1743–1759, 2009.
- [241] A. Vlachostergiou, G. Caridakis, Ph. Mylonas, and A. Stafylopatis. Learning representations of natural language texts with generative adversarial networks at document, sentence, and aspect level. 2018.
- [242] Aggeliki Vlachostergiou, George Caridakis, and Stefanos Kollias. Context in affective multiparty and multimodal interaction: Why, which, how and where? In *Proceedings of the 2014 workshop on Understanding and Modeling Multiparty, Multimodal Interactions*, pages 3–8. ACM, 2014.
- [243] Aggeliki Vlachostergiou, George Caridakis, and Stefanos Kollias. Context in affective multiparty and multimodal interaction: Why, which, how and where? pages 3–8. Workshop on Understanding and Modeling Multiparty, Multimodal Interactions (UMMMI 2014), 2014.
- [244] Aggeliki Vlachostergiou, George Caridakis, and Stefanos Kollias. Investigating context awareness of affective computing systems: a critical approach. *Procedia Computer Science*, 39:91–98, 2014.
- [245] Aggeliki Vlachostergiou, George Caridakis, Amaryllis Raouzaïou, and Stefanos Kollias. Hci and natural progression of context-related questions. In *International Conference on Human-Computer Interaction*, pages 530–541. Springer, 2015.
- [246] Aggeliki Vlachostergiou, Mark Dennison, Catherine Neubauer, Stefan Scherer, Peter Khooshabeh, and Andre Harrison. Unfolding the external behavior and inner

- affective state of teammates through ensemble learning: Experimental evidence from a dyadic team corpus. 2018.
- [247] Aggeliki Vlachostergiou, George Marandianos, and Stefanos Kollias. Context incorporation using context—aware language features. In *Proceeding of the 25th European Signal Processing Conference (EUSIPCO)*, pages 568–572. IEEE, 2017.
- [248] Aggeliki Vlachostergiou, George Marandianos, and Stefanos Kollias. From conditional random field (crf) to rhetorical structure theory (rst): Incorporating context information in sentiment analysis. In *European Semantic Web Conference*, pages 283–295. Springer, 2017.
- [249] Jin Wang, Liang-Chih Yu, K Robert Lai, and Xuejie Zhang. Dimensional sentiment analysis using a regional cnn-lstm model. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 225–230, 2016.
- [250] Peng Wang, Jiaming Xu, Bo Xu, Chenglin Liu, Heng Zhang, Fangyuan Wang, and Hongwei Hao. Semantic clustering and convolutional neural network for short text categorization. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, volume 2, pages 352–357, 2015.
- [251] Xingyou Wang, Weijie Jiang, and Zhiyong Luo. Combination of convolutional and recurrent neural network for sentiment analysis of short texts. In *Proceedings of the 26th International Conference on Computational Linguistics (COLING): Technical Papers*, pages 2428–2437, 2016.
- [252] Yequan Wang, Minlie Huang, Li Zhao, et al. Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 606–615, 2016.
- [253] Cynthia Whissell. The dictionary of affect in language. *Emotion: Theory, research, and experience*, 4(113-131):94, 1989.
- [254] Janyce Wiebe and Ellen Riloff. Creating subjective and objective sentence classifiers from unannotated texts. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 486–497. Springer, 2005.
- [255] Janyce Wiebe, Theresa Wilson, and Claire Cardie. Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2-3):165–210, 2005.
- [256] Anna Wierzbicka. *Emotions across languages and cultures: Diversity and universals*. Cambridge University Press, 1999.

- [257] Martin Wöllmer, Florian Eyben, Björn Schuller, Ellen Douglas-Cowie, and Roddy Cowie. Data-driven clustering in emotional space for affect recognition using discriminatively trained lstm networks. In *Proc. Interspeech 2009, Brighton, UK*, pages 1595–1598, 2009.
- [258] Martin Wöllmer, Florian Eyben, Björn W. Schuller, and Gerhard Rigoll. Temporal and situational context modeling for improved dominance recognition in meetings. In *Proceedings of Interspeech*, pages 350–353. ISCA, 2012.
- [259] David H Wolpert. Stacked generalization. *Neural Networks*, 5(2):241–259, 1992.
- [260] Dongrui Wu, Thomas D Parsons, Emily Mower, and Shrikanth Narayanan. Speech emotion estimation in 3d space. In *Proceeding of the International Conference on Multimedia and Expo (ICME)*, pages 737–742. IEEE, 2010.
- [261] Alec Yenter and Abhishek Verma. Deep cnn-lstm with combined kernels from multiple branches for imdb review sentiment analysis. In *8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*, pages 540–546. IEEE, 2017.
- [262] Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. Seqgan: Sequence generative adversarial nets with policy gradient. In *AAAI*, pages 2852–2858, 2017.
- [263] Yang Yu, Zhiqiang Gong, Ping Zhong, and Jiaxin Shan. Unsupervised representation learning with deep convolutional neural network for remote sensing images. In *International Conference on Image and Graphics*, pages 97–108. Springer, 2017.
- [264] Robert B Zajonc. Feeling and thinking: Preferences need no inferences. *American psychologist*, 35(2):151, 1980.
- [265] Zhihong Zeng, Yuxiao Hu, Glenn I Roisman, Zhen Wen, Yun Fu, and Thomas S Huang. Audio-visual spontaneous emotion recognition. In *Artificial intelligence for human computing*, pages 72–90. Springer, 2007.
- [266] Zhihong Zeng, Maja Pantic, Glenn I Roisman, and Thomas S Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):39–58, 2009.
- [267] Zhihong Zeng, Zhenqiu Zhang, Brian Pianfetti, Jilin Tu, and Thomas S Huang. Audio-visual affect recognition in activation-evaluation space. In *Proceedings of the International Conference on Multimedia and Expo, 2005. (ICME)*, pages 4–pp. IEEE, 2005.

- [268] Lei Zhang and Bing Liu. Extracting resource terms for sentiment analysis. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 1171–1179, 2011.
- [269] Yizhe Zhang, Zhe Gan, and Lawrence Carin. Generating text via adversarial training. In *NIPS workshop on Adversarial Training*, volume 21, 2016.
- [270] Junbo Zhao, Michael Mathieu, and Yann LeCun. Energy-based generative adversarial network. *arXiv preprint arXiv:1609.03126*, 2016.
- [271] Andreas Zimmermann, Andreas Lorenz, and Reinhard Oppermann. An operational definition of context. In *Modeling and Using Context*, pages 558–571. Springer, 2007.
- [272] Cécilia Zirn, Mathias Niepert, Heiner Stuckenschmidt, and Michael Strube. Fine-grained sentiment analysis with structural features. In *IJCNLP*, pages 336–344, 2011.