ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Πτυχιακή Εργασία

Αναγνώριση Κειμενικής Συνεπαγωγής με Βαθιά Μάθηση

Ιωάννης – Ίων Κουτσικάκης

Επιβλέπων: Ίων Ανδρουτσόπουλος

Βοηθός επιβλέπων: Πρόδρομος Μαλακασιώτης

Περιεχόμενα

1 Εισαγωγή 1
1.1 Αντικείμενο της εργασίας 1
1.2 Διάρθρωση της εργασίας 2
1.3 Ευχαριστίες 2
2 Βιβλιογραφική επισκόπηση 4
3 Τα μοντέλα της εργασίας 6
3.1 Μοντέλα Κωδικοποίησης Πρότασης 6
3.1.1 Κωδικοποίηση με Κεντροειδή – CENT 7
3.1.2 Κωδικοποίηση με Ανατροφοδοτούμενα Νευρωνικά Δίκτυα - RNNENC 8
3.2 Αποσυντιθέμενο Μοντέλο Προσοχής - DAM 10
4 Πειραματική αξιολόγηση 15
4.1 Δεδομένα 15
4.2 Λεπτομέρειες υλοποίησης 18
4.3 Αποτελέσματα 20
4.3.1 Μοντέλο Κωδικοποίησης Πρότασης με Κεντροειδή – CENT 20
4.3.2 Μοντέλο Κωδικοποίησης Πρότασης με Ανατροφοδοτούμενα Νευρωνικά Δίκτυα - RNNENC 21
4.3.3 Αποσυντιθέμενο Μοντέλο Προσοχής - DAM 23
5 Συμπεράσματα 26
6 Αναφορές 28

1 Εισαγωγή

1.1 Αντικείμενο της εργασίας

Σκοπός της εργασίας αυτής είναι η μελέτη, η υλοποίηση και η επαλήθευση των αποτελεσμάτων διαφόρων μοντέλων Βαθιάς Μάθησης (Deep Learning), που έχουν προταθεί πρόσφατα, για την Αναγνώριση της Κειμενικής Συνεπαγωγής (Recognizing Textual Entailment). Τα μοντέλα αυτά δέχονται ως είσοδο ένα ζευγάρι προτάσεων, τις οποίες ονομάζουμε προϋπόθεση (premise) και υπόθεση (hypothesis) αντίστοιχα, και το κατατάσσουν στην κατηγορία συνεπαγωγής (entailment) όταν η υπόθεση αποτελεί συμπέρασμα της προϋπόθεσης, στην κατηγορία αντίφασης (contradiction) όταν η υπόθεση αποτελεί αντίφαση της προϋπόθεσης ή στην κατηγορία ουδετερότητας (neutral) όταν η υπόθεση δε σχετίζεται σημασιολογικά με την προϋπόθεση.

Προϋπόθεση	*Υπόθεση*	*Κατηγορία*
Αν κλέψεις θα τιμωρηθείς από τη δικαιοσύνη.	Αν κλέψεις θα τιμωρηθείς.	συνεπαγωγή
Αν κλέψεις θα τιμωρηθείς από τη δικαιοσύνη.	Αν κλέψεις δεν θα έχεις καμία επίπτωση.	αντίφαση
Αν κλέψεις θα τιμωρηθείς από τη δικαιοσύνη.	Κινηματογραφική ληστεία με ομηρία σε υποκατάστημα τράπεζας.	ουδετερότητα

Πίνακας 1.1-1 Παραδείγματα Αναγνώρισης Κειμενικής Συνεπαγωγής.

Η κατασκευή μιας αξιόπιστης μεθόδου που να επιτυγχάνει υψηλή ακρίβεια (accuracy) στο πρόβλημα της Αναγνώρισης Κειμενικής Συνεπαγωγής είναι σημαντική, καθώς τέτοιες μέθοδοι μπορούν να χρησιμοποιηθούν σε πολλές εφαρμογές. Τα Σύστηματα Ερωταποκρίσεων (Question Answering) επιστρέφουν συνήθως πολλές υποψήφιες απαντήσεις. Χρησιμοποιώντας ένα σύστημα Αναγνώρισης Κειμενικής Συνεπαγωγής μπορούμε να βαθμολογήσουμε κάθε υποψήφια απάντηση και να επιλέξουμε την καλύτερη. Για να το πετύχουμε αυτό χρησιμοποιούμε κάθε δυνατή απάντηση ως προϋπόθεση και μετατρέπουμε κάθε ερώτηση σε δήλωση την οποία χρησιμοποιούμε ως υπόθεση. Έτσι, το πρόβλημα Ερωταποκρίσεων μετατρέπεται σε πρόβλημα Αναγνώρισης Κειμενικής Συνεπαγωγής [Har06]. Για παράδειγμα, μετατρέπουμε την ερώτηση 'Ποιος ζωγράφισε

τη Μόνα Λίζα;' στη δήλωση 'Κάποιος ζωγράφισε τη Μόνα Λίζα.' και αναγνωρίζουμε ότι η υποψήφια απάντηση 'Ο Λεονάρντο ντα Βίντσι ζωγράφισε τη Μόνα Λίζα' συνεπάγεται τη δήλωση αυτή. Επίσης, σε ένα σύστημα Κατασκευής Περιλήψεων (Summarization) το οποίο παράγει υποψήφιες προτάσεις της περίληψης, μπορούμε να χρησιμοποιήσουμε ένα σύστημα Αναγνώρισης Κειμενικής Συνεπαγωγής προκειμένου να αποφύγουμε την επανάληψη των ίδιων πληροφοριών [Llo08]. Για να το πετύχουμε αυτό, πριν συμπεριλάβουμε μια πρόταση στην κατασκευαζόμενη περίληψη, ελέγχουμε αν από την ήδη υπάρχουσα περίληψη μπορούμε να βγάλουμε το συμπέρασμα ότι ισχύει η υποψήφια για ένταξη στην περίληψη πρόταση. Σε περίπτωση που ισχύει, τότε δεν την συμπεριλαμβάνουμε καθώς θα ήταν πλεονασμός. Τέλος, σε ένα σύστημα Μηχανικής Μετάφρασης (Machine Translation) υπάρχουν συνήθως πολλές υποψήφιες μεταφράσεις για μια πρόταση. Μπορούμε λοιπόν να χρησιμοποιήσουμε ένα σύστημα Αναγνώρισης Κειμενικής Συνεπαγωγής προκειμένου να επιλέξουμε τη μετάφραση που ταιριάζει περισσότερο. Για να το πετύχουμε αυτό, ελέγχουμε, για κάθε υποψήφια μετάφραση, κατά πόσο αποτελεί συμπέρασμα (entailment) της αρχικής πρότασης και το αντίστροφο [Pad09].

1.2 Διάρθρωση της εργασίας

Στο δεύτερο κεφάλαιο αναφέρονται συνοπτικά οι νέες τάσεις για την αντιμετώπιση του προβλήματος της Αναγνώρισης Κειμενικής Συνεπαγωγής. Στο τρίτο κεφάλαιο περιγράφονται εκτενώς τα μοντέλα που υλοποιήθηκαν. Στο τέταρτο κεφάλαιο περιγράφονται τα πειράματα, τα αποτελέσματά τους καθώς και το σώμα κειμένων (corpus) που χρησιμοποιήθηκε για την εκπαίδευση, την επικύρωση και την εξέταση των μοντέλων. Τέλος, στο πέμπτο κεφάλαιο αναφέρονται τα συμπεράσματα που εξήχθησαν από τα αποτελέσματα των πειραμάτων.

1.3 Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Ίωνα Ανδρουτσόπουλο για την καθοδήγηση, τις συμβουλές και την ευκαιρία

που μου έδωσε να ασχοληθώ με ένα γνωστικό αντικείμενο εξαιρετικού ενδιαφέροντος. Επίσης, ευχαριστώ το Μεταδιδακτορικό Ερευνητή Μάκη Μαλακασιώτη για τη συνεχή επίβλεψη και τις ιδέες που μου παρείχε κατά τη διάρκεια της εκπόνησης της εργασίας. Τέλος, ευχαριστώ το Μεταδιδακτορικό Ερευνητή Γιάννη Παυλόπουλο για τις ουσιαστικές συζητήσεις που είχαμε πάνω σε θέματα της εργασίας.

2 Βιβλιογραφική επισκόπηση

Μέχρι πρότινος, οι μέθοδοι αναγνώρισης Κειμενικής Συνεπαγωγής χρησιμοποιούσαν σε μεγάλο βαθμό μεθόδους εξαγωγής χειρωνακτικά κατασκευασμένων χαρακτηριστικών (hand-crafted feature extraction) και μεθόδους επιβλεπόμενης μηχανικής μάθησης βασισμένες κυρίως σε γραμμικούς διαχωριστές, για τον προσδιορισμό της μορφολογικής, συντακτικής και σημασιολογικής ομοιότητας των δύο προτάσεων [And10]. Πρόσφατα, όμως, χάρη στην αύξηση της υπολογιστικής δύναμης των υπολογιστών καθώς και του μεγάλου όγκου πληροφορίας που υπάρχει στο Διαδίκτυο, η χρήση Βαθιών Νευρωνικών Δικτύων (Deep Neural Networks) είναι πλέον δυνατή και δίνει τη δυνατότητα κατασκευής πιο περίπλοκων μη γραμμικών μοντέλων τα οποία συχνά επιτυγχάνουν εντυπωσιακές επιδόσεις. Έτσι, υπάρχουν πολλές πρόσφατες δημοσιεύσεις στις οποίες παρουσιάζονται ποικίλες αρχιτεκτονικές βασισμένες στη Βαθιά Μάθηση (Deep Learning)[Goo16] οι οποίες έχουν ξεπεράσει τις συμβατικές μεθόδους Αναγνώρισης Κειμενικής Συνεπαγωγής.[1]

Πολλά από τα μοντέλα αυτά ανήκουν στην οικογένεια των Μοντέλων Κωδικοποίησης Πρότασης (Sentence Encoding Models). Τα μοντέλα της κατηγορίας αυτής κωδικοποιούν τις προτάσεις ως διανύσματα, στοχεύοντας να συλλάβουν τη σημασιολογική πληροφορία κάθε πρότασης. Η κωδικοποίηση αυτή γίνεται συνήθως με χρήση κάποιου Νευρωνικού Δικτύου όπως, για παράδειγμα, με Ανατροφοδοτούμενο Νευρωνικό Δίκτυο (Recurrent Neural Network) [Bow15] ή με Αμφίδρομο Ανατροφοδοτούμενο Νευρωνικό Δίκτυο (Biderictional Recurrent Neural Network) [Liu16] ή με Συνελικτικό Νευρωνικό Δίκτυο (Convolutional Neural Network) [Mou16].

Επίσης, μια τεχνική που δίνει πολλά υποσχόμενα αποτελέσματα είναι αυτή της Προσοχής (Attention). Πρόκειται ουσιαστικά για έναν μηχανισμό ο οποίος υποδεικνύει στο μοντέλο σε ποιες λέξεις των προτάσεων να εστιάσει, είτε κατά τη φάση της κωδικοποίησης της πρότασης στα Μοντέλα Κωδικοποίησης Πρότασης (Sentence Encoding

[1] Βλ. π.χ. τον κατάλογο δημοσιεύσεων στη διεύθυνση http://nlp.stanford.edu/projects/snli

Models) [Liu16], είτε κατά την φάση εξαγωγής χαρακτηριστικών (feature extraction) σε άλλα μοντέλα [Roc15] [Wan16] [Par16].

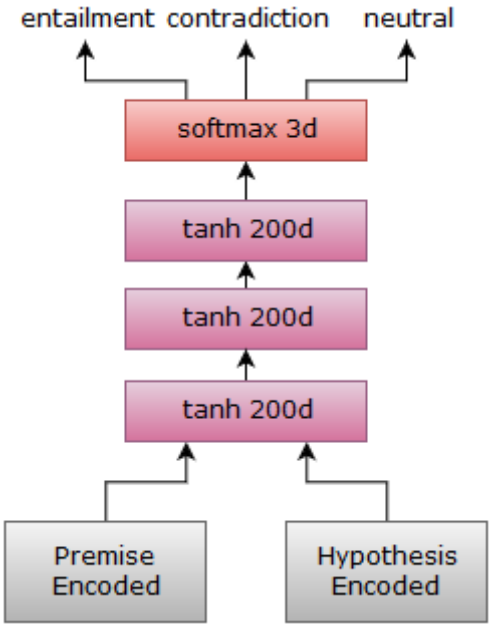
Στην εργασία αυτή, υλοποιήθηκαν δύο Μοντέλα Κωδικοποίησης Πρότασης και ένα μοντέλο που χρησιμοποιεί μηχανισμό Προσοχής, όπως θα δούμε στο επόμενο κεφάλαιο.

3 Τα μοντέλα της εργασίας

Στο πλαίσιο της εργασίας αυτής υλοποιήθηκαν τρία μοντέλα. Κάθε μοντέλο δέχεται ως είσοδο την προϋπόθεση (premise) και την υπόθεση (hypothesis) και έχει ως έξοδο τρεις πιθανότητες που αντιστοιχούν στις κατηγορίες του προβλήματος. Σε όλα τα μοντέλα οι λέξεις αντικαταστάθηκαν με τις αντίστοιχες διανυσματικές παραστάσεις τους (word embeddings) τριακοσίων (300) διαστάσεων, όπως αυτές παρήχθησαν με τον αλγόριθμο GloVe [Pen14].

3.1 Μοντέλα Κωδικοποίησης Πρότασης

Στα Μοντέλα Κωδικοποίησης Πρότασης (Sentence Encoding Models), το πρώτο βήμα είναι η κωδικοποίηση και των δυο προτάσεων σε διανύσματα. Στη συνέχεια, τα διανύσματα συνενώνονται και εισέρχονται σε ένα Νευρωνικό Δίκτυο με προς τα εμπρός τροφοδότηση (feed-forward neural network) το οποίο έχει ως έξοδο τις πιθανότητες που αντιστοιχούν σε κάθε κατηγορία [Bow15].



Εικόνα 3.1-1 Παράδειγμα Νευρωνικού Δικτύου Κωδικοποίησης Πρότασης.

Υποθέτοντας ότι έχουμε έναν τρόπο να κωδικοποιούμε τις προτάσεις, συνενώνουμε τις κωδικοποιήσεις αυτές και τις περνάμε στο Νευρωνικό Δίκτυο (Εικόνα 3.1-1). Το νευρωνικό δίκτυο της Εικόνας 3.1-1 αποτελείται

από τρία κρυφά επίπεδα διακοσίων (200) διαστάσεων τα οποία έχουν ως συνάρτηση ενεργοποίησης την *tanh.*

$$tanh(x) = \frac{1- e^{-2x}}{1+ e^{-2x}} \quad \text{(Σχέση 3.1-1)}$$

Το επίπεδο εξόδου είναι τριών (3) διαστάσεων με συνάρτηση ενεργοποίησης τη *softmax.*

$$softmax(x_i) = \frac{e^{x_i}}{\sum_{k=1}^{K} e^{x_k}} \quad \text{(Σχέση 3.1-2)}$$

Έτσι, οι τρείς τιμές της εξόδου είναι μεταξύ μηδέν και ένα και αθροίζουν στη μονάδα. Επομένως, μπορούν να θεωρηθούν ως οι πιθανότητες το ζευγάρι των προτάσεων να ανήκει σε κάθε κατηγορία.

Στη συνέχεια περιγράφονται τα δύο μοντέλα της κατηγορίας αυτής.

3.1.1 Κωδικοποίηση με Κεντροειδή – CENT

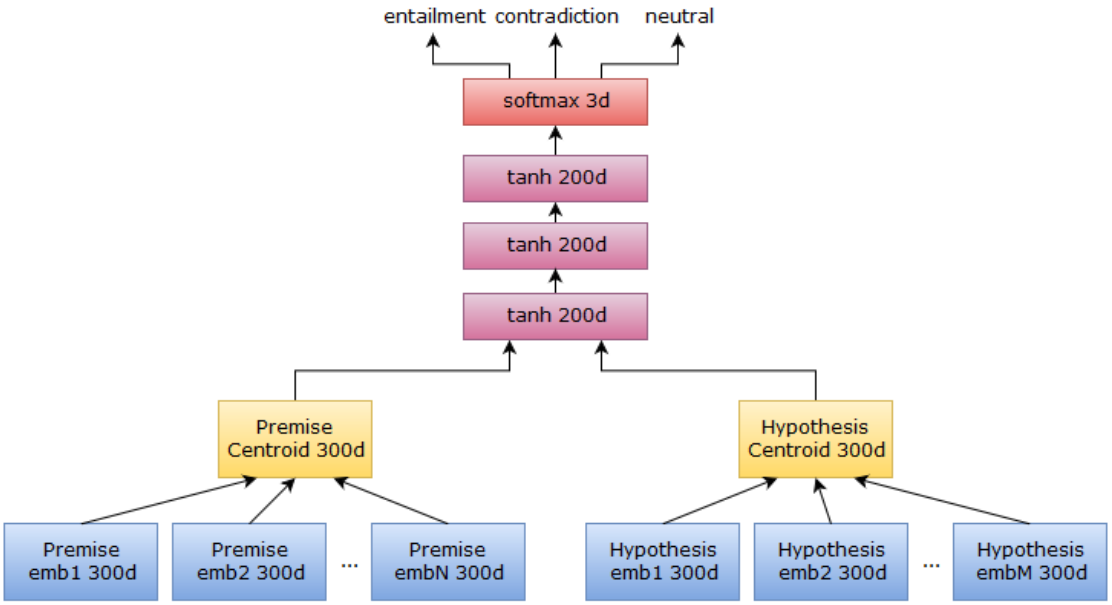
Στο μοντέλο αυτό, οι προτάσεις αντικαθίστανται από τα κεντροειδή τους. Το κεντροειδές υπολογίζεται από τον ακόλουθο τύπο:

$$\vec{t} = \frac{\sum_{j=1}^{|V|} \vec{w_j} \cdot TF(w_j,t) \cdot IDF(w_j)}{\sum_{j=1}^{|V|} TF(w_j,t) \cdot IDF(w_j)} \quad \text{(Σχέση 3.1-3)}$$

όπου:

- $\vec{t}$: το κεντροειδές
- V : το λεξιλόγιο
- $\vec{w_j}$: η διανυσματική παράσταση της λέξης j του λεξιλογίου
- t : η πρόταση
- $TF(w_j, t)$: ο αριθμός των εμφανίσεων της λέξης w_j στην πρόταση t
- $IDF(w_j) = \log\frac{N}{n_j}$ όπου N ο αριθμός των προτάσεων του σώματος κειμένων και n_j ο αριθμός των προτάσεων του σώματος κειμένων που περιέχουν τη λέξη j του λεξιλογίου.

Στην Εικόνα 3.1-2 φαίνεται το ολοκληρωμένο Μοντέλο Κωδικοποίησης Πρότασης που χρησιμοποιεί κεντροειδή. Ν είναι το πλήθος των λέξεων της προϋπόθεσης και Μ το πλήθος των λέξεων της υπόθεσης.



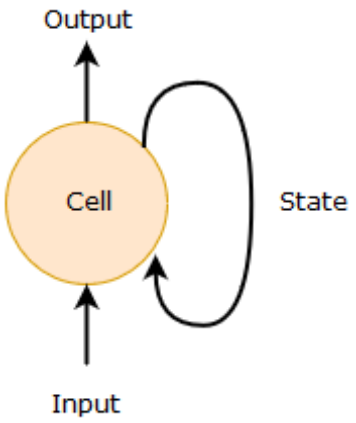
Εικόνα 3.1-2 Μοντέλο Κωδικοποίησης Πρότασης με Κεντροειδή.

3.1.2 Κωδικοποίηση με Ανατροφοδοτούμενα Νευρωνικά Δίκτυα - RNNENC

Στο Μοντέλο Κωδικοποίησης Πρότασης με κεντροειδή, κάθε πρόταση κωδικοποιείται ως το σταθμισμένο άθροισμα των διανυσματικών παραστάσεων (word embeddings) των λέξεων που την απαρτίζουν. Στο σταθμισμένο άθροισμα αυτό, τα βάρη εξαρτώνται από τον αριθμό εμφανίσεων των λέξεων στο σώμα κειμένων καθώς και στην πρόταση αυτή καθαυτή. Αυτή η προσέγγιση έχει το μειονέκτημα ότι δεν λαμβάνεται καθόλου υπόψη η σειρά εμφάνισης των λέξεων. Για να άρουμε αυτόν τον περιορισμό, θα κάνουμε χρήση ενός ειδικού τύπου νευρωνικών δικτύων, των ανατροφοδοτούμενων (recurrent).

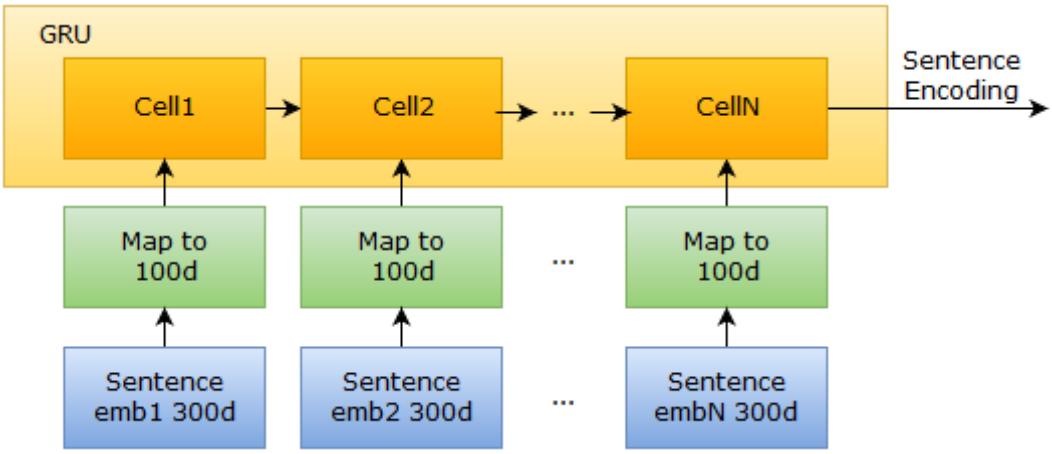
Στα Ανατροφοδοτούμενα Νευρωνικά Δίκτυα (Recurrent Neural Networks) οι συνδέσεις μεταξύ των κρυφών επιπέδων σχηματίζουν έναν κατευθυνόμενο κύκλο (Εικόνα 3.1-3). Αυτό δημιουργεί μια εσωτερική κατάσταση η οποία δίνει τη δυνατότητα στο δίκτυο να παρουσιάζει

συμπεριφορά που εξαρτάται όχι μόνο από την τρέχουσα είσοδο αλλά και από τις προηγούμενες εισόδους. Στην περίπτωση όπου η είσοδος είναι μια πρόταση, καθώς καταναλώνονται σειριακά οι λέξεις της πρότασης, υπάρχει μια εσωτερική κατάσταση η οποία συνοψίζει τις λέξεις που έχουν καταναλωθεί.



Εικόνα 3.1-3 Παράδειγμα Ανατροφοδοτούμενου Νευρωνικού Δικτύου.

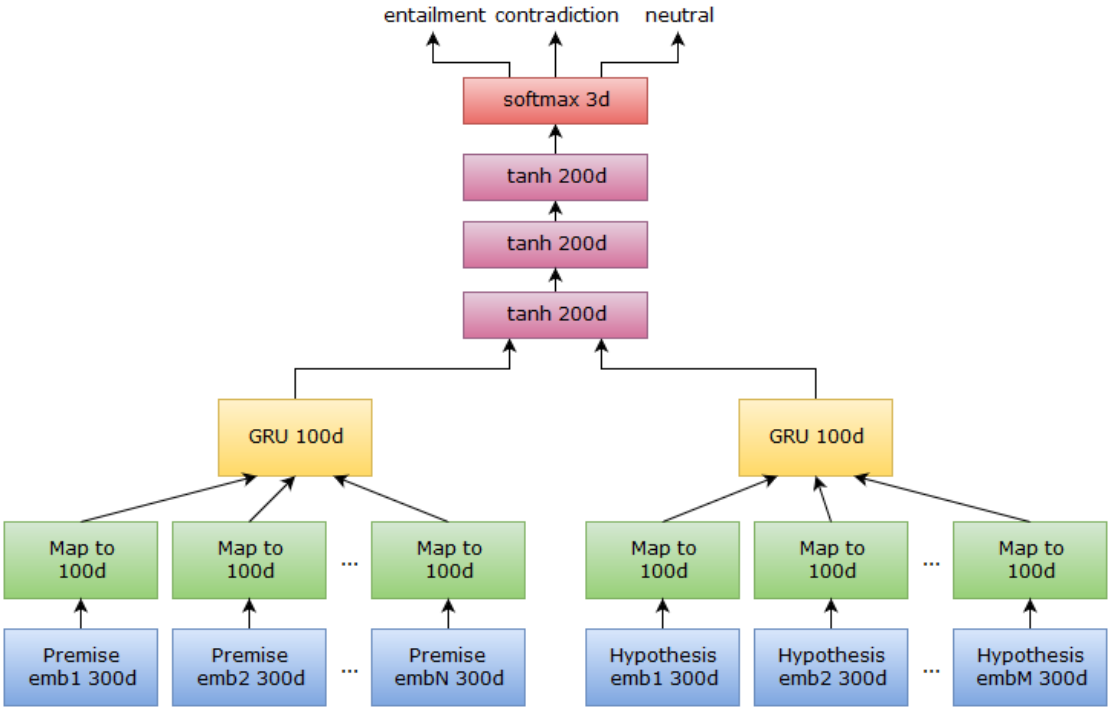
Κατά την εκπαίδευση του κλασικού είδους ανατροφοδοτούμενου νευρωνικού δικτύου, όμως, παρουσιάζεται εντονότερα το πρόβλημα της εξαφανιζόμενης κλίσης (vanishing gradient problem) [Pas13], με αποτέλεσμα οι λέξεις που βρίσκονται στην αρχή της πρότασης να μην επηρεάζουν αρκετά τη μάθηση. Ένας τρόπος για την καταπολέμηση αυτού του προβλήματος είναι η χρήση ειδικών κελιών τύπου GRU [Cho14].



Εικόνα 3.1-4 Παράδειγμα Ανατροφοδοτούμενου Νευρωνικού Δικτύου με κελιά GRU, ξεδιπλωμένο στο χρόνο.

Στο μοντέλο αυτής της ενότητας, που βασίζεται στην εργασία [Bow15], για την κωδικοποίηση των προτάσεων χρησιμοποιούνται δύο

Ανατροφοδοτούμενα Νευρωνικά Δίκτυα (RNN) με κελιά GRU. Αρχικά, γίνεται μια προβολή των διανυσματικών των λέξεων κάθε πρότασης (προϋπόθεση και υπόθεση) από τις τριακόσιες (300) διαστάσεις στις εκατό (100). Στη συνέχεια, τα διανύσματα που προκύπτουν εισέρχονται σειριακά στο αντίστοιχο RNN και ως κωδικοποίηση της κάθε πρότασης χρησιμοποιούμε την τελική κατάσταση του κελιού GRU του αντίστοιχου RNN (Εικόνα 3.1-4). Είναι σημαντικό να σημειωθεί ότι τα δύο RNN (της προϋπόθεσης και της υπόθεσης) μοιράζονται τα βάρη τους. Το ολοκληρωμένο μοντέλο φαίνεται στην Εικόνα 3.1-5.



Εικόνα 3.1-5 Μοντέλο Κωδικοποίησης Πρότασης με Ανατροφοδοτούμενα Νευρωνικά Δίκτυα.

3.2 Αποσυντιθέμενο Μοντέλο Προσοχής - DAM

Το Αποσυντιθέμενο Μοντέλο Προσοχής (Decomposable Attention Model) [Par16] ακολουθεί μια διαφορετική προσέγγιση. Συγκεκριμένα, χωρίζει το πρόβλημα σε υποπροβλήματα, όπου κάθε υποπρόβλημα συγκεντρώνει πληροφορίες για το πόσο και προς ποια κατεύθυνση συμβάλλει μία λέξη των προτάσεων στην κατηγοριοποίηση. Αυτό επιτυγχάνεται ελέγχοντας με ποιες λέξεις της άλλης πρότασης σχετίζεται / ευθυγραμμίζεται (aligns) η λέξη αυτή και αναγνωρίζοντας το είδος της

σχέσης αυτής. Στη συνέχεια, οι πληροφορίες αυτές συγκεντρώνονται και παράγουν την τελική κατηγοριοποίηση.

Αυτή η προσέγγιση είναι αποτελεσματικότερη, κυρίως στις περιπτώσεις όπου μια πρόταση είναι μεγάλη και είναι δύσκολο να συγκεντρωθεί ολόκληρη η σημασία της σε ένα διάνυσμα.

Ας δούμε ένα παράδειγμα:

- *Bob is in his room, but because of the thunder and lightning outside, he cannot sleep.*
- *Bob is awake.*
- *It is sunny outside.*

Η πρώτη πρόταση έχει αρκετά πολύπλοκη δομή και είναι δύσκολο να κατασκευαστεί μια συνοπτική διανυσματική παράσταση που να εκφράζει όλο το νόημά της. Όμως, είναι αρκετά εύκολο να καταλήξουμε στο ότι από την πρώτη πρόταση μπορούμε να συμπεράνουμε τη δεύτερη, παρατηρώντας ότι η λέξη *Bob* ευθυγραμμίζεται με την λέξη *Bob*, ότι η λέξη *awake* ευθυγραμμίζεται με τη φράση *cannot sleep* και αναγνωρίζοντας ότι οι τελευταίες είναι συνώνυμες. Ομοίως, μπορούμε να καταλήξουμε ότι η πρώτη πρόταση είναι αντιφατική ως προς την τρίτη, παρατηρώντας ότι η λέξη *outside* ευθυγραμμίζεται με τη λέξη *outside*, η λέξη *sunny* ευθυγραμμίζεται με τη φράση *thunder and lightning* και αναγνωρίζοντας ότι οι τελευταίες είναι αντιφατικές.

Το μοντέλο αυτό χωρίζεται στα εξής βήματα:

- **Βήμα προσοχής (Attention)**
- **Βήμα σύγκρισης (Comparison)**
- **Βήμα συσσωμάτωσης (Aggregation)**

Παρακάτω θεωρούμε ότι τα **a** = $(a_1, \dots, a_{l_a})$ και **b** = $(b_1, \dots, b_{l_b})$ είναι οι δύο προτάσεις (προϋπόθεση και υπόθεση), με μήκος l_a και l_b αντίστοιχα. Επίσης, θεωρούμε ότι τα a_i, $b_j \in \mathbb{R}^{200}$ είναι διανυσματικές παραστάσεις λέξεων τριακοσίων (300) διαστάσεων οι οποίες έχουν προβληθεί στις διακόσιες (200) διαστάσεις.

Βήμα προσοχής (Attention)

Στο βήμα αυτό, υπολογίζεται ένας πίνακας βαρών, όπου κάθε βάρος υποδεικνύει το κατά πόσο ευθυγραμμίζεται μια λέξη της προϋπόθεσης (premise) με μια λέξη της υπόθεσης (hypothesis). Τα βάρη αυτά, που αποτελούν ένα μηχανισμό προσοχής, υπολογίζονται ως εξής:

$$e_{ij} := F'(a_i, b_j) := F(a_i)^T \cdot F(b_j) \quad \text{(Σχέση 3.2-1)}$$

Η συνάρτηση *F* της Σχέσης 3.2-1 είναι ένα νευρωνικό δίκτυο με προς τα εμπρός τροφοδότηση (feed-forward neural network). Το νευρωνικό δίκτυο αυτό αποτελείται από δύο επίπεδα διακοσίων (200) διαστάσεων με συνάρτηση ενεργοποίησης την *Rectified Linear Unit (ReLU)*.

$$relu(x) = \max(0, x) \quad \text{(Σχέση 3.2-2)}$$

Η αποσύνθεση αυτή της *F'* μας απαλλάσσει από υπολογισμούς καθώς απαιτείται ο υπολογισμός της *F* $l_a + l_b$ φορές. Διαφορετικά, θα χρειαζόταν να υπολογίσουμε την *F'* $l_a \times l_b$ φορές.

Στην συνέχεια, για κάθε λέξη της μιας πρότασης κατασκευάζεται ένα σταθμισμένο άθροισμα των διανυσματικών παραστάσεων λέξεων της άλλης. Το άθροισμα αυτό καλείται υποφράση (subphrase) και αποτελεί το μέρος της σημασιολογικής πληροφορίας της άλλης πρότασης που σχετίζεται / ευθυγραμμίζεται χαλαρά (softly aligned) με τη λέξη αυτή.

$$\beta_i := \sum_{j=1}^{l_b} \frac{\exp(e_{ij})}{\sum_{k=1}^{l_b} \exp(e_{ik})} b_j \quad \text{(Σχέση 3.2-3)}$$

$$\alpha_j := \sum_{i=1}^{l_a} \frac{\exp(e_{ij})}{\sum_{k=1}^{l_a} \exp(e_{kj})} a_i \quad \text{(Σχέση 3.2-4)}$$

Έτσι, στη Σχέση 3.2-3, το β_i είναι η υποφράση της πρότασης **b** που ευθυγραμμίζεται με την λέξη a_i. Αντίστοιχα, στη Σχέση 3.2-4, το α_j είναι η υποφράση της πρότασης **a** που ευθυγραμμίζεται με την λέξη b_j.

Με αυτόν τον τρόπο, το αρχικό πρόβλημα διαιρείται σε υποπροβλήματα σύγκρισης (comparison) μεταξύ των λέξεων και των ευθυγραμμιζόμενων υποφράσεων.

Βήμα σύγκρισης (Comparison)

Στο βήμα αυτό, συγκρίνουμε ξεχωριστά τις ευθυγραμμισμένες υποφράσεις $\{(a_i, \beta_i)\}_{i=1}^{l_a}$ και $\{(b_j, \alpha_j)\}_{j=1}^{l_b}$ με την χρήση της συνάρτησης *G* ως εξής:

$$v_{1,i} := G([a_i; \beta_i]), \forall i \in \{1, \dots, l_a\} \quad \text{(Σχέση 3.2-5)}$$

$$v_{2,j} := G([b_j; \alpha_j]), \forall j \in \{1, \dots, l_b\} \quad \text{(Σχέση 3.2-6)}$$

Οι αγκύλες $[* ; *]$ σημαίνουν συνένωση.

Η συνάρτηση *G,* που εμφανίζεται στη Σχέση 3.2-5 και στη Σχέση 3.2-6, είναι κι αυτή ένα νευρωνικό δίκτυο, πανομοιότυπο με αυτό της *F*.

Είναι σημαντικό να παρατηρήσουμε ότι σε αυτήν την περίπτωση έχουμε γραμμικό, ως προς τον αριθμό των λέξεων, αριθμό όρων, οπότε δεν χρειάζεται να κάνουμε αντίστοιχη αποσύνθεση με εκείνη που κάναμε για την *F'*.

Βήμα συσσωμάτωσης (Aggregation)

Έχουμε πλέον υπολογίσει τα δύο σύνολα διανυσμάτων σύγκρισης, $\{v_{1,i}\}_{i=1}^{l_a}$ και $\{v_{2,j}\}_{j=1}^{l_b}$. Στην συνέχεια, αθροίζουμε τα διανύσματα κάθε συνόλου και υπολογίζουμε το αποτέλεσμα ως εξής:

$$v_1 = \sum_{i=1}^{l_a} v_{1,i} \quad \text{(Σχέση 3.2-7)}$$

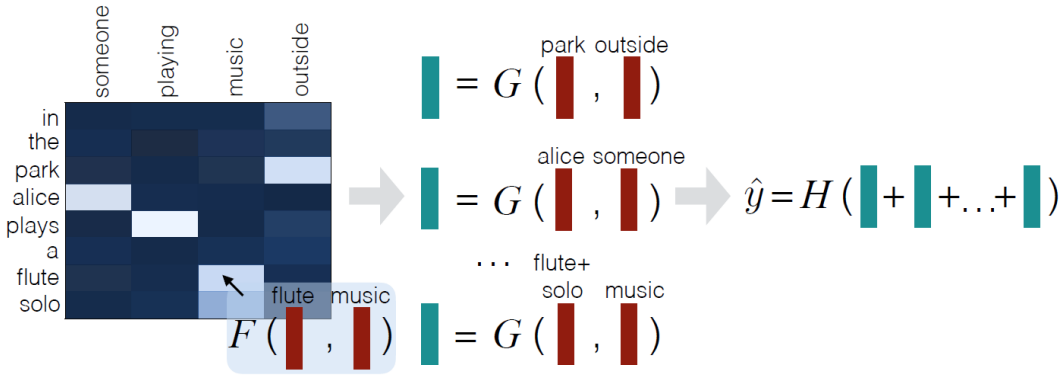
$$v_2 = \sum_{j=1}^{l_b} v_{2,j} \quad \text{(Σχέση 3.2-8)}$$

$$\hat{y} = H([v_1; v_2]) \quad \text{(Σχέση 3.2-9)}$$

Στη Σχέση 3.2-9, η συνάρτηση *H* είναι και αυτή νευρωνικό δίκτυο, όμοιο με αυτό της *F* και της *G*, αλλά έχει και ένα επιπλέον επίπεδο τριών (3) διαστάσεων με συνάρτηση ενεργοποίησης τη softmax.

Το $\hat{y} \in \mathbb{R}^3$ είναι το αποτέλεσμα της κατηγοριοποίησης και περιέχει τις πιθανότητες για κάθε κατηγορία.

Το ολοκληρωμένο μοντέλο φαίνεται στην Εικόνα 3.2-1. Στον πίνακα της Εικόνας 3.2-1, όσο πιο ανοιχτό είναι το χρώμα ενός κελιού τόσο πιο πολύ σχετίζονται οι αντίστοιχες λέξεις μεταξύ τους. Για παράδειγμα, η λέξη “music” σχετίζεται περισσότερο με τις λέξεις “flute” και “solo” απ’ ότι με τις υπόλοιπες λέξεις.



Εικόνα 3.2-1 Αποσυντιθέμενο Μοντέλο Προσοχής.

4 Πειραματική αξιολόγηση

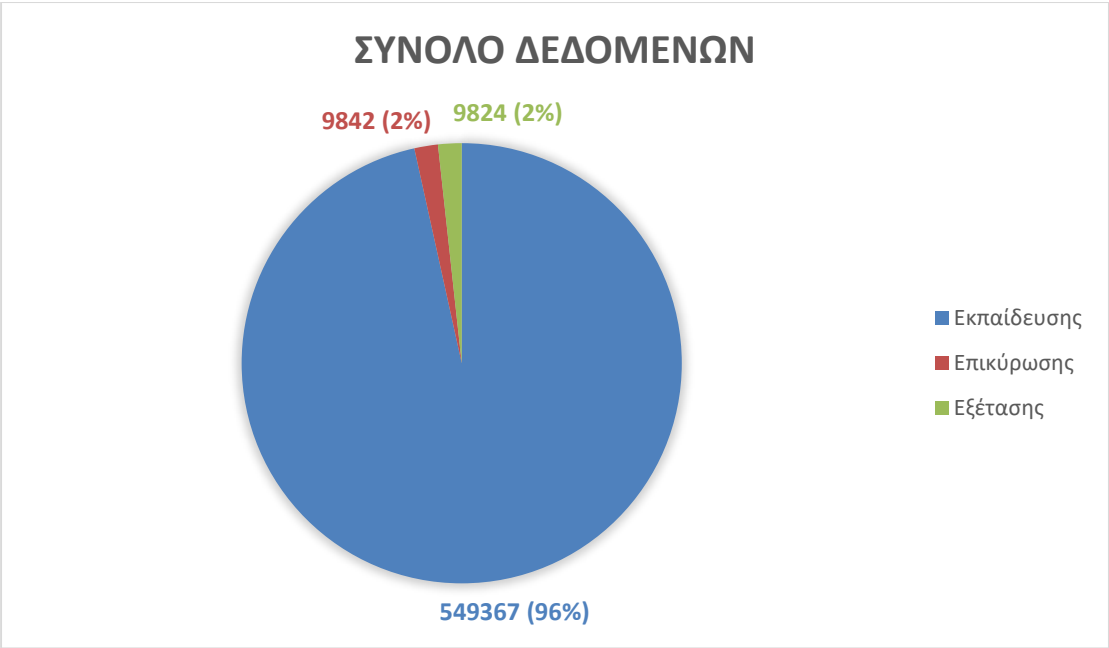
4.1 Δεδομένα

Για την εκπαίδευση (train), επικύρωση (validation) και εξέταση (test) των μοντέλων χρησιμοποιήθηκε το σώμα κειμένων SNLI (Stanford Natural Language Inference) [Bow15].[1] Το σώμα αυτό περιέχει 570.152 παραδείγματα. Κάθε παράδειγμα αποτελείται από μια πρόταση προϋπόθεσης (premise), μια πρόταση υπόθεσης (hypothesis) και τη σωστή κατηγορία. Κάθε παράδειγμα αξιολογήθηκε από πέντε κριτές κατά τη δημιουργία του σώματος και ως σωστή κατηγορία επιλέχθηκε αυτή με τις περισσότερες ψήφους. Η χρήση παραδειγμάτων όπου δεν υπάρχει πλειοψηφούσα κατηγορία δεν ενδείκνυται για μοντέλα αυστηρής κατηγοριοποίησης (hard classification) και ως εκ τούτου αφαιρέθηκαν από το σώμα κειμένων αφήνοντας 569.033 παραδείγματα.

Το σώμα κειμένων είναι ήδη χωρισμένο σε σύνολο εκπαίδευσης (train set), σύνολο επικύρωσης (validation set) και σύνολο εξέτασης (test set) όπως φαίνεται στην Εικόνα 4.1-1. Η κατανομή των παραδειγμάτων κάθε συνόλου στις τρεις κατηγορίες φαίνεται στις Εικόνες 4.1-2, 4.1-3 και 4.1-4 αντίστοιχα.

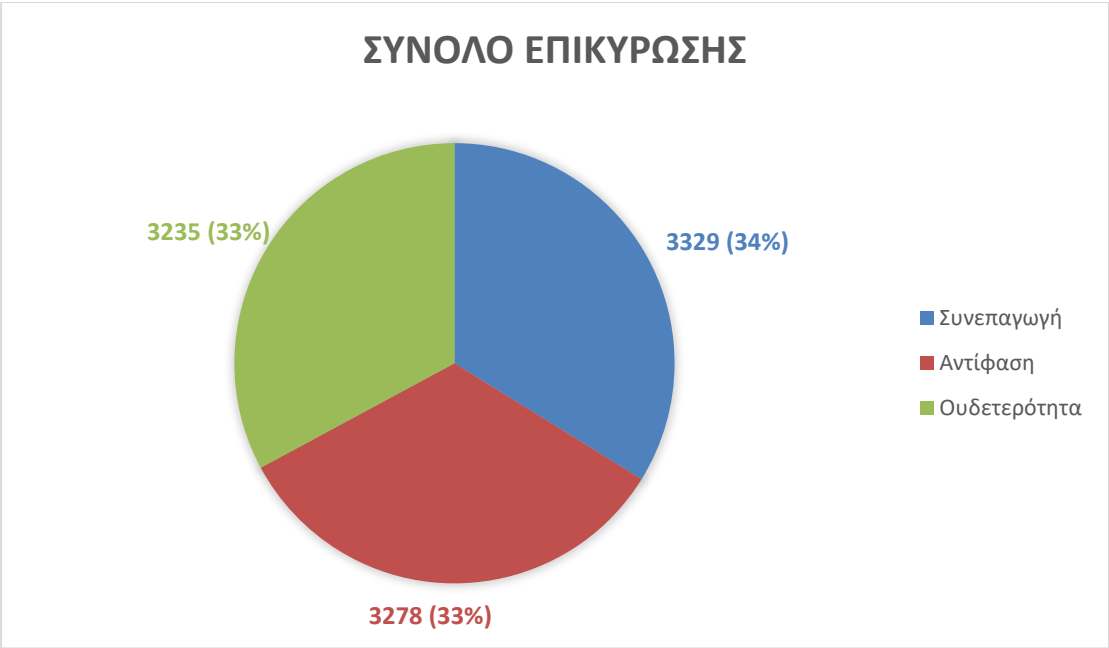
Είναι σημαντικό να σημειωθεί, ότι το σώμα κειμένων αυτό είναι δυο τάξεις μεγαλύτερο από κάθε άλλο παρόμοιου είδους και ότι τα παραδείγματα φτιάχτηκαν εξολοκλήρου από ανθρώπους χωρίς καμία χρήση αυτοματοποιημένων μεθόδων.

[1] http://nlp.stanford.edu/projects/snli



Εικόνα 4.1-1 Κατανομή παραδειγμάτων σε σύνολα δεδομένων.

Εικόνα 4.1-2 Κατανομή του συνόλου εκπαίδευσης σε κατηγορίες.



Εικόνα 4.1-3 Κατανομή του συνόλου επικύρωσης σε κατηγορίες.

Εικόνα 4.1-4 Κατανομή του συνόλου εξέτασης σε κατηγορίες.

Τυχαία επιλεγμένα παραδείγματα φαίνονται στον Πίνακα 4.1-1.

Προϋπόθεση	*Υπόθεση*	*Κατηγορία*
A man inspects the uniform of a figure in some East Asian country.	The man is sleeping	αντίφαση
An older and younger man smiling.	Two men are smiling and laughing at the cats playing on the floor.	ουδετερότητα
A black race car starts up in front of a crowd of people.	A man is driving down a lonely road.	αντίφαση
A soccer game with multiple males playing.	Some men are playing a sport.	συμπερασμός
A smiling costumed woman is holding an umbrella.	A happy woman in a fairy costume holds an umbrella.	ουδετερότητα

Πίνακας 4.1-1 Τυχαία επιλεγμένα παραδείγματα του σώματος κειμένων.

Τα χαρακτηριστικά κάθε συνόλου δεδομένων φαίνονται στον Πίνακα 4.1-2.

Χαρακτηριστικά

	Σύνολο Εκπαίδευσης	***Σύνολο Επικύρωσης***	***Σύνολο Εξέτασης***	***Συνολικά***
Μέσος αριθμός λέξεων προϋπόθεσης	13,02	14,18	14,15	13,06
Μέσος αριθμός λέξεων υπόθεσης	7,25	7,35	7,31	7,25
Αριθμός λεκτικών μονάδων (tokens)	11.141.420	211.979	210.938	11.564.337
Μέγεθος λεξιλογίου	40.512	6.588	6.742	41.243

Πίνακας 4.1-2 Χαρακτηριστικά του σώματος κειμένων.

4.2 Λεπτομέρειες υλοποίησης

Όλα τα μοντέλα υλοποιήθηκαν κατά τη διάρκεια της εργασίας στη γλώσσα προγραμματισμού Python έκδοση 2.7.6, με τη βοήθεια της βιβλιοθήκης tensorflow [Aba15] για GPU, έκδοση 0.10.

Οι υλοποιήσεις εκτελέστηκαν σε υπολογιστή (server) με τα εξής χαρακτηριστικά:

- CPU: Intel Xeon E5-2630V3 2.4 GHz 8-core LGA2011-v3 Socket
- RAM: 64GB RAM (8 x 8Gb RDIMM, 1600MHz)
- HDD: Seagate Barracuda 7200.12 3TB SATA 6Gb/s 64MB Cache
- SSD: Samsung 850 Evo Series, 250GB, 2.5-inch SSD, SATA III

- GPU: NVIDIA Titan X GPU 12-GB
- Λειτουργικό Σύστημα: Ubuntu 14.04

Για την εκπαίδευση των μοντέλων χρησιμοποιήθηκε η Categorical Cross Entropy ως συνάρτηση σφάλματος (loss function):

$$L(x) = -\sum_{c \in C} p(x = c) log \hat{p}(x = c) \quad \text{(Σχέση 4.2-1)}$$

και η Adam [Kin14] ως μέθοδος βελτιστοποίησης.

Χρησιμοποιήθηκε το σώμα κειμένων (corpus) SNLI για την εκπαίδευση (train), επικύρωση (validation) και εξέταση (test) των μοντέλων, αφού πρώτα αφαιρέθηκαν τα παραδείγματα χωρίς πλειοψηφούσα ετικέτα (Ενότητα 4.1). Για το διαχωρισμό κάθε πρότασης σε λεκτικές μονάδες (tokens) χρησιμοποιήθηκε η βιβλιοθήκη nltk.[1]

Σε όλα τα μοντέλα, οι διανυσματικές παραστάσεις λέξεων αρχικοποιούνται με αυτές του Common Crawl και παραμένουν σταθερές κατά την διάρκεια της εκπαίδευσης.[2] Αυτές οι διανυσματικές παραστάσεις λέξεων παρήχθησαν με τον αλγόριθμό GloVe [Pen14] σε σώμα κειμένων (corpus) 840 δισεκατομμυρίων λεκτικών μονάδων (tokens) με λεξιλόγιο μεγέθους 2,2 εκατομμυρίων, στο οποίο γίνεται διάκριση πεζών και κεφαλαίων. Τέλος, οι λέξεις που δεν υπάρχουν στο λεξιλόγιο απεικονίζονται τυχαία σε μια από 100 διανυσματικές παραστάσεις, οι οποίες αρχικοποιούνται τυχαία.

Κάθε μοντέλο εκπαιδεύεται μέχρις ότου να μη βελτιώνεται πλέον η ακρίβεια (accuracy) στο σύνολο επικύρωσης (validation set), με περίοδο χάριτος 20 εποχών.

$$\alpha\kappa\rho\acute{\iota}\beta\varepsilon\iota\alpha = \frac{\alpha\rho\iota\theta\mu\acute{o}\varsigma\ \sigma\omega\sigma\tau\acute{\alpha}\ \kappa\alpha\tau\eta\gamma o\rho\iota o\pi o\iota\eta\mu\acute{\varepsilon}\nu\omega\nu\ \pi\alpha\rho\alpha\delta\varepsilon\iota\gamma\mu\acute{\alpha}\tau\omega\nu}{\alpha\rho\iota\theta\mu\acute{o}\varsigma\ \pi\alpha\rho\alpha\delta\varepsilon\iota\gamma\mu\acute{\alpha}\tau\omega\nu} \quad \text{(Σχέση 4.2-2)}$$

Τα αποτελέσματα και οι παράμετροι κάθε μοντέλου περιγράφονται στις υποενότητες που ακολουθούν.

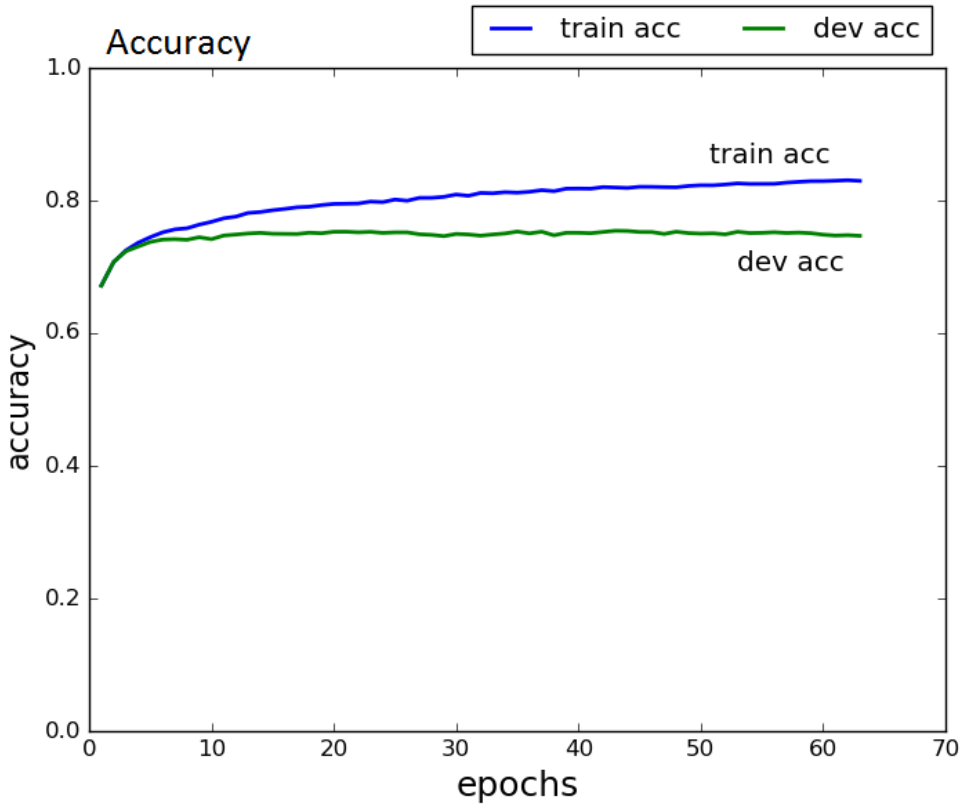
[1] http://www.nltk.org/
[2] http://nlp.stanford.edu/projects/glove/

4.3 Αποτελέσματα

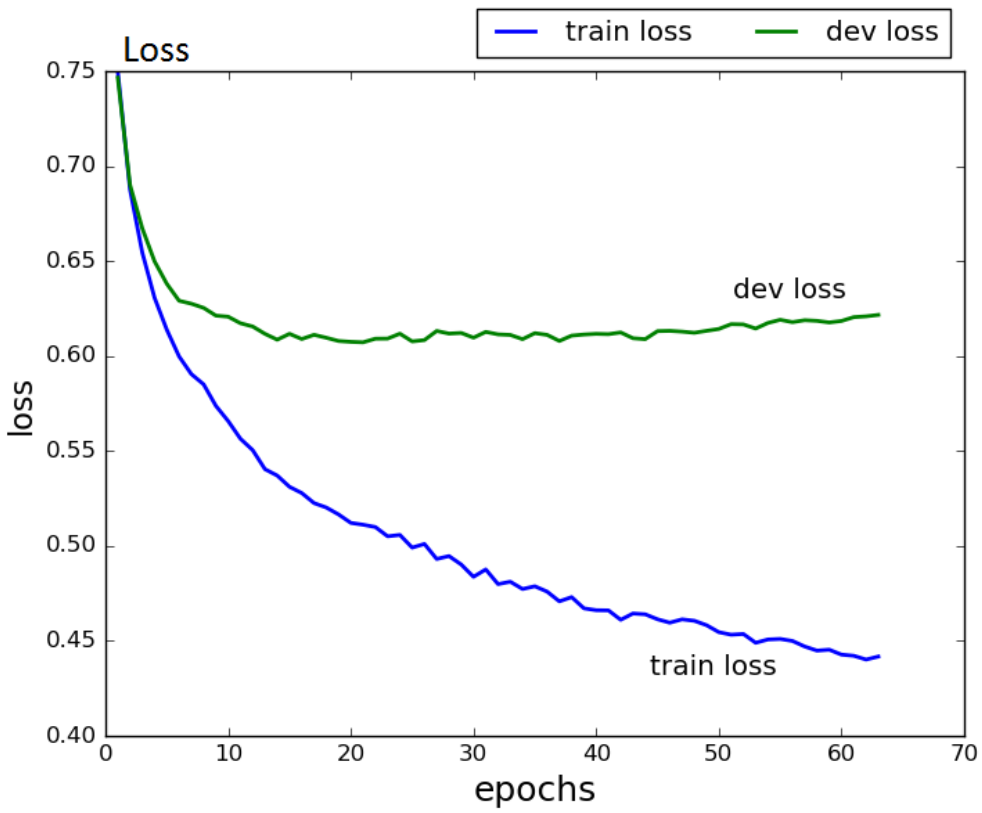
4.3.1 Μοντέλο Κωδικοποίησης Πρότασης με Κεντροειδή – CENT

Το μοντέλο αυτό πέτυχε ακρίβεια 75,16% και σφάλμα (loss) 0,6021 στο σύνολο εξέτασης (test set) αφού πρώτα εκπαιδεύτηκε στο σύνολο εκπαίδευσης (training set) για 43 εποχές με ρυθμό μάθησης (learning rate) 0,001 και μέγεθος παρτίδας (batch size) 3000. Επίσης, χρησιμοποιήθηκαν Dropout επίπεδα με πιθανότητα απόρριψης 20%.

Η ακρίβεια και το σφάλμα του μοντέλου στο σύνολο εκπαίδευσης (training set) και στο σύνολο επικύρωσης (validation set), συναρτήσει των εποχών, φαίνονται στην Εικόνα 4.3-1 και στην Εικόνα 4.3-2 αντίστοιχα.



Εικόνα 4.3-1 Διάγραμμα ακρίβειας του CENT.



Εικόνα 4.3-2 Διάγραμμα σφάλματος του CENT.

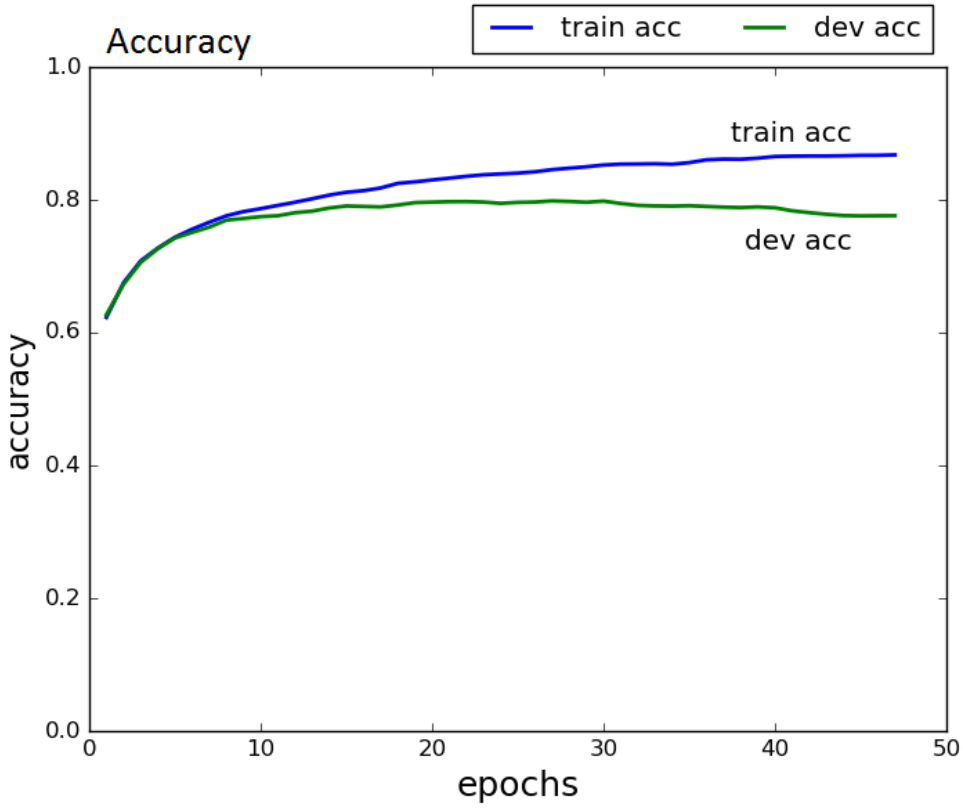
Παρόλο που το μοντέλο αυτό είναι αρκετά απλό, επιτυγχάνει αρκετά καλή ακρίβεια. Το αποτέλεσμα αυτό είναι ένδειξη της καλής ποιότητας του σώματος κειμένων και των διανυσματικών παραστάσεων λέξεων που χρησιμοποιήθηκαν.

4.3.2 Μοντέλο Κωδικοποίησης Πρότασης με Ανατροφοδοτούμενα Νευρωνικά Δίκτυα - RNNENC

Το μοντέλο αυτό πέτυχε ακρίβεια 78,91% και σφάλμα 0,5419 στο σύνολο εξέτασης (test set) αφού πρώτα εκπαιδεύτηκε στο σύνολο εκπαίδευσης (training set) για 27 εποχές με ρυθμό μάθησης (learning rate) 0,001 και μέγεθος παρτίδας (batch size) 3000. Στο μοντέλο αυτό δεν χρησιμοποιήθηκαν Dropout επίπεδα.

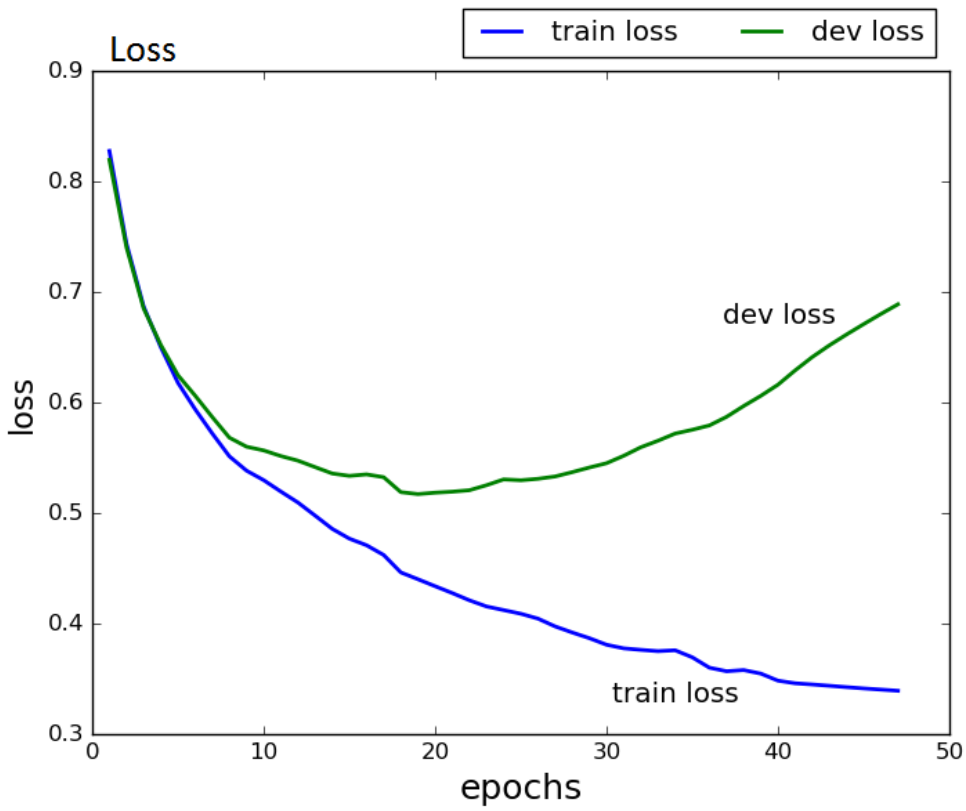
Η ακρίβεια και το σφάλμα του μοντέλου στο σύνολο εκπαίδευσης (training set) και στο σύνολο επικύρωσης (validation set), συναρτήσει

των εποχών, φαίνονται στην Εικόνα 4.3-3 και στην Εικόνα 4.3-4 αντίστοιχα.



Εικόνα 4.3-3 Διάγραμμα ακρίβειας του RNNENC.

Το μοντέλο αυτό ξεπέρασε σε ακρίβεια το Μοντέλο Κωδικοποίησης Πρότασης με Κεντροειδή κατά 3,75%. Από αυτό, μπορούμε να συμπεράνουμε ότι η χρήση Ανατροφοδοτούμενων Νευρωνικών Δικτύων οδηγεί σε καλύτερη κωδικοποίηση των προτάσεων. Ένα σημαντικό πλεονέκτημα που παρέχει η χρήση των Ανατροφοδοτούμενων Νευρωνικών Δικτύων για την κωδικοποίηση, το οποίο και εξηγεί τα καλύτερα αποτελέσματα, είναι η δυνατότητά τους να λαμβάνουν υπόψη τη σειρά των λέξεων της πρότασης.

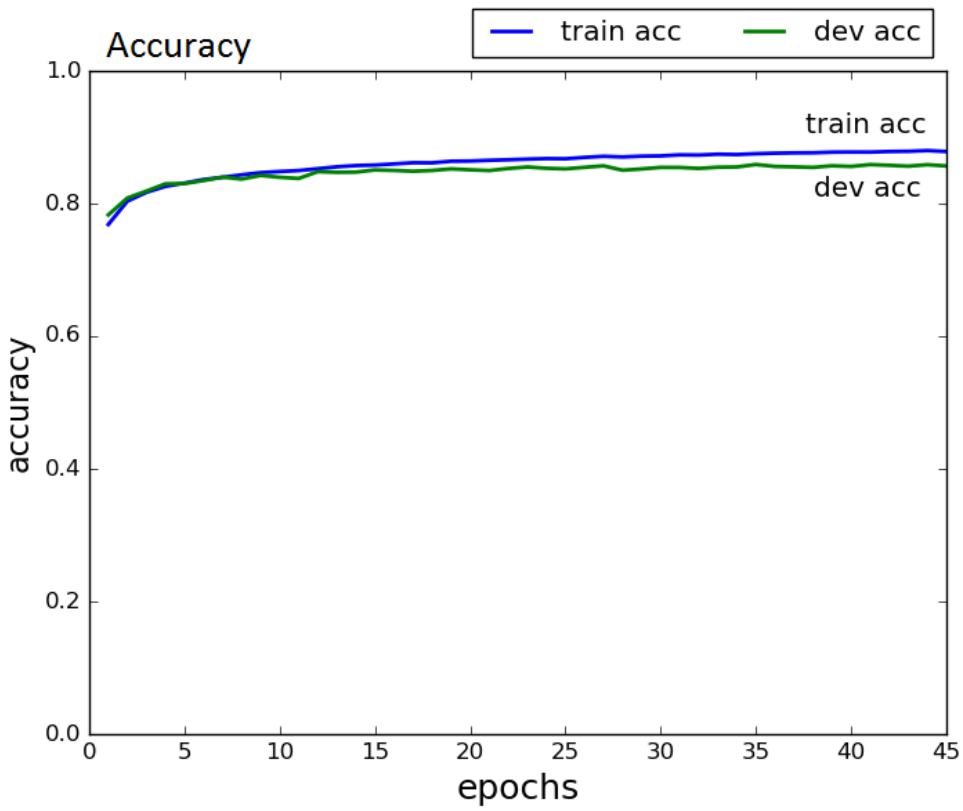


Εικόνα 4.3-4 Διάγραμμα σφάλματος του RNNENC.

4.3.3 Αποσυντιθέμενο Μοντέλο Προσοχής - DAM

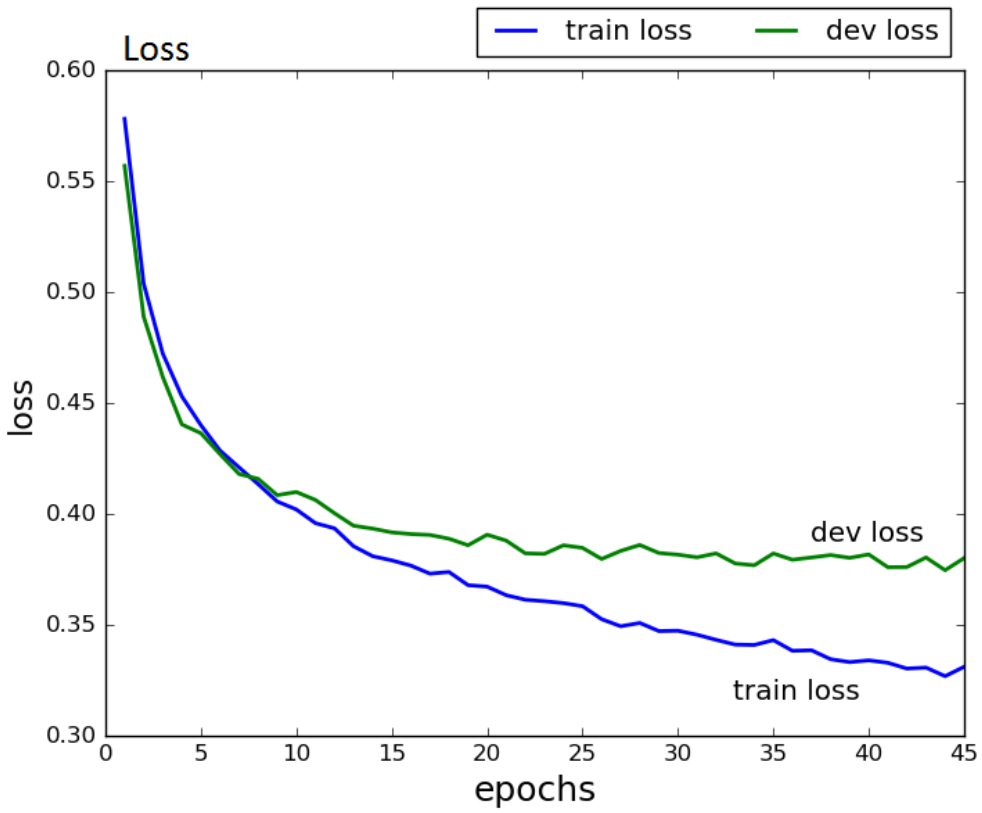
Το μοντέλο αυτό πέτυχε ακρίβεια 85,49% και σφάλμα 0,3868 στο σύνολο εξέτασης (test set) αφού πρώτα εκπαιδεύτηκε στο σύνολο εκπαίδευσης (training set) για 35 εποχές με ρυθμό μάθησης (learning rate) 0,001 και μέγεθος παρτίδας (batch size) 1000. Επίσης, χρησιμοποιήθηκαν Dropout επίπεδα με πιθανότητα απόρριψης 20%.

Η ακρίβεια και το σφάλμα του μοντέλου στο σύνολο εκπαίδευσης (training set) και στο σύνολο επικύρωσης (validation set), συναρτήσει των εποχών, φαίνονται στην Εικόνα 4.3-5 και στην Εικόνα 4.3-6 αντίστοιχα.



Εικόνα 4.3-5 Διάγραμμα ακρίβειας του DAM.

Το μοντέλο αυτό ξεπέρασε κατά πολύ σε ακρίβεια και τα δύο Μοντέλα Κωδικοποίησης Πρότασης. Όταν μια πρόταση είναι μεγάλη, είναι δύσκολο να συγκεντρωθεί ολόκληρη η σημασία της σε ένα διάνυσμα. Επίσης η χρήση Ανατροφοδοτούμενων Νευρωνικών Δικτύων, αν και βελτιώνει τα αποτελέσματα σε σχέση με το μοντέλο που χρησιμοποιεί κεντροειδή, δεν οδηγεί σε μεγάλη βελτίωση. Ο λόγος είναι ότι τα Ανατροφοδοτούμενα Νευρωνικά Δίκτυα τείνουν να χάνουν την πληροφορία που περιέχεται στις πρώτες λέξεις της εισόδου, ειδικά όταν το μήκος της πρότασης είναι μεγάλο. Σύμφωνα με τον Πίνακα 4.1-2, ο μέσος όρος αριθμού λέξεων της προϋπόθεσης (premise) είναι διπλάσιος σε σχέση με αυτόν της υπόθεσης (hypothesis), πράγμα που οδηγεί σε χειρότερη κωδικοποίηση της πρώτης σε σχέση με τη δεύτερη. Η χρήση κελιών GRU μετριάζει το φαινόμενο αυτό αλλά δεν το εξαφανίζει. Το Αποσυντιθέμενο Μοντέλο Προσοχής ξεπερνάει αυτούς τους περιορισμούς, αφού εξετάζει κάθε λέξη ξεχωριστά εστιάζοντας στην υποφράση με την οποία σχετίζεται.



Εικόνα 4.3-6 Διάγραμμα σφάλματος του DAM.

Οι παράμετροι και τα αποτελέσματα των μοντέλων συνοψίζονται στον Πίνακα 4.3-1 και στον Πίνακα 4.3-2 αντίστοιχα.

Παράμετροι

	CENT	***RNNENC***	***DAM***
Ρυθμός Μάθησης	0,001	0,001	0,001
Μέγεθος Παρτίδας	3000	3000	1000
Dropout	0,2	0	0,2
Αριθμός Εποχών	43	27	35

Πίνακας 4.3-1 Παράμετροι των μοντέλων.

Αποτελέσματα

	CENT	***RNNENC***	***DAM***
Ακρίβεια συνόλου εκπαίδευσης	0,8196	0,8454	**0,8751**
Ακρίβεια συνόλου εξέτασης	0,7516	0,7891	**0,8549**
Σφάλμα συνόλου εκπαίδευσης	0,4644	0,3974	**0,3431**
Σφάλμα συνόλου εξέτασης	0,6021	0,5419	**0,3868**

Πίνακας 4.3-2 Αποτελέσματα των μοντέλων.

5 Συμπεράσματα

Αντικείμενο της εργασίας αυτής ήταν η μελέτη, η υλοποίηση και η επαλήθευση των αποτελεσμάτων τριών μοντέλων Βαθιάς Μάθησης (Deep Learning) τα οποία έχουν προταθεί πρόσφατα για την Αναγνώριση της Κειμενικής Συνεπαγωγής (Recognizing Textual Entailment). Τα δύο από αυτά (CENT και RNNENC) ανήκουν στην κατηγορία Μοντέλων Κωδικοποίησης Πρότασης (Sentence Encoding Models), ενώ το τρίτο (DAM) κάνει χρήση ενός μηχανισμού Προσοχής (Attention). Επίσης, περιγράφηκε και χρησιμοποιήθηκε το σώμα κειμένων SNLI.

Το μοντέλο CENT αντικαθιστά την προϋπόθεση και την υπόθεση με τα κεντροειδή τους και στη συνέχεια εισάγει τη συνένωση τους σε ένα νευρωνικό δίκτυο με προς τα εμπρός τροφοδότηση. Σκοπός αυτού του μοντέλου ήταν να χρησιμοποιηθεί ως βάση σύγκρισης για τα υπόλοιπα μοντέλα που χρησιμοποιούν πιο περίπλοκες τεχνικές. Το μοντέλο αυτό, παρά την απλότητά του, πέτυχε αρκετά καλή ακρίβεια (accuracy), πράγμα που αποτελεί ένδειξη της καλής ποιότητας του σώματος κειμένων και των διανυσματικών παραστάσεων λέξεων που χρησιμοποιήθηκαν.

Το μοντέλο RNNENC χρησιμοποιεί δύο Ανατροφοδοτούμενα Νευρωνικά Δίκτυα (Recurrent Neural Networks) με κελιά GRU για να κωδικοποιήσει την προϋπόθεση και την υπόθεση ως διανύσματα. Το μοντέλο αυτό πέτυχε καλύτερα αποτελέσματα από το CENT αλλά με μικρή διαφορά. Το πλεονέκτημά του σε σχέση με το CENT είναι ότι λαμβάνει υπόψη τη σειρά των λέξεων. Το γεγονός ότι η βελτίωση είναι μικρή οφείλεται στην αδυναμία των Ανατροφοδοτούμενων Νευρωνικών Δικτύων να διατηρούν την πληροφορία που περιέχεται στις πρώτες λέξεις της εισόδου, καθώς και στην αδυναμία που έχουν όλα τα μοντέλα που κωδικοποιούν τις προτάσεις ξεχωριστά να συγκεντρώνουν όλη τη σημασία κάθε πρότασης σε ένα διάνυσμα.

Το μοντέλο DAM εξετάζει κάθε λέξη της πρότασης ξεχωριστά, εστιάζοντας κάθε φορά στην υποφράση της άλλης πρότασης με την οποία σχετίζεται, χρησιμοποιώντας ένα μηχανισμό Προσοχής (Attention). Έτσι, ξεπερνάει τον περιορισμό της ξεχωριστής

κωδικοποίησης των προτάσεων η οποία παρουσιάζεται στα προηγούμενα μοντέλα. Το μοντέλο αυτό ξεπερνάει σε ακρίβεια τα προηγούμενα μοντέλα με σημαντική διαφορά και αποτελεί ένα από τα καλύτερα μοντέλα για την Αναγνώριση Κειμενικής Συνεπαγωγής που υπάρχουν αυτή τη στιγμή.

Τέλος, μια ιδέα για μελλοντική εργασία είναι η χρήση αυτών των μοντέλων σε πιο σύνθετα προβλήματα. Για παράδειγμα, μπορεί να κατασκευαστεί ένα σύστημα Ερωταποκρίσεων (Question Answering) που να χρησιμοποιεί κάθε δυνατή απάντηση ως προϋπόθεση και να μετατρέπει κάθε ερώτηση σε δήλωση την οποία να χρησιμοποιεί ως υπόθεση. Επίσης, μια άλλη ιδέα θα ήταν η περαιτέρω μελέτη της τεχνικής της Προσοχής (Attention) και η κατασκευή μοντέλων που να την χρησιμοποιούν για να λύσουν άλλα προβλήματα. Για παράδειγμα, είναι πιθανό ένα μοντέλο σύγκρισης παραγράφων που θα χρησιμοποιεί την τεχνική αυτή να πετύχει καλά αποτελέσματα. Για το σκοπό αυτό θα μπορούσε να χρησιμοποιηθεί το μοντέλο DAM ελαφρώς παραλλαγμένο, έτσι ώστε να συγκρίνει τις προτάσεις των παραγράφων αντί τις λέξεις τους. Άρα, θα έπρεπε να προηγηθεί και κάποιου είδους κωδικοποίηση των προτάσεων, για παράδειγμα με Κεντροειδή ή με Ανατροφοδοτούμενα Νευρωνικά Δίκτυα.

6 Αναφορές

[Aba15] Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mane, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viegas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu and Xiaoqiang Zheng, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems". arXiv preprint arXiv:1603.04467, 2015.

[And10] Ion Androutsopoulos and Prodromos Malakasiotis, "A Survey of Paraphrasing and Textual Entailment Methods". *Journal of Artificial Intelligence Research*, 38:135-187, 2010.

[Bow15] Samuel R. Bowman, Gabor Angeli, Christopher Potts and Christopher D. Manning, "A large annotated corpus for learning natural language inference". Proceedings of the *Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, pp. 632-642, 2015.

[Cho14] Kyunghyun Cho, Bart van Merrienboer, Dzmitry Bahdanau and Yoshua Bengio, "On the Properties of Neural Machine Translation: Encoder--Decoder Approaches". Proceedings of the *Conference on Syntax, Semantics, Structure in Statistical Translation*, Doha, Qatar, pp. 103-111, 2014.

[Goo16] Goodfellow Ian, Bengio Yoshua and Courville Aaron, Deep Learning, The MIT Press, 2016.

[Har06] Harabagiu Sandra and Hickl Andrew, "Methods for Using Textual Entailment in Open-Domain Question Answering". Proceedings of the *21th International Conference on Computational Linguistics* and the *44th Annual Meeting of the Association for Computational Linguistics*, pp. 905-912, 2006.

[Kin14] Diederik P. Kingma and Jimmy Ba. "Adam, A Method for Stochastic Optimization". arXiv preprint arXiv:1412.6980, 2014.

[Liu16] Yang Liu, Chengjie Sun, Lei Lin and Xiaolong Wang, "Learning Natural Language Inference using Bidirectional LSTM model and Inner-Attention". arXiv preprint arXiv:1605.09090, 2016.

[Llo08] Lloret Elena, Ferrandez Oscar, Munoz Rafael and Palomar Manuel, "A Text Summarization Approach under the Influence of Textual Entailment". Proceedings of

the *5th International Workshop on Natural Language Processing and Cognitive Science*, Barcelona, Spain, pp. 22-31, 2008.

[Mou16] Lili Mou, Rui Men, Ge Li, Yan Xu, Lu Zhang, Rui Yan and Zhi Jin, "Natural Language Inference by Tree-Based Convolution and Heuristic Matching". Proceedings of the *54th Annual Meeting of the Association for Computational Linguistics*, Berlin, Germany, pp. 130-136, 2016.

[Pad09] Pado Sebastian, Galley Michel, Jurafsky Dan amd Manning Chris, "Robust Machine Translation Evaluation with Entailment Features". Proceedings of the *Joint Conference of the 47th Annual Meeting of the ACL* and the *4th International Joint Conference on Natural Language Processing of the AFNLP*, Suntec, Singapore, pp. 297-305, 2009.

[Par16] Ankur Parikh, Oscar Täckström, Dipanjan Das and Jakob Uszkoreit, "A Decomposable Attention Model for Natural Language Inference". Proceedings of the *Conference on Empirical Methods in Natural Language Processing*, Austin, Texas, pp. 2249-2255, 2016.

[Pas13] Razvan Pascanu, Tomas Mikolov and Yoshua Bengio, "On the difficulty of training recurrent neural networks". Proceedings of the *30th International Conference on Machine Learning*, Atlanta, USA, 28(3):1310-1318, 2013.

[Pen14] Jeffrey Pennington, Richard Socher and Christopher Manning, "Glove: Global Vectors for Word Representation". Proceedings of the Conference on *Empirical Methods in Natural Language Processing*, Doha, Qatar, pp. 1532-1543, 2014.

[Roc15] Tim Rocktäschel, Edward Grefenstette, Karl Moritz Hermann, Tomáš Kočiský and Phil Blunsom, "Reasoning about entailment with neural attention". arXiv preprint arXiv:1509.06664, 2015.

[Wan16] Shuohang Wang and Jing Jiang, "Learning Natural Language Inference with LSTM". Proceedings of the *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, California, pp. 1442-1451, 2016.