

Game-Theoretic Approach to WSD (M902 Team Project)

> Slide 1 - Theatina

— Καλησπέρα σας.

— Είμαστε οι Θεατίνα Κυλάφη και Τζουλιάννα Πιργιάσι και θα σας παρουσιάσουμε το άρθρο με τίτλο «A Game – Theoretic Approach to Word Sense Disambiguation», το οποίο αφορά τον τομέα της επεξεργασίας φυσικής γλώσσας και συγκεκριμένα την άρση της αμφισημίας των λέξεων (AAL), εφαρμόζοντας τεχνικές της θεωρίας παιγνίων.

> Slide 2 - Theatina

— Παρουσιάζεται λοιπόν ένα νέο μοντέλο, που αξιοποιεί την εξελικτική προσέγγιση του τομέα.

— Αρχικά θα ορίσουμε το WSD. Κάθε γλώσσα περιλαμβάνει λέξεις οι οποίες μπορεί να είναι αμφίσημες, δηλαδή να έχουν περισσότερες από μια έννοιες, ανάλογα με το περιεχόμενο στο οποίο βρίσκονται. Η αποσαφήνιση αυτών των εννοιών είναι η διαδικασία η οποία εντοπίζει τη συγκεκριμένη έννοια που λαμβάνει η λέξη μέσα σε αυτήν την πρόταση δηλαδή το περιεχόμενο.

— Για παράδειγμα, η λέξη bank είναι αμφίσημη. Οι κύριες έννοιες τις οποίες μπορεί να λάβει είναι η οικονομική και η περιβαλλοντική / νατουραλιστική .

— Η AAL είναι ακριβώς αυτό, προσπαθεί να βρει ποια από τις δύο σημασίες κυριαρχεί κάθε φορά ανάλογα με την πρόταση μέσα στην οποία βρίσκεται η λέξη «bank»

#Notes

Το άρθρο παρουσιάζει ένα νέο μοντέλο που προσεγγίζει το θέμα της αποσαφήνισης των εννοιών μιας λέξης από τη σκοπιά της θεωρίας των παιγνίων (Game theory) και συγκεκριμένα της εξελικτικής θεωρίας (evolutionary game theory).

> Slide 3 - Theatina

Το θέμα του WSD είχε προσεγγιστεί ήδη από τις αρχές του τομέα επεξεργασίας φυσικής γλώσσας (NLP), ωστόσο οι μελέτες αυτές παρέμεναν στην επιφάνεια του θέματος, παραμελώντας τη συνοχή του κειμένου το οποίο εξέταζαν.

Οι μελέτες προσεγγίζονταν από τη σκοπιά επιβλεπόμενων, μη επιβλεπόμενων ή ημι-επιβλεπόμενων μοντέλων μάθησης, από ευρετικές τεχνικές (και αλγορίθμους) καθώς και αλγορίθμους βασισμένους σε γραφήματα (graph based) ή γνώση (knowledge based).

#Notes

Supervised = μοντέλα τα οποία μαθαίνουν από ένα σύνολο δεδομένων με παραδείγματα και μετά είναι ικανά να χαρακτηρίσουν νέα παραδείγματα.

Unsupervised= μέσα από μη χαρακτηρισμένα/επισημειωμένα δεδομένα προσπαθούν να ανακαλύψουν μια πιθανή δομή.

Semi-supervised = Ένα μικρό ποσό χαρακτηρισμένων και ένα μεγάλο μη χαρακτηρισμένων παραδειγμάτων.

Heuristics = είναι οποιαδήποτε προσέγγιση στην επίλυση που χρησιμοποιεί μια πρακτική μέθοδο που δεν είναι εγγυημένη ότι είναι βέλτιστη αλλά είναι ωστόσο επαρκής για την επίτευξη ενός άμεσου, βραχυπρόθεσμου στόχου.

Knowledge based =χρησιμοποιώντας μια υπάρχουσα λεξική βάση γνώσεων, χρησιμοποιεί αλγόριθμους γραφημάτων για να αποσαφηνίσει τις λέξεις με βάση τις σχέσεις που έχουν έννοιες αυτών των λέξεων στο network.

Graph based= μοντελοποιεί τις σχέσεις μεταξύ λέξεων και αισθήσεων ενός κειμένου, σε γραφήματα, που αντιπροσωπεύουν λέξεις και αισθήσεις ως κόμβους και τις σχέσεις μεταξύ τους ως άκρα

> Slide 4 - Theatina

— Η θεωρία παιγνίων ασχολείται με την ανταγωνιστική αλληλεξάρτηση μεταξύ δύο ή περισσότερων αντιπάλων, ώστε να μελετηθεί και ιδανικά να προβλεφθεί ο τρόπος με τον οποίο γίνεται η λήψη των αποφάσεων, υπό συγκεκριμένες συνθήκες. Το παίγνιο, αποτελείται από ένα σύνολο παικτών, των αντίστοιχών τους στρατηγικών καθώς και τις αποδόσεις τους.

- Οι παίκτες, ως αυτόνομη μονάδα ο καθένας και βασιζόμενοι σε κανόνες και πληροφορίες, προσπαθούν να λάβουν μια απόφαση που θα βελτιστοποιήσει τη δική τους ευημερία έναντι των αντιπάλων τους.
- Οι κανόνες που ορίζουν τις εφικτές επιλογές των παικτών, αποτελεί το σύνολο των δυνατών στρατηγικών τους, που καθορίζονται είτε απόλυτα και ονομάζονται καθαρές στρατηγικές, είτε μέσω μιας κατανομής πιθανοτήτων επί των αντίστοιχων καθαρών και σε αυτή την περίπτωση ονομάζονται μικτές.
- Η «απόδοση» εκφράζει το αριθμητικό αποτέλεσμα του παιγνίου για έναν συνδυασμό στρατηγικών.
- Η λύση του παιγνίου ή αλλιώς ισορροπία κατά Nash, είναι ο συνδυασμός εκείνος των στρατηγικών όπου η απόδοση της αντίστοιχης στρατηγικής για κάθε παίκτη είναι η καλύτερη απάνκριση στη στρατηγική των άλλων παικτών και κανένας από αυτούς δεν θα επωφεληθεί από κάποια αλλαγή (στρατηγικής του) .
- Η εξελικτική μέθοδος της θεωρίας παιγνίων έχει δυναμικό χαρακτήρα και σε αντίθεση με την παραδοσιακή που ορίζει «λύση» σε παίγνιο που παίζεται μόνο μια φορά, εδώ οι στρατηγικές των παικτών ανταγωνίζονται μεταξύ τους καθώς ο χρόνος εξελίσσεται, προσομοιώνεται μια μέθοδος φυσικής επιλογής και τελικά επικρατούν αυτές με τα πιο σωστά «εξελικτικά» χαρακτηριστικά.

#Notes

Simplex -> πλέγμα που εκφράζει τον χώρο μεικτών στρατηγικών του παίκτη και οι γωνίες του τις καθαρές στρατηγικές του

> Slide 5 - Theatina

- Το σύστημα που προτείνει η παρούσα έρευνα λοιπόν, εφαρμόζει για πρώτη φορά τεχνικές της θεωρίας παιγνίων και συγκεκριμένα της εξελικτικής, με σκοπό την επίλυση του προβλήματος της ΑΑΛ.
- Για να μπορέσει να μοντελοποιηθεί, μετατρέπεται σε εργασία εννοιολογικής (σημασιολογικής) επισημείωσης, με σκοπό τη συμμόρφωση σε προκαθορισμένους περιορισμούς.
- Σε όρους θεωρίας παιγνίων, έχουμε τις αντιστοιχίες που φαίνονται στη διαφάνεια. Για παράδειγμα, αντί για τις γενικές έννοιες «παίκτης», «στρατηγική» και «απόδοση», έχουμε αντίστοιχα τη «λέξη», την «έννοια» και την «ομοιότητα» μεταξύ αυτών.
- Όπως προαναφέρθηκε, κατά την επαναληπτική εκτέλεση παιγνίων μεταξύ των λέξεων, λαμβάνει χώρα μια διαδικασία επιλογής των πιο κατάλληλων εννοιών, οι οποίες επικρατούν έναντι των λιγότερο κατάλληλων, οδηγώντας στην ανάθεση μίας μόνο έννοιας σε κάθε λέξη

> Slide 6 - Julie

Η συγκεκριμένη προσέγγιση βασίζεται στην σταθερή επισημείωση των δεδομένων αλλά και στο Θεώρημα του Nash κατά το οποίο για κάθε πεπερασμένο παίγνιο υπάρχει τουλάχιστον ένα σημείο ισορροπίας (αν όχι με καθαρές, με μεικτές στρατηγικές), οπότε και η λύση του προβλήματος επιτυγχάνεται σε κάθε περίπτωση (Θεώρημα Nash, 1951). Επίσης, κάθε λέξη σχετίζεται με το περιεχόμενό της. Λόγω της θεωρίας παιγνίων, το task της ΑΑΛ, μετατρέπεται σε πρόβλημα (-κατάσταση) δυναμικής βελτιστοποίησης, εκμεταλλευόμενη τις πληροφορίες του περιεχόμενου. Το προτεινόμενο σύστημα είναι πολύπλευρο, καθώς μπορεί να χρησιμοποιηθεί και από μη επιβλεπόμενα μοντέλα μάθησης αλλά και από ημι-επιβλεπόμενα.

> Slide 7 - Julie

Η διαδικασία μοντελοποίησης των δεδομένων, έγινε βηματικά και περιγράφεται στις επόμενες διαφάνειες.

> Slide 8 - Julie

Αρχικά, συλλέγονται N λέξεις, των οποίων υπολογίζεται η συσχέτιση ανά δύο, δημιουργώντας έτσι έναν πίνακα γειννίας ενός γράφου με βάρη, μέσω 8 διαφορετικών μεθόδων, η απόδοση των οποίων και αξιολογείται στη συνέχεια. Έπειτα, η ομοιότητα μεταξύ ζευγών λέξεων επαυξάνεται με βάση τη συντακτική δομή της πρότασης και την απόσταση της κάθε λέξης από τους η γείτονές της (δεξιά και αριστερά). Όπως είναι φυσικό, με μικρές τιμές του η αποτυπώνονται στερεότυπες φράσεις, ενώ αντίθετα με μεγάλες τιμές του η, ευρύτερες σημασιολογικές ιδέες.

> Slide 9 - Julie

Οι διαδικασίες που περιγράφηκαν προηγουμένως, παράγουν τα 3 διαγράμματα που βλέπουμε.

Το σχήμα 1 αποτελεί γραφική αναπαράσταση του πίνακα συσχέτισης μεταξύ των λέξεων, το σχήμα 2 αναπαριστά τις σχέσεις τους με τις n γειτονικές τους λέξεις (εδώ $n=1$) και ο γράφος της ενισχυμένης ομοιότητας μεταξύ των γειτόνων τους που φαίνονται στο n -gram, όταν αυτές δεν περιλαμβάνουν stop-words (πχ. το "be"), παρουσιάζεται στο σχήμα 3.

> Slide 10 - Theatina

— Στο επόμενο βήμα συλλέγονται οι k έννοιες που συνδέονται με την εκάστοτε λέξη i , ώστε να δημιουργηθεί ο χώρος στρατηγικής του παιχνιδιού, αποτελούμενος από την ένωση των προαναφερθέντων συνόλων και να υπολογιστεί η πιθανότητα επιλογής της εκάστοτε έννοιας από την i -οστή λέξη.

— Ο κάθε παίκτης διαθέτει ένα σύνολο μεικτών στρατηγικών, το οποίο εκφράζεται ως ένα διάνυσμα s_r διάστασης n , δηλαδή με n μη αρνητικές συντεταγμένες, οι οποίες αθροίζουν στη μονάδα, όπως θα δούμε αργότερα και εκφράζουν την πιθανότητα επιλογής της αντίστοιχης έννοιας με την οποία αυτή η λέξη σχετίζεται.

— Για την καλύτερη κατανόηση του χώρου των στρατηγικών, θα δούμε τη γραφική αναπαράστασή του, η οποία αποδίδεται ως ένα κανονικό πολύγωνο ακτίνας 1, με τις ακμές που ενώνουν το κέντρο του με κάθε κορυφή να δηλώνουν την πιθανότητα επιλογής της έννοιας που αντιπροσωπεύει η αντίστοιχη κορυφή.

> Slide 11 - Theatina

— Για την αρχικοποίηση λοιπόν του χώρου στρατηγικών του παιγνίου, αποδίδεται σε κάθε λέξη μια τιμή πιθανότητας επιλογής των εννοιών της.

— Στην περίπτωση της μη επιβλεπόμενης μάθησης, ακολουθείται η ομοιόμορφη κατανομή, καθώς δεν υπάρχει πρότερη γνώση για τις προτιμήσεις των λέξεων σχετικά με την επιλογή εννοιών, ενώ στην ημι-επιβλεπόμενη μάθηση, αξιοποιούνται πληροφορίες από εννοιολογικά επισημειωμένα δεδομένα και ακολουθείται η γεωμετρική κατανομή. Με αυτό τον τρόπο, είτε λαμβάνουν μεγαλύτερη πιθανότητα οι συχνότερες έννοιες που αφορούν την i -οστή λέξη, είτε μετά από τεχνικές συσταδοποίησης, αποδίδονται ίσες τιμές πιθανότητας στις έννοιες εντός μιας συστάδας.

#Notes

Ομοιόμορφη κατανομή (μέση τιμή: $(a+b)/2$, διασπορά: $(n^2 - 1)/12$), n =οι έννοιες

Μικρό rank r_{ih} μεγάλη συχνότητα, και όσο μικραίνει το rank, τόσο μεγαλώνει και η πιθανότητα γιατί είναι φθίνουσα κατανομή (μέση τιμή: $(1-p)/p$, διασπορά: $(1-p)/(p^2)$)

Όσο μικραίνει το p , τόσο μεγαλύτερη διασπορά τιμών έχουμε

Διαιρούμε με το άθροισμα του συνόλου των τιμών, ώστε να αθροίζει στη μονάδα

> Slide 12 - Julie

Οι ομοιότητες ανά ζεύγη εννοιών χρησιμοποιούνται για να φτιαχτεί ο πίνακας Z . Οι αποδόσεις (των στρατηγικών) για κάθε παιχνίδι ανάμεσα σε μια λέξη i και j , λαμβάνονται από τον αντίστοιχο υποπίνακα $Z_{(|M_i| \times |M_j|)}$.

(διάστασης $m_i \times m_j$ όπου τα m_i και m_j είναι ο πληθικός αριθμός των συνόλων εννοιών M_i και M_j , αντίστοιχα.)

Για τον υπολογισμό της σημασιολογικής ομοιότητας των εννοιών, χρησιμοποιήθηκαν τα μέτρα wup και jcn .

Αυτά, υπολογίζουν την ομοιότητα μεταξύ των δύο εννοιών c_i, c_j σύμφωνα με το βάθος των δύο εννοιών στη λεξική βάση δεδομένων και τον πιο συγκεκριμένο πρόγονο κόμβο (msa) των δύο εννοιών.

Το Information Content είναι η πληροφορία ενός περιεχομένου μιας ευρύτερης έννοιας c που προέρχεται από ένα corpus και πρόκειται για εναλλακτική έκφραση πιθανότητας που δηλώνει το βαθμό έκπληξης σχετικά με την εμφάνιση μιας λέξης.

#Notes

Οι γραμμές επιλέγονται με βάση τις έννοιες που σχετίζονται με τη λέξη i και οι στήλες με τις έννοιες που σχετίζονται με τη λέξη j .

Ο πίνακας μερικής απόδοσης περιέχει παιχνίδια ανάμεσα σε 2 λέξεις i, j .

> Slide 13 - Julie

Η συσχέτιση, είναι η ομοιότητα ανάμεσα στους ορισμούς δύο ευρύτερων εννοιών, οπότε και τα μέτρα συσχέτισης είναι πιο γενικά.

Οι ορισμοί των εννοιών χρησιμοποιούνται για να φτιάξουν ένα διάνυσμα συνεμφάνισης $v_i = (w_{i1}, w_{i2}, \dots, w_{in})$ όπου για κάθε έννοια i , το w αντιπροσωπεύει τον αριθμό των φορών που εμφανίζεται η λέξη και n είναι ο συνολικός αριθμός

διαφορετικών λέξεων (τύπων) στο corpus. Αυτή η αναπαράσταση επιτρέπει την προβολή τους στο διανυσματικό χώρο για τη διεξαγωγή πειραμάτων και μελετών.

Στα πειράματα της παρούσας έρευνας, χρησιμοποιήθηκε το μέτρο ομοιότητας συνημιτόνου της γωνίας θ ανάμεσα στα διανύσματα συνεμφάνισης δύο διαφορετικών ευρύτερων εννοιών i, j , η οποία εκφράζει τις κοινώς συνεμφανιζόμενες λέξεις των 2 concepts.

#Notes

Και $\|v\|$ είναι η νόρμα των διανυσμάτων.

Όσο περισσότερο μεγαλώνει η γωνία τόσο μεγαλύτερη γίνεται κάθε φορά η ομοιότητα.

Οι ορισμοί των οποίων προέρχονται από τα glosses των synsets [όπου gloss είναι μια σύντομη σημειογραφία, της έννοιας μιας λέξης και synset είναι ένα σύνολο λέξεων που μοιράζονται παρόμοιες έννοιες, δηλαδή ένα σύνολο συνωνύμων].

Συνημίτονο της γωνίας θ ανάμεσα σε δύο διανύσματα συνεμφάνισης δύο διαφορετικών λέξεων v_i, v_j .

Στην περίπτωση μας, επιστρέφει τιμές που κυμαίνονται από 0 έως 1 επειδή οι τιμές στα διανύσματα συνεμφάνισης είναι όλες θετικές

Όσο πιο κοντά στο 1 τόσο πιο όμοια είναι τα διανύσματα.

> Slide 14 - Theatina

— Όπως αναφέρθηκε και στις εισαγωγικές διαφάνειες, για την εύρεση του σημείου ισορροπίας του παιγνίου, πραγματοποιείται μια διαδικασία επιλογής των επικρατέστερων εννοιών.

— Αναλυτικότερα, σε κάθε επανάληψη, κάθε λέξη «παίζει» ένα παιχνίδι με το σύνολο των γειτονικών της.

— Με αυτό τον τρόπο ενημερώνει την πιθανότητα επιλογής της κάθε συσχετιζόμενης με αυτήν έννοιας, λαμβάνοντας υπόψιν την ομοιότητα με τις γειτονικές της, την ομοιότητα μεταξύ των εννοιών τους καθώς και τις προτιμήσεις του συμπαίκτη της για τη συγκεκριμένη έννοια.

— Καθώς λοιπόν εξελίσσεται ο χρόνος, σε κάθε επόμενη χρονική στιγμή οι λέξεις «μιμούνται» επιτυχημένες στρατηγικές συμπαίκτων τους, δηλαδή αυξάνουν την πιθανότητα επιλογής μιας έννοιας μέσω της εξίσωσης αντιγραφής, δίνοντας μεγαλύτερο βάρος σε εκείνη που επιφέρει διαφορά απόδοσής της από τη μέση με θετικό πρόσημο, ενώ παράλληλα μειώνει την πιθανότητα στην αντίθετη περίπτωση (όταν δηλαδή η διαφορά από τη μέση απόδοση έχει αρνητικό πρόσημο). Έτσι, ενισχύονται και επικρατούν συνολικά οι έννοιες εκείνες που προσφέρουν μεγαλύτερο κέρδος στις αντίστοιχες λέξεις.

#Notes

$\llbracket e \rrbracket \wedge h \rightarrow$ μοναδιαίο διάνυσμα

$u_i(x) = \sum_{(j \in N_i)} \llbracket x_i \wedge T(w_{ij} Z_{ij} x_j) \rrbracket \rightarrow$ average fitness of population i

> Slide 15 - Julie

Τελικά, μετά το πέρας της προηγούμενης επαναληπτικής διαδικασίας, προκύπτει μία και μόνο επικρατούσα έννοια που αποδίδεται σε κάθε λέξη-παίκτη.

Αυτό επιτυγχάνεται μέσω της συνάρτησης \argmax , η οποία επιστρέφει την έννοια με τη μεγαλύτερη πιθανότητα από το διάνυσμα των μεικτών στρατηγικών της κάθε λέξης.

Επίσης, όπως πειραματικά αποδείχθηκε, σε ελάχιστες περιπτώσεις δεν επιτυγχάνεται η επίλυση της αμφισημίας μιας λέξης, λόγω αδυναμίας ενημέρωσης του χώρου των μεικτών στρατηγικών της, γεγονός που μπορεί να οφείλεται για παράδειγμα στην ανυπαρξία κοινών εννοιών με άλλες λέξεις.

> Slide 16 - Theatina

— Όπως είδαμε και πριν, στο σχήμα αυτό παρουσιάζεται γραφικά ο στρατηγικός χώρος των λέξεων, δηλαδή τα κανονικά πολύγωνα με τις αντίστοιχες πιθανότητες επί των συσχετιζόμενων εννοιών τους.

— Εδώ λοιπόν γίνεται πιο ξεκάθαρα αντιληπτή η εξέλιξη των στρατηγικών στο χρόνο που αναφέρθηκε στην προηγούμενη διαφάνεια, για τις λέξεις “be”, “institution” και “bank”.

— Συγκεκριμένα, στη χρονική στιγμή t_1 , δηλαδή την 1η γραμμή, όλες οι έννοιες που σχετίζονται με τις προηγούμενες λέξεις είναι ισοπίθανες.

Στο επόμενο χρονικό βήμα t_2 απαλοίζονται αρκετές έννοιες, είτε λόγω μικρής συσχέτισης, είτε λόγω κυρίαρχης επίδρασης κάποιων άλλων, όπως ισχύει για τη λέξη «institution», η οποία λόγω των λέξεων «financial» & «bank», κατευθύνεται αμέσως προς μια συγκεκριμένη έννοια (οικονομική).

— Ένα ενδιαφέρον παράδειγμα από την άλλη, είναι εκείνο της αμφίσημης λέξης «bank», η οποία "παίζοντας παιχνίδια" με τις λέξεις "financial" & "institution" κατευθύνεται προς την οικονομική έννοια, ενώ όταν παίζει με τη λέξη «river», ενισχύεται η περιβαλλοντική / νατουραλιστική της έννοια.

— Στο χρονικό βήμα t12 το σύστημα τελικά συγκλίνει (δηλαδή οι αλλαγές στις καταστάσεις του παιχνιδιού είναι πλέον αμελητέες), και η έννοια της αμφίσημης λέξης «bank» με τη μεγαλύτερη πιθανότητα καταλήγει να είναι η περιβαλλοντική (s20) έναντι της οικονομικής (s22), επιλογή που όπως περιγράφηκε σε προηγούμενες διαφάνειες, οφείλεται σε ποικίλους παράγοντες και μέτρα, ισχυροποιώντας τη σταθερότητα του τελικού αποτελέσματος.

> Slide 17 - Theatina

— Οι μεταβλητές που πρόκειται να παραμετροποιηθούν είναι:

1. Τα μέτρα που αφορούν τη συσχέτιση και τη σημασιολογία, ώστε να σταθμιστεί η ομοιότητα ανάμεσα σε λέξεις και έννοιες.
2. Το n-gram γράφημα που αξιοποιείται για την αύξηση των βαρών των κοντινών λέξεων (proximity-εγγύτητα) και
3. Η παράμετρος p της γεωμετρικής κατανομής, από την οποία εξαρτάται η διασπορά των τιμών.

> Slide 18 - Theatina

— Τα αποτελέσματα, δείχνουν ότι όσον αφορά την σχετικότητα και την ομοιότητα, τα μέτρα της συσχέτισης (relatedness) λειτουργούν καλύτερα απ' ό,τι αυτά της σημασιολογικής ομοιότητας (semantic). Συγκεκριμένα, τα mdice & tf-idf αποδείχθηκαν ως τα πιο κατάλληλα.

— Όσον αφορά το n-gram, όταν το $n=5$ (5 κοντινότερες λέξεις - neighbours), τότε τα αποτελέσματα είναι τα βέλτιστα.

— Τέλος, σχετικά με τη γεωμετρική κατανομή, τα καλύτερα αποτελέσματα παρουσιάζονται με $p=0.4$

> Slide 19 - Julie

Στη συνέχεια για να λάβουμε τα ποσοστά της αξιολόγησης χρησιμοποιείται το μέτρο F-score και συγκεκριμένα το balanced, δηλαδή το F1, το οποίο καθορίζει τον σταθμισμένο αρμονικό μέσο της ακρίβειας και της ανάκλησης.

Αρχικά έγιναν πειράματα που είχαν ως knowledge base είτε το wordnet είτε το babelnet.

Τα κύρια αποτελέσματα είναι αυτά που φαίνονται στη διαφάνεια.

> Slide 20 - Julie

Έπειτα έγινε σύγκριση του κάθε αλγορίθμου με τα state of the art συστήματα, όπου όντως φάνηκε στα αποτελέσματα να λειτουργούν καλύτερα αλλά και να παρουσιάζουν μια σταθερότητα στα ποσοστά τους σε κάθε dataset.

Κάτι άλλο που μπορεί να παρατηρηθεί είναι ότι το semi-supervised μοντέλο, έχει κοντινές τιμές με τα άλλα supervised συστήματα.

-> τα δείχνουμε παράλληλα με την εικόνα στην Διαφάνεια 21

> Slide 21 - Julie

> Slide 22 - Julie

Κλείνοντας, θα συνοψίσουμε τις μέχρι τώρα πληροφορίες.

Έγινε λόγος για μια νέα μέθοδο που άρει την αμφισημία των λέξεων, μέσω της εξελικτικής θεωρίας παιχνιδιών η οποία θέτει κάποια μέτρα ομοιότητας που αποδίδουν καλύτερα και μπορεί να θεωρηθεί η συνέχιση των knowledge-based & graph-based αλγορίθμων.

Η αποσαφήνιση των εννοιών των λέξεων αντιμετωπίστηκε ως ένα διαρκές παιχνίδι [έλεγε πρόβλημα] περιορισμών και ικανοποίησης αυτών, οδηγώντας σε σταθερή συσχέτιση εννοιών και λέξεων

Η ανάθεση της κάθε έννοιας στην αντίστοιχη λέξη βασίστηκε στο περιεχόμενο (στη κειμενική συνοχή), χαρακτηριστικό που λείπει από τα state of the art συστήματα.

Η χρήση του replicator dynamics equation είναι υπεύθυνη για την καλύτερη ανάθεση έννοιας σε κάθε λέξη

Το σύστημα αυτό όπως ειπώθηκε είναι πολύπλευρο και ευπροσάρμοστο

Και τέλος θεωρείται πλήρως ανταγωνιστικό σε σχέση με τα προϋπάρχοντα συστήματα. Η σημασία μιας λέξης εξαρτάται μόνο από λέξεις που μοιράζονται μια σχέση εγγύτητας (βρίσκονται πιο κοντά) και από εκείνες που απολαμβάνουν υψηλή ομοιότητα.

> Slide 23 - Theatina

- After so much talking we got hungry, therefore a toast is what we need the most !
- Context: talking, got, hungry, toast, need
- Stop-words: got need
- Ambiguous: toast

> Slide 24 - Theatina

> Slide 25 - Theatina

! DISCLAIMER !

- Αν και λατρέψαμε το θέμα και τον τρόπο προσέγγισής τους, αρκετές ήταν οι φορές που βρεθήκαμε σε αδιέξοδο λόγω εκφραστικών ή και μαθηματικών λαθών (τύποι, διαγράμματα, κλπ), όπως και ασάφεια ή έλλειψη συνοχής κατά την περιγραφή διαδικασιών και εννοιών, με ένα από τα κυριότερα προβλήματα να αποτελεί η έλλειψη συνέπειας στη σημειογραφία.
- Επιπλέον, επειδή η θεωρία παιγνίων αποτελεί καινούριο αντικείμενο με το οποίο ασχολούμαστε πρώτη φορά, τα προαναφερθέντα λάθη είχαν μεγαλύτερη βαρύτητα.