

Θέματα Διπλωματικών Εργασιών Εαρινό Εξάμηνο Ακαδημαϊκού Έτους 2021-2022

- Αν σας ενδιαφέρει κάποιο από τα θέματα στείλτε μου email μαζί με ένα σύντομο βιογραφικό ή/και την αναλυτική σας βαθμολογία **μέχρι και τις 20.1.2022** για εκπόνηση διπλωματικής το **εαρινό εξάμηνο**. Όσοι ενδιαφέρονται για την επόμενη ακαδημαϊκή χρονιά μπορεί να επικοινωνήσουν μαζί μου και αργότερα.

Για όλες τις εργασίες χρειάζεται πολύ καλή γνώση Αγγλικών και προγραμματισμού καθώς και κατανόησης θεωρητικών εννοιών.

Διπλωματική 1: Μελέτη διάφορων μορφών πόλωσης στο δίκτυο Reddit

Σε πολλά κοινωνικά δίκτυα δημιουργείται πόλωση όπου κυριαρχούν δυο ακραίες απόψεις για κάποιο θέμα (π.χ., υπέρ και κατά του εμβολιασμού). Σε αυτήν την εργασία θα μελετήσουμε διαφορετικούς τύπους δομικής πόλωσης στο δίκτυο Reddit, συγκεκριμένα (α) με πρόσημο (θετικό, αρνητικό) και μη, και (β) στην ίδια ή σε διαφορετικές κοινότητες (subreddits).

Έχουμε δομική πόλωση χωρίς πρόσημο όταν δημιουργούνται δύο ομάδες, η επικοινωνία ανάμεσα στα μέλη της ίδιας ομάδας είναι συχνή, αλλά τα μέλη διαφορετικών ομάδων επικοινωνούν σπάνια. Δομική πόλωση με πρόσημο έχουμε όταν πάλι δημιουργούνται δυο ομάδες, τα μέλη μιας ομάδας συμφωνούν με τα μέλη της ίδιας ομάδας (θετικό πρόσημο) και διαφωνούν με τα μέλη της άλλης ομάδας (αρνητικό πρόσημο).

Το Reddit είναι χωρισμένο σε κοινότητες (subreddits) ανάλογα με τη θεματολογία των συζητήσεων (π.χ., Turkey, Greece). Θα εξετάσουμε τη δημιουργία πόλωσης για κάποιο θέμα μέσα σε μία κοινότητα και ανάμεσα σε δύο ή περισσότερες κοινότητες.

Η εργασία θα βασιστεί (επαναλάβει) προηγούμενη μελέτη μας στο θέμα (εργασία [1]). Για τη μέτρηση της πόλωσης, θα χρησιμοποιηθούν βιβλιοθήκες που υλοποιούν σχετικές μετρικές.

Αναφορές

[1 Chrisoula Terizi, Evaggelia Pitoura, Polarized Groups in Discussion Forums: The Case of Reddit, Unpublished, May 2021

Διπλωματική 2: Χρήση τεχνικών μηχανικής μάθησης σε σχεσιακά δεδομένα

Η χρήση ML (machine learning – μηχανικής μάθησης) τεχνικών έχει δώσει εξαιρετικά αποτελέσματα σε δύσκολα προβλήματα σε πολλές ερευνητικές περιοχές, όπως η επεξεργασία φυσικής γλώσσας και η επεξεργασία εικόνας.

Πρόσφατα ML τεχνικές χρησιμοποιούνται και σε προβλήματα σχεσιακών βάσεων δεδομένων.

Σε αυτή τη διπλωματική θα εξεταστεί μια τεχνική βασισμένη σε graph embeddings που έχει προταθεί στην εργασία [2].

Θα μελετηθεί η εργασία και θα επαναληφθούν τα πειράματα που παρουσιάζονται σε αυτές. Επίσης, θα εξεταστούν πιθανές παραλλαγές. Αν υπάρξει χρόνος, θα εξεταστούν και πρόσφατες επεκτάσεις με χρήση πιο προχωρημένων ML τεχνικών.

Αναφορές

[2] Riccardo Carruzzo, Paolo Papotti, Saravanan Thirumuruganathan: Creating Embeddings of Heterogeneous Relational Datasets for Data Integration Tasks. SIGMOD Conference 2020: 1335-1349

Διπλωματική 3: Μελέτη του επίδρασης των ελλειπουσών τιμών (missing values) στη δικαιοσύνη

Ένα από τα συχνά προβλήματα με τα δεδομένα, είναι ότι δεν είναι πλήρη, αλλά λείπουν οι τιμές από κάποια πεδία τους. Οι αλγόριθμοι που τα χρησιμοποιούν είτε αγνοούν αυτά τα δεδομένα, είτε συμπληρώνουν τις τιμές που λείπουν χρησιμοποιώντας διάφορες τεχνικές.

Σε αυτή την διπλωματική εργασία θα εξετάσουμε κατά πόσο ο χειρισμός των τιμών που λείπουν επηρεάζει την δικαιοσύνη των αποτελεσμάτων των αλγορίθμων. Για την μέτρηση της δικαιοσύνης, θα χρησιμοποιήσουμε έναν ή περισσότερους από τους πολλούς τρόπους που έχουν προταθεί στη βιβλιογραφία.

Θα αρχίσουμε με απλά δεδομένα και αλγορίθμους ταξινόμησης, όπως στην αναφορά [3]. Στη συνέχεια, θα επικεντρωθούμε και θα μελετήσουμε βάσεις γνώσεων και αλγορίθμους embedding για αυτές, όπως στην αναφορά [4].

Αναφορές

[3] Fernando Martínez-Plumed, Cèsar Ferri, David Nieves, José Hernández-Orallo: Missing the missing values: The ugly duckling of fairness in machine learning. Int. J. Intell. Syst. 36(7): 3217-3258 (2021)

[4] Styliani Bourli, Evaggelia Pitoura: Bias in Knowledge Graph Embeddings. ASONAM 2020: 6-10