Προτεινόμενα Θέματα Διπλωματικής Εργασίας

Από

Γεώργιο Γιαννακόπουλο

Α. Διασταύρωση ισχυρισμών με χρήση μηχανικής μάθησης

Συνοπτική περιγραφή:

Στη δημόσια σφαίρα (δημόσιος λόγος και επικοινωνίες) είναι πολύ συχνό να εμφανίζονται ειδήσεις και ισχυρισμοί με αδυναμία επαλήθευσης, είτε σκόπιμα είτε από παραδρομή. Αυτή η εργασία θα μελετήσει και θα προτείνει μεθόδους υπολογιστικής επαλήθευσης[1] ειδήσεων και ισχυρισμών, συνδυάζοντας πρότερη γνώση με πολλαπλές επιγραμμικές (online) πηγές περιεχομένου. Για το σκοπό αυτό θα αξιοποιηθούν μέθοδοι ανάλυσης φυσικής γλώσσας και μηχανικής μάθησης και υπάρχοντα δεδομένα από σχετικές ερευνητικές εργασίες[2].

Β. Μελέτη της επίδρασης κωδικοποίησης εισόδου σε μεθόδους μηχανικής μάθησης

Συνοπτική περιγραφή:

Σε σύγχρονα συστήματα μηχανικής μάθησης αξιοποιούνται διάφοροι τρόποι για κωδικοποίηση των δειγμάτων εισόδου[3]. Η παρούσα εργασία καλείται να μελετήσει και να αναδείξει πώς η κωδικοποίηση επηρεάζει το αποτέλεσμα εργασιών μηχανικής

---

[1] Ciampaglia, G. L., Shiralkar, P., Rocha, L. M., Bollen, J., Menczer, F., & Flammini, A. (2015). Computational fact checking from knowledge networks. PloS one, 10(6).

Επίσης: Rashkin, Hannah, et al. "Truth of varying shades: Analyzing language in fake news and political fact-checking." Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017.

[2] Vlachos, A., & Riedel, S. (2014, June). Fact checking: Task definition and dataset construction. In Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science (pp. 18-22).

[3] Potdar, K., Pardawala, T. S., & Pai, C. D. (2017). A comparative study of categorical variable encoding techniques for neural network classifiers. International journal of computer applications, 175(4), 7-9.

μάθησης (π.χ. ταξινόμηση ή συσταδοποίηση) σε διάφορες κατηγορίες αλγορίθμων. Στο πλαίσιο της εργασίας, οι φοιτητές θα κληθούν να κατανοήσουν τις ιδιαιτερότητες διαφορετικών κωδικοποιήσεων, να περιγράψουν και να υλοποιήσουν πειράματα αξιολόγησης σε διαφορετικά σύνολα δεδομένων με διαφορετικούς αλγορίθμους μάθησης, και να στηρίξουν στατιστικά τα ευρήματά τους.