



Προτεινόμενα Θέματα Διπλωματικής Εργασίας

Από

Γεώργιο Γιαννακόπουλο

A. 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 επαλήθευσης¹ 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 εργασίες².

B. Μελέτη της επίδρασης κωδικοποίησης εισόδου σε μεθόδους μηχανικής μάθησης

Συνοπτική περιγραφή:

Σε σύγχρονα συστήματα μηχανικής μάθησης αξιοποιούνται διάφοροι τρόποι για κωδικοποίηση των δειγμάτων εισόδου³. Η παρούσα εργασία καλείται να μελετήσει και να αναδείξει πώς η κωδικοποίηση επηρεάζει το αποτέλεσμα εργασιών μηχανικής μάθησης (π.χ. ταξινόμηση ή συσταδοποίηση) σε διάφορες κατηγορίες αλγορίθμων.

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.

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.

Γ. Βελτιστοποίηση Διαδρομών Πλοίων

Qualifications required: C++, Python,

- Machine learning, Deep learning

Το πρόβλημα της βελτιστοποίησης της διαδρομής πλοίου αποτελεί πρόβλημα μείζονος ενδιαφέροντος για τις ναυτιλιακές εταιρείες και ερευνητικό πεδίο σταθερού ενδιαφέροντος για τις Επιστημες Μηχανικού, την Πληροφορική και τα Μαθηματικά. Στο παραπάνω πλαίσιο παρέχουμε την ευκαιρία σε έναν/μία φοιτητή/φοιτήτρια να ασχοληθεί αρχικά με την διαδικασία συλλογής και επεξεργασίας δεδομένων τα οποία προέρχονται από αισθητήρες (sensors) επί του πλοίου ή δορυφόρους και αφορούν είτε δεδομένα καιρού είτε στοιχεία λειτουργίας του πλοίου.

Σε δεύτερη φάση θα δοθεί η δυνατότητα στον/στην φοιτητή/φοιτήτρια να αναπτύξει μοντέλα Machine Learning (ML, Μηχανικής Μάθησης) με σκοπό την εύρεση της υποκείμενης μαθηματικής δομής (π.χ. εξίσωσης) που διέπει το πρόβλημα της βελτιστοποίησης της διαδρομής του πλοίου.

Λαμβάνοντας υπόψη πληθώρα ανεξάρτητων μεταβλητών που αφορούν τα προαναφερθέντα δεδομένα το πρόβλημα ανάγεται τελικώς στον προσδιορισμό της κατάλληλης σύνθεσης αυτών με απώτερο στόχο την υλοποίηση και αξιολόγηση εναλλακτικών μοντέλων τα οποία είναι σε θέση να δώσουν εκτιμήσεις, με ικανοποιητικό ποσοστό ακρίβειας, της κατανάλωσης καυσίμου του πλοίου, σε διαστήματα της διαδρομής του.

Ο/η φοιτητής/φοιτήτρια θα έχει επίσης την ευκαιρία να ασχοληθεί και να αξιοποιήσει:

- Μεθόδους ομαδοποίησης (clustering) δεδομένων και να διερευνήσει πώς αυτές θα μπορούσαν να συνδυαστούν κατάλληλα με μοντέλα ML έτσι ώστε να επιτευχθούν καλύτερα αποτελέσματα.
- Αναλυτικές μεθόδους τετραγωνικών εκτιμητών πάνω σε επιφάνειες που έχουν προέλθει από τριγωνοποίηση δεδομένων και πώς μπορούν αυτές να συνδυαστούν κατάλληλα με μεθόδους ML ή/και NN (Neural Nets, Νευρωνικά Δίκτυα).

Δ. Semantic Retrieval of Legal Texts

Qualifications required: Πολύ καλή γνώση προγραμματισμού, Data Mining, Machine Learning, Python/Java

Qualifications desired: Εξοικείωση με αναπαραστάσεις κειμένου και συσταδοποίηση (clustering), Εξόρυξη κειμένου, Graph theory, Natural Language Processing, Deep Learning

Συνοπτική περιγραφή:

Σε μία πληθώρα νομικών εφαρμογών, υπάρχει η ανάγκη να χρησιμοποιηθεί ένα κείμενο ως ερώτημα σε μία βάση γνώσης, ώστε να ανακληθούν τα πιο σχετικά κείμενα. Για παράδειγμα, στο εθιμικό δίκαιο, κάθε περίπτωση έχει σχετιζόμενες προϋπάρχουσες περιπτώσεις, στις οποίες ανατρέχουν δικηγόροι και δικαστικοί. Η

ανάλυση φυσικής γλώσσας μπορεί να υπολογίσει σημασιολογική ομοιότητα κειμένων, ενώ η ανάκληση πληροφορίας επιφορτίζεται με την αποδοτική ευρετηρίαση και ανάκληση σχετικών εγγράφων. Αυτή η εργασία, θα λειτουργήσει στο όριο των 2 τομέων, προσπαθώντας να κτίσει αποδοτικές αναπαραστάσεις, μεθόδους υπολογισμού ομοιότητας και ευρετήρια για να επιτρέψει ανάκληση νομικών εγγράφων, με αξιοποίηση της σημασιολογικής εγγύτητας.

Μετά το πέρας της εργασίας, οι φοιτητές θα μπορούν:

- να περιγράψουν τι είναι η σημασιολογική ομοιότητα,
- να χρησιμοποιήσουν εργαλεία ανάλυσης φυσικής γλώσσας,
- να χρησιμοποιήσουν μηχανική μάθηση / βαθιά μάθηση για αναπαράσταση κειμένων,
- να χρησιμοποιήσουν πληθώρα τεχνικών ανάκλησης για την αναζήτηση παρόμοιων κειμένων.