

Συγκριτική μελέτη μετρικών εγγύτητας χρηστών στον αλγόριθμο του συνεργατικού φιλτραρίσματος

(Comparative analysis of user similarity metrics for the collaborative filtering algorithm)

Περιγραφή

Οι αλγόριθμοι συνεργατικού φιλτραρίσματος (collaborative filtering algorithms) λαμβάνουν υπόψη τα γούστα και τα ενδιαφέροντα των χρηστών, εκφραζόμενα ως αξιολογήσεις (ratings), προκειμένου να δημιουργήσουν εξατομικευμένες συστάσεις (personalized recommendations). Αυτοί οι αλγόριθμοι, αρχικά, βρίσκουν τους “κοντινούς γείτονες” (near neighbors) κάθε χρήστη, δηλαδή χρήστες που έχουν πολύ παρόμοια γούστα (βάσει των αξιολογήσεων τους στα κοινά αξιολογούμενα αντικείμενα). Στη συνέχεια, χρησιμοποιούνται οι ήδη καταχωρημένες αξιολογήσεις τους, προκειμένου να διατυπωθούν προβλέψεις αξιολόγησης (rating predictions), οι οποίες προβλέψεις οδηγούν στη δημιουργία συστάσεων (recommendations), προτείνοντας τα αντικείμενα με τις υψηλότερες τιμές προβλέψεων (rating prediction values). Όπως είναι αναμενόμενο, η ποιότητα των προβλέψεων αξιολογήσεων επηρεάζει σημαντικά την ποιότητα των συστάσεων που δημιουργούνται.

Στα πλαίσια αυτής της πτυχιακής εργασίας, ο/οι φοιτητής/ές θα υλοποιήσουν τον βασικό αλγόριθμο συνεργατικού φιλτραρίσματος (χωρίς χρήση επιπρόσθετης πληροφορίας, όπως χαρακτηριστικά χρηστών και αντικειμένων, σχέσεις μεταξύ των, κλπ), καθώς και τις γνωστότερες μετρικές εγγύτητας χρηστών (π.χ. Pearson Correlation Coefficient, Cosine Similarity, κλπ), τα οποία θα εφαρμόσουν πάνω σε γνωστά και παγκοσμίως αποδεκτά, στην κοινότητα των συστημάτων συστάσεων, σύνολα δεδομένων (datasets), όπως αυτά της Amazon, της MovieLens, της Netflix, κ.λπ., προκειμένου να υπολογίσουν την απόκλιση των προβλέψεων αξιολογήσεων (rating prediction error) από τις πραγματικές τιμές.

Στα ανωτέρω σύνολα δεδομένων περιλαμβάνονται και μικρά και μεγάλα (μεγέθους αρκετών GB), καθώς και αραιά και πυκνά, οπότε η υλοποίηση που θα γίνει, θα πρέπει να είναι πλήρως βελτιστοποιημένη/αποδοτική, και χρονικά (χρόνος εκτέλεσης), αλλά και χωρικά (χώρος στην κύρια μνήμη).

Η μέγιστη διάρκεια ολοκλήρωσης της πτυχιακής είναι ένα ημερολογιακό έτος.

Παραδοτέα

- Ο Κώδικας σε ενότητες (modules)
- Αναλυτικές τιμές μετρήσεων
- Αναφορά πτυχιακής εργασίας (σε αρχείο word), που θα περιλαμβάνει τις τιμές των μετρήσεων (σε γραφήματα excel), με σχολιασμό τους

Απαραίτητες γνώσεις

Πάρα πολύ καλή γνώση Προγραμματισμού (προτιμάται η γλώσσα C), Δομών Δεδομένων, Αλγοριθμικής Πολυπλοκότητας, καθώς και η δυνατότητα προγραμματισμού σε πολλούς πυρήνες ή/και νήματα (multithreading and/or multiprocessing programming)

Βιβλιογραφία και αναφορές

- Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. In The adaptive web (pp. 291-324). Springer, Berlin, Heidelberg.
- Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. Advances in artificial intelligence, 2009.
- Herlocker, J. L., Konstan, J. A., & Riedl, J. (2000, December). Explaining collaborative filtering recommendations. In Proceedings of the 2000 ACM conference on Computer supported cooperative work (pp. 241-250).
- . . .

Πλήθος φοιτητών

1 ή 2 άτομα του Τμήματος Ψηφιακών Συστημάτων.

Επιβλέπων

Διονύσης Μάργαρης (margaris@uop.gr), Επίκουρος Καθηγητής ΤΨΣ.

Απαραίτητη η επικοινωνία με τον επιβλέποντα, πριν δηλώσετε το παρόν θέμα, ώστε να σας είναι ξεκάθαρη η ακριβής έκταση της εργασίας και οι απαιτήσεις της.