

Εργασία 5 – Recommender Systems

Θα αναπτύξετε ένα σύστημα συστάσεων με χρήση των δεδομένων από το MovieLens σύνολο δεδομένων (<https://grouplens.org/datasets/movielens/latest/>). Θα χρησιμοποιήσετε το μικρό σύνολο με τις 100000 βαθμολογίες.

Θα αναπτύξετε υβριδικό σύστημα συστάσεων με δύο recommenders χρησιμοποιώντας pipeline, και συγκεκριμένα cascade, έτσι ώστε η έξοδος του 1ου recommender να χρησιμοποιείται ως είσοδος στον 2ο (δλδ. Ο δεύτερος recommender δεν θα εισάγει νέα αντικείμενα στις συστάσεις).

Οι δύο recommenders θα υλοποιούν συνεργατικό φιλτράρισμα (collaborative filtering), και συγκεκριμένα:

RecA: user-user collaborative filtering, με pearson correlation και συνάρτηση πρόβλεψης το σταθμισμένο μέσο όρο των k κοντινότερων γειτόνων αφαιρώντας το bias του κάθε γείτονα (slide 25, L10).

RecB: item-item collaborative filtering, με cosine similarity και συνάρτηση πρόβλεψης το σταθμισμένο μέσο όρο των k κοντινότερων γειτόνων.

Και τα δύο συστήματα θα προτείνουν τα N κορυφαία αντικείμενα.

Θα συγκρίνετε την απόδοση δύο υβριδικών συστημάτων. Το πρώτο θα χρησιμοποιεί το pipeline RecA → RecB, και το δεύτερο RecB → RecA.

Για την αποτίμηση των συστημάτων, χωρίστε τα δεδομένα σας σε δύο μέρη, σύνολο εκπαίδευσης και σύνολο ελέγχου. Το δεύτερο σύνολο κρατήστε το σταθερό στο 10% των αρχικών δεδομένων.

Τα συστήματα θα συγκριθούν ως προς

α) MAE

β) average precision και

γ) liftindex.

Πάρτε μέσους όρους για όλες τις προβλέψεις στο σύνολο ελέγχου.

Πραγματοποιήστε τα εξής πειράματα:

A. Για σταθερό $K=30$ και $N=100$ εκπαιδεύστε το σύστημα σας χρησιμοποιώντας το 10% των δεδομένων σας για εκπαίδευση, το 20%, το 30%, ..., το 90%. Σε όλες τις περιπτώσεις διατηρείται ένα άλλο 10% για τον έλεγχο. (Όλα τα ποσοστά είναι επί του αρχικού).

B. Για σταθερό ποσοστό εκπαίδευσης 90% και $N=100$ δοκιμάστε με διαφορετικά K , $K=10$, 50 και 100.

Γ. Για σταθερό ποσοστό εκπαίδευσης 90% και $K=30$ δοκιμάστε διαφορετικά N , $N=20$, 50, 100, 150 και 200.

Παρουσιάστε και σχολιάστε τα αποτελέσματα των τριών πειραμάτων. Σε σχέση με τον κώδικα σας, σχολιάστε όποια βελτιστοποίηση ή υπόθεση κάνατε για ότιδήποτε δεν καθορίζεται από την εκφώνηση.