

Large Scale Recommender Systems (Collaborative filtering, k nearest neighbors)

Naman Goel

One of the algorithms used for implementing recommender systems is called the “user based collaborative filtering”. The input to this algorithm is a user, item matrix r . Entry $r_{i,j}$ of the matrix represents the rating given by user i to item j . The algorithm finds k nearest neighbors of every user by computing a similarity matrix of the users. Once the set of users that are similar to a given user are known, recommendations can be made based on the assumption that similar user like similar items. However, in a naive implementation, as the number of users and items grows, this pairwise similarity computation becomes costly and storing the large similarity matrix in memory also becomes difficult.

The goal of this project is to make large scale recommendation systems faster. In the first part of the project, student will implement the classic user based collaborative filtering and understand the scalability problem with large datasets. Optionally, the student may implement a distributed version of the algorithm with several machines and understand the difficulty in distributing this algorithm. In second part of the project, the student will study the sparse nature of several real world datasets. The third part will involve providing an empirical proof of the possible speedups (and simplifications in distributed implementation) assuming that there exists an efficient way to exploit sparsity. The last part of the project will be devoted to developing an efficient way to exploit sparsity to actually achieve this speedup in practice. This speedup should not be at the cost of drop in quality of recommendations.

Required Skills : Good in programming. Knowledge of basic machine learning concepts like k-fold cross-validation, k-nearest neighbors and familiarity with basic graph theory.

Preferred : Experience with handling large datasets.

Duration : 3-4 months

References :

1. Ekstrand, Michael D., John T. Riedl, and Joseph A. Konstan. "Collaborative filtering recommender systems." *Foundations and Trends in Human-Computer Interaction* 4.2 (2011): 81-173.
(Available at : <http://files.grouplens.org/papers/FnT%20CF%20Recsys%20Survey.pdf>)