

CS532 Project/Lab Proposal (Fall 2015)

Tentative Title: Techniques for developing Recommender Systems

Team Members

- Lokesh Jindal
- Mehreen Ali
- Urmish Thakker

Brief Overview and Motivation

The goal of a Recommender System is to generate meaningful recommendations to a user for items or products that might interest him/her. Suggestions for music on Spotify, books on Amazon, or movies on Netflix, are real world examples of widely used recommender systems. Recommender Systems differ in the way they analyze these data sources to develop notions of affinity between users and items which can be used to identify well-matched pairs. The design of such recommendation engines depends on the domain and the particular characteristics of the data available.

Recommender Systems could be developed using three methods. These include Collaborative Filtering, Content Based Filtering and Hybrid Recommender Systems. We will look into Collaborative Filtering and Content Based Filtering in this project.

Collaborative Filtering explores techniques for matching people with similar interests and making recommendations based on this. A key problem in collaborative filtering is to come up with an approach to combine and weigh the existing preferences of various users for different items to generate a recommendation for a new user or a new item. We will study two methods addressing this issue - User Based Collaborative Filtering and Item Based Collaborative Filtering.

Content Based Filtering, also referred to as cognitive filtering, recommends items based on a comparison between the content of the items and a user profile. The content of each item is represented as a set of descriptors or terms, for example, the words that occur in a document. The user profile is represented with the same terms and built up by analyzing the content of items which have been seen by the user. Here, we would look into techniques like - Bayesian Networks.

Core Concepts and tools

- **Pearson Coefficient, Vector Cosine for similarity:** These techniques are used to estimate similarity between users and items.
- **Matrix Factorization:** This technique can be used to discover the latent features underlying the interactions between users and items. This helps in making a new recommendation to a user based on his/her existing preferences. It is also used in Low Rank Matrix Completion.
- **Singular Value Decomposition (SVD) and Principal Component Analysis (PCA):** In the case of recommender systems, the data is represented as user-product relationships, with potentially millions of users and hundred of thousands of products. Using the tools like SVD and PCA, dimensionality reduction will allow concise representation of data for efficient predictions.
- **Least Squares:** Basic tool that will be used for solving system of equations and developing prediction models.

Related Papers, Datasets, or Resources

- Book Crossing Dataset - <http://www2.informatik.uni-freiburg.de/~ctiegle/BX/>
- Movie Lens Dataset - <http://grouplens.org/datasets/movielens/>
- Coursera Course - <https://www.coursera.org/learn/recommender-systems>
- Content-based Recommender Systems: State of the Art and Trends - <http://facweb.cs.depaul.edu/mobasher/classes/ect584/Papers/ContentBasedRS.pdf>

- Collaborative Filtering Recommender Systems - <http://files.grouplens.org/papers/FnT%20CF%20Recsys%20Survey.pdf>
- Keshavan R. H., Montanari A., Oh S (2009). Lowrank Matrix Completion with Noisy Observations: a Quantitative Comparison.
- Wikipedia: https://en.wikipedia.org/wiki/Collaborative_filtering
- Toby Segaran. 2007. *Programming Collective Intelligence* (First ed.). O'Reilly.