CS533 Project Plan

Group Members: Basri Kahveci, Christina Kirchner, Burak Kocuroğlu

a) Description of the Problem

Recommendation systems are a subclass of information filtering system that try to predict the 'rating' or 'preference' that a user would give to an item (such as music, books, or movies) or social element (e.g. people or groups) they had not yet considered, using a model built from the characteristics of an item or the user's social environment.

b) Motivation

Obtaining recommendations from trusted sources is a critical component of the natural process of human decision-making. With growing consumerism and increasing popularity of the Internet, buyers are being presented with an increasing range of choices. When a query is made in amazon.com, such as “Java books”, system returns hundreds of books. Or, when a user goes to a movie renting web site, she can take a look at top ten movies. But if most of those movies are the movies the user does not like, that list is not a good list for that user. Sellers must offer relevant products to consumers. The more relevant products are offered to buyers, the more likely they buy products. So it has become important for enterprises to collect large volumes of data that allows for deeper analysis of how a customer base interacts with the products. Recommendation Systems have evolved to produce automated generation of recommendations based on data analysis. Netflix Prize [1] shows how important could be to increase the performance of recommendation systems. They had offered $1 million to any group that can beat Netflix’s own algorithm by 10%.

We are going to make a movie recommendation system using a dataset taken from an online movie web site. We aim to find relevant recommendations for users. Thus we increase customer satisfaction of an online service.

c) Methodology

There are two different kinds of recommendation system approaches: collaborative filtering and content-based filtering. Collaborative filtering systems recommend items based on similarity measures between users and/or items. The items recommended to a user are those preferred by similar users. Content-based systems examine properties of the items previously preferred. For instance, if a Netflix user has watched many comedy movies, then a movie not seen by the user having the comedy genre is recommended.

We are planning to use a hybrid recommendation system that combines both approaches. We will implement both approaches separately then combine their rankings.

We are going to implement our system using Java and Apache Mahout [2].

d) Expected results

We are going to build our recommendation system using 31289 ratings for training and 8322 ratings for testing. We are going to evaluate our results by using root mean square error, precision and recall values and compare results of content based and collaborative filtering approaches.

e) References

[1]: http://en.wikipedia.org/wiki/Netflix_Prize

[2]: http://mahout.apache.org/

Jeffrey D. Ullman, Mining of Massive Datasets (http://infolab.stanford.edu/~ullman/mmds.html)

Prem Melville , Vikas Sindwani, Recommender Systems