

628 Project 2

Chenxi He, Shengqian Chen, Wenxue Zhang

Introduction:

Recommender systems are usually classified according to their approach to rating estimation. In general we can distinguish between content-based and collaborative recommender systems. Content-based recommendations are typically based on item similarity to objects the user preferred in the past. In contrast, collaborative recommendation systems depend on the ratings given by individuals with similar taste and preference. However, both techniques exhibit specific drawbacks.

Recommendations based on content-based techniques tend to overspecialize, because only items with a high similarity to those already rated will be suggested to the individual user. For example a teenager clicks like button on several cartoon, the only thing recommended by the system is much likely to be cartoon again instead of other kind of movies. Another problem with content-based recommenders is that a user first has to rate a sufficient number of items before the system is able to make accurate recommendations. Collaborative systems have the drawbacks that they suffer from the item cold-start problems which occur when recommendations must be made on the basis of few recorded ratings.

A common approach to solve the problems of the above techniques is to combine both content-based and collaborative information into a hybrid recommender system.

Key questions:

How to build a hybrid recommender system to improve the accuracy?

Plan of approach:

First we try to detect the connection between rating tendency and users' personal information and movie category. Will older people give an overall lower rating for Adventure? Does gender significantly affect the rating trend? And what kind of movie tend to have higher rating? Based on the initial data visualization, we find that various occupation and age of viewers tend to have different taste. And thus it's reasonable to assume that if we take the personal information into account and build a hybrid recommender system would result a better prediction for personal preference.

For the movie content, it is natural that we will recommend a movie with the same director, movie stars or type as the movie someone rated high scores. So we also want to consider this part as one part of the prediction ratings.

So we want to include the user information and movie content along with the ratings into the prediction algorithm. The hybrid algorithm is intended to combine user-based collaborative filtering and content based filtering.

To realize this algorithm, we need to know how to actually take them into account to compute a rating score. Another problem will be how to code the algorithm.

