

There has been a good deal of research over the last 20 years on how to automatically recommend things to people and a wide variety of methods have been proposed. For our group, the research is primarily focused on the recommender methods and system building.

Key Question: How to build a trusty movie recommender system?

Data sources are obtained from the database of MovieLens, and composed of three parts: Users information, movies' genre and rating. (Users included in our data are at least rating for 20 movies.) The next several weeks, by mining users' history and stated preferences, we want to develop our recommending system separately for different groups. Also, some other conditions are discussed: how can we implement the recommendation for people with no movie records? If we base a recommendation on ratings of others, how to deal with people whose overall rating is quite low?

Therefore, first we divide people into two parts: new users and old users.

For those new users, people who have no movie records on this website, their preference on movies are unknown. However, according to other users' movie records and rating data, we can deduce those new users' preference from the sharing characteristics of people. For example, ages, genders, occupations. It is assumed that people of similar characteristics might have similar preferences. We then can find different patterns for different groups and fit models to predict new users' preference on movies and perform the recommendation. Another more intuitively way is to recommend movies by other users' historical data. Movie that has been rated highest and watched most times should be recommended to others.

For those older users, people who have already rated some movies before, we'll mainly concentrate on their ratings and watching history. Here we have two methods as well: The first one is simpler, only focus on the similarity of movies—finding movies similar to the ones liked by a user using textual similarity. And to determine which movie the user actually likes, we should first consider the rating data, filter out those low-rating movies. The second one is similar to the method we mentioned above for the new users. By dividing people into different clusters, people in the

same cluster might have similar preference on movies. We'll recommend movies based on group preference. The fundamental assumption behind this method is that other users' opinions can be selected and aggregated in such a way as to provide a reasonable prediction of the active user's preference.

