협동적 필터링 기반 고도화된 추천 시스템

김택헌° 박석인 양성봉

연세대학교 컴퓨터과학과

{kimthun, psi93, yang}cs.yonsei.ac.kr

Collaborative Filtering-based High Quality Recommender Systems

Taek-Hun Kim° Seok-In Park Sung-Bong Yang

Dept. of Comupter Science, Yonsei University

There are a myriad of goods traded in commercial transactions both online and off-line. These goods include physical products as well as digital contents. It is not easy for a customer to search the best suitable goods among almost countless items in the Internet. In order to provide valuable information on the items to the customers who have various preferences, we need an effective personalized recommender system.

A personalized recommender system predicts the best suitable goods for the customers according to their individual preferences and recommends the predicted results to them. Thus the customers can save time and efforts in searching the items they want.

It is very important for a recommender system to have a capability to predict accurately by analyzing the preferences of the customers. A recommender system utilizes in general an information filtering technique called collaborative filtering, which is based on the ratings of other customers who have similar preferences and is widely used for many online commercial web sites[1][2].

A recommender system using collaborative filtering which we call it CF, calculates the similarity between the test customer who is supposed to obtain a recommendation from the recommendation system and each of other customers who have rated the items that are already rated by the test customer. Since CF is based on the ratings of the neighbors who have similar preferences, it is very important to select the neighbors properly to improve prediction quality.

There have been many investigations in selecting proper neighbors based on neighbor selection methods such as the k-nearest neighbor selection, the threshold-based neighbor selection, and the clustering-based neighbor selection. They are quite popular techniques for recommender systems[2][3]. These techniques then predict customer's preferences for the items based on the results of the neighbors' evaluation on the same items.

In clustering-based CF, the performance can be quite good, since the size of a cluster that must be analyzed is much smaller. Therefore, the clustering-based method can solve the very large scale problem in recommender systems[3][4]. Although the clustering method is fast and useful

to cluster a large scale dataset, it has a limitation for improving prediction quality.

In this paper we present that the recommender systems using the refined neighbor selection and the attributes of the items for high prediction quality. The proposed method consists of two stages for the prediction. In the first stage we use the refined neighbor selection algorithm that exploits the transitivity of similarities and dissimilarities using a graph approach. And then we exploit the attributes of each item and apply those to the prediction process in the second stage. We also show a neighbor selection method that uses an alternate neighbor of the test customer who may have an unusual preferences or who is the first rater.

We use the MovieLens dataset for the experiments[5]. The experimental results show that the proposed systems select meaningful neighbors for the high prediction quality and the prediction quality is affected by using the attribute information for the test customer and the neighbors. Therefore the clustering-based recommender systems using the proposed algorithms could resolve the very large scale dataset problem without deteriorating prediction quality and could solve the first rater problem in CF-based recommender systems.

From the experimental results, we could observe the followings:
- The neighbor selection based on the transitivity of similarity can select meaningful customers as the neighbors.
- The neighbor selection using a virtual neighbor and an alternate neighbor is useful to select valuable neighbors in collaborative filtering.
- The recommender systems using the attribute information have improved the prediction quality, compared with the ones without using it.
- The clustering-based collaborative filtering using the proposed method can be an answer to the large scale dataset problem and also the first rater problem without deteriorating prediction quality.

References

[1] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. T. Riedl, "Application of Dimensionality Reduction in Recommender System - A Case Study," Proc. ACM WebKDD 2000 Web Mining for E-Commerce Workshop, 2000.

[2] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. T. Riedl, "Analysis of Recommendation Algorithms for E-Commerce," Proc. ACM E-Commerce 2000, 2000.

[3] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. T. Riedl, "Recommender Systems for Large-Scale E-Commerce: Scalable Neighborhood Formation Using Clustering," Proc. The Fifth International Conference on Computer and Information Technology, 2002.

[4] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," Proc. UAI 1998 Conference, 1998.

[5] MovieLens dataset, GroupLens Research Group, url: http://www.grouplens.org/.