An Algorithmic Framework for Adaptive Collaborative Filtering

KiHong Kim, SangKu Kim, Sung Joo Park

Department of Management Engineering,
KAIST Graduate School of Management
email : gammir@kgsm.kaist.ac.kr
Tel : 02-958-3647
Addr : 207-43 KAIST, ChonRyangRi Dong,
DonDaeMun Gu, Seoul (130-012)

Many of us are already overwhelmed with the amount of information we must process every day. For this reason, there is a growing interest in recommendation systems that suggest music, film, books, and other products and services to users in internet shopping malls. Collaborative Filtering (CF) and its variations have seen considerable success on the Internet, being used at sites like Amazon.com, CDNow, and so forth.

Despite the usefulness of typical Collaborative Filtering techniques, they have some limitations. In traditional CF process all the ratings in database are treated with equal importance no matter whether it is rated today or one year ago. This is problematic, because ratings reflect user's interest in a specific circumstance, and the circumstance may change easily, and this change often makes a user's past rating obsolete. Therefore each rating has different level of importance in estimating user's current preference.

In this paper, we present a variation of Collaborative Filtering that captures the difference of importance among the ratings, and utilize the captured knowledge to increase the adaptability of Collaborative Filtering.

Most CF recommendation consists of two steps. First, we find the Neighbors of active user. Active user is one to whom a recommendation is to be made, and Neighbors are other users who have similar interest with the active user. Second, we predict the ratings of the active user based on the ratings of Neighbors. We focused on the first step.

The original CF user database consist of a set of ratings $v_{i,j}$ corresponding to the rating of user i on item j. We added one more variable $\theta_{i,j}$ for each $v_{i,j}$, which represents the relative importance of corresponding rating $v_{i,j}$, and used it to decrease the gap between user's current interest and his/her rating set. By this way, $\theta_{i,j}$ is used as the adaptive weight of $v_{i,j}$.

$\theta_{i,j}$ is obtained by using one of the Machine Learning Techniques. Hebbian is widely used in fields where too many variables exist and no model-based approach can be applied. Under the concept of Hebbian Learning, the adaptive weight $\theta_{i,j}$ is rewarded if $v_{i,j}$ contributes to successful recommendation, and it is penalized if not. Whenever new rating is entered, the adaptive weight of the rater's items are updated

Once we have the adaptive weights, we apply it to our new version of CF algorithm to find active user's neighbors. Pearson's Correlation (PC) is known to be a best performing method for finding neighbors, but does not discriminate ratings with different importance. For this reason, we modified PC to reflect the adaptive weights. In modified PC, we multiply $\theta_{a,j}$ by the ratings of active user and neighbor candidates before calculating the correlation. As a result, the weighted correlation is more influenced by the ratings with higher importance, and the influence of less important ratings can be reduced, and this leads to better recommendation performance.

To verify our algorithm's performance, we implemented and experimented it on EachMovie dataset. The experiment result was compared to original collaborative filtering based on various performance metrics. Suggested algorithm over-performed the original CF by the measure of NDPM and MAE.