

A New Collaborative Filtering Method for Movie Recommendation Using Genre Interest

Soojung Lee

Dept. of Computer Education, Gyeongin National Univ. of Education

영화 추천을 위한 장르 흥미도를 이용한 새로운 협력 필터링 방식

이수정

경인교대 컴퓨터교육과

Abstract Collaborative filtering has been popular in commercial recommender systems, as it successfully implements social behavior of customers by suggesting items that might fit to the interests of a user. So far, most common method to find proper items for recommendation is by searching for similar users and consulting their ratings. This paper suggests a new similarity measure for movie recommendation that is based on genre interest, instead of differences between ratings made by two users as in previous similarity measures. From extensive experiments, the proposed measure is proved to perform significantly better than classic similarity measures in terms of both prediction and recommendation qualities.

Key Words : Recommender system, Similarity measure, Collaborative filtering, Memory-based collaborative filtering, Content-based filtering

요약 협력 필터링은 상업적 추천 시스템에서 널리 사용되어 왔는데, 고객의 사회적 행태를 구현하여 사용자의 흥미에 부합하는 항목들을 제안하기 때문이다. 현재까지 적절한 항목을 추천하기 위한 가장 보편적인 방법은 유사한 사용자들을 찾아 그들의 평가치를 참조하는 방법이다. 본 논문은 영화를 추천하기 위해서 장르 흥미도를 기반으로 하는 새로운 유사도 공식을 제안하는데, 이는 기존 공식에서 사용자들의 평가등급 차이를 기반으로 하는 것과 대비된다. 광범위한 실험 결과에 따르면, 제안한 공식은 정확도와 추천의 질에 있어서 전통적인 유사도 공식의 성능을 크게 향상시키는 것으로 확인되었다.

주제어 : 추천 시스템, 유사도 공식, 협력 필터링, 메모리 기반 협력 필터링, 내용 기반 필터링

1. Introduction

Collaborative Filtering (CF) is a popular method for

recommending items in commercial systems to help online customers by suggesting items that might fit to

Received 16 May 2014, Revised 21 July 2014
Accepted 20 August 2014
Corresponding Author: Soojung Lee(Dept. of Computer Education, Gyeongin National Univ. of Education)
Email: sjlee@gin.ac.kr

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ISSN: 1738-1916

the interests of a user. It has been successfully utilized and implemented in various systems such as the Tapestry system, GroupLens, Video Recommender, Ringo, Amazon.com, the PHOAKS system, and the Jester system [1].

At large, two classes of algorithms have been used for collaborative recommendation, memory-based (or heuristic-based) and model-based [1]. This classification is made based on the method of estimating ratings for unrated items. In order to make rating predictions, memory-based systems refer to the history of ratings made by the users. Therefore, these systems may result in poor recommendations in case past rating records are scarce, which is referred to as the *data sparsity* problem. In contrast, model-based algorithms try to establish a model by learning complex patterns from a collection of ratings and then make predictions based on the learned model. They use various methods from data mining, statistical probabilities, or machine learning [1]. Some of the techniques reported in the literature include clustering [2], Bayesian networks [3], logistic regression [4], probabilistic latent semantic analysis [5], and genetic algorithm[6][7]. Although model-based systems can alleviate the data sparsity problem of memory-based systems, they require expensive model-building procedures and their performance is highly dependent on the estimated parameters and assumptions used in the model.

This paper focuses on memory-based algorithms, especially on the application of movie recommendation. One may use various characteristics related to movies, such as actors, director, genre, year of release, etc., in making recommendations, which obviously requires additional expenses from general collaborative filtering methods. Recently, user preference or importance of items and users has been calculated to be reflected on similarity estimation [8][9][10]. However, these works are usually integrated into the traditional similarity measures or require extra process to convert the

user-item rating matrix. The proposed method does not require extra information on user-preferred types of movies, but simply utilizes the number of ratings made by users for each genre, which is interpreted as the genre preference of the users. With this simple treatment, i.e., considering ratings not as a whole, but grouped into each genre, we confirmed that performance of recommendations is greatly enhanced, as proved through several experiments. Its main contribution is to reveal that utilization of only the features other than rating differences between users significantly improves performance of CF methods which have been solely dependent on rating differences.

2. Memory-based Collaborative Filtering

In order to recommend a yet unrated item to a user, memory-based CF methods predict the rating of the item with respect to the user. If the predicted rating is among the highest, the corresponding item should be recommended to the user. Prediction is usually made in two different ways, *user-based* and *item-based*. The former method incorporates ratings given to the item by those users who are most similar to the user, whereas the latter refers to the ratings given by the user, of those items similar to the item. Therefore, similarity calculation is one of the most critical aspects of CF systems, which influences recommendation performance.

Most common approaches to measure similarity between two users or two items are correlation-based and vector cosine-based. The Pearson correlation and its variants, constrained Pearson correlation and Spearman rank correlation, are examples of widely-used correlation-based approaches [1]. Cosine-based method treats each user or item as a vector and measures an angle between the vectors. Although these approaches are popular in CF systems, they suffer from several drawbacks that they show

poor performance with a few co-rated items or when a user gives an identical rating for all co-rated items. Detailed analysis on traditional similarity measures can be found in [11].

Once similarity is calculated between every pair of users or items, rating prediction for an unrated item is usually made by taking the similarities as a weighted average. A well-known prediction formula is suggested by [12] as

$$\hat{r}_{u,x} = \bar{r}_u + \frac{\sum_{v \in NN(u)} \text{sim}(u,v) \cdot (r_{v,x} - \bar{r}_v)}{\sum_{v \in NN(u)} |\text{sim}(u,v)|},$$

where $NN(u)$ is a set of nearest neighbors, i.e., the users most similar to user u , who has rated item x . $r_{v,x}$ is the rating of item x given by user v , and \bar{r}_u and \bar{r}_v denote the mean of the ratings of all the co-rated items of users u and v , respectively.

3. Proposed Similarity Measure

3.1 Motivation

In this section, we investigate how preference for genres is related to similarity, i.e., whether genre interests between similar users are more in common than between less similar users. The dataset used for this experiment is MovieLens (<http://www.grouplens.org>), where information on genres which a movie belongs to is also provided. Since no specific information regarding genre preference of users can be obtained from the dataset, we assume that the number of ratings made by a user to movies belonging to a genre indicates the level of preference of the user for that genre; the more ratings obviously implies the higher interest.

Let I be the set of all items or movies and G be the set of genres provided in the dataset. We define the following terms to introduce metrics used in this

experiment.

- $I(u) = \{i \in I \mid r_{u,i} \neq \text{null}\}$
- $G_i = \{g \in G \mid i \in I \text{ belongs to genre } g\}$
- $G(u) = \{g \in G_i \mid i \in I(u)\}$

To explain the above terms, $I(u)$ represents the set of items to which user u has given a rating and G_i refers to the set of genres where item i belongs. $G(u)$ is the set of genres, at least one of whose items user u has rated.

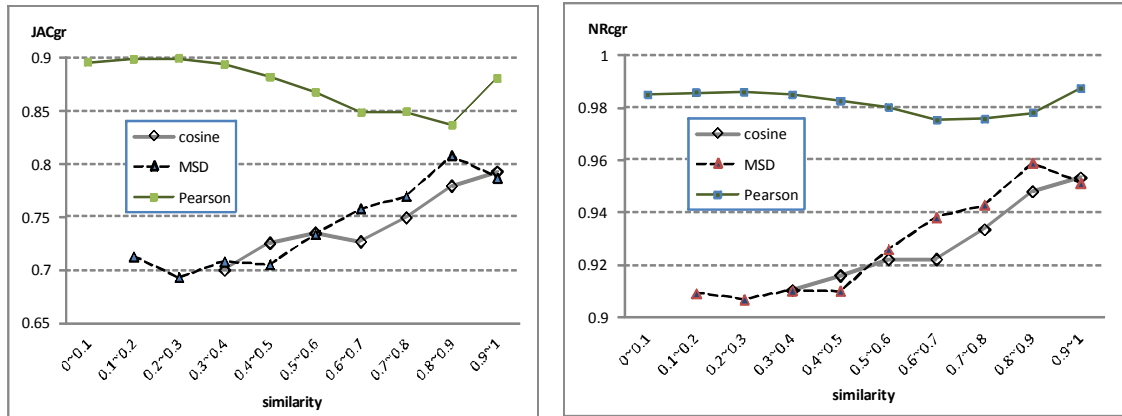
With the above definitions, two metrics are defined as follows.

$$JAC_{gr}(u,v) = \frac{|G(u) \cap G(v)|}{|G(u) \cup G(v)|}$$

$$NR_{cgr}(u,v) = \frac{1}{|I(u)| + |I(v)|} \cdot (|\{i \in I(u) \mid G_i \cap (G(u) \cap G(v)) \neq \emptyset\}| + |\{i \in I(v) \mid G_i \cap (G(u) \cap G(v)) \neq \emptyset\}|)$$

$JAC_{gr}(u, v)$ represents the relative number of genres commonly interested by both of the two users. While JAC_{gr} measures the common interests of the two users in terms of genres, $NR_{cgr}(u, v)$ measures those in terms of the number of ratings. i.e., how many ratings are made on the genres of common interests of the two users.

[Fig. 1] shows the changing patterns of the two metrics according to similarity. Three popular similarity measures are used in this experiment on MovieLens dataset; cosine similarity, Pearson correlation, and Mean squared difference (MSD). It is observed that there seems no big difference between the behavior of JAC_{gr} and NR_{cgr} . This is an obvious result, considering that a user gives ratings to more movies in his interested genres than to movies out of his interested genres. Moreover, cosine similarity and MSD tend to behave almost similarly and justify our conjecture that similar users would have more common interests in genres. However, Pearson correlation seems to act negatively on this argument, since JAC_{gr}



[Fig. 1] Relationship between traditional similarity measures and proportions of common genre interests in terms of JAC_{gr} (left) and NR_{cgr} (right)

and NR_{cgr} are almost stable or show no positive relation with Pearson similarity. This is probably because of the shortcomings of Pearson correlation, as described in [11], that it may yield abnormal results in case of a very few co-rated items. Overall from these experiment results, we believe that genre preference may be utilized in similarity calculation for collaborative filtering, especially for non-scarce datasets.

3.2 Description

Following the observation from the previous subsection, the proposed method takes difference between genre preferences of users into account, in calculating similarity between users. The genre preference is measured through the number of ratings; high number of ratings to those movies in a specific genre implies that much preference for that genre. For instance, assume that users *u* and *v* have similarity of 0.6 with respect to ratings of items in genre *g*₁. Also assume users *u* and *w* have the same similarity of 0.6 with respect to those items in genre *g*₂. However, if user *u* rates more items in genre *g*₁ than those in genre *g*₂, then we regard that user *u* has higher similarity with user *v* than user *w*.

Using the notations described in the previous

subsection, genre preference of user *u* for genre *g*, denoted as P_{*g*}(*u*), is defined as follows.

$$N_g(u) = \frac{|\{i \in I(u) | g \in G_i\}|}{\max_{g \in G} |\{i \in I(u) | g \in G_i\}|}$$

$$P_g(u) = \frac{2}{1 + \exp(-6 \cdot N_g(u) + 6)}$$

N_{*g*}(*u*) represents the relative number of ratings made to items in genre *g*, out of maximum number of ratings made among all the genres. Then preference for genre *g*, P_{*g*}(*u*), is defined by using the logistic function, as an attempt to place drastically higher weight to those genres having ratings close to the maximum.

Besides the genre preference, we incorporate Jaccard index into the proposed similarity measure. It quantifies how many ratings are given to items in common by the two users. Although the index does not take the rating difference between two users into account, it is reported to increase performance when implemented with a traditional similarity measure [13]. Hence, we believe that its inclusion into the proposed measure would affect the performance favorably. This claim is verified through the experiments described in the next section. However, different from the formula suggested in [14], we adopt Jaccard index by confining it to each

genre separately. That is, the index for genre g between two users u and v , denoted as $Jaccard_g(u, v)$, is defined as

$$Jaccard_g(u, v) = \frac{|\{i \in I(u) \cap I(v) \mid g \in G_i\}|}{|\{i \in I(u) \cup I(v) \mid g \in G_i\}|}$$

Consequently, the final proposed similarity between two users u and v , denoted as $GPRF(u, v)$, is defined as the combination of the two factors, preference for each genre and relative number of co-rated movies in that genre, as follows.

$$GPRF(u, v) = \sum_{g \in G} P_g(u) \cdot Jaccard_g(u, v)$$

4. Performance Experiments

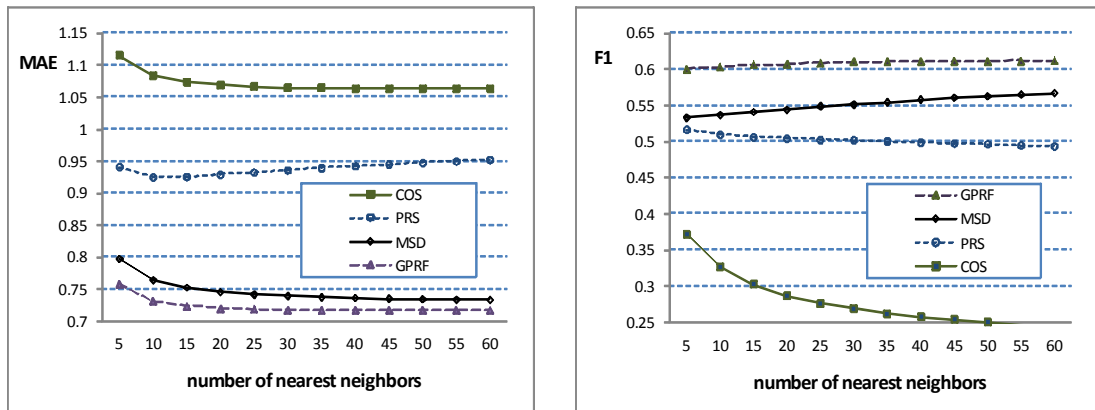
To evaluate performance of the proposed measure, various experiments were conducted. As our idea is for movie recommendation, the MovieLens dataset is selected, as it has been popularly used in the related study and publicly open for research purpose. The dataset consists of 1,000,209 ratings of 1~5 scale made by 6,040 users on 3,900 movies approximately. To secure the reliability of the result, we implemented 5-fold cross validation of the results where each experiment was repeated 5 times with different training and testing data. 80% of each user's ratings were used for training, i.e., for similarity calculation, and prediction of ratings was made on the rest 20% of testing set.

The accuracy of prediction is usually measured through several metrics in literature such as MAE (Mean Absolute Error) and RMSE (Root Mean Squared Error) [15]. In this paper, we use MAE to estimate the prediction precision of our method. $MAE = \sum_i |r_i - \hat{r}_i| / N$ for N predictions \hat{r}_i for the corresponding real ratings r_i . Besides measuring prediction quality, we also evaluate the methods in

terms of recommendation quality through F1 metrics [16]. It is a comprehensive indicator that combines both the precision and the recall with equal weights. Precision is the percentage of truly relevant ratings among those predicted as relevant by a recommendation system. Recall is the percentage of truly relevant ratings predicted by the system among all of the truly relevant ratings.

We conducted experiments to compare performance of most commonly used measures in related studies, Pearson correlation (PRS), the cosine similarity (COS), Mean squared difference (MSD), and the proposed metric GPRF. In [Fig. 2], it is observed that every measure demonstrates gradually stabilized MAE along with increasing number of NNs, which implies that consulting large number of neighbors does not contribute to better prediction accuracy. MAE results between the previous measures differ greatly, where MSD significantly outperforms the other two. However, GPRF performs best by almost 33% over COS, 19~24% over PRS, and 3~5% over MSD. This is quite notable considering that different from traditional measures, GPRF does not take rating differences of co-rated items into consideration.

To investigate F1 performance of the measures, it is seen that the results are much similar to those of MAE. However, COS demonstrates far lower results than all of the others, while GPRF yields much higher performance compared to even MSD, where around 9~13% improvement has been achieved throughout varying number of nearest neighbors. Consequently, all of these results demonstrate that considering only preference for movie genres and number of co-rated items in each genre to estimate similarity, as suggested in this research, surely enhances both prediction and recommendation accuracy in collaborative filtering systems.



[Fig. 2] Performance comparison results of various similarity measures in terms of MAE (left) and F1 (right) for varying number of nearest neighbors

5. Conclusions

Collaborative filtering-based recommendation systems have been very popular these days. For the effectiveness of these systems, the underlying similarity measures play a critical role. The present work suggests a new similarity measure for user-based collaborative systems. Its distinctive property that is differentiated from the classic measures is that it reflects only users' preference for genres and number of co-ratings per genre onto similarity calculation. With such approach, it is quite notable that the proposed measure significantly outperforms previous similarity measures in terms of both prediction and recommendation qualities in our experiments, considering that it does not utilize differences between ratings made by two users as in traditional similarity measures. This implies that there is much room for further improvement of performance in collaborative filtering methods. Therefore, our future work includes case analysis of low performance in both proposed and traditional measures to integrate them effectively while supplementing shortcomings of each.

REFERENCES

- [1] G. Adomavicius, & A. Tuzhilin, Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, *IEEE Transactions on Knowledge & Data Engineering*, IEEE, Vol. 17, No. 6, pp. 734 - 749, 2005.
- [2] S. Han, S. Chee, J. Han, & K. Wang, RecTree: An efficient collaborative filtering method, *The 3rd International Conference on Data Warehousing and Knowledge Discovery*, pp. 141-151, 2001.
- [3] X. Su, & T. M. Khoshgoftaar, Collaborative filtering for multi-class data using belief nets algorithms, *The 13th International Conference on Tools with Artificial Intelligence*, IEEE, pp. 497 - 504, 2006.
- [4] R. Greiner, X. Su, B. Shen, & W. Zhou, Structural extension to logistic regression: Discriminative parameter learning of belief net classifiers, *The Eighteenth Annual National Conference on Artificial Intelligence*, pp. 167 - 173, 2002.
- [5] T. Hofmann, Latent semantic models for collaborative filtering, *ACM Transactions on Information Systems*, ACM, Vol. 22, No. 1, pp. 89-115, 2004.
- [6] D. Anand, & K. K. Bharadwaj, Adaptive user similarity measures for recommender systems: A

- genetic programming approach, The 3rd IEEE International Conference on Computer Science and Information Technology, pp. 121-125, 2010.
- [7] J. Bobadilla, F. Ortega, A. Hernando, & J. Alcalá, Improving collaborative filtering recommender system results and performance using genetic algorithms, Knowledge-Based Systems, Elsevier Science Inc., Vol. 24, No. 8, pp. 1310 - 1316, 2011.
- [8] C. Baoxian, M. Fei, & L. Sujuan, A collaborative filtering recommendation algorithm based on user topic preference, International Journal of Advancements in Computing Technology, AICIT, Vol. 4, No. 14, pp. 342-351, 2012.
- [9] J. Bobadilla, A. Hernando, F. Ortega, & A. Gutiérrez, Collaborative filtering based on significances, Information Sciences, Elsevier Science Inc., Vol. 185, No. 1, pp. 1 - 17, 2012.
- [10] X. Tang, W. Deng, & J. Liu, A personalized recommendation method based on comprehensive interest, International Journal of Advancements in Computing Technology, AICIT, Vol. 5, No. 5, pp. 157-164, 2013.
- [11] H. J. Ahn, A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem, Information Sciences, Elsevier Science Inc., Vol. 178, No. 1, pp. 37-51, 2008.
- [12] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, & J. Riedl, GroupLens: An open architecture for collaborative filtering of netnews, Proceedings of the ACM Conference on Computer Supported Cooperative Work, pp. 175 - 186, 1994.
- [13] J. Bobadilla, F. Serradilla, & J. Bernal, A new collaborative filtering metric that improves the behavior of recommender systems, Knowledge-Based Systems, Elsevier Science Inc., Vol. 23, No. 6, pp. 520-528, 2010.
- [14] G. Koutrica, B. Bercovitz, & H. Garcia-Molina, FlexRecs: Expressing and combining flexible recommendations, Proceedings of the 2009 ACM SIGMOD International Conference on Management of data, ACM, pp. 745 - 758, 2009.
- [15] J. L. Herlocker, J. A. Konstan, L. G. Terveen, & J. T. Riedl, Evaluating collaborative filtering recommender systems, ACM Transactions on Information Systems, ACM, Vol. 22, No. 1, pp. 5-53, 2004.
- [16] M. Gao, Z. Wu, & F. Jiang, Userrank for item-based collaborative filtering recommendation, Information Processing Letters, Elsevier Science Inc., Vol. 111, No. 9, pp. 440-446, 2011.

이 수 정(Lee, Soojung)



- 1985년 2월 : 이화여자대학교 과학교육과(이학사)
- 1990년 12월 : 미국 Texas A&M 대학교 컴퓨터공학과(석사)
- 1994년 8월 : 미국 Texas A&M 대학교 컴퓨터공학과(박사)
- 1994년 9월 ~ 1998년 4월 : 삼성전자 통신개발실 선임연구원
- 1998년 9월 ~ 현재 : 경인교육대학교 컴퓨터교육과 교수
- 관심분야 : 컴퓨터교육, 추천시스템, 웹마이닝
- E-Mail : sjlee@gin.ac.kr