International Journal of Advanced Smart Convergence Vol.9 No.3 42-48 (2020)
http://dx.doi.org/10.7236/IJASC.2020.9.3.42

IJASC 20-3-6

Using Fuzzy Rating Information for Collaborative Filtering-based Recommender Systems

Soojung Lee

Professor, Department of Computer Education, Gyeongin National University of Education, Korea
sjlee@ginue.ac.kr

Abstract

These days people are overwhelmed by information on the Internet thus searching for useful information becomes burdensome, often failing to acquire some in a reasonable time. Recommender systems are indispensable to fulfill such user needs through many practical commercial sites. This study proposes a novel similarity measure for user-based collaborative filtering which is a most popular technique for recommender systems. Compared to existing similarity measures, the main advantages of the suggested measure are that it takes all the ratings given by users into account for computing similarity, thus relieving the inherent data sparsity problem and that it reflects the uncertainty or vagueness of user ratings through fuzzy logic. Performance of the proposed measure is examined by conducting extensive experiments. It is found that it demonstrates superiority over previous relevant measures in terms of major quality metrics.

Keywords: *Collaborative filtering, Fuzzy logic, Similarity measure, Recommender system*

1. Introduction

Recommender systems have been regarded as essential in many of current commercial systems, where customers get assistance in searching products by inquiring the recommendation list. In literature, there have been various techniques of recommender systems developed, collaborative filtering, content-based filtering, demographic filtering, knowledge-based filtering, hybrid filtering, and social network-based filtering [1]. Among these, collaborative filtering (CF) has been the topic of great concern in recent research studies and successfully implemented in many practical commercial sites. The main reason is that it has a big advantage of requiring no information on user or product features which is in reality hard to acquire.

The basic principle of CF technique is to find likeminded people with the current user and recommend items that they have preferred in the past. While this strategy is called user-based CF, an analogous idea applied with respect to products is named item-based CF. These two types of CF constitute so-called memory-based CF [1]. Memory-based CF systems are named as such because they maintain all user ratings in memory.

Most existing similarity measures basically rely on the number of ratings for common items of two users [2]. When two users have few co-rated items, the resulting similarity may not be trustworthy. In reality, most commercial systems maintain an enormous volume of products, thus making the user-item ratings matrix used by CF extremely sparse, which makes it highly probable to generate few co-rated items. Hence, this data

Manuscript Received: June. 16, 2020 / Revised: June. 20, 2020 / Accepted: June. 27, 2020
Corresponding Author: sjlee@gin.ac.kr
Tel: +82-31-470-6298, Fax: +82-31-470-6299
Professor, Department of Computer Education, Gyeongin National University of Education, Korea

sparsity is one of the challenging issues in CF researches.

This study addresses the data sparsity issue and suggests a new similarity measure. We attempt to alleviate the problem inherent in CF systems by utilizing information in addition to ratings of co-rated items. This information includes user rating patterns which consist of user entropy and the probability of each numeric rating. A user showing similar user rating pattern as well as similar numeric ratings for co-rated items will have a high priority as a neighbor. In addition, we take a further strategy for trustworthy similarity that reflects uncertainty of user ratings. This strategy comes from the fact that users typically show vagueness when giving ratings to items [3]. The specific method for this is to transform the user rating pattern and ratings for items into fuzzy values for computing similarity. We conducted extensive experiments to investigate performance of CF systems with the proposed measure applied and found superiority of the proposed measure in terms of prediction and recommendation qualities compared to those of various relevant existing similarity measures.

The remainder of this paper is organized as follows: In the next section, we discuss how previous works handle the issues of CF systems. Section III presents the proposed similarity measure, followed by experimental studies in Section IV. Section V concludes this paper.

2. Related Studies

2.1 Data Sparsity

Memory-based CF systems compute similarity between users based on the ratings they have assigned to the items. Traditional similarity measures such as Pearson correlation, the cosine similarity, and the mean squared differences, are in need of commonly rated items to compute similarity [4]. The main drawback of these measures is that when the number of co-rated items is low, the computed similarity becomes far unreliable.

The above mentioned drawback is mainly caused by data sparsity. This situation occurs inherently in a recommender system which maintains enormous number of items and users can normally rate only a few items. Several approaches have been developed to handle the sparsity problem, e.g., default voting [1] and imputation-boosted algorithms utilizing machine learning, Bayesian multiple imputation, and linear regression [4]. Dimensionality reduction techniques such as matrix factorization and singular value decomposition also address the data sparsity problem by maintaining only significant ratings information and reducing the user-item matrix size, but may require high computational cost and lose some useful information [4].

Simpler solutions for data sparsity reflect the size of intersection item sets onto similarity computation. Herlocker et al. reflect the number of items co-rated by two users on the similarity [5]. T*animoto coefficient* and Jaccard index refer to the ratio of the number of items co-rated by two users to the total number of items rated by them [6]. These indices are usually employed as weight factors for computing similarity. For instance, Bobadilla et al. multiply Jaccard index with the mean squared differences [7] and Saranya et al. combine the index with Pearson correlation [4]. As a result, the resulting measures are reported to bring performance improvement in various aspects. An enhancement of Jaccard index is studied in [8] where it is segmented into three indices depending on the magnitude of ratings of co-rated items.

2.2 Context of User Ratings

User ratings are essentially imprecise and subjective in nature, which are rarely taken into account by similarity measures in the past. Recently, there are some researches which reflect the uncertainty of ratings onto similarity measures. One of the representative schemes to achieve it is by utilizing the fuzzy logic. Son integrated the fuzzy similarity based on the users' demographic data with Pearson coefficient, reporting that his method obtains higher accuracy than other relevant methods [9]. In [10], the authors fuzzified ratings and rating deviations and incorporated them into existing similarity metrics as weights. Other studies applied fuzzy logic to CF systems by introducing a fuzzy linguistic concept [11] or linguistic terms for user ratings.

Some contextual information is extracted from user ratings to reflect it onto similarity. Desrosiers and Karypis compute similarity by using all the ratings instead of using only ratings of common items [12]. In the work by [13], the concept of singularity is defined and incorporated into the mean squared differences to compute similarity, so that two similar ratings both with higher singularities contribute to higher similarity. In [14], the prediction time of CF systems is improved by combining the numerical relevance of the ratings with non-numerical information based on the votes structure.

The information entropy is another approach to deduce context information from user ratings [4]. In [15], the rating of an unrated item is complemented by the entropy. Kwon et al. combined the entropy difference of user ratings with the previous similarity measures [6]. Wang et al. also estimated the entropy of the rating differences to incorporate it with Pearson correlation through the weighted average [16].

3. Proposed Similarity Estimation

3.1 User Rating Behavior and its Fuzzification

The purpose of this study is to overcome two major problems present in current memory-based similarity measures. Firstly, these measures heavily rely on items co-rated by two users to compute similarity. Secondly, they make use of exact user ratings, which often fail to reflect imprecise or vague rating behavior of users.

Our approach to the first problem is to make use of all of the user ratings as well as those ratings for common items when computing similarity. By using this additional information, the resulting similarity can be more trustworthy. To elicit information of user rating behavior from all the ratings of a user, we represent user rating pattern with two factors, the probability of each numeric rating and the entropy of user ratings.

Table 1 illustrates a user rating pattern. Assuming that the system allows the integer rating scale from one to ten, this table shows the ratio of each rating of this user and the information entropy of the user ratings. Consider two users who assign the same rating to each common item. The cosine similarity yields one in this case, which indicates that the two users are perfectly similar, regardless of the number of common items. This is obviously absurd for very few common items. However, by considering the rating patterns of two users in computing similarity, similarity value of one may not be obtained, thus ensuring higher reliability of similarity.

Let us formally define the user rating pattern. The probability of each user rating is estimated from all the ratings given by the user. Specifically, let I be the set of all items in the system, $r_{u,i}$ the rating given by user u to item i, and $[r_{mn}, r_{mx}]$ the rating range allowed in the system. Then probability of rating r by user u is defined as

$$P_{u,r} = \frac{\left|\{i \in \mathrm{I} \;\mid\; r_{u,i} = r\}\right|}{\left|\{i \in \mathrm{I} \;\mid\; r_{u,i} \in [r_{mn},\ r_{mx}]\}\right|}$$

The information entropy can be used to deduce useful context information from user ratings. As in [6][16], we define the entropy of user ratings (Hu) as follows and normalize it using the sigmoid function (H'_u).

$$H_u = -\sum_{r=r_{mn}}^{r_{mx}} P_{u,r} log_2(P_{u,r}), \qquad H'_u = \frac{1}{1+e^{-H_u}}$$

In order to reflect the essential uncertainty of user ratings, we transform the user rating pattern into a fuzzy one, so that similarity between two users is computed using fuzzy membership values of their patterns. In addition, we also fuzzify user ratings to compute similarity.

3.2 Proposed Similarity Measure

We propose a new similarity measure which is developed on the basis of the mean squared differences (MSD). Basically the proposed measure considers differences of all the fuzzy membership values of the user rating pattern and those of user ratings between two users. Let $\hat{r}_{u,i}$ be a normalized rating of $r_{u,i}$ into [0, 1] and Iu,v the set of items co-rated by users u and v. Then MSD between users u and v is defined as follows [1].

Table 1. Illustration of a user rating pattern

rating	1	2	3	4	5	6	7	8	9	10
ratio	0	0.1	0.15	0.25	0.2	0	0.15	0.05	0.1	0
entropy						2.3463				

$$MSD(u,v) = 1 - \frac{1}{|I_{u,v}|}\sum_{i\in I_{u,v}} (\hat{r}_{u,i} - \hat{r}_{v,i})^2$$

Similarity with respect to fuzzy user rating patterns is formulated as follows. Here, we assume a discrete rating within the range $[r_{mn},\ r_{mx}]$ of the rating scale. For a non-integer rating *r*, it can be easily transformed into the nearest integer, floor(r+0.5). Let *k* be the number of membership values of the fuzzy function with the *j*th membership function m^j. Then the probability of rating *r* given by user *u*, $P_{u,r}$, is represented with $m^j_{u,r}$'s and the normalized user entropy H'_u with $m^j_{u,h}$'s. We denote similarity of the probability of each rating by $MSD_P(u,v)$ and that of the normalized user entropy by $MSD_H(u,v)$. These definitions are provided below.

$$MSD_P(u,v) = 1 - \frac{1}{k(r_{mx} - r_{mn} + 1)}\sum_{r=r_{mn}}^{r_{mx}}\sum_{j=1}^{k}(m^j_{u,r} - m^j_{v,r})^2$$

$$MSD_H(u,v) = 1 - \frac{1}{k}\sum_{j=1}^{k}(m^j_{u,h} - m^j_{v,h})^2.$$

Likewise, a rating $r_{u,i}$ is transfomed with fuzzy values $m^j_{u,i}$'s. Then similarity with respect to ratings of items common between two users is

$$MSD_r(u,v) = 1 - \frac{1}{k|I_{u,v}|}\sum_{i\in I_{u,v}}\sum_{j=1}^{k}\left(m^j_{u,i} - m^j_{v,i}\right)^2$$

Finally we propose a similarity measure denoted as $PROP_{FEQ}$ that combines all of the three types of similarities defined above together with equal weight. Specifically,

$$PROP_{FEQ}(u,v) = \frac{1}{3}(MSD_P(u,v) + MSD_H(u,v) + MSD_r(u,v))$$

4. Performance Evaluation

4.1 Design of Experiments

Among publicly available datasets in the related field, we choose one with widely varying rating patterns among users to cautiously examine performance behavior of similarity measures. Jester dataset suits such purpose with a large range of rating scale from -10.0 to +10.0. This dataset contains 70,675 number of ratings by 1000 users for 100 items and the sparsity level is 0.2936. We divide the whole dataset into 80% of training set and the rest for testing.

Performance of a CF-based recommender system is typically measured by how much a target user prefers the items recommended by the system. In order to recommend an item, the system needs to first select users similar to the target user and calculate preferences of items yet unseen by the target user from similar users' rating records. Then items of highest preferences, i.e., highest predicted ratings, are to be recommended.

Therefore, quality of the CF systems is basically determined by similar users, i.e., by the similarity measure. We use representative metrics to evaluate prediction, recommendation, and ranking accuracies [4]. Typical metrics for these are adopted, which include MAE (mean absolute error), F1, and nDCG (normalized discounted cumulative gain). F1 is a harmonic mean of precision and recall.

Similarity measures are selected for experiments as follows. First, the mean squared differences (MSD) are selected as baseline. As other baselines, we experimented with the method reported in [7] (JMSD), the measure using entropy reported in [6] (KWON), and the measure using singularities in [13] (SING). In addition, we investigated MSD with fuzzy ratings, i.e., $MSD_r(u,v)$, denoted by MSD_FZ. A fuzzy function with five membership functions each of a trapezoid type is used for experimentation.

4.2 Performance Results

Fig. 1 presents MAE of the similarity measures. In general, MAE tends to decrease to a stable condition as more similar users are referenced. There is found a big difference among the measures mainly due to KWON and SING. KWON takes a similar strategy as ours in that it combines entropy and MSD, which, however, leads to the worst performance. The outcome of SING also seems rather unexpected, although it incorporates a distinguishing factor, i.e., singularity, into MSD. We used the value of zero as a singularity bound.

Beside the two measures, relatively small differences are observed in MAE among the rest measures in the figure. However, incorporating Jaccard index into MSD appears undesirable, as shown in the results of JMSD. JMSD performs still worse than the original MSD. When comparing MSD and MSD_FZ results, it is observed that just by replacing exact ratings by fuzzy ratings, unignorable performance improvement is obtained. This proves the existence of imprecise rating behavior of users. The proposed measure PROP_FEW, yields the best accuracy, although MSD_FZ is very competitive to ours.

Rank accuracy of the measures is also depicted in the figure. Similar to the MAE results, KWON performs worst, followed by SING. In particular, these two measures are far outperformed by the others. Note that MSD is surprisingly one of the best along with PROP_FEW and MSD_FZ, a different outcome from the MAE results. The reason for this seems that nDCG only concerns the ranking order of the recommendation list, whereas MAE computes the precise rating differences. Therefore, MSD is proved quite suitable for recommending items to users when the system only presents their order.

Regarding recommendation quality, different from the MAE and nDCG results, only SING yields significantly low performance. The reason might be that item ratings are not that much singular in this dataset, thus the strategy of SING of utilizing singularity information for measuring similarity contributes negatively and instead degrades the original MSD performance. It is also noted that MSD leads to relatively poor outcomes and is worse than KWON which has consistently shown the lowest performance throughout the experiments in terms of other performance metrics.

Although MSD_FZ demonstrated unexpectedly good performance in terms of MAE and nDCG, its F1 results are not that satisfactory as observed in the figure. On the other hand, PROP_FEW outperforms all the others as in the MAE results. Consequently, the proposed approach of considering the distribution of ratings and fuzzifying ratings to compute similarity demonstrates its superiority.

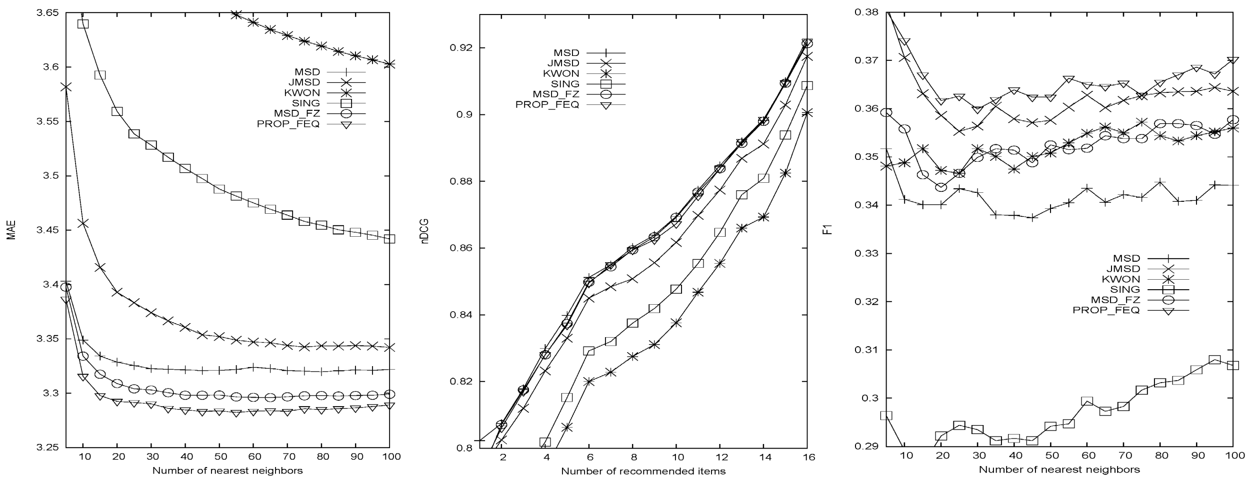

Figure 1. Performance of similarity measures in terms of MAE, nDCG, and F1

5. Conclusions

Similarity computation is very critical to performance of collaborative filtering systems, because it determines the set of similar users and the recommendation list. This paper addresses the serious problem of data sparsity inherent in the systems by proposing a new similarity measure. Outstanding features of the proposed measure is that it considers the whole user rating behavior to compute similarity between users, not just the set of co-rated items as in most previous measures. In addition, it reflects the uncertainty or vagueness of user ratings by transforming them into fuzzy ratings with which similarity computation is made. Through extensive experiments, the proposed measure demonstrates superior performance in terms of various metrics, whereas no single measure in previous related studies exhibits such outcome.

References

[1] M. Jalili, S. Ahmadian, M. Izadi, P. Moradi, and M. Salehi, “Evaluating Collaborative Filtering Recommender Algorithms: A Survey,” IEEE Access, Vol. 6, pp. 74003-74024, 2018.
DOI: https://doi.org/10.1109/ACCESS.2018.2883742.

[2] S. Bag, A. Ghadge, and M.K. Tiwari, “An Integrated Recommender System for Improved Accuracy and Aggregate Diversity”, Computers and Industrial Engineering, Vol. 130, pp. 187-197, 2019.
DOI: https://doi.org/10.1016/j.cie.2019.02.028.

[3] Y. Wu, X. Zhang, H. Yu, S. Wei, and W. Guo, “Collaborative Filtering Recommendation Algorithm based on User Fuzzy Similarity,” Intelligent Data Analysis, Vol. 21, pp. 311–327, 2017. DOI: https://doi.org/10.3233/IDA-150438.

[4] K.G. Saranya, G.S. Sadasivam, and M. Chandralekha, “Performance Comparison of Different Similarity Measures for Collaborative Filtering Technique,” Indian Journal of Science and Technology, Vol. 9, No. 29, pp. 1–8, Aug. 2016. DOI: https://10.17485/ijst/2016/v9i29/91060.

[5] J.L. Herlocker, J.A. Konstan, A. Borchers, and J. Riedl, “An Algorithmic Framework for Performing Collaborative Filtering,” The Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 230–237, 1999. DOI: http://doi.org/10.1145/312624.312682.

[6] H.-J. Kwon, T.-H. Lee, J.-H. Kim, and K.-S. Hong, “Improving Prediction Accuracy using Entropy Weighting in Collaborative Filtering,” In 2009 Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing, pp. 40-45, 2009. DOI: https://doi.org/10.1109/UIC-ATC.2009.50.

[7] J. Bobadilla, F. Serradilla, and J. Bernal, “A New Collaborative Filtering Metric that Improves the Behavior of Recommender Systems,” Knowledge-Based Systems, Vol. 23, No. 6, pp. 520–528, 2010. DOI:

https://doi.org/10.1016/j.knosys.2010.03.009.
[8] S. Lee, "Improving Jaccard Index using Genetic Algorithms for Collaborative Filtering," Lecture Notes in Computer Science, Vol. 10385, pp. 378-385, 2017. DOI: https://doi.org/10.1007/978-3-319-61824-1_41.
[9] L.H. Son, "HU-FCF: A Hybrid User-based Fuzzy Collaborative Filtering Method in Recommender Systems," Expert Systems with Applications, Vol. 41, pp. 6861–6870, 2014. DOI: https://doi.org/10.1016/j.eswa.2014.05.001.
[10] M.Y.H. Al-Shamri and N.H. Al-Ashwal, "Fuzzy-weighted Similarity Measures for Memory-based Collaborative Recommender Systems," Journal of Intelligent Learning Systems and Applications, Vol. 6, pp. 1–10, 2014. DOI: https://doi.org/10.4236/jilsa.2014.61001.
[11] E.S.G. Herrera-Viedma, J.A. Olivas, A. Cerezo, and F.P. Romero, "A Google Wave-based Fuzzy Recommender System to Disseminate Information in University Digital Libraries 2.0," Information Sciences, Vol. 181, No. 9, pp. 1503–1516, 2011. DOI: https://doi.org/10.1016/j.ins.2011.01.012.
[12] C. Desrosiers and G. Karypis, "A Novel Approach to Compute Similarities and Its Application to Item Recommendation," In PRICAI 2010: Trends in Artificial Intelligence, Lecture Notes in Computer Science, Vol. 6230/2010, pp. 39–51, 2010. DOI: https://doi.org/10.1007/978-3-642-15246-7_7.
[13] J. Bobadilla, F. Ortega, and A. Hernando, "A Collaborative Filtering Similarity Measure based on Singularities," Information Processing and Management, Vol. 48, No. 2, pp. 204-217, 2012. DOI: https://doi.org/10.1016/j.ipm.2011.03.007.
[14] B. Zhu, R. Hurtado, J. Bobadilla, and F. Ortega, "An Efficient Recommender System Method based on the Numerical Relevances and the Non-numerical Structures of the Ratings," IEEE Access, Vol. 6, pp. 49935-49954, 2018. DOI: https://doi.org/10.1109/ACCESS.2018.2868464.
[15] C.H. Piao, J. Zhao, and L.J. Zheng, "Research on Entropy-based Collaborative Filtering Algorithm and Personalized Recommendation in Ecommerce," Service Oriented Computing and Applications, Vol. 3, No. 2, pp. 147–157, 2009. DOI: https://doi.org/10.1007/s11761-008-0034-3.
[16] W. Wang, G. Zhang, and J. Lu, "Collaborative Filtering with Entropy Driven User Similarity in Recommender Systems," International Journal of Intelligent Systems, Vol. 30, No. 8, pp. 854–870, Aug. 2015. DOI: https://doi.org/10.1002/int.21735.