Enhancing Recommendation Quality of Content-based Filtering through Collaborative Predictions and Fuzzy Similarity Measures

Vibhor Kanta,* Kamal K. Bharadwajb

*Research Scholar, School of Computer & Systems Sciences, Jawaharlal Nehru University, New Delhi-110067, India
bProfessor, School of Computer & Systems Sciences, Jawaharlal Nehru University, New Delhi-110067, India

Abstract

Recommender systems (RSs) provide personalized suggestions about items to users while interacting with the large spaces on the web. Content based recommender systems (CB-RSs) offer personalized recommendations to a user mainly based on his past history and representations of the items. Although CB-RSs have been applied successfully in various domains, however recommendation diversity, representation of items as well as users’ modeling are still major concerns. Our work in this paper is an attempt towards developing effective content based filtering (CBF) by introducing an item representation scheme, fuzzy similarity measures and incorporating collaborative diverse predictions for alleviating its recommendation diversity. Experimental results show that the proposed hybrid scheme Fuzzy-CF-CBF outperforms hybrid CF-CBF, as well as both the fuzzy collaborative filtering (Fuzzy-CF) and the fuzzy content based filtering (Fuzzy-CBF).

Keywords: Recommender system; Collaborative and content based filtering; Recommendation diversity; Fuzzy similarity measure

1. Introduction

It is getting more difficult to generate automatic appropriate recommendations to a user related to his/her preferences; not only because of the unprecedented proliferation of textual contents, such as online news, research papers, blog articles and other things like movies, books, restaurants etc. on the web but also because of the difficulty of automatically grasping his/her interests. Recommender Systems (RSs)
have emerged as the most essential tool to deliver personalized recommendations for users in response to the above challenges [1]. There are two widely used approaches among recommender systems, collaborative filtering (CF) and content based filtering (CBF). The traditional task in the former one is to predict the utility of a particular item for the active user from the opinions of other similar users, and thereby make appropriate recommendations; on the other hand, later approach provides recommendations by comparing representations of content contained in an item to those of a user's interest ignoring opinions of other similar users [1, 2].

CBF systems analyze item descriptions to identify items that are of particular interest to the user. Features of items and users’ behaviour are highly subjective in nature [3]. Therefore, the representation of items as well as user modeling are major problems for constructing CB-RSs. Yager [11] presents reclusive methods for RSs with justifications and rules of the recommendations based on fuzzy set. A lot of work has been done in the field of RSs [3, 11, 12] based on fuzzy logic. In addition to above issues, we argue that CBF is flawed in some application domains. For example, consider a digital camera recommender: a user submits his preferences about a digital camera less than £ 200, with security digital card, 16.2 MP and full HD video recording capabilities. The top recommendation returned is for a Sony Cyber-Shot DSC containing above features. A good recommendation perhaps, but what if second, third and forth recommendations are all slight modifications of the same model? If the user decides to avoid Sony, then none of CB-RSs will suffice. Hence diversity problem is also recognized a shortcoming of CBF systems [4]. A common solution is to combine CF technique with CBF systems so that the varieties of items can be recommended by CBF systems. The main contributions of the proposed work are three fold:

- First of all, a fuzzy collaborative filtering (Fuzzy-CF) using GIM feature is designed by utilizing the local and global fuzzy distances.
- Second, a fuzzy CBF (Fuzzy-CBF) is presented by introducing a fuzzy item representation scheme and similarity measures.
- Finally, a hybrid Fuzzy-CF-CBF is developed utilizing the set of diverse items predicted by fuzzy-CF into the fuzzy-CBF.

The rest of the paper is structured as follows: Section 2 provides related works on RSs. In Section 3, the proposed work is presented. Computational experiments and results are given in Section 4. Finally, in the last section we conclude our work with some future research directions.

2. Background

The most widely used techniques in RSs are CF and CBF. CF relies on the fact that user preferences are stable and is based on the premise of “people who have agreed in the past tend to agree in future”. The great power of CF relative to CBF is its ‘outside the box’ recommendation ability [5, 6]. While in CBF, an item is recommended to a user mainly based on the characteristics of the item and the users’ past actions like purchases, queries, and ratings. CBF system is highly overspecialized to recommended items whereas CF system has new item and sparsity problem [1, 2, 3]. Therefore many researchers [5, 7, 8, 9, 10] have chosen different ways to combine CF technique with CBF techniques for alleviating their individual weaknesses. The Fab [9] system combines collaborative and content-based filtering in its recommendations by measuring similarity between users after computing a profile for each user. In Pazzani’s [10] approach predictions are made by applying CF directly to the matrix of user-profiles.

3. Proposed recommendation framework

In this section, we describe Fuzzy-CF, diverse algorithm, Fuzzy-CBF and proposed scheme.
3.1 Fuzzy collaborative filtering (Fuzzy-CF)

Sometimes, crisp description of features does not reflect the actual case of human decisions. For example, two users of age 18 and 20 have the age difference of 2 years, while both users belong to the same age group, i.e., young. Therefore, a great advantage can be gained by fuzzifying the features of the users in order to retrieve most similar users. Different fuzzy sets are used to fuzzify the various features of a user. First of all, age is fuzzified into three fuzzy sets, young, middle-aged, and old as shown in Fig. 1(a). Furthermore, the genre-interestingness measure (GIM) [5] is fuzzified into six fuzzy sets, very bad (VB), bad (B), average (AV), good (G), very good (VG), and excellent (E). Fig. 1(b) shows the membership value of GIM in different fuzzy sets.

Fig. 1. (a) Membership function for age (b) Membership function for genre interestingness measure (GIM)

The similarity measurement between users is an important task for CF systems to retrieve the best users. Since features of users are fuzzified into different fuzzy sets. Therefore, to compute the similarity between two users \( a \) and \( b \), the local fuzzy distance (LFD) between each feature \( i \) is computed as

\[
LFD(a_i, b_i) = \frac{\sum_{j=1}^{k} |a_{i,j} - b_{i,j}|}{k}
\]

where \( a_{i,j} \) is the membership value of the feature \( i \) in its \( j \)th fuzzy set for a user \( a \) and \( k \) is the number of fuzzy sets. Consequently, to obtain the global fuzzy distance (GFD) between users we employ an aggregation operator. Therefore the GFD between two users \( a \) and \( b \), can be defined as

\[
GFD(a, b) = \frac{\sum_{i=1}^{l} LFD(a_i, b_i)}{l}
\]

where \( l \) is the number of features of user \( a \) and \( b \). Using the GFD, the similarity between users is computed as

\[
Sim(a, b) = 1 - GFD(a, b)
\]

Now CF system generates the neighbourhood set \( N \) of similar users for an active user \( a \) to predict the rating \( p_{a,s} \) on item \( s \) using Resnick formula

\[
p_{a,s} = \bar{a} + \frac{\sum_{u \in N} Sim(a, u) \cdot (r_{u,s} - \bar{u})}{\sum_{u \in N} Sim(a, u)}
\]

where \( \bar{a} \) denotes the average rating of user \( a \) and \( N \) denotes the set of similar users.

3.2 Bounded diverse algorithm

Once top \( bk \) items are recommended to an active user by using proposed Fuzzy-CF method, bounded diverse item selection algorithms selects \( k \) most dissimilar items from predicted \( bk \) items. The rationale behind this approach is improving the recommendation diversity among the predicted items.
3.3 Fuzzy content-based filtering (Fuzzy-CBF)

We proposed fuzzy-CBF framework in which fuzzy modelling is used for the representation of items.

3.3.1 Item representation

An item can be represented in the terms of their features. In MovieLens dataset, movies are represented in the terms of genres. If a genre is present in a movie then it is represented by 1 otherwise 0. But it is not a real scenario because a movie cannot contain equal amounts of different genres. Let $S = \{s_1, s_2, \ldots, s_n\}$ be the set of items and $f$ be a feature of an item which can take the multiple values from a set $A = \{f_1, f_2, \ldots, f_m\}$. The Gaussian membership function of an item $s_i$ to value $f_j$ is described as follows

$$
\mu_{f_j}(s_i) = \frac{r_k}{2^{\rho \gamma(r_k - 1)}}
$$

(5)

where $\gamma$ the number of values of $A$ that is is associated to an item $s_i$, $r_k$ is the rank position of $f_j$ and $\rho > 1$.

3.3.2 Similarity measures between items

Let items $s_i$ and $s_p$ are defined as $\{(f_j, \mu_{f_j}(s_i)) | j = 1,2, \ldots, m\}$ and $\{(f_j, \mu_{f_j}(s_p)) | j = 1,2, \ldots, m\}$ respectively. The similarity between these items as computed as

$$
Sim(s_i, s_p) = \left( 1 - \frac{\sum_{j=1}^{m} |\mu_{f_j}(s_i) - \mu_{f_j}(s_p)|}{m} \right)
$$

(6)

In order to predict the rating of an unseen item $s$ to an active user $a$ by employing this approach, the following formula is used

$$
P_{a,s} = \frac{\sum_{s \in L} Sim(s,s) \times r_{a,s}}{\sum_{s \in L} Sim(s,s)}
$$

(7)

where $L$ is the set of liked items by user $a$.

3.4 Proposed hybrid approach (Fuzzy-CF-CBF)

The main steps involved in our proposed hybrid approach Fuzzy-CF-CBF are as follows:

**Step1:** Obtain top $bk$ recommendations with their predicted ratings using Fuzzy-CF method by utilizing the hybrid features of GIM and age.
Step2: Select top k diverse items among predicted bk items by bounded diverse item selection algorithm.

Step3: Generate quality recommendations utilizing the set of diverse items predicted by Fuzzy-CF into Fuzzy-CBF.

4. Experimental results

To demonstrate the effectiveness of our proposed work, we conducted experiments on publicly available MovieLens (ML) dataset.

4.1 Design of experiments

The dataset contains movie attributes, user ratings and user demographic features. The dataset consists of 100,000 ratings provided by 943 users on 1682 movies in the range 1-5. Each user has rated at least 20 and at most 737 movies. For our experiments, we select three subsets from the data, containing 50, 100 and 150 users called ML50, ML100 and ML150 respectively. This is to demonstrate the effectiveness of the proposed scheme under varying number of participating users. Each of the selected subsets was randomly split into 60% training data and 40% test data. We selected 25% items from test set for the value of k. In order to test the performance of our scheme, we measure the predictive accuracy using Mean Absolute Error (MAE) and coverage.

4.2 Results

To illustrate the ability of the proposed scheme by offering better recommendation accuracy, we compare the MAE and coverage with Fuzzy-CBF, Fuzzy-CF and CF-CBF. The results are presented in Table 1. The MAE and coverage are computed based on the average over 20 runs of the experiments over different datasets. A lower value of MAE and higher value of coverage imply the better performance of the proposed scheme. It is clear from Table 1, the proposed scheme considerably performed better than any of the other techniques in the terms of predictive accuracy as well as coverage. The MAE for the different 20 runs of the experiment for ML100 is shown in Fig. 2.

Table 1. Comparison of MAE and Coverage of proposed Fuzzy-CF-CBF with Fuzzy-CBF, Fuzzy-CF and CF-CBF

| Dataset | Algorithms          | MAE  | Coverage (%) | MAE  | Coverage (%) | MAE  | Coverage (%) | MAE  | Coverage (%) |
|---------|---------------------|------|--------------|------|--------------|------|--------------|------|--------------|
|         | Fuzzy-CBF           | .8968| 94.50        | .8915| 87.60        | 1.7240| 98.40        | .8791| 98.40        |
|         | Fuzzy-CF            | .8689| 97.40        | .8620| 86.40        | 1.7605| 97.40        | .8469| 98.50        |
|         | CF-CBF              | .8830| 97.60        | .8744| 86.50        | 1.6800| 97.60        | .8630| 98.20        |

Fig.2 MAE for ML100 over 20 runs
5. Conclusion and future work

The proposed work deals with the use of fuzzy modeling for the representation of items and user behavior model and also addresses the recommendation diversity problem of CB-RSs. Incorporation of collaborative predictions into fuzzy CB-RSs has resulted in alleviating the recommendation diversity problem, thereby producing quality recommendations. To evaluate the effectiveness of our hybrid scheme Fuzzy-CF-CBF, we conducted an experimental study comparing the proposed approach with the fuzzy-CBF, fuzzy-CF and hybrid CF-CBF. Our results indicate that proposed scheme consistently outperforms all the three approaches.

In our future work, we plan to integrate the current approach by including other features of items for further enhancing the accuracy of content based filtering. The current framework is specific to movie RSs and it would be interesting to explore the feasibility of extending the framework to other domains e.g. jokes, books, music etc.

References

[1] Adomavicius, G., Tuzhilin, A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans. Knowledge Data Eng. 17 (6); 2005, p. 734-749.
[2] Burke, R. Hybrid recommender systems: survey and experiments. User Model. User-Adapted Interaction 12 (4); 2002, p. 331-370.
[3] Zenebe, A., Norcio, A.F. Representation, similarity measures and aggregation methods using fuzzy sets for content-based recommender systems. Fuzzy Sets and Systems 160 (1); 2009, p. 76-94.
[4] Smyth, B., McClave P. Similarity vs. diversity. In Proceedings of the Fourth International Conference on Case-Based Reasoning. ICCBR, LNAI 2080, Springer-Verlag Berlin Heidelberg, 2001, p. 347-361.
[5] Al-Shamri, M.Y.H., Bharadwaj, K.K. Fuzzy-genetic approach to recommender system based on a novel hybrid user model. Expert Systems with Applications 35(3); 2008, p. 1386-1399.
[6] Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T. Evaluating collaborative filtering recommender systems. ACM Trans. Inform. Syst. 22 (1); 2004, p. 5-53.
[7] Basu, C., Hirsh, H., Cohen, H. Recommendation as classification: using social and content-based information in recommendation. In Proceedings of the 15th Nat. Conf. on Artificial Intelligence; 1998, p. 714-720.
[8] Antonopoulus, N., Salter, J. Cinemascreen recommender agent: combining collaborative and content-based filtering. IEEE Intelligent Systems; 2006, p. 35-41.
[9] Balabanovic, M., Shoham, Y. Fab: content-based, collaborative recommendation. Communications of the ACM 40 (3); 2007, p. 66-72.
[10] Pazzani, M. J. Framework for collaborative, content-based and demographic filtering. Artificial Intelligence Review 13(5-6); 1999, p. 393-408.
[11] Yager, R.R. Fuzzy logic methods in recommender systems. Fuzzy Sets and Systems 136; 2003, p. 133-149.
[12] Kant, V., Bharadwaj, K.K. Incorporating fuzzy trust in collaborative filtering based recommender systems. In Proceedings of the 2nd international conference of SEMCCO, Part I, LNCS 7076, Springer-Verlag Berlin Heidelberg 2011, p. 433-440.