Improvement of Collaborative Filtering Based on Fuzzy Reasoning Model

Ryosuke FUJIOKA* and Toshihiko WATANABE**

* System Development Section, Kobe Sogo Sokki, Ltd.
** Faculty of Engineering, Osaka Electro-Communication University

Abstract: Though various contents are provided through the internet recently, it is not easy to find favorite contents among huge amounts of contents in terms of user’s preference. In this paper, we focus on the collaborative filtering algorithm in the recommender system. We propose a fuzzy modeling approach for preference similarity model in collaborative filtering. In our approach, valid simplified fuzzy reasoning model is constructed through optimization of MAE (Mean Absolute Error). The model decides the weight of preference similarity from the value of correlation coefficient and the number of items. Through numerical experiments compared with conventional correlation coefficient based approach using Movie Lens data, the approach is found to be promising for improvement of collaborative filtering model accuracy.

Keywords: Collaborative filtering, Fuzzy reasoning model, Recommender system, Optimization, Preference similarity.

1. Introduction

Recently, various contents are provided through the internet. Especially the contents of e-Commerce such as music, movies, and books are indispensable for modern life style. However, it is not easy to find favorite contents among huge amounts of contents in terms of user’s preference. An effective approach for suppliers and customers of the contents to solve such a problem is to develop “Recommender System.” The “Recommender System”, the system of Amazon site[14] is well known, selects and recommends the contents item to meet user’s preference automatically using various data stored in database. The important essential technique in the Recommender System is information filtering. Among the various types of information filtering that have been proposed, the techniques fall into two categories: content-based filtering[1] and collaborative filtering[2].

In content-based filtering, a user preference model is constructed for the individual based on the descriptions of items(contents) as well as user’s ratings. On the other hand, collaborative filtering tries to find desirable items based only on the preference of a set of similar users corresponding to the user’s ratings. Since it is not necessary to analyze the contents of items, collaborative filtering can be applied to many kinds of domains. Furthermore, the software structure of collaborative filtering is comparatively simple for application to real-world problems. However increase of the size of database for collaborative filtering leads to increase of the computation time[3]. In order to reduce the computation time, applications of clustering are proposed[4,5]. On the other hand, improvement of collaborative filtering based on prediction model such as the simple Bayesian classifier[6] and predictive algorithm[7] are proposed.

In this paper, we focus on performance improvement of the collaborative filtering algorithm in the recommender system. We propose a collaborative filtering based on preference similarity model. The model is expressed by linear regression form or simplified fuzzy reasoning model. It is constructed through optimization of the objective function based on prediction error. We also discuss about the per-
formance of proposed method through numerical experiments using Movie Lens data[8].

The paper is organized as follows. We summarize the collaborative filtering concepts and technique in section 2. In section 3, the basic concepts of collaborative filtering based on preference similarity model are proposed. In section 4, constructions of preference similarity model by linear regression form are introduced for comparative study. In section 5, preference similarity model by fuzzy reasoning model is proposed. Results of numerical experiments are shown in section 6. Finally, conclusions are drawn in section 7.

2. Collaborative Filtering

The basic idea of collaborative filtering is to recommend new items of interest for a particular user based on other users’ opinion(evaluation). The algorithm is based on a simple intuitive sense: predictions for a user should be based on the preference patterns of other people who have similar interests. The conceptual diagram of collaborative filtering is shown in Fig.1. An example of the rating table(matrix) is shown in Table.1. Where rows correspond to users, columns correspond to items, and the entries are ratings. Typically, the rating matrix is sparse matrix. The last row in Table.1 represents the ratings of a user for whom the system will make predictions. In the example, the rating of item F of Ken is estimated by collaborative filtering algorithm.

The first step of the algorithm is to estimate similarities between user ratings. Let us assume that users’ ratings are collected and stored in a database. In conventional approach, the Pearson correlation coefficient is used as a measure of preference similarity. The correlation between users t and u is calculated as:

$$r_{tu} = \frac{\sum_j (x_{ij} - \bar{x}_t)(x_{uj} - \bar{x}_u)}{\sqrt{\sum_j (x_{ij} - \bar{x}_t)^2 \sum_j (x_{uj} - \bar{x}_u)^2}}$$  \hspace{2cm} (1)

where $x_{ij}$ is the rating of user t for item j, $\bar{x}_t$ is the mean rating value of user t, and all summations over j are over the items that have been rated by both user t and user u. The predicted rating of user i for item j is computed as a weighted sum of the other users’ ratings:

$$\hat{x}_{ij} = \bar{x}_i + \frac{\sum_{u=1}^{n} (x_{uj} - \bar{x}_u) \times r_{iu}}{\sum_{u=1}^{n} |r_{iu}|}$$  \hspace{2cm} (2)

where n is the number of users in the database. This correlation-based prediction manner has been shown to have valid performance in terms of computation time and prediction accuracy.

The collaborative filtering is based on the assumption that there exist a significant quantity of rating data. However the assumption is not always held in actual applications. In order to overcome this problem, known as the “cold-start problem”[17], and to reduce the requirements of information, knowledge based system approaches are also proposed[18] based on expertise of system application. In this study, we assume that the assumption is held good, in order to deal with improvement of essential performance of collaborative filtering.

![Fig.1. Conceptual Diagram of Collaborative Filtering](image)

![Table.1. Example of Rating Table](image)
3. Collaborative Filtering Based on Preference Similarity Model

As the progressive approaches of the basic algorithm, combination of considerable metrics adopted and their improvement methods of processing are also studied[9,10]. On the other hand, clustering based approaches have been studied[4,5,15,16], in order to improve the performance of collaborative filtering especially in terms of computational efficiency. Furthermore machine learning approaches are applied considering the collaborative filtering problem to be the classification problem[6,11]. Though the approaches are effective and significant, it can be said that it is inadequate in the viewpoint of formulation of precise user preference model, i.e. basic concept of the collaborative filtering. Our basic motivation of this study is to construct simple and precise user preference model for collaborative filtering. In other words, we deal with the collaborative filtering problem as “modeling problem” instead of “classification problem” or “clustering problem.”

The conventional filtering algorithm with correlation coefficient as similarity measure described above is useful because of its simple structure and valid performance. However, the preference similarity measure using the only correlation coefficient value might be inadequate in order to improve the recommendation performance. Furthermore it is well known that the correlation coefficient is not appropriate when it has nonlinear property or there exist only a few data pair. Our basic idea is to construct preference similarity model for collaborative filtering. In this study, we focus on the collaborative filtering that use only rating matrix, in order to tackle to essential improvement of the prediction performance in collaborative filtering. From the reason, the inputs of the model are restricted to the correlation coefficient value and the number of the both evaluated items which is used in calculation of correlation coefficient as Eq.(1). The filtering is computed as follows:

\[ \hat{x}_{ij} = \frac{1}{n} \sum_{u=1}^{n} \left( x_{uj} - \bar{x}_u \right) \times W_{iu} \]

\[ W_{iu} = f(r_{iu}, n_{iu}) \]

where \( f \) denotes the preference similarity model, \( W \) denotes the output of the preference similarity model, and \( n_{iu} \) is the number of items that have been rated by both user \( i \) and user \( u \). Since our research objective is to construct simple and good similarity model in order to improve recommendation performance maintaining the simple structure of the filtering technique, primitive important issue is to decide the structure of the model \( f \).

4. Preference Similarity Model by Linear Regression Model

In system identification techniques, it is also a difficult and essential issue to decide appropriate model structure that is not clearly defined. The most relevant strategy is to assume the linear structure that is simple, robust, and easy to handle. First, as the basic model structure in Eq.(4), the following three types of linear regression form are considered in this study:

\[ f(r,n) = qr + bn \]

\[ f(r,n) = qr + bn + crn \]

\[ f(r,n) = qr + bn + crdn^e \]

where \( q, b, c, d, \) and \( e \) are parameters. The parameters are identified by optimization algorithm using collected data as described in the following section.

5. Preference Similarity Model by Fuzzy Reasoning Model

5.1. Basic Idea

Preference similarity is not easy to represent as the model generally. It could be thought that it depends on the target area such as movie contents, music contents, and so on. Furthermore, the assumption of linear property of the model might be not valid. In order to deal with such nonlinear property of similarity, the simplified fuzzy reasoning model[12] is used to construct the preference similarity model.
5.2. Simplified Fuzzy Reasoning Model

Let \( x_i (i=1,2,\ldots,v) \) denote the input variable. The rule of the simplified fuzzy reasoning model is expressed as follows:

\[
R_j: \text{If } x_1 \text{ is } A^1_j, x_2 \text{ is } A^2_j, \ldots, x_v \text{ is } A^v_j \quad \text{Then } y = \omega_j \quad (j = 1, \ldots, m)
\]  

(8)

where \( A^i_j \) is the fuzzy set of the j-th rule defined in the i-th input variable and \( \omega \) is the “non-fuzzy” real number value. Let \( x^0_i \) denote the input value to the reasoning model. The membership degree \( h \) of the rule is calculated as:

\[
h_j = \prod_{i=1}^{v} \mu_{A^i_j}(x^0_i)
\]

(9)

where \( \mu_{A^i_j}(x^0_i) \) denotes the value of membership function of fuzzy set \( A \). Examples of the membership function are shown in Fig.2. In this paper, triangular type membership function is used for the model. The membership function is defined as:

\[
\mu_{A^i_j}(x^0_i) = \begin{cases} 
\frac{x^0_i - a^i_{k-1}}{a^i_k - a^i_{k-1}}; & \text{if } a^i_{k-1} \leq x^0_i \leq a^i_k \\
\frac{a^i_{k+1} - x^0_i}{a^i_{k+1} - a^i_k}; & \text{if } a^i_k \leq x^0_i \leq a^i_{k+1} \\
0; & \text{otherwise}
\end{cases}
\]

(10)

where \( a^i_k \) is the center of triangle in k-th fuzzy set. The output of the simplified fuzzy reasoning model is calculated as follows.

\[
W = \frac{\sum_{i=1}^{m} p_i \omega_i}{\sum_{i=1}^{m} p_i}
\]

(11)

In this study, the input of the fuzzy model is set as only the correlation coefficient value \( r \) or both the correlation coefficient value \( r \) and the number of evaluated items \( n \) in Eq.(4). The output \( W \) is used as the preference similarity in collaborative filtering. Figure 3 shows the flowchart of collaborative filtering based on the fuzzy reasoning model.

5.3. Parameter Identification Algorithm

We should decide above described parameters appropriately to construct precise preference similarity model. The necessary parameters are composed of real values \( \omega \) in the consequent part in Eq.(8), and the parameters of membership functions in the antecedent part. In this study, we assume that the number of fuzzy sets in the antecedent part is decided in advance. The center of membership function \( a \) in the antecedent part is tuned along with the real values in the consequent part by optimization procedure. The width of each membership function is decided spontaneously by difference of the center values of adjacent membership functions as formulated in Eq.(10).
The parameter identification is performed minimizing the objective function of MAE (Mean Absolute Error), i.e. prediction error of ratings, from collected data as:

\[
E(P^t) = \frac{1}{s} \sum_{j} \sum_{j} |x_{ij} - \hat{x}_{ij}(P^t)|
\]

where \( s \) is the number of predictions and \( P^t \) denotes the parameter vector (composed of the parameters of the fuzzy reasoning model) of \( t \)-th iteration.

Nonlinear optimization (search) method is applied in this study, as it is difficult to use the differential value of MAE. The flowchart of the parameter identification algorithm is shown in Fig.4. The process of the parameter adjustment with searching directions is employed as the following equations:

\[
P^{t+1} = P^t + \Delta P, \quad \Delta P = \left[ p_1, p_2, \ldots, p_z \right]^T
\]

\[
p_j = \begin{cases} 
\varepsilon_j & \text{if } E(P^t + \varepsilon_j d_j) - E(P^t) < 0 \\
-\varepsilon_j & \text{if } E(P^t - \varepsilon_j d_j) - E(P^t) < 0 \\
0 & \text{otherwise}
\end{cases}
\]

5.4. A Heuristic Model for Preference Similarity

As a model having middle degree of complexity compared with the above described linear model and the fuzzy model, the following heuristic model is empirically constructed:

\[
f(r, n) = n^e \cdot g(r)
\]

where \( g(r) \) denotes the simplified fuzzy reasoning model of input \( r \) and \( e \) is a parameter. The basic idea of the heuristic model is to represent the preference similarity as weighted formulation by the number of both rated items based on fuzzy reasoning model of correlation coefficient.

6. Numerical Experiments

In order to evaluate our proposed method, numerical experiments using Movie Lens data[8] are performed. The data is collected from 943 users about ratings to 1682 popular movies. The number of data is 100,000. The rating matrix is sparse as described in the previous section. As the measure of recommendation performance, MAE (Mean Absolute Error) as described in Eq.(12), ROC (Receiver Operating Characteristic)[13], and NROC are used in this experiments. ROC is the rate of proper recommendation by the filtering algorithm. NROC is defined as the rate of proper “no recommendation” by the algorithm. Low MAE value, high ROC value, and low NROC value denote good recommendation performance respectively. The measures of ROC
and NROC are directly corresponding to actual performance of the collaborative filtering model. In Movie Lens database, the users evaluated the movie using integer rating from one to five as SD (Semantic Difference) method. In the actual experiments, ROC is calculated as rate of the number of items that the collaborative filtering model decides to recommend positively coincident with the actual ratings. NROC is calculated as rate of the number of items that the collaborative filtering model decides to recommend positively though the actual value is negative value of recommendation.

In order to evaluate the predictive accuracy systematically, the experiments are performed through 5-fold Cross Validation.

Figure 6-8 show the results by linear regression forms. In the figures, “Conventional” denotes the conventional filtering with only the correlation coefficient, “Linear-linear” corresponds to Eq.(5), “Linear-bilinear” corresponds to Eq.(6), and “Linear-exponential” corresponds to Eq.(7). “Ave” in x-axis stands for the average value of each samples in Cross Validation process. From the results, the performance is found to be improved especially by using the model of Eq.(7).

Figures 9-11 show the results by fuzzy reasoning models. “Fuzzy Model(Correl)” denotes the fuzzy model of only correlation coefficient input used. “Fuzzy Model(Correl,Num)” denotes the fuzzy model using inputs of correlation coefficient and the number of both rating items. The accuracy of recommendation is improved by the preference similarity model based on the fuzzy reasoning models. The output examples of the constructed fuzzy model under the different data samples are shown in Fig.12. The measure of the preference similarity expressed as fuzzy model is formulated as high value in positive range of the correlation coefficient. In contrast, it is formulated as low value in negative range. This can be explained that it is desirable for collaborative filtering not to take the other user having opposite preference into consideration.

Figure 13-15 show the results by the heuristic model expressed by Eq.(14) compared with conventional filtering and the “Linear-exponential” model(Eq.(7)). The proposed heuristic model show good performance in terms of NROC compared with the “Linear-exponential” model. We fulfilled further experiments of different samplings and assured that the proposed method can improve the recommendation performance in collaborative filtering system.

From the results of the numerical experiments, the proposed approach is found to be promising for improvement of collaborative filtering model accuracy. However, a problem of computation time remains for applying the method to actual commercial systems. Though the amount of the calculation of preference similarity model proposed in this paper is relatively small compared with the calculation of the correlation coefficient, there exists trade-off problem of computation time and results accuracy. We consider that further development such as clustering based computation is needed for computational efficiency.
Fig. 8. Results by Linear Form Model (NROC)

Fig. 9. Results by Fuzzy Reasoning Model (MAE)

Fig. 10. Results by Fuzzy Reasoning Model (ROC)

Fig. 11. Results by Fuzzy Reasoning Model (NROC)

Fig. 12. Output from Fuzzy Reasoning Model

Fig. 13. Results by Heuristic Model (MAE)
7. Conclusion

In this paper, we proposed a collaborative filtering technique based on the preference similarity model. The model is expressed by linear regression form, the simplified fuzzy reasoning model, and heuristic form based on fuzzy reasoning model. It is constructed through optimization of the objective function based on prediction error(MAE). Through the numerical experiments using Movie Lens data, the proposed technique is found to be promising for improvement of collaborative filtering model accuracy. Our future plan includes application the method to actual sized problem of recommender system.

References

Knowledge-Based Systems, pp.417-419.


Ryosuke FUJIOKA
He received the M.E. degree in industrial engineering from Osaka Prefecture University in 1992. He joined Electronics Research Laboratory, Kobe Steel, Ltd in 1992. In 1994-1995 he was affiliated with Laboratory for International Fuzzy Engineering, in 1995 was back to Kobe Steel, Ltd. Production system Research Laboratory. He joined Kobe Sogo Sokki, Ltd in 1998, as Chief Information Officer. His current research interests include Soft Computing, Data Mining, and their application to manufacturing system and biomedical system. He is a member of Japan Institute of Fuzzy Theory and Intelligent Informatics.

Toshihiko WATANABE
He received the M.E. degree in industrial engineering from Osaka Prefecture University in 1990. He joined Electronics Research Laboratory, Kobe Steel, Ltd in 1990. He had been engaged in R&D for control system and manufacturing system especially in Steel plant and Mechanical plant. He received his D.E. from Osaka Prefecture University in 2000. He joined Osaka Electro-Communication University in 2001. He is an associate professor in Faculty of Engineering. His current research interests include Soft Computing, Data Mining, and their application to manufacturing system and biomedical system. He is a member of BMFSA.