A User Preference Classification Method in Information Recommendation System

Chentung Chen¹, Weishen Tai²

¹ Department of Information Management, Da-Yeh University, Taiwan, China
² Institute of Management, School of Management, Da-Yeh University, Taiwan, China
¹, ² {chtung, wstai}@mail.dyu.edu.tw

ABSTRACT

As information overload problem more serious on the Internet, it becomes an important issue for users to retrieve information effectively. An information recommendation system is helpful for providing user information meet he/she requirements appropriately. However the traditional recommendation concepts usual classify a user into one preference class. It seems unreasonable because a user may possess interests about many information classes generally. In this study we propose a new recommendation concept in information recommendation system, namely club member, differs from content-based and collaboration filter method. It can classify a user into some clubs which he/she interests with different preference degrees.

In order to classify users into multi-club with different preference degrees, fuzzy association rule based on data mining technology is applied in this study. Fuzzy association rules are generated by discovering and analyzing the members’ feature in the same club. According to fuzzy association rules, a new user can be classified into some clubs that he/she may interest with preference degrees. It is helpful for an information recommendation system to provide user more suitable information in accordance with their preferences precisely.

Keywords: Preference classification, Recommendation system, Data Mining, Fuzzy association rule

1. INTRODUCTION

The Internet, with continuous progress in computer technology and the pervasion of network environments, has become a major channel for modern people to acquire information (Chen and Tai, 2003). As amounts of information increase on the Internet, Users also face the serious problem of information overload simultaneously (Ribak et al., 2003; Sugiyama et al., 2004; Fang and Sheng, 2004). Under this circumstance, Web tools have become more important for developing better solutions to assist people in acquiring information easily and effectively (Sugiyama et al., 2004).

Thus far, Recommendation Systems (RSs) are regarded as one of effective solutions for assisting people to acquire information on the Internet (Garcia et al., 2004). Furthermore, RSs differ from traditional search engine tools or portal websites; it applies active information recommendation methods to replace passive keyword search methods. It is helpful for user to acquire suitable information in less time and effort (Chen and Tai, 2003). Nowadays, contented-based and collaborative filter are two popular means applied in the RSs. Although they really emerge excellent performance in information recommendation, there are some problems still cannot be resolve. In order to provide user more abundance and suitable information, a new recommendation concept is proposed in this study, namely club member, differs from content-based and collaboration filter method. It allows a user can acquire a variety of information from some clubs that he/she interests.

In order to classify users into multi-club with different preference degrees, fuzzy association rule based on data mining technology is applied in this study. Fuzzy association rules are generated by discovering and analyzing the members’ feature in the same club. Those fuzzy association rules can classify a new coming user into those clubs that he/she maybe interests with preference degrees. It is helpful for an information RS to provide user more suitable information in accordance with their preferences precisely.

2. RECOMMANDATION SYSTEM

Generally speaking, RS is applied to recommend user suitable information, e.g. news, products or services, which meet he/she interests or requirements (Chen and Tai, 2003). For the sake of achieving this goal, it must rely on several effective methods to analyze or predict user’s preference and purchase trend to improve the accuracy ratio of recommendation (Garcia et al., 2004; Zhang and Iyenger, 2002). Recently, most of recommendation means can be divided into collaborative filter and contented-based. Their definitions, advantages and drawbacks are described as follows (Weng and Liu, 2004; Wang et al., 2004)

2.1. Content-based approach

It recommends user information according to the results of similarity measurement between item-feature and target-feature. In general, a term vector is applied to represent information content and relevant profile (e.g. category, user and information) (Shih and Karger, 2004;
Wang et al., 2004). A recommendation system will process similarity measurement and determine how to recommend information to user (as shown in Fig. 1).

There are several ways can be applied as follows:
1. According to the result of similarity measurement between each piece of information and each category, information can be classified to appropriate category. A user can acquire recommendation information from his/her preference categories.
2. According to the result of similarity measurement between each piece of information and each user profile, information will be recommended to those users whose feature is similar to information feature directly.
3. According to the result of similarity measurement between each piece of information and the other information, information will be recommended to those users who had read relevant information in the past.

The club member concept can predict a user’s interest not only from the other members in the same club, but also from the other interesting clubs (as shown in Fig. 3). According to the result of similarity measurement in the same group (as shown in Fig. 2). Collaborative filter analysis possesses advantages include:
1. The RS recommends information to all members in the same group; meanwhile, item-feature need not be analyzed in advance. This advantage makes multi-media information can be recommended without content definition or similarity measurement.
2. The recommended item could be different from users’ past preference or experiments. Therefore, it is possible to draw up potential needs and interests of users (Herlocker et al., 2000).
3. Even though a user has not any browsing record or purchasing history, he/she can still acquire information from the other members in the same group.

It also possesses drawbacks include:
1. The RS must perform user clustering to locate user-group with similar interest or pattern.
2. Information (item) not yet rated or purchased could not be recommended to other users (Weng and Liu, 2004).
3. As usual, clustering method makes the centroid unstable while a new user joins. Although it is an effective method to locate user’s group, it makes the members of user-group will unstable in any time.
4. When the RS perform user clustering, a user should belong to a specific group at least. Whereas a user cannot understand or realize the meaning of group which he/she stays.

2.2. Collaborative filter analysis

In this concept, a preliminary supposition is “a set of users possess similar feature (e.g. preference, taste, purchasing trend and browsing pattern) will possess the same interests. So a user will interest these information comes from the other members’ past browsing record or purchasing history in the same group.” Therefore, users are clustered to several groups according to their feature. The RS will recommend information to all members in the same group (as shown in Fig. 2). Collaborative filter analysis possesses advantages include:
1. The RS recommends information to all members in the same group; meanwhile, item-feature need not be analyzed in advance. This advantage makes multi-media information can be recommended without content definition or similarity measurement.
2. The recommended item could be different from users’ past preference or experiments. Therefore, it is possible to draw up potential needs and interests of users (Herlocker et al., 2000).
3. Even though a user has not any browsing record or purchasing history, he/she can still acquire information from the other members in the same group.

It also possesses drawbacks include:
1. The RS must perform user clustering to locate user-group with similar interest or pattern.
2. Information (item) not yet rated or purchased could not be recommended to other users (Weng and Liu, 2004).
3. As usual, clustering method makes the centroid unstable while a new user joins. Although it is a effective method to locate user’s group, it makes the members of user-group will unstable in any time.
4. When the RS perform user clustering, a user should belong to a specific group at least. Whereas a user cannot understand or realize the meaning of group which he/she stays.

2.3. Club member concept

In a real world, people may possess several interests or hobbies with different preference degree at the same time. This concept differs from traditional content-base approach and collaborative filter analysis. It emphasizes that a user can be recommended to join some interesting clubs or groups under rational conditions. According to the multi-club concept, a user can acquire information not only from the other members in the same club, but also from the other interesting clubs (as shown in Fig. 3). The club member concept can predict a user’s interest intend and provide information meet he/she interests more flexible and effective in a RS. Several advantages of this method are described as follows:
1. Each club has clearly topic and meaning to represent itself, it is different from traditional ambiguity
in collaborative filtering analysis.
2. Once a user is classified to a specific club, he/she will stay in the club until he/she loses interest.
3. Each user can possess more interest choices with different preference degree. It can satisfy user’s requirements for multi-interest.
4. A user can acquire information shared by the other members in the same club.
5. It is helpful for finding potential users from new coming users, and also eliciting new products consuming target from existed members.
6. Collaborative filtering and content-based are still support this concept, users can share information each other in the same club and acquire information that classified to this club by RS.

In this paper, each quantitative attribute is partitioned by simple fuzzy partition methods with triangle membership function. For example, the attribute “Age” is divided into linguistic values (low, medium, high) and its membership function is predefined as follows (shown as Fig. 4). These membership functions can be represented by triangle fuzzy numbers. For example, the linguistic variable $A_{Age}^{med}$ can be represented as $A_{Age}^{med} = (l, m, u)$, where $m$ is the hedge value of linguistic variable $A_{Age}^{med}$.

![Fig. 4 Linguistic values of “Age”](image)

### 3. FUZZY ASSOCIATION RULE

In general, the interesting or preference of each user is difficult to describe clearly. In order to provide users multi-club service in the RS, fuzzy association rules are applied to classify users into appropriate clubs. For the sake of obtaining valuable fuzzy association rule for classification problem, data mining technology based on Apriori algorithm (Hu et al., 2004) is applied in this study.

#### 3.1. Fuzzy partition methods

Fuzz sets and linguistic variables concepts were proposed by Zadeh (1965, 1975ab, 1976); they are feasible to represent the diversity of degree of the feature with different linguistic value in many applications (Hu et al., 2004). Formally, a linguistic variable is characterized by a quintuple (Pedrycz and Gomide, 1998; Chen and Jong, 1997; Zimmermann, 1991) denoted by $(x, T(x), U, G, M)$, in which $x$ is the name of the variable; $T(x)$ denotes the set of names of linguistic values or terms, which are linguistic words or sentences in a natural language, of $x$; $U$ denotes a universe of discourse; $G$ is syntactic rule for generating values of $x$; and $M$ is semantic rule for associating a linguistic value with a meaning. Using the simple fuzzy partition methods, each attribute can be partitioned by various linguistic values.

In addition, each qualitative attribute is partitioned by a finite number of linguistic values in accordance with its categories. For example, the attribute “Gender” is divided into 2 linguistic values and its membership function is predefined as follows.

\[
\mu_{\text{male}}(x) = \begin{cases} 
1, & x = \text{male} \\
0, & \text{otherwise}
\end{cases} 
\]

\[
\mu_{\text{female}}(x) = \begin{cases} 
1, & x = \text{female} \\
0, & \text{otherwise}
\end{cases} 
\]

Each attribute is partitioned into appropriate linguistic value and obtained its membership in accordance with membership function, respectively. These linguistic...
values are regarded as initial candidate fuzzy grids to generate the other large fuzzy grids and fuzzy association rule for classification in following tasks.

### 3.2. Fuzzy association rule

In order to generate valuable fuzzy association rules for classification, the large fuzzy grid concept is applied in this paper. If a database possesses \( k \) attributes and \( n \) tuples of member profile, the fuzzy classification rules are generated as follows.

Step 1. Perform the fuzzy partition to generate initial candidate fuzzy grids.

Step 2. Scan the database, and the fuzzy support (FS) value is the membership function of is calculated as follows (Ishibuchi et al., 2001a; Ishibuchi et al., 2001b; Hu et al., 2002).

\[
FS(A_{s_i}^{x_i}) = \frac{1}{n} \sum_{j=1}^{n} \mu_{s_i}^{x_i}(d_j^{x_i})
\]

(6)

where, \( x_i \) is the \( i \)th attribute, \( s_i \) is the term of degree in \( x_i \), \( A_{s_i}^{x_i} \) is the linguistic value of \( s_i \) in the \( i \)th attribute, \( \mu_{s_i}^{x_i} \) is the membership function of \( s_i \) in the \( i \)th attribute, \( d_j^{x_i} \) is the original value of the \( i \)th attribute in the \( j \)th tuple.

Step 3. If the FS value of linguistic value is larger than or equal to the user-specified min FS, it will be regarded as an eligible fuzzy grid.

Step 4. Those eligible fuzzy grids are combined to calculate their FS value as follows.

\[
FS(A_{s_1}^{x_1}, A_{s_2}^{x_2}, ..., A_{s_n}^{x_n}) = \frac{1}{n} \sum_{j=1}^{n} [\mu_{s_1}^{x_1}(d_j^{x_1}) \cdot \mu_{s_2}^{x_2}(d_j^{x_2}) \cdot \ldots \cdot \mu_{s_n}^{x_n}(d_j^{x_n})]
\]

(7)

Step 5. Go to step 3 to check the FS value of each fuzzy association rule. It will stop when only one rule is larger than or equal to the user-specified min FS. Otherwise, the processes between steps 3 to 5 will repeat again.

Step 6. Filter out ineffective fuzzy association rules with fuzzy confidence. The general fuzzy classification rule denoted by \( R \) is stated as

\[
R : A_{s_1}^{x_1} \times A_{s_2}^{x_2} \times \ldots \times A_{s_n}^{x_n} \Rightarrow A_{s_k}^{x_k} \text{ with } CF(R)
\]

(8)

where, \( x_\alpha \) (\( 1 \leq \alpha \leq k \)) is the class label and CF(R) is the certainty grade of \( R \). It means that: if \( x_1 \) is \( A_{s_1}^{x_1} \) and \( x_2 \) is \( A_{s_2}^{x_2} \) and ... and \( x_k \) is \( A_{s_k}^{x_k} \), then \( x_\alpha \) is \( A_{s_\alpha}^{x_\alpha} \) with certainty CF(R). The fuzzy confidence (Ishibuchi et al., 2001a; Ishibuchi et al., 2001b; Hu et al., 2002) of \( R \) is defined as follows:

\[
FC(R) = \frac{FS(A_{s_1}^{x_1} \times A_{s_2}^{x_2} \times \ldots \times A_{s_n}^{x_n})}{FS(A_{s_1}^{x_1} \times A_{s_2}^{x_2} \times \ldots \times A_{s_n}^{x_n})}
\]

(9)

When the FC(R) of rule \( R \) is not larger than or equal to the user-specified min FC, the rule \( R \) will be regarded as ineffective fuzzy association rule and filtered out.

Step 7. Prune the redundant fuzzy association rules. If there exist two rules, denoted by \( R \) and \( S \), possessing same consequence, and the antecedence of \( R \) is contained in \( S \), then \( R \) is redundant and can be discarded.

### 4. EXAMPLE

There are 10 tuples of members profile in an exercise club as shown in Table 1. In this paper, the fuzzy classification rule is applied to classify a new coming user into multi-club with different preference degree. Hence, it is essential to locate effective association rule from the profile of existing member in each club. According to these classification rules, a new user can be classified into interesting clubs more effectively. The processes of generate fuzzy classification rule are described as follows.

Step 1. Perform fuzzy partition to transform original value of each attribute into linguistic value. We apply membership mapping table (shown as Table 2) to assign each education degree with a suitable membership in accordance with linguistic value, respectively. The other attributes are partitioned into linguistic values and obtained membership by membership functions. The linguistic value of each attribute is shown as Table 3. For example, the original value of attribute “Age” of member 10 is 43, and then membership degrees are 0.65, 0.35 and 0 for \( A_{med}^{Age} \), \( A_{low}^{Age} \) and \( A_{high}^{Age} \), respectively.

Step 2. Scan the database, and the FS of each linguistic value is calculated (shown as Table 4). For example, the FS of \( A_{high}^{Age} \) is calculated as.

\[
FS(A_{high}^{Age}) = \frac{1}{10} (1 + 1 + 0.2 + 0 + 0.1 + 0.3 + 1 + 0 + 0 + 0.65) = 0.43
\]

Step 3. Find the eligible fuzzy grids that FS value is larger than or equal to the user-specified min FS (In this case, the min FS is 0.3).

Step 4. In this case, the fuzzy classification rules aim to judge a new coming user whether interests this club or not. Therefore, the eligible linguistic value of “Preference” will be regarded as classification label to mining the association rules between preference degrees with other attributes.

Step 5. After the third scan, only one rule is larger than or equal to the user-specified min FS and the eligible large fuzzy grids are shown in Table 5.

Step 6. Calculate FC value of those eligible large fuzzy grids, and then filter out those ineffective fuzzy grids rules that FC value is smaller than the
user-specified min FC (in this case, the min FC is 0.5).

Step 7. When the redundant fuzzy association rules are pruned, the fuzzy classification rule is remained only one as “If Age is medium, Education is high, then Preference is medium.”

Table 1 Member’s profile

<table>
<thead>
<tr>
<th>Member</th>
<th>Age</th>
<th>Education</th>
<th>Income(K)</th>
<th>Preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>Doctor</td>
<td>56</td>
<td>0.96</td>
</tr>
<tr>
<td>2</td>
<td>56</td>
<td>Bachelor</td>
<td>67</td>
<td>0.89</td>
</tr>
<tr>
<td>3</td>
<td>34</td>
<td>Master</td>
<td>34</td>
<td>0.56</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>Bachelor</td>
<td>34</td>
<td>0.34</td>
</tr>
<tr>
<td>5</td>
<td>32</td>
<td>Master</td>
<td>23</td>
<td>0.52</td>
</tr>
<tr>
<td>6</td>
<td>36</td>
<td>Master</td>
<td>78</td>
<td>0.73</td>
</tr>
<tr>
<td>7</td>
<td>53</td>
<td>Bachelor</td>
<td>74</td>
<td>0.62</td>
</tr>
<tr>
<td>8</td>
<td>28</td>
<td>Master</td>
<td>32</td>
<td>0.55</td>
</tr>
<tr>
<td>9</td>
<td>25</td>
<td>Bachelor</td>
<td>55</td>
<td>0.47</td>
</tr>
<tr>
<td>10</td>
<td>43</td>
<td>Doctor</td>
<td>60</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Table 2 Education membership mapping

<table>
<thead>
<tr>
<th>Education</th>
<th>Top</th>
<th>High</th>
<th>Medium</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doctor</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Master</td>
<td>0.8</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Bachelor</td>
<td>0.5</td>
<td>0.8</td>
<td>0.3</td>
<td></td>
</tr>
</tbody>
</table>

Table 3 Linguistic value of each attribute

<table>
<thead>
<tr>
<th>Attribute</th>
<th>High (50,50,30)</th>
<th>Medium (50,30,10)</th>
<th>Low (30,10,10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>(90,90,50)</td>
<td>(90,50,20)</td>
<td>(50,20,20)</td>
</tr>
<tr>
<td>Preference</td>
<td>(0.8,0.8,0.04)</td>
<td>(0.8,0.4,0)</td>
<td>(0.4,0,0)</td>
</tr>
</tbody>
</table>

Table 4 FS value of each linguistic value

<table>
<thead>
<tr>
<th>Attribute</th>
<th>High</th>
<th>Medium</th>
<th>Low</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.43</td>
<td>0.51</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>0.23</td>
<td>0.52</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>Preference</td>
<td>0.46</td>
<td>0.50</td>
<td>0.04</td>
<td></td>
</tr>
</tbody>
</table>

Table 5 Fuzzy classification rules

<table>
<thead>
<tr>
<th>Fuzzy association rule</th>
<th>FS(R)</th>
<th>FC(R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_{Age} \rightarrow A_{Pref}</td>
<td>0.34</td>
<td>0.47</td>
</tr>
<tr>
<td>A_{Age} \rightarrow A_{Prof}</td>
<td>0.35</td>
<td>0.49</td>
</tr>
<tr>
<td>A_{Age} \rightarrow A_{Med}</td>
<td>0.43</td>
<td>0.47</td>
</tr>
<tr>
<td>A_{Age} \rightarrow A_{Prof}</td>
<td>0.46</td>
<td>0.50</td>
</tr>
<tr>
<td>A_{Age} \times A_{Age} \rightarrow A_{Prof}</td>
<td>0.33</td>
<td>0.64</td>
</tr>
</tbody>
</table>

5. CONCLUSION

As information explosive grows, people must face the information overload more serious than before. RSs provide several solutions for people to acquire information meet their requirements. In order to achieve this goal, RSs need more effective and precisely method to improve the quality of information recommendation.

In this paper, we propose a club-member concept to allow a user can acquire more information from a variety of interesting clubs. In addition, fuzzy association rule based on data mining technology is applied to classify users into multi-club with different preference degrees in accordance with his/her personal features. Those fuzzy association rules can assist RSs to classify a new user into some suitable or interesting clubs for user. It is helpful for user to acquire more information from RSs more easily and effectively.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support of the National Science Council, Taiwan, China, under project numbers NSC 92-2213-E-212-008.

REFERENCES


