ICEB 2004 Proceedings

International Conference on Electronic Business (ICEB)

Winter 12-5-2004

A User Preference Classification Method in Information Recommendation System

Chentung Chen

Weishen Tai

A User Preference Classification Method in Information Recommendation System

Chentung Chen[1], Weishen Tai[2]

[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.

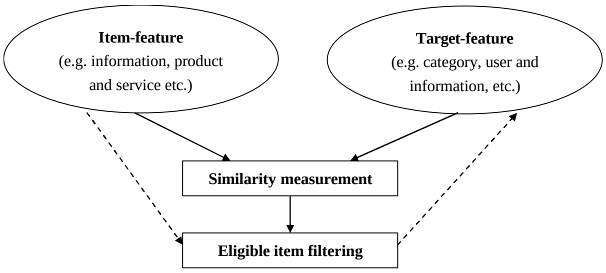

Fig. 1 Concept of content-based approach

Contented-based approach possesses advantages include:
1. A user will acquire information meet his preference feature or category appropriately.
2. Even though a piece of information cannot interest anyone in the past, it is still possible to be recommended to someone.
3. When user accepts recommendation information and makes feedback to the system, collaborative filter can support the RS (Weng and Liu, 2004).

It also possesses drawbacks include:
1. Multi-media information, such as image, sound and video data cannot be transformed into term vector. It makes the RS cannot define the content of those data type and process similarity measurement.
2. If the RS recommends user information according to user past browsing record or purchasing history, unique or different information will hardly be recommended to user (Weng and Liu, 2004).
3. Both the content of term vector and the number of term will determine the similarity degree between two features.
4. A user cannot acquire any information come from the other user with similar interests.

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.

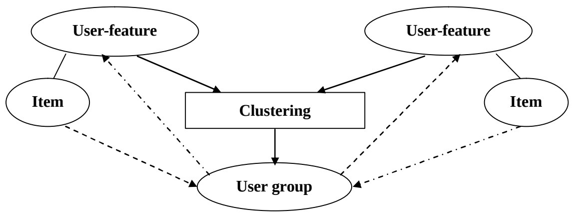

Fig. 2 Concept of collaborative filter analysis

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 loss 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.

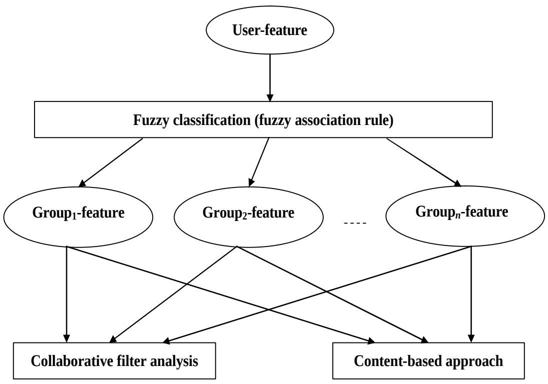

Fig. 3 Concept of club member

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 attributes can be partitioned by various linguistic values.

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 3 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_{med}^{Age} can be represented as $A_{med}^{Age} = (l, m, u)$, where m is the hedge value of linguistic variable A_{med}^{Age}.

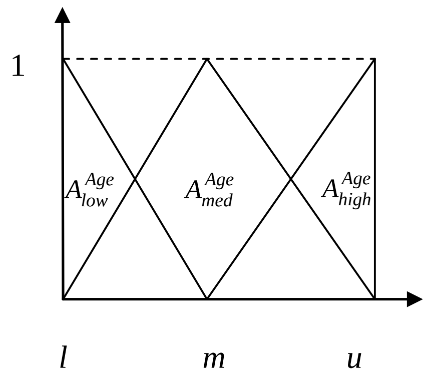

Fig. 4 Linguistic values of "Age"

$$m_{low}^{Age}(x) = \begin{cases} 1 & , x < l \\ \dfrac{m-x}{m-l} & , l \le x \le m \\ 0 & , x > m \end{cases} \qquad (1)$$

$$m_{med}^{Age}(x) = \begin{cases} \dfrac{x-l}{m-l} & , l < x \le m \\ \dfrac{u-x}{u-m} & , m < x < u \\ 0 & , otherwise \end{cases} \qquad (2)$$

$$m_{high}^{Age}(x) = \begin{cases} 0 & , x < m \\ \dfrac{u-x}{u-m} & , m \le x \le u \\ 1 & , x > u \end{cases} \qquad (3)$$

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.

$$m_{male}^{Gender}(x) = \begin{cases} 1 & , if\ x = male \\ 0 & , otherwise \end{cases} \qquad (4)$$

$$m_{female}^{Gender}(x) = \begin{cases} 1 & , if\ x = female \\ 0 & , otherwise \end{cases} \qquad (5)$$

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 rule 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 fuzzy support (FS) value is calculated as follows(Ishibuchi et al., 2001a; Ishibuchi et al., 2001b; Hu et al., 2002).

$$\mathrm{FS}(A_{s_i}^{x_i}) = \frac{1}{n}\sum_{j=1}^{n} m_{s_i}^{x_i}(d_j^{x_i}) \tag{6}$$

where, x_i is the ith 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 ith attribute, $m_{s_i}^{x_i}$ is the membership function of s_i in the ith attribute, $d_j^{x_i}$ is the original value of the ith attribute in the jth 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.

$$\mathrm{FS}(A_{s_1}^{x_1}, A_{s_2}^{x_2}, \ldots, A_{s_k}^{x_k}) = \frac{1}{n}\sum_{j=1}^{n}[m_{s_1}^{x_1}(d_j^{x_1}) \bullet m_{s_2}^{x_2}(d_j^{x_2}) \bullet \ldots \bullet m_{s_k}^{x_k}(d_j^{x_k})] \tag{7}$$

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

$$R: A_{S_1}^{x_1} \times A_{S_2}^{x_2} \times \ldots \times A_{S_k}^{x_k} \Rightarrow A_{S_a}^{x_a} \quad \text{with CF}(R) \tag{8}$$

where, $x_a\ (1 \le a \le k)$ 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_a is $A_{s_a}^{x_a}$ with certainty grade CF(R). The fuzzy confidence (Ishibuchi et al., 2001a; Ishibuchi et al., 2001b; Hu et al., 2002) of R is defined as follows:

$$\mathrm{FC}(R) = \mathrm{FS}(A_{s_1}^{x_1} \times A_{s_2}^{x_2} \times \ldots \times A_{s_k}^{x_k} \times A_{s_a}^{x_a}) / \mathrm{FS}(A_{s_1}^{x_1} \times A_{s_2}^{x_2} \times \ldots \times A_{s_k}^{x_k}) \tag{9}$$

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_{high}^{Age}, A_{med}^{Age} and A_{low}^{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 association 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

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

Table 2 Education membership mapping

Education	Top	High	Medium
Doctor	1	1	0
Master	0.8	1	0
Bachelor	0.5	0.8	0.3

Table 3 Linguistic value of each attribute

Attribute	High	Medium	Low
Age	(50,50,30)	(50,30,10)	(30,10,10)
Income	(90,90,50)	(90,50,20)	(50,20,20)
Preference	(0.8,0.8,04)	(0.8,0.4,0)	(0.4,0,0)

Table 4 FS value of each linguistic value

Attribute	High	Medium	Low
Age	0.43	0.51	0.07
Income	0.23	0.52	0.26
Preference	0.46	0.50	0.04
	Top	High	Medium
Education	0.72	0.92	0.12

Table 5 Fuzzy classification rules

Fuzzy association rule	FS(R)	FC(R)
$A_{top}^{Edu} \Rightarrow A_{high}^{pref}$	0.34	0.47
$A_{top}^{Edu} \Rightarrow A_{med}^{pref}$	0.35	0.49
$A_{high}^{Edu} \Rightarrow A_{high}^{pref}$	0.43	0.47
$A_{high}^{Edu} \Rightarrow A_{med}^{pref}$	0.46	0.50
$A_{med}^{Age} \Rightarrow A_{med}^{pref}$	0.33	0.64
$A_{med}^{Age} \times A_{high}^{Edu} \Rightarrow A_{med}^{pref}$	0.30	0.63

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

[1] Chen, C.T. , W.S. Tai, "An information push-delivery system design for personal information service on the Internet," *Information Processing and Management*, Vol. 39, Issue 6, pp. 873-888, 2003.

[2] Chen, S.M., W.T. Jong, "Fuzzy query translation for relational database systems," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 27, No. 4, pp. 714-721, 1997.

[3] Fang, X., O.R. Liu Sheng, "LinkSelector: A Web mining approach to hyperlink selection for Web portals," *ACM Transactions on Internet Technology (TOIT)*, Vol. 4, Issue 2, pp. 209-237, 2004.

[4] Garcia, A.C.B., M. Ekstrom, H. Björnsson, "HYRIWYG: leveraging personalization to elicit honest recommendations," *Proceedings of the 5th ACM conference on Electronic commerce*, New York, USA, pp. 232-233, 2004.

[5] Herlocker, J.L., J.A. Konstan, J. Riedl, "Explaining collaborative filtering recommendations," *Proceedings on the ACM 2000 Conference on Computer Supported Cooperative Work*, Philadelphia, USA, pp. 241-250, 2000.

[6] Hu, Y.C., J.S. Hu, R.S. Chen, G.H. Tzeng, "Assessing weights of product attributes from fuzzy knowledge in a dynamic environment," *European Journal of Operational Research*, Vol. 154, Issue 1, pp. 125-143, 2004.

[7] Hu, Y.C., R.S. Chen, G.H. Tzeng, "Mining fuzzy association rules for classi?cation problems," *Computers and Industrial Engineering*, Vol. 43, Issue 4, pp. 735-750, 2002.

[8] Ishibuchi, H., T. Nakashima, T. Yamamoto. "Fuzzy association rules for handling continuous attributes," *Proceedings of IEEE International Symposium on Industrial Electronics*, Pusan, Korea, pp. 118-121, 2001b.

[9] Ishibuchi, H., T. Yamamoto, T. Nakashima, "Fuzzy data mining: e? ect of fuzzy discretization," *Proceedings of the 1st IEEE International Conference on Data Mining*, San Jose, USA, pp. 241-248, 2001a.

[10] Pedrycz, W., F. Gomide, *An Introduction to Fuzzy Sets: Analysis and Design*, Cambridge: MIT Press, 1998.

[11] Ribak, A., M. Jacovi, V. Soroka, ""Ask before you search": peer support and community building with reachout," *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, New Orleans, Louisiana, USA, pp. 126-135, 2002.

[12] Shih, L.K., D.R. Karger, "Using urls and table layout for web classification tasks" *Proceedings of the 13th conference on World Wide Web*, New York, USA, pp. 193-202, 2004.

[13] Sugiyama, K., K. Hatano, M. Yoshikawa, "Adaptive web search based on user profile constructed without any effort from users," *Proceedings of the 13th conference on World Wide Web*, New York, USA, pp. 675-684, 2004.

[14] Wang, Y.F., Y.L. Chuang, M.H. Hsu, H.C. Keh, "A personalized recommender system for the cosmetic business," *Expert Systems with Applications*, Vol. 26, Issue 3, pp. 427-434, 2004.

[15] Weng, S.S., M.J. Liu, "Feature-based recommendations for one-to-one marketing," *Expert Systems with Applications*, Vol. 26, Issue 4, pp. 493-508, 2004.

[16] Zadeh, L.A. "The concept of a linguistic variable and its application to approximate reasoning," *Information Science*, Vol. 8, Issue 3, pp. 199-249, 1975a.

[17] Zadeh, L.A. "The concept of a linguistic variable and its application to approximate reasoning," *Information Science*, Vol. 8, Issue 4, pp. 301-357, 1975b.

[18] Zadeh, L.A. "The concept of a linguistic variable and its application to approximate reasoning," *Information Science*, Vol. 9, Issue 1, pp. 43-80, 1976.

[19] Zadeh, L.A., "Fuzzy sets," *Inform. and Control*, Vol. 8, pp. 338-353, 1965.

[20] Zhang, T., V.S. Iyengar, "Recommender systems using linear classifiers," *The Journal of Machine Learning Research*, Vol. 2, pp. 313-334, 2002.

[21] Zimmermann, H.J. *Fuzzy Sets, Decision Making, and Expert Systems*, Boston: Kluwer Academic Publisher, 1991.