Proceedings of 2007 IEEE International Conference on Grey Systems
and Intelligent Services, November 18-20, 2007, Nanjing, China

Personality Mining Method in Web Based Education System Using Data Mining

Jin Huang,Aiqin Zhu,Qi Luo

Abstract—**Web based education has become the new growth point of education development and primary developing trend of modern education technology. To meet the personalized needs of web based education; an improved association mining rule was proposed in the paper. First, data cube from database was established. Then, frequent item-set that satisfies the minimum support on data cube was mined out. Furthermore, association rules of frequent item-set were generated. Finally, redundant association rules through the relative method in statistics were wiped off. The algorithm had two advantages, and the first was that the execution time was short while searching for the frequent item-set; the second was that the precision of the rules was high. The algorithm was also used in personality mining system based on web based education model. The result manifested that the algorithm was effective.**

I. INTRODUCTION

WEB based education has become the new growth point of education development and primary developing trend of modern education technology. There are many problems in current web based education, such as singularity teaching mode, educational resource stacking simply, low intelligence level etc [1]. But different student has different learning capability, personal interest, personal learning foundation and struggling level. Giving corresponding learning content and tactics to realize teaching learners according to their needs is very difficult. Its basic reason lies in being difficult by obtaining the relations between the learner's personality characteristics and learning behavior patterns accurately and automatically. Data mining provides a strong support for personalized network education [2].In this way, it is necessary to mine out the association rules between personality characteristics and learning behavior patterns. The subsequent learners' personality and characteristics are deduced from their learning behaviors by using the rules above. Based on it, personality-learning model is set up and interesting groups are formed. Thus, personalized learning and cooperative learning are realized.

At present, many scholars have carried on a great deal of researches on association algorithms, such as such as Apriori algorithm or improved Apriori algorithm FP-tree [3]. The execution time of their algorithms was long while searching for the frequent item-set and the rule's interest degree was low. Even a large number of redundant rules are also included [4]. Besides, the complexity of space and time is high. According to this, an improved association rule mining algorithm is proposed in the paper. It has improved the traditional algorithm, and introduced data cube, relative method in statistics and adaptive adjusting mechanism in intelligence control [5]. The algorithm had two advantages. The first was that the execution time was short while searching for the frequent item-set; the second was that the precision of the rules was high. The algorithm was also used in personality mining system based on Web based education model. The result manifested that the algorithm was effective.

II. AN IMPROVED ASSOCIATION RULES

The steps of an association rules mining based on data cube as follows:

Step1: establishing data cube from database

Step2: mining frequent item-set which satisfies the minimum support on data cube

Step3: generating association rules of frequent item-set.

Step4: redundant association rules are wiped off through the relative method

A. Establishing Data Cube

Supposed in personality model database, we observe the data by three dimensional angles. These three dimensions separately are learning behavior patterns dimension, learner's personality characteristics dimension, and time dimension. Then data cube is obtained through OLAP operation from data base [6]. Fig.1 is data cube by 3-D.

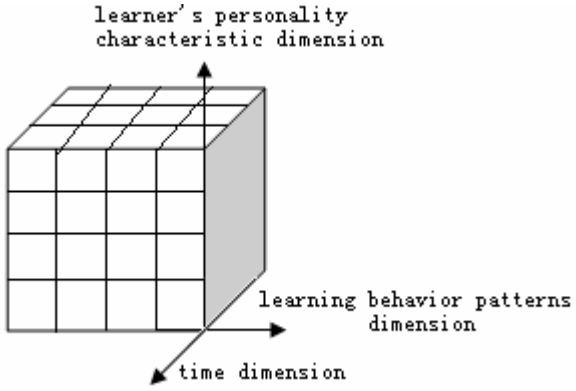

Fig. 1 Data cube by 3-D

Each dimension relates to one table that is called dimension table. It carries on further description to the dimension. For example, personality characteristics dimension contains name, sex, age, income, occupation, hobby, character and so on.

Above the specific angle (dimension) also has many different descriptions in detail degree. This kind of description is called dimension level. One dimension generally has the many levels. For example, the description of time dimension may describe different level from year, quarter, month, date and so on. Year, quarter, month, and date are the dimension level of time.

.

B. Mining Frequent Item-set on Data Cube

The paper has improved Apriori algorithm. The algorithm searching for the frequent item-set on data cube is called Personal_ Cube_Apriori algorithm. L_k Represents set of frequent k_ item-set, C_k Represents set of candidate k_ item-set. The idea of adaptive adjusting of the Personal_ Cube_Apriori algorithm as follow:

Firstly, the algorithm searches for frequent k_ item-set for each dimension. If some dimensions don't have frequent k_ item-set, it shows that the dimension level is excessively low and we should drill through above and improve the dimension level. If in some dimensions, all frequent k_ item-set is frequent k_ item-set, it shows that the dimension level is excessively high and we should drill through under and lower dimension level.

Fig .2 is shown as the flow of Personal_ Cube_Apriori

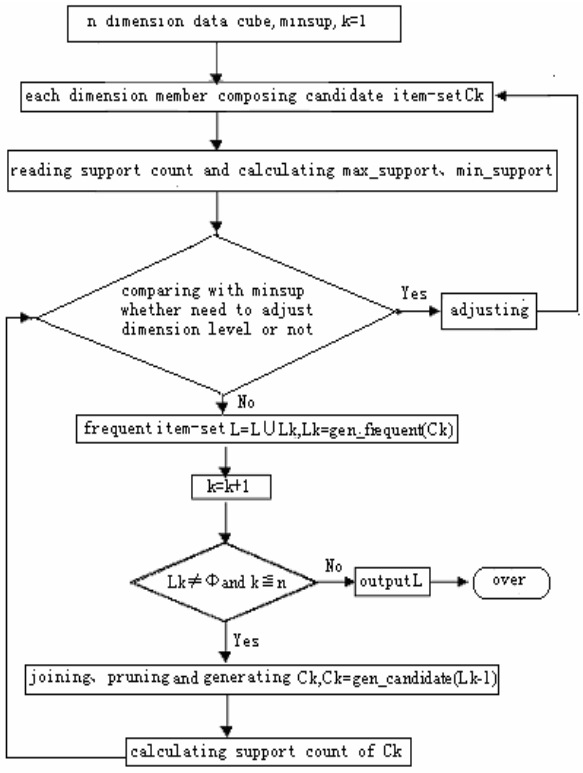

Fig.2. The flow of personal cube aprior.

Fig .2 is shown as the flow of Personal_ Cube_Apriori

Input: n dimension data cube $(d_1, d_2 \cdots d_n, count)$, minimum support min_sup, k=1

Output: set of frequent item-set $L = L \bigcup L_k$

$L_k = gen\_frequent(C_k)$ Represents set of candidate k_ item -set generating frequent k_ item-set

$C_K = gen\_candidate(L_{k-1})$ Represents set of frequent (k-1) _ item-set joining, pruning and generating k_ item-set

C. Generating Irredundant Association Rules

After mining out frequent item-set, the process of generating association rules is composed of two steps:

Step1: regarding to each frequent item-set *l*, all non-spatial subsets are generated.

Step2: regarding to each non- spatial subset of frequent item-set *l*, if

Then the rule *s*=> (*l*-*s*) is generated. min_conf represent the minimum confidence thresholds, support count(*l*) represents the number of transaction containing item-set *l*, support count(*s*) represents the number of transaction containing item-set *s*.

Lower interest degree rules are obtained in this way, so the redundant rules must be wiped off. Thus, the paper introduces

$$\frac{\sup port\_count(l)}{\sup port\_count(s)} \geq \min\_conf \qquad (1)$$

the relative analysis method in statistics to wipe off the redundant rules. After mining out one association rule, we calculate the correlation among frequent item-set. Concept of correlation is formula 2

$$CORR_r = \frac{O(r)}{E(r)} \qquad (2)$$

$CORR_r$ More approaches to 1, the independence among set of r item-set is better. For example, the association rule X => Y, X, Y are the item-set. If its correlation is worse, it must be the false strong rule.

Relative analysis method is applied in k dimension data cube. If item or dimension Y is $Y_1, Y_2, \cdots, Y_r$ r values, X is $X_1, X_2, \cdots X_s$ s values, n_{ij} represents the number of X *is* Y_i, and X *is* X_j.

$$n_i = \sum_{j=1}^{s} n_{ij} \quad i = 1, 2, \cdots, r \qquad (3)$$

$$n_j = \sum_{i=1}^{r} n_{ij} \quad j = 1, 2, \cdots, s \qquad (4)$$

$$n = \sum_{i=1}^{r}\sum_{j=1}^{s} n_{ij} \quad m_{ij} = \frac{n_i n_j}{n} \qquad (5)$$

$$\chi^2 = n\sum_{i=1}^{r}\sum_{j=1}^{s} \frac{(n_{ij} - m_{ij})^2}{m_{ij}} \qquad (6)$$

χ^2 Is correlation degree among items, if all items are independent, χ^2 is 0. Supposed the critical value is assigned. If χ^2 is bigger than the critical value, X and Y are statistical correlation. Otherwise, they are not statistical correlation. Namely, they are independent.

III. PMSEM

The technologies of personality, data mining are applied in the paper [7], while the model of personality mining system based on web based education is shown in Fig.3

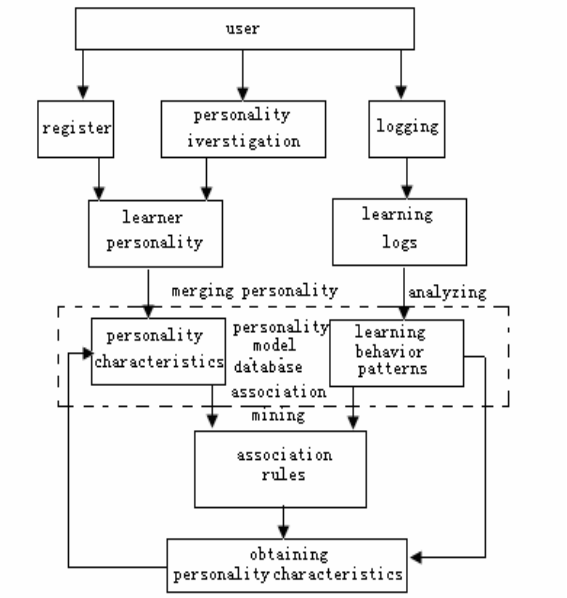

Fig.3. The model of Personality mining system based on Web based education

The main function of Personality mining system based on Web based education is to find out the association rules between personality characteristics and learning behavior patterns. The subsequent learners' personality characteristics can deduce from their learning behavior by using the rules above.

The personality model is composed of personality characteristics and learning behavior patterns. Personality characteristics including name, sex, age, income, occupation, hobby, character and so on. Learning behavior patterns including learning courseware, issue information on BBS and so on. A web sever log records user's IP address, visiting time, visiting page, visiting way, requested the document URL ,HTTP version, return codes, transmission bytes, protocol, wrong codes and so on.

The work process of model is as follows: firstly, learner personality information is collected though personality investigation and register information. Then, personality characteristics are obtained though personality merging. Redundant personality characteristics are wiped off though personality merging. Furthermore, learning behavior patterns are obtained by analyzing learning logs. Finally, behavior-personality rules are mining out by analyzing the association rules between personality characteristics and learning behavior patterns. Thus, the subsequent learners' personality characteristics are deduced from their learning behavior according to the rules.

IV. APPLICATION

Based on the above research, we combine with the cooperation item of personalized knowledge service system in network education .The author constructs a personality mining system website. The system is also applied in network institute of some university. The results manifest the system support personalized Web based education better the experiment is that we carry on the investigation and construct the database for personality attributes of 360 learners, 300 learners are selected to regard as the data source and other 60 learners data are selected to regard as the examination data.

The steps of experiment are shown as Fig.4

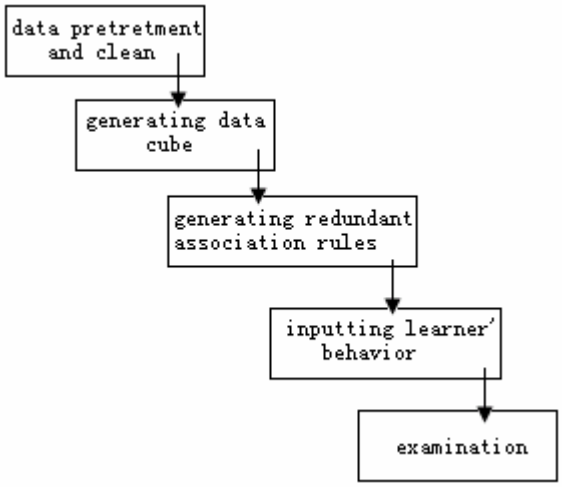

Fig.4. The steps of experiment

Step1: Data pretreatment and clean. First, not complete attributes values in database are filled in though rough theory. Then, according to the distribution of each attribute extreme value, sparser values is removed by minimum support min_sup =30%. Date converses to Boolean attributes according to the actual value scope of each attribute .thus, 100 attributes are obtained.

Step2: Generating data cube. Data cube is generated through OLAP operation. These are three dimensions such as learning behavior patterns dimension, learner's personality characteristics dimension, and time dimension.

Step3: Generating association rules. The frequent item-set is searched out though Personality_ Cube_Apriori. Then, χ^2 is calculated and some false strong rules are removed. As a result of the application environment limited, the partial learning behavior patterns dimension and learner's personality characteristics dimension are carried on analysis though improved association rule algorithm. Learning behavior patterns selects two aspects such as learning courseware and issue information on BBS. 9766 multi-dimensional association rules are obtained. 6794 association rules are obtained though $\chi^2 < 166.7$. There are 10 former and back integrity rules of them.

Step4: Inputting the following learners' behavior matches above rules, so learning characteristics is obtained. On the basis, best learning content and learning strategy are provided to them. So, personalized learning and cooperative learning are realized

Step5: 60 Learners' data is regarded as the examination data. Compared with above data, the result manifested that the system is effective.

V. CONCLUSION

In summary, an improved association algorithm is proposed in the paper. The results manifest that the algorithm

is effective through above algorithm performance research. Meanwhile, the model of personality mining system is proposed. The results manifest that the system mines out the association rules between personality characteristics and learning behavior patterns. The subsequent learners' personality characteristics can be deduced accurately and automatically from their learning behavior by using the rules above. I wish that this article's work could give some references to certain people.

REFERENCES

[1] Yizhi Wang.Research on Web personalized recommendation based on correlation analysis of association rule. Journal of Chongqing University of Posts and Telecommunications(Natural Science).2007.02

[2] Dongshan Xing. A construction of personalized network education system basedon web usage mining. Computer Application Software.2002.08.

[3] Liu Jun, Li Renhou and Zheng Qinhua, "Study on the Personality Mining Method for Learners in Network Learning". Journal of Xian Jiaotong University, 2004, 38(6), pp.575-576.

[4] Han J and Kamber M, "Data Mining Concept and Techniques". Academic Prints, 2001, pp.100-103.

[5] Lu Lina and Chen yaping, "Research on algorithm Apriori of mining association rules". MINI-MICRO SYSTEM, vol 21, 2001.9, pp.942-944.

[6] Yuan Wei, "data mining center of Statistics department of china renmin university", Statistics and information forum, 2002(1), pp.5-9.

[7] Xianghua Fu.Design and implementation of P2P based personalized Web search system. Computer Engineering and Applications.2007.07