

Electronic Book Recommendation Method Based on Group User Behavior Analysis

Li Peng^{1,2}, Zhang Ming-yue², LiangTian-ge² and Zhang Kai-hui³

¹*School of Software, Harbin University of Science and Technology, 150080 Harbin, China*

²*School of Computer Science and Technology, Harbin University of Science and Technology, 150080 Harbin, China*

³*Journal Center, HeiLongJiang University, 150080 Harbin, China
e_roc@126.com*

Abstract

With the rapid development of the Internet, for-profit site need to analyze the user's behavior and provide more satisfactory service. Therefore, the classification of network behavior analysis and the further research based on it are more and more urgent on the agenda. In this paper, a method based on similar aggregation user behavior analysis algorithm is proposed. Addressing the recommendation of personalized books problems is solved by this method. Firstly, the user behavior is analyzed by using the RFM model. Secondly, the Apriori algorithm based on weight increment is applied to mining association rules between users in line with the recent habits of users. Similarity is calculated by Apriori algorithm with using VSM model. In this paper, readers' browsing history of e-library which is provided by Harbin University of Science and Technology is used as experimental data. This method is compared with the method which does not use the weight increment and similar aggregation. Comparison of results showed that the method of our paper can meet the requirements of the Book Recommendation system.

Keywords: *Book recommendation; Behavior analysis; Incremental weight; Apriori algorithm*

1. Introduction

With the rapid popularization of the Internet, digital library resources are more abundant. Among the many library resources, users do not know how to quickly find the resources they really need [1]. Therefore, personalized recommendation system becomes an urgent demand products. Related research has become a hot issue in recent years in field of computer research, and received wide attention of many researchers at home and abroad [2].

Personalized recommendation system get the reader's interest hobby by online analyzing, personalized recommendations. Wherein, collaborative filtering is one of the most successful technology recommendation [3]. The technology has been widely used in various recommendation system. For example GroupLens, it can automatically provide their users with information which comes from people with similar interests. Users make evaluation to books and the system also provides users with a list of books that comes from other users. It is different from recommended system that needs to show the user's evaluation information. Phoaks application implicit user rating of user recommendation information, application in the business world, the famous e-commerce site Amazon.com evaluated according to the customer's purchase history or customer information application collaborative filtering policies recommended for users to buy products. Overall, the application of

collaborative filtering technology goal is for a particular user, based on his previous evaluation and other things would like Bear and Tiger evaluation of things, calculate the user's interest degree for something new, systems based on these recommended products [4].

The key of this mechanism is how to find and identify groups with similar preferences, and how to filter out their points of interest. At present, many domestic and foreign researchers have made a number of high-level methods and strategies to solve this problem, and have made tremendous progress. For example, Literature proposed a method which is based on other users' information and use Joint Model to predict the user point of interest [5]. Literature proposed a multi-feature based on the book personalized recommendation algorithm based on feature vector diagram classification and calculation of similarity books, readers and one feature vector computing similarity borrowing record readers[6]. Literature describes a recommended method based on social network, the author of this literature believes the social network friends should have similar preferences[7]. However, the number of FaceBook users accounted for less than 40% according to the statistics of YouTube users. Therefore, this method is only preliminary studies social networking application development[8]. Although the use of collaborative filtering methods to analyze the user's row has achieved great success, it still has some shortcomings need to be addressed and improved. In the preferred system, the biggest difficulty comes from namely book data and reading user data. The cost of classification and clustering in these two big data are very high, and even if successful it is difficult to change dynamically. Therefore, how can we avoid to directly process the two big data, and use other tactics to bypass it has become a direction of our exploration. We use behavior analysis method to analyze this problem and operating data will be converted to user behavior data. User behavior data is automatically generated by users within a period of time to book file operations (such as flip, draw marks, *etc.*), and behavioral data is a formal data to facilitate analysis and processing, but also adapt to the dynamic changes.

User behavior analysis early come from the field of management guidance through the analysis of customer behavior business operations management[9]. In recent years, scholars have thought this method is introduced to study the computer field, Liu apply the method of user behavior analysis on search engine performance auto-evaluation[10]. Chen research the user behavior analysis model, effectively curb invasive behavior untrusted cloud end users[11]. We think users' reading behavior reflect users' interest, according to this view further assumptions are put forward-users with similar browsing behavior have similar interests. Thus proposes to conduct a series of books with similar operations should have assumed that a similar user preferences and points of interest. All the technology used in this paper are designed to test this hypothesis is established or not.

2. Books Recommended Process Based on User Behavior Analysis

The basic flow chart of book recommendation system, as shown in Figure 1, is mainly to provide users with personalized book recommendation service. Users read electronic tag information of books by browsing interface, such as the number of words, style, name, country, and publication date and so on. The user can view book list and read their favorite books, database record book number, category, style and other information. Paper by three kinds of modules stage to present books recommended procedure:

(1) User Grouping module analyze users' behavior, this module records book data and the corresponding readers' data, and then it will translate that into data that

could reflect users' browser behavior. Rely on Log Data, this module classify congenial readers as a group.

(2) Data mining module the user log data for analysis and obtain user association rules frequent items improved Apriori algorithm based on the weight increment, so that the user can tap the rules of customary in recent behavior.

(3) Collaborative module is based on the ratio of similarity after similarity of users, users gather similar rules, and finally collaborative, will be similar to the results than do top-N recommendation stage.

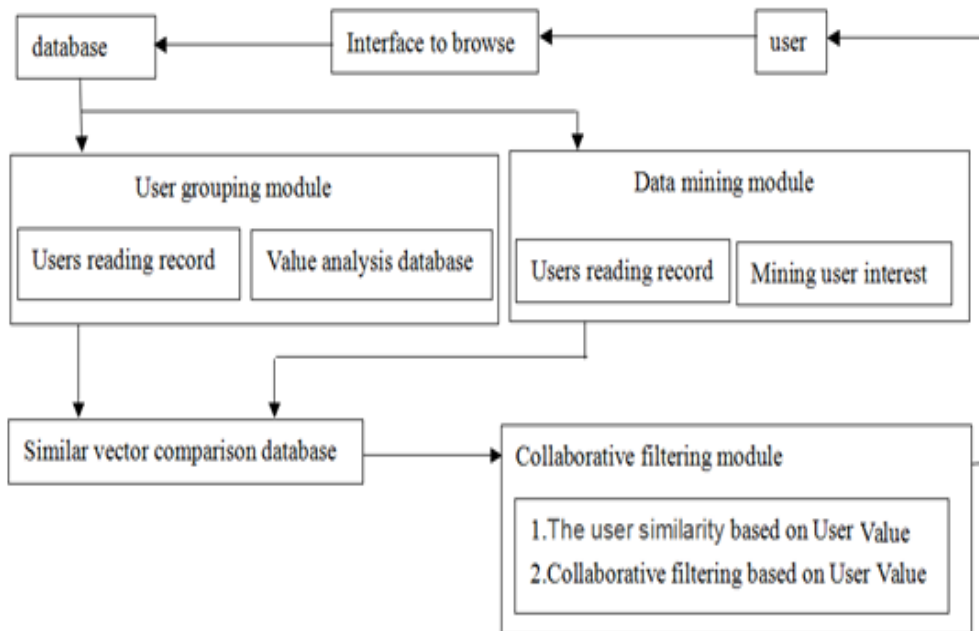


Figure 1. Book Recommended Process based on Behavior Analysis

3. User Behavior Analysis Algorithms Based on Hybrid Strategy

3.1. User Behavior Analysis Based on RFM Model

Behavior Indicators are composed by the general rules, through statistics and analysis of user behavior in the reading process, analyze users' behavior and master the rules of users' behavior, the behavior of the user can be predicted. By understanding characteristics and laws of behaviors, we achieve personalized recommendation. User behavior analysis indicator mainly analyzed from the following aspects.

According to research, analysis of the users' data required indicator, the indicator is made up of 3 special elements: Recentness, Frequency, Monetary Amount, these three elements be united to RFM models [12].

(1) Recentness refers to the length of time that from the user consumption for the last time to the analysis time. When its value is small, the probability of a user and then consumption is relatively large, and thus its Recentness eigenvalues is higher.

(2) Frequency refers to the number of products which are consumed by users in a Period of time. In general, the numbers of users consume more, the higher the customer value and loyalty. Conversely, the lower the customer value and loyalty.

(3) Monetary refers to the total amount of user spends on this product in a period of time. In general, the amount of consumption is more, the higher the value of the user.

This article analyzes the users' behavior, and maps it into three elements of as shown in Figure2. The last reading time as Recentness, Frequency of reading as Frequency, reading quantity as monetary. However, this paper will change the calculation of monetary into calculates the number of item sets. The number is larger, the more time spend on these kinds of Item sets, and each file is a category unit amount.

By analyzing the users' behavior, users can be divided into eight groups. According to RFM value of each user, on the basis of mean value of all users, \uparrow represented by a value greater than the overall average, while \downarrow less than the overall average. Using this notates can be divided into eight groups ($\uparrow\uparrow\uparrow$, $\uparrow\uparrow\downarrow$, $\uparrow\downarrow\uparrow$, $\uparrow\downarrow\downarrow$, $\downarrow\uparrow\uparrow$, $\downarrow\uparrow\downarrow$, $\downarrow\downarrow\uparrow$, $\downarrow\downarrow\downarrow$). By comparing RFM value and average, it can be confirmed that the group each user belongs and put user' data in corresponding group, systems for each group is assigned a different recommendation strategies.

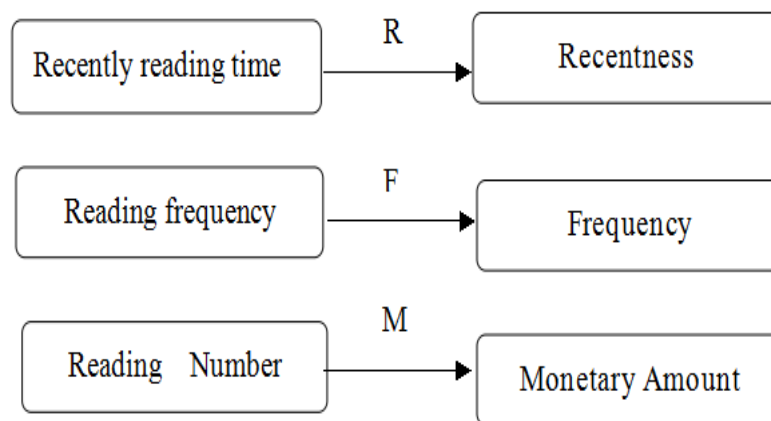


Figure 2. The Map of User Behavior Analysis and RFM

3.2. The Apriori Algorithm Based on Weight Increment

As traditional database need to calculate all the read data, so to get the users' high frequent file is bound to cause the system to perform time and increased costs, which affects the immediacy of the book recommendation system. Furthermore, the user has recently read selection will not necessarily have been around the same style. Therefore, this article uses incremental mining based on weight thought to identify the users' interests in the last time. The incremental data mining excavation can not only shorten the time it can also be dynamic mine the user's most recent habits.

Apriori algorithm as the mining Boolean association rule frequent item sets algorithm is by far one of the most important and influential association rules algorithm, its core is based on a two-stage recursive algorithm for frequent item set thinking[13]. In the weight increment thought, we set a support threshold, when the weight of support exceeds the support threshold to stop increment calculation. When the number of different increment calculation, arrangement results obtained will be different based on the iterative threshold setting, eliminating the need to set support threshold and minimum support threshold, so as to achieve the purpose of simplifying the calculation efficiency.

In order to calculate each category of transaction, whether the project is frequently set to define a right article weight support (*Weight Support - WS*). In case the value (β^{j-1}) of the option W_j reach the threshold value, or to determine whether a log transaction data is read, then stop calculating and delete option (W_j)'s value is zero. ws_i^j is weight support

when the i file reads the j transaction time; $Count_i^j$ is the sum of the number of data occurs when the i file appears in the j transaction time, j is the number of incremental mining, W size weight value is calculated, β^{j-1} is the $j-i$ power of β , is a constant, where $\beta < 1$

$$WS_i^j = WS_i^{j-1} + (Count_i^j \times W_j)$$

Where
$$WS_i^0 = 0, W_j = \beta^{j-1}, j = 1, 2, \dots, n \quad (1)$$

This method is the idea of the weight increment was added to the Apriori algorithm and make improvements in order to obtain the ideal study the rules, here are the steps described in mining rules.

1. Assume first deal with n transactions from back to front in database and count the number of each set of i .

2. In equation (1) to calculate the value of each category of WS , judge whether the value (β^{j-1}) of the option W_j reach the threshold value, or to determine whether a log transaction data is read.

3. If the data does not match, then continue to read the next set of read data, and recalculate the value of WS , until the value (β^{j-1}) of the option W_j reaches the threshold value, then stop calculating.

4. Delete the value of option (WS_i^j) is zero or delete the value of option is less than the $Min - Support$, and through a set of high-frequent category, a combination of two items candidate set, repeat the above manner in order to calculate each of the two categories set i .

5. Two candidate set based on the previous incremental number ($j = 1, 2, \dots, n$), continue to delete the value of option (WS_i^j) is zero or delete the value of option is less than the $Min - Support$.

6. The last remaining two items set categories will be considered customary rules of recent users.

$$Rbr = \{[i \rightarrow j] | Min - Support \leq WS([i \rightarrow j]), < 0.1\} \quad (2)$$

In equation (2), file WS greater than $Min - Support$ classified as recent habits and interests rule. Where $[i \rightarrow j]$ represents a possible project category.

3.3. User Similar Aggregation based on VSM Model

By the vector space model (VSM) the user similarity calculation, and based on user habits and interests has recently drawn the rules do user clustering of similar users gather again. Its purpose is to gather like-minded groups' project users find the user more similar between groups, to lower real collaborative filtering method to share information role defined as follows:

$SM(X)$ for the user U_x recently customary rules (*Recent behavior rules*) launched together in a similar matrix (*Similar Matrix*), Rbr was recently customary rules $[i \rightarrow j]$ category items may be set, as shown in equation (3).

$$SM(X) = [sm_{ij}]_{m \times n}; (m, n = 1, 2, 3, \dots, p)$$

Where
$$sm_{ij} = \begin{cases} 1, & \text{if } [i \rightarrow j] \in Rbr \\ 0, & \text{Otherwise} \end{cases} \quad (3)$$

The similarity calculating vector similarity vector is defined as follows:

$$SV(X) = [sm_{12}, sm_{13}, \dots, sm_{1n}, sm_{23}, sm_{24}, sm_{2n}, \dots, sm_{(m-1)n}] \quad (4)$$

In this paper, the similarity vector space model comparison between users. In the vector space model, the behavior of the two user similarity between D_1 and D_2 . $Sim(D_1, D_2)$ cosine of the angle between the vectors used, said the formula is:

$$Sim(D_1, D_2) = \cos \theta = \frac{\sum_{k=1}^n W_{1k} \times W_{2k}}{\sqrt{\left(\sum_{k=1}^n W_{1k}^2\right) \left(\sum_{k=1}^n W_{2k}^2\right)}} \quad (5)$$

Wherein, W_{1k} , W_{2k} represent text weight D_1 and D_2 K -characteristic items, $1 \leq k \leq N$. If the n-dimensional vector is $V\{v_1, v_2, v_3, \dots, v_n\}$, the modulus :

$$|v| = \text{sqrt}(v_1 \times v_1 + v_2 \times v_2 + \dots + v_n \times v_n),$$

the product of the two vectors $m \times n = n_1 \times m_1 + n_2 \times m_2 + \dots + n_n \times m_n$, then the similarity of two vectors is $(m \times n) / (|m| \times |n|)$.

The user can be calculated the most recent vector similarity by means of VSM model and vector of each element as a vector of feature items. If two users' categories like high similarity vector, then it will be recommended to a user's books to a high proportion of the number of other users, if the similarity of the two categories of users' similar vector is low, so it will be a user's Books in a low proportion of the recommended amount to other users.

3.4. Book Recommend Based on Collaborative Filtering

This article adopts the strategy of collaborative filtering to recommend books. In the grouping module, using RFM model for user behavior analysis, and divided into eight groups of them. This method will regard books which users have selected as recommended books. If the user is part of a group of high-value, it is recommended to book on mining results by collaborative filtering. On the contrary, if the user is a low value of the group, that according to the recommended books on mining results and random are not related books.

For example: user U_1 belongs to a group of high-value, its mining results is the $\{A \rightarrow C\}$ and $\{A \rightarrow D\}$ two rules. We will all users select books about the A, C, D and recommended to the user U_1 , to achieve accurate personalized book recommendations. If the user U_1 belongs to the low value of the group, we not only recommend related books, also add other types of books. This principle is similar to marketing advertising strategy and the purpose is to allow this group of users can log in several times, augment browsing history so as to provide users with better service.

For the same group, we conducted a similar user clustering within the group, gathering the similar style within the same group of users. This method utilizes the RFM model to classify and recommend books for different users. The purpose of this approach is through the second classification aggregation.

U_1 is assumed with other groups with similar preferences of the user U_2, U_3 , users can know the three selected book number (where v_t represents the first t products). U_2 and U_3 put in the same category in the books, do not repeat the recommendation to U_1 . For example, U_2 recommend $\{A : v_1, v_2, B : v_6, C : v_9\}$ to U_1 . Same category of U_3 is also recommended to U_1 . U_3 recommend $\{A : v_2, v_4, B : v_6, v_8\}$ to U_1 . Finally, combining the results of both U_2 and U_3 will not be repeated recommend to U_1 , that is $\{A : v_1, v_2, B : v_6, v_8, C : v_9\}$.

4. Experimental Validation and Analysis

The recommended method of evaluation for the book is a more difficult problem for the recommended subject is human. People's preference for books is also various in different times, under different circumstances, the same individual choices will be different. Therefore, one can not build a unified common data set to measure the pros and cons of various methods. The vast majority of experimental data comes from the users' feedback data, which to verify that the policy or the technology used is effective and stable. We also use these experimental strategies, by organizing part of the staff as the experimenters, to collect feedback information and verify the stability of the technology.

4.1. The Construction of Experimental Platform

To verify the effectiveness and stability of the proposed method, the paper built a Experimental platform. Experimental data are divided into five categories nearly 500 books, where the first 70 as training corpus for each class and 30 as the test corpus. In this study, a total of 20 the experimenter, they generate 1843 log data within 30 days.

Table 1. The Data and Style of Books

Num	Style	Total
101-200	Literature	100
201-300	live-aimed	100
301-400	Children	100
401-500	History	100
501-600	Politics	100

In order to prove the effectiveness and stability of the proposed method, we construct four methods to verify the RFM model, weight increment and similar aggregation in the course of the role. Four methods are as follows:

Method 1: using the weight increment and RFM model (IMW+RFM).

Method 2: using the weight increment, RFM model and similar aggregation (IMW+RFM+ Similarity).

Method 3: does not use the weight increment but using RFM model and similar aggregation (Non-IMW +RFM + Similarity).

Method 4: using the weight increment but not a member of any grouping method (IMW+ Non-Cluster).

The study use accuracy, recall and F value to measure the effectiveness of the experimental method. In order to calculate the accuracy and recall of the experimental data, the experimental data and matrix combination will be combined. Such performance can better describe the system as shown in Table 2.

Table 2. The Confusion Matrix of Classification

	Preference	No-preference
Proposal	<i>TP</i>	<i>FP</i>
NO-proposal	<i>FN</i>	<i>TN</i>

TP represents recommended and user real favorite books, *FP* represents recommended but not the user's favorite books, *FN* represents not recommended but users actually prefer books, *TN* represents neither a recommendation and users do not like books. Evaluation is calculated as follows:

$$Precision = \frac{TP}{(TP + FP)} \quad (6)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (7)$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (8)$$

4.2. Experimental Results and Analysis

By 15 experimenter behavioral data collection and analysis recommend books to readers with four methods which have been constructed, we obtained the following results:

Table 3. The Compare Results of four Methods

	IMW +RFM	IMW +RFM + Similarity	Non-IMW +RFM + Similarity	IMW +Non-Cluster
Average Precision	0.65	0.75	0.56	0.58
Average Recall	0.66	0.78	0.63	0.63
Average F-Measure	0.65	0.78	0.58	0.62

Compared method 2 with other methods, If we only consider the RFM model group, there maybe more other users' recommendation Therefore, the system may recommend some books that users are not interested. This experiment designed to prove the two groups was better than a packet. The third Method is to recommend books without using incremental mining. On the recommendation will be regardless of the user's preferences, entirely arbitrary recommend other books selected by the user, resulting in more messy recommended. Users may not like this recommended method. The accuracy of this method is significantly lower than the incremental mining methods, which demonstrates the importance of the weight of the incremental mining. In addition, the method 4 does not use packet technology and the average accuracy rate is 60 percent, significantly lower than the method using packet technology accuracy. Therefore, more efficient packet technology, which allows users to get the desired, rather than a bunch of books, causing the user does not know how to choose. Above experimental data prove the effectiveness 2, but also need to verify the stability of the method.

Based on the above analysis of the data, we can see the method of incremental weight and use the recommended method similar RFM model aggregation, better able to find the user's preferences and thus better compared to other basic methods recommended capacity, with relatively high at the same accuracy, to a certain

extent, also showed stability in the process. Therefore, we can prove that the three technologies, namely the weight of the incremental method in this study involved the excavation, in the group of users and user behavior similar to gather RFM model analyzes were based on the books promote positive recommendation is an effective means. In addition, further evidence of the previous assumption of this article is to set up, that is a series of books on the behavior of users with similar actions should have similar preferences and points of interest.

5. Conclusion

In this paper, firstly by RFM model, users with the same behavior are classified as a group; second, the Apriori algorithm based on weight increment is applied to mining association rules between users in line with the recent habits of users, and by using the VSM model for similarity calculation, the user similarity aggregation is realized; finally, the whole process of personalized book recommendation is completed by means of collaborative. It can recommend the book closer to the reader's preference. The above experimental results prove that the three techniques described herein may be used in a combination of methods is an effective strategy recommended books, achieves the desired effect.

The main contribution of this paper is to propose the method which is based on user behavior analysis algorithm. Now there is no literature to use the same method recommend books. By automatically collecting behavioral data, and analysis of these data through technical means to find a user with the same preferences, and then we collaborate recommended. Behavioral data enables dynamic real-time acquisition, behavioral data belonging to the formal data, little difficulty handling, faster processing speed, thereby enabling real-time updates, at great cost, but also to avoid large data directly processed. In the practical application of the system, the timeliness of the recommended behavior is often more important than the accuracy of the recommended methods, and therefore the application of the system does not only focus on the complexity of the algorithm, while the opposite should be looking for a simple, stable strategy. In future studies, we will continue to thoroughly explore this algorithm, actively study the behavior of the user's deep attribute characteristics and constantly enrich the connotation of behavior patterns.

Acknowledgment

This paper is partially supported by Natural Science Foundation of Heilongjiang Province (QC2013C060) and Youth Science Fund of Heilongjiang University (QL201404).

References

- [1] Q. H. Zeng and Y. H. Qiu, "An e-book recommender system with collaborative filtering", *Computer Science*, vol. 16, no. 6, (2005), pp. 55-62.
- [2] X. J. Zhao, J. Yuan and M. Wang, "Video recommendation over multiple information sources", *Multimedia Systems*, vol. 19, no. 1, (2011), pp. 3-15.
- [3] V. J. De, N. Degrande and M. Verhoeyen, "Video content recommendation: an overview and discussion on technologies and business models", *Bell Labs Technical Journal*, vol. 16, no. 2, (2011), pp. 235-250.
- [4] J. Park, S. Lee and K. Kim, "Online video recommendation through tag-cloud aggregation", *IEEE MultiMedia*, vol. 18, no. 1, (2011), pp. 78-87.
- [5] C. R. Su, Y. W. Li and R. Z. Zhang, "An adaptive video program recommender based on group user profiles", *Smart Innovation, Systems and Technologies*, vol. 21, no. 2, (2013), pp. 499-509.
- [6] K. C. Li and Z. Y. Liang, "Personalized book recommendation algorithm based on multi-feature", *Computer Engineering*, vol. 11, no. 2, (2012), pp. 11-17.
- [7] S. D. A. Da and L. K. Wives, "POI enhanced video recommender system using collaboration and social networks", *Proceedings of the 8th International Conference on Web Information Systems and Technologies*, (2012).

- [8] X. Q. Ma, H. Y. Wang and H. T. Li, "Exploring sharing patterns for video recommendation on YouTube-like social media", *Multimedia Systems*, vol. 11, no. 3, (2013), pp. 1-17.
- [9] D. E. Rapach and M. E. Wohar, "Forecasting the recent behavior of US business fixed investment spending: An analysis of competing models", *Journal of Forecasting*, vol. 26, no. 1, (2007), pp. 33-51.
- [10] Y. Q. Liu, R. W. Cen and M. Zhang, "Automatic search engine performance evaluation based on user behavior analysis", *Journal of Software*, vol. 19, no. 11, (2008), pp. 3023-3032.
- [11] Y. R. Chen, L. Q. Tian and Y. Yang, "Model and analysis of user behavior based on dynamic game theory in cloud computing", *ACTA Electronica Sinica*, vol. 39, no. 8, (2011), pp. 1818-1823.
- [12] T. Chen, "The RFM-FCM approach for customer clustering", *International Journal of Technology Intelligence and Planning*, vol. 8, no. 4, (2012), pp. 358-373.
- [13] M. H. Awadalla and S. G. Elfar, "Aggregate function based enhanced apriori algorithm for mining association rules", *International Journal of Computer Science Issues*, vol. 9, no. 3, (2012), pp. 277-287.