Personalized Recommendation System Based on Cloud Computing

Weiwei Jiao, Xin Li and Jingji Li

***Abstract*—**With the advent of cloud computing applications, the amount of data increases rapidly, personalized recommendation technology is becoming more and more important, however, because of the characteristics of cloud computing and large scale distributed processing architecture, the traditional recommendation techniques are applied directly to the cloud computing environment, will be faced with the recommended efficiency, accuracy and low personal problems, aiming at these problems, an improved method is proposed. This method combine's user's implicit query and context search query when users create interest models, so as to generate user personalized interest models, achieve immediate updating and accurately reflect user interest. In this paper, the principles and advantages of the improved method are analyzed in detail, and the flow chart of its recommendation steps is given.

***Keywords*—**cloud computing, personalized recommendation, implicit feedback, user model, information retrieval system.

I. INTRODUCTION

THE rapid growth of the Internet brings information overload. It is difficult for users to get valuable information from massive data, and the efficiency of information utilization is decreasing rapidly. Personalized recommendation is an important means of information filtering. By analyzing users' interests and historical behaviors, it recommends potential projects to users, effectively solving the problem of information overload on the Internet[1]. Recommendation system is based on recommendation technology and data mining theory. According to users' interest characteristics, we can dig out resources that users may be interested in or need from mass information, and make corresponding recommendations for users. The mainstream recommendation technology mainly includes the following contents: content based recommendation, recommendation based on user statistical information, recommendation based on collaborative filtering technology, and recommendation based on association rules. Recommendation based on collaborative filtering technology is one of the most successful technologies in the application of personalized recommendation system. It uses user preferences of interest or similar experience to recommend articles or resources of interest to users, whose biggest advantage is low data dependence. There are three main types of collaborative filtering based on collaborative filtering: collaborative filtering based on users, collaborative filtering based on objects and collaborative filtering based on model.

This work was supported in part by the Science and Technology Foundation of Liaoning Province (No. 2013020028).

Weiwei Jiao is with the Department of Computer Center, Liaoning University of Technology, Jinzhou 121001, Liaoning, China (corresponding author; e-mail: lgvivian@126.com).

Xin Li is with the Department of Computer Center, Liaoning University of Technology, Jinzhou 121001, Liaoning, China.

Jingji Li is with the Department of State Grid Liaoning Electric Power Corp Jinzhou Branch, Jinzhou 121001, Liaoning, China.

Because of the large scale and distributed processing architecture of cloud computing, the Internet application deployed in cloud computing environment is different from the traditional mode in data storage location, data storage capacity and application scale. Therefore, the direct application of the recommendation algorithm based on collaborative filtering technology to the cloud computing environment will produce the following problems: Firstly, it is the problem of data sparsity. In the face of massive data, the general users' evaluation information only accounts for 1%~2% of the total number of goods. After two or more than two users' evaluation, the items are less. The evaluation matrix data is rather sparse, it is difficult to find the nearest neighbor of the user or item, and the recommendation precision is low. Secondly, it is extensibility. Collaborative filtering recommendation based on the need for global data in the whole space calculation, when the amount of data is small, can obtain better recommendation effect at the right time, but the system has millions of users or items, the algorithm will encounter serious scalability problems, real-time and accuracy of the problem the direct influence of recommendation system. Finally, it is the problem of hardware and equipment constraints. In the face of super large data sets, it is easy to be subject to CPU speed, storage capacity and even the constraints of network bandwidth.

To sum up, in order to solve the problems of low recommendation accuracy, high recommended delay and hardware device constraints, based on collaborative filtering technology applied to the cloud computing environment, this paper proposes an improved method. Firstly, the system can dynamically generate user forms to return the related records from different user tables to handle keyword queries. On the basis of this, we propose a user interest model construction method, which combines context search query with user implicit feedback to generate user personalized interest model, achieve immediate updating and accurately reflect user interest. The experimental results show that the method has good scalability and greatly improves the real-time performance and accuracy of the proposed system.

II. RELATED WORK

The information to determine the trusted service is an effective method for application of recommended open system, recommendation process can refer to consumer scores, recommended by the expert level and social level, but the existing recommendation system in cloud environment is the lack of adequate support of the user's preference, on the recommendation of the credibility of information has not been evaluated effectively.

In [2], for instance, a cloud service selection middleware is designed with multi attribute utility theory, and the service recommendation is realized by solving the multi-objective decision problem. Also, in [3] presents an approach that addresses the sparsity problem using Bhattacharyya similarity. The approach utilizes the numerical values of all ratings made by a pair of users, not just the ratings submitted for common items. Thus, it combines local and global similarity to obtain the final similarity value between a pair of users. Experiments using MovieLens, Netflix, and Yahoo datasets showed that this approach outperformed other approaches in terms of the evaluation metrics used. In [4], the original cognitive network method is applied to solve the multi-objective decision-making problem of SaaS service product reliability, but its evaluation process depends on expert scoring, and it only applies to the procurement process of some larger software services. Wasid and Kant [5] applied fuzzy sets to represent the imprecise features of users. Also in [6], an incremental CF approach based on Mahalanobis distance was proposed. This approach builds clusters of similar users using an incremental algorithm based on the Mahalanobis radial basis function. The prediction phase employs fuzzy sets to determine the membership degree of each user to the formed clusters. The MovieLens dataset was used to show the effectiveness of the proposed approach. In [7], the authors presented an approach that deals with the sparsity and computational issues of CF systems by constructing nonredundant subspaces that represent each user. Then, a tree is built to represent these users' commonality with the active user, followed by a new recommendation method to predict unknown ratings. Experiments performed using the MovieLens and Jester datasets showed that the proposed approach was effective in most cases. Moreover, in [8], an interest intensity model that decayed over time was built to represent how users' preferences are correlated and change with time. A scoring matrix is then constructed that can alleviate problems with the sparse user-item scoring matrix. Also, in [9], a multi-criteria approach for representing imprecise aspects of user behavior with respect to items is introduced. The approach applies the Adaptive Neuro-Fuzzy Inference System combined with subtractive clustering and Higher Order Singular Value Decomposition to deal with imprecise and high-dimensional data. These experiments used the Yahoo! Movies dataset and showed improved prediction accuracy compared to similarity-based approaches. A reversed CF was also presented in [10]. This approach utilized a k-nearest neighbor (k-NN) graph, which, instead of finding k similar neighbors of unrated items, finds the k-nearest neighbors of rated items. Experimental results using the MovieLens dataset show that this approach outperforms traditional user-based/item-based CF algorithms in terms of preprocessing and query processing time without sacrificing the level of accuracy. Also, in [11], a multi-level recommendation method that modifies similarities between users using Pearson Correlation Similarity based on some constraints is proposed. The effectiveness of the presented approach is tested on three datasets.

III. SYETEM STRUCTURE

Cloud computing is a new term begun in 2007 has been mentioned. Wikipedia in 2010 gives the new definition: cloud computing refers to a kind of dynamic, easy to expand, is calculated to provide virtualized resources via the Internet, users do not have to know the internal details of the cloud, also need not have the expertise or direct control infrastructure. Cloud computing, including infrastructure, software, services and other platforms rely on the Internet to meet customer demand for computing technology trends. Cloud computing aims to integrate multiple network computing entity relatively low cost into a perfect system has powerful computing ability, and with the aid of SaaS (Software as a service), PaaS (Platform as a service), IaaS (Infrastructure as a Service), and other advanced business model of the distribution of the powerful computing capability to the terminal the hands of users. Aiming at the problems faced by traditional recommendation algorithms in cloud computing environment, this paper proposes a personalized recommendation mechanism based on collaborative filtering in cloud computing environment, which provides efficient and accurate recommendation services for all kinds of Internet applications in data centers. This section first introduces the collaborative filtering recommendation model in cloud computing environment, describes the interaction relationship between each module of recommendation mechanism, and then gives the related terms definition.

Personalized Recommendation System Based on Cloud Computing (PRC) can be divided into four layers according to the function, as shown in figure 1.

This system is mainly reflected in the intelligent personalized proactive services, mainly to complete the following functions:

The bottom layer of the cloud services layer for the infrastructure services layer is the infrastructure (computer resources, server resources, storage resources) as a service rental, can be a simple understanding of cloud devices. Through the Internet can be from a sound computer based equipment to obtain services, will be based on the integration of equipment. This means that the virtual computer not only has the ability to deal with fast, but also has the ability to store a stable.

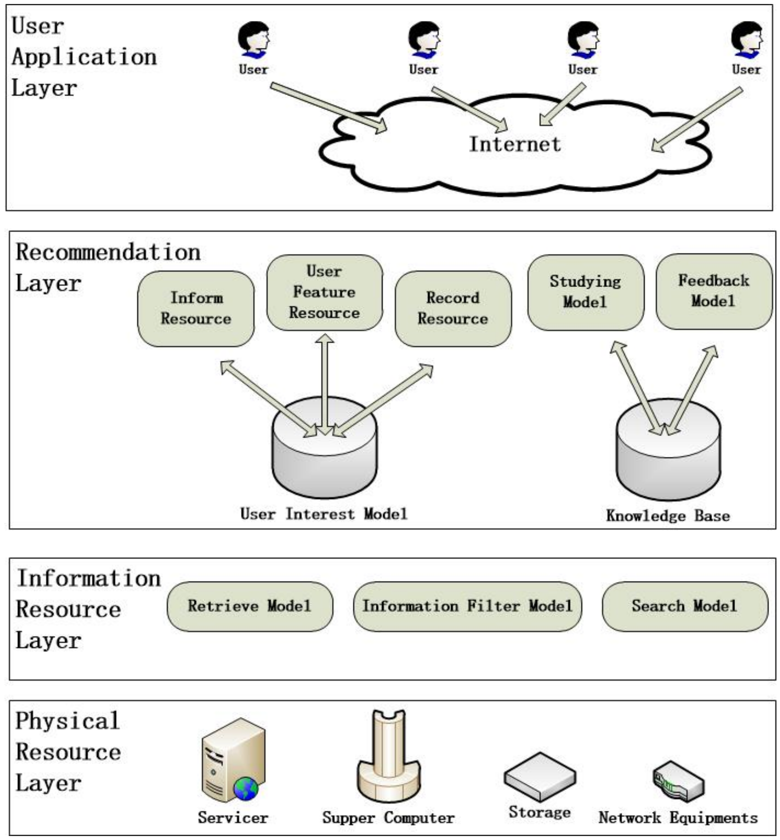

Fig. 1. The System Framework

The Information Resource layer is used to store obtained web pages information, dictionary, and training documents. The dictionary allows the model to compare and discover relationships between a pair of words or concepts. Each concept has a list of supporting documents or links. It is responsible for the mapping from information source base to knowledge base, and organizes the knowledge of knowledge management system. Search model begins searching and transmits the searching results to information filter Model. Its function is to realize personal and intelligent searching according to user's interest and requirements. Retrieval model obtains the necessary methods according to the different retrieval task, and it also may seek the services of a group of model that work cooperatively and synthesize the integrated result. Then it searches the corresponding knowledge from knowledge base. Finally it submits the results to user interaction model. Information filter model deals with the delivery of information that is relevant to the user in a timely manner. This model filters and selects web pages obtained by search model.

Recommendation layer usually include a method for classifying users into groups, and relating these groups to domain concepts, and a software user modeling component to be incorporated into an application. User modeling components generally include a component to store the user model, and an interface to write and read from that store. The write interface supports the acquisition of user models, and the read interface supports the querying of the user model for domain or application information. The store itself supports the heuristic classification either at write time or read time. Knowledge base is responsible for the mapping from information source base to knowledge base, and organizes the knowledge of knowledge management system.

Finally, the top of the cloud service layer for application services layer, it can be simply understood as the cloud software. It is a model for providing software through the Internet, through a large number of public or private service software or programs, by the user's free choice or combination, for its services. It is the most widely used cloud computing service layer, it provides the most close to the user's service. It breaks the original software development, installation and use patterns, according to personal preferences and business needs to be customized to provide a wide range of software services.

IV. PERSONALIZED RECOMMENDATION MECHANISM OF CLOUD COMPUTING

A. *User Modeling*

The user's personalized interest model is composed of a webpage set browsed by the user, a page classification set, a page preprocessing module, a user interest generating set and an interest updating module.

(1) Page Preprocessing: Each page browsed by the user needs to be preprocessed, and the corresponding parser is used to generate a text file.

(2) Page Category: A single page cannot fully reflect the user's interest, but after the corresponding classification, but

can reflect the user's interest. Therefore, page classification is crucial for getting the current user's interests.

(3) Part of speech tagging: to determine a word is an adjective, noun or verb, and then noun extraction.

(4) Weight calculation: calculate the weight value of each noun through the weight calculation formula.

(5) Generation of interest: The user interest file is generated by combining the context of the query term and the implicit feedback feature of the user, and the user's personalized interest model is updated instantly according to the improved updating mechanism.

B. Page Classification

The page features represent Boolean logic model, probabilistic model, vector space model, etc. In this paper, the cosine based vector space model is used to classify the page sets browsed by users, in which the user browsed page features extraction Methods as below:

(1) Convert the pages browsed by the user to text form, and keep some important information.

(2) Cut the text files accordingly.

(3) At the same time remove the browser page has nothing to do with the content of some function words.

(4) Remove some low-frequency words accordingly.

(5) The remaining words as the user's characteristic words, and calculate the frequency of their appearance.

The formula is as follows:

$$sim(p, u_c) = \frac{\sum_{i=1}^{n} ap\,(i) u_c(i)}{\sqrt{\sum_{i=1}^{n} ap\,(i)^2 \sum_{i=1}^{n} au_c(i)^2}} \qquad (1)$$

C. User Interest Weight Calculation

User interest weight is the degree to which a user is interested in a particular object. Weighting the keywords is to confirm the most important keywords in the webpage. The number of keywords extracted from a single webpage is strictly limited. Therefore, any page with only the top N keyword ranking is conducive to the construction of personalized user interest model. The degree of web page interest refers to the degree of relevance between the web document set and the user's interest or query term, and the source can be obtained by observing the browsing behavior in the relevant webpage of the user.

In a period of time T during which a user browses a webpage set, assuming that a total of M web pages are browsed, denoted as pi, i=1,2,…,M. if the user views the same web page multiple times and is recorded as DTi in the T time period, then the number of times to visit m different resources is respectively Ni, i=1,2,…,m, then there is a formula:

$$M = \sum_{i=1}^{m} N_i \qquad (2)$$

If a user repeatedly visits a webpage to indicate that the user is interested in the information of a webpage, the characteristic word extracted from the webpage may be used as the user's favorite item. Otherwise, the user's viewing behavior *PageDegree*(DT_i) can be considered from the browsing behavior of the user (for example, when the page is resided, whether it is saved, printed, or collected). *ScanTotal*(DT_i) is the total time that the page resides, *Save*(DT_i) is whether to save the page, *Print*(DT_i) is whether to print the page, and *Collect*(DT_i) is whether to collect the page.

$$PageDegree(DT_i) = (ScanTotal(DT_i), Save(DT_i), Print(DT_i), Collect(DT_i)) \qquad (3)$$

$$ScanTotal(DT_i) = \sum_{i=0}^{num} Scan(DT_i) \qquad (4)$$

A comprehensive analysis of the browsing behavior of the above-mentioned users can find that if the user saves, prints, and stores the webpage, it can indicate that the user is more interested in the webpage and does not need to analyze it again. The user's interest weight is:

$$W_i = wt_i * PageDegree(DT_i) \qquad (5)$$

D. User Interest Model Update

The update system of the user interest model refers to the continuous interaction between the user and the user, and the continuous learning process. As users 'interests change over time, users' interests need to be divided into short-term interests and long-term interests. At the same time in order to adapt to changes in user interest at any time, the user interest model needs to be updated immediately, real-time supplement. The update algorithm is as follows:

(1) Forgetting the key query terms in the user's personalized interest model system to adapt to changes in user interest at any time, the formula is as follows:

$$F(x) = e^{-\frac{log\,2}{hl}(cur - est)} \qquad (6)$$

(2) If the information comes from the query submitted by the user, the query is extracted to obtain the set of interest words ($q_1, q_2, ..., q_n$)

(3) If the information is from the document set browsed by the user, then the document is processed by word segmentation to obtain the set of interest words ($r_1, r_2, ..., r_n$), and the corresponding term weights ($d_1, d_2, ..., d_n$) are calculated.

(4) Determine whether the terms of interest q_i ($1 \le i \le n$) and r_i ($1 \le i \le n$) satisfy the conditions in the set I_m of interest words of the user, and increase the weights of the user interest words q_i and r_i in the interest set value. Do not meet the conditions, the interest words q_i and r_i directly into the user interest set I_m.

(5) Sort all the interest words by weight, and eliminate interest words with smaller weights to form a new user interest model.

We should get the complete user interest from all aspects of user. After the study, we found that in order to gain a more complete reaction of a user's interest, we need analysis by the page content of user browse and user behavior. Different from the search engine, the personalized recommendation system needs to rely on the user's behavior data, recommend the system to obtain the log and table data, analyze the user's behavior data, excavate and calculate, get the recommended item list, and record to the database, and finally display the page on. Therefore, in this paper, the user's interest is calculated from these two aspects of the data. The formula is:

$$w'_{ij} = \left(\frac{w'_{ij}}{\sum_{m=1}^{|v_i|} w'_{im}} + 1\right) \times BI_i \qquad (7)$$

BI_i expressed interest degree of user behavior on the page. Thus, it provides a more complete user data for the user interest categories on the future. The user's behavior data is mainly stored in the preference behavior table Preference; the last recommended result is stored in the recommendation table Recommend.

V. EXPERIMENT EVALUATION

User behavior data is mainly stored in the preference behavior table Preference, the table is mainly three fields: UserID, ItemID, PreferenceValue. PreferenceValue comes from the user's implicit feedback and explicit feedback. The implicit feedback behavior includes finishing, collecting, downloading. The dominant feedback behavior is mainly scoring, and the system uses a two-level scoring mechanism.

The last recommended result is stored in the Recommend Model. When the recommended list needs to be displayed, it is sorted according to the recommended value, and the first N are taken to make the TopN recommendation.

We apply the method described to computer optimal weights at different instants and for several values of the parameter p. Results are summarized in figure 2, figure 3 and figure 4.

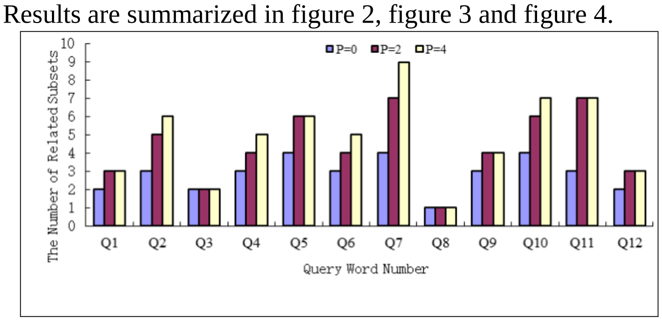

Fig. 2. Relevant answer sub-tables for each query on three values of parameter *p*

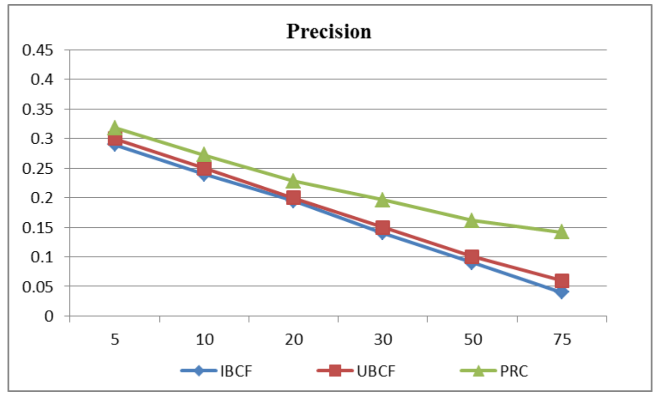

Fig. 3. The Average Precision

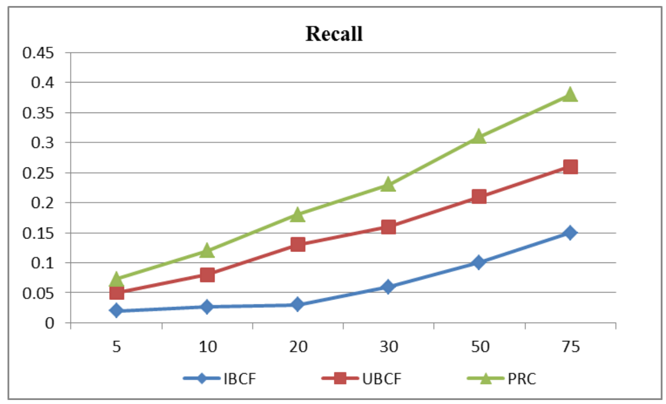

Fig. 4. The Average Recall

VI. CONCLUSIONS

The establishment and updating of user interest model is the foundation and key of personalized service, which is directly related to the effect of personalized recommendation. Based on the widely popular and powerful e-commerce platform, this paper takes full advantage of the computing and data processing capabilities provided by cloud computing. It proposes to use cloud computing-based user interest model representation and update mechanism, The initial interest preferences to establish interest model, the use of user feedback on the results of the recommendation to achieve the interest model automatically updates. The effectiveness and feasibility of this method are demonstrated through experimental simulation analysis. Under complex cloud computing environment, the proposed method can support the preference of users and achieve high-quality personalized recommendation, which has high application value.

REFERENCES

[1] Tunkelang D, "Recommendations as a conversation with the user", in Proc. Conf. On Recommender System, pp. 11–12, 2011.

[2] Silas S, Rajsingh E B, Ezrz K., "Efficient service selection middleware using ELECTRE methodology for cloud environments", Information Technology Journal, vol. 11, no.7, pp. 868-875, 2012.

[3] Patra BK, Launonen R, Ollikainen V, Nandi S, "A new similarity measure using Bhattacharyya coefficient for collaborative filtering in

sparse data", Knowledge-Based Systems, vol. 82, no.5, pp.163-177, 2015.

[4] Yuen K K F, "Software are as a service evaluation in cloud paradigm: primitive cognitive network process approach", in Proc. Conf. On Signal Processing, Communications and Computing, pp.119-124, 2012.

[5] Wasid M, Kant V, "A Particle Swarm Approach to Collaborative Filtering based Recommender Systems through Fuzzy Features", Procedia Computer Science, vol. 54, no.8, pp.40-48, 2015.

[6] Komkhao M, Lu J, Li Z, Halang WA, "Incremental collaborative filtering based on Mahalanobis distance and fuzzy membership for recommender systems", International Journal of General Systems, vol.42, no.1, pp.41-66, 2013.

[7] Ramezani M, Moradi P, Akhlaghian F, "A pattern mining approach to enhance the accuracy of collaborative filtering in sparse data domains", Physica A: Statistical Mechanics and its Applications, vol.40, no.8, pp.72-84, 2014.

[8] Liu Z, "Collaborative Filtering Recommendation Algorithm Based on User Interests", International Journal of U-& E-Service, Science & Technology, vlo.8, no.4, pp.67-81, 2015.

[9] Nilashi M, Bin Ibrahim O, Ithnin N, "Multi-criteria collaborative filtering with high accuracy using higher order singular value decomposition and Neuro-Fuzzy system", Knowledge-Based Systems, vol.60, no.4, pp.82-101, 2014.

[10] Park Y, Park S, Jung W, Lee S-g, "Reversed CF: A fast collaborative filtering algorithm using a k-nearest neighbor graph", Expert Systems with Applications, vol. 42, no.8, pp.4022-4028, 2015.

[11] Polatidis, Christos K, Nikolaos, "A dynamic multi-level collaborative filtering method for improved recommendations", Computer Standards & Interfaces, vol. 51, no.3, pp.14-21, 2017.

Weiwei Jiao was born on Sep. 23, 1981**.** She received the master's degree in computer science and technology from Northeast Normal University of China. Currently, she is a researcher at Liaoning University of Technology, China. Her major research interests include information retrieval and personalized service.

Xin Li was born on Oct. 18, 1966**.** He received the PhD degree in computer science and technology from Liaoning University of Engineering and Technology. Currently, he is a professor at Liaoning University of Technology, China. His major research interests include information retrieval and Database query.

Jingji Li was born on May. 30, 1980**.** He received the Bachelor's degree in Economic and administrative management from Chinese people's Liberation Army Nanjing Political College. Currently, he is a researcher at State Grid Liaoning Electric Power Corp Jinzhou Branch, China. His major research interests include information security.