Collaborative Filtering Method of Web Service Recommendation Based on Location Aware and Personalized

Li Ziman

Henan Logistics Vocational College,ZhengZhou,Henan,China

Abstract

With the rapid growth of the number of Web services, it is necessary to build an efficient web service recommendation system in the face of massive web services. In order to recommend high-quality services to users, the key problem is how to obtain the s value of Web services. This paper proposes a collaborative web service recommendation method based on location clustering. Firstly, users are clustered according to the autonomous system by using the correlation between QoS and user location. According to the clustering results, the system fills in the vacancy Qos value; Then, the vacancy Qos value is filled in in advance and the similarity between active users and each user is calculated. Based on this, to P-K algorithm is used to obtain the most similar Qos value to predict the unknown service for active users to complete the recommendation. The method proposed in this paper can effectively solve the problem of data sparsity and cold start of Web services. At the same time, a better balance between accuracy and coverage is obtained.

Keywords: Collaborative Filtering, Data Fusion, QoS, P-K algorithm

I. Introduction

In recent years, with the rapid rise of e-commerce, web applications have developed from localization to globalization, from B2C (business to customer) to B2B (business to business), from centralized to distributed [1-2]. As a new web application mode, web service is a new distributed computing model and an effective mechanism for data and information integration on the web. From the perspective of e-commerce application field, complex application connection and program code cause the cost of e-commerce application maintenance and update, and web services can solve this problem, and become the most reasonable solution in the current application environment [3]. Therefore, web service has become a hot spot and attracted the attention of industry and academia.

With the development and popularization of information superhighway, people have been surrounded in the vast ocean of information. Internet is the source of massive information, and the organization of its information is heterogeneous, pluralistic and distributed [4-5]Due to the continuous updating and increasing of information, the amount of information increases and expands rapidly with exponential law, thus forming the "information explosion". In addition, the vast amount of information on the Internet is not only far beyond the scope of people's browsing, but also intermingled [6]. This seriously affects the efficiency and effect of information utilization, and also consumes a lot of users' valuable time and energy.

With the continuous development of web technology and the maturity of technology, the types and number of Web services on the network are growing at a very fast speed, and different service providers provide a large number of services. For service users, they not only can't understand all the services at a glance, but also can't directly check the service quality. Therefore, service selection is a difficult problem for service users.

In view of the increasing variety and quantity of Web services on the network, this paper studies the web service recommendation technology, mainly aiming at the personalized we service recommendation system and the personalized recommendation combined with the QoS geographical location aware nature of Web services.

II. Introduction of basic theory of Web service recommendation based on QoS

2.1 Web Service Technology

At present, there are many organizations and institutions for the research of Web services, and the research of Web services technology is also quite mature. However, different organizations have different understanding and understanding of the definition of Web services. This article mainly lists the definition of Web services by IBM and other companies [7].

IBM's definition: Web service is a set of operations described by XML that can be used for message passing through XML. These operations can be accessed by the network to complete the target task. The service description provides all the details necessary to interact with the service, including message format, transport protocol, and location.
Microsoft definition: Web service can be defined as an application logic unit, which provides services and data to other applications. Applications can access and call web services through XML, HTTP, OAP and other ubiquitous data formats and web protocols. In the process of accessing, we don't need to care about the specific implementation of the web service, just need to know the function of the web service.

Definition of Sun: Web service is such a software component, which has the characteristics of being discovered, recombined and reusable, which is convenient for users to solve the needs of users.

W3C definition: Web service is a software system that supports the operation of different machines on the network. Services on the network are usually composed of many application interface (API), which execute the service requests submitted by users through the network, such as the remote server side on the Internet.

The main architecture of Web services is SOA (Service Oriented Architecture) model, as shown in Figure 1. Service Oriented Architecture (SOA) is such a component model, which connects different functional units (also known as services) of an application through well-defined protocols and interfaces between these functional units (services). The interface is defined in a neutral way, which is completely independent of the hardware platform, programming language and operating system. This feature enables services in various systems to interact with each other in a common and unified way. SOA architecture of Web services mainly includes three kinds of participants and three basic operations. As can be seen from Figure 1, these three types of participants are: service provider, service client and service registry; The three basic operations are: publish, find and bind.
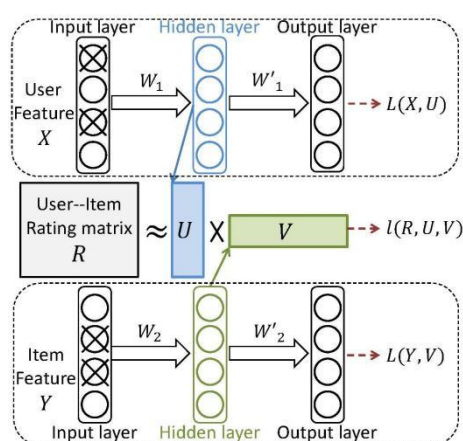
Fig 1: Web Services Architecture

2.2 Research on Recommender System

In 1993, R. Agrawal and others first proposed the problem of mining association rules between item sets in customer transaction data. The core idea is a recursive algorithm based on the idea of two-stage frequent sets. The basic purpose of association rules mining is to find the correlation in the sales records, so as to better guide the formulation of sales strategy. A typical rule is: "43% of customers who buy Nestle instant coffee will buy Nestle coffee partner.". Based on this rule, in the physical supermarket, the two products should be put in the same place. In the online supermarket, if customers buy Nestle instant coffee but do not buy coffee partner, they can add the corresponding recommendation in the column of related commodities. Now many enterprises have realized the importance of detailed original purchase records, and established a standardized data warehouse, which has laid a good foundation for the application of association rule mining technology.

Content filtering mainly uses natural language processing, artificial intelligence, probability statistics and machine learning techniques to filter. The items or objects are defined by the attributes of related features. The system learns users' interests based on the features of users' evaluation objects, makes recommendations according to the matching degree between users' data and the items to be predicted, and tries to recommend products similar to their previous favorite products to customers. Such as newsgroup filtering system News Weeder. The system based on content filtering is simple and effective. Its disadvantage is that the ability of feature extraction is limited and excessively detailed, and the content-based recommendation system can not find new interesting resources for customers, but can only find resources similar to those that customers are already interested in. This method is usually limited to the recommendation of products whose content is easy to analyze, but for some products whose content is difficult to extract, such as music CD, movies, etc., it cannot produce satisfactory recommendation effect.

Collaborative filtering algorithm is rapidly becoming a popular technology in information filtering and information system. Different from the traditional content-based filtering, collaborative filtering analyzes the user's interest, finds the similar (interest) users of the specified user in the user group, synthesizes the evaluation of these similar users on a certain information, and forms the system's preference degree prediction of the specified user on this information. Compared with traditional text filtering, collaborative filtering has the following advantages:

(1) It can filter the information that is difficult to do automatic content-based analysis. Such as art, music;

(2) Be able to filter based on some complex and difficult concepts (information quality and grade);

(3) The novelty of recommendation. Because of this, collaborative filtering has achieved good results in commercial applications. Amazon, cdnow and moviefinder all adopt collaborative filtering technology to improve service quality.
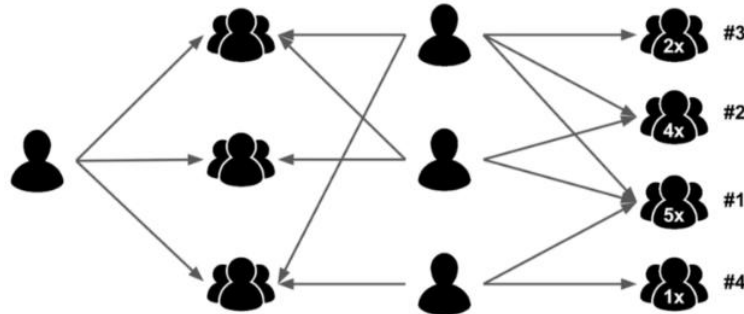
2.3 Research on Web Service personalized recommendation system

1) User based collaborative filtering recommendation algorithm

The basic principle of collaborative filtering recommendation based on users is to find "neighbor" user groups similar to current users' tastes and preferences according to all users' preferences for items or information. In general applications, the algorithm of computing "K-neighbor" is used. Then, based on the historical preference information of the K neighbors, the recommendation is made for the current user. The schematic diagram is shown in Figure 2.

Figure 2 illustrates the basic principle of collaborative filtering recommendation mechanism based on users, assuming that user A likes item A, item C, user B likes item B, and user C likes item A, item C, and item D. From these users' historical preference information, this section can find that user A and user C have similar tastes and preferences, and user C also likes item D. so this section can infer that user A may also like item D, so item D can be recommended to user A.

User based collaborative filtering recommendation mechanism and demography based recommendation mechanism both calculate the similarity of users, and calculate the recommendation based on "neighbor" user group, but the difference between them is how to calculate the similarity of users. Demography based recommendation mechanism only considers the characteristics of users. The collaborative filtering mechanism based on users can calculate the similarity of users on the data of users' historical preferences. Its basic assumption is that users who like similar items may have the same or similar tastes and preferences.
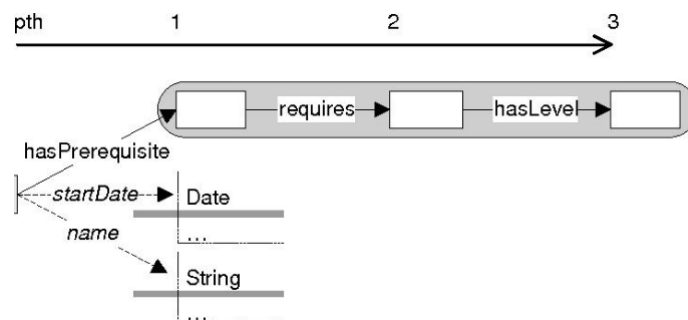

Fig 2: Basic principle of user based collaborative filtering recommendation mechanism

2) Recommendation algorithm of collaborative filtering based on Project

The basic principle of item based collaborative filtering recommendation is similar, except that it uses all users' preferences for items or information to find the similarity between items, and then recommends similar items to users according to users' historical preference information. Figure 3 illustrates its basic principle.

Suppose user A likes item A and item C, user B likes item A, item B and item C, and user C likes item A. From these users' historical preferences, we can find that item A and item C are similar. People who like item A like item C. based on this data, we can infer that user C is likely to like item C, so the system will recommend item C to user C.

Fig 3: The basic principle of collaborative filtering recommendation mechanism based on Project

Similar to the above, both item based collaborative filtering recommendation and content-based recommendation are based on item similarity prediction recommendation, but the similarity calculation method is different. The former is based on the user's historical preferences, while the latter is based on the attribute information of the item itself.

3) Recommended system evaluation criteria

(1)Absolute average error and standardized absolute average error

Absolute mean error (MAE) is widely used to measure the prediction quality of collaborative filtering method, and its definition is shown in formula (1) [8-10]:

$$MAE = \frac{\sum_{i,j} \left| r_{i,j} - \hat{r}_{i,j} \right|}{N} \quad (1)$$

In which $r_{i,j}$ represents the expected QoS value of service user i for service j, $r_{i,j}$ represents the predicted QoS value, and n represents the number of predicted values. because different QoS attributes of Web services have different value ranges, similar to the literature, this paper uses standardized absolute mean error (NMAE) to measure the prediction accuracy of the method in this section, and the definition of NMAE is shown in formula (2):

$$NMAE = \frac{MAE}{\sum_{i,j} r_{i,j} / N} \quad (2)$$

The smaller the standardized absolute mean value is, the smaller the prediction error is, the higher the prediction quality is, and vice versa.

(2) Algorithm execution time
The execution time of the algorithm should be measured by the time that the program based on the algorithm runs on the computer. The execution time of the algorithm directly reflects the efficiency of the algorithm, and is an important evaluation index of the algorithm.

III. Personalized web service recommendation method based on QoS location aware

3.1 QoS location aware personalized web service recommendation framework

For any active user, the recommendation system records his historical QoS records, and the geographic location information of the active user (the autonomous system to which the active user belongs can be determined according to the IP address) can also be obtained. QoS location aware personalized web service recommendation framework mainly includes the following modules: (1) service location information processing module: clustering all services of the website according to the country or autonomous system; (2) user location information processing module: clustering all users of the website according to the country or autonomous system. (3) User service matrix: record the historical QoS record of each user calling the service. (4) Search for similar user module: calculate the similarity between users according to the results of user location information processing module. (5) Search for similar service module: calculate the similarity between services according to the result of service location information processing module. (6) User-based prediction module: according to the result of searching similar user module, the predicted QoS value of active users to target service is calculated by using user-based collaborative filtering algorithm. (7) Service-based prediction module: according to the results obtained by searching for similar service modules, the service-based collaborative filtering algorithm is adopted to calculate the predicted QoS value of active users to target services. (8) Joint prediction module: Combine the results of the user-based prediction module and the service-based prediction module to determine the final prediction value of the target service by the active users. (9) Recommendation module: sort all services according to the QoS value predicted by the joint prediction module, recommend services to active users, and complete the recommendation.
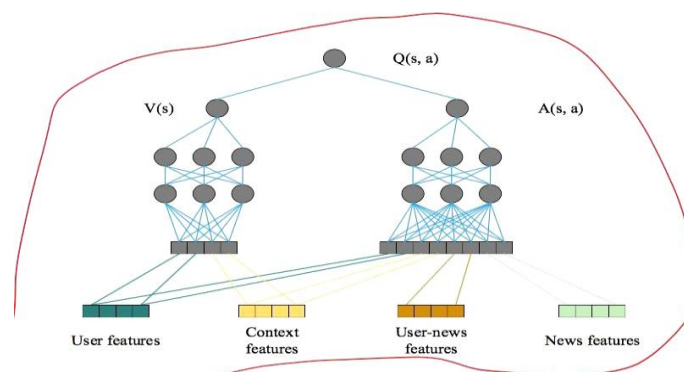
Fig 4: QoS location aware personalized web service recommendation framework

3.2 Personalized web service recommendation algorithm based on QoS location aware

From the structure and composition of the Internet, we can know that the Internet is composed of many Autonomous Systems. In the Internet, an autonomous system (AS) is a small unit that has the right to decide which routing protocol should be adopted in its own system. This network unit can be a simple network or a network group controlled by one or more ordinary network administrators. It is a single manageable network unit (such as a university, an enterprise or a company individual). An autonomous system is sometimes referred to as a routing domain. An autonomous system will be assigned a globally unique 16-digit number, which is sometimes called Autonomous System Number (ASN) in this section.

In this section, users and services are mainly classified into different AS and different countries according to their IP addresses, and the decimal representation of IP addresses can be calculated by using Formula (3). Generally, IP address is represented by A,B,C,D. the conversion from a.b.c.d. form to decimal number form of IP address is realized by the formula:

$$IP(A.B.C.D) = (((A \times 256 + B) \times 256) + C) \times 256 + D \quad (3)$$

By calculating the decimal system of IP address, the IP addresses of users and services can be mapped to get the AS and country where users and services are located. At present, there are such Internet measurement projects that publish mapping information from IP addresses or prefixes of IP addresses to autonomous systems AS and countries for free. For example, in the RouteViews project, data sets can be downloaded freely from the Internet. Route Views analyzes the routing information collected from BGP routers and the network topology diagram at AS level of Internet autonomous system, so AS to obtain the mapping of IP address or IP prefix to corresponding AS number and country.

3.3 Personalized QoS prediction based on regional model

Assuming that the active user is ui, the user-based collaborative filtering algorithm adopts the following steps to predict the QoS missing value of service sj for him:

Step 1: According to the IP address of the active user ui, convert the IP address into decimal representation, and calculate the autonomous system AS and the country where the active user is located by mapping with AS and countries.

Step 2: Find the users in the same AS as the active user ui, calculate the Pearson similarity between the active user and each user by using formula (4), and if the number of similar users in the same autonomous system is greater than topk, execute step 6.

$$Sim(a,u) = \frac{\sum_{i \in I} \frac{|I| di^{\lambda_1}}{\sum_{i \in I} di^{\lambda_1}} (r_{a,i} - \overline{r_a})(r_{u,i} - \overline{r_u})}{\sqrt{\sum_{i \in I}(r_{a,i} - \overline{r_a})^2} \sqrt{\sum_{i \in I}(r_{u,i} - \overline{r_u})^2}} \quad (4)$$

If $\lambda_1 = 0$, this situation does not consider the personalized characteristics of the service, which is equivalent to the traditional user-based method to calculate the user similarity. Because the personalization degree of services in different data sets is different, the parameter λ_1 makes the personalization method more suitable for different environments.

Step 3: Find the users in the same country as the active user ui, calculate Pearson similarity between the active user and each user by using formula (4), and if the number of similar users in the same country is more than topk, execute step 6.

Step 4: Find users in the same country as the active user ui globally, and calculate Pearson similarity between the active user and each user by using formula (4.2). If the number of similar users in the global scope is greater than topk, then step 6 is executed.

Step 5: If the service sj has been called before, the average QoS of all users to the service sj is used as the predicted value, otherwise, the prediction fails.

Step 6: Use formula (5) to predict the missing value:

$$P_u\left(r_{u,i}\right) = \bar{u} + \frac{\sum_{u_a \in S(u)} Sim'\left(u_a, u\right)\left(r_{u_a, i} - \bar{u}_a\right)}{\sum_{u_a \in S(u)} Sim'\left(u_a, u\right)} \quad (5)$$

IV. Conclusion

With the rapid development of Internet and the maturity of web service technology, there are more and more applications of Web services on the network. A large number of Web services with the same or similar functions appear on the network, resulting in the problem of information overload. At the same time, users are faced with the problem of service selection, and the competition of service providers has become very fierce. A well-designed recommendation and selection algorithm can effectively solve the problems of service users, strengthen the loyalty and trust of service users to the website, and improve the competitiveness of service providers. Collaborative filtering algorithms have been applied to web service recommendation. Most of these methods are based on the QoS records of users with similar historical QoS experience on some web services. However, these methods have defects in understanding the characteristics of QoS. According to the observation of this section, QoS is recognized as the attribute of user perception, which is highly related to the user's geographical location. According to the real Web Service QoS data set, this section verifies the geographic location aware performance of QoS.

References

[1] Hwong, N. C., Caswell, A., Johnson, D. W.,&Johnson, R. T. . Effects of cooperative and individualistic learning on prospective elementary teachers' music achievement and attitudes. Journal of Social Psychology, vol.133, no. 1, pp. 53, 1993.

[2] Lyster, &Norman, C., The use of the music operating system to supplement the teaching of cobol. ACM Sigcse Bulletin, vol.18, no. 4, pp.46-49, 1986.

[3] Oliveira, & A. Music teaching as culture: introducing the pontes approach. International Journal of Music Education, vol.23, no. 3, pp. 205-216, 2005.

[4] Ilhan Özgül. An analysis of the elementary school music teaching course in turkey. International Journal of Music Education, vol.27, no. 2, pp. 116-127, 2009.

[5] Boulet, & Marie-Michele. Designing and developing an intelligent advisor system for transfer tasks in music. Computers & Education, vol.19, no. 4, pp. 341-357, 1992.

[6] Hodge, G. M., Jupp, J. J.,&Taylor, A. J. . Work stress, distress and burnout in music and mathematics teachers. British Journal of Educational Psychology, vol.64, no. 1, pp. 65-76, 2011.

[7] Yang, L., Ketner, K., Luker, S.,&Patterson, M. . A complete system for publishing music-related etds. Library Hi Tech, vol.34, no. 1, pp. 151-163, 2016.

[8] Maganioti, A.E., Chrissanthi, H.D. "Cointegration of Event-Related Potential (ERP) Signals in Experiments with Different Electromagnetic Field (EMF) Conditions". Health, vol.12, no.2, pp.400-406, 2016.

[9] Tianyi Qin, Drivers drowsiness detection in embedded system, IEEE International Conference on Vehicular Electronics and Safety, 2007.ICVES

[10] Yan Chen, Shunqing Zhang, Shugong Xu, G.Y. Li, "Fundamental tradeoffs on green wireless networks," Communications Magazine, IEEE , vol.49, no.6, pp.30,37, June2011.