ISSN: 2688-8645

A Survey of Web Service QoS Prediction

Tingting Zhang, Guihua Huang and Huitong Liao*

Guangdong University of science&technology, Dongguan, 523000, China

Abstract

With the increase in number of services, with the same or similar functions and different quality of services are more and more features. Many researchers have proposed service should not only consider selecting features and services, and services should be considered non-functional properties that is QoS (Quality of Service, QoS), such as response time, throughput and availability. Based on the study discusses the background and status of the service on the prediction, first introduced the tradition of service prediction algorithm. In addition, in-depth analysis of one of the core methods: collaborative filtering algorithms. Then, the disadvantage of this algorithm exists for, introduces two optimization algorithms, namely: network-based location regularization collaborative filtering algorithm based on clustering smooth scalable collaborative filtering algorithm. Finally, further prospects of further research in this field.

Keywords

Web Service; Quality of Services Prediction; Collaborative filtering.

1. Introduction

With the rapid increase of Web services, researchers in the field of service-oriented computing pay more and more attention to the learning of service quality. The research based on QoS is mainly applied in the fields of Web service selection, recommendation and combination. The most common assumption in these research fields is that the QoS attribute values of Web services are available and accurate. However, this premise is incorrect. Due to the large number of service developers, there are many services that provide similar or identical functions. How to make a reasonable choice among these services has always been one of the issues of widespread concern in the research community. Many researchers believe that when a service consumer When choosing a service, not only the functional requirements that the service can meet, but also the service quality requirements that the service can meet should also be considered.

On the basis of discussing the research background and current status of service prediction, this paper deeply analyzes one of the core methods of service quality prediction: collaborative filtering algorithm. Then, in view of the shortcomings of the algorithm, two optimization algorithms are introduced: a collaborative filtering algorithm based on network location regularization [1] and a scalable collaborative filtering algorithm based on clustering smoothing [2]. Among them, the collaborative filtering algorithm based on network location regularization introduces the user's network location information, determines a similar neighbor set based on the current active user's network location information, and then uses the regular term based on the similar neighbor set to modify the matrix decomposition model to be the current active user Predict missing values of QoS. The scalable collaborative filtering algorithm based on clustering and smoothing combines the advantages of memory-based and model-based, and uses clustering to calculate groups and smooth users who have not been rated. By using closely related user rating information in the group, predictions are made for unrated users in the group, which allows missing values to be filled in, which makes the system scalable. Table 1 is a simple example. The numbers in the table indicate the response time for

ISSN: 2688-8645

the corresponding user to call the corresponding service, and it means that the corresponding user has not called the corresponding service. In this article, we call it lost data. Since there is no data for the user u1 to call the service s3 in the table, it is difficult to select the service with the shortest response time for the user u1. Therefore, predicting missing data is very important for service selection.

Table 1. User-service matrix (response time)

Numble	S1	S2	S3
u1	0.4	1.6	null
u2	2.8	null	3.5

2. Research Status

Collaborative filtering is one of the most widely used and successful methods in the field of service QoS prediction. The basic idea of this method is to find the relationship between users (services) based on the existing records, and predict the QoS value of the target users calling the target service through the records of users (services) that are closely related to the target user (target service).

Rich et al. [3] first proposed the collaborative filtering algorithm, which is currently widely used in various commercial recommendation systems for personalized recommendation. Subsequently, Bress et al.proposed a user-based collaborative filtering algorithm-UPCC. Generally, Pearson Correlation Coefficient (PCC) is used to calculate the similarity between users, and then similar users are selected, and missing values are predicted to obtain a recommendation list. This traditional user-based collaborative filtering algorithm has some shortcomings: it overestimates the similarity of service users and ignores that the personalized characteristics of the service will affect the predicted value and the accuracy of the recommendation [4]. Resnick et al. [5] proposed an item-based collaborative filtering algorithm-IPCC. This method also has some shortcomings: it ignores the user's personalized characteristics. The Chinese University of Hong Kong Zheng et al. [6] proposed a Web service recommendation method WSRec, which uses a hybrid collaborative filtering method that combines user-based and item-based collaborative filtering. Chen et al. [7] of Beijing University of Aeronautics and Astronautics believed that QoS is related to the user's geographic location, and they proposed a collaborative filtering algorithm RegionKNN that integrates location information. In RegionKNN, all users are divided into regions based on the similarity of IP addresses, and then the historical QoS records and partition information are combined to improve the QoS prediction of the service. Literature [8] combines IP addresses with autonomous systems, and proposes a collaborative filtering method that considers location information to predict the QoS of Web services.

3. Traditional Collaborative Fitering Algorithm based on QoS Prediction

3.1. Memory-based Collaborative Filtering

Memory-Based Collaborative Filtering Algorithm: Find a set of similar neighbors through similarity calculation, and then predict the missing value based on the information of the set of similar neighbors. Typical memory-based collaborative filtering algorithms mainly include: user-based and item-based collaborative filtering algorithms.

3.1.1. User-based Collaborative Filtering

The user-based collaborative filtering algorithm [9] calculates the similarity between the active user and all other users, obtains the neighbor set similar to the current active user's hobbies,

ISSN: 2688-8645

and then predicts and recommends it based on the information of the similar user set. In the user-based collaborative filtering algorithm recommendation process, the most critical step is to calculate the similarity between different users and find a set of similar neighbors for active users. The calculation of similarity between different users mainly includes cosine similarity, Pearson correlation coefficient and modified cosine similarity.

3.1.2. Item-based Collaborative Filtering Algorithm

Item-based collaborative filtering algorithm [10] is currently the most widely used algorithm in the field of e-commerce. User-based collaborative filtering algorithm has been applied to a certain extent, but this algorithm has a big flaw. When the number of users increases, it becomes more and more difficult to calculate the similarity between different users, and its time complexity and space complexity increase in a quadratic order. In addition, if a user has not rated the item, the user-based collaborative filtering algorithm cannot find a set of similar users and cannot make recommendations. Based on these shortcomings, Amazon proposed an item-based collaborative filtering algorithm. The item-based collaborative filtering algorithm is also divided into three steps: calculate the similarity between different items, make predictions for active users based on the similarity set information of the target items, and finally determine the recommendation list to recommend to the user.

3.2. Model-based Collaborative Filtering Algorithm

The model-based collaborative filtering algorithm obtains a model by training and learning the user-item score matrix, and then predicts the missing value. The model-based collaborative filtering algorithm uses mathematical calculations, machine learning, and data mining methods to mine the potential relationship between users and items, build a predictive model based on the user's historical rating data, and then predict missing values, and finally recommend.

Model-based collaborative filtering algorithms mainly include matrix factorization, cluster analysis, factorization machines, Bayesian networks, etc. These models are applied in different application scenarios. The algorithm generally completes the construction of the model through offline calculation, and when the model construction is completed, it can quickly complete the recommendation for the user. However, when a new user or item is added to the recommendation system, the user-item matrix needs to be retrained and learned. Therefore, the model-based collaborative filtering algorithm is not suitable for online real-time recommendation. In addition, the model-based collaborative filtering algorithm has poor interpretability. It recommends based on the hidden correlation characteristics of users and items, and it is difficult to explain these users or items. What exactly does the implicit feature mean?

4. Improved Collaborative Filtering Algorithm

4.1. Collaborative Filtering Algorithm Based on Regularization of Network Location

The algorithm introduces the user's network location information, determines a similar neighbor set based on the current active user's network location information, and then uses the regular term based on the similar neighbor set to modify the matrix factorization model to predict the missing QoS value of the current active user. Figure 1 is a flowchart of a collaborative filtering algorithm based on network location regularization.

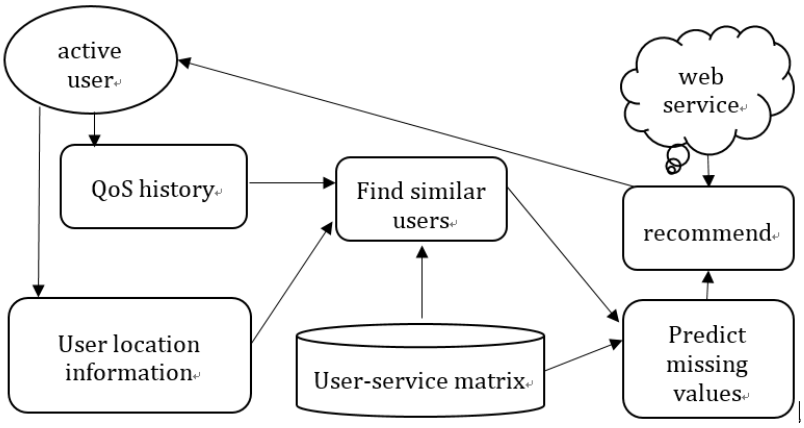

Figure 1. Flow chart based on regularization of network location

4.1.1. Improved Matrix Factorization Model

The matrix factorization model is a popular method that can effectively predict missing values. The core idea of the matrix factorization model is: to decompose the matrix R into two low-dimensional matrices, one is m × d representing the hidden feature matrix of users, and the other is n × d representing the hidden feature matrix of items (Web services). Then calculate the predicted value through the inner product.

Assuming that each user and Web service contains d hidden features, the matrix R is decomposed into two low-dimensional matrices U and S:

$$R \approx U \times S (U \in R_{d\times m}, S \in R_{d\times n}) \tag{1}$$

The singular value decomposition method approximates the original matrix R through continuous training and learning of two low-latitude matrices U and S, and approximates the original matrix R by reducing the square of the error between the predicted matrix and the original matrix, in order to avoid overfitting Phenomenon, adding the term to prevent over-fitting, the entire loss function is defined as follows:

$$L = \min \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}(R_{ij} - U_i^T S_j)^2 + \frac{\lambda_1}{2}\|U\|_F^2 + \frac{\lambda_2}{2}\|S\|_F^2 \tag{2}$$

In order to minimize the above objective function L, a stochastic gradient descent method can be used. The stochastic gradient descent method is applied to the objective function L, which is integrated with the traditional matrix factorization model, such as formula (3):

$$L_1 = \min_{U,S} \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}(R_{ij} - U_i^T S_j)^2 + \frac{\lambda_1}{2}\|U\|_F^2 + \frac{\lambda_2}{2}\|S\|_F^2 + \frac{\alpha_1}{2}\sum_{i=1}^{m}\|U_i - \frac{1}{|G(i)|}\sum_{g\in G(i)} U_g\| \tag{3}$$

The gradient descent method is used to calculate the local minimum. The same as the traditional matrix factorization model. Due to the intrinsic structure of the objective function, the global minimum cannot be reached. Improved model: The basic model puts different neighbors with equal importance. This model provides the neighbors with larger weights that are closer and the neighbors with smaller weights that are farther away. Therefore, we have:

$$L_2 = \min_{U,S} \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}(R_{ij} - U_i^T S_j)^2 + \frac{\lambda_1}{2}\| U \|_F^2 + \frac{\lambda_2}{2}\| S \|_F^2 + \frac{\alpha_2}{2}\sum_{i=1}^{m}\sum_{g\in G(i)} w(i,g)\| U_i - U_g \|_F^2 \quad (4)$$

Again, the gradient descent method searches for the local minimum L2.

4.1.2. Calculation of Similar Neighbors

Pearson correlation coefficient and vector space similarity are often used to calculate the similarity between different Web service users. In the study [17], it is shown that the PCC method takes into account the differences in the observed values of Web service users to obtain higher accuracy, so the PCC method can obtain higher performance than the VSS method. Therefore, this chapter uses PCC to calculate the similarity between different users. Based on the QoS value observed by user u and user v calling the common web service, the similarity between the two is calculated.

4.2. Scalable Collaborative Filtering Algorithm Based on Clustering Smoothin

We combined the advantages of memory-based and model-based to propose a new framework in order to be able to make recommendations for related groups. Our method uses clustering to calculate groups and smooth users who have not been rated. Use clustering to smoothly integrate the advantages of memory-based and model-based methods. Predictions are made for unrated users in the group by using information about the ratings of closely related users in the group, which allows missing values to be filled in. In addition, assuming that for active users, the nearest neighbors should also be the top-k most similar clusters, we only need to select the nearest neighbors among the top-k most similar clusters, which makes the system scalable.

4.2.1. Collaborative Filtering Framework Based on Clustering

We first define the symbols used throughout this article. Let $T = \{t_1, t_2, ..., t_m\}$ is the set of items, $U = \{u_1, u_2, ...u_n\}$ is the set of users, u_a is an active user-for them, our users need to provide recommendations for users who have not seen items before. $\{(u_{(1)}, i_{(1)}, r_{(1)}), ..., (u_{(k)}, i_{(k)}, r_{(k)})\}$ is a score in the training set. Each triple represents that the item is rated by the user. For each user u, the user u's rating of item t is, and the user's average rating is $\overline{R_u}$, and the rating ranges from 1 to $r_{\max}$. Our cluster-based smoothing algorithm is shown in Figure 1 and the framework is quite common, including memory-based and model-based methods. In this algorithm, the selection of a cluster () corresponds to the selection of a similar user group. The scoring step and the prediction step integrate smooth operations and suggestions.

4.2.2. Clustering Algorithm

There are many algorithms that can be used to create clusters. In this article, the K-means algorithm is chosen as the basic clustering algorithm. The number k is the expected number of specified clusters as the input to the algorithm. In the first pass, the algorithm takes the first user as the centroid of k unique clusters. Each remaining user is compared with the nearest centroid in turn. Recalculate the cluster centers on the basis of re-evaluation through cluster members before forming the cluster centers. This algorithm clusters the total number N of users whose running time of each pass is linear; that is, the calculation time is $O(k^2 N)$.

4.2.3. Data Smoothing

As we discussed above, data sparseness is a fundamental problem of collaborative filtering. In order to fill in the missing values in the data set, we explicitly use clustering as a smoothing mechanism. Natural language processing clusters are successfully used based on smoothing technology [18] to estimate the probability of which topic (cluster) a word belongs to. It helps

us solve the sparsity problem of collaborative filtering. Based on the results of the clustering, we use a smoothing strategy to smooth the data without scores.

4.2.4. Neighbor Selection and Prediction

An important step of the collaborative filtering algorithm is to search for the neighbors of active users. By adopting the concept of cluster, the function of using user groups in the cluster is represented by the center of mass of the cluster. This centroid represents the average rating of all users in the cluster. After calculating the similarity between the groups and active users, we select the group with the most similar users as candidates. From this process, clustering can help speed up the calculation of similarity and delete some irrelevant information. After the pre-selection is over, we also need to recalculate the similarity between the users in the candidate set and the smoothly rated active users. After the cluster information is smoothed, the user's score contains two parts: the original score and the group score. In this paper, the user's original score and the group's score are given different weights when calculating the similarity between the candidate ensemble user and the active user. When calculating similarity, we adjust the weights of different levels by assigning different parameter values. Finally, for prediction, a subset of k most similar users is the selection of active users based on similarity, and the weighted total score is used to generate predicted active users.

5. Conclusion

With the rapid development of web service-related technologies, there are more and more web service applications on the Internet. Different service providers provide service users with a large number of web services with the same or similar functions on the Internet. This is a problem of information overload. Service users need to spend a lot of time and energy to choose a Web service that meets their needs. Even if a user finds a Web service that meets their needs, the Web service provider only provides a functional description of the service. QoS is obtained directly, so the user may not call the web service with the best QoS. At this time, service users are faced with a difficult problem of service selection. QoS is often used to describe and evaluate the non-functional attributes of Web services, so this article focuses on the prediction methods based on QoS missing values in the field of Web service recommendation.

Project source: Guangdong University of Science and Technology school-level scientific research project, project number: GKY-2019KYQN-15、GKY-2019KYYB-37

References

[1] He P, Zhu J, Xu J, et al. A Hierarchical Matrix Factorization Approach for Location-Based Web Service QoS Prediction[C]//Service Oriented System Engineering (SOSE), 2014 IEEE 8th International Symposium on. IEEE, 2014: 290-295.

[2] Xue G R, Lin C, Yang Q, et al. Scalable collaborative filtering using cluster-based smoothing[C]//Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2005: 114-121.

[3] Rich E.User modeling via stereotypes.Cognitive Service,Vol.3,No.4,1979.

[4] M.R.McLaughlin and J.L.Herlocker. A collaborative filtering algorithm and evaluation metric that accurately model the user experience. In SIGIR , 2004.

[5] P.Resnick, N.Iacovou, M.Suchak, P.Bergstorm, and J.Riedl. Grouplens: An open architecture for collaboration filtering of netnews. In Proc. of ACM Confidence on Computer Supported Cooperative work,1994.

[6] Z.Zheng, Hao Ma, Michael R.Lyu et. al. WSRec: A Collaborative Filtering Based Web Service Recommender System[C]. Proc. 7th International Conference on Web Sercices (ICWS)2009,pp. 437-444,2009.

[7] Chen X, Liu X D, Huang Z C ,et al. RegionKNN: A Scalable Hybrid Collaborative Filtering Algorithm for Personalized Web Service Recommendation[C].Proc. 8th International Conference on Web Services(ICWS 2010),pp.9-16,2010.

[8] M.D.Tang, Y.C.Jiang, J.X.Liu. QoS Prediction Method for Web Services Based on User Location Awareness [J]. Small Microcomputer System, 2012.12(33):2664-2668.

[9] Resnick P, Iacovou N, Suchak M, et al. GroupLens: an open architecture for collaborative filtering of netnews[C]//Proceedings of the 1994 ACM conference on Computer supported cooperative work. ACM, 1994: 175-186.

[10] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms[C]//Proceedings of the 10th international conference on World Wide Web. ACM, 2001: 285-295.