# Neighborhood-based Uncertain QoS Prediction of Web Services via Matrix Factorization

Guobing Zou[1], Shengye Pang[1], Pengwei Wang[2], Huaikou Miao[1,3], Sen Niu[4], Yanglan Gan[2(✉)], and Bofeng Zhang[1(✉)]

[1] School of Computer Engineering and Science, Shanghai University, China
{guobingzou, shengyepang}@gmail.com, bfzhang@shu.edu.cn
[2] School of Computer Science and Technology, Donghua University, China
{wangpengwei, ylgan}@dhu.edu.cn
[3] Shanghai Key Laboratory of Computer Software Evaluating and Testing, China
[4] School of Computer and Information Engineering, Shanghai Polytechnic University
sens306314@gmail.com

**Abstract.** With the rapidly overwhelming number of services on the internet, QoS-based web service recommendation has become an urgent demand on service-oriented applications. Since there are a large number of missing QoS values in the user historical invocation records, accurately predicting these missing QoS values becomes a hot research issue. However, most existing service QoS prediction research assumes that the transactional process of the service was stable, and its QoS doesn't change as time goes. In fact, service invocation process is usually affected by many factors (e.g., geographical location, network environment), leading to service invocations with QoS uncertainty. Therefore, QoS prediction based on traditional methods can not exactly adapt to the scenarios in real-world applications. To solve the issue, combined with the collaborative filtering and matrix factorization theory, we propose a novel approach for prediction of uncertain service QoS under the dynamic Internet environment. Extensive experiments have been conducted on a real-world data set and the results demonstrate the effectiveness and applicability of our approach for QoS prediction.

**Keywords:** Service-Oriented Computing · Uncertain QoS Prediction · Collaborative Filtering · Matrix Factorization.

## 1 Introduction

Web services are self-contained and self-describing computational Web components designed to support machine-to-machine interaction by programmatic Web method calls [2]. With the development of Web services on the Internet, Quality of Services (QoS) has become a very important criterion as it can distinguish services with the same functionality. In general, QoS criteria can be divided into user-independent and user-dependent properties. User-independent QoS criteria (e.g., price, popularity) are usually defined by service providers, which are identical to different service requesters. Due to the influence of unpredictable network connections and heterogeneous user environments, QoS values of those user-dependent criteria (e.g., failure probability, response time, throughput) can fluctuate widely among different service requesters or different invocations by the same service requester.

QoS as nonfunctional property plays an important role in many research branches of services computing, such as service selection under QoS constraints [5,7,8], dynamic composition of web services with QoS optimality [17], and QoS prediction for service recommendation [4,9]. However, it is difficult and sometimes impossible to trigger actual service invocation transactions on the client-side. Moreover, it is also impractical to release QoS information from service providers or third-communities when publishing their web services on the Internet for use. Therefore, there are usually a large number of missing QoS values in the user-service historical invocation records and how to effectively predict these missing service QoS has become solid foundations of further service-oriented applications.

In recent years, existing QoS-aware approaches have been done by applying matrix factorization theory for QoS prediction of web services. The traditional matrix factorization based methods decompose user-service transactional QoS matrix formed by the historical invocations into two characteristic sub-matrices, thereby predicting the missing values in the original QoS matrix [10]. On the basis of the traditional matrix factorization method, there are researchers who have made modifications on QoS matrix with external heuristic information to improve the accuracy of service QoS prediction. By calculating user similarity, some works combined user neighborhood with traditional matrix factorization to more accurately predict service QoS values [6,14,15]. Besides considering similar users, another group of works further considered similar services to optimize matrix factorization process for more precise QoS prediction [19].

However, most existing approaches mainly focused on how to accurately implement QoS matrix factorization and they rarely took the uncertainty of QoS transactions into consideration for effective service recommendation. During actual service invocation processes, users often make a lot of invocations to a same service over a long period of time. As a result, a pair of user-service combination always generates multiple QoS transactional records. Therefore, it is mandatory to design a novel approach for effectively solving the research issue on service QoS prediction with uncertainty. To solve the above issue, we have fully considered the features of uncertain QoS of web services and proposed a new method to achieve more accurate uncertain service QoS prediction satisfying the demands on real-world service-oriented applications.

The main contributions of this paper are threefold as below.

- First, user uncertain QoS model is structured as a three-layer tree leveraging the historical QoS transaction logs, which depicts the QoS uncertainty of users invoking web services many times under a dynamic application environment.
- Second, considering the feature of QoS uncertainty, we propose a new similarity calculation method to identify the neighborhood in terms of user side

and service side, respectively. Furthermore, by fusing the similar users and services into a unified matrix factorization framework in an uncertain service invocation environment, we propose three QoS prediction strategies called U_UMF S_UMF and US_UMF.

- Third, we implement a prototype system for QoS prediction with uncertainty and conduct extensive experiments on a large-scale real-world data set that has more than 1.5 million QoS transaction logs from service invocations of users among 27 countries. The experimental results validate the effectiveness of the proposed approach for uncertain QoS prediction.

The remainder of this paper is organized as follows. Section 2 gives a motivating example. Section 3 elaborates our uncertain QoS prediction approach. Section 4 shows extensive experiments and analyzes the results. Section 5 reviews the related work and Section 6 concludes the paper.

## 2 Motivating Example

In this section, we give a motivating example to illustrate the uncertainty of service QoS during user invocation process. It is observed from Fig.1 that there are a set of services and users in the service-oriented application environment. Each user has invoked a subset of services at different times, generating uncertain service QoS transaction logs. The nonfunctional QoS values (e.g., response time) are observed by each service requester.

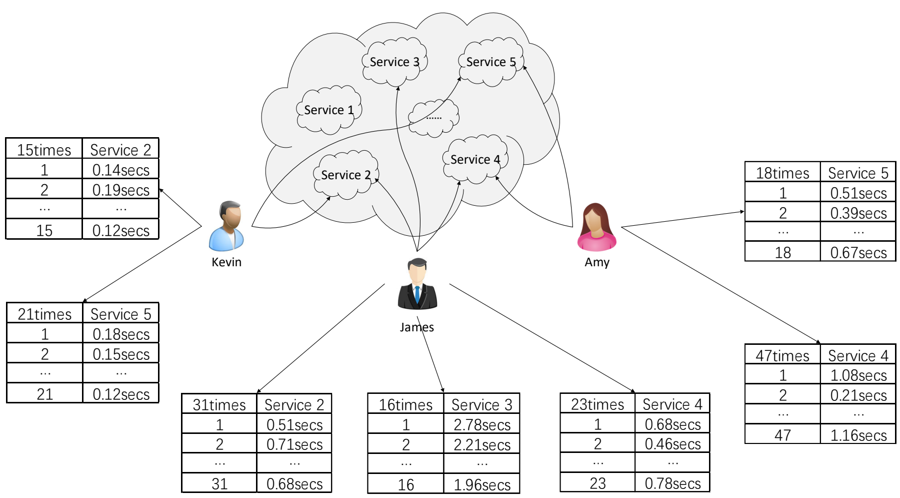


**Fig. 1.** The motivating scenario of service invocations with QoS uncertainty

In this scenario, let's take James as an example. As a service requester, James has invoked service 2, service 3 and service 4 for 31, 16 and 23 times, respectively. Those invocations of the same service reflect different variations on response time in an uncertain application environment. Now we aim to predict his missing QoS values for service 5. In traditional matrix factorization methods, they mostly consider the situation where each user invokes a service only once. However, in real-world service-oriented applications, a user may invoke a web service a bunch of times, each of which consists of multiple QoS invocation records. Based on above hypothesis, we should take the QoS uncertainty of web services into consideration so as to improve prediction accuracy and usability for real-world application demands.

## 3 Uncertain QoS Prediction of Web Services

In this section, we first illustrate the overall framework of our approach for uncertain QoS prediction of web services. Then, we detailedly present each component in the framework.

### 3.1 Framework of the approach

We design an approach for uncertain QoS prediction of web services via improved matrix factorization integrated with neighborhood information of services and users. The overall framework is shown in Fig. 2.

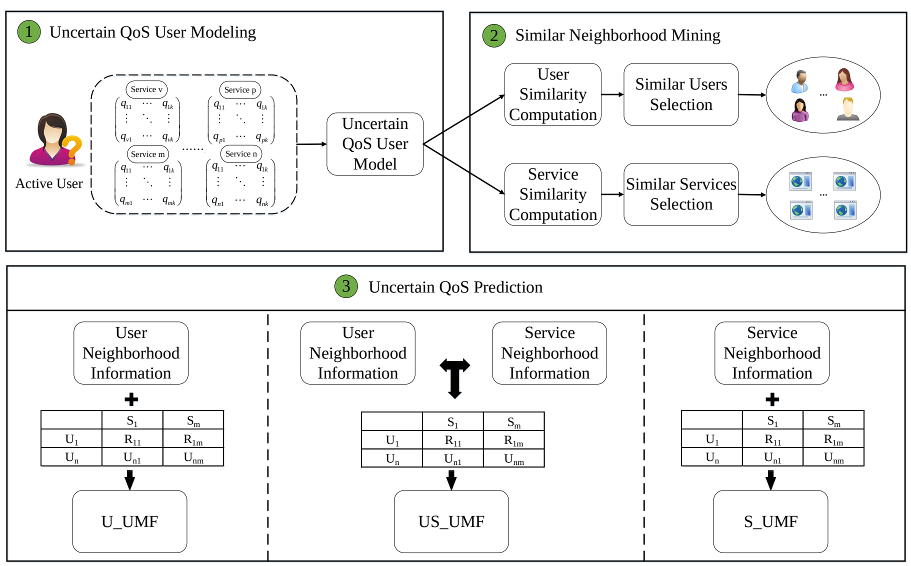


**Fig. 2.** The framework of QoS prediction of web services with uncertainty

The whole framework is composed of three key steps, including uncertain QoS user modeling, similar neighborhood mining, and uncertain QoS prediction. Initially, we build up uncertain QoS user model leveraging the historical QoS transaction logs. On the basis of nonfunctional QoS criteria modeling of a service user, we evaluate the neighborhood relationship between all of the users and services. Then a neighborhood set can be mined for an active user or target service by the similarity computation among uncertain QoS user models. Finally, we fuse user neighborhood and/or service neighborhood into traditional matrix factorization algorithm. Consequently, we propose three different uncertain QoS prediction strategies: U_UMF (User_Uncertain Matrix Factorization), S_UMF (Service_Uncertain Matrix Factorization) and US_UMF (User Service_Uncertain Matrix Factorization).

### 3.2 Uncertain QoS User Modeling

It is observed that a service user may invoke a set of web services, each of which corresponds to multiple QoS transaction logs that can be formalized as a QoS matrix with uncertainty. Formally, an uncertain QoS user model is defined as a four-tuple $\langle Auser, Lservices, Smatrices, f \rangle$. $Auser$ represents an active user; $Lservices$ is a list of all the services which the active user has ever invoked; $Smatrices$ consists of a number of different matrices where each matrix includes all of the QoS transaction logs invoked by the active user on a web service in $Lservices$; $f$ is a mapping function from a service to its corresponding uncertain QoS matrix denoted by $f : Lservices \rightarrow Smatrices$.

We visualize uncertain QoS user model of web services as a tree with three layers. It can be illustrated in Fig. 3, including user layer, invocation service layer and QoS transaction matrixes layer.

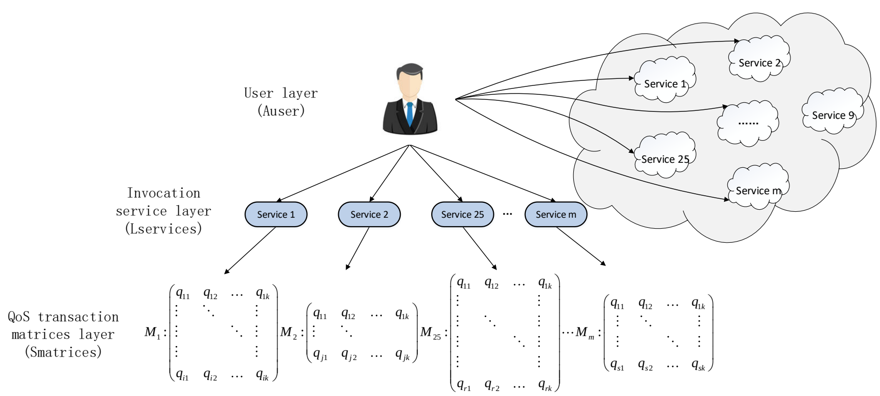


**Fig. 3.** Uncertain QoS user model of web services

### 3.3 Similar Neighborhood Mining

To integrate external heuristic information with uncertainty into a matrix factorization framework, the neighborhood set with similar users or services needs to be identified for making QoS prediction.

**User Neighborhood Mining** Given an active user $a$, its neighborhood set can be mined by evaluating the similarity between $a$ and each candidate user $u$ with their corresponding uncertain QoS user model. Given an uncertain QoS transaction matrix, it is transformed into a vector by averaging the invocation

QoS values within each corresponding column. To simplify the calculation, we only take the $i$th QoS criterion into consideration.

$$\overline{Val(a,s)} = \frac{\sum_{j=1}^{n} V(q_{ji})}{n} \tag{1}$$

Where $n$ counts the number of invocation times on web service $s$ by user $a$, $V(q_{ji})$ is the QoS value on the $i$th criterion in the $j$th invocation on service $s$ by user $a$.

After the transformation, a modified Pearson Correlation Coefficient is employed to calculate the similarity between an active user $a$ and a candidate user $u$. Here, two weighting factors that correspond to user $a$ and $u$ are calculated as below.

$$\lambda_a = \frac{N_a(s)}{N_a(s) + N_u(s)} \qquad \lambda_u = \frac{N_u(s)}{N_a(s) + N_u(s)} \tag{2}$$

Where $N_a(s)$ counts the number of invocation times on web service $s$ by user $a$, $N_u(s)$ counts the number of invocation times on web service $s$ by user $u$ .

$$Sim(a,u) = \frac{\sum_{s \in S(\cap)} \lambda_a(\overline{Val(a,s)} - Ave(\overline{Val(a)})) * \lambda_u(\overline{Val(u,s)} - Ave(\overline{Val(u)}))}{\sqrt{\sum_{s \in S(\cap)} (\lambda_a(\overline{Val(a,s)} - Ave\overline{Val(a)}))^2} * \sqrt{\sum_{s \in S(\cap)} (\lambda_u(\overline{Val(u,s)} - Ave(\overline{Val(u)})))^2}} \tag{3}$$

Where $S(a)$ and $S(u)$ represent web services that $a$ and $u$ have invoked, respectively. $S(\cap) = S(a) \cap S(b)$ denotes services that both $a$ and $u$ have invoked, $\overline{Val(a,s)}$ represents the average QoS on $s$ invoked by $a$ and $\overline{Val(u,s)}$ represents the average QoS on $s$ invoked by $u$. $Ave(\overline{Val(a)})$ represents the average QoS on all the services invoked by user $a$ and $Ave(\overline{Val(u)})$ represents the average QoS on all the services invoked by user $u$.

Although the similarity from the modified Pearson Correlation Coefficient calculation is evaluated by the difference of co-invoked web services between two users, it may still incur overestimation on the similarity calculation as there exists such a situation where both users invoke very small amount of services with high QoS similarity. Based on this observation, the similarity calculation can be further improved by a weighting factor that devalues the overestimated similarity.

$$Sim'(a,u) = \frac{2 * |S(a) \cap S(u)|}{|S(a) \cup S(u)|} * Sim(a,u) \tag{4}$$

By doing so, if an active user provides more QoS transaction logs, it is probably associated with more accurate neighborhood set. Along this way, similar neighborhood set can be mined by the traditional Top-K algorithm as below.

$$T(a) = \{u | u \in U \wedge u \in Top{-}K(a), u \neq a\} \tag{5}$$

Where $T(a)$ is a set of top K similar users for the active user $a$.

**Service Neighborhood Mining** Given two web services $s$ and $v$ , their similarity is evaluated based on the established uncertain QoS user model. To that end, we modify Pearson Correlation Coefficient to calculate the similarity degree by the following two weighting factors, indicating the importance on each web service.

$$\lambda_s = \frac{N_u(s)}{N_u(s) + N_u(v)} \qquad \lambda_v = \frac{N_u(v)}{N_u(s) + N_u(v)} \tag{6}$$

Where $N_u(s)$ counts the number of invocation times on web service $s$ by user $u$, and $N_u(v)$ counts the number of invocation times on web service $v$ by user $u$.

$$Sim(s,v) = \frac{\sum\limits_{u \in U(\cap)} \lambda_s(\overline{Val(u,s)} - Ave(\overline{Val(s)})) * \lambda_v(\overline{Val(u,v)} - Ave(\overline{Val(v)}))}{\sqrt{\sum\limits_{u \in U(\cap)} (\lambda_s(\overline{Val(u,s)} - Ave\overline{Val(s)}))^2} * \sqrt{\sum\limits_{u \in U(\cap)} (\lambda_v(\overline{Val(u,v)} - Ave\overline{Val(v)}))^2}} \tag{7}$$

Where $U(s)$ and $U(v)$ represent the users who have invoked service $s$ and service $v$, respectively. $\overline{U(\cap) = U(s) \cap U(v)}$ denotes the users who have both invoked service $s$ and $v$, $\overline{Val(u,s)}$ represents the average QoS on $s$ invoked by $u$ and $\overline{Val(u,v)}$ represents the average QoS on $v$ invoked by $u$. $Ave(\overline{Val(s)})$ represents the average QoS on service $s$ invoked by all of the users and $Ave(\overline{Val(v)})$ represents the average QoS on service $v$ invoked by all of the users.

To further improve the reliability, an enhanced PCC for the similarity calculation between two web services is used to avoid the overestimation case.

$$Sim'(s,v) = \frac{2 * |U(s) \cap U(v)|}{|U(s) \cup U(v)|} * Sim(s,v) \tag{8}$$

After the above adjustment, similar neighborhood set can be mined by Top-K algorithm.

$$T(s) = \{v | v \in S \wedge v \in Top - K(S), v \neq s\} \tag{9}$$

Where $T(s)$ is a set of top K similar services for the target service $s$.

### 3.4 Uncertain QoS Prediction

On the basis of similar neighborhood mining with QoS uncertainty, we apply the idea of collaborative filtering to improve the traditional matrix factorization algorithm and propose three kinds of QoS prediction strategies:U_UMF (User_Uncertain Matrix Factorization), S_UMF (Service_Uncertain Matrix Factorization) and US_UMF (User Service_Uncertain Matrix Factorization)

**U_UMF QoS Prediction** In collaborative filtering based recommender systems, the interactive experience of service users inside a neighborhood should be highly similar. As a result, they contribute the similar patterns on web services invocations. Based on this intuition, we improve the traditional matrix factorization prediction algorithm as U_UMF (User_Uncertain Matrix Factorization), considering similar neighborhood set of active user based on uncertain QoS of web services.

First, we define user relational regularization terms.

$$EU_{iu} = \frac{Sim'(i,u)}{\sum\limits_{g \in T(i)} Sim'(i,g)} \tag{10}$$

$$min \sum_{i=1}^{m} \left\| U_i - \sum_{u \in T(i)} EU_{iu} * U_u \right\|_F^2 \tag{11}$$

Where $T(i)$ is a set of top K similar neighborhood users for the user $i$. $Sim'(i,u)$ represents the similarity between user $i$ and $u$. $U_i$ represents the eigenvector of user $i$. The goal of this paradigm is to minimize the intrinsic behavioral commonality between the user $i$ and its corresponding similar neighborhood users $T(i)$. That is given similar neighborhood set, the eigenvector of $i$ is similar with the average eigenvector of all users. After that, we combine the paradigm with the traditional matrix factorization, forming a new objective function for uncertain QoS prediction model [10].

$$\begin{aligned} \mathcal{L}_U(R,U,S) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}(R_{ij} - U_i^T S_j)^2 + \frac{\lambda_1}{2}\left\|U\right\|_F^2 \\ + \frac{\lambda_2}{2}\left\|S\right\|_F^2 + \frac{\alpha_1}{2}\sum_{i=1}^{m}\left\| U_i - \sum_{u \in T(i)} EU_{iu} * U_u \right\|_F^2 \end{aligned} \tag{12}$$

Where $R = U^T S$ is the user-service matrix. $I_{i,j}$ is the indicator function that is equal to 1 if user $i$ invoked web service $j$ and is equal to 0 otherwise. $\left\|\cdot\right\|_F^2$ denotes the Frobenius norm. $\alpha_1$ controls the importance of this constraint paradigm in the U_UMF matrix factorization model. Two regularization terms related to $U$ and $S$ are involved to avoid the problem of overfitting, where $\lambda_1$ and $\lambda_2$ are the learning rates. In many recommendation systems, this formula was widely adopted.

To solve the modeled uncertain QoS prediction problem, we leverage the most commonly used gradient descent algorithm to find the optimum value of the objective function.

$$\frac{\partial \mathcal{L}_U}{\partial U_i} = \sum_{j=1}^{n} I_{i,j}(R_{i,j} - U_i^T S_j)(-S_j) + \lambda_1 U_i + \alpha_1 (U_i - \sum_{u \in T(i)} EU_{iu} * U_u) \tag{13}$$

$$\frac{\partial \mathcal{L}_U}{\partial S_j} = \sum_{i=1}^{m} I_{i,j}(R_{i,j} - U_i^T S_j)(-U_i) + \lambda_2 S_j \tag{14}$$

**S_UMF QoS Prediction** From the perspective of web services, similar neighborhood services with the target service can be combined with traditional matrix factorization to improve the accuracy of the service QoS prediction. We propose an improved matrix factorization prediction algorithm called S_UMF (Service_Uncertain Matrix Factorization), considering similar services with QoS uncertainty.

It is observed that similar services share high similarity when invoked by different users. Under this assumption, the regularization of service relations can be defined as below.

$$ES_{jv} = \frac{Sim'(j,v)}{\sum_{h \in T(j)} Sim'(j,h)} \tag{15}$$

$$min \sum_{j=1}^{n} \left\| S_j - \sum_{s \in T(j)} ES_{js} * S_s \right\|_F^2 \tag{16}$$

Where $T(j)$ is a set of top K similar neighborhood services for the service $j$. $Sim'(j,v)$ represents the similarity between service $j$ and $v$. $S_j$ represents the eigenvector of service $j$. The goal of this paradigm is to minimize the intrinsic behavioral commonality between the service $j$ and its corresponding similar neighborhood services $T(j)$. In the same way, we combine the paradigm with the traditional matrix factorization, forming a new objective function for uncertain QoS prediction model.

$$\begin{aligned}\mathcal{L}_U(R,U,S) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}(R_{ij} - U_i^T S_j)^2 + \frac{\lambda_1}{2}\|U\|_F^2 \\ + \frac{\lambda_2}{2}\|S\|_F^2 + \frac{\alpha_2}{2}\sum_{j=1}^{n}\left\| S_j - \sum_{s \in T(j)} ES_{js} * S_s \right\|_F^2\end{aligned} \tag{17}$$

Where $\alpha_2$ controls the importance of this constraint paradigm in the S_UMF matrix factorization model.

Finally, gradient descent algorithm is applied to solve the modeled uncertain QoS problem and find the optimum value of the objective function.

$$\frac{\partial \mathcal{L}_S}{\partial U_i} = \sum_{j=1}^{n} I_{i,j}(R_{i,j} - U_i^T S_j)(-S_j) + \lambda_1 U_i \tag{18}$$

$$\frac{\partial \mathcal{L}_S}{\partial S_j} = \sum_{i=1}^{m} I_{i,j}(R_{i,j} - U_i^T S_j)(-U_i) + \lambda_2 S_j + \alpha_2(S_j - \sum_{s \in T(j)} EU_{js} * S_s) \tag{19}$$

**US_UMF QoS Prediction** When comprehensively considering both similar users and services, we propose an improved matrix factorization prediction algorithm called US_UMF (User Service_Uncertain Matrix Factorization) with the consideration of uncertain QoS. Combining formula (12) and (17), the objective function on QoS prediction problem for US_UMF is as below.

$$\begin{aligned}\mathcal{L}_U(R,U,S) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}(R_{ij} - U_i^T S_j)^2 + \frac{\lambda_1}{2}\|U\|_F^2 + \frac{\lambda_2}{2}\|S\|_F^2 + \\ \frac{\alpha_1}{2}\sum_{i=1}^{m}\left\| U_i - \sum_{u \in T(i)} EU_{iu} * U_u \right\|_F^2 + \frac{\alpha_2}{2}\sum_{j=1}^{n}\left\| S_j - \sum_{s \in T(j)} ES_{js} * S_s \right\|_F^2\end{aligned} \tag{20}$$

Where $\alpha_1$ and $\alpha_2$ control the importance of similar user and service, respectively. As with the above two methods, the gradient descent algorithm is used to derive an optimum solution to the uncertain QoS prediction problem.

$$\frac{\partial \mathcal{L}_U}{\partial U_i} = \sum_{j=1}^{n} I_{i,j}(R_{i,j} - U_i^T S_j)(-S_j) + \lambda_1 U_i + \alpha_1(U_i - \sum_{u \in T(i)} EU_{iu} * U_u) \tag{21}$$

$$\frac{\partial \mathcal{L}_S}{\partial S_j} = \sum_{i=1}^{m} I_{i,j}(R_{i,j} - U_i^T S_j)(-U_i) + \lambda_2 S_j + \alpha_2(S_j - \sum_{s \in T(j)} EU_{js} * S_s) \tag{22}$$

## 4 Experimental Evaluation

To validate the effectiveness and applicability of our proposed approach for uncertain QoS prediction, we conduct a set of experiments on a large-scale real-world service data set.

### 4.1 Experimental Settings and Data Set

The experiments are carried on a PC with Intel(R) core(TM) i5-5200U processor 2.20 GHz and 8G RAM in Windows 10 operating system. Our large number of data set of QoS transaction logs is from WS-DREAM[1], collecting approximately 1.5 million historical QoS invocation records from multiple execution processes of 101 web services. Furthermore, 150 service requesters from more than 20 countries participated in the service invocation processes of these services. The distributions of users and services in different countries is illustrated in Fig. 4. More specifically, Table 1 shows the situation of a user access to each service with multiple QoS transaction logs.

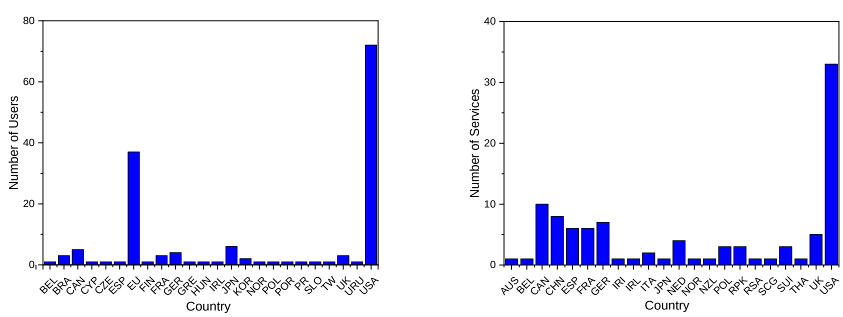


**Fig. 4.** Data distributions of users and services in different countries

[1] http://wsdream.github.io

**Table 1.** Uncertain QoS transaction logs for a user invoking a set of web services

| UserIP | WSID | Response Time | DataSize | HttpCode | HttpMessage |
|---|---|---|---|---|---|
| 35.9.27.26 | 1521 | 1101 | 1457 | 200 | Bad types |
| 35.9.27.26 | 1521 | 2032 | 1457 | 200 | Bad types |
| 35.9.27.26 | 1521 | 1408 | 1457 | 200 | Bad types |
| 35.9.27.26 | 6405 | 415 | 620 | 200 | OK |
| 35.9.27.26 | 6405 | 470 | 620 | 200 | OK |
| 35.9.27.26 | 8953 | 20009 | 2624 | -1 | Timed Out |
| 35.9.27.26 | 8953 | 20002 | 2624 | -1 | Timed Out |
| 35.9.27.26 | 8953 | 20024 | 2624 | -1 | Timed Out |

### 4.2 Competitive Methods and Evaluation Metrics

We carried out experiments on four different methods for QoS prediction of web services including the traditional matrix factorization method and the three improved ones in support of uncertain Internet application environment. They are described as below.

1) MF (Matrix Factorization): The most primitive matrix factorization prediction algorithm.
2) U_UMF (User_Uncertain Matrix Factorization): An improved matrix factorization prediction algorithm for QoS prediction that combines similar neighborhood users with uncertainty and the traditional matrix factorization model.
3) S_UMF (Service_Uncertain Matrix Factorization): An improved matrix factorization prediction algorithm for QoS prediction that fuses similar neighborhood services with uncertainty and the traditional matrix factorization model.
4) US_UMF (User Service_Uncertain Matrix Factorization): An improved matrix factorization prediction algorithm for QoS prediction that integrates both similar neighbor users and services with uncertainty into the traditional matrix factorization model.

In order to measure the accuracy on QoS prediction among different approaches, Mean Absolute Error(MAE) and NMAE are used as the evaluation metrics. They are defined as below.

$$MAE = \frac{\sum_{i,j} |R_{i,j} - \hat{R}_{i,j}|}{N} \tag{23}$$

Where $R_{i,j}$ represents the real QoS value and $\hat{R}_{i,j}$ represents the predicted QoS value, when user $i$ invokes service $j$. Note that NMAE is a normalized form of MAE.

$$NMAE = \frac{MAE}{\sum_{i,j} R_{i,j}/N} \tag{24}$$

**Table 2.** Experimental results of uncertain QoS prediction among four competitive service approaches at three different matrix densities

| | Density=5% | | Density=15% | | Density=25% | |
|---|---|---|---|---|---|---|
| | MAE | NMAE | MAE | NMAE | MAE | NMAE |
| MF | 1085.744 | 0.7828 | 841.346 | 0.6043 | 719.8431 | 0.5109 |
| U_UMF | 1084.241 | 0.782 | 838.5726 | 0.6011 | 714.051 | 0.5099 |
| S_UMF | 1083.203 | 0.7812 | 832.8425 | 0.5901 | 710.5599 | 0.5092 |
| US_UMF | **1080.589** | **0.7791** | **814.4709** | **0.5851** | **706.9142** | **0.5087** |

### 4.3 Experimental Results and Analysis

In order to make the experiments closer to the actual situation, we preprocess the QoS transactions by randomly pruning invocation records from user-service QoS matrix to imitate real-world service-oriented application scenario. We dilute uncertain QoS matrix density to 5%, 10%, 15%, 20% and 25%, respectively. The experiments on uncertain QoS prediction of web services are conducted among four competitive approaches. The experimental results are shown in Table 2 and Fig. 5, respectively.

From the experimental results, three observations can be concluded in Table 2 and Fig. 5. Furthermore, among them US_UMF method shows the best accuracy in QoS prediction, because more information is taken into account. The experimental result validates our theory well.

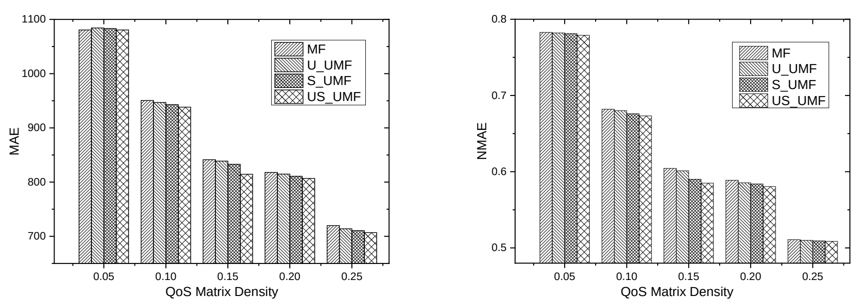


**Fig. 5.** Experimental results on MAE and NMAE among four competitive service QoS prediction approaches at five different matrix densities

To further verify the effectiveness and applicability of the improved matrix factorization approach integrated with similar user and services for uncertain QoS prediction we adjust three crucial parameters involved in the approach and test them how to influence the performance on prediction accuracy. These parameters include the number of similar neighborhood users selected for an active user (TOP-K_U), the number of similar neighborhood services selected for a target service (TOP-K_S), and the number of hidden topics in the matrix factorization framework (Dimensionality). We performed three sets of experiments on US_UMF approach each of which aims to adjust one of the three parameters.

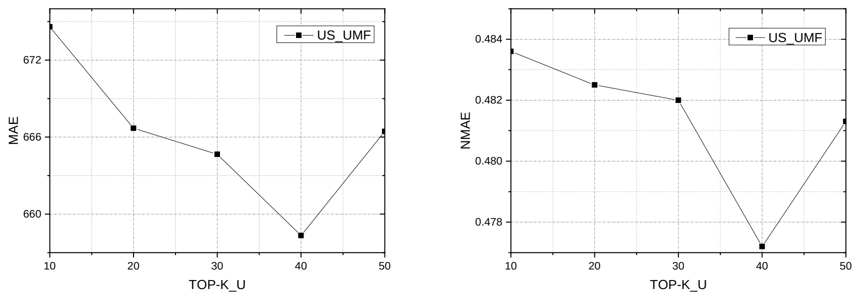


**Fig. 6.** The results of MAE and NMAE along with the changes of TOP-K_U

In experiment one we set the uncertain QoS matrix sparsity to 25%, Dimensionality to 10, and TOP-K_S to 10, while TOP-K_U varies from 10 to 50 with 10 interval. We can find from the experimental results in Fig. 6 as the number of similar users increases from 10 to 40, the MAE and NMAE of US_UMF gradually decrease and they reach the best when TOP-K_U equals 40. That is, the prediction accuracy of uncertain service QoS becomes better and better. However, when TOP-K_U reaches 50, the effectiveness of QoS prediction begins to decline. The main reason is that we have selected some users with low similarity in the neighborhood set for an active user, which reduces the uncertain QoS prediction accuracy. Therefore, as the number of similar neighborhood users increases, the performance on QoS prediction becomes better and better within a certain range.

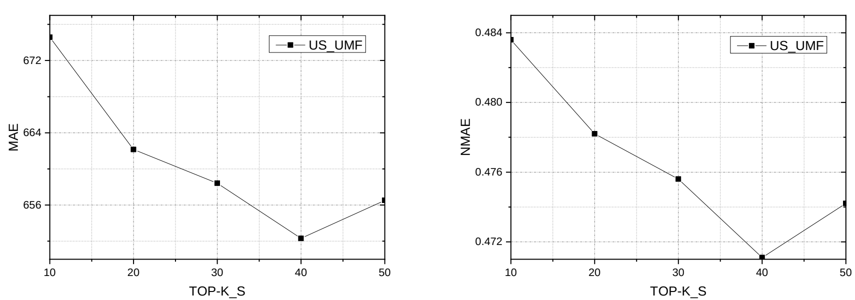


**Fig. 7.** The results of MAE and NMAE along with the changes of TOP-K_S

In experiment two we set the uncertain QoS matrix sparsity to 25%, Dimensionality to 10, and TOP-K_U to 10, while TOP-K_S varies from 10 to 50 with 10 interval. The results in Fig. 7 demonstrate that as the number of similar users increases from 10 to 40, the MAE and NMAE of US_UMF gradually decrease and they reach the best when TOP-K_S arrives at 40. In other words, the prediction accuracy of uncertain service QoS becomes better and better. However, when TOP-K_S reaches 50, the effectiveness of QoS prediction begins to decline. The main reason is that we have selected some services with low similarity in the neighborhood set for a target service, which reduces the uncertain QoS prediction accuracy. Therefore, as the number of neighborhood users increases, performance on QoS prediction becomes better and better within a certain range.

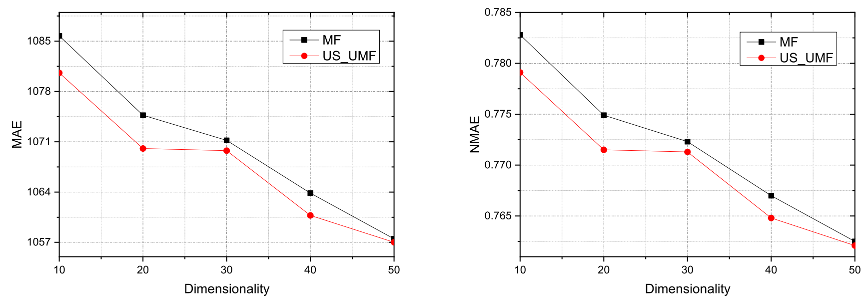


**Fig. 8.** The results of MAE and NMAE along with the changes of TOP-K_S

Dimensionality indicates how many feature factors are considered in matrix factorization model and it is an important parameter that influences the prediction accuracy. In experiment three we set the uncertain QoS matrix sparsity to 25%, Top-K_S to 10, and TOP-K_U to 10, while Dimensionality varies from 10 to 50 with 10 interval. We can observe from Fig. 8 that as the number of Dimensionality increases, performance on QoS prediction becomes better and better within a certain range.

## 5 Related Work

In recent years, QoS-aware services recommendation based on the prediction of missing QoS values and the technology of collaborative filtering and matrix factorization has gained a lot of attentions [12, 16, 18].

There are two kinds of collaborative filtering method for QoS prediction of web services: neighborhood-based and model-based. The neighborhood-based methods can be further divided into similar-user based [3], similar-service based [11], and the combination of them [13]. Neighborhood-based approaches often apply PCC as the calculation method when evaluating the relationship between two users or services, since it shows excellent performance in similarity measurement [1]. Different with the neighborhood-based method, model-based method usually train a predefined model with large-scale service repository. By using the trained model, it can recommend similar users or services for an active user or a target service. In more detail, clustering and classification algorithms are the most commonly used machine learning methods for model-driven collaborative filtering service recommendation [20].

In recent years, matrix factorization methods have been proposed for QoS prediction of web services. These methods focus on using the historical QoS transaction logs to model them as a QoS matrix that is decomposed by two characteristic submatrices. As a result, those missing values can be predicted in the original QoS matrix. Recently, some researchers tried to combine the techniques of collaborative filtering and matrix factorization together to make better QoS prediction for service recommendation. Zheng et al. integrated similar neighborhood users to traditional matrix factorization [19]. Besides considering similar users, Wei et al. further applied similar neighborhood services to optimize the process of matrix factorization [14].

However, most existing approaches mainly count on the QoS transaction logs without the consideration of uncertainty, which may reduce the prediction accuracy and its applicability on service-oriented systems.

## 6 Conclusion

In this paper, we fully consider the QoS uncertainty of real-world service invocations in dynamic Internet environment. To solve QoS prediction problem with uncertainty, we propose an approach consisting of three steps. First, uncertain QoS user model is constructed by the representation of three-layer tree. Second, we evaluate the similarity among uncertain QoS user models and mine similar neighborhood users and services. Finally, we integrate the generated similar users and services with QoS uncertainty into traditional matrix factorization model and present three QoS prediction strategies, including User_Uncertain Matrix Factorization (U_UMF), Service_Uncertain Matrix Factorization (S_UMF), and User Service_Uncertain Matrix Factorization (US_UMF). Comprehensive experiments conducted on large-scale real-world web services validate the effectiveness and applicability of the proposed approach for service-oriented systems.

## Acknowledgement

This work was partially supported by Shanghai Natural Science Foundation (No. 18ZR1414400, 17ZR1400200), National Natural Science Foundation of China (No. 61772128, 61303096), Shanghai Sailing Program (No. 16YF1400300), and Fundamental Research Funds for the Central Universities (No. 16D111208).

## References

1. Alshamri, M.Y.H., Alashwal, N.H.: Fuzzy-weighted similarity measures for memory-based collaborative recommender systems. Journal of Intelligent Learning Systems & Applications **6**(1), 1–10 (2014)
2. Bichier, M., Lin, K.J.: Service-oriented computing. Wiley, (2006)
3. Breese, J.S., Heckerman, D., Kadie, C.: Empirical analysis of predictive algorithms for collaborative filtering. Uncertainty in Artificial Intelligence **98**(7), 43–52 (2013)
4. Deng, S., Wang, D., Li, Y., Cao, B., Yin, J., Wu, Z., Zhou, M.: A recommendation system to facilitate business process modeling. IEEE Transactions on Cybernetics **47**(6), 1380–1394 (2016)

5. Deng, S., Wu, H., Hu, D., Zhao, J.L.: Service selection for composition with qos correlations. IEEE Transactions on Services Computing **9**(2), 291–303 (2016)
6. Ding, S., Li, Y., Wu, D., Zhang, Y., Yang, S.: Time-aware cloud service recommendation using similarity-enhanced collaborative filtering and arima model. Decision Support Systems **107**, 103–115 (2018)
7. Hadad, J.E., Manouvrier, M., Rukoz, M.: TQoS: Transactional and QoS-Aware selection algorithm for automatic web service composition. IEEE Transactions on Services Computing **3**(1), 73–85 (2010)
8. Haddad, J.E., Manouvrier, M., Ramirez, G., Rukoz, M.: QoS-Driven selection of web services for transactional composition. In: IEEE International Conference on Web Services. pp. 653–660 (2008)
9. Kuang, L., Xia, Y., Mao, Y.: Personalized services recommendation based on context-aware QoS prediction. In: IEEE International Conference on Web Services. pp. 400–406 (2012)
10. Salakhutdinov, R., Mnih, A.: Probabilistic matrix factorization. In: International Conference on Neural Information Processing Systems. pp. 1257–1264 (2007)
11. Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Item-based collaborative filtering recommendation algorithms. In: International Conference on World Wide Web. pp. 285–295 (2001)
12. Shao, L., Zhang, J., Wei, Y., Zhao, J., Xie, B., Mei, H.: Personalized QoS prediction forweb services via collaborative filtering. In: IEEE International Conference on Web Services. pp. 439–446 (2007)
13. Wang, Jun, Vries, D., Arjen, P., Reinders, Marcel, J.T.: Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In: ACM SIGIR Conference on Information Retrieval. pp. 501–508 (2006)
14. Wei, L., Yin, J., Deng, S., Li, Y., Wu, Z.: An extended matrix factorization approach for QoS prediction in service selection. In: IEEE International Conference on Services Computing. pp. 162–169 (2012)
15. Wu, X., Cheng, B., Chen, J.: Collaborative filtering service recommendation based on a novel similarity computation method. IEEE Transactions on Services Computing **10**(3), 352–365 (2017)
16. Xu, Y., Yin, J., Deng, S., Xiong, N.N., Huang, J.: Context-aware QoS prediction for web service recommendation and selection. Expert Systems with Applications **53**, 75–86 (2016)
17. Yilmaz, A.E., Karagoz, P.: Improved genetic algorithm based approach for QoS aware web service composition. In: IEEE International Conference on Web Services. pp. 463–470 (2014)
18. Zheng, Z., Ma, H., Lyu, M.R., King, I.: QoS-Aware web service recommendation by collaborative filtering. IEEE Transactions on Services Computing **4**(2), 140–152 (2011)
19. Zheng, Z., Ma, H., Lyu, M.R., King, I.: Collaborative web service QoS prediction via neighborhood integrated matrix factorization. IEEE Transactions on Services Computing **6**(3), 289–299 (2013)
20. Zou, G., Li, W., Zhou, Z., Niu, S., Gan, Y., Zhang, B.: Clustering-based uncertain QoS prediction of web services via collaborative filtering. International Journal of Web and Grid Services **13**(4), 403–424 (2017)