Research Article

From Reputation Perspective: A Hybrid Matrix Factorization for QoS Prediction in Location-Aware Mobile Service Recommendation System

Shun Li [ID],[1] Junhao Wen [ID],[1] and Xibin Wang [ID][2]

[1]*School of Big Data and Software Engineering, Chongqing University, Chongqing 400044, China*
[2]*School of Data Science, Guizhou Institute of Technology, Guiyang 550003, China*

Correspondence should be addressed to Junhao Wen; jhwen@cqu.edu.cn

With the great development of mobile services, the Quality of Services (QoS) becomes an essential factor to meet end users' personalized requirement on the nonfunctional performance of mobile services. However, most of the QoS values in real cases are unattainable because a service user would only invoke some specific mobile services. Therefore, how to predict the missing QoS values and recommend high-quality services to end users becomes a significant challenge in mobile service recommendation research. Previous QoS prediction researches demonstrate that the nonfunctional performance of mobile services is closely related to users' location information. However, most location-aware QoS prediction methods ignore the premise that the obtainable QoS values observed by different users in same location region would probably be untrustworthy, which will lead to inaccurate and unreliable prediction results. To make credible location-aware QoS prediction, we propose a hybrid matrix factorization method integrated location and reputation information (LRMF) to predict the unattainable QoS values. Our approach firstly cluster users into different locational region based on their geographical distribution, and then we compute users' reputation to identify untrustworthy users in every locational region. Finally, the unknown QoS values can be predicted by integrating locational cluster information and users' reputation into a hybrid matrix factorization model. Comprehensive experiments are conducted on a public QoS dataset which contains sufficient real-world service invocation records. The evaluation results indicate that our LRMF method can effectively reduce the impact of unreliable users on QoS prediction and make credible mobile service recommendation.

1. Introduction

Based on the flexibility and expansibility of mobile application development technique, tens of thousands of hybrid mobile services with similar function have been developed and provided in mobile application store. However, this phenomenon automatically leads to information overload problem in mobile service retrieval system. To tackle this challenge, Quality of Service (QoS) is used in service-oriented system to analyse the nonfunctional performance of mobile services [1–4]. QoS has been widely used in service selection, composition, and recommendation research [5–9]. In real-world service invocation scenario, users would only search and select some specific mobile services under the unpredictable Internet environment. For lots of unknown mobile services, it is impractical to make users invoke each of them and evaluate their nonfunctional performance. Therefore, how to make accurate QoS prediction for unknown mobile services is a critical step to make high-quality service recommendation in mobile service computing paradigm.

Collaborative filtering (CF) is widely utilized in most e-commerce recommender systems to predict miss rating values. Traditional CF models generally fall into two categories: memory-based and model-based. The memory-based CF of QoS prediction process would generally find a subset of similar users for the target user and recommend high-quality mobile services shared by these similar users to the

target user [10]. The model-based methods will train a model by learning users' historical QoS performance and then predict the QoS values for unknown mobile services [11, 12]. Although CF model proved to be effective in QoS prediction on different mobile services, the prediction accuracy is still unsatisfactory because of cold-start problem.

To reduce the impact of cold-start problem, more contextual information is introduced into QoS prediction model for the QoS values are greatly affected by some context factors (e.g., location distribution, invocation time, and so on) in Internet. Based on this realization, the location-aware CF model is proposed to predict unknown QoS values in service recommendation [13]. As we all know, different users in one location region generally share the same set of IT infrastructure and they would suffer from similar Internet usage experience when they invoke mobile services, as it is reported in the work of [14], in which the QoS performance is strongly correlated with the location information of users. In Figure 1, we give an example of service invocation with location information. As mentioned above, the user 1 would have similar QoS records (such as response time) with other users in the US; meanwhile, user 2 may share similar QoS records with other users in India when they invocated services such as YouTube, Twitter, etc.

Previous location-aware CF approaches usually compute the similarity of all different users in one specific location region to find most Top-K similar neighbours for the target user [12, 15]. However, some unreliable users who would submit untrustworthy QoS values will be indiscreetly included in the neighbourhood set. Unreliable users would randomly provide some QoS values or better ones to improve the visibility of their own services and worse values for others' applications [16]. Those untrustworthy QoS values would have a marked negative effect on the prediction accuracy. Therefore, it is essential to introduce credibility of available QoS values in prediction process to enhance the prediction accuracy and persuasiveness of service recommendation mechanism.

Based on above realizations, a hybrid matrix factorization algorithm is proposed by integrating users' reputation and locational information to predict the unattainable QoS values in this paper. Complementary to previous service recommendation method which only adopts available QoS values, our study tends to make credible and accurate QoS prediction for mobile service recommendation by considering the reputation of different users' QoS usage experience. We then exploit personal geographical distance and QoS values to find the locational similarity neighbourhoods and discover the latent connectivity between the target user and his/her neighbours. Meanwhile, we use users' reputation to control the weight of users' latent feature learning. Finally, these constraints are integrated into matrix factorization model to make credibly personalized QoS prediction. The following contributions are achieved in this paper:

(1) We firstly cluster users into different locational regions based on their geographical distribution and design an iterative method to compute users' reputation score by their provided QoS usage data. Then
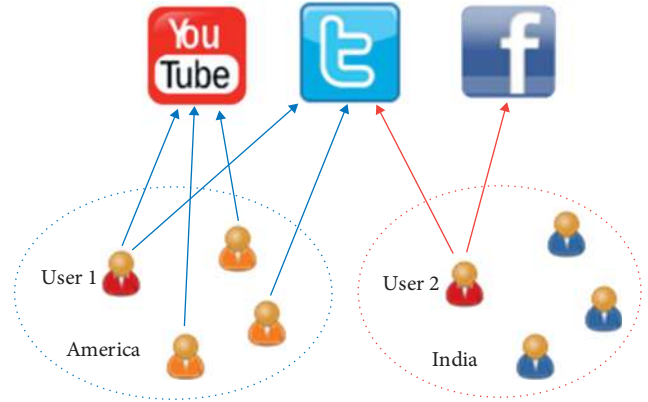

Figure 1: An example of mobile service invocation scenario.

a subset of trustworthy users in each locational region can be identified by the rank of reputation score.

(2) In the next step, a trustworthy neighbourhood can be identified by incorporating both users' location distribution and reputation score. By integrating the latent feature of both available QoS values and those shared by the neighbourhood, a hybrid matrix factorization model is proposed to make high-quality QoS prediction.

(3) Results of experiments conducted on a public QoS dataset show that by considering data credibility, our method can achieve higher prediction accuracy than other previous studies which involve the untrustworthy impact of available QoS values.

The remainder of this paper is organized as follows: Section 2 introduces related studies; Section 3 shows the basic principles of our method; Section 4 elaborates how to design our proposed method; Section 5 discusses experiments and results analysis; finally, Section 6 concludes the paper and draws future studies.

2. Related Work

2.1. QoS-Based Service Computing. QoS plays a control role in service-oriented architectures, especially in the service discovery and recommendation research. Al-Masri and Mahmoud [17] firstly calculated users' preferences on their historical QoS data, and then proposed a service discovery approach by ranking users' QoS parameters. Kritikos and Plexousakis [18] extracted the contextual information of QoS from the description file of service and designed a service discovery method. Rosario et al. [19] proposed soft probabilistic contracts on QoS parameters to composite web services and validated their method on TOrQuE tool to show its outperformance than other previous studies. Hadad et al. [20] proposed a web service composition framework by exploiting both transactional properties and QoS values. This framework composite plenty of existing web services into a workflow which can satisfy users' preferences on nonfunctional requirements. However, these methods only conducted experiments on synthetic datasets and lack authenticity in real-world service invocation.

2.2. Collaborative Filtering. Collaborative filtering is a common algorithm adopted by many recommendation systems, such as the famous commercial system Amazon [21]. Memory-based collaborative filtering approaches make prediction and recommendation by calculating the similarity of users or items [21, 22]. These methods utilize the whole entire user-item matrix as the input data, which will take a lot of time and memory spaces in the online recommendation system. Model-based collaborative filtering methods will generally train a predefined model on available data and predict the missing values in the test dataset, and then select the appropriate items as candidate list to the target user [23, 24]. Model-based approaches can learn the model quickly with little need for runtime and memory space, which will be often adopted in online recommendation systems.

Matrix factorization model is now widely adopted in many online recommendation systems for its effectiveness and efficiency. Mnih and Salakhutdinov [11] introduced the mathematical theory of matrix factorization in probabilistic analysis and validated the performance of this method on a famous film recommendation system. Zhang et al. [25] designed a personalized recommendation approach by integrating original matrix factorization with a constraint item extracted from their personal information. It identifies users into different clusters by the statistics of user behaviours on different tags and considers this constraint as a regularization term in matrix factorization model to enhance its prediction accuracy. Ma et al. [26] improved the matrix factorization approach with users' social information to enhance the prediction accuracy for social recommendation. This approach uses users' social relationship as an additional constraint which can reflect users' latent judgment of interest on items in the user-item matrix factorization. Recently, the matrix factorization methods have been widely introduced into service recommendation research [10, 15]. Although matrix factorization methods make some improvements in prediction accuracy, none of them realize that the QoS credibility deserve serious consideration.

2.3. Location-Based QoS Prediction. Location information has been widely used for service recommendation in recent years. Ali and Solis [27] presented a novel distributed service architecture that can adapt to the changes of Internet resources and location topology. Wei et al. [15] firstly calculated users' similarity with their real-world distances and then clustered users into geographical sets as a constraint item of matrix factorization to generate the location-aware QoS prediction approach. Tang et al. [28] solved the sparsity issue in QoS aware service recommendation by integrating collaborative filtering with users' geography data. Lee et al. [29] adopt the preference propagation through users in same location region to improve prediction accuracy. This work clusters users and service into different groups by the locational information and then use preference propagation to compute the similarity between different users and services, respectively. Finally, a matrix factorization model is introduced to predict missing QoS values by integrating these

constrains. Gonsalves and Patil [30] exploited users' location information and QoS values to cluster users and web services and then proposed a CF algorithms to make personalized web service recommendation. It firstly uses Pearson correlation coefficient (PCC) to identify different users and service regions and then exploit K-nearest neighbour (KNN) and support vector machine (SVM) in CF algorithm framework to predict missing QoS values. However, above studies do not consider the users' reputation, and neglect the fact that available QoS values may be untrustworthy even though these values are provided by users in same location region.

2.4. Reputation. Based on the achievement of reputation in applications (e.g., YouTube and Twitter) to avoid possible deception risk, some academics introduce the reputation into QoS prediction to enhance the reliability of service-oriented computing. The reputation values evaluated from QoS data can measure whether the available QoS values are trustworthy or not. Qiu et al. [31] designed a QoS prediction method by calculating users' reputation to obtain higher accuracy for service recommendation. In their work, reputation of different users will be computed and ranked to find the subset of unreliable users. Followed by this, the memory-based collaborative filtering model combined with reputation for QoS prediction becomes more remarkable. Based on Qiu's work, Xu et al. [32] presented an improved QoS prediction method with the users' reputation (RMF), which introduced users' reputation weight into of a matrix factorization approaches to make QoS prediction for unknown services and then recommend high-quality services to the target user. Mehdi et al. [33] introduced a stochastic approach to evaluate the reputation of services by leveraging the correlation information among different QoS metrics. Comi et al. [34] proposed a hybrid service composition method by exploiting users' reputation of QoS to help users discover and select high-quality services in multicloud environment. However, these studies do not take location information into consideration on the QoS prediction.

3. Principles and Reputation Analysis

In this section, we would introduce the main principles our LRMF method at the beginning and then present the reputation analysis on different users.

3.1. Principles of LRMF. Previous CF-based QoS prediction approaches (e.g., [7, 35]) only utilize the available QoS values in collaborative filtering model to make personalized service recommendation. However, these methods ignore that users' reputation and location will make great impact on the prediction results. Therefore, we design a novel mobile service recommendation system by considering both of users' reputation and location information when predicting missing QoS values. As presented in Figure 2, historical invocations with QoS and location data will be submitted to server database in the service invocation process. Then, the reputation and location information could be calculated by
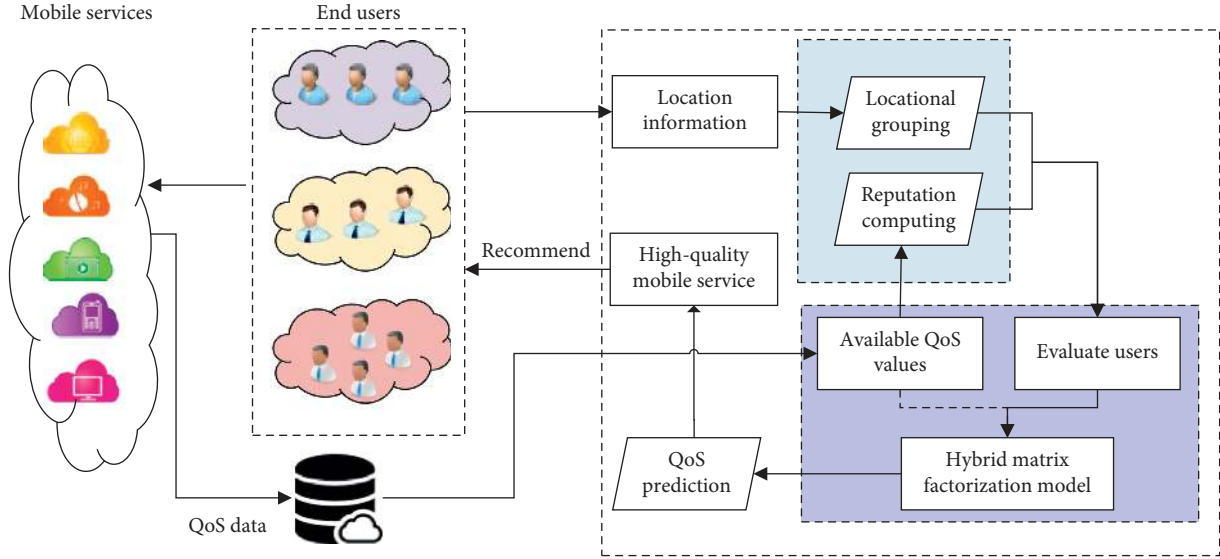
Figure 2: Mobile service recommendation framework.

the collected historical QoS dataset, and finally, we can predict missing QoS values with reputation and location information. The main workflow of LRMF is demonstrated as following:

(1) The multisource invocation records will be collected and submitted into the database when users invoke different mobile services.

(2) In the data preprocess, the available historical QoS values and users' location info will be extracted as the input data of our method.

(3) Based on the available QoS values and location data, users' reputation score can be computed by our iterative method and the location region can also be identified by their real-word location distribution. Then, the trustworthiness of users in different location region can be evaluated.

(4) A hybrid matrix factorization model is proposed by integrating location grouping and users' reputation to predict missing QoS values of unknown mobile services.

(5) Finally, by combining the predicted results and available QoS data, the high-quality services will be discovered and recommended to the target user.

3.2. User Reputation Analysis. In order to identify untrustworthy users in a given geographical region, we analyse the QoS values from users' historical service invocation records. Figure 3 demonstrates the response time of users in a same region (i.e., the United States in this example) of three randomly selected services from the real-world QoS dataset [36]. As a demonstration, Figure 4 describes the response time of 100 randomly selected services which are invoked by 5 randomly selected users in a same region.

Figure 3 shows that the response time of service invocation varies among users even they are in the same
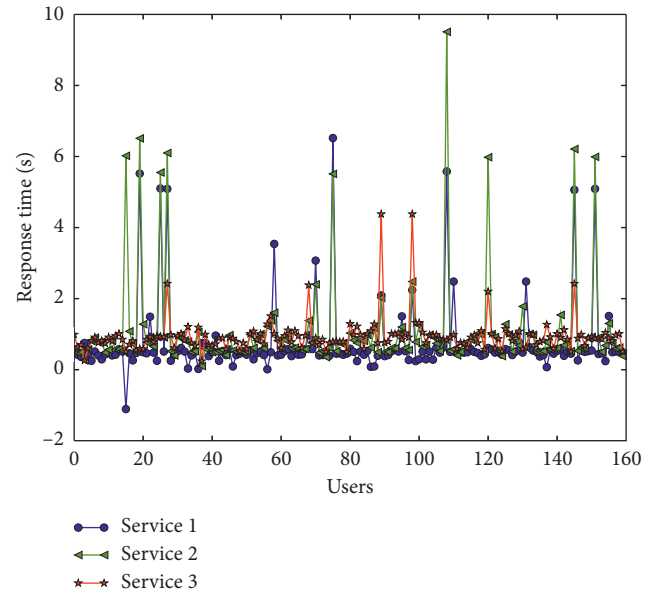

Figure 3: The distribution of response time of different users on 3 services in a location region.

location group. Although most of the response time values falls into the normal range, i.e., [0, 2], some users still submit outlier QoS records when they invoke services. It is unlikely that a QoS item would deviate from the normal value too much. For a specific user, if most QoS values significantly deviated from the normal range, he/she is probably an unreliable user. In order to test whether there are some unreliable users or not, we analyse the QoS data submitted by 5 randomly selected users on 100 services.

It is obviously in Figure 4 that the user 4 is an unreliable user because all of his submitted values deviated greatly from the normal range [0, 2]. Based on this analysis, we can compute the reputation of users by his past QoS value records and measure whether these users should be regarded
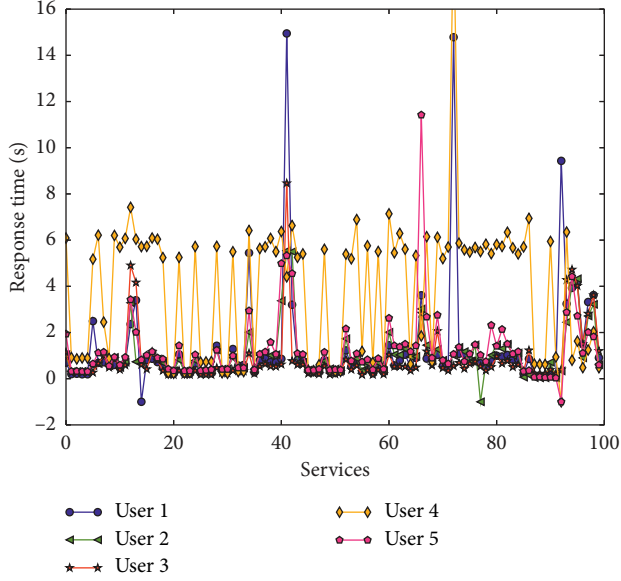
FIGURE 4: The distribution of response time submitted by 5 users on 100 services.

as unreliable users or not. In next steps, the algorithm of users' reputation computation will be introduced in detail.

4. Hybrid Matrix Factorization Based on Location and Reputation

We will firstly present the original matrix factorization approach in this part and then propose our improved matrix factorization model.

4.1. Matrix Factorization Prediction. Matrix factorization utilized low-rank approximations to fit the sparse matrix of user and item. It factorizes original sparse matrix into two low-rank matrices with small number of factors. This factorization is based on the hypothesis that users' latent preference on QoS values would be significantly affected by some latent factors. Then an objective function can be defined as the sum error of original values and the predict values by the conducts of the two low-rank matrices.

We suppose there is a QoS matrix where users in the rows and service in the columns and two low-rank matrices represent user-specific feature and service-specific feature, respectively. The QoS matrix can be regarded as a product of matrix multiplication on low-rank matrix and approximately as follows:

$$R \approx \widetilde{R} = U^T S, \tag{1}$$

where $U \in \mathbb{R}^{d \times m}$ and $S \in \mathbb{R}^{d \times n}$. d $(d \ll \min(m, n))$ is the number of latent factors.

Then, the objective function can be defended by minimizing the sum error of available values in original matrix R and the corresponding predicted values in matrix $\widetilde{R}$:

$$\min_{U,S} \psi(U, S) = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} \left\| R_{ij} - U_i^T S_j \right\|_F^2, \tag{2}$$

where R_{ij} represents the available QoS value provided by user i on service j; U_i denotes the ith row of U; S_j is j^{th} column of S. However, the available QoS values are limited in real invocations scenario, so an optimal objective function can be defined to solve this issue:

$$\min_{U,S} \psi(U, S) = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{ij} \left(R_{ij} - U_i^T S_j \right)_F^2, \tag{3}$$

where $I_{ij} = 0$ indicates the QoS value provided by user i on service j is unknown and $I_{ij} = 1$ otherwise. Two regularization terms are introduced in Equation (3) to avoid the overfitting problem as follows:

$$\min_{U,S} \psi(U, S) = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{ij} \left(R_{ij} - U_i^T S_j \right)_F^2 + \frac{\lambda_u}{2} |U|^2 + \frac{\lambda_s}{2} |S|^2, \tag{4}$$

where $\| \cdot \|_F$ is the Frobenius norm. Equation (4) is the objective function of matrix factorization approach to minimize the squared error between the predicted values and original values. The gradient decent algorithm is adopted to train our model and get a local minimum value of (4) as follows:

$$U_i' = U_i - \alpha \frac{\partial \psi}{\partial U_i}, \tag{5}$$

$$S_j' = S_j - \alpha \frac{\partial \psi}{\partial S_j}. \tag{6}$$

4.2. Locational Grouping. In this part, we identify different users by their real-world location to acquire a subset of users who have geographical similarity with the target user. Since the location information significantly affects the QoS, it should be considered a significant factor in QoS prediction [13]. We calculate users' physical distance to generate the location region. The Euclidean distance between users i on j can be computed by following definition:

$$\text{dis}(i, j) = \sqrt{(\text{lon}(i) - \text{lon}(j))^2 + (\text{lat}(i) - \text{lat}(j))^2 \times \delta}, \tag{7}$$

where $\text{lon}(i)$ and $\text{lat}(i)$ represent the longitude and latitude of user i in the real world, respectively. δ converts the unit of degree into 2D meter with a constant value. In our study, δ takes the value of 111,261.

The geographical region then could be generated by selecting a set of users who are with small distance calculated by Equation (7). On one hand, the size of this set cannot be too small; otherwise, too many similar users would be filtered. On the other hand, it cannot be too large; otherwise the different locations would not be correctly recognized. For a target user i, the region $G(i)$ can be defined as follows:

$$G(i) = \left\{ u_j | \text{dis}(i,j) \le \varepsilon, i \ne j \right\}, \tag{8}$$

where ε is a positive variable locational threshold which affects the region size, and $G(i)$ denotes the subset of users who are in the same geographical region with user i. Here, we use the real-world distance (i.e., distance based on longitude and latitude) other than the country-level type (i.e., differentiate users based on their countries).

4.3. Reputation Algorithm.

We firstly give a definition r_i as the reputation score of user i. If user i gets a higher reputation score, he/she can be considered as a reliable user. Then we propose an iterative and incorporative method to compute different users' reputation score:

$$
\begin{cases}
r_i^{k+1} = 1 - \dfrac{1}{1 + e^{-(d/|\Phi_i|)\sum_{j \in \Phi_i} \left| R_{ij} - A_j^{k+1} \right|}}, \\[3mm]
A_j^{k+1} = \dfrac{\sum_{i \in \Gamma_j} R_{ij} \times r_i^k}{\left| \Gamma_j \right|},
\end{cases}
\tag{9}
$$

where k is the kth iteration and d represents the damping factor in [0, 1], r_i^k is the kth reputation iteration of user i, Φ_i are user i's invocated services, A_j^{k+1} represents average value of service j, and Γ_j denotes users who invoked the service j. Each user is assumed to be trustworthy, and the reputation will be assigned an initial value of 1 in first step of the iteration.

From the above discussion, it is obvious that it is an iterative process to calculate users' reputation. The reputation score of each user is computed by the difference between the available QoS values in original matrix R and the average QoS values provided by the target user on all invocated services.

4.4. QoS Prediction Based on Location and Reputation.

The method in the previous section is the original matrix factorization method to predict QoS values with historical invocation records. To take full advantage of users' location data and repudiation score, a high-performance hybrid matrix factorization method is proposed as follows:

$$
\min_{U,S} \psi(U,S) = \frac{1}{2} \sum_{i=1}^{m} r_i \sum_{j=1}^{n} I_{ij} \left(R_{ij} - U_i^T S_j \right)_F^2
$$
$$
+ \frac{\gamma}{2} \sum_{i=1}^{m} \left\| U_i - \frac{\text{loc\_sim}(i,j)}{|G(i)|} \sum_{U_j \in G(i)} U_j \right\|
$$
$$
+ \frac{\lambda_u}{2} |U|^2 + \frac{\lambda_s}{2} |S|^2, \tag{10}
$$

where parameter r_i denotes the reputation score of user i. The previous studies make QoS prediction based on the premise that users in $G(i)$ suffer similar service invocations and observe similar QoS usage experience [15, 28]. However, there are some unreliable users who would provide untrustworthy QoS values and make bad impact on prediction result. Therefore, we introduce r_i into Equation (10) to regulate the credibility of different users. Meanwhile, users' location information should be introduced into our method to enhance the prediction accuracy as is mentioned in the previous section. Therefore, a locational constraint item is defined as follows:

$$
\frac{\gamma}{2} \sum_{i=1}^{m} \left\| U_i - \frac{\text{loc\_sim}(i,j)}{|G(i)|} \sum_{U_j \in G(i)} U_j \right\|, \tag{11}
$$

where $\gamma > 0$ denotes the relative proportion of the location grouping, U_i represents the latent factors of user i, $|G(i)|$ represents the subset of users who are near to user i, and $\text{loc\_sim}(i,j)$ represents the similarity between user i and user j, defined as

$$
\text{loc\_sim}(i,j) = 1 - \frac{1}{1 + e^{-\text{dis}(i,j)}}, \tag{12}
$$

where $\text{dis}(i,j)$ is the real-world distance between user i and user j, calculated by Equation (7).

The gradient parts of objective function Equation (10) could be calculated by employing the gradient descent method in U_i and S_j:

$$
\frac{\partial \Psi}{\partial U_i} = r_i \sum_{j=1}^{n} I_{ij} \left(R_{ij} - U_i^T S_j \right) \left(-S_j \right) + \lambda_u U_i
$$
$$
+ \gamma \left(U_i - \frac{\text{loc\_sim}(i,j)}{|G(i)|} \sum_{U_j \in G(i)} U_j \right), \tag{13}
$$

$$
\frac{\partial \Psi}{\partial S_j} = r_i \sum_{i=1}^{m} I_{ij} \left(R_{ij} - U_i^T S_j \right) \left(-U_i \right) + \lambda_j S_j. \tag{14}
$$

Based on the update process in Equations (5) and (6), we can update U_i and S_j with the two derivative Equations (13) and (14) until we get the local minimum of objective function Equation (10).

5. Experiments

5.1. Experiment Setup.

A series of experiments are conducted on a well-known public QoS dataset, which is provided in the previous related work [7]. The dataset contains 1,974,675 response time records of service invocation, which is collected from 339 users on 5,825 services. Figure 5 shows the distribution of all values in the dataset and nearly 90% of all response time values are in the range [0, 2]. To simulate a real-world service invocation scenario, several values of response time records are randomly removed to generate random unreliable users, and different numbers of unreliable users will be introduced into the dataset to study the impact of users' reputation on the prediction accuracy.
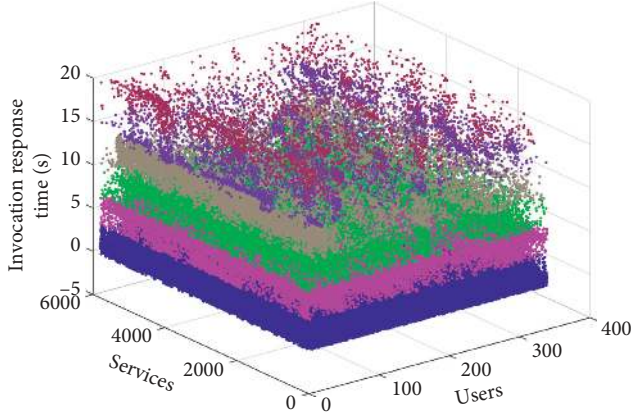
FIGURE 5: The distribution of response time values in real-world service invocation.

5.2. Evaluation Metrics.

In the experiment, we utilize two famous statistical metrics, i.e., mean absolute error (MAE) and root mean squared error (RMSE) to evaluate our predicted result. If the prediction result is closer to the actual QoS value, the smaller value of MAE and RMSE will be generated and higher prediction accuracy can be achieved. MAE is given by

$$\text{MAE} = \frac{\sum_{i,j} \left| R_{ij} - \widetilde{R}_{ij} \right|}{N}, \tag{15}$$

and RMSE is calculated as follows:

$$\text{RMSE} = \frac{\sum_{i,j} \sqrt{\left| R_{ij} - \widetilde{R}_{ij} \right|^2}}{N}, \tag{16}$$

where R_{ij} represents the available QoS value in original matrix, $\widetilde{R}_{ij}$ denotes the predicted QoS value, and N represents the number of unknown QoS values.

5.3. Comparison Study.

The proposed approach is compared with following CF approaches:

(i) UPCC. This approach utilizes PCC to calculate similarity between users [37]. In service computing researches, this approach can be employed to predict QoS values.

(ii) IPCC. This approach is a common commercial recommendation method. Service recommendation systems usually utilize this method to calculate service similarities and make prediction [38].

(iii) UIPCC. Both UPCC and IPCC are employed synchronously in the prediction framework [39].

(iv) PMF. This approach uses probability theory to explain how to use matrix factorization make prediction [11].

(v) LBR1. This approach predicts QoS values by combining geographical information and matrix factorization approach [15].

(vi) RMF. This approach utilizes available QoS values to calculate users' reputation and adopt matrix factorization model to predict missing QoS values of unknown services [32].

To simulate service invocation cases, a set number of elements are removed from the original QoS matrix. After this data preprocess, the density of the final matrix is set to 5%, 10%, 15%, and 20%, respectively. The parameter d is set 0.1 to compute users' reputation. We also set $\lambda_u = \lambda_u = 0.001$, dimensionality = 50, and $\gamma = 0.001$. The number of unreliable users is equal to 10 about 2.79 percentage of all users. The overall comparison details are presented in Tables 1 and 2.

The comparison results in Tables 1 and 2 show that the proposed LRMF can achieve higher prediction accuracy than other state-of-the-art approaches, which indicates LRMF has greatly improved QoS prediction accuracy. According to the comparison result, LRMF could achieve the best prediction performance.

5.4. Impact of Unreliable Users.

The number of unreliable users determines the untrustworthy QoS values in the training dataset, which will produce great influence on the prediction method. In the experiment, the number of unreliable users is set from 10 to 80 under the condition of dimensionality in 80 and matrix density in 5%, 10%, 15%, and 20%, respectively.

As shown in Figure 6, both MAE and RMSE values of LRMF are significantly smaller when matrix density becomes denser under different conditions. It can also be proven that more available QoS values will make better prediction result. When the number of unreliable users changes in the range of [30, 80], both MAE and RSME do not increase so much, which demonstrates that our LRMF method could minimize the negative impact produced by unreliable users and improve QoS prediction accuracy.

5.5. Impact of Parameter γ.

The parameter γ determines the proportion of the location region factor in our proposed method. For one thing, too large value of γ will create a strong relation between the prediction accuracy and the geographical region; for another, if γ is assigned too small, the location region cannot generate enough contribution to the objective function. The comparative study for γ is introduced with the condition of dimensionality = 50 and matrix density in 5% and 20%. Also, we add 10 unreliable users in experiment settings. The details of the experiment are presented in Figure 7.

As presented in Figures 7(a) and 7(b), the MAE can achieve a minimum when γ reaches a certain value 10^{-3} and when γ is smaller or larger than this specific value, the MAE will fluctuate wildly. It is similar in Figures 7(c) and 7(d), the RMSE will reach a certain threshold when γ increases to 10^{-2} at beginning, but it will increase slightly after the threshold. The analysis shows that users' location data

TABLE 1: Prediction accuracy comparison result of mean squared error (MAE).

Methods	Matrix density			
	5%	10%	15%	20%
UPCC	0.5931	0.5497	0.5195	0.4908
IPCC	0.6239	0.5901	0.5519	0.5227
UIPCC	0.5962	0.5319	0.5029	0.4749
PMF	0.5797	0.5201	0.4872	0.4621
LBR1	0.5647	0.5021	0.4709	0.4498
RMF	0.5639	0.4897	0.4641	0.4529
LRMF	**0.5514**	**0.4719**	**0.4493**	**0.4384**

TABLE 2: Prediction accuracy comparison result of root mean squared error (RMSE).

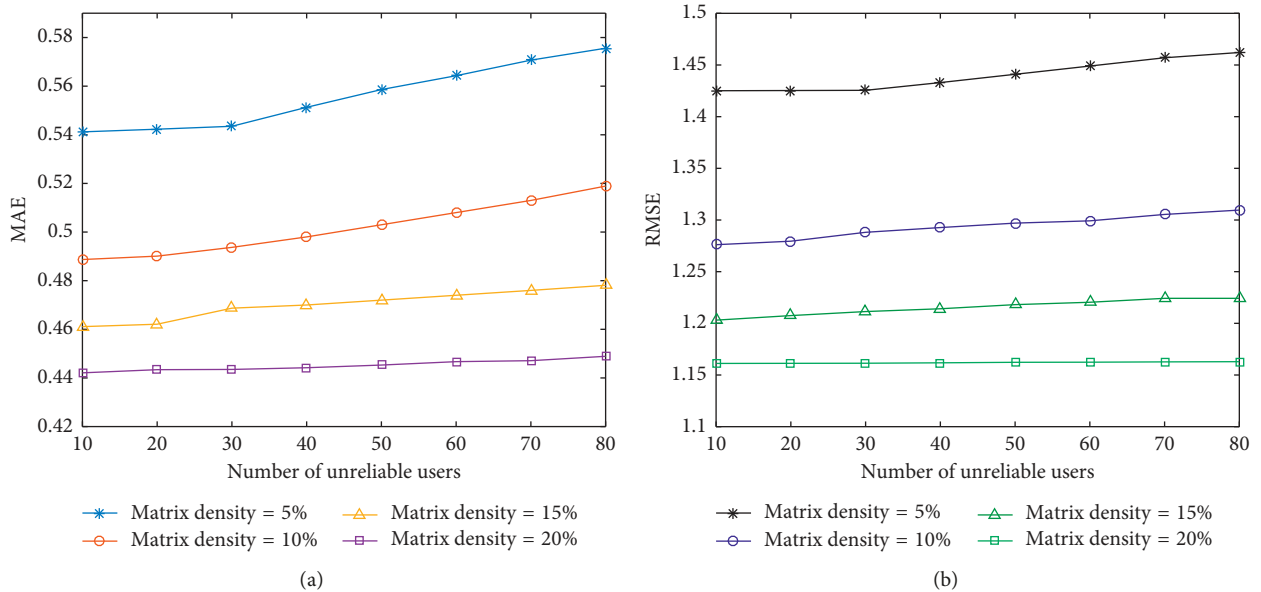| Methods	Matrix density			
	5%	10%	15%	20%
UPCC	1.4127	1.3301	1.2697	1.2492
IPCC	1.4371	1.3561	1.2898	1.2139
UIPCC	1.3969	1.3071	1.2598	1.1815
PMF	1.4481	1.2894	1.2319	1.1839
LBR1	1.4431	1.2891	1.2208	1.1641
RMF	1.4372	1.2792	1.2105	1.1621
LRMF	**1.4152**	**1.2584**	**1.1956**	**1.1498**

FIGURE 6: Impact of the number of unreliable users.

would make a remarkable impact on the prediction accuracy of LRMF.

5.6. Impact of Parameter ε. In LRMF, the location group threshold ε determines the geographical region size. If ε is assigned too small, the region would be very small and users in the region would have a very short distance. If ε is assigned too large, much more users would be identified into the same geographical region, which would lead to more noise data and neglect the local factor.

To study how parameter ε impacts on the prediction approach, the value of dimensionality is assigned 50 under the condition of matrix density in 10% and 30%. We also add 10 unreliable users to experiment settings.

Figure 8 illustrates how parameter ε affects the prediction accuracy. It is obvious that the evaluation values decrease when parameter ε increases firstly, but both two kinds of evaluation values gradually increase when parameter ε passes over a threshold. The observed phenomenon could be considered as when parameter ε is smaller
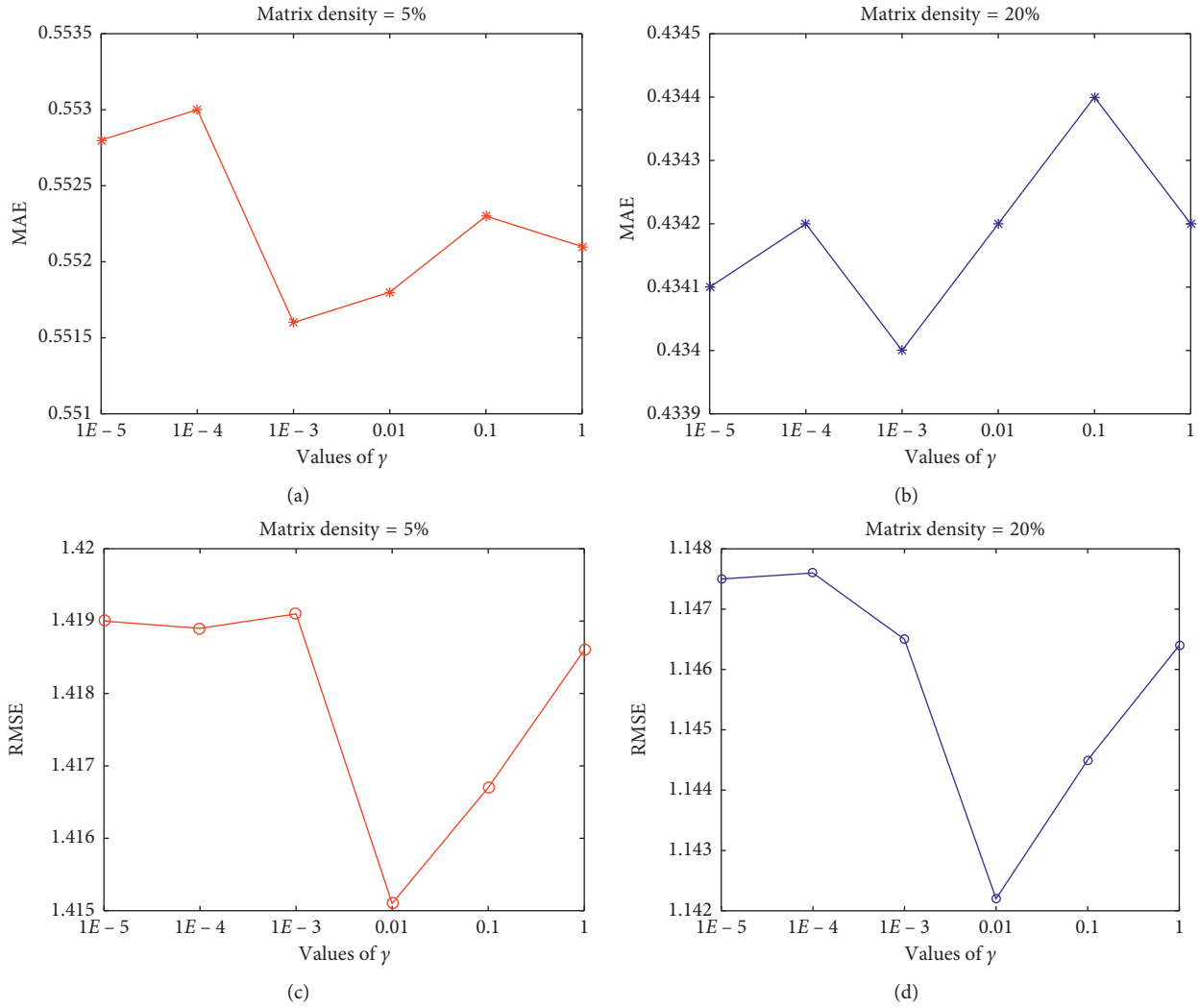
FIGURE 7: Impact of parameter γ.

than a certain value, the location region lacks enough similar users for active user, which prevents the crowds to contribute their collective intelligence. When parameter ε is assigned to be a large value, too much noise data will be included in the location region. Both two cases would produce negative impacts on the prediction curacy.

5.7. Impact of Density. To study what impacts does matrix density have on the prediction result, we add 10 unreliable users and set dimensionality to 10 and 50 when assigning matrix density from 5% to 20%.

As shown in Figure 9, both MAE and RMSE decrease firstly when the matrix density increases at first. Then, the curve becomes flat when the matrix density continues to increase. These comparative details show that the sparsity of original data would have great impact on the prediction accuracy. If more additional entries are available, the proposed method could get better prediction result. This observation demonstrates that when the original sparse

matrix becomes denser by collecting more QoS values, the prediction accuracy can be greatly enhanced in our proposed method.

5.8. Impact of Dimensionality. The number of latent feature vectors is regulated by dimensionality in our LRMF method. We vary the dimensionality from 10 to 100 under the condition of matrix density in 5% and 10 and 30 unreliable users, respectively, to conduct the comparative experiments.

Figure 10 shows that a proper value of dimensionality can achieve better prediction result as we can get a smallest prediction error. Both of MAE and RMSE decrease firstly because of more latent factors are added into the factorization process. When the dimensionality overpasses a certain threshold, more noised data may be brought into the training model with overfitting problem. As a result, the threshold of dimensionality in our model is approximately assigned 80.
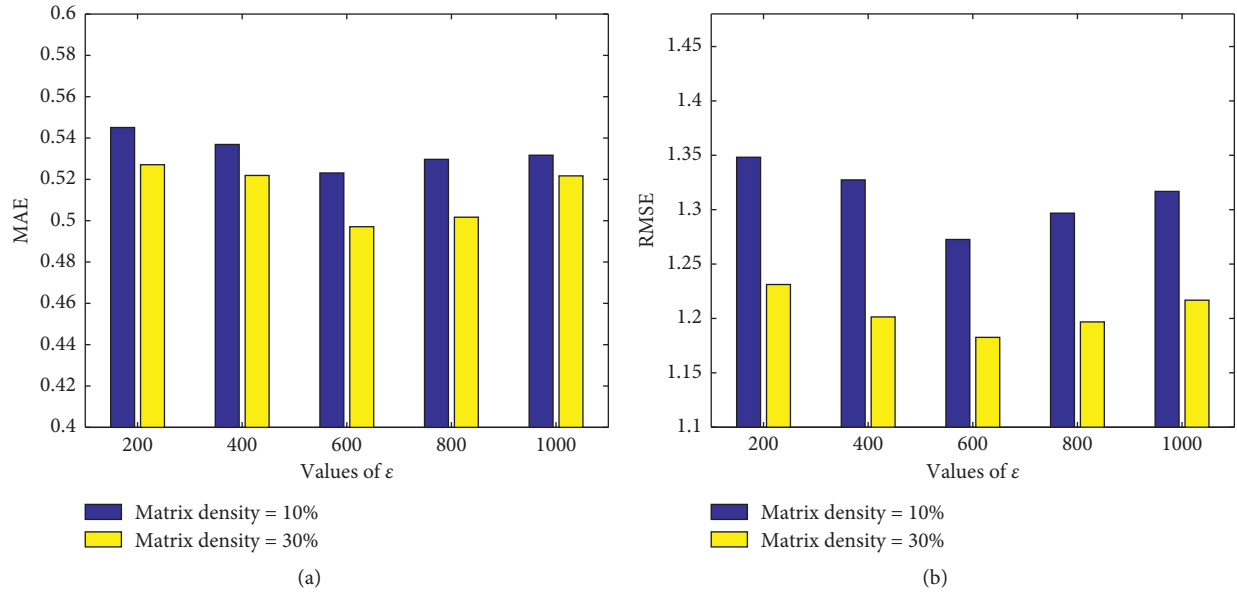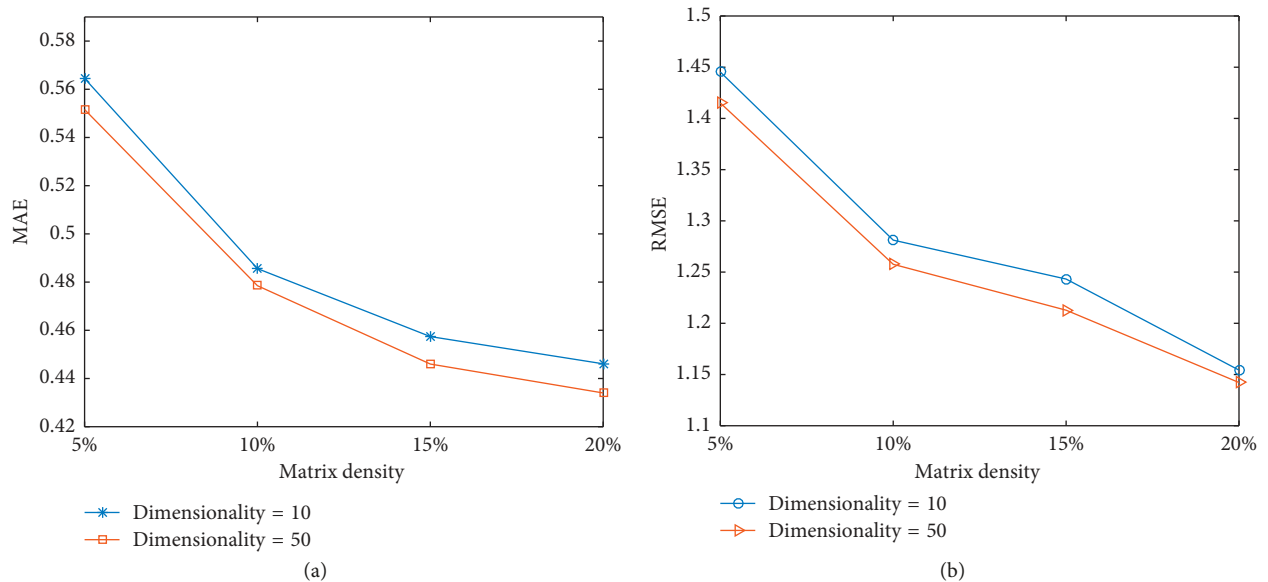
FIGURE 8: Impact of parameter ε.

FIGURE 9: Impact of matrix density.

6. Conclusions and Future studies

Based on the premise that the reputation of mobile service users would make great effect on the unknown QoS prediction, this paper propose an efficient prediction method by simultaneously exploring users' reputation and geographic distribution to make personalized service recommendation. We firstly cluster service users into different locational groups by their real-world geographical information and then calculate their reputations by the historical QoS values. At last, a hybrid matrix factorization model is proposed by integrating users' reputation and geographic data to predict

unknown QoS values. Experimental analysis on public QoS dataset demonstrates the high-performance and effectiveness of our LRMF on QoS prediction. The analysis shows that there are some unreliable uses in some location region and they submitted untrustworthy QoS values to gain benefits for their own services. The mobile service recommendation approach proposed in this study could reduce the poor effect of unreliable users and recommend high-quality and credible mobile service to end users.

In this paper, we only consider users' information as a significant issue to predict unknown QoS values for mobile service recommendation. In fact, the geographical distribution
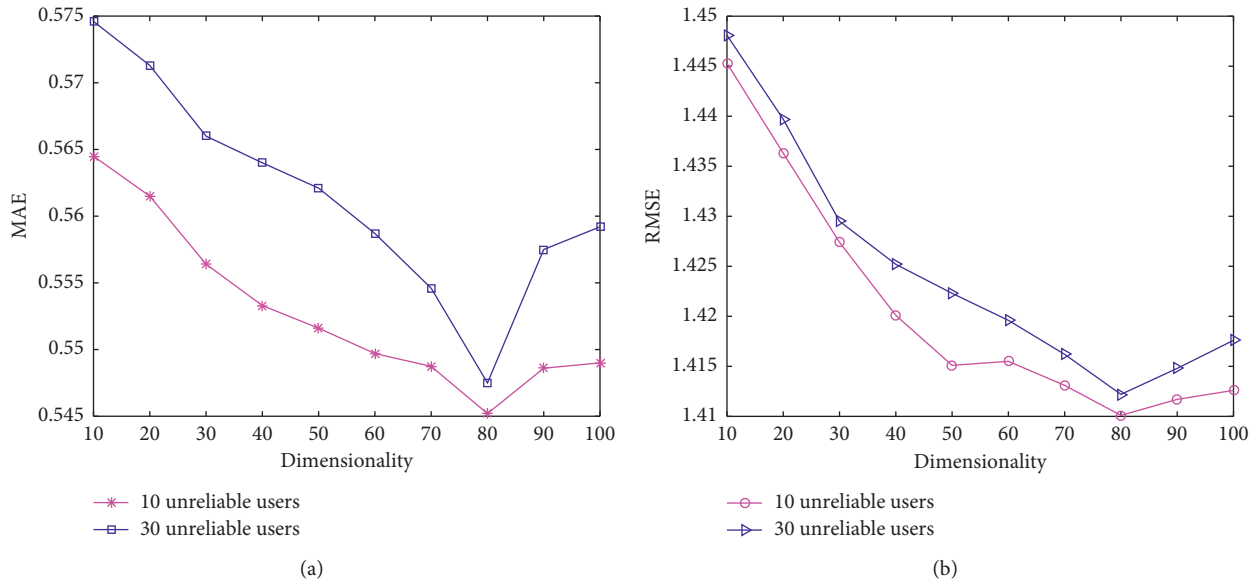
Figure 10: Impact of dimensionality.

of services would also provide useful facilities when identifying users' location region based on the distance between user and service. Therefore, it is potential to design more accurate location grouping model by combining both users' and service's location information. Besides, if the algorithms of users' reputation computing have deficiencies, then we would try to introduce intelligence methods, e.g., deep learning, reinforcement learning to design optimal algorithms for QoS prediction. Furthermore, the reliability of user may be affected by their trusted friends, so we will continue to track users' reputation in their social network to make high-quality service recommendation in our future work.

Data Availability

The real number data WS-DREAM used to support the findings of this study are included within the article.

Conflicts of Interest

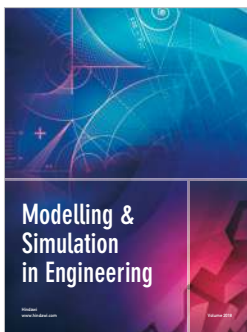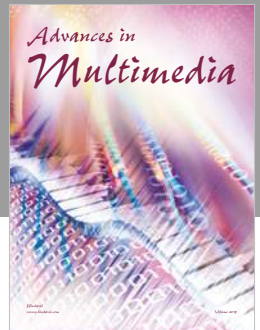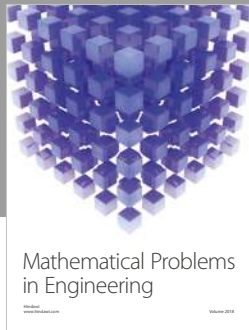The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] Y. Yin, W. Xu, Y. Xu, H. Li, and L. Yu, "Collaborative QoS prediction for mobile service with data filtering and SlopeOne model," *Mobile Information Systems*, vol. 2017, Article ID 7356213, 14 pages, 2017.

[2] L. Li, S. Li, and S. Zhao, "Qos-aware scheduling of services-oriented internet of things," *IEEE Transactions Industrial Informatics*, vol. 10, no. 2, pp. 1497–1505, 2014.

[3] Y. Ma, S. Wang, P. C. K. Hung, C. H. Hsu, Q. Sun, and F. Yang, "A highly accurate prediction algorithm for unknown web service QoS values," *IEEE Transactions on Services Computing*, vol. 9, no. 4, pp. 511–523, 2016.

[4] L. Xin, M. C. Zhou, L. Shuai, Y. N. Xia, Z. H. You, and Q. S. Zhu, "Incorporation of efficient second-order solvers into latent factor models for accurate prediction of missing qos data," *IEEE Transactions on Cybernetics*, vol. 48, no. 4, pp. 1216–1228, 2017.

[5] N. Xi, D. Lu, C. Sun, J. Ma, and Y. Shen, "Distributed secure service composition with declassification in mobile clouds," *Mobile Information Systems*, vol. 2017, Article ID 7469342, 13 pages, 2017.

[6] S. Wang, L. Sun, Q. Sun, X. Li, and F. Yang, "Efficient service selection in mobile information systems," *Mobile Information Systems*, vol. 2015, Article ID 949436, 10 pages, 2015.

[7] Z. Zheng, H. Ma, M. R. Lyu, and I. King, "Qos-aware web service recommendation by collaborative filtering," *IEEE Transactions on Services Computing*, vol. 4, no. 2, pp. 140–152, 2010.

[8] M. Silic, G. Delac, and S. Srbljic, "Prediction of atomic web services reliability for QoS-aware recommendation," *IEEE Transactions on Services Computing*, vol. 8, no. 3, pp. 425–438, 2015.

[9] S. Li, J. Wen, F. Luo, M. Gao, J. Zeng, and Z. Y. Dong, "A new qos-aware web service recommendation system based on contextual feature recognition at server-side," *IEEE Transactions on Network and Service Management*, vol. 14, no. 2, pp. 332–342, 2017.

[10] G. Kang, J. Liu, M. Tang, B. Cao, and Y. Xu, "An effective web service ranking method via exploring user behavior," *IEEE Transactions on Network and Service Management*, vol. 12, no. 4, pp. 554–564, 2015.

[11] A. Mnih and R. Salakhutdinov, "Probabilistic matrix factorization," in *Proceedings of Advances in Neural Information Processing Systems 2007*, pp. 1257–1264, Kitakyushu, Japan, December 2007.

[12] M. Tang, Z. Zheng, G. Kang, J. Liu, Y. Yang, and T. Zhang, "Collaborative web service quality prediction via exploiting matrix factorization and network map," *IEEE Transactions on Network and Service Management*, vol. 13, no. 1, pp. 126–137, 2017.

[13] X. Chen, X. Liu, Z. Huang, and H. Sun, "Regionknn: a scalable hybrid collaborative filtering algorithm for personalized web service recommendation," in *Proceedings of IEEE 17th International Conference on Web Services (ICWS'10)*, pp. 9–16, Miami, FL, USA, July 2010.

[14] http://www.google.com/transparencyreport/.

[15] L. Wei, J. Yin, S. Deng, Y. Li, and Z. Wu, "Collaborative web service QoS prediction with location-based regularization," in *Proceedings of IEEE 19th International Conference on Web Services (ICWS'12)*, pp. 464–471, Honolulu, HI, USA, June 2012.

[16] I. Gunes, C. Kaleli, A. Bilge, and H. Polat, "Shilling attacks against recommender systems: a comprehensive survey," *Artificial Intelligence Review*, vol. 42, no. 2, pp. 767–799, 2014.

[17] E. Al-Masri and Q. H. Mahmoud, "Qos-based discovery and ranking of web services," in *Proceedings of IEEE 16th International Conference on Communications and Networks (ICCCN'07)*, pp. 529–534, Honolulu, HI, USA, August 2007.

[18] K. Kritikos and D. Plexousakis, "Requirements for QoS-based web service description and discovery," *IEEE Transactions on Services Computing*, vol. 2, no. 4, pp. 320–337, 2009.

[19] S. Rosario, A. Benveniste, S. Haar, and C. Jard, "Probabilistic QoS and soft contracts for transaction-based web services orchestrations," *IEEE Transactions on Services Computing*, vol. 1, no. 4, pp. 187–200, 2009.

[20] J. E. Hadad, M. Manouvrier, and M. Rukoz, "TQoS: transactional and QoS-aware selection algorithm for automatic web service composition," *IEEE Transactions on Services Computing*, vol. 3, no. 1, pp. 73–85, 2010.

[21] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," *IEEE Internet Computing*, vol. 7, no. 1, pp. 76–80, 2003.

[22] D. A. Adeniyi, Z. Wei, and Y. Yongquan, "Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method," *Applied Computing and Informatics*, vol. 12, no. 1, pp. 90–108, 2016.

[23] Y. Zuo, J. Zeng, M. Gong et al., "Tag-aware recommender systems based on deep neural networks," *Neurocomputing*, vol. 204, pp. 51–60, 2016.

[24] B. Engelbert, M. B. Blanken, R. Kruthoff-Bruwer, and K. Morisse, "A user supporting personal video recorder by implementing a generic Bayesian classifier based recommendation system," in *Proceedings of IEEE 19th International Conference on Communications and Networks (ICCCN'11), Workshops*, pp. 567–571, Seattle, WA, USA, March 2011.

[25] C. X. Zhang, Z. K. Zhang, L. Yu, C. Liu, H. Liu, and X. Y. Yan, "Information filtering via collaborative user clustering modeling," *Physica A Statistical Mechanics and Its Applications*, vol. 396, no. 2, pp. 195–203, 2014.

[26] H. Ma, H. Yang, M. R. Lyu, and I. King, "SoRec:social recommendation using probabilistic matrix factorization," in *Proceedings of ACM Conference on Information and Knowledge Management*, pp. 931–940, ACM, Napa Valley, CA, USA, November 2008.

[27] N. Ali and C. Solis, "Self-adaptation to mobile resources in service oriented architecture," in *Proceedings of 2015 IEEE International Conference on Mobile Services*, pp. 407–414, New York, NY, USA, June 2015.

[28] M. Tang, T. Zhang, J. Liu, and J. Chen, "Cloud service qos prediction via exploiting collaborative filtering and location-based data smoothing," *Concurrency and Computation Practice and Experience*, vol. 27, no. 18, pp. 5826–5839, 2015.

[29] K. Lee, J. Park, and J. Baik, "Location-based web service QoS prediction via preference propagation for improving cold start problem," in *Proceedings of IEEE 22nd International Conference on Web Services (ICWS '15)*, pp. 400–407, New York, NY, USA, June 2015.

[30] B. Gonsalves and V. Patil, "Improved web service recommendation via exploiting location and QoS information," in *Proceedings of International Conference on Information Communication and Embedded Systems IEEE*, pp. 1–5, Boca Raton, FL, USA, December 2016.

[31] W. Qiu, Z. Zheng, X. Wang, X. Yang, and M. R. Lyu, "Reputation-aware QoS value prediction of web services," in *Proceedings of the IEEE 10th International Conference on Service Computing (SCC'13)*, pp. 41–48, Santa Clara, CA, USA, June–July 2013.

[32] J. Xu, Z. Zheng, and M. R. Lyu, "Web service personalized quality of service prediction via reputation-based matrix factorization," *IEEE Transactions on Reliability*, vol. 65, no. 1, pp. 28–37, 2015.

[33] M. Mehdi, N. Bouguila, and J. Bentahar, "Trust and reputation of web services through QoS correlation lens," *IEEE Transactions on Services Computing*, vol. 9, no. 6, pp. 968–981, 2016.

[34] A. Comi, L. Fotia, F. Messina, G. Pappalardo, and D. Rosaci, "A reputation-based approach to improve QoS in cloud service composition," in *Proceedings of IEEE, International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises*, pp. 108–113, Paris, France, June 2015.

[35] K. K. Fletcher and X. F. Liu, "A collaborative filtering method for personalized preference-based service recommendation," in *Proceedings of IEEE 22nd International Conference on Web Services (ICWS'15)*, pp. 400–407, New York, NY, USA, June 2015.

[36] Z. Zheng and M. R. Lyu, *QoS Management of Web Services*, Springer, Berlin, Germany, 2013.

[37] L. H. Son, "H-FCF: a hybrid user-based fuzzy collaborative filtering method in recommender systems," *Expert Systems with Applications*, vol. 41, no. 15, pp. 6861–6870, 2014.

[38] K. Y. Chung, D. Lee, and K. J. Kim, "Categorization for grouping associative items mining in item-based collaborative filtering," *Multimedia Tools and Applications*, vol. 71, no. 2, pp. 889–904, 2014.

[39] M. Papagelis and D. Plexousakis, "Qualitative analysis of user-based and item-based prediction algorithms for recommendation agents," *Engineering Applications of Artificial Intelligence*, vol. 18, no. 7, pp. 781–789, 2005.