Exploring the Effectiveness of True Abnormal Data Elimination in Context-aware Web Services Recommendation

Xiaoliang Fan, Yujie Wang, Yakun Hu
School of Information Science and Engineering
Lanzhou University
Lanzhou, China
{fanxiaoliang,huyk14}@lzu.edu.cn; wangyujie15@lzu.cn

You Ma
National Satellite Meteorological Center, China Meteorological Administration
Beijing, China
mayou@cma.gov.cn

Xiao Liu
School of Information Technology
Deakin University
Melbourne
Australia
xiao.liu@deakin.edu.au

Abstract—**Recent years have witnessed a growing interest in context-aware recommender system (CARS), which explores the impact of context factors on personalized Web services recommendation. Basically, the general idea of CARS methods is to mine historical service invocation records through the process of context-aware similarity computation. It is observed that traditional similarity mining process would very likely generate relatively big deviations of QoS values, due to the dynamic change of contexts. As a consequence, including a considerable amount of deviated QoS values in the similarity calculation would probably result in a poor accuracy for predicting unknown QoS values. In allusion to this problem, this paper first distinguishes two definitions of Abnormal Data and True Abnormal Data, the latter of which should be eliminated. Second, we propose a novel CASR-TADE method by incorporating the effectiveness of True Abnormal Data Elimination into context-aware Web services recommendation. Finally, the experimental evaluations on a real-world Web services dataset show that the proposed CASR-TADE method significantly outperforms other existing approaches.**

Keywords- context awareness; Web services recommendation; QoS; true abnormal data elimination; QoS prediction.

I. INTRODUCTION

With the explosive growth of Web services published on the Internet, users are facing with massive and functionally equivalent Web services for both effective and efficient services recommendation [1]. For most recommender systems, Quality of Services (QoS) is designed to represent the non-functional characteristics of Web services [2], such as response time, throughput, and invocation failure rate. The most popular method is collaborative filtering (CF), which usually conducts Pearson Correlation Coefficient (PCC) measurement for user/item-based similarity computation with the QoS matrix [7]. Specifically, the user-dependent QoS properties generally are tied to various context factors such as network distance, bandwidth, user client performance, etc. Under this circumstance, context-aware recommender system (CARS) that take context factors of users into account has aroused a great deal of interests in recent years [8, 10].

Basically, the objective of CARS methods is to mine historical service invocation records under the similar contexts for a specific user [20]. For example, the more similar between the current user and another user's context (e.g., the location of the user), the higher probability of the two users will have similar QoS on the same Web service. However, such effect of context similarity is not well discussed by the traditional CF-based methods. In short, the CARS method is generally superior to traditional CF-based methods for its better QoS prediction accuracy as well as lower computational complexity.

As user contexts especially the network environments are dynamic, the user-dependent QoS properties often change accordingly. Therefore, uncommon changes of user contexts (e.g. bandwidth, DNS delay and client load) may result in a considerable amount of deviations of QoS values. In order to explore the effect of abnormal QoS values, we made summary statistics for response times in the WS-Dream dataset[1]. The mean value and the standard deviation of response times is 0.3435 and 0.741, respectively.

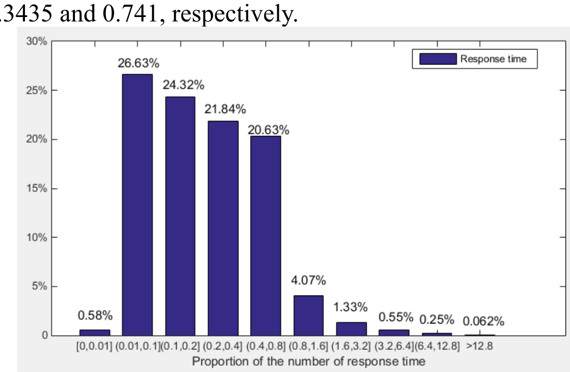

Figure 1. Proportion of the Differentiate QoS values (response times, 150 users×150 services)

The maximum and the minimum value of response times in WS-Dream dataset vary significantly from 0.003 to 19.61

1 WS-Dream dataset, http://www.wsdream.net/dataset.html

978-1-5090-2675-3/16 $31.00 © 2016 IEEE
DOI 10.1109/ICWS.2016.46

second. As shown in Fig. 1, we observed that the proportion of [>12.8] is 0.062%, but the value of such proportion is ten times than the mean value. In another word, potentially poor prediction accuracy would be expected, when deviated QoS values are used to calculate the similarity among users for predicting unknown QoS values. Certainly, abnormal QoS values could also come from abnormal context changes of server clients. This issue has been explored by our previous work [19]. However, the problem of abnormal QoS values from user clients has still not been well solved in both CF-based and CARS methods. Deviated QoS values usually account for a small fraction of total values, but they would have a great impact on the accuracy of QoS prediction. To clearly explain the impact of abnormal data, we distinguish two definitions as follows.

Definition 1: Abnormal Data –This represents those QoS values that have big deviations with the average value of all QoS values in the dataset.

Definition 2: True Abnormal Data –For item-based similarity computation, this represents each pairs of two QoS values of two "*similar*" services that have a big deviation by a specific user. Here, "*similar*" indicates that QoS values of those two services invoked by most other users are approximately equal. The analogous definition could be made for user-based similarity computation.

TABLE I. DATA SAMPLES OF RESPONSE TIME VALUES IN WS-DREAM

	WS_1	WS_2	WS_3	WS_4	WS_5	WS_6	WS_7
U_8	**0.247**	**0.257**	**0.243**	**0.266**	**0.921**	**0.336**	**0.424**
U_9	**0.212**	**0.217**	**0.207**	**0.2**	**0.716**	**0.438**	**0.376**
U_{10}	**0.207**	**0.221**	**0.2**	**0.199**	**0.626**	**0.389**	**0.34**
U_{11}	**0.278**	**0.294**	**0.264**	**0.278**	**1.015**	***5.26***	$\boxed{\mathbf{0.477}}$

As shown in Table I, the average response time of four Web services $\{WS_1, WS_2, WS_3, WS_4\}$ invoked by four users $\{U_8,\ U_9,\ U_{10},\ U_{11}\}$ under similar contexts is around 0.2 seconds. The observations from Table I are: 1) four response times of WS_5 are not true abnormal because they are all bigger than the average value due to the poor QoS of WS_5; 2) there is a relatively big deviation for the QoS value of WS_6 invoked by U_{11} (i.e., *5.26*), compared with $\{U_8,\ U_9,\ U_{10}\}$. In order to avoid poor prediction accuracy before each similarity computation we hypothesize that this deviation is the True Abnormal Data that we should eliminate each pair of U_{11} with other three users QoS values of WS_6; and 3) to predict the unknown QoS values of WS_7 invoked by U_{11}, we only use the historical invocation records from $\{WS_1, WS_2, WS_3, WS_4, WS_5\}$, excluding all QoS value of WS_6. By eliminating the True Abnormal Data, we finally obtain the prediction value of 0.4982 that is close to the real value (0.477 in Table I) and is better than the traditional CF prediction (i.e., 0.3639).

In order to make personalized QoS prediction accurately, we believe that it is essential to identify and then eliminate those true abnormal QoS values before the similarity computation. However, due to both large deviations among users' historical invocation records and dynamic changes of contexts as time goes on, it is quite difficult to identify true abnormal values from all possible abnormal QoS values. In this study, we hypothesize that the accuracy of Web services recommendation could be improved if incorporating the true abnormal data elimination effect into the context-aware Web services recommendation. The major contributions of the paper are threefold:

- Based on real Web service QoS data and a number of experiments, we distinguished two definitions of *Abnormal Data* and *True Abnormal Data*. The latter definition represents those QoS values that should be eliminated before the similarity computation.
- We proposed a novel CASR-TADE method to incorporate the effect of True Abnormal Data Elimination with context-aware Web services recommendation, allowing the unknown QoS values to be predicted accurately.
- Employing a real-world Web service dataset, we conducted a set of comprehensive experiments, which proved that the proposed method outperforms existing well-known approaches significantly.

Hereafter, the paper is organized as follows. Section II introduces the related works. Section III proposes the CASR-TADE method. Section IV shows the implementation, experiment and discussions. Finally, the general conclusion and perspectives in Section V concludes the paper.

II. RELATED WORK

Collaborative filtering (CF) is considered as a method to make predictions automatically and collect preferences from many collaborating users based on the target user's interests [3]. CF techniques usually consist of two types [4]: model-based and memory-based CF. Memory-based CF could further be classified into item-based [5] and user-based [6], respectively.

Similarity computation measurements, such as Pearson Correlations Coefficient (PCC) and Cosine Similarity, have been widely applied in QoS prediction and Web services recommendation [7]. Basically, the objective of PCC methods is to mine historical services invocation experiences under the same or similar context for a specific user. Furthermore, various extensions such as UPCC [16] and IPCC [17] were proposed. Ma et al. [19] proposed a Web service QoS value prediction method HAPA to realize the characteristics of objective QoS datasets, allowing the unknown QoS values to be predicted accurately.

However, existing PCC-based similarity measurements assume that the influence of abnormal QoS values on the target user are equal, thus it may probably result in poor prediction accuracy. We believe it is essential to treat deviations of QoS values differently from the normal values. Specifically, the extremely deviated values, considered as *True Abnormal Data* defined in Section I, should be eliminated before the similarity calculation.

Context-aware recommender system (CARS) has been widely employed to explore the significant role of contextual factors for personalized recommendation over the years [8]. Preliminary benefits have been seen in recommending Web services when taking contextual factors into account. Specifically, temporal [9, 11, 12, 20], spatial [13, 18] and social [14] contexts are extracted separately for personalized QoS-based Web services recommendation.

In fact, QoS attributes of Web services, such as response times, rely much on the contextual information, such as network conditions, availability of services, etc. For example, the unbearable long response time is probably due to the network failure where the targeted service is deployed [2]. However, such effect of context similarity (i.e., where and when a user invokes a service) is not fully explored by existing CF-based Web services recommendation methods.

III. CASR-TADE Method

In this section, we first introduce the problem definition in Section III.A. Then, we present the four steps of proposed Context-aware Services Recommendation based on True Abnormal Data Elimination (CASR-TADE) method in Section III.B, III.C, III.D, and III.E, separately.

A. Problem Definition

Suppose a context-aware Web services recommender system contains a set of users $\mathrm{U} = \{\mathrm{u}_1, \mathrm{u}_2, \ldots, \mathrm{u}_\mathrm{n}\}$ and a set of Web services $S = \{s_1, s_2, \ldots, s_m\}$. $u_i (1 \le i \le n)$ denotes a user and he must have invoked a Web service $s_j (1 \le j \le m)$ from S at least once.

$I_{u_i, u_j} = \{s_1, s_2, \ldots, s_k\}$ is a set of commonly invoked Web services by user u_i and u_j.

$U_{s_i s_j} = \{u_1, u_2, \ldots, u_k\}$ is a set of users invoking Web services s_i and s_j.

$R = \{r_{u_i, s_k}\}$ is the set of QoS values of Web service s_k by the user u_i, where $(1 \le i \le n)$ and $(1 \le k \le m)$.

$M = \{r_{u_1,s_k}, r_{u_2,s_k}, \ldots, r_{u_i,s_k};\ r_{u_1,s_l}, r_{u_2,s_l}, \ldots, r_{u_i,s_l}\}$ or $\{r_{u_i,s_1}, r_{u_i,s_2}, \ldots, r_{u_i,s_k}; r_{u_j,s_1}, r_{u_j,s_2}, \ldots, r_{u_j,s_k}\}$ is a matrix of users invoking Web services s_k and s_l or commonly invoked Web services by user u_i and u_j, respectively.

$\hat{R} = \{\bar{r}_1, \bar{r}_2, \ldots, \bar{r}_i, \ldots, \bar{r}_n\}$ is the set of predicted value of all Web services invoked by $\mathrm{U} = \{\mathrm{u}_1, \mathrm{u}_2, \ldots, \mathrm{u}_\mathrm{n}\}$.

$\mathrm{Q} = \{\mathrm{q}_1, \mathrm{q}_2, \ldots, \mathrm{q}_\mathrm{k}\}$ is a set of QoS properties recording a service invocation.

When a user u_j invoke a service s_k, it presents a set of QoS properties. We will have $\mathrm{Q}_{\mathrm{i,j}} = < \mathrm{q}_1^{\mathrm{i,j}}, \mathrm{q}_2^{\mathrm{i,j}}, \ldots, \mathrm{q}_\mathrm{k}^{\mathrm{i,j}} >$, which is a *l*-tuple denoting service invocation records of s_k invoked by the user u_j, where $q_l^{i,j}$ $(1 \le i \le n, 1 \le j \le m)$ denotes the value of *l-th* property recorded during the invocation of s_i called by u_j.

B. Context-aware Similarity Mining and True Abnormal Data Elimination

1) Context-aware Similarity Mining

In this step, we conduct a context-aware similarity mining to find similar users or services with the current user or services.

We use the location of users or services as the context factor. We filter the users or services of similar context to structure similar users or services dataset, respectively. For example, all American user will be composed a subset.

2) True Abnormal Data Elimination

Similarity measurement methods, such as Pearson Correlation Coefficient (PCC) and its variants [16, 17], are widely used in Web services recommendation. However, those similarity models mostly ignore the impact of true abnormal data, which will downgrade the accuracy of similarity computation.

Inspired by existing mathematic methods of data elimination, we propose an approach to identify and then eliminate true abnormal QoS values. First, we establish a linear regression model of QoS value of Web service s_i and similar service s_j by a set of users U. The linear regression model is defined as:

$$\begin{cases} b_1 = \dfrac{\sum_{v \in U} r_{v,s_i} \cdot r_{v,s_j} - |U| \cdot \bar{r}_{s_i} \cdot \bar{r}_{s_j}}{\sum_{v \in U} r_{v,s_j}^2 - |U| \cdot \bar{r}_{s_j}^2} \\ b_0 = \bar{r}_{s_i} - b_1 \cdot \bar{r}_{s_j} \end{cases} \quad (2)$$

Where b_0 and b_1 represent two coefficients of the linear relationship between Web service s_i and similar service s_j respectively. r_{v,s_i} and r_{v,s_j} represent the QoS values (i.e., response time) of the Web service. s_i and s_j invocated by the user v. $\bar{r}_{s_i}$ and $\bar{r}_{s_j}$ describe the average QoS value of two users who invocate the Web services s_i and s_j, respectively. Using such two coefficients, we obtain a rough QoS prediction of r_{v,s_i} as:

$$Rp(r_{v,s_i}) = b_1 \cdot r_{v,s_j} + b_0 \quad (3)$$

By $Rp(r_{v,s_i})$, we can determine a set of abnormal values r_{v,s_i}. While $|r_{v,s_i} - Rp(r_{v,s_j})| \ge nS$ means that r_{v,s_i} has a big deviation with $Rp(r_{v,s_j})$, because s_i and s_j are linearly correlated according to the Equation (3). Thus r_{v,s_i} and r_{v,s_j} are identified as two true abnormal values which could be eliminated. In addition, n is a constant coefficient. In Section IV.D, we will discuss the influence on the results of different *n*, in order to find the optimal value of *n*. And S is the standard deviation of service s_i, which can be determined by the *Bessel Formula* as follows:

$$S = \sqrt{\frac{\sum_{v \epsilon U} (r_{v,s_i} - b_1 \cdot r_{v,s_j} - b_0)^2}{|U| - 2}} \quad (4)$$

Through the steps above, we have eliminated a pair of true abnormal QoS values. Then we employ this method iteratively, until all true abnormal values are identified and then eliminated. Finally, we obtain a new matrix M_{item} (i.e., a variation of the matrix M) to predict the unknown value of QoS before the similarity computation.

C. Item-based CASR-TADE

Suppose r_{u,s_i} is the QoS value of service s_i invoked by the user u which is an unknown value. Let x represent the predicted value of r_{u,s_i}, but r_{u,s_j} is known value.

The PCC between service s_i and s_j is calculated as:

$$sim(s_i, s_j) = \frac{\sum_{v \in U}(r_{v,s_i} - \bar{r}_{s_i})(r_{v,s_j} - \bar{r}_{s_j})}{\sqrt{\sum_{v \in U}(r_{v,s_i} - \bar{r}_{s_i})^2}\sqrt{\sum_{v \in U}(r_{v,s_j} - \bar{r}_{s_j})^2}} \tag{5}$$

We add r_{u,s_i} and r_{u,s_j} into the matrix M_{item} in Section III.B. Then we re-calculate the similarity between s_i and s_j. Let the new similarity be $sim'(s_i, s_j)$. If the $sim(s_i, s_j)$ is considerable high, then it will fluctuate slightly even if more invocations are added by other users. As a result, $sim'(s_i, s_j)$ should be approximately the same with $sim(s_i, s_j)$[19]. Two similarities can be presented as:

$$sim'(s_i, s_j) = sim(s_i, s_j) \tag{6}$$

Where

$$sim'(s_i, s_j) = \frac{\sum_{v \in U'}(r_{v,s_i} - \bar{r}'_{s_i})(r_{v,s_j} - \bar{r}'_{s_j})}{\sqrt{\sum_{v \in U'}(r_{v,s_i} - \bar{r}'_{s_i})^2}\sqrt{\sum_{v \in U'}(r_{v,s_j} - \bar{r}'_{s_j})^2}} \tag{7}$$

where $U' = U \cup \{u\}$. $\bar{r}'_{s_i}$ and $\bar{r}'_{s_j}$ are the average QoS values of service s_i and s_j invoked by different users. r'_{s_i} is a function of x, which can be written as:

$$\bar{r}'_{s_i} = \frac{\bar{r}_{s_i}|U| + x}{|U| + 1} \tag{8}$$

With (5), (6), (7) and (8), the equation (9) about x could be represented as:

$$\frac{a \cdot x + b}{c\sqrt{d \cdot x^2 + e \cdot x + f}} = sim(s_i, s_j) \tag{9}$$

Equation (9) is an item-based QoS prediction equation, and we use linear regression to determine the optimal root of it. The linear regression models are defined as:

$$b_1 = \frac{\sum_{v \in U} r_{v,s_i} \cdot r_{v,s_j} - |U| \cdot \bar{r}_{v,s_i} \cdot \bar{r}_{v,s_j}}{\sum_{v \in U} r_{v,s_j}^2 - |U| \cdot \bar{r}_{v,s_j}^2}$$
$$b_0 = \bar{r}_{v,s_i} - b_1 \cdot \bar{r}_{v,s_j} \tag{10}$$

The rough prediction of r_{u,s_i} made by similar service s_j can be obtained from:

$$Rp(r_{u,s_i}) = b_1 \cdot r_{u,s_j} + b_0 \tag{11}$$

Let x_1 and x_2 denote the two roots of prediction equation. The optimal root is selected as follows:

$$\begin{cases} pre_{sj}(r_{u,s_i}) = x_1, if\ |x_1 - Rp(r_{u,s_i})| < |x_2 - Rp(r_{u,s_i})| \\ pre_{sj}(r_{u,s_i}) = x_2, else \end{cases} \tag{12}$$

Where $pre_{sj}(r_{u,s_i})$ is the optimal value of r_{u,s_i} given by the similar service s_i.

In order to obtain higher accuracy, we use TOP K similar Web services with higher similarity to calculate predicted value of Web service s_i for users. Let KI represent most similar services in the set of *item_topK*. Hereafter, every similar item in KI has its predicted value r_{u,s_i}, and we calculate the final QoS prediction as:

$$\widehat{r_{u,s_i}} = \sum_{s_j \in KI} pre_{sj}(r_{u,s_i}) \cdot con(s_j) \tag{13}$$

Where $\widehat{r_{u,s_i}}$ is the final QoS prediction of r_{u,s_i}, and $con(s_j)$ denotes the weight in the prediction given by s_i:

$$con(s_j) = \frac{sim(s_i, s_j)}{\sum_{s_j \in KI} sim(s_i, s_j)} \tag{14}$$

D. User-based CASR-TADE

In the Section III.C, we introduced the item-based CASR-TADE to predict the unknown QoS values. Now we will explain the user-based CASR-TADE method, which is quite similar to the item-based method that we simplify the process in the following steps.

First, we assume that two users u_i and u_j obtain the value of response times when they have invoked services $\{s_1, s_2, \dots, s_k\}$. We use the true abnormal data elimination method in Section III.B to obtain a new matrix M_{user} (i.e., a variation of the matrix M) to predict the unknown value of QoS before the similarity computation.

Second, we use the new matrix M_{user} to compute the similarity between users.

Third, we suppose r_{u_i,s_i} is an unknown value, and let x represent the predicted value of r_{u_i,s_i}. User u_j is a similar user to u_i. The similarity between two user u_i and u_j is presented by $sim(u_i, u_j)$. We add r_{u,s_i} and r_{u,s_j} into the matrix M_{user}. Then we recalculate $sim'(u_i, u_j)$, which is the similarity of u_i and u_j. If two users u_i and u_j are highly similar, thus:

$$sim'(u_i, u_j) = sim(u_i, u_j) \tag{15}$$

In addition, the equation (15) is formula of x.

Furthermore, we obtain a rough prediction value of r_{u_i,s_i} according to linear regression, and the optimal root is determined by the rough prediction value.

Finally, Let KU represents most similar users in the set of *user_topK*, Hereafter, every similar user in KU has its predicted value r_{u,s_i}. We can calculate the final QoS prediction $\widehat{r_{u,s_i}}$ by every predicted value and their weight.

E. Item-based and User-based CASR-TADE

We have obtained two prediction values through item-based and user-based TADE in IV.C and IV.D. In this section, we will combine two methods to further improve the prediction accuracy.

Let $r_{u,i}$ be the unknown QoS value, pre_{user} and pre_{item} denote the predicted values given by user-based and item-based CASR-TADE respectively. We use following three indexes to evaluate the quality of these predicted values:

1) Max Similarity (MS): For pre_{user} and pre_{item}, MS represents the maximum similarity between users in KU and services in *KI*, denoted by $ms(pre_{user})$ and $ms(pre_{item})$, respectively.

2) Average Similarity (AS): For pre_{user} and pre_{item}, the index represents the average similarity between users in KU and services in *KI*, denoted by $as(pre_{user})$ and $as(pre_{item})$, respectively.

3) Reciprocal of Standard Deviation (RSD): pre_{user} and pre_{item}, the index represents the reciprocal of the standard deviation of similarities between users in KU and Web services in *KI*, denoted by $rsd(pre_{user})$ and $rsd(pre_{item})$, respectively. These indexes can be written as:

$$ms(pre_{user}) = max\{sim(u_i, u_j) | u_j \in KU\} \tag{16}$$

$$ms(pre_{item}) = max\{sim(s_i, s_j) | s_j \in KI\} \tag{17}$$

$$as(pre_{user}) = \frac{\sum_{u_j \in KU} sim(u_i, u_j)}{|KU|} \tag{18}$$

$$as(pre_{item}) = \frac{\sum_{s_j \in KI} sim(s_i, s_j)}{|KI|} \tag{19}$$

$$rsd(pre_{user}) = \sqrt{\frac{|KU|}{\sum_{u_j \in KU}\left[sim(u_i, u_j) - as(pre_{user})\right]^2}} \tag{20}$$

$$rsd(pre_{item}) = \sqrt{\frac{|KU|}{\sum_{s_j \in KI}\left[sim(s_i, s_j) - as(pre_{item})\right]^2}} \tag{21}$$

We use these three indexes to predict final value of $r_{u,i}$ through following equations:

$$Q(pre_{user}) = \frac{ms(pre_{user})}{ms(pre_{user}) + ms(pre_{item})} + \frac{as(pre_{user})}{as(pre_{user}) + as(pre_{item})} + \frac{rsd(pre_{user})}{rsd(pre_{user}) + rsd(pre_{item})} \tag{22}$$

$$Q(pre_{item}) = \frac{ms(pre_{item})}{ms(pre_{user}) + ms(pre_{item})} + \frac{as(pre_{item})}{as(pre_{user}) + as(pre_{item})} + \frac{rsd(pre_{item})}{rsd(pre_{user}) + rsd(pre_{item})} \tag{23}$$

Hence, we obtain the final predicted value of $r_{u,i}$ according to the following equation:

$$pre = \frac{pre_{user} \times Q(pre_{user}) + pre_{item} \times Q(pre_{item})}{Q(pre_{item}) + Q(pre_{item})} \tag{24}$$

The entire procedure of CASR-TADE algorithm is shown as follows.

Algorithm: Context-Aware Services Recommendation based on True Abnormal Data Elimination (CASR-TADE)

Input: **k**: the number of neighbors; **U**: the testing dataset; **Dataset:** training dataset

Output: **MAE/RMSE**: the error between the real values and predicted values

for every test user u_i in **U do**
  for k= 2:1:10
    $D1(u_i)$ = CASM ($u_i$dataset)
    $D2(u_i)$ = TADE ($D1(u_i)$)
    //TADE represents the True Abnormal Data Elimination //
    $UD1(u_i)$ = UPCC ($D2(u_i)$)
    // UPCC represents similarity mining according to dataset of the current user u_i *//*
    P_{ui} = DP (k, $UD1(u_i)$)
    //use the DP to get the different predicted P_{ui}. *//*
  end for
end for
for every test service s_i in **I do**
  for k= 2:1:10
    $D1(s_i)$ = CASM ($s_i$dataset)
    $D2(s_i)$ = TADE ($D1(s_i)$)
    //TADE represents the True Abnormal Data Elimination //
    $UD1(s_i)$ = IPCC ($D2(s_i)$)
    // IPCC represents similarity mining according to dataset of the current service s_i *//*
    P_{s_i} = DP (k, $UD1(s_i)$)
    //use the DP to get the different predicted P_{us_i}. *//*
  end for
end for
PreQoS =FinalPre (P_{ui}, P_{s_i})
// get the final predicted preQoSs of different weights of P_{ui}, P_{s_i} *//*
    MAE = MAEfun (preQoS, real QoS)
    RMSE = RMSEfun (preQoS)
    // Compute both MAEs and RMSEs//
  end for
end for

F. Computational Complexity Analysis

In this section, we discuss the computational complexity of predicting an unknown QoS value with the proposed CASR-TADE method. For the convenience of computational complexity analysis, we first assume that the dataset under the similar contexts is a $m \times n$ matrix which contains *m* Web service users and *n* Web services.

In the proposed method, we execute the TADE algorithm before calculating the similarity. The TADE method is a recursive method, thus the complexity is m or n for item-based or user-based method respectively. Because we eliminate t pair of true abnormal data for every pair values, $n - t$ services are invoked by user u_i and u_j. Therefore the complexity of all similarity calculations for user u_i is $O(m(n-t)(2n-t)t/2)$, and for all users is $O(m^2(n-$

$t)(2n-t)t/2)$. Similarly, the complexity of all similarities calculations for all services is $O(n^2(m-t)(2m-t)t/2)$.

In order to predict r_{u,s_i} in the item-based algorithm, we need KI services with high similarity. Then we use heap sorting method to sort all the similarities between services s and the other $m-1$ services. Thus the complexity of KI is $n\,log_2\,n$. Similarly, the complexity of KU in the user-based algorithm is $m\,log_2\,m$.

As appropriate roots in KI or KU equations are determined in a linear regression process, the complexity of predicting one similar user is $O(n-t)$. Therefore, the complexity of user-based TADE is $O(|KU|(n-t)) + O(m\,log_2\,m) + O(m^2(n-t)(2n-t)t/2)$. Similarly, the computational complexity of item-based TADE is $O(|KI|(m-t)) + O(n\,log_2\,n) + O(n^2(m-t)(2m-t)t/2)$.

Hence, the total computational complexity of the proposed CASR-TADE algorithm is $O(|KU|(n-k)) + O(m\,log_2\,m) + O(m^2(n-k)(2n-k)k/2) + O(|KI|(m-k)) + O(n\,log_2\,n) + O(n^2(m-k)(2m-k)k/2)$.

IV. Experiments

In this section, we conduct extensive experiments to evaluate the performance of proposed CASR-TADE method. Six compared methods including CASR-TADE are evaluated on WS-Dream dataset by evaluation metrics of both MAE and RMSE.

A. Datasets and Data Processing

We have adopted the WS-Dream [15] dataset that comprises 1,974,675 Web service invocation records of response time by 339 distributed service users on 5825 Web services. The raw data is a user-item matrix for response time. In addition, the simulation and experiment are developed by the MATLAB 2014b, and conducted on an ASUS K55V PC with Windows 7 operating system, 32 GB RAM and Intel Core I7 3.6 GHz CPU.

B. Evaluation Metrics

In WS-Dream dataset, we divide the dataset into 15 segments, in which various ratios of training and testing dataset will be discussed later. The mean absolute error (MAE) and root mean squared error (RMSE) are frequently used to measure the difference between values predicted by a model or estimator and observed values. We adopted MAE and RMSE to measure the prediction accuracy of our method through comparisons with other methods. MAE is defined as:

$$MAE = \frac{\sum |R_{i,j} - \hat{R}_{i,j}|}{N} \tag{25}$$

where $R_{i,j}$ denotes the QoS value of service j observed by user i, $\hat{R}_{i,j}$ is the predicted QoS value, and N is the number of predicted values. RMSE is defined as:

$$RMSE = \sqrt{\frac{\sum (R_{i,j} - \hat{R}_{i,j})^2}{N}} \tag{26}$$

C. Evaluation

1) Comparative Methods

We conducted series of experiments to compare our CASR-TADE method with the following five methods:

- **UPCC** [16]: the method recommends services to a user collected by other users sharing the similar preference. The similarity between users is calculated by PCC based on user profiles.
- **IPCC** [17]**:** the method recommends services to a user similar to the ones the user preferred in the past. The similarity between services is calculated by PCC based on services.
- **HAPA** [19]: the method makes services recommendation based on user similarity. The method includes User-based HAPA and Item-based HAPA.
- **ADE (Abnormal Data Elimination):** the method recommends services based on the similarity computation, which eliminates all possible abnormal QoS values. The method is different from the proposed CASR-TADE method, which only eliminates the true abnormal values.
- **CASR-TE** [20]**:** the method makes recommendation for users considering the temporal effectiveness.

2) Performance Comparison

For different methods, we firstly show the results of MAE and RMSE which are generated in different number *top-k* (from 2 to 10) of similar users or items in the ratio 14:1 of training dataset and test dataset in Fig. 2.

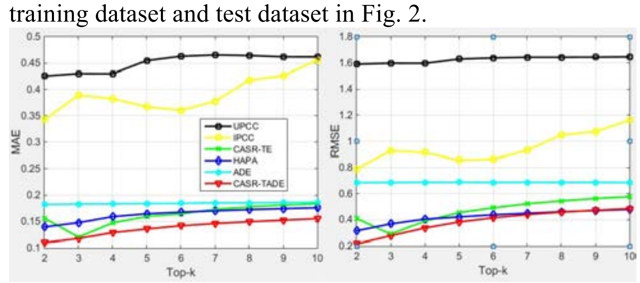

Figure 2. MAE and RMSE results of compared methods (14:1)

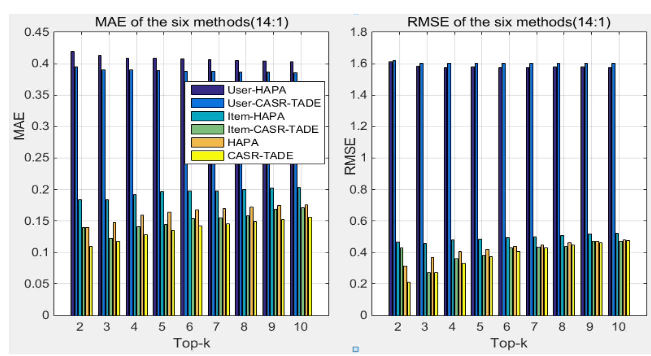

Figure 3. MAE and RMSE results of user/item-based methods (14:1)

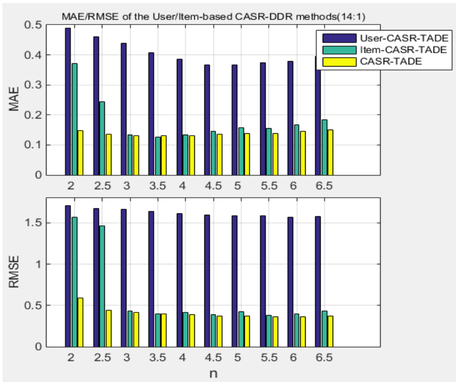

Figure 4. MAE and RMSE results of proposed methods in different n

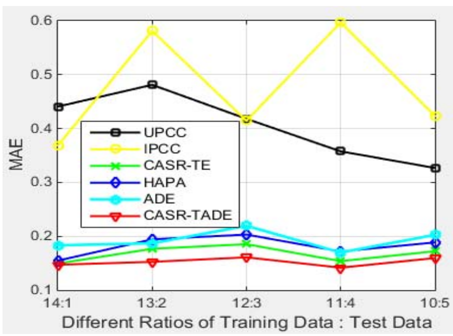

Figure 5. Evaluation results of different ratios of training and testing data

As shown in Fig. 2, we can see that: 1) generally speaking, when the number k of similar users or items increases from 2 to 10, the accuracy of the proposed CASR-TADE is better than five other baselines in terms of both MAE and RMSE; and 2) when the k increases, the MAEs and RMSEs of all methods will increase slowly. In general, the results in different k demonstrate that the significant accuracy of the CASR-TADE method in recommending personalized Web services.

Second, as shown in Fig. 3, we compare the MAE/RMSE results of proposed CASR-TADE method with the existing user-based and item-based methods (i.e., User-HAPA, Item-HAPA and HAPA). We can observe from Fig. 3 that: 1) the results of User-CASR-TADE are better than User-HAPA; 2) the results of Item-CASR-TADE are better than Item-HAPA; 3) the results of CASR-TADE are better than HAPA; and 4) the results of CASR-TADE are better than User-CASR-TADE or Item-CASR-TADE. Generally speaking, three methods including TADE (i.e., User-CASR-TADE, Item-CASR-TADE and CASR-TADE) outperform existing user/item-based methods significantly.

In Fig. 4, we further show the average MAE/RMSE results of Item-CASR-TADE, User-CASR-TADE and CASR-TADE in different values of n. Here, n has the same meaning with the "n" in the formula $|r_{v,s_i} - \mathrm{Rp}\left(r_{v,s_j}\right)| \geq nS$. When n is small, more values will be included as true abnormal values. We could see from Fig. 4 that: 1) the accuracy of CASR-TADE outperforms both Item-CASR-TADE and User-CASR-TADE; and 2) the optimal value n of three methods is around $\{3.5, 4.5\}$.

Finally in Fig. 5, we show the MAE results of six comparative methods in the different ratios (10:5, 11:4, 12:3, 13:2 and 14:1) of training and testing dataset. We conclude from Fig. 5 that: 1) the MAE and RMSE results of six methods have an approximately rising trend as the ratio of training and test data decreases; and 2) the result of our CASR-TADE method still performs better than other methods in different ratios.

D. Discussion

Comparing the performance of the proposed CASR-TADE method with five representative methods, we now discuss two aspects in our experiments: adjustment of the trade-off parameters and the effectiveness of TADE.

1) Trade-off parameters

The parameter k (i.e., the number of most similar users or items) has an important influence on the prediction results. Fig. 2 shows the MAEs of CASR-TADE are smaller than other methods in all different k values. While the RMSEs results between CASR-TADE and HAPA are close when the value $7 \leq k \leq 10$. The reason is that with the enlargement of the value k, users or services of lower similarity with the current user or service will be included. Thus, users or services of lower similarity have more nagative influence on the prediction accuracy which weakens the effect of TADE.

2) Impacts of TADE

To verify the impacts of TADE, we show the results of comparsion bewteen HAPA and CASR-TADE of user-based, item-based and combination of them in Fig. 3. The results show that TADE methods outperform existing user/item-based methods, which demonstrates the significant impact of introducing TADE before similarity computation.

The results shown in Fig. 2 and Fig. 5 could collectively show that: 1) in different k and different ratios of training dataset and testing dataset, TADE results in the great influence of context similarity in datasets, allowing a good prediction accuracy; 2) the TADE method is effectively integrated into the context-aware Web services recommendation model; and 3) the fusion of TADE and context-aware Web services recommendation model could further improve the accuracy of recommendation.

V. CONCLUSION

In this paper, we propose a novel CASR-TADE method by incorporating the effect of true abnormal data elimination into context-aware Web services recommendation, in order to downgrade the deviated impact of *True Abnormal Data* on the accuracy of QoS prediction. Then, we evaluate the CASR-TADE method on a real-world Web services dataset. Finally, experimental results show that the proposed method outperforms existing approaches significantly.

Despite the significant progress of proposed personalized Web services recommendation method, there still remain numerous avenues to explore. According to the experimental results in Section IV, our future works might include: 1) associating with other context factors, such as social context, to further improve the accuracy of QoS prediction; and 2) incorporating the spatio-temporal correlations with other similarity measurement methods for personalized Web services recommendation.

ACKNOWLEDGMENT

The work was supported in part by grants from National Natural Science Foundation of China (No. 61300232); Gansu Provincial Science and Technology Support Program-International Scientific and Technological Cooperation Projects (1504WKCA087); China Postdoc Foundation (141439); Fundamental Research Funds for the Central Universities (lzujbky-2015-100, lzujbky-2016-br04).

REFERENCES

[1] L.J. Zhang, J. Zhang, and H. Cai, Service Computing. Springer and Tsinghua University Press, 2007.

[2] 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, 2011, pp. 140-152, doi: 10.1109/TSC.2010.52.

[3] C. Wu, W. Qiu, Z. Zheng, X. Wang, and X. Yang, "QoS Prediction of Web Services based on Two-Phase K-Means Clustering," Proc. IEEE International Conference on Web Services (ICWS 15), IEEE Press, June 2015, pp. 161-168, doi: 10.1109/ICWS.2015.31.

[4] G. Adomavicius, and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," IEEE Transactions on Knowledge and Data Engineering, 2005, vol. 17, no. 6, pp.734-749, doi: 10.1109/TKDE.2005.99.

[5] G. Linden, B. Smith, and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," IEEE Internet Computing, 2003, vol. 7, no. 1, pp. 76-80.

[6] R. Jin, J. Y. Chai, and L. Si, "An automatic weighting scheme for collaborative filtering," Proc. of the 27th annual international ACM SIGIR conference on Research and Development in Information Retrieval (SIGIR 04), 2004. doi:10.1145/1008992.1009051.

[7] Z. Zheng, and M.R. Lyu, "Collaborative Reliability Prediction for Service-Oriented Systems," Proc. ACM/IEEE 32nd International Conference on Software Engineering Web (ICSE 10), vol. 1, 2010, pp. 35-44, doi: 10.1145/1806799.1806809.

[8] G. Adomavicius and A. Tuzhilin, Context-aware recommender systems: Recommender systems handbook, Springer, US, 2011.

[9] Y. Hu, Q. Peng, X. Hu, and R. Yang, "Web Service Recommendation Based on Time Series Forecasting and Collaborative Filtering," Proc. IEEE International Conference on Web Services (ICWS 15), IEEE Press, June 2015, pp. 233-240, doi: 10.1109/ICWS.2015.40.

[10] L. Kuang, Y. Xia, and Y. Mao, "Personalized Services Recommendation Based on Context-Aware QoS Prediction," Proc. IEEE International Conference on Web Services (ICWS 12), IEEE Press, June 2012, pp. 400-406, doi: 10.1109/ICWS.2012.12.

[11] B. Bai, Y Fan, K. Huang, W. Tan, B. Xia, and S. Chen, "Service Recommendation for Mashup Creation Based on Time-Aware Collaborative Domain Regression," Proc. IEEE International Conference on Web Services (ICWS 15), IEEE Press, June 2015, pp. 209-216, doi:10.1109/ICWS.2015.37.

[12] W. Zhang, H. Sun, X. Liu, and X. Guo, "Incorporating Invocation Time in Predicting Web Service QoS via Triadic Factorization," Proc. IEEE International Conference on Web Services (ICWS 14), IEEE Press, June 2014, pp.145-152, doi: 10.1109/ICWS.2014.32.

[13] J. Liu, M. Tang, Z. Zheng, X. Liu, and S. Lyu, "Location-Aware and Personalized Collaborative Filtering for Web Service Recommendation", IEEE Transactions on Services Computing, vol. PP, issue 99, 2015, pp. 1, doi: 10.1109/TSC.2015.2433251.

[14] F. Hao, S. Li, G. Min, H. Kim, S. Yao, and L. Yang, "An Efficient Approach to Generating Location-Sensitive Recommendations in Ad-hoc Social Network Environments," IEEE Transactions on Services Computing, vol. 8, no. 3, 2015, pp. 520-533, doi: 10.1109/TSC.2015.2401833.

[15] Z. Zheng, Y. Zhang, and M.R. Lyu, "Investigating QoS of Real-World Web Services", IEEE Transactions on Services Computing, vol. 7, no. 1, 2014, pp. 32-39, doi: 10.1109/TSC.2012.34.

[16] J. Breese, D. Heckerman, and C. Kadie, Empirical analysis of predictive methods for collaborative filtering, San Francisco, CA: Morgan Kaufmann Publishers Inc., 1998.

[17] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. R. Grouplens, "An open architecture for collaborative filtering of netnews," Proc. ACM Conference on Computer Supported Cooperative Work (CSCW 94), 1994, pp. 175-186, doi: 10.1145/192844.192905.

[18] K. Lee, J. Park, and J. Baik, "Location-based Web Service QoS Prediction via Preference Propagation for Improving Cold Start Problem," Proc. IEEE International Conference on Web Services (ICWS 15), IEEE Press, June 2015, pp. 177 - 184, doi: 10.1109/ICWS.2015.33.

[19] Y. Ma, S. Wang, P. Hung, C. Hsu, Q. Sun, and F. Yang, "A Highly Accurate Prediction Method for Unknown Web Service QoS Values," IEEE Transactions on Services Computing, vol. 99, no. 99, 2015, pp. 1, doi: 10.1109/TSC.2015.2407877.

[20] X. Fan, Y. Hu, R. Zhang, W. Chen, and P. Brézillon, "Modeling Temporal Effectiveness for Context Aware Web Services Recommendation," Proc. IEEE International Conference on Web Services (ICWS 15), IEEE Press, June 2015, pp. 225-232, doi: 10.1109/ICWS.2015.39.