Knowledge-Based Systems 115 (2017) 55–65

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

TAP: A personalized trust-aware QoS prediction approach for web service recommendation

Kai Su[a,*], Bin Xiao[a], Baoping Liu[a], Huaiqiang Zhang[a], Zongsheng Zhang[b]

[a] Department of Equipment Economics and Management, Naval University of Engineering, Mailbox 220, Jiefang Avenue 717, Wuhan 430033, China
[b] PLA, Academy of National Defense Information, Wuhan, China

ARTICLE INFO

Article history:
Received 29 May 2016
Revised 14 September 2016
Accepted 18 September 2016
Available online 18 October 2016

Keywords:
Web service
QoS prediction
Collaborative filtering
Data credibility
Clustering
Beta reputation system
Recommender systems

ABSTRACT

With the rapid development of service-oriented computing, cloud computing and big data, a large number of functionally equivalent web services are available on the Internet. Quality of Service (QoS) becomes a differentiating point of services to attract customers. Since the QoS of services varies widely among users due to the unpredicted network, physical location and other objective factors, many Collaborative Filtering based approaches are recently proposed to predict the unknown QoS by employing the historical user-contributed QoS data. However, most existing approaches ignore the data credibility problem and are thus vulnerable to the unreliable QoS data contributed by dishonest users. To address this problem, we propose a trust-aware approach TAP for reliable personalized QoS prediction. Firstly, we cluster the users and calculate the reputation of users based on the clustering information by a beta reputation system. Secondly, a set of trustworthy similar users is identified according to the calculated user reputation and similarity. Finally, we identify a set of similar services by clustering the services and make prediction for active users by combining the QoS data of the trustworthy similar users and similar services. Comprehensive real-world experiments are conducted to demonstrate the effectiveness and robustness of our approach compared with other state-of-the-art approaches.

1. Introduction

Web service is platform-independent software component designed to support interoperable machine-to-machine communication over a network [1]. Strongly promoted by the development of cloud computing and service-oriented computing technologies, web services are widely developed and employed to build service-oriented applications over the Internet. With the increasing number of functionally equivalent web services emerging on the network, Quality of Service (QoS) becomes a major concern for users to select services [2].

QoS is a set of properties describing the non-functional characteristics of web services, such as cost, response time, throughput, reliability, availablity, etc. In real world, the value of some user-dependent QoS properties (e.g., response time, throughput, failure rate, etc.) may vary widely among different users caused by their different physical positions, network conditions and other objective factors [3]. Personalized QoS prediction on client-side and web service recommendation based on the predicted QoS values are thus essential for supporting users to select optimal services to build high quality applications.

Collaborative Filtering (CF) plays an important role in the recommendations of items, such as movies, songs, books, etc [4–6]. Inspired by the success of CF techniques achieved in some commercial recommender systems such as Amazon, YouTube, Netflix, etc., many CF based QoS prediction approaches are recently proposed [1,2,7–9]. These CF based QoS prediction approaches commonly utilize the historical QoS information collected from other similar users or services to predict the unknown QoS of current users. Obviously, the prediction accuracy of CF based approaches is highly influenced by the credibility of user-contributed QoS data. Most existing CF based QoS prediction approaches are based on the hypothesis that all the users are honest and will provide reliable QoS data. However, dishonest users have been found in many web service recommender systems [10–12]. These dishonest users may submit constant values or random values as their observed QoS [10,11]. For immediate gain, some malicious users even exaggerate their partners' services while badmouthing their competitors' services, which are also known as push attack and nuke attack in shilling attacks [13]. Therefore, an effective QoS prediction approach, which could not only achieve high prediction accuracy

* Corresponding author.
E-mail address: keppelsue@163.com (K. Su).

http://dx.doi.org/10.1016/j.knosys.2016.09.033
0950-7051/

but also deal with unreliable user-contributed data, is required for a robust web service recommender system.

To improve the robustness of QoS prediction system, Qiu et al. [10] propose a reputation-aware prediction framework RAP, in which the users with low reputation rank are considered to be untrustworthy and filtered out from the system. Similar to RAP, Xu et al. [11] propose a reputation-based matrix factorization approach RMF, where the reputation of users is incorporated in prediction framework to reduce the influence of low reputed users. Both of RAP and RMF, which employ L1-AVG algorithm [14] to iteratively calculate the reputation of users, are sensitive to the parameter settings (e.g., iteration number and damping factor), inappropriate parameters may lead to low prediction accuracy due to the imprecise reputation evaluation.

In this paper, we present a trust-aware QoS prediction approach (TAP) by fully utilizing the observed QoS data of trustworthy similar users and similar services. Firstly, we employ K-means clustering algorithm to identify the honest user cluster on each service and classify the QoS feedback submitted by each user as positive or negative feedback based on the honest cluster. Secondly, the reputation of each user is calculated by statistically updating Beta probability density functions based on the number of positive and negative feedback he has provided in the past. Finally, based on the estimated user reputation, we employ Pearson Correlation Coefficient (PCC) to identify a set of trustworthy similar users and employ K-means clustering again to identify a set of similar services. The QoS data of the trustworthy similar users and similar services are then combined to make prediction for missing values. The main contributions of our work are three-fold:

(1) We employ unsupervised K-means clustering and Beta distribution based methods to calculate the reputation of users and ensure the credibility of the QoS data. Sections 3.3 and 4 demonstrate that our approach is very efficient and robust without any extra parameter settings.
(2) We propose a highly accurate and reliable QoS prediction approach by systematically combing the information of trustworthy similar users and similar services.
(3) We conduct several extensive real-world experiments to study the prediction accuracy and robustness of our approach compared with other state-of-the-art approaches.

The rest of this paper is organized as follows. Section 2 reviews the related work. Section 3 describes our TAP QoS prediction approach. Section 4 evaluates our approach with extensive experiments. Section 5 concludes the paper.

2. Related work

QoS has been widely discussed in Service-Oriented Computing domain and Cloud Computing domain, covering the topic of service selection [15–18], service recommendation [19–21], service discovery [22], service composition [23,24], composite service reliability prediction [25,26], etc. Most existing QoS-based approaches assume that the QoS of services are pre-existing and easily obtained from a third-party organization or service providers with guaranteed accuracy. However, the QoS data obtained from a third-party organization or service providers is not suitable for every user, since the QoS of services may vary widely among users. Therefore, an accurate personalized QoS values prediction approach is essential for the successful implementation of QoS-based approaches. In this paper, we focus on how to accurately predict the unknown QoS values and recommend optimal services for active users by CF approaches.

CF is the most popular technique employed in recommender systems. Recently, there are two main types of CF approaches widely discussed to predict the QoS of web services: memory-based CF and model-based CF. Memory-based CF predict the missing values by employing the neighbor information of observed QoS data. Zheng et al. [9] conduct a large-scale distributed QoS evaluation for real-world services and propose a hybrid CF approach UIPCC, which systematically combines the similar users' information and the similar services' information to make accurate prediction. Chen et al. [27] propose a modified memory-based CF approach which integrates the physical locations of users into neighbor selection and QoS prediction process. Hu et al. [28] propose a time-aware CF approach, in which the time information is integrated into similarity calculation to deal with the time-varying problem of QoS.

Model-based CF usually employs the observed data to train a predefined factor model and make predictions. Zheng et al. [1] propose a neighborhood integrated matrix factorization approach, in which the prediction model is learned by employing the user's information and the user's neighbors' information. Su et al. [2] propose a non-negative matrix factorization model for QoS prediction, in which the factor model is learned by an Expectation-Maximization (EM) based approach. Wei et al. [29] propose a local neighborhood matrix factorization QoS prediction approach, in which the geographical domain knowledge is integrated into factor model learning.

Although these recent research works improve the prediction accuracy to some extent, most of them presume that the QoS data contributed by users are reliable, causing them easily be attacked by the dishonest users, which commonly appear in many recommender systems. To improve the prediction accuracy under unreliable data condition, several reputation-based approaches are recently proposed. In [10], Qiu et al. propose a reputation-aware memory-based CF approach RAP, in which the data contributed by low reputed users is excluded and the missing values are predicted using the purified QoS data. Similarly, Xu et al. [11] propose a reputation-aware matrix factorization based CF approach, in which the user reputation is integrated into the factor model learning process to reduce the impact of dishonest users. However, the reputation evaluation method of above approaches is highly influenced by many parameter settings. Inappropriate parameters may lead to low performance. To overcome this limitation, Wu et al. [12] propose a credibility-aware QoS prediction approach CAP based on two-phase K-means clustering algorithm, in which the dishonest users are identified by user clustering and missing values are predicted based on the credible clustering information. K-means clustering algorithm is a common technique used to group data into different clusters, which is capable to detect the outlier points [30]. Although CAP is not sensitive to the parameter settings and significantly improves the robustness of prediction approach, the prediction accuracy is still not satisfactory, since CAP only employs the trustworthy similar users to make prediction and ignores the valuable information of similar services.

Based on the previous research works, we employ K-means clustering algorithm and Beta reputation system based approach to accurately evaluate the reputation of users and predict the missing values by fully utilizing the trustworthy similar users' information and similar services' information. Our approach can dynamically evaluate the reputation of users simply according to the statistical information of users without any extra parameter settings. The extensive experiments conducted in Section 4 demonstrate the effectiveness and robustness of our approach.

3. Trust-aware QoS prediction

In this section, we will introduce our proposed trust-aware QoS prediction approach TAP. Firstly, the prediction framework is given.