Information Sciences 265 (2014) 68–84

Contents lists available at ScienceDirect

Information Sciences

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

Colbar: A collaborative location-based regularization framework for QoS prediction

Jianwei Yin [a], Wei Lo [a], Shuiguang Deng [a], Ying Li [a], Zhaohui Wu [a], Naixue Xiong [b,*]

[a] School of Computer Science and Technology, Zhejiang University, China
[b] School of Computer Science, Colorado Technical University, USA

ARTICLE INFO

Article history:
Received 16 July 2013
Accepted 7 December 2013
Available online 22 December 2013

Keywords:
QoS prediction
Collaborative Filtering
Matrix Factorization
Regularization

ABSTRACT

Quality-of-Service (QoS) is a fundamental element in Service-Oriented Computing (SOC) domain. At the ongoing age of Web 2.0, predicting the missing QoS values becomes more and more important since it is an indispensable preprocess of numerous service-oriented applications. Previous research works on this task underestimate the importance of users' geographical information, which we argue would contribute to improving prediction accuracy in Web services invocation process. In this paper, we propose a novel collaborative location-based regularization framework (**Colbar**) to address the problem of personalized QoS prediction. We first leverage the personal geographical and QoS information to identify robust neighborhoods. And then, we collect the wisdom of crowds to construct two location-based regularization terms, which are integrated to build up an unified Matrix Factorization framework. Finally we make intermediate fusions to generate better prediction results. The experimental analysis on a large-scale real-world QoS dataset shows that the prediction accuracy of Colbar outperforms other state-of-the-art approaches in various criteria.

1. Introduction

Web services are software applications designed to support interoperable machine-to-machine interactions over a heterogeneous environment via a set of standard communication protocols [1]. With the exponential growth of Web services deployed on Internet, numerous users enjoy high-quality services to get connected, and this driven force makes World Wide Web more flourishing. Quality-of-Service (QoS) is usually employed to describe the non-functional aspects of Web services. Thanks to the fact that companies and organizations raise unprecedent needs for Web services at present, studies on QoS have raised the concerns of Service-Oriented Computing (SOC) researchers. A number of QoS-based paradigms have been applied to the area of service selection [2,3], service discovery [4,5], service recommendation [6,7], service composition [8,9], service trust [10,11], service routing [12,13], etc.

The common hypothesis of above research areas is that QoS values of all Web services are available. However, this premise sometimes may be challenged in real-world cases for the following reasons: (1) The Web services hierachy becomes more complex to fit in different situations. As a result, it is time-consuming for an end user to explore massive QoS records. (2) Most Web services are operated by commercial companies. Gathering QoS information by execution can be too costly to an end user. (3) Nowadays the Internet environment becomes more dynamic yet vulnerable. It turns out impractical and

* Corresponding author. Tel.: +1 404 645 4067.
E-mail addresses: zjuyjw@zju.edu.cn (J. Yin), spencer_w_lo@zju.edu.cn (W. Lo), dengsg@zju.edu.cn (S. Deng), cnliying@zju.edu.cn (Y. Li), wzh@zju.edu.cn (Z. Wu), xiongnaixue@gmail.com (N. Xiong).

0020-0255/$ - see front matter © 2013 Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.ins.2013.12.007

impossible to collect QoS records all the time. In order to satisfy the basic requirements from service-oriented research domains mentioned above, a necessary preprocess is to predict those missing QoS values.

Inspired by the idea of user-collaboration in the era of Web 2.0, Collaborative Filtering (CF) techniques are widely used to fulfill this task. The core idea behind is to identify a set of similar neighbors and to collect the wisdom of crowds. In this process, the measurement of similarity becomes vital. Previous CF approaches on QoS prediction mainly conduct *Pearson Correlation Coefficient* (PCC) [14] to build up user-similarity neighborhoods [15,16]. However, we argue that this kind of pure PCC-based neighborhood selection mechanism is inappropriate in Web service scenarios for the following reasons: (1) The historical records on QoS are sometimes incorrect on account of the fragile nature in Internet environments. However, PCC relies heavily on precise QoS values to measure similarity. (2) Classic CF approaches are broadly applied in Recommender Systems which contain a lot of missing user ratings [17]. Each rate marks a user's objective preference towards a specific item. Meanwhile in Web service domain, the QoS record is heavily affected by subjective environments. This factor makes PCC calculation less effective in capturing similarity relationship.

In this paper, we present an elegant neighborhood selection mechanism by leveraging the personal geographical and QoS information. To our common sense, it is natural to consider that users in the same local region are very likely to receive similar QoS information [18,19]. The premise behind this intuition is that local users share the same or similar IT infrastructures (network routers, switches, etc.). In general, this phenomenon is prevalent in the real world.[1] Because of this local factor, these neighbors tend to present similar Web services usage information. Sometimes even though each user inside a neighborhood would choose different network settings, we observe that these fluctuations is a minor element compared with the geographical factor [20]. And to deal with the issue on various network settings, we then employ PCC calculations on users' existing QoS usage experience. The integration of personal geographical information and QoS history records promises to generate a robust neighborhood in prediction process.

Based on the above intuition, we propose a novel collaborative location-based regularization framework (**Colbar**) to address the QoS prediction problem in this paper. Instead of using pure PCC computations in classic CF approaches, we first leverage personal geographical and QoS information to generate a list of user-similarity neighborhoods. We then elaborate how to collect the wisdom of crowds to construct two novel location-based regularization terms to capture latent connectivity among users. These constraints are employed to revamp the traditional Matrix Factorization (MF) [21] model as a unified framework. Finally, we make personalized QoS predictions by fusing neighbors' intermediate results to emphasize neighborhoods effects. The experimental analysis on a large-scale real-world QoS dataset [22] shows that the prediction accuracy of our proposed Colbar framework outperforms other state-of-the-art approaches in various metrics.

The major contributions of this paper are three-folds:

- We leverage the geographical and QoS information to generate robust neighborhoods, which is the key element in Collaborative Filtering.
- We elaborate how to build location-based regularization terms to capture the user latent connectivity inside a neighborhood.
- We employ local wisdom to revamp the traditional Matrix Factorization model as a unified QoS prediction framework. We also show that our approach can efficiently scale up to large-scale datasets from theoretical and practical perspectives.

The rest of this paper is organized as follows: Section 2 describes the scope of problem in this paper. Section 3 details how to use Colbar framework to make QoS prediction. Section 4 presents the results of an empirical analysis. Section 5 reviews some related works on the topic of QoS prediction. Finally Section 6 concludes the paper.

2. Problem description

The problem we study in this paper is how to leverage users' geographical information and history records to make personalized QoS predictions. In practice, there are m users and n Web services. They contribute to an $m \times n$ user-service QoS matrix R, and each entry r_{ui} records a piece of QoS usage information of Web service i executed by the user u. Note that there are many missing QoS values in matrix R. Our goal is to build up a unified framework to fill in these missing records effectively with higher accuracy.

To give a better understanding of our goal, we introduce a typical case to highlight the benefits of harnessing the users' geographical information. Fig. 1 illustrates a real-world Web service invocation scenario. All active users and Web services are located in a Cartesian coordinate system. Each user is identified by a tuple: (name, x-location, y-location), and it randomly makes service invocation at different time intervals. The matrix in Fig. 2 records the corresponding response time of Web services invoked by users. As mentioned above, if users live very close to each other and they are somehow similar in terms of QoS records, we can view these users as neighbors. For instance, user Park and Sun might build up a neighborhood N_2 because their Euclidean distance is smaller than a certain threshold. This local characteristic embeds latent connectivity among neighbors. As a result, if we want to make QoS prediction on user Park, we can incorporate user Sun's

[1] http://www.akamai.com/stateoftheinternet/.