A Time Series QoS Prediction Method for Time-aware Cloud Service Recommendation

Shuai Ding
School of Management
Hefei University of Technology, Hefei, China
Tel: (+86) 13855169247, Email: dingshuai@hfut.edu.cn

Ye-Qing Li
School of Management
Hefei University of Technology, Hefei, China
Tel: (+86) 15256911908 ext. 2382, Email: 760242570@qq.com

Abstract. With the rapid development of cloud computing, plenty of cloud services are implemented on the cloud platform. Thus, how to select the most appropriate cloud services candidates for users is becoming an essential issue. Being compared with traditional services, instantaneity reflects the different performances of cloud services in different time. This paper presents time-aware cloud service recommendation based on time series. Our proposed method estimates the time-aware Pearson Correlation Coefficient to find neighboring users, and we generate time series model to predict the prospective performances of cloud services for recommendation. The experimental results show that our approach outperforms other competing methods.

Keywords: cloud computing, cloud service, time series

1. INTRODUCTION

Cloud computing is becoming popular, which makes cloud services become more and more mature and dominant. There has been a rising demand for cloud services and many scholars focus on how to promote quality of service (QoS) which is the non-functional evaluation indicator to describe the quality of cloud services, such as response time, throughput, security etc. Therefore, how to recommend high-quality cloud services has become an essential issue under the background of rapidly increasing demand of cloud services.

However, it is worth to note that the QoS values are variable in different time even though users invoke the same cloud service, which leads to the various user experiences, including the delay of invoking, insensitivity of reaction and etc. Due to the lack of consideration of time factor, the traditional recommendation systems cannot meet the dynamic demand in the process of cloud service selection. Meantime, the existing time-aware methods usually utilize time weighted function to address this issue. These approaches pay attention to the dynamics of the similarity attributing for the time sensibility and improve the accuracy of prediction partly. However, it is harder to mine the connection relationship in different time due to neglect the time series characteristics of data.

Thus, this paper proposes the time series QoS prediction method in the view of time dynamics of QoS. Firstly, we integrate Pearson Correlation Coefficient (PCC) to find neighboring users. Secondly, based on the top-K similar neighbors, we utilize collaborative filtering (CF) to fill the cloud user-service-time matrix (CUSTM) in the existing time. Thirdly, in the light of the stationarity of data, we build the time series model to predict the QoS values in future. Finally, we rank and recommend the cloud services for user selection according to the predicted results. The experimental results show that our method is more accurate than traditional approaches in OoS prediction and the cloud service recommendation.

2. LITERATURE REVIEW

In recent years, with the continual development of cloud services, as the cloud services category gets more diverse, the study of QoS turns to be a research hotspot in the field of cloud computing. In 1999, Xiao et al. (1999) defined the emerging Internet QoS. Then the investigation on QoS of Cloud services grows significant gradually and mainly covers the optimization of cloud service composition, resource matching, security, etc. For instance, Zhi-Zhong Liu (2015) proposed a specific social learning optimization algorithm to solve the problem of QoS-aware cloud service composition; Tiejiang Liu et al. (2012) presented a service deployment management system for optimization to improve deployment efficiency while guaranteeing the users' QoS requirements; Rizwana Shaikh recommends a trust model which acts as a benchmark and ranking service to measure security in a cloud computing environment (2015). Nonetheless, these researches have not applied QoS to cloud service prediction.

How to select the suitable cloud services is becoming a spotlight with the exponential growth of cloud services. Plenty of studies have focused on the prediction of QoS values, as a significant issue in cloud services selection. Yueshen Xu et al. (2016) employed the context information, from both the user side and service side, to achieve superior QoS prediction accuracy; Ali Asghar Pourhaji Kazem (2015) presented BNQM, a Bayesian network based probabilistic QoS Model for Grid service composition which is efficient in predicting the QoS values. Furthermore, the emerging artificial intelligence technology has also been utilized in this field. Mao et al. structure a prediction framework based on the modified particle swarm optimization algorithm for QoS ranking prediction.

To some extent, these methods improve the recommendation accuracy and reflect the importance of the time factor. Yet, these researches do not pay attention to the dynamics of user similarity and they ignore the application of time series in this field. Hence, we propose a QoS-based cloud service selection model which combines the time-aware user similarity and time series.

3. Time-aware Similarity Estimation

While the cloud services are invoked, due to the uncertainty of cloud service in time dimension, the dynamics of QoS proves to be the major contributory element in QoS prediction. Therefore, to predict the QoS of cloud services which are not invoked by users is significant in the processes of cloud services selection, dynamic combination, computing property, etc. Considering the time, each QoS attribute has CUSTM, respectively, including cloud users, cloud services and time slices.

Suppose U={u_i|$i\in\{1,2,\ldots,m\}$} is a finite set of m cloud users, CS={cs_j|$j\in\{1,2,\ldots,n\}$} is a finite set of n cloud services, P={p_t|$t\in\{1,2,\ldots,h\}$} is a finite set of h time slices, and Q={q_k|$k\in\{1,2,\ldots,l\}$} is a finite set of l QoS attributes.

In the time-aware QoS prediction, firstly, we extract each cloud user-service matrix (CUSM) in every time slice to reduce the dimension. Secondly, we estimate the user similarity in CUSMs by Pearson Correlation Coefficient, and utilize collaborative filtering (CF) to predict the null values of CUSMs. Finally, we build the model for every user in each service employed time series method to make cloud services recommendation.

3.1 Pearson Correlation Coefficient

PCC is a kind of widely used method in calculating the degree of linear association between two entities. Therefore, we estimate the similarity between u_p and u_q in one time slice as following:

$$Sim_{p,q}^{t}=\frac{\sum_{j=1}^{n}(Q_{pjk}^{t}-\bar{Q}_{pk}^{t})(Q_{qjk}^{t}-\bar{Q}_{qk}^{t})}{\sqrt{\sum_{i=1}^{n}(Q_{pjk}^{t}-\bar{Q}_{pk}^{t})^{2}}\sqrt{\sum_{i=1}^{n}(Q_{qjk}^{t}-\bar{Q}_{qk}^{t})^{2}}} \tag{1}$$

where Q_{pjk}^{t} and Q_{qjk}^{t} are the QoS values, concerning k attribute, which u_p and u_q invoke cs_j in designated p_t, respectively. $\bar{Q}_{pk}^{t}$ and $\bar{Q}_{qk}^{t}$ are the average QoS values of cloud services invoked by u_p and u_q.

3.2 Similarity Combination based on Time

In this paper, we take the time factor into consideration in the cloud user similarity calculation, which represents the dynamics of similarity. Our method integrates the separate similarities of each time slice into a holistic similarity by weighted function.

The similarity in the last time slice has more weight for its instantaneity, because the action of users in the recent time reflects future behavioral tendencies more exactly. The similarity combination based on time $Sim_{p,q}$ can be computed as following:

$$Sim_{p,q}=\frac{t*\mathrm{Sim}_{p,q}^{t}}{\sum_{t=0}^{h}t} \tag{2}$$

4 the Cloud Service Recommendation based on Time Series Prediction

Based on the time-aware similarity $Sim_{p,q}$, we build the model to predict the QoS value in future time slice according to the features of the data in the past several time slices, utilizing the time series method which is a statistical method focusing on dynamic data process.

The serialization of QoS values is formed with the time, which is impacted by two elements: the first one is the fluctuation of the QoS serialization and the second one is the change of the external environment influenced by the invoking cloud services.

Meantime, the feature of time-aware QoS values is consistent with the time series. Therefore, we adopt ARMA model to do prediction and recommendation. It is worthy to note that we fill the CUSTM with CF to supply enough data support firstly in order to improve the prediction accuracy.

4.1 Collaborative Filtering Prediction

We employ the Top-K similar of users, avoiding the negative influence caused by dissimilar users or less similar users, to predict the missing values in the CUSTM as following:

$$pred(\mathrm{p},\mathrm{j}) = \bar{\mathrm{Q}}^t_{pk} + \frac{\sum_{i=1}^{k} Sim_{p,q} * (\mathrm{Q}^t_{qjk} - \bar{\mathrm{Q}}^t_{qk})}{\sum_{i=1}^{k} Sim_{p,q}} \tag{3}$$

where *pred(p,j)* is the predicted QoS generated by u_p invoking cs_j, and we fill the CUSTM for the structure of time series model.

4.2 Time Series Analysis

As we estimate all of the values in the existing time slices in section 4.1, we use the time series method to analyze. When the next QoS value is simply affected by the fluctuation of the previous QoS, the serialization is in good agreement with Auto-Regression Model (AR). If the QoS values are just influenced by the external environment, it meets the characters of Moving Average Model (MA). Once the QoS serialization is both affected by these two factors, it matches the ARMA model better. The specific details are described in following sections.

4.2.1 Data Preprocessing

1) Data Preprocessing

Data preprocessing can be called as data smoothing. Before we build the model, it is necessary to analyze the stationarity of the QoS serialization. Once the serialization does not meet the conditions of the model, we should smooth the serialization by some methods such as difference equation. As conditions of the QoS serialization satisfy the time series model condition, we recognize the characters of QoS series and identify the model including AR, MA and ARMA.

2) Model Identification

We identify the QoS serialization attributions adopting the Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) and determine which model fits the QoS serialization condition, which includes AR, MA and ARMA. The attributions can be expressed as following in Table 1:

Table 1. ACF and PACF of three models

	AR(*p*)	MA(*q*)	ARMA(*p*,*q*)
ACF	trailing	truncation after *q* steps	trailing
PACF	truncation after *p* steps	trailing	trailing

(3)

The calculation of ACF and PACF can be described as following:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (\mathrm{Q}^t_{ijk} - \bar{Q}^t_{ijk})(\mathrm{Q}^{t+k}_{ijk} - \bar{Q}^t_{ijk})}{\sum_{t=1}^{n} (\mathrm{Q}^t_{ijk} - \bar{Q}^t_{ijk})^2} \tag{4}$$

$$\phi_{kk} = \frac{\hat{D}_k}{\hat{D}}, \forall 0 < k < n \tag{5}$$

Based on the value of ACF and PACF, we build the the following models:

① AR(*p*) model:

As the PACF truncates in *p* steps, the QoS serialization applies to AR(*p*) model, which means that the QoS values are just impacted by p historical QoS values and the impact factor is φ_ε. Accordingly, the AR model can be represented as following:

$$Q^t_{ijk} = \varphi_0 + \sum_{\varepsilon=1}^{p} \varphi_\varepsilon Q^{t-\varepsilon}_{ijk} + u_t \tag{6}$$

where $Q^{t-\varepsilon}_{ijk}$ is the QoS value in the time slice, $\varepsilon \in (1, p)$.

② MA(*q*) model:

When the ACF truncates in *q* steps, QoS values match the MA(*q*) model. In cloud environment, the future QoS values just depend on the external random interferences and the interference factor is expressed as $u^{t-\xi}_{ijk}$. The MA model can be described as following:

$$Q^t_{ijk} = u + \sum_{\xi=1}^{q} \theta_\xi u^{t-\xi}_{ijk} + u_t \tag{7}$$

where $u^{t-\xi}_{ijk}$ is the external random interference factor on the cloud services, and $\xi \in (1, \mathrm{q})$.

③ ARMA(*p*,*q*) model:

When both of the ACF and PACF show the character of tailing, the QoS serialization applies to ARMA(p,q) model. It indicates that QoS values are not affected by single factor but the p historical values and external disturbances. The ARMA model can be described as following:

$$Q_{ijk}^{t} = \varphi_0 + \sum_{\varepsilon=1}^{p} \varphi_{\varepsilon} Q_{ijk}^{t-\varepsilon} + \sum_{\xi=1}^{q} \theta_{\xi} u_{ijk}^{t-\xi} + u_t \quad (8)$$

3) Order Determination

During the procedure of model identification, we fix the model preliminarily. In order to improve the degree of agreement and reduce the difficulty in practical application, we usually employ Akaike Information Criterion (AIC) or Bayesian Information Criterions (BIC) to revise the model as following:

$$AIC(k) = \ln \hat{\sigma}_e^2 + \frac{2k}{N}, k = 0,1,\cdots \quad (9)$$

$$BIC(k) = \ln \hat{\sigma}_e^2 + \frac{k \ln N}{N}, k = 0,1,\cdots \quad (10)$$

where $\hat{\sigma}_e^2 = \hat{\rho}_k - \sum_{j=1} \hat{\phi}_j \hat{\rho}_j$ is uesd to estimate the σ_e^2 of various algorithms and N is the size of QoS sample used in the time series model.

4) Parameter Estimation and Prediction

After confirming a certain model, we use least square estimating the model parameters. Following the steps above, we define the time series model eventually. Based on the model, we predict the QoS value in the short future for the recommendation of cloud service.

5. Experiment Results

In this section, we compare Time-aware Cloud Service (TSCC) we proposed with user-based collaborative filtering (UBCF) and ARMA.

5.1 Data Sources

Cloud services are similar with web services, particularly in QoS properties. Because of lack of the real QoS dataset of cloud services, we used the QoS of web services instead.

As it is impossible to invoke thousands of web services for large-scale experiments, we perform experiments on the QoS dataset from WS-DREAM team, which collected 4,532 web services from public sources on the web and 142 distributed computers in 64 time intervals from Planet-Lab. Meanwhile, this dataset includes response-time and throughput QoS value.

5.2 Evaluation Criteria

We evaluate the effectiveness of three methods with Mean Absolute Error (MAE) which is defined as:

$$MAE = \frac{\sum_{p,i} \left| Q_{p,i} - Q_{p,i}^{pre} \right|}{N} \quad (11)$$

where $Q_{p,i}$ is the real QoS value of the cloud service cs_i invoked by user u_p. $Q^{pre}{}_{p,i}$ is the predicted QoS value by experimental approaches, and N is the number of predicted QoS values. The accuracy of prediction increases with the decreasing of the MAE.

5.3 Performance Comparison

In this section, we compare the MAE of our proposed method with UBCF and ARMA

UBCF: this method predicts the QoS values in the 64th time slice directly only based on the information on the 64th time slice.

ARMA: this method is a classically statistical approach which only considers the statistical property of the data and predicts the QoS values in the 64th time slice based on the top 63 time slices.

TSCC: our proposed method combines UBCF and ARMA, and it also takes the time-aware similarity into the consideration.

Table 2. MAE values of three methods

	TSCC	***UBCF***	***ARMA***
First time	1.452	1.652	1.488
Second time	1.501	1.601	1.521
Third time	1.428	1.632	1.463

In Table 2, we compare our approach with the UBCF and ARMA. As the Table illustrates, TSCC and ARMA are better than the UBCF in prediction accuracy obviously, and the performance our method is better than ARMA slightly, which proves that time factor has significant impact on the cloud service recommendation.

6. CONCLUSION

In this paper, we propose a time-aware cloud service recommendation which estimates the time-aware Pearson Correlation Coefficient to find neighboring users and build time series model to predict the future QoS values. The experiments show that our method has better predicted results which take the impact of time into consideration.

REFERENCES

Xiao X, Ni LM. Internet QoS: A Big Picture. (1999) *IEEE Network*, 8–18.

Liu Z Z, Chu D H, Song C, et al (2015). Social learning optimization (SLO) algorithm paradigm and its application in QoS-aware cloud service composition. *Information Sciences An International Journal*, 315-333.

Liu T, Lu T, Wang W, et al (2012). SDMS-O: A service deployment management system for optimization in clouds while guaranteeing users' QoS requirements, F*uture Generation Computer Systems*, 1100-1109.

Shaikh R, Sasikumar M (2015). Trust Model for Measuring Security Strength of Cloud Computing Service *Procedia Computer Science*, 380-389.

Xu YS, Yin J, Deng S, et al (2016). Context-aware QoS prediction for web service recommendation and selection. *Expert Systems with Applications*, 75-86.

Kazem A A P, Pedram H, Abolhassani H. BNQM (2015), A Bayesian Network based QoS Model for Grid service composition. *Expert Systems with Applications*, 6828-6843.