CLOUD SERVICE QoS PREDICTION ALGORITHM BASED ON TIME PERCEPTION AND USER'S INTEREST

Zhou ZHOU, Xinyuan ZHOU*

In this study, a cloud service QoS algorithm based on the time perception and the user's interest (denoted by CSQI) is proposed to recommend the cloud service precisely to the users for meeting their demand. Different from the current QoS value prediction, CSQI processes the data in the earlier stage, removes the negative correlation data, and improves the predictive data ratio of the positive correlation by fully considering the multiple time-slice perceptual ranking. Moreover, the effect of the user's interest of the regional similarity in the period time slice on QoS is emphasized. In addition, the service curve of the regional similarity user in the multiple time-slice is fitted based on gray relation theory, and the first K users of the ranking synthesis similarity are weighted to avoid the loss of the QoS history data. Experimental data indicate that the CSQI algorithm has better precision than other algorithms.

Keywords: Cloud computing; Cloud service; Recommended algorithm; Time perception; QoS

1. Introduction

Cloud computing is widely used at present, and cloud-based services are also increasing [1-2]. Obtaining the cloud service resources and pushing them to users are problems that need to be solved urgently. The existence of QoS (Quality-of-Services) provides a corresponding strategy to solve this problem. In other words, some outstanding sequences can be screened out by predicting cloud services in advance and improving the accuracy of the QoS prediction algorithm can significantly avoid time consumption and costly resource waste [3-4].

QoS is a comprehensive indicator of cloud services. The calculation on the QoS value, especially the study corresponding to the various attribute values, is only recently established [5]. The valid conclusions also show that the uniform criteria are not formed [6]. A large number of cloud services with the same functions are available on the Internet. Calculating and estimating the future QoS dynamic conditions based on various types of information collected are important tasks at present. The effects of the dynamic nature of the historical QoS datasets, the sparseness of the data matrix, and the use periodicity of the user need to be considered as much as possible when considering the user's satisfaction and accuracy of the cloud service QoS. At present, three types of research directly affect this aspect. The three types of research are the cloud service QoS

[1] Department of Mathematics and Computer Science, Changsha University, China, E-mail: xyzhouccsu@126.com

predictions based on the time perception ranking, the user's interest, and the hybrid system.

The first type of research is the cloud service QoS prediction based on the time perception ranking [7-8]. The time perception-based recommendation established a TaCR framework, estimated the service used by the customer on the basis of the active customers, and conducted the deficiency prediction of QoS to recommend the similar cloud service products to them. The detailed steps were as follows. First, the historical user dataset was obtained, the record for the user to call the service was divided into multiple parts, and the queue order was processed separately. Second, the similarity of the active and historical customers was calculated to select the customers who met the requirements. The prediction algorithm based on time perception was simple and fast, and it was accurate for products with periodic characteristics. However, it could not recommend the hidden concerns to the customers, and this incapability was not conducive to solving non-text information, such as sound and images.

The second type of research is the cloud service QoS prediction based on the user's interest [9-11]. The characteristics of the user's interest could be used to solve the dynamic changes and the personalization of the user's needs in the cloud service field. Introducing the user's interest strategies into the service prediction direction is a current research area. For example, WSRec, which is a cloud service recommendation system based on the user's contribution mechanism, was proposed by Zibin Zheng of the Chinese University of Hong Kong in 2009 [9]. Its policy was deployed in a typical environment by JAVA for experiments. When it was classified into WS-Dream dataset, it could be accessed arbitrarily, which assisted in promoting the application and evolution of the user's interest in the direction of the cloud service prediction algorithms. The service discovery system based on the collaborative filtering architecture was proposed in [10], which was a great resource for the customers to screen to meet the demand for the resources. Its distinctive feature was that it did not require the interaction of community services and the evaluation of cloud services. Karta analyzed and sorted out an agent-based algorithm for the cloud service prediction experiments based on the concept of the user's interest. However, the video information library MovieLens could not accurately and effectively feedback the required resources because this algorithm needed to call this information library.

The third type of research is the cloud service QoS prediction in hybrid systems [12-13]. All prediction algorithms contained their unique attributes, for which most researchers gradually started to study the hybrid prediction algorithms. In the fields of online shopping malls and virtual media libraries and other fields, the hybrid recommendation has been proven to have better recommendation performance [14]. The hybrid algorithm was widely used because it could better avoid the shortcomings of a single algorithm in a certain aspect. For example, the prediction model based on the user’s interest was combined with the prediction model based on time perception. The recommendation of the user’s interest could make up for the restriction of the recommended method based on the time perception on the key time attributes. Meanwhile, the recommended method based on the time perception was used to handle the data sparseness, new services/users, the cold start, and other issues in the recommended method of the user’s interest.

Overall, the main contributions of this study can be summarized as follows:

(1) A novel cloud service QoS algorithm (CSQI) based on the time perception and the user’s interest is proposed to recommend the cloud service to the users precisely.

(2) Different from other QoS algorithms, CSQI fits the service curve of the regional similarity user in the multiple time-slice by using gray relation theory, and the first K users of the ranking synthesis similarity are weighted to avoid the loss of the QoS history data.

(3) A large number of experiments are conducted to verify the correctness and effectiveness of the proposed algorithm.

The rest of the paper is organized as follows. In Section 2, we propose the cloud service QoS prediction algorithm CSQI. Experimental results and analysis are presented in Section 3. Section 4 elaborates the conclusions of this study.

2. Cloud service QoS prediction algorithm based on time perception and user’s interest (CSQI)

The main idea of CSQI can be concluded as follows. First, the dataset was ranked, and the time slice operation was done for all the services. Thereafter, the segment was weighted and combined again for the subsequent service ranking. It focused on the head users where the regional service similarity was associated with the time slice. To reduce the effect of the noise data, the top K cloud service users with strong correlation were selected through calculation after the time attenuation factor was added. At the end of the experiment, the selected users and their QoS were calculated to obtain the missing prediction, as shown below:

Algorithm 1: CSQI

Input: User’s service matrix E, the similar neighbor set K, the curve correlation weighting factor θ, time decay factor τ

Output: The missing value of the target user Ru was predicted

Begin

For t=1 to T-1

//The user’s data were divided and performed per time slot.

```
INT N= t × Num(E)
For (INT T0=0; T0 < N; T0++)
    Weight (θ)
//The response time for each time slot was assigned per weight θ.
For(i=1;i<n-1;i++)
For(j=1;j<=n;j++)
    S[i][j]=Sim(u, v)   //The similarity between u and v were calculated by the formula Sim(u, v).
For(U>i; U <=j;U++)
    //The top K users of U with the highest similarity to the user U were found,
        //and the missing calculation was performed combined with the time decay factor τ.
    For (S=i; S<= τ; S ++)
        If (Q[u,i]==0)
            Weight (τ)
  If (G[i][j]>=u and X[i]=0)
    then ri= Aggregate (ri, rj)  // Aggregate (ri, rj) was a clustering function
End
```

Different from other algorithms, CSQI set the ideal value and measured the closeness of the regional service similarity based on the relative time ratio for the regional service similarity model and gray correlation theory. The pull factor in the system was set by the gray correlation between all sub-modules to obtain the influence of the ideal correlation degree and the multiple time-slice periodic trend index (the fitting curve slope of the gray correlation). Thus, the influence of the curve slope was considered in the weighting of each time slice, that is, the periodic service similarity trend index–curve slope × 100.

To collect the dataset information of different customers who called the same cloud service, the round-trip time in the cloud service customer data was collected and fed back using the matrix E_u, i, and the round-trip data information was relatively concentrated. $Denote(u,i)$ represents the customer i, n represents the $n-th$ cloud service being called (evaluated), and $t_{ui,jn}$ represents the round-trip time consumed by the customer i who used a cloud service n, as shown in Eq. (1).

$$E_{u,j} = \begin{bmatrix} t_{u1,j1} & \cdots & t_{u1,jm} \\ \vdots & \ddots & \vdots \\ t_{ui,j1} & \cdots & t_{ui,jm} \end{bmatrix} \tag{1}$$

In the first step, the time of the filtered data was fragmented. After multiple comparisons, the experimental result was optimal when the divided time slices were seven segments.

In the second step, the weight was conducted on different time slices according to their corresponding weights.

$$S(u,v) = \chi_1 S_{U1}(u,v) + \chi_2 S_{U2}(u,v) + \Lambda + \chi_i S_{Ui}(u,v) \tag{2}$$

The weighting factor x_i was weighted according to the slope of the curve at different time points of the model statistics $\sum x_i = 1$.

$$\chi_i = \frac{Sim(u,v)}{\sum_{v \in S(u)} Sim(u,v)} \tag{3}$$

$Sim(u,v)$ was used to find the similarity between the variables u and v.

In the third step, all datasets obtained after weighting were simply ranked. The "similarity" processing was conducted gradually on the customer in the effective feedback customer set according to the "Kendall harmony coefficient." The top k users were selected to be included in a similar result set. For some types of services with obvious differences, the small data were processed. The corresponding reference thresholds for the associated k values were preset in this study. If it was lower than the preset threshold, then it was considered invalid. Thus, the validity of the experimental input was guaranteed.

$$Sim_{attr}(u,v) = \begin{cases} The\ \ actual\ \ value, calculated\ \ similarity\ is\ greater\ than\ the\ threshold \\ 0, The\ calculated\ \ similarity\ value\ is\ smaller\ than\ the\ threshold \end{cases} \tag{4}$$

At the later stage of the experiment, the "product" process was performed per the data and similarity values according to a preset time attenuation factor. Thus, the corresponding similarity was output. The final S value was generated on the basis of the set weights to facilitate further missing prediction.

$$S = s_1 q_1 + s_2 q_2 + \Lambda + s_i q_i \tag{5}$$

$$s_i = e^{-\tau^{t_i}} \tag{6}$$

Finally, the error processing was performed on S to determine the recommended accuracy of the model.

3. Experimental results and analysis

In this part, the performance of the algorithm CSQI will be evaluated, and experiments will be performed by the dataset included in WS-Dream by Zheng et al [15]. In this dataset, the QoS was monitored, and its related information was recorded by calling WSMonitor on the PlanetLab2 platform. Nearly 1.5 million QoS records were collected, and their statistics are shown in Table 1.

Table 1

Statistics of QoS data set

Name	Value
Number of users	150
Number of Web service	100
Country of users	24
Country of services	22
Number of QoS records	⩽1500000

Each piece of data in the statistical information contained the multi-dimensional attribute values produced by certain a user who used certain a cloud service, as shown in Table 2. Among them, the user IP represents the caller of various cloud services. The Service ID (denoted by SID) represents various cloud services called by the customer. Other items in Table 2 represent the QoS attributes of each dimension.

Table 2

Data sample of the Web QoS

User IP	SID	Round-trip time (ms)	Data size	HTTP code	HTTP message
192.33.90.66	9693	887	837	200	OK
129.10.120.193	9566	617	627	200	OK
129.10.120.193	9217	765	1053	200	OK
129.10.120.193	4785	340	595	200	OK
129.33.90.66	5	395	1416	401	Unauthorized
129.33.90.66	8742	20005	2624	-1	Overtime
129.10.120.193	9566	20009	2621	-1	Overtime

(1) Determination of the optimal number of fragments

The eight sets of records were randomly selected from the dataset, and the number of effective service calls by all customers was 5. The ranking per the time period was conducted on QoS after being used by all customers in the record. Specifically, the 64 time slices of all records were divided into 2, 4, 6, 8, and 16 segments one by one. The experiment was ranked according to the round-trip time. One of all combination records was randomly selected as an active customer. Then, the experiment of the cloud service ranking similarity was started by other customers, as shown in Table 3.

Table 3

WS-Dream user data was sorted in four segments

User ID	1…16	17…32	33…48	49…64
User 2	Service2	Service2	Service5	Service1
	Service1	Service1	Service2	Service2
	Service3	Service4	Service4	Service3
	Service4	Service5	Service1	Service4
	Service5	Service3	Service3	Service5

The records in Table 3 were based on five services used by certain a customer at all times.

Fig. 1 displays five sets of QoS records. Fig. 1 shows that the experimental MAE had more prominent effects after multiple fragments. Sixteen sets of records were extracted again, and the service number was increased to 50. The experimental process was repeated, as shown in Fig. 2.

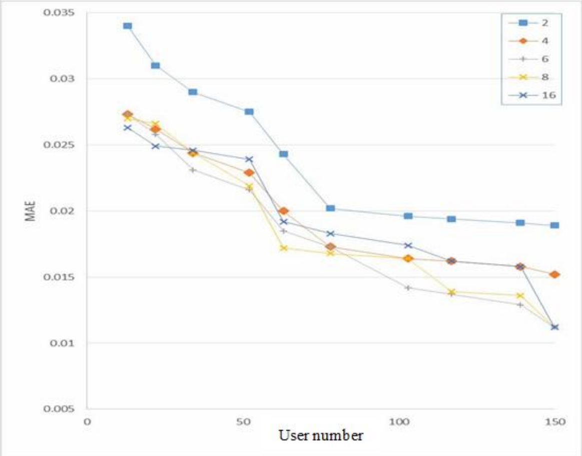

Fig. 1. MAE performance comparison 1

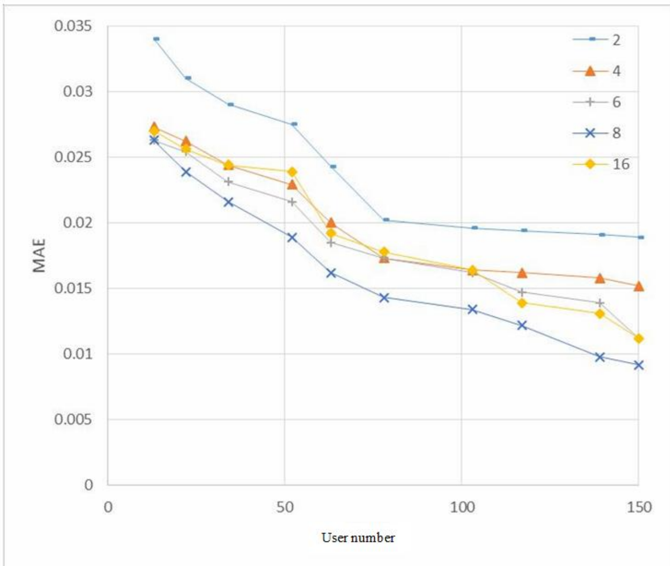

Fig. 2. MAE performance comparison 2

The experimental results from Figs. 1 and 2 showed that the CSQI algorithm had the best performance when the time slice was divided into eight slices.

(2) Comparison with other algorithms

Four prediction algorithms (UPCC, UVS, WRec, and TaCR.) were chosen for comparison [16-18]. Table 4 records the round-trip time and the fault rate of each algorithm. For the convenience of analysis, the “dimensionless method” was conducted on these data, and the error information was enlarged to a hundred times.

Table 4

Comparison with other algorithms

Algorithm	Round-trip time		Fault rate	
	MAE	MRE	MAE	MRE
UPCC	11.630	0.122	21.384	0.192
UVS	9.155	0.087	19.280	0.127
WSRec	9.073	0.083	15.073	0.113
TaCR	9.025	0.072	11.124	0.098
CSQI	8.037	0.066	10.250	0.087

With the continuous addition of the training set in the experiment, an increasing amount of historical information was involved again in the experiment. Thus, the content was generated, as shown in Fig. 3.

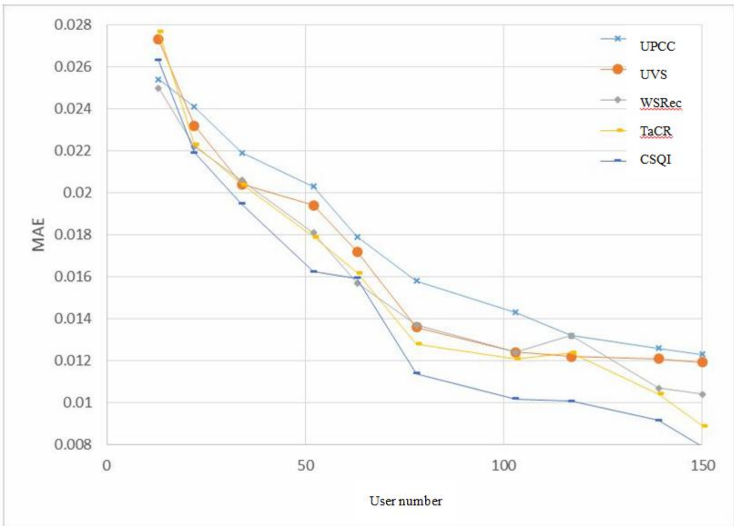

Fig. 3. MAE performance comparison for the five algorithms

In general, UVS and UPCC had higher performances in MAE value than TaCR and CSQI. Compared with TaCR algorithm, the CSQI algorithm exhibited some fluctuations when the experimental data was less. However, the CSQI algorithm had better performance when the experimental data reached a certain threshold. The reason was that the CSQI algorithm considered to remove the influence of noise data and add the regional service similarity factors in the early experiment. Accordingly, the accuracy of the cloud service prediction algorithm could be improved. Thus, it performed well.

4. Summary and prospect

The cloud service QoS prediction algorithm based on the time perception and the user's interest (CSQI) is proposed in this paper. This algorithm could pre-process the dataset, such as removing the negative or weakly related data, which reduces the workload of the prediction. At the same time, the regional similarity factor of the user's interest is considered. The real experimental data show that the proposed algorithm has higher accuracy than other algorithms.

The cloud service prediction is always based on scenarios. The existing method or the model described could not guarantee that the need of complex scenarios could be met in the long run, especially when the real-time recommendation and prediction abilities are insufficiently strong. This aspect will be studied in the future.

Acknowledgements

The research presented in this paper was supported by the scientific research project of education department of Hunan Province (Grant:18B412), the Natural Science Foundation of Hunan Province (Grant: 2019JJ50689), China Postdoctoral Science Foundation (Grant: 2018M642974).

R E F E R E N C E S

[1]. *R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg, I. Brandic*, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility", Future Generation Computer Systems, vol. 25, no. 6, 2009, pp. 599-616

[2]. Z. Zhou, J. Abawajy, L. Fangmin, H. Zhigang, L. Keqin, et al, "Fine-grained Energy Consumption Model of Servers Based on Task Characteristics in Cloud Data Center, IEEE Access, vol. 6, no. 1, 2018, pp. 27080-27090

[3]. *A. A. D. Souza, T. S. Barcelos, R. Munoz, et al*, "Data Mining Framework to Analyze the Evolution of Computational Thinking Skills in Game Building Workshops", IEEE Access, vol. 86, no. 1, 2018, pp. 82848-82866

[4]. *Y. Al-Dhuraibi, F. Paraiso, N. Djarallah, et al*, "Elasticity in Cloud Computing: State of the Art and Research Challenges", IEEE Transactions on Services Computing, 2017

[5]. *R. Wang, P. Shang, J. Zhang, Q. Wang, T. Liu, J. Wang,* "MAR: A novel power management for computer systems in data-intensive environment, IEEE Transactions on Computers, vol. 65, no. 6, 2016, pp. 1816-1830

[6]. *H. Yicun, J. Yaochu, H. Kuangrong,* "A Clustering-Based Adaptive Evolutionary Algorithm for Multi-objective Optimization with Irregular Pareto Fronts", IEEE Transactions on Cybernetics, vol. 49, no. 7, 2019, pp. 2758-2770

[7]. *B. Li, B. Q. Cao, K. M. Wen, et al,* "Trustworthy assurance of service interoperation in cloud environment", International Journal of Automation and Computing, vol. 8, no. 3, 2011, pp. 297-308

[8]. *G. Kang, J. Liu, M. Tang, et al,* "An effective web service ranking method via exploring user

behavior", IEEE Transactions on Network and Service Management, vol. 12, no. 4, 2015, pp. 554-564

[9]. *Z. B. Zheng, H. Ma, M. Lyu, and I. King,* "QoS-Aware Web Service Re commendation by Collaborative Filtering", IEEE Transactions on Service Computing (TSC), vol. 4, no. 2, 2011, pp. 140-152

[10]. *H. Sun, Z.B. Zheng, J. Chen, W. Pan, C. Liu and W. Ma,* "Personalized Open API Recommendation in Clouds via Item-based Collaborative Filterring", In Proc. 4th IEEE International Conference on Utility and Cloud Computing, Melbourne, Australia, 2011, pp. 5-8

[11] *K. Karta,* "An investigation on personalized collaborative filtering for web service selection", Honours Programme thesis, 2005, University of Western Australia: Brisbane.

[12]. *Z. Zheng, M. Lyu*, "Collaborative reliability prediction of service-oriented systems", In Proc. of ICSE'10, Cape Town: ACM, 2010, pp. 35-44

[13]. *L. Chen, L. Hu, Z.B. Zheng, and J. Wu*, "Utilizing Tags for clustering Web services", In Proc. 9th International Conference on Service Oriented Computing, Paphos, Cyprus, 2011, pp. 5-8

[14]. Y. Xu, J. Yin, W. Lo, et. al, "Personalized Location-Aware QoS Prediction for Web Services Using Probabilistic Matrix Factorization", In Web Information Systems Engineering; Springer: Berlin, Germany, 2013, pp. 229–242

[15]. *Y. Zhang, Z.B. Zheng, and Michael R. Lyu*, "Real-Time Performance Pre dictions for Cloud Components", In Proc. the 5th IEEE International Workshop on Real-Time Service-Oriented Architecture and Applications, Shenzhen, China, 2012

[16]. *J. JIN, L. ROTHROCK, L.MCDERMOTTP, et al,* "Using the analytic hierarchy process to examine judgment consistency in a complex multi-attribute task", IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, vol. 40, no. 5, 2010, pp. 1105-1115

[17]. *Haas and A. Brown*, Web Services glossary [EB/OL]. W3C Working Group Note 11, http://w3.org/TR/ws-gloss, 2012

[18]. *C. Lei-lei，Y. Jian，Z. Liang,* "Time based QoS modeling and prediction for Web services", Proceedings of International Conference on Service-Oriented Computing, 2011, pp. 532-540