A CONTEXT-AWARE SERVICE EVALUATION APPROACH OVER BIG DATA FOR CLOUD APPLICATIONS

B. Katharin prema[1], S. Noordeen[2]

[1]*Dept of Computer Science, School of Arts and Science, Vinayaka Mission's Research Foundation, Chennai, India.*
Email: kathrinprema8@gmail.com

[2]*Assistant Professor Grade-II, Dept of Computer Science, School of Arts and Science, Vinayaka Mission's Research Foundation, Chennai, India.*
Email: noordeen@avit.ac.in

ABSTRACT

Cloud computing has promoted the success of big data applications such as medical data analyses. With the abundant resources provisioned by cloud platforms, the QoS (quality of service) of services that process big data could be boosted significantly. However, due to unstable network or fake advertisement, the QoS published by service providers is not always trusted. Therefore, it becomes a necessity to evaluate the service quality in a trustable way, based on the services' historical QoS records. However, the evaluation efficiency would be low and cannot meet users' quick response requirement, if all the records of a service are recruited for quality evaluation. Moreover, it may lead to 'Lagging Effect' or low evaluation accuracy, if all the records are treated equally, as the invocation contexts of different records are not exactly the same. In view of these challenges, a novel approach named Partial-HR (Partial Historical Records-based service evaluation approach) is put forward in this paper. In Partial-HR, each historical QoS record is weighted based on its service invocation context. Afterwards, only partial important records are employed for quality evaluation. Finally, a group of experiments are deployed to validate the feasibility of our proposal, in terms of evaluation accuracy and efficiency.

Keywords—Big data, cloud, context-aware service evaluation, historical QoS record, weight.

1. INTRODUCTION

In recent years, cloud computing and big data have gained increasing attention, from academic research to industrial application domains. With the great advantages of virtualized computing resources, cloud computing technology provides users a novel and attractive resource use model, in an easy-to-access and pay-per-use manner. While the abundance and elasticity of cloud resource provision have enabled cloud users to deploy their various business applications in a scalable manner, from small-scale personal daily use to large-scale big data processing, such as medical data analysis, weather data processing and shopping trend prediction. Currently, a variety of cloud service providers including Sa-

lesforce.com, Amazon Web Services and Microsoft Azure, have provided nearly infinite computing resources to the public community. By utilizing these infinite and powerful cloud computing resources, the services that process big data (abbreviated as 'big data service' in the rest of paper, for the convenient discussions) could be delivered to their users with better QoS (quality of service) performance, e.g., faster execution time, higher availability, and so on.

2. RELATED WORKS

"An Evaluation Method of Outsourcing Services for Developing An Elastic Cloud Platform"

To gain and retain competitive advantages in a competitive business arena, a business cloud-computing platform should continuously strive to offer new services and remain competitive. Unfortunately, it becomes more and more recognized by the industry that a cloud-computing platform could not cover all aspects of IT layers engaged in infrastructure, platform and application. In practice, end users' requests are nearly unlimited; while the services held by a cloud-computing platform is relatively limited, no matter in service category or in service capacity. In view of this challenge, an elastic cloud platform is investigated by recruited outside services that are absent from the cloud platform. Concretely, through dynamically hiring a qualified service on Internet to replace the absent service inside a cloud platform, an elastic cloud platform could nearly provide unlimited capabilities in an outsourcing service way, e.g., computing power, storage, application functions, etc. At last, the validity of the method is evaluated by a case study.

"QoS-Driven Service Selection for Multi-Tenant SaaS"

Cloud-based software applications (Software as a Service - SaaS) for multi-tenant provisioning have become a major development paradigm in Web engineering. Instead of serving a single end-user, a multi-tenant SaaS provides multiple end-users with the same functionality but with potentially different quality-of-service (QoS) values. The service selection for such a SaaS is a complex decision-making process which involves a number of stakeholders with different QoS requirements. SaaS developers need to compose services with different QoS values to meet end-users' different multidimensional QoS constraints for the SaaS. Furthermore, they also need to satisfy SaaS providers' optimization goals for the SaaS, such as least resource cost and best system performance. Existing QoS-aware service selection approaches are oriented at a single tenant. They do not consider the characteristics of multi-tenant SaaS and hence are ineffective and inefficient when applied to compose multi-tenant SaaS. In this paper, we introduce a novel QoS-driven approach for helping SaaS developers select the services for composing multi-tenant SaaS, which achieves SaaS providers'optimisation goals while fulfilling the end-users' different levels of QoS constraints. The proposed approach is evaluated using an example SaaS synthetically generated based on a dataset of real-world Web services. Experimental results show that our approach significantly outperforms existing approaches in terms of both effectiveness and performance.

"Towards Enabling Cyber Infrastructure as A Service in Clouds"

The number of surveillance video data has grown enormously since the popularity of the smart cities, which brings the difficulties to users for searching

and analyzing the content of the videos. These video data are not limited to data anymore, which provide the effective information for criminal investigation systems, intrusion detection system, and many others. However, as the number of available cloud services increases, the problem of data discovery and selection arises. The semantic technology is an effective choice for enhancing the accurate of the data searching and analyzing process. In this paper, a semantic based cloud environment is proposed to facilitate the analyzing and searching process of surveillance video data. An architecture integrating ontology building, semantic annotation, and semantic search is proposed to leverage the semantic description of the video data to find them from concept-based level.

"Dimensionality Reduction of Medical Big Data Using Neural-fuzzy Classifier"

Massive and complex data are generated every day in many fields. Complex data refer to data sets that are so large that conventional database management and data analysis tools are insufficient to deal with them. Managing and analysis of medical big data involve many different issues regarding their structure, storage and analysis. In this paper, linguistic hedges neuro-fuzzy classifier with selected features (LHNFCSF) is presented for dimensionality reduction, feature selection and classification. Four real-world data sets are provided to demonstrate the performance of the proposed neuro-fuzzy classifier. The new classifier is compared with the other classifiers for different classification problems. The results indicated that applying LHNFCSF not only reduces the dimensions of the problem, but also improves classification performance by discarding redundant, noise-corrupted, or unimportant features. The results strongly suggest that the proposed method not only help reducing the dimensionality of large data sets but also can speed up the computation time of a learning algorithm and simplify the classification tasks.

3. EXISTING SYSTEM

The run-time quality of service is heavily affected by the service invocation context, e.g., user input. Third, the big data services often do not run on local servers, but run on re-mote cloud platform. Therefore, user location, or the distance between service and cloud user is a key factor for service quality, because a huge amount of data needs to be transmitted from cloud user to remote service, or from remote service to cloud user. With the above observations, the QoS data advertised by service providers is not always trusted, which makes it necessary to explore more credible evaluation approaches for the big data services in cloud.

Disadvantages of Existing System

- In, the problem of QoS credibility is firstly put forward, and the historical QoS records are suggested to be considered for evaluating the real quality of service.
- In the literature, the service's QoS credibility is calculated, by comparing the historical QoS data with the SLA (Service Level Agreement) promised by service providers.
- However, in the above literatures, the weight problem of different historical QoS records is not discussed.

4. PROPOSED SYSTEM

In this profile, the proposed Partial-HR approach is evaluated from another point of view. Concretely, we com-pare the weight distribution of historical QoS records of Partial-HR with other four approaches: Average, Last-K, AP] and DWF. In AP and DWF approaches, the weights of historical QoS records are assigned in the manners of arithmetical progression and geometric progression respectively. In order to obtain the weight distribution, we calculate the weight of each historical QoS record, based on the five approaches. Afterwards, we select top K records with larger weights, so that their weight sum is at least 0.8 (i.e., K important records occupy 80% of the total weight). By observing the value of K, we can analyze the weight distribution of historical QoS records and further evaluate the rationality of different approaches.

Advantages of the Proposed System

- We cannot always guarantee the advantages of our proposed Partial-HR approach; however, as, we can ensure the advantages of Partial-HR in 'most' cases.
- The above literatures only consider the services' invocation time in the weight model, while omitting other important context elements, e.g., user input and user location.

5. ARCHITECTURE

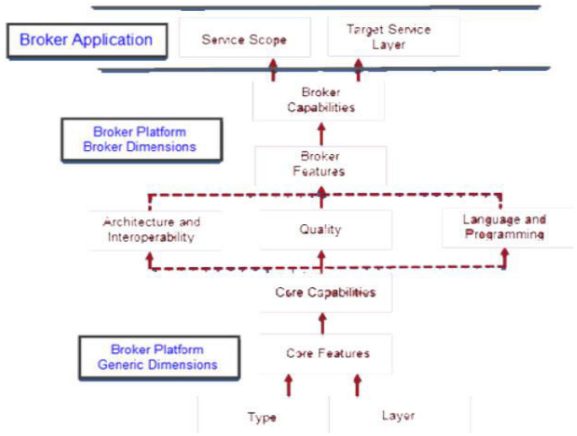

Fig.1.A Context-Aware Based Key Architecture

6. MODULE DESCRIPTION

6.1 Cloud computing

In recent years, cloud computing and big data have gained increasing attention, from academic research to industrial application domains. With the great advantages of virtualized computing resources, cloud computing technology provides users a novel and attractive resource use model, in an easy-to-access and pay-per-use manner. While the abundance and elasticity of cloud resource provision have enabled cloud users to deploy their various business applications in a scalable manner, from small-scale personal daily use to large-scale big data processing, such as medical data analysis , weather data processing and shopping trend prediction. Currently, a variety of cloud service providers including Sa-lesforce.com, Amazon Web Services and Microsoft Azure, have provided nearly infinite computing resources to the public community.

6.2 Invocation time based Weight Modeling

In this subsection, we model the weight of a historical QoS record based on its corresponding invocation time. For the convenient discussions, we assume that big data service BD Service own L historical QoS records, and the invocation time of these L records are denoted by respectively. Without loss of generality, we assume that. For example, means that occurred before occurred at the same time). Next, we introduce how to design weight W for historical QoS record, based on invocation time tk. Actually, our weight model refers to several classic strategies adopted in E-commerce, as service evaluation is similar to commodity evaluation to some extent.

6.3 Input size-based Weight Modeling

In this subsection, the weight of a historical QoS record is modeled based on the record's input size (actually, be-sides input size, input type and other input items can also affect the service quality; however, to facilitate the sub sequential calculation, we only consider input size here). For the convenient discussions, we assume the input sizes of L historical QoS records of BD Service are denoted by size1, size L respectively, and the input size of upcoming service invocation from New User is de-noted by size new. Next, we introduce how to model weight, based on hrk's input size k.

6.4 User location-based Weight Modeling

As big data services often run in remote cloud platform, user location becomes an important factor that affects the running quality (e.g., time cost) of the services. For example, if a user is 'close' to his/her invoked service BD Service (e.g., user and BD Service locate in an identical university), then the time cost spent on data input, data processing, data transmission and data output would be small. To depict this kind of correlation, in this subsection, we study the relationship between weight W(user location)k of historical QoS record hrk and hrk's user location information location k.

7. RESULTS

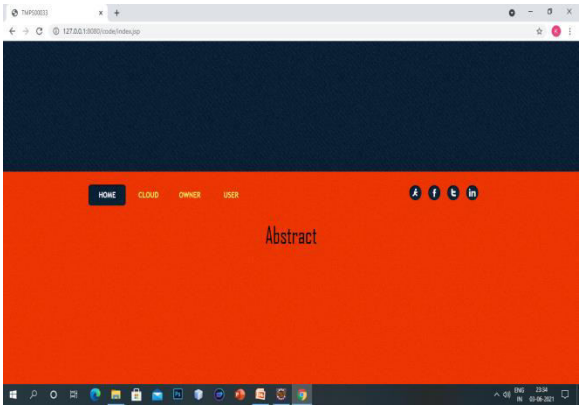

Fig2: Home page

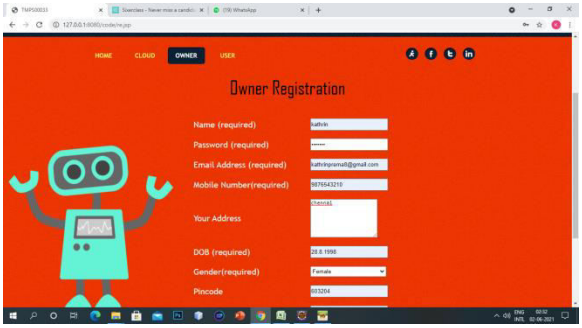

Fig3: Owner Registration

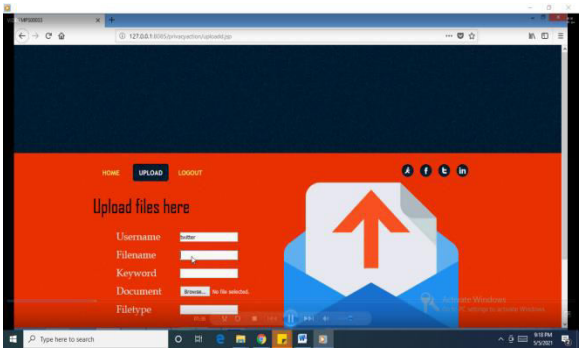

Fig4: Uploading Files

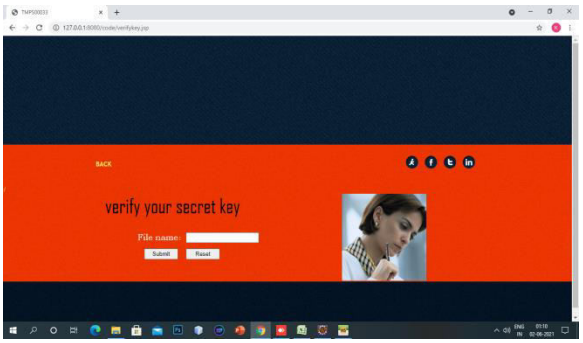

Fig5: Verify Secret Key

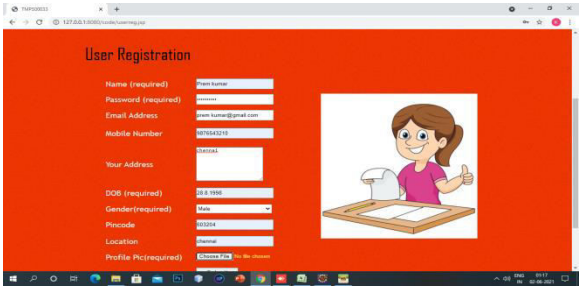

Fig6: User Registration

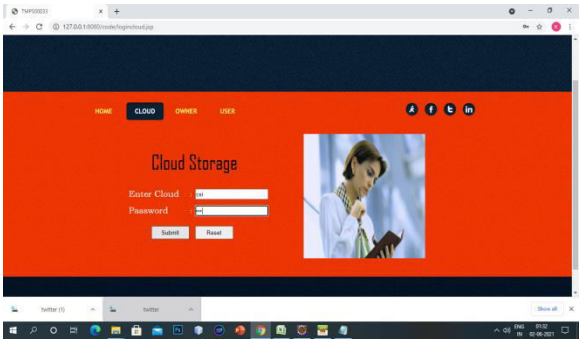

Fig7: Cloud Storage

8. CONCLUSION

Due to the unstable network or fake advertisement, the QoS information of services that process big data in cloud, is not always trustable as advertised by service providers. Therefore, it becomes a necessity to evaluate the service quality in a trustable way based on the historical QoS records. However, it may lead to low efficiency if all the records are considered in service quality evaluation. Moreover, evaluation accuracy would be low if all the historical QoS records are treated equally, as their service invocation contexts are not exactly the same. In view of these challenges, a novel evaluation approach named Partial-HR is proposed in this paper, which not only considers the service invocation context, but also satisfies 'Volatility Effect' and 'Marginal Utility' simultaneously. Through a set of experiments, we validate the feasibility of Partial-HR in terms of evaluation accuracy and efficiency. In the future, we will introduce more context elements into our weight model for historical QoS records, so as to further improve the evaluation accuracy of big data services in cloud.

REFERENCES

[1] W. Dou, L. Qi, X. Zhang, and J. Chen, "An Evaluation Method of Outsourcing Services for Developing An Elastic Cloud Platform", The Journal of Supercomputing, vol. 63, no. 1, pp. 1-23, January 2013.

[2] Q. He, J. Han, Y. Yang, J. Grundy, and H. Jin, "QoS-Driven Service Selection for Multi-Tenant SaaS", Proc. 5th IEEE Int'l Conf. on Cloud Computing (CLOUD'12), pp. 566-573, June 2012.

[3] L. Wang, D. Chen, Y. Hu, Y. Ma, and J. Wang, "Towards Enabling Cyber Infrastructure as A Service in Clouds", Com-puters& Electrical Engineering, vol. 39, no. 1, pp. 3-14, 2013.

[4] A. Azar, and A. Hassanien, "Dimensionality Reduction of Medical Big Data Using Neural-fuzzy Classifier", Soft Computing, 2014. (DOI: 10.1007/s00500-014-1327-4).

[5] T. Murdoch, and A. Detsky, "The Inevitable Application of Big Data to Health Care", JAMA, vol. 309, no. 13, pp. 1351-1352, April 2013.

[6] G. Kim, S. Trimi, and J. Chung, "Big-data Applications in the Government Sector", Communications of the ACM, vol. 57, no. 3, pp. 78-85, March 2014.

[7] V. Schönberger, and K. Cukier, “Big Data: A Revolution that Will Transform how We Live, Work, and Think”, published by Houghton Mifflin Harcourt, 2013.

[8] W. Dou, X. Zhang, J. Liu, and J. Chen, “HireSome-II: Towards Privacy-Aware Cross-Cloud Service Composi-tion for Big Data Applications”, IEEE Transactions on Pa-rallel and Distributed Systems, 2013.