Cloud Service Recommendation: State of the Art and Research Challenges

Lantian Guo[1,4], Xianrong Zheng[2], Chen Ding[3], Dejun Mu[1], Zhe Li[1]
[1]*School of Automation, Northwestern Polytechnical University, Xi'an, Shaanxi, China 710072*
[2]*Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, China 518055*
[3]*Department of Computer Science, Ryerson University, Toronto, Ontario, Canada M5B 2K3*
[4]*School of Computing, Queen's University, Kingston, Ontario, Canada K7M 2W8*
guolt0211@gmail.com, xr.zheng@siat.ac.cn, cding@ryerson.ca, mudejun@nwpu.edu.cn

***Abstract*—Cloud computing is an attractive platform which offers on-demand resources as services. When many cloud services are available, some may have the same or similar functionalities. So cloud service recommendation become an important technique for cloud services. It can help users to select services that meet their preferences, In this paper, we review relevant techniques for conducting cloud service recommendation. First, we introduce web service recommendation techniques. Next, we describe QoS models recommendation methods for cloud services. Third, we discuss research challenges for cloud service recommendation.**

***Keywords*-Cloud computing; service recommendation; QoS measurement; collaborative filtering**

I. Introduction

Cloud computing is an attractive paradigm which provides computing, storage and networking resources as services. IT providers (e.g., Google, Microsoft, Amazon, etc.) strive to offer powerful and reliable cloud services to their customers in an on-demand pay-as-you-go model [7], [25]. At the same time, both individual and enterprise customers apply cloud services to their commercial information systems or personal applications.

As the increasing number of cloud services provide the same or similar functionalities, it is not easy for services users to find a suitable cloud service among different providers [7]. In order to differentiate them, Quality of Service (QoS) which are non-functional attributes (e.g., response time, throughput, reliability, availability, security, etc.) is used to evaluate services [3].

There is an increasing need to assist non-expert users to select cloud services cloud services that meet their QoS preferences [29]. A key approach to address above mentioned is recommendation systems. Recommendation systems have been proven to be an effective approach to deal with information overload and make use of available information for providing users with rating prediction and recommendation lists [1]. In this paper, we conduct a survey on cloud service recommendation, highlighting its key concepts, state-of-art and research challenges. Its main contributions are as follows:

1) We argue that cloud service recommendation is important. Our work may draw more attention for cloud service recommendation.

2) We review recommendation methods both for both web service and cloud service, and focus on Collaborative Filtering (CF).

3) We identify some research problems for cloud service recommendation are surveyed respectively, and give some ideas to solve them.

The rest of this paper is organized as follows: Section II introduces web service selection and recommendation technologies because of the relevance of cloud services and web services. Section III discusses QoS models, recommendation methods, and research challenges for cloud service recommendation. Section IV concludes the paper.

II. Web Service Recommendation

Web services are software systems designed for machine-to-machine interaction over networks [18]. QoS is an important factor for web service selection and recommendation [28]. Some QoS attributes (e.g., reliability, security etc.) are difficult to obtain, since a number of long duration observations are required. Besides those, service providers may not provide the QoS that their Other QoS attributes (e.g., response time) are affected by the location and network conditions of service users [3].

Accurate QoS prediction is important in web service recommendation [17]. QoS prediction approaches include clustering-based [30], content-based [21], CF-based [28] and hybrid approaches [23]. Most work on service selection and recommendation are CF-based. CF identifies similar users for an active user and predicts QoS values based on the preferences of its neighbors, i.e., a group of similar users [24].

CF methods include neighborhood-based, model-based and hybrid approaches. Neighborhood-based methods include user-based, service-based and hybrid approaches [29]. A user-based approach predicts missing values of a user-service matrix based on the similarity of two users, while a service-based approach is based on the characteristics of services similar to those selected by an active users

978-1-4799-8006-2/15 $31.00 © 2015 IEEE
DOI 10.1109/CCGrid.2015.144

[4]. A hybrid approach is the combination of user-based and service-based, which can achieve a higher QoS prediction accuracy compared with the alone one [29]. Most neighborhood-based approaches employ the Pearson Correlation Coefficient (PCC) to compute similarity, while others use the Cosine Correlation Coefficient or constrained PCCs.

Model-based approaches provide QoS prediction results by creating a model of user preferences. One model-based approach is establishing the model by probabilistic and learning algorithms (e.g., K-means clustering, latent-factor, etc.) [3]. Another popular model-based approach is matrix factorization methods, which employ latent factor models to detect implicit information [9]. By matrix factorization, users and services can be characterized by latent factors, where a user and a service are mapped onto a latent feature space. As a result, a user-service matrix can be divied into a user factor matrix and a service factor matrix. Model-based approaches can generate prediction results efficiently, but a model should be updated when new users or services enter.

Moreover, there are many work integrating some interaction information (e.g., location [4], [5], reputation [15], time series [20], etc.). The location factor could help to make more achieve accurate prediction. Chen, et al. (2014) proposed a LoRec recommendation approach employing both location and QoS information to cluster users and services [5]. Qiu, et al. (2013) proposed a reputation-aware approach that utilizes reputation to identify untrustworthy user [15]. In addition, service invocation time may affect QoS prediction.

III. Cloud Service Recommendation

A. *QoS Model in Cloud Service*

A fundamental task of cloud services selection and recommendation system is to help users select their desired ones from many cloud services and dynamic cloud environment. Cloud services are relevant to web services. For example, SaaS that provides approaches to access and manage applications as a service offers web-based interfaces [19]. As with web service recommendation, a main issue in cloud service selection and recommendation is QoS evaluation. Li, et al. (2013) argue that QoS model provides multiple QoS attributes to assess a service [10], and so it enables a userto find the most suitable one for user [6].

Li, et al. (2013) presented a QoS model considering five attributes, i.e., response time, pricing, throughput, security and availability [10]. Its quality dimensions contain price but reliability is missed out.

Ardagna, et al. (2013) proposed a cloud service QoS model for interactive cloud resource process and service provision capabilities. He argue that considering QoS of cloud services to providers and consumers, QoS model attributes should contain response time, price, availability, reliability and reputation [14].

In order to measure and compare the quality of the cloud providers, Zheng et al. (2014) proposed a quality model for cloud services named CLOUDQUAL [27]. In this model, six attributes, i.e., usability, availability, reliability, responsiveness, security, and elasticity are used to describe cloud services. The model can differentiate service quality, and an evaluation on storage clouds is carried out.

Nevertheless, measuring QoS attributes is still a difficult task. Although many QoS attributes and model are proposed, only few QoS attributes are quantified for model assessment. For instance, Wang, et al. (2014) used response time to show model can distinguish cloud storage [22]. Li, et al. (2013) employs both respond time and throughput to evaluate cloud service ranking model [10]. In the future, measuring models for more attributes should be carried out to achieve the differentiation of service quality.

B. *Research Problems, Methods and Expected Contributions*

Unlike web services, cloud services provide scalable access to applications and resources. Cloud computing refers to a large pool of virtualized resources that can be dynamically configured to provide elastic services that meet the needs of its users [27]. Collaborative Filtering, which is an effective approach for exploiting users' preferences are widely adopted in personalized recommendation systems [4], [24]. However, due to the characteristics of cloud services, recommendation approaches need to be improved for cloud services. New challenges need to be addressed when conducting cloud service selection and recommendation.

One of the challenges is data sparsity. In recommendation system, the QoS rating is represented as a user-service rating matrix generally. However, since many users have only used a limited number of cloud services, the QoS data obtained are usually very limited. As a result, recommendation systems are unable to predict the ratings of a user for a service accurately [24], since the QoS information is limited.

One approach to addressing the data sparsity is exploiting content data to enhance the CF. The content data can extend data source for a user-service matrix. User review is an important content auxiliary data source to acquire information about services. Nevertheless, there is little work on how user reviews can be transferred and incorporated into cloud service recommendation. we plan to utilize user reviews to enhance the cloud service recommendation.

For transferring the review data to traditional CF, we plan to employ user feature priority weighting method to model users opinion and distinguish them. A way of learning a model using weighted word counts as the feature space is TF-IDF (Term Frequency–Inverse Document Frequency). TF-IDF is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus [16]. With that scheme, service QoS features influencing users overall opinions on different services can be measured. Therefore, users' review data is able to make different ratings in recommendation system.

Beside the user review, there are many contextual data sources (e.g. location, weather and profile, etc.) that affect the prediction performance in recommendation need to be considered as well. However, existing context-aware recommendation approaches cannot effectively combine various contextual data sources (e.g., discrete value, versus continuous value, etc.). Social network, where users share similar tastes and interests, brings an additional contextual information to enhance recommendation models. However, a problem hinders the application of either the user review method or the social network method is how to transfer diverse types of contextual data sources to user-service matrix.

Recently, transfer learning [13] has been proposed to address the data sparsity problem by using the data from other related domains. Traditional learning algorithms use statistical models that are trained on previously historical data, and make predictions on the future data. In contrast, transfer learning applies knowledge learned from one context to other related contexts [12]. Many transfer learning methods have been developed to treat auxiliary data as constraints to improve the predictive performance of collaborative filtering [13], [26] and applied to predict missing values by heterogeneous user feedbacks.

Transfer learning is widely employed to model the rating or ranking prediction problem in CF and allow knowledge to be transferred across different domains. Service recommendation system aim at providing a personalized list of services ranked according to the preferences of service users. Services are typically ranked by decision making strategies [11]. Generating a personalized recommendation list can be cast as a ranking problem. Current research work on CF have exploited learning to rank technique for improving the ranking of recommendations. Learning to rank employs information retrieval and machine learning techniques to automatically obtain the ranking model using a labeled training dataset [8].

CF methods make predictions on users' preferences, and then learn a ranking model to recommend a small set of services to the users [2]. However, recommendation system is difficult to learn a great ranking model by a sparse data source. To address that problem, transfer learning is explored to help a given learning task using a related but slightly different auxiliary task. Learning to rank problem is composed by a set of training queries where for each query we are given a set of retrieved documents and their relevance labels. Transfer learning for ranking method gives individual queries together with its related documents. This method is able to transfer the knowledge contained in one ranking dataset to another ranking dataset. With this method, useful knowledge in source domain is bridged to help the target domain via several features ,and a personalized ranking model for individual uses can be achieved.

Beside data sparsity, cold start and scalability are two problems for cloud service recommendation too. The cold start problem occurs when new users or cloud services entered a recommendation system. The matrix factorization approach has been adopted user and service communities, since it can deal with a high-dimensional matrix [9]. Considering applying nonnegative matrix factorization directly without learning models may be difficult to cluster the communities. At the same time, although many methods such as nonnegative matrix factorization can achieve the clustering of users and services, computational complexity will be extremely high when systems have large-scale matrix. Thus, an future direction for learning algorithms is to design schemes to parallel the workload of these algorithms, since there are a large amount of user and services in recommendation systems.

Yu, et al. (2014) proposed a CloudRec method that employ the nonnegative matrix factorization to extract the inherent features of cloud services [24], which help to cluster the users and services into groups. Whereas Yu, et al. (2014) design an iterative algorithm to deal with the an incomplete QoS matrix, the development of other learning models need to be addressed [24].

IV. Conclusion

In this paper, we conduct a literature review on QoS and service recommendation methods for cloud services. We also identify some research problems, and propose some ideas to solve them.

V. Acknowledgements

We would like to thank anonymous reviewers for their constructive comments. This work was supported in part by the Natural Science Foundation of China (NSFC) under Grant 61303224, and NSFC under Grant 61403311.

References

[1] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-Based Systems*, vol. 46, pp. 109–132, 2013.

[2] D. Chen, Y. Xiong, J. Yan, G.-R. Xue, G. Wang, and Z. Chen, "Knowledge transfer for cross domain learning to rank," *Information Retrieval*, vol. 13, no. 3, pp. 236–253, 2010.

[3] X. Chen, Z. Zheng, X. Liu, Z. Huang, and H. Sun, "Personalized qos-aware web service recommendation and visualization," *IEEE Transactions on Services Computing*, vol. 6, no. 1, pp. 35–47, 2013.

[4] X. Chen, Z. Zheng, and M. R. Lyu, *QoS-aware Web Service Recommendation via Collaborative Filtering*. New York,USA: Springer, 2014.

[5] X. Chen, Z. Zheng, Q. Yu, and M. R. Lyu, "Web service recommendation via exploiting location and

qos information," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 7, pp. 1913–1924, 2014.

[6] D. Y. Cheng, K. M. Chao, C. C. Lo, and C. F. Tsai, "A user centric service-oriented modeling approach," *World Wide Web*, vol. 14, no. 4, pp. 431–459, 2011.

[7] S. Han, M. Hassan, C. Yoon, H. Lee, and E. Huh, "Efficient service recommendation system for cloud computing market," *Grid and Distributed Computing*, vol. 63, pp. 117–124, 2009.

[8] A. Karatzoglou, L. Baltrunas, and Y. Shi, "Learning to rank for recommender systems," in *Proceedings of the 7th ACM Conference on Recommender Systems*, 2013, pp. 493–494.

[9] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, no. 8, pp. 30–37, 2009.

[10] J. Li, L. Meng, Z. Zhu, X. Li, J. Huai, and L. Liu, "Cloudrank: A cloud service ranking method based on both user feedback and service testing," *Principles, Methodologies, and Service-Oriented Approaches for Cloud Computing*, pp. 230–259, 2013.

[11] T.-Y. Liu, *Learning to rank for information retrieval*. Springer, 2011.

[12] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[13] W. Pan, E. W. Xiang, and Q. Yang, "Transfer learning in collaborative filtering with uncertain ratings." in *Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI 2012)*, pp. 39–55.

[14] A.-S. K. Pathan, M. M. Monowar, and Z. M. Fadlullah, *Building Next-generation Converged Networks: Theory and Practice*. CRC Press, 2013.

[15] W. Qiu, Z. Zheng, X. Wang, X. Yang, and M. R. Lyu, "Reputation-aware qos value prediction of web services," in *Proceedings of the 10th IEEE International Conference on Services Computing (SCC 2013)*, 2013, pp. 41–48.

[16] A. Rajaraman and J. D. Ullman, *Mining of massive datasets*. Cambridge University Press, 2011.

[17] L. Shao, J. Zhang, Y. Wei, J. Zhao, B. Xie, and H. Mei, "Personalized qos prediction forweb services via collaborative filtering," in *Proceedings of the 5th IEEE International Conference on Web Services (ICWS 2007)*, pp. 439–446.

[18] H. Sun, Z. Zheng, J. Chen, and M. R. Lyu, "Personalized web service recommendation via normal recovery collaborative filtering," *IEEE Transactions on Services Computing*, vol. 6, no. 4, pp. 573–579, 2013.

[19] L. Sun, H. Dong, F. K. Hussain, O. K. Hussain, and E. Chang, "Cloud service selection: State-of-the-art and future research ddirections," *Journal of Network and Computer Applications*, vol. 45, pp. 134–150, 2014.

[20] G. Tian, J. Wang, K. He, P. C. Hung, and C. Sun, "Time-aware web service recommendations using implicit feedback," in *Proceedings of the 21st IEEE International Conference on Web Services (ICWS 2014)*, pp. 273–280.

[21] M. Voigt, S. Pietschmann, L. Grammel, and K. Meißner, "Context-aware recommendation of visualization components," in *Proceedings of the 4th International Conference on Information, Process, and Knowledge Management (eKNOW 2012)*, pp. 101–109.

[22] S. Wang, Z. Liu, Q. Sun, H. Zou, and F. Yang, "Towards an accurate evaluation of quality of cloud service in service-oriented cloud computing," *Journal of Intelligent Manufacturing*, vol. 25, no. 2, pp. 283–291, 2014.

[23] Q. Xie, K. Wu, J. Xu, P. He, and M. Chen, "Personalized context-aware qos prediction for web services based on collaborative filtering," in *Advanced Data Mining and Applications*, 2010, vol. 6441, pp. 368–375.

[24] Q. Yu, "Cloudrec: A framework for personalized service recommendation in the cloud," *Knowledge and Information Systems*, vol. 43, no. 2, 2014.

[25] Q. Zhang, L. Cheng, and R. Boutaba, "Cloud computing: state-of-the-art and research challenges," *Journal of internet services and applications*, vol. 1, no. 1, pp. 7–18, 2010.

[26] Y. Zhang, B. Cao, and D. Y. Yeung, "Multi-domain collaborative filtering," *arXiv*, 2012.

[27] X. Zheng, P. Martin, K. Brohman, and L. D. Xu, "Cloudqual: A quality model for cloud services." *IEEE Transaction on Industrial Informatics*, vol. 10, no. 2, pp. 1527–1536, 2014.

[28] Z. Zheng, H. Ma, M. R. Lyu, and I. King, "Wsrec: A collaborative filtering based web service recommender system," in *Proceedings of the 7th IEEE International Conference on Web Services (ICWS 2009)*, pp. 437–444.

[29] Z. Zheng, H. Ma, M. R. Lyu, and I. King, "Qos-aware web service recommendation by collaborative filtering," *IEEE Transactions on Services Computing*, vol. 4, no. 2, pp. 140–152, 2011.

[30] J. Zhu, Y. Kang, Z. Zheng, and M. R. Lyu, "Wsp: A network coordinate based web service positioning framework for response time prediction," in *Proceedings of the 19th IEEE International Conference on Web Services (ICWS 2012)*, pp. 90–97.