International Journal of Security and Its Applications
Vol.9, No.8 (2015), pp.35-44
http://dx.doi.org/10.14257/ijsia.2015.9.8.04

Personalized Context-aware Recommendation Approach for Web Services

Zhang Xue-Jie[1,2], Wang Zhi-Jian[1] and Zhang Wei-Jiang[1]

[1]College of Computer and Information, Hohai University, Nanjing 210098, China
[2]College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
xuejie_zh@hhu.ed.cn

Abstract

With the increasing number of Web services, the goal of consumers becomes to discover and use services that lead to their experiencing the highest quality. Quality of Service (QoS) is important to evaluate the QoS performance of services to differentiate the qualities of service candidates. QoS is highly related to context information since service consumers are typically distributed in different geographical locations. Their experience is usually different. Invoking a huge number of Web services for consumers to predict the quality is time-consuming, resource- consuming, and sometimes even impractical. To address the challenge, this paper proposes a personalized context-aware recommendation approach for predicting the QoS of Web services and designs a prediction framework. This algorithm is a hybrid of the model-based and memory-based collaborative filtering algorithms. In our experiment, we collect QoS information from geographically distributed service consumers through the framework. Based on the QoS and context information, we predict the quality of services. As a result, we can obtain a list of recommended services for selection. Finally, the experiment shows that the algorithm using context information achieves better prediction.

Keywords: *Web services, QoS, personalization, context-aware, recommendation*

1. Introduction

Web service is defined as a software system designed to support interoperable machine-to-machine interaction over a network [1]. It is great importance in Service Oriented Architecture (SOA). In reality, the goal of service consumers is to discover and use services which lead to them experiencing the highest quality, so that their expectations and needs are satisfied. Quality of Service (QoS) is usually employed to describe the non-functional properties of services. It is important to evaluate the QoS performance of services to differentiate the qualities of service candidates. Web services are usually deployed in remote servers and accessed by consumers through Internet connections. Moreover service consumers are typically distributed in different geographical locations. For different consumers, QoS performance of Web services they experienced is different. QoS is a set of properties, including response time, throughput, correctness, etc. It's time-consuming and resource-consuming to invoke all Web services each requested, as there may be a large number of function-equivalent candidate Web services [2-3]. Therefore, it's necessary to predict the QoS before invoking the service.

There is several works being done to apply collaborative filtering (CF) in web service recommendation [4-7]. In these approaches, QoS values are predicted for an active user based on the QoS records provided by users who have similar historical QoS experiences on some web services. However, it didn't consider the context factor of consumers or services in these recommendation approaches. Context-free

ISSN: 1738-9976 IJSIA

is a common problem in the service selection and recommendation. Quality of Web services is closely related to the context of services, consumers and the providers. Therefore, to improve the efficient and effective of recommendation, it should take the context information into account during the process of predicting QoS.

To address this problem, this paper presents a Personalized Context-aware Prediction Approach (PCPA) to predicting QoS for Web services. QoS is regarded as a set of consumer-perceived properties, so it highly relates to consumers' physical locations. According to our observation, consumers often have different experiences on the quality of the same web service. Generally, the consumers, who are at the same or close locations, would have similar experiences on the same service. Our approach is based on the method of Clustering-Based QoS Prediction (CBQP), taking the consumers' location information into consideration. It is a hybrid of the model-based and memory-based collaborative filtering algorithms. CBQP has divided the consumers into several clusters according to the delivery QoS [8]. Our algorithm will first cluster consumers in the same subset into several regions based on their physical locations. Then region-sensitive services are identified and the similarity among regions is calculated. After that, the similar and insensitive region is aggregated to be used to predict the QoS of the candidate web services for an active consumer. Based on the predicted QoS, the top-k services will be recommended to the active consumer.

The remainder of this paper is organized as follows. Section 2 presents the personalized context-aware recommendation system. Section 3 describes the QoS prediction approaches in detail. Section 4 discusses the experiments and results. Finally, the section "Conclusion" concludes the paper.

2. Personalized Context-aware Recommendation System

Context is any information that can describe the characteristics of an entity [9]. Context-aware computing affects the capability which applications discover and response the environment changes [10]. Context-based prediction method is based on the preferences of similar users with similar projects in similar situations [11]. Context information includes the physical locations of users and services, time etc. This paper involves two kinds of context, that is, user-context and web service context.

Figure 2.1 shows the personalized context-aware recommendation framework for web services. Within the prediction platform, there are seven modules implemented for managing the web services. Task Scheduler is responsible for task scheduling. QoS Monitor is responsible for monitoring the QoS performance of web services from the consumers' perspective. Context Manager is responsible for capturing the context information of consumers and services. Similar consumers/services Aggregation is used to cluster consumers/services into several regions based on their physical locations. QoS Predictor is supposed to provide personalized QoS value prediction for different consumers. Recommender returns the top-k web services to the active consumer. Quality Database (QDB) is used to store the quality data from monitor.

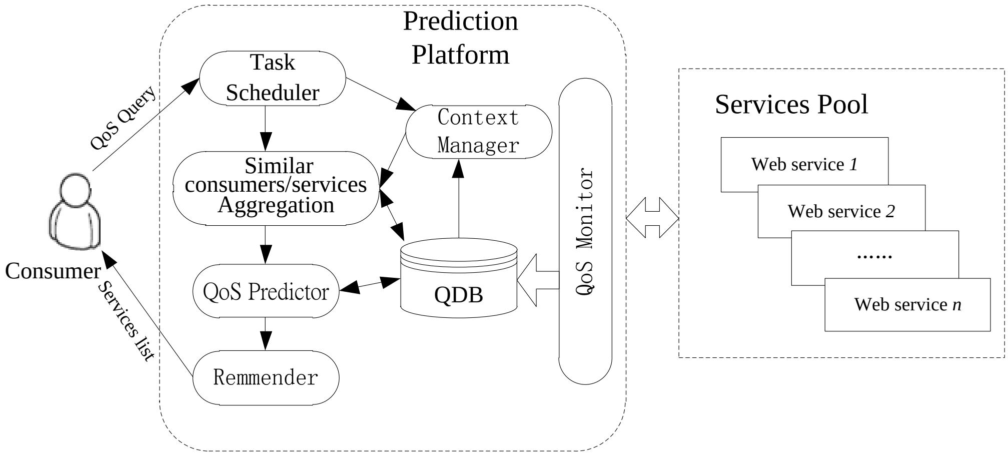


Figure 2.1 Personalized context-aware Web services recommendation Framework

Personalized context-aware services recommended procedure is as follows.

i. **QoS Query:** Consumer, namely services user, sends query request to prediction platform.
ii. **Similar consumers/services Aggregation:** Context Manager provides personalized context information about consumers and services. According to this information, our algorithm will cluster consumers/services into several regions.
iii. **Prediction:** The probable QoS value is predicted by the proposed method PCPA.
iv. **Ranking:** Candidate services are ordering by the prediction result.
v. **Recommendation:** According to the ranking result, the top-k web services will be recommended to the active consumer.

Similar consumers aggregation and the QoS prediction are the two key parts of the framework. Next, we present the personalized context-aware prediction approach in detail.

3. Personalized Context-aware Prediction Approach

This approach (PCPA) has two phases: region model building (offline) and the QoS prediction (online).

3.1 Region Model Building

Region model is a model for compressing QoS data by clustering consumers into different regions. We define a region as a group of consumers who are closely located with each other and have similar experience on a certain property. There are three parts in building region model: region-sensitive service detection, region similarity computing and region aggregation.

3.1.1 Region-Sensitive Service Detection: From large number of QoS records, we observe that for a certain number of web services, their QoS value of some properties vary greatly from region to region. Our detection objective is to distinguish those services from others.

Given a service s, the set of non-zero $A = \{A_1(s), A_2(s), \cdots, A_k(s)\}$ $(1 \le k \le n)$ is a sample from property A, collected from all users of service s. To estimate the mean $\overline{A}$ and the standard deviation σ of A, we use median instead of average value, since average value is strongly affected by isolated point. The mean $\overline{A}$ and the standard deviation σ are calculated by equation (1) and (2) separately.

$$\overline{A} = median_i(A_i(s)), \quad i = 1, \cdots, k \tag{1}$$

$$\sigma = \sqrt{\frac{1}{k}\sum_{i=1}^{k}\left(A_i(\mathrm{s}) - \mu\right)^2} \tag{2}$$

Definition1. **Region-Sensitive Service:** Let $A = \{A_1(s), A_2(s), \cdots, A_k(s)\}$ $(1 \le k \le n)$ be the set of *A* of service *s* provided by consumers from all regions. Service *s* is a sensitive service to region *R iff* $\exists A_i(s) \in A, (A_i(s) > \overline{A} + 3\sigma) \wedge region(i) = R$,

where $\overline{A} = median(A)$ and $\sigma = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(A_i(\mathrm{s}) - \mu)^2}$.

The decision is based on 3 sigma criteria.

Definition2. **Region Sensitivity:** The sensitivity of region *R* is the fraction between the numbers of sensitive services in region *R* over the total number of services.

Definition3. **Sensitive Region:** Region *R* is a sensitive Region *iff* its region sensitivity exceeds the sensitivity threshold λ.
Definition4. **Region center:** The center of region *R* is defined as the median of all the QoS of *A* provided by consumers in region *R*.

3.1.2 Region Similarity Computing

The similarity of two regions M and N is measured by the similarity of their region centers. In our algorithms, we employ Pearson Correlation Coefficient (PCC) to calculate the similarity of two regions. PCC has a range of [-1, 1]. PCC>0 means the two region consumers have similar preferences, while PCC<0 indicates that their preferences are opposite. Based on PCC, the similarity of the two region centers m and n is calculated by equation (3) as follow:

$$Sim(m,n) = \frac{\sum_{s \in S(m) \cap S(n)} \left(A_m(s) - \overline{A_m}\right)\left(A_n(s) - \overline{A_n}\right)}{\sqrt{\sum_{s \in S(m) \cap S(n)} \left(A_m(s) - \overline{A_m}\right)^2}\sqrt{\sum_{s \in S(m) \cap S(n)} \left(A_n(s) - \overline{A_n}\right)^2}} \tag{3}$$

where S(m) is the set of services invoked by consumers in region M, and S(n) is the set of services invoked by consumers in region N. Since PCC only accounts for the QoS difference between the services invoked by both regions, it often overestimates the similarities when the two regions have few co-invoked services [12]. To address the limit, we use Jaccard Coefficient to adjust the equation (3), as equation (4):

$$Sim'(m,n) = \frac{\left|S(m) \cap S(n)\right|}{\left|S(m) \cup S(n)\right|} Sim(m,n) \tag{4}$$

where $\left|S(m) \cap S(n)\right|$ is the number of web services invoked by both regions, and $\left|S(m) \cup S(n)\right|$ is the number of web services invoked either by region *M* or by region *N*. The fraction between them is the Jaccard Coefficient.

3.1.3 Region Aggregation: Since each consumer only provides a limited number of QoS values, the QoS dataset of each region is very sparse, which always leads to poor recommendation. To address this problem, we propose region aggregation, which is a bottom-up hierarchical clustering algorithm [13]. In each consumer cluster, the algorithm treats consumers with similar IP addresses as a region and then successively aggregates pairs of regions until the stopping criterion has been met.

Algorithm 1：Region Aggregation

Input：N regions：$r_1,\ldots,r_N$

Output：k aggregation regions

//Computing the similarity of two regions and judging the region sensitivity

FOR i = 1 TO N-1

  //Computing the center m of region r_i

  $m = median(r_i)$

  FOR j = i+1 TO N

    // Computing the center n of region $=r_j$

    $n = median(r_j)$

      //Computing the similarity between region r_i and r_j using equation 4

    $S[i][j] = Sim(m,n)$

  END FOR

  //Judging whether the region r_i is sensitive, the result is stored in vector P

  $P[i] = Sensitive(r_i)$

  //Aggregating the similar regions

  IF $S[i][j] \geq \mu$ and $P[i] = 0$

  THEN $r_i = Aggregate(r_i, r_j)$

END FOR

The algorithm 1 presents the process of region aggregation. S is an $N \times N$ matrix storing similarities between each two regions. *P* is a vector indicates which regions are non-sensitive and still available to be aggregated. The result is *k* regions. Function *Sim (m, n)* computes the similarity between two regions using adjusted PCC. *Sensitive(r)* judges whether region *r* is sensitive, and *Aggregate* (r_i, r_j) aggregates regions r_i and r_j. In each iteration, the two most similar and non-sensitive regions are selected and aggregated, if their similarity exceeds threshold μ. The algorithm executes at most *N-1* steps, in the worst case that all regions are non-sensitive and extremely correlates to each other. Finally all regions aggregate into one region. The time complexity of region aggregation algorithm is $O(N^2)$, and *N* is the number of regions at the outset.

3.2 QoS Prediction

After building the region model, we conduct to predict QoS. Traditional CF algorithm [4-5,14] is to search the entire dataset, which suffers from low efficiency. Different from that, after the phase of region aggregation, thousands of users are clustered into a certain number of regions based on their physical locations and historical QoS similarities. In other words, after determining the consumers' expectation range according to the historical QoS, we cluster the consumers based on their physical locations. In doing so, consumers in each region have the similar expectation and context information.

The characteristic of the consumers group in a region is represented by the region center. So the similarity between the active consumer and consumers of a region is computed by the similarity between the active consumer and the region center. Moreover, it is more reasonable to predict the QoS value for an active consumer based on his region, because consumers in the same region are more likely to have similar QoS experiences on web services, especially on those region-sensitive ones. This algorithm has the advantage of greatly reduced the search scope.

Some previous CF-based web service recommendation algorithms [4, 5] are based on the assumption that each consumer's rating range is subjective and comparatively fixed. However the range of QoS value varies largely from service to service. As a result, the average QoS value of all services provided by consumer *u* cannot reveal the performance of a specific web service s_i. So this kind of algorithm is not applicable in such context. Taking the predication of *A* value of service s_i for example, the steps are as follow.

i. Obtaining the active consumer's (u) IP address and expectation *Eu*. Finding the region *Ru* the consumer belongs to. If no appropriate region is found, the active consumer will be treated as a member of a new region.

ii. Determining the services which are used by consumers in the same region through space mapping [8], and then confirming the corresponding rating region of those services.

iii. Identifying whether service s_i is sensitive to the specific region. If it is region-sensitive, the prediction is generated from the region center, as equation 5. The process is finished.Otherwise, skip to step iv.

$$RA_u\left(s_i\right) = RA_{center}\left(s_i\right) \tag{5}$$

Where $RA_u\left(s_i\right)$ denotes the prediction of attribute *A* of services s_i by consumer *u*. $RA_{center}\left(s_i\right)$ is the center of rating region corresponding to *Ru*.

iv. Determining the range that *Eu* is belonging to. And then obtaining the region that all consumers in that range have rated the service s_i. After that using adjusted PCC to compute the similarity between the active consumer *u* and each region center c_j（$1 \leq j \leq k$）. If the active consumer's region center has the non-zero value, $RA_{center}\left(s_i\right) \neq 0$, then he prediction is computed using equation 6. Otherwise, using equation 7, this indicates that there is no consumer rating the service s_i in the active region

$$RA_u\left(s_i\right) = RA_{center}\left(s_i\right) + \frac{\sum_{j=1}^{k}\left(RA_{cj}\left(s_i\right) - \overline{RA_{cj}}\right)Sim'\left(u, c_j\right)}{\sum_{j=1}^{k} Sim'\left(u, c_j\right)} \tag{6}$$

$$RA_u\left(s_i\right) = \frac{\sum_{j=1}^{k} RA_{cj}\left(s_i\right) Sim'\left(u, c_j\right)}{\sum_{j=1}^{k} Sim'\left(u, c_j\right)} \tag{7}$$

Where $RA_{cj}\left(s_i\right)$ denotes the rating value of service s_i by region center c_j. $\overline{RA_{cj}}$ denotes the average of the rating value for service s_i by similar region center. $Sim'\left(u, c_j\right)$ is the similarity between the active consumer *u* and each region center c_j.

4. Experiment

4.1 Data Collection

In this experiment, the data is from WS-DREAM[1]。The data set contains 150 files, where each file includes Web service invocations on 100 Web services by a service

[1] http://www.wsdream.net

consumer. So the record is more than 150, 0000. Figure 4.1 and 4.2 shows the region of consumers and services separately. We mainly study the impact of personalized context to prediction method. We remove the unusable data caused by unauthorized or connecting time out. So we collect 1,538,730 QoS data。 We take Round-trip Time (RTT) for example, comparing the method CBQP with PCRA. Figure 4.3 shows the distribution of the RTT value.

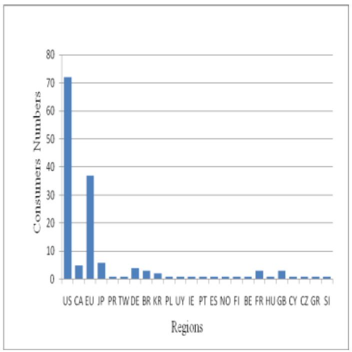


Figure 4.1. Service Consumer Location

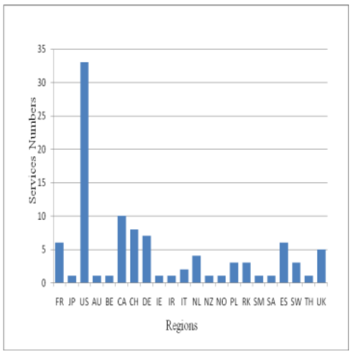


Figure 4.2. Service Location

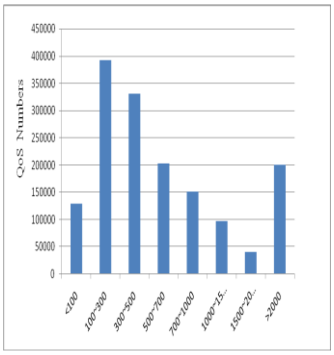


Figure 4.3. RTT Distribution (ms)

4.2 Evaluation Metric

Prediction accuracy is one of the important evaluation criteria of researchers. To evaluate the QoS prediction performance, we employ the metrics Mean Absolute Error (MAE) [4] for our experiments. MAE metric is widely employed to measure the prediction quality, which is defined as equation 8.

$$MAE = \frac{\sum_{u,i} \left| r_{ui} - p_{ui} \right|}{N} \tag{8}$$

where rui denotes the measured QoS value between service consumer u and Web service i, pui denotes the predicted QoS value, and N is the number of predicted services. The lower the value is, the higher the prediction accuracy is.

4.3 Impact of Context

In this section, in order to show the effectiveness of our proposed prediction approach (PCPA), we compare the prediction accuracy of two methods. One method is PCPA which considers the context information and another is CBQP which doesn't consider the context information. Taking RTT for example, context information involves only physical location. In PCPA, we set parameter $\lambda = 0.8$, $\mu = 0.2$, $k = 50$.

We select consumers' data from ten regions as a new dataset. In each region, there are 1~72 consumers, shown as Table 1. The dataset is divided into two parts: a training set and a test set. We employ 2-Fold Cross Validation to examine the approaches. In the case there is only one region, we select data from each region. The system predicts QoS value 20 times. The final result is the average of 20 error values. When there are two regions, we select data from two regions randomly. Similarly, we select region 10 times and compute 20 times. The result is also the average of error values. In turn, one region is increased every time. At the last time, there are ten regions, so computing directly.

Table 1. The Statistics of Services Consumers Regions

Region	FR	US	CA	EU	DE	BR	PL	KR	SI	JP
Number of consumers	3	72	5	37	4	3	1	2	1	6

Figure 4.4 shows the impact of context on the prediction method. X axis represents the number of service consumers region and Y axis represents the value of MAE. It can be seen from the figure that MAE value of two methods is very close in less region. It indicates that in that case the context is less influence to the prediction result. With the increasing number of regions, the difference between two methods is obvious. The PCPA is better than CBQP, because PCPA is using the context information about consumers, doing so the similarity between consumers is more accuracy.

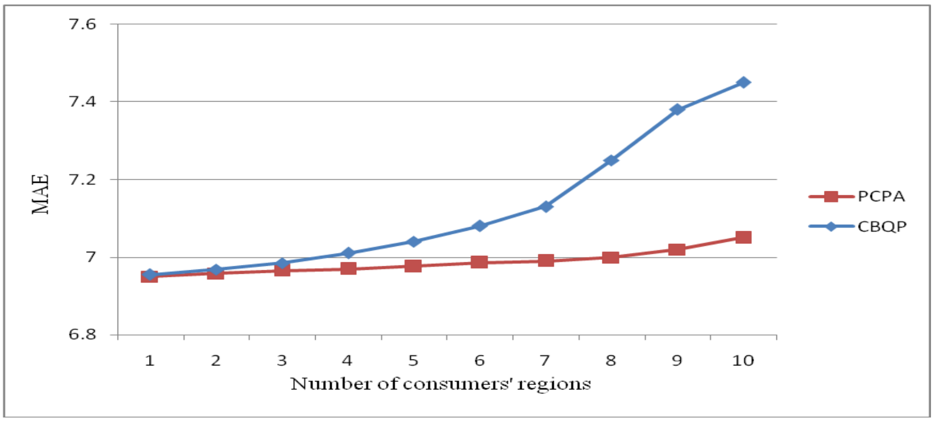


Figure 4.4. The Impact of Context

5. Conclusion

Context is one of the important factors affecting the performance of Web service recommendation. In this paper, we propose a personalized context-aware recommendation approach. The algorithm builds the region model using the QoS and the context information which is mainly consumers' physical location. The process includes consumer clustering, region-sensitive identifying and data region aggregation. In addition, we design a personalized context-aware recommendation framework for consumer. Experiment demonstrates the approach is effective. For future work, we will extend the dimension of context information for improving the prediction accuracy.

Acknowledgements

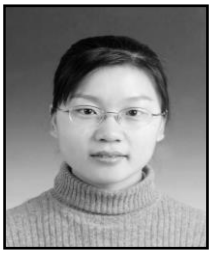
This paper is supported by the National Key Technology Research and Development Program (NO. 2013BAB05B01, 2013BAB06B04), the Central University Basic Scientific Research Project (NO. 2013B16114) and the Open Foundation of Huaian Research Institute of Hohai University (NO. 2014502512).

References

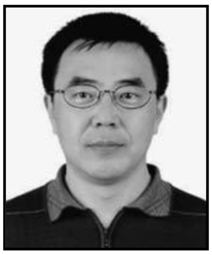
[1] L.J. Zhang, J. Zhang and H. Cai, "Services computing", Springer and Tsinghua University Press, Beijing **(2007).**
[2] M. N. Huhns and M. P. Singh, "Service-oriented computing: Key concepts and principles", IEEE Internet Computing, vol. 9, no. 75, **(2005).**
[3] E. A. Parasuraman, V. A. Zethaml, and L. L. Berry, "More on Improving Service Quality Measurement", Journal of retailing, vol. 69, no. 140, **(1993)**
[4] Z. Zheng, H. Ma, M. R. Lyu, and I. King, "WSRec: a collaborative filtering based web service recommendation system", Proc. 7th International Conference on Web Services, **(2009)** July 7-10, Los Angeles, CA, USA.
[5] L. Shao, J. Zhang, Y. Wei, J. Zhao, B. Xie, and H. Mei, "Personalized QoS prediction for web services via collaborative filtering", Proc. 5th International Conference on Web Services, **(2007)** July 9-13, Salt Lake City, Utah, USA.
[6] W. Rong, K. Liu, and L. Liang, "Personalized web service ranking via user group combining association rule", Proc. 7th International Conference on Web Services, **(2009)** July 7-10, Los Angeles, CA, USA.

[7] R. M. Sreenath and M. P. Singh, "Agent-based service selection", Journal of Web Semantics, vol. 261, **(2003).**
[8] X. J. Zhang, Z. J. Wang, X. Lv and R. Z. Qi, "A Clustering-based QoS Prediction Approach for Web Service Selection", Proceeding of the 3rd International Conference on Information Science and Cloud Computing, **(2013)** December 7-8, Guangzhou, China.
[9] A. K. Dey, "Understanding and using context", Personal and Ubiquitous Computing, vol. 5, no. 4, **(2001).**
[10] G. C. Roman, C. Julien and A. L. Murphy, "A Declarative Approach to Agent-Centered Context-Aware Computing in Ad Hoc Wireless Environments", Proceedings of the 2nd Inter. Work. Software Engineering for Large-Scale Multi-Agent Systems, **(2002)**, Orlando, Florida, USA.
[11] M. Gao and Z. Wu, "Personalized Context-aware Collaborative Filtering based on Neural Network and Slope One", Lecture Notes in Computer Science, vol. 5738, no. 109, **(2009).**
[12] M. R. McLaughlin and J. L. Herlocker, "A collaborative filtering algorithm and evaluation metric that accurately model the user experience", Proceedings of the 27th Annual International ACM SIGIR Conference, **(2004)** July 25-29, Sheffield, UK.
[13] C. D. Mining, P. Raghavan, and H. Schütze, "An Introduction to Information Retrieval", Cambridge University Press, Cambridge **(2009).**
[14] B. N. Miller, I. Albert, S. K. Lam, J. A. Konstan, and J. Riedl, "MovieLens unplugged: experiences with an occasionally connected recommender system", Proc. Int'l Conf. Intelligent User Interfaces, **(2003)** January 12-15, Miami, Florida, USA.

Authors



Zhang Xue-Jie, she is a Ph.D. candidate at Nanjing University of Aeronautics and Astronautics. She is a engineer at Hohai University. Her research interests include services recommendation and evaluation of QoS, etc.



Wang Zhi-Jian, he is a professor and doctoral supervisor at Hohai University. His research interests include Cloud Computing, Grid Computing and Big Data Processing, etc.

Zhang Wei-Jian, he is an engineer at Hohai University. His main research interest is computer network.

