

## Utility and Collaborative Filtering-Based Evaluation Method

Xiaolong Zhu<sup>1,2</sup>, Weidong Zhu<sup>1</sup> and Shuai Ding<sup>1</sup>

<sup>1</sup>*School of Management, Hefei University of Technology, Box 270, Hefei 230009, Anhui, PR China*

<sup>2</sup>*Anhui Business Vocational College, Hefei 231131, Anhui, PR China*

<sup>1</sup>*zhuxiaolong2008@163.com, <sup>2</sup>gdzwd@mail.hf.ah.cn, <sup>3</sup>174304135@qq.com*

### Abstract

*The theory and technology of cloud computing have been widely adopted by many small and medium enterprises and individuals for their business systems or personal affairs. To facilitate the evaluation processes of cloud service creditability, the present paper proposed a new multi-attribute evaluation theory. In addition, a utility and collaborative filtering-based evaluation method was presented, which targeted actual problems existing in the processes, such as missing data or inconsistency of data dimensions. The new method utilizes Enhanced Lance and Williams Distance, which is based on Jaccard similarity coefficient, to measure the similarities between different cloud services; it also applies utility theory to data unification and integration. In the later part of this paper, simulation experiments were conducted to test the validity and rationality of the proposed method.*

**Key words:** *cloud service; creditability; similarity; collaborative filtering; utility; data integration*

### 1. Introduction

Cloud service is an online application based on cloud computing platforms where users rent cloud services from software vendors through on-demand billing; information stored in cloud services can be expanded, updated, and released through simple webpage operations [1, 2]. With variables such as the software vendor's skill level, reliability of computing ability, and network platform, the cloud services provided by different vendors typically differ in creditability. Information security incidents have recently occurred due to the lack of creditability of cloud services. In August 2011, the cloud-computing network of Amazon and Microsoft went through a downtime, which resulted in the shutdown of hundreds of websites that totally relied on the Amazon EC2 and Microsoft BPOS. Salesforce's servers went down twice in a large scale in January 2009 and 2010, exerting direct influence on the 177 transactions of hundreds of enterprises that rent CRM cloud services.

The creditability of a cloud service is an extension of the software creditability [3, 4] and quality-of-service (QoS) concepts in cloud computing. Creditability is a multi-attribute measurement of the overall QoS. Creditability not only weighs the general quality of a cloud service from various aspects including reliability, safety, performance, maintainability and privacy, but also shows the users or decision-makers' subjective perceptions [5] of quality from an objective perspective. In actual applications, cloud services are mainly used in open, dynamic, and shared external cloud-computing platforms. Numerous factors and complex incentives can affect cloud services, so their quality has higher standards to meet compared with that of traditional localized and networked applications [6]. The popularization of cloud-computing technology involves

two key factors: determining the multi-dimensional attributes affecting the quality of cloud services and rationally evaluating their creditability to improve users' perceptions. These two aspects have long been the concern of local and international scholars. Abbadi [7] illustrated several attributes that can be used to evaluate the creditability of a cloud service; he also discussed the complex internal relations between different attributes. Wu [8] proposed several evaluation indicators, such as safety of data backup, stability of services and compatibility of the application; he also provided an evaluation method that targeted SaaS cloud service for decision making. Skopik *et al.*, addressed problems concerning the creditability of hybrid cloud service applications. They utilized personal trust preferences, multiple parties' trust, and partial trust to conduct a comprehensive assessment of creditability.

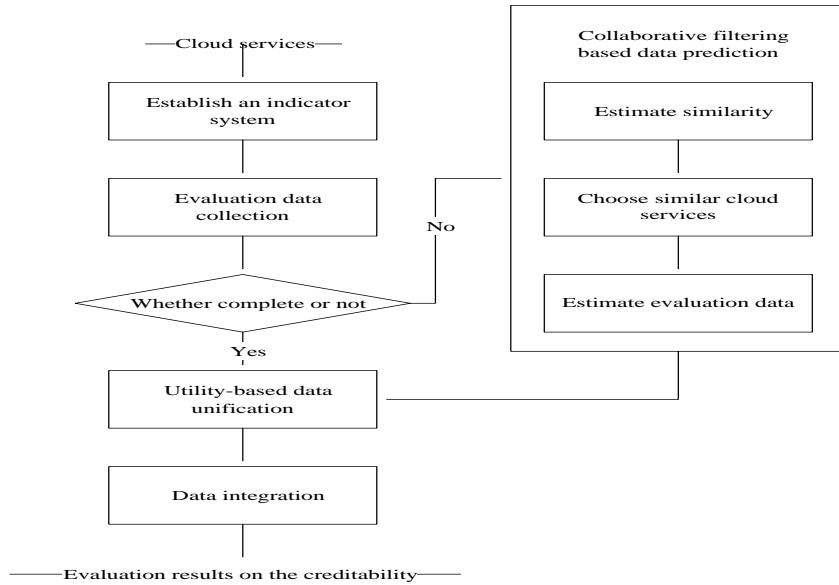
Given problems such as the dynamic nature and high evaluation cost of cloud services, one can hardly acquire complete evaluation data during the processes, leading to lack of information and constraint on the solving performance of an evaluation model. To address this dilemma, based on the multi-attribute evaluation method we have studied, the present research probes further into a utility and collaborative filtering-based evaluation method for the creditability of cloud services. By integrating evaluation data from similar cloud services, the new method applies approximate estimation to determine the predicted value of the attributes of the missing data. In addition, the general assessed value of a cloud service can be ascertained by transforming and integrating useful data. Different from existing collaborative filtering techniques, this method adopts Enhanced Lance and Williams Distance (ELWD) to estimate the similarity of several cloud services. Through this process, the prediction accuracy of missing data is improved. Meanwhile, the effect of data inconformity on the prediction is significantly reduced.

## 2. Processes of Cloud Service Evaluation

In the actual application of a cloud service, several factors will result in either part of or all data missing, such as limited evaluation techniques, high evaluation cost and changing application processes. Existing studies on Web Service recommendation have mostly focused on cases of missing data. To solve the problem, collaborative filtering, whose recommendation performance is widely demonstrated, is an effective method [11]. Based on this perspective, we proposed utility and collaborative filtering-based evaluation processes for the evaluation of the creditability of cloud services, as shown in Figure 1.

To evaluate the creditability of a cloud service, a comprehensive evaluation from various perspectives on key factors that affect the creditability is needed. Moreover, designing or choosing indicators that target specific cloud services is necessary. These indicators are used to establish a system for decision making. The processes involve collecting and dealing with multi-source and isomeric evaluation data. The indicator system is the basis in ascertaining whether the data is complete or incomplete. By using missing data estimation recommended by collaborative filtering and utility-based data unification, complete and unified data sets will be guaranteed. Therefore, the integration of multi-source data helps facilitate a comprehensive evaluation of the creditability of a cloud service.

Figure 1 illustrates a collaborative filtering-based data prediction. Utility-based data unification is necessary to guarantee the accuracy of evaluation results. Based on these two methods, we will present the actual evaluation processes.



**Figure 1. Processes used to Evaluate the Creditability of Cloud Services**

### 3. Collaborative Filtering-based Data Prediction

#### 3.1. Calculating Similarity of Various Cloud Services

We utilize collaborative filtering recommendation to solve missing evaluation data problems. The core idea is to make use of the concept that “the same user will make a similar evaluation on similar cloud services”. Based on this premise, we can estimate the similarity between two cloud services by analyzing users’ preferred evaluation data on a similar cloud service. Through this step, we can determine the most similar cloud services to improve the accuracy of predicting the missing data.

One of the basic features for evaluating cloud services is the inconsistency of data dimensions. For instance, in the evaluation presented in a previous study [12], two indicators are typically used: “response-time” and “throughput”. However, their units are respectively “s” and “kpbs”, making the data rather disparate. The conventional collaborative filtering technology mainly adopts Pearson Correlation Coefficient (PCC) to calculate similarity. However, PCC fails to make good recommendations when dealing with multi-dimension data. For this reason, the present paper adopts ELWD, which can calculate similarity and effectively eliminate the negative influence of disparate and multi-dimension data. Consequently, a more effective evaluation is carried out.

Suppose we have a collection  $U$  of users and a collection  $I$  of cloud services, then a  $U \times I$  user–cloud service matrix is established. In this matrix,  $e_{u,i}$  represents how user  $u$  evaluates cloud service  $i$  on a specific creditability attribute. If user  $u$  has not yet used cloud service  $i$ , then  $e_{u,i} = \text{null}$ .

Definition 1. Let  $i$  and  $j \in I$  be two independent cloud services in the evaluation system. The similarity between  $i$  and  $j$  can be calculated through ELWD as follows:

$$Sim(i, j) = \sum_{u \in U_{ij}} \frac{|e_{u,i} - e_{u,j}|}{e_{u,i} + e_{u,j}}, \quad (1)$$

In this formula,  $U_{ij} \in U$  is the intersection of the non-null evaluation data on  $i$  and  $j$  from  $U$ .  $|U_{ij}|$  represents the number of users in  $U_{ij}$ , whereas  $e_{u,i}$  and  $e_{u,j}$  are user  $u$ ’s actual evaluation of cloud services  $i$  and  $j$ .

During the collaborative filtering processes based on the  $U \times I$  matrix, cloud services can be divided into two types according to a historical perspective: positive and negative

cloud services. Negative services refer to neighboring cloud services that can be effectively used, but have minimal records. A previous paper [12] pointed out that negative cloud services impair the overall predicting outcomes for the target cloud services. Therefore, we proposed ELWD on the basis of Jaccard similarity coefficient, that is, to make use of the diverse historical records of different cloud services to recalculate their similarity. Through this method, we can reduce the impact of negative cloud services on missing data predicting.

Definition 2. Let  $i$  and  $j \in I$  be two independent cloud services in the evaluation system. Taking diverse historical records into account, their similarity can be calculated by ELWD as follows:

$$Sim^J(i, j) = J_{i,j} \times Sim(i, j), \quad (2)$$

In the formula above,  $Sim(i, j)$  is the Lance distance between  $i$  and  $j$ .  $J_{i,j}$  represents the Jaccard similarity coefficient between  $i$  and  $j$ , and the coefficients meet the following formula:

$$J_{i,j} = \frac{|U_{i,j}|}{|U_i| + |U_j| - |U_{i,j}|} \quad (3)$$

In this formula,  $|U_i|$  and  $|U_j|$  represent the number of users of cloud services  $i$  and  $j$ , respectively.  $|U_{ij}|$  is the number of users in the intersection  $U_{ij}$ .

In the  $U \times I$  matrix, if the Lance distance of the neighboring cloud services  $j$  and  $k$  equals that of target cloud service  $i$ , then  $Sim(i, j) = Sim(i, k)$ . If and only if the Jaccard similarity coefficient of  $j$  is bigger than that of  $k$ , namely,  $J_{i,j} \geq J_{i,k}$ , is  $Sim^J(i, j) \geq Sim^J(i, k)$  workable.

### 3.2. Choosing Similar Cloud Services

Data prediction is made possible by using historical data about neighboring cloud services from the  $U \times I$  matrix. The prediction performance has a positive correlation with the accuracy of estimating similarity. If any cloud service has little or no ELWD similarity with others in the training data ( $Sim < 0$ ), the accuracy of the prediction is significantly impaired. Therefore, using a similarity threshold is necessary to rationally decide which cloud services to choose. Similarity threshold refers to the users' or decision makers' subjective judgment on the similarity between cloud services; it exerts a direct influence on the scale of the training data, predicted cost, and accuracy of the prediction, so users and decision-makers need to set the threshold in advance according to the actual requirements.

Suppose  $i \in I$  is the target cloud service whose creditability will be evaluated. After forming the  $U \times I$  matrix, we need to determine the similarity between the target cloud service and each neighboring service separately, and then choose the one with the highest similarity.

Definition 3. Let  $Sim^J(i, j)$  represent the ELWD similarity between the target cloud service and its neighboring ones. Then, we can work out the combination  $I_{cs}$  of similar cloud services by following the formula

$$I_{cs} = \{j \mid j \in I, j \neq i, Sim^J(i, j) > \xi_j\}, \quad (4)$$

In this formula,  $\xi_j$  represents the similarity threshold.

### 3.3. Missing Data Prediction

An incomplete evaluation data will significantly limit research on creditability evaluations. The creditability evaluations of cloud services are different from conventional credibility evaluations for software. Given the openness, inconstancy, and

large scope of cloud services, the historical evaluation of data in cloud computing has increased exponentially, thus enabling easy sharing. These factors have made the credibility evaluation of cloud services possible.

By using historical data from the collection  $U$  on the combination  $I_{cs}$  of similar cloud services, we can estimate the predicted value of the target cloud service even when data are missing. Suppose  $A_u = \{a_1, \dots, a_T\}$  is one of the attributes for the credibility evaluation of the target cloud service  $I$  with missing evaluation data at  $\in A_u$  (i.e.,  $e_{u,I}$  is unavailable and further estimation is needed for the evaluation). We can then use the definition below to conduct the estimation.

Definition 4. Let  $Sim^J(i,j)$  represent the ELWD similarity between cloud service  $i$  and a similar cloud service  $j$ ;  $e_{u,j}$  is the historical data. It is the evaluation given by user  $u \in U$  on the similar cloud service  $j \in I_{cs}$  concerning the attribute  $a_t$ . Thus, the predicted value of  $a_t$  is expressed as follows:

$$e_{u,i} = \bar{e}_i + \frac{\sum_{j \in I_{cs}} Sim^J(i,j)(e_{u,j} - \bar{e}_j)}{\sum_{j \in I_{cs}} Sim^J(i,j)}, \quad (5)$$

In this formula,  $\bar{e}_i$  and  $\bar{e}_j$  are the average value for service  $i$  and  $j$ , respectively;  $|I_{cs}|$  represents the scope of similar services.

The problem of missing data can be effectively solved by using the above definition. The application of ELWD also helps improve the prediction accuracy of missing data. Thus, the following discussion focuses on the utility-based unification and integration of evaluation data to acquire the comprehensive assessed value of the target cloud service.

## 4. Utility-based Data Integration

### 4.1. Data Unification

The aforementioned analysis indicates that addressing the inconsistency of data dimensions is necessary when collecting and predicting evaluation data from a multi-attribute indicator system. Utility is considered a quantized value that uses the subjective experience of decision-making experts as criteria and predicts the satisfaction of experts with the service [3]. Utility theory has been successfully applied in previous software evaluations to address multi-attribute evaluation problems. The present paper proposed a unification and integration method based on the utility of previous software evaluations.

Suppose  $A_u = \{a_1, \dots, a_T\}$  is one of the attributes for the credibility evaluation of the target cloud service  $i$ . The actual value or predicted value of  $A_u$  is  $e_t$ . The utility-based data unification is shown below.

Definition 5. Let the decision level of the evaluation be  $H_1, \dots, H_5$ . The piecewise utility function should be  $u(H_n)$  ( $n = 1, \dots, 5$ ). The function of the actual value given by experts on the attribute is  $f(H_n)$  ( $n = 1, \dots, 5$ ). The utility value of the evaluation  $e_t$  is as follows: 1) if  $e_t \leq f(H_1)$ , then  $u_t = u(H_1)$ ; 2) if  $e_t \geq f(H_5)$ , then  $u_t = u(H_5)$ ; 3) if  $f(H_n) \leq e_t \leq f(H_{n+1})$  ( $n = 1, \dots, 4$ ), then

$$u_t = \frac{(e_t - f(H_n))u(H_n)}{f(H_{n+1}) - f(H_n)} + \frac{(f(H_{n+1}) - e_t)u(H_{n+1})}{f(H_{n+1}) - f(H_n)}, \quad (6)$$

In this formula, the utility value is  $u_t \in [0, 1]$ .  $u(H_n)$  and  $f(H_n)$  are both monotonous increasing functions.

After the calculation, evaluation data with different dimensions could be unified into a utility value within  $[0, 1]$ . This approach can effectively avoid the adverse effects of different dimensions on data integration.

## 4.2. Data Integration

For users who have insufficient professional knowledge, choosing a cloud service that not only meets specific requirements but is also within a specific level of creditability is difficult. Choosing an improper service may lead to “information overload.” Therefore, an evaluation system should be used to reduce the cost of wrong choices. After obtaining the data, one needs to integrate the data by combining the relative weights of different attributes to provide users with accurate assessed values.

Definition 6. Suppose  $Au = \{a_1, \dots, a_T\}$  is one of the attributes for the credibility evaluation of the target cloud service  $i$ .  $u_1, \dots, u_T$  is the utility value for each attribute, and  $w(a_t)$  ( $t = 1, \dots, T$ ) is the relative weight between attributes. Thus, the accurate assessed value of the target cloud service is expressed as follows:

$$cst_i = \sum_{t=1}^T w(a_t)u_t, \quad (7)$$

In this formula, relative weight follows the equation  $\sum tw(a_t) = 1$ .

The estimation method for the relative weights is relatively mature with available methods divided into three types: subjective weighting method, objective weighting method, and objective and subjective weighting method. No further discussion about the three methods will be presented in this paper. Cloud services reduce enterprises or individual users' initial costs on an application and require only simple daily maintenance. Limited resources can be mostly used on the R&D and popularization of user products, thus increasing profits. The aforementioned utility and collaborative filtering-based evaluation method maximize the use of cloud service features such as large amount of client information and real time information sharing to address the problems caused by missing data and dimension inconsistency.

## 5. Case Study

### 5.1. Results

To verify whether the evaluation method was effective, we designed a prototype system of the proposed method by using the Java platform and a set of simulation experiments. The relevant parameters are as follows: attribute  $T = 6$ , collection of cloud services  $|I| = 20$ , and user collection  $|U| = 50$ . Combined with the basic understanding of attributes in [7], we choose six attributes to evaluate the creditability of the target cloud service (Table 1).

**Table 1. Attributes**

Attributes (a1–a6)	Dimensions	Evaluations	Interval of the actual value
Throughput	kbps	190	[100, 250]
Response time	s	1.75	[1, 2]
Survivability	s	6.5	[0, 10]
Accuracy	%	99.9	[99, 99.99]
Failure rate	%	2	[0.5, 3.5]
Stable time	h	1.21	[0.5, 1.5]

During the simulation experiments, we used the open QoS database proposed in [12] because of the typicality and rationality of the data. Approximately 100 cloud services were randomly chosen from the QoS database to analyze the observed data on two indicators: throughput and response time. Two attributes, a1 and a2, was used to verify whether the method in the present study was effective. The data simulation method in [15] also contributed to our case study. This data simulation method randomly sets the values of attributes a3 to a6 within a certain range. The results of the simulation experiments are shown in Table 1.

To calculate the actual utility value of the data, we set the piecewise utility function within {0.05, 0.3, 0.5, 0.7, 0.95}, and the interval of the actual value function f(Hn) is show in Table 1. We then determined the utility value u1,..., u6 (Table 2) for each attribute by Definition 5. By combining the relative weight w(at), data integration was made possible to obtain the comprehensive assessed value csti = 0.7651.

**Table 2. Relative Weight and Utility Value**

Attributes	w(at)	ut
Throughput	0.13	0.723
Response time	0.15	0.6927
Survivability	0.21	0.8725
Accuracy	0.13	0.8272
Failure rate	0.14	0.6485
Stable time	0.24	0.7737

## 5.2. Discussions

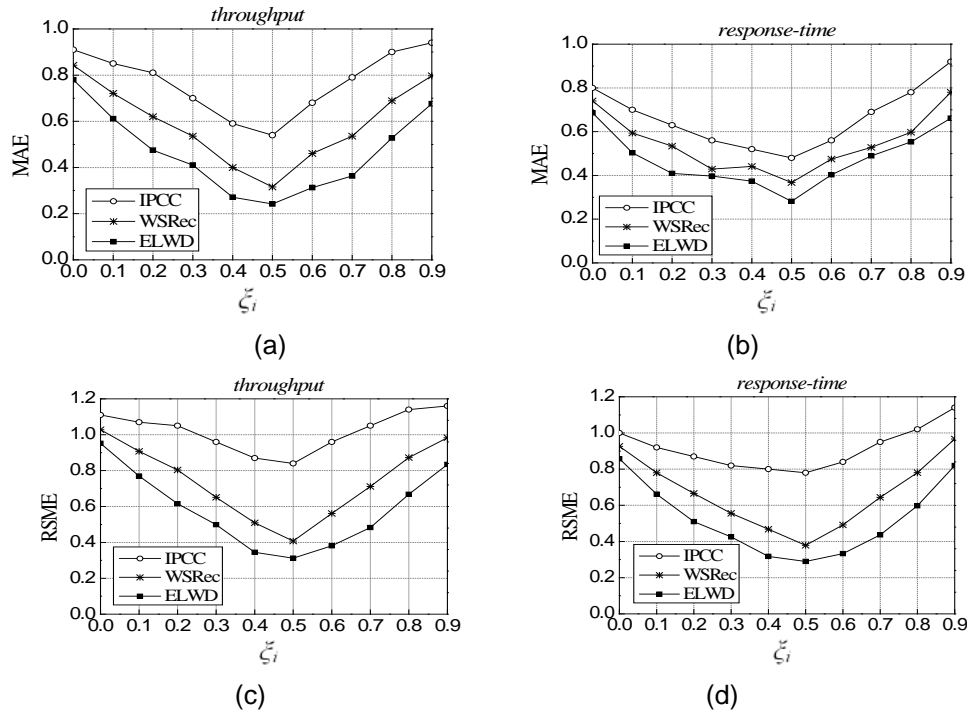
The prediction accuracy on missing data has a direct influence on the effectiveness of the evaluation results. The present method differs in this aspect from other methods. Therefore, we designed simulation experiments for a1 and a2 compared with the commonly used IPCC and WSRec methods [12]. The two factors for measuring accuracy, that is, mean absolute error (MAE) and rank scoring measure (RSM), were used to test the effectiveness of the present method. IPCC and WSRec are two frequently used methods for predicting missing data. These two methods are combined with the Pearson correlation coefficient to calculate the similarity between two cloud services. The WSRec uses the relative weights of the sample importance to adjust the Pearson correlation coefficient; thus, the overall accuracy is increased by reducing the negative influence of collaborative filtering recommendations. The method proposed by the present paper is based on ELWD (*i.e.*, the ELWD method).

The calculation formula for MAE and RSME are as follows:

$$\begin{cases} MAE = \frac{\sum_{u,i} |\hat{e}_{u,i} - e_{u,i}|}{M}, \\ RMSE = \sqrt{\frac{\sum_{u,i} (\hat{e}_{u,i} - e_{u,i})^2}{M}}, \end{cases} \quad (8)$$

In this formula,  $\hat{e}_{u,i}$  is the predicting value,  $e_{u,i}$  is the actual value, and M is the scope of prediction.

We used the data presented in Section 4.1 to construct a  $U \times I$  matrix ( $339 \times 100$ ). We chose 80% and 20% of the historical data as the training set and test set, respectively. By considering the effect of the similarity threshold  $\xi_i$ , we set the step length of  $\xi_i$  as 0.1. The comparison of the results of the three methods is shown in Figure 2. Figures 2(a) and 2(c) show the MAE and root-mean-square error (RMSE) for each method used on a1 (throughput), whereas Figures 2(b) and 2(d) show the MAE and RMSE used on a2 (response time).



**Figure 2. Results**

Compared with IPCC and WSRec methods, the present ELWD method significantly improves the prediction accuracy on missing data by reducing the similarity of negative cloud services and by using ELWD to avoid data and dimension inconsistencies (Figure 2). Therefore, the ELWD method is appropriate for complex evaluations on the creditability of cloud services. Figure 2 also shows that the optimal similarity threshold is  $\xi_i = 0.5$ . When  $\xi_i \in [0, 0.5]$ , an increase in  $\xi_i$  will reduce the effect of less similar neighboring services ( $\xi_i < 0.5$ ); When  $\xi_i \in [0.5, 0.9]$ , the accuracy of the ELWD method will be impaired because of the sharp reduction of historical data.

## 6. Conclusion

The creditability of a cloud service is directly related to the actual performance and success rate of the application. Thus, credibility has become the major concern in making decisions concerning cloud services. In the open and inconstant cloud-computing context, credibility evaluation is limited by missing data and dimension inconsistencies. To solve these problems, this paper used ELWD to estimate the similarity between two services. We also used collaborative filtering recommendation to predict missing data. Utility theory is also adopted for data unification and integration.

Our future research will still focus on the mechanism and technology of the evaluation model to probe into the key factors that influence evaluation accuracy. Emphasis will be placed on the sparse matrix and cold start in predicting missing data to improve the feasibility of our evaluation method.



## Acknowledgements

This paper is sponsored by Key Program of Anhui Provincial Humanities and Social Science Fund (No.SK2013A027).

## References

- [1] B. Hayes, "Cloud computing", *Communications of the ACM*, vol. 51, no. 9, (2008).
- [2] Z. Yunqiao and D. Lognzheng, "Based on the research of flexible thought assembly services in the cloud", *Computer Engineering and Applications*, vol. 49, no. 225, (2013).
- [3] S. Ding, X. J. Ma and S. L. Yang, "A software trustworthiness evaluation model using objective weight based evidential reasoning approach", *Knowledge and Information Systems*, vol. 33, no. 171, (2012).
- [4] T. Yongxin and L. Zengliang, "Software reliability measurement model is reviewed", *Computer Engineering and Applications*, vol. 46, no. 12, (2010).
- [5] M. Mohammd and V. Alagar, "A formal approach for the specification and verification of trustworthy component-based systems", *Journal of Systems and Software*, vol. 84, no. 77, (2011).
- [6] M. Alhamad, T. Dillon and E. Chang, "A trust-evaluation metric for cloud applications", *International Journal of Machine Learning and Computing*, vol. 1, no. 416, (2011).
- [7] I. M. Abbadi and A. Martin, "Trust in the Cloud", *Information Security Technical Report*.doi:10.1016/j.istr.2011.08.006, (2011).
- [8] W. J. Li, L. D. Ping, "Trust model to enhance security and interoperability of cloud environment", *Lecture Notes in Computer Science*, vol. 593, no. 69, (2009).
- [9] F. Skopik, D. Schall and S. Dustdar, "Modeling and mining of dynamic trust in complex service-oriented systems", *Information Systems*, vol. 35, no. 7335, (2010).
- [10] D. Shuai, L. Fujun, Y. Shanlin and X. Chengyi, "A Requirement-Driven Software Trustworthiness Evaluation and Evolution Model", *Journal of Computer Research and Development*, vol. 4, no. 647, (2011).
- [11] W. Uoxia and L. Huoping, "Personalized recommendation systems were reviewed", *Computer Engineering and Applications*, vol. 48, no. 66, (2012).
- [12] Z. B. Zheng, H. Ma and M. R. Lyu, "QoS-aware web service recommendation by collaborative filtering", *IEEE Transactions on Services Computing*, vol. 4, no. 140, (2011).
- [13] H. Fushing, H. Wang and K. Vander, "Multi-scale clustering by building a robust and self-correcting ultrametric topology on data points", *Plos One*, vol. 8, no. 1, (2013).
- [14] Y. T. Liu and Y. J. Liang, "A sentence semantic similarity calculating method based on segmented semantic comparison", *Journal of Theoretical and Applied Information Technology*, vol. 48, no. 231, (2013).
- [15] N. Limam and R. Boutab, "Assessing software service quality and trustworthiness at selection time", *IEEE Transactions on Software Engineering*, vol. 36, no. 559, (2010).

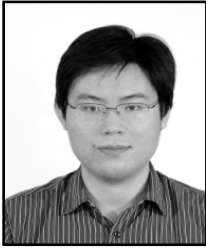
## Authors



**Xiaolong Zhu** received his B.S. degree from Hefei University of Technology, Hefei, China in 1999. Now he is a doctoral candidate in School of Management, Hefei University of Technology. Currently, Xiaolong Zhu is serving as associate professor in Anhui Business Vocational College, Hefei Anhui, China. His current interests include cloud computing, capital operation and management.



**Weidong Zhu** is a Doctor of Management, Hefei University of technology school of management professor, doctoral supervisor, Association for Information System of individual members. Mainly engaged in intelligent decision-making theory, accounting theory research and enterprise management information.



**Shuai Ding**, Hefei University of Technology in 2008; and the Ph.D. degree from School of Management, Hefei University of Technology in 2011. He is currently an Assistant Professor with the School of Management, Hefei University of Technology. He is currently leading a research project supported by National Natural Science Foundation of China. His publications have appeared in some well-known conference proceedings and international journals. He also served as a Reviewer for some international journals. His research interests include cloud service and trustworthy software.