The Prediction of Trust Rating Based on the Quality of Services Using Fuzzy Linear Regression

M. Hadi Mashinchi, Lei Li, Mehmet A. Orgun, and Yan Wang

Department of Computing
Macquarie University
Sydney, NSW 2109, Australia
Emails: {hadi.mashinchi, lei.li, mehmet.orgun, and yan.wang}@mq.edu.au

Abstract—With the advent of service-oriented computing, the issue of trust and Quality of Service (QoS) have become increasingly important. In service-oriented environments, when there are a few service providers providing the same service, a service client would be keen to know the trustworthiness of each service provider in the forthcoming transaction. The trust value of a delivered service from a service provider can be predicted according to a set of advertised QoS data collected by the trust management authority. Although trust and QoS are qualitative by nature, most data sets represent trust and QoS in the ordinal form for the sake of simplicity.

This paper introduces a new approach based on Fuzzy Linear Regression Analysis (FLRA) to extract qualitative information from quantitative data and so use the obtained qualitative information for better modeling of the data. For verification purposes, the proposed approach can be applied for the trust prediction in the forthcoming transaction based on a set of advertised QoS in service-oriented environments.

I. INTRODUCTION

In Service-Oriented Computing (SOC), a service is an autonomous and platform-independent computational entity, which can be described, published, discovered and dynamically assembled for developing massively distributed systems [1]. In fact, any piece of code or application component deployed on a system can be taken as a service [1], [2].

In SOC environments, QoS (Quality of Service) refers to various nonfunctional characteristics. These specified characteristics should be measurable and constitute a description of what a service can offer. The QoS of an IT service is often expressed in terms of capacity, latency, bandwidth, the number of service requests, and the number of incidents [3]. QoS is essential when a set of quality metrics have to be achieved during service provision [3]. In contrast, in SOC environments, the general quality of a delivered service can be represented by a trust rating given by a service client. As trust and QoS both focus on the quality of service, the trust value of a service or service provider can be taken as a function of a set of advertised QoS values reflecting the quality of the service or the service provider, and the trust value can be predicted by the trust management authority.

When a service client looks for a service from a large set of services offered by different service providers, in addition to functionality and QoS, reputation-based trust is also a key factor for service selection. It is also critical for the trust management authority to be responsible for maintaining the list of reputable and trustworthy services and service providers with their advertised QoS values, and providing them to clients as required [2].

Conceptually, trust is the measure taken by one party on the willingness and ability of another party to act in the interest of the former party in a certain situation [4]. If the trust value is in the range of [0,1], it can be taken as the subjective probability with which, one party expects that another party performs a given action [5].

The trust issue has been widely studied in many applications. In e-commerce environments, a trust management system can provide valuable information to buyers and prevent certain typical attacks [6], [7]. In Peer-to-Peer information-sharing networks, binary ratings work pretty well as a file is either the definitively correct version or not [8]. In SOC environments, an effective trust management system is critical to identify potential risks, provide objective trust results to clients and prevent malicious service providers from easily deceiving clients [9]. Different from P2P information-sharing networks or the eBay reputation management system, where a binary rating system is used [8], in SOC environments, trust data is usually captured in its ordinal form, such as: *5-star*, *gold*, *high-class* and etc., due to the abstraction or even the simplicity of representation. This abstraction makes the trust value more understandable for the end users and service clients, but it is not really suitable for any further inference. An intuitively appealing approach towards trust, which is qualitative by nature, is to not quantize the trust but to take the uncertainty which comes from vagueness into consideration in the level of inference [10]. In contrast, the quantization only takes place for representing the trust to the end users or service clients when it is deemed to become necessary [10].

As mentioned earlier, the trust rating, representing the subjective belief of a service client on the general quality of the delivered service, is both subjective and qualitative. However, by quantization both subjective and qualitative information in the trust data set would be overlooked. For example, two services with the same trust level of *5-star* may not be necessarily the same for a specific purpose or the requirement

of a service client. In fact, one service could be much more suitable than the other one, but due to the quantization such kind of important information may be ignored in the service selection process.

This paper proposes a QoS based approach to trust prediction using Fuzzy Linear Regression Analysis (FLRA). This approach is built up with the ordinal trust data which can be extracted, so that the hidden fuzzy information can be taken into consideration where it has been overlooked in the original trust data set. The motivations of applying FLRA for predicting trust are summarized as follows.

- Capturing the hidden fuzziness: we want to capture the fuzziness overlooked in the ordinal trust data set due to the reasons such as abstraction. The other machine learning approaches such as neural networks, decision trees and support vector machines in their original form, do not take the fuzziness into consideration and only model the original trust data set.
- Having a transparent model: we want to seek for a transparent relationship between the delivered trust value and advertised QoS where the underlying function is extractable. In industry, neural networks or any other approaches, which have the black-box nature, are not suitable for marketing purposes.
- More detailed information: we want to provide more information than only a few trust levels. For example, a web service is not necessarily either *platinum* or *gold*; it could be for instance *better than gold,* but *worse than platinum* - which can be represented by a fuzzy value.

FLRA is a powerful tool to model the relationships in situations where the attributes are inexact. FLRA is widely applied in economy, finance, marketing, social sciences, healthcare and other domains [11], [12], [13]. In addition, trust is also a widely studied and complex issue, including many uncertain factors [6]. Hence, with the fuzziness, fuzzy regression based on fuzzy set theory can deal with QoS based trust prediction in a reasonable manner. In this paper, we propose a method for the prediction of trust rating based on fuzzy linear regression for building up a mapping from a set of advertised QoS values to the trust value of the corresponding delivered service of a service provider. Then, the service client can predict the trustworthiness of the service provider in the forthcoming transaction with its advertised QoS values.

This paper is organized as follows. In Section II, we review trust management, trust evaluation and uncertainty. Section IV presents our QoS based trust prediction method with fuzzy regression. Some experiments are presented in Section V for illustrating how our proposed method can predict trust with QoS values. Finally Section VI concludes our work.

II. RELATED WORK

A. Trust Evaluation in E-Commerce Environments

Trust is an important issue in e-commerce (EC) environments. At eBay [14], after each transaction, a buyer can give feedback with a rating of "positive", "neutral" or "negative" to the system according to the service quality of the seller. eBay calculates the feedback score $S = P - N$, where P is the number of positive ratings left by buyers and N is the number of negative ratings. Then the positive feedback rate $R = \frac{P}{P+N}$ (e.g., $R = 99.1\%$) is calculated and displayed on web pages. This is a simple trust management system providing valuable reputation information to buyers.

In [7], Sporas system is introduced to evaluate the trust for EC applications based on the ratings of transactions in a recent time period. In this method, the ratings of later transactions are given higher weights as they are more important in trust evaluation. In [6], Wang et al. propose an approach to evaluate the situational transaction trust, which binds the trust ratings of previous transactions with a new transaction. Since the situational trust includes service specific trust, service category trust, transaction amount category specific trust and price trust, it can deliver more accurate trust information to buyers and prevent some typical attacks.

B. Trust Evaluation in P2P Information Sharing Networks

The trust issue has been actively studied in Peer-to-Peer (P2P) information sharing networks as a client peer needs to know prior to download actions which serving peer can provide complete files. In [15], Damiani et al. propose an approach for evaluating the reputation of peers through distributed polling algorithm and the *XRep* protocol before initiating any download action. This approach adopts a binary rating system and it is based on the Gnutella query broadcasting method. EigenTrust [16] adopts a binary rating system as well and aims to collect the local trust values of all peers to calculate the global trust value of a given peer.

Some other earlier studies also adopted the binary rating system. In [8], Xiong et al. propose a *PeerTrust* model which has two main features. First, they introduce three basic trust parameters (i.e., the feedback that a peer receives from other peers, the total number of transactions that a peer performs, the credibility of the feedback sources) and two adaptive factors in computing trustworthiness of peers (i.e., transaction context factor and the community context factor). Second, they define some general trust metrics and formulas to aggregate these parameters into a final trust value.

C. Trust Evaluation in Service-Oriented Environments

In the literature, the issue of trust has also received much attention in the field of service-oriented computing (SOC). In [9], Vu et al. presents a model to evaluate the service trust by comparing the advertised service quality and the delivered service quality. If the advertised service quality is as good as the delivered service quality, the service is reputable.

The application of machine learning and soft-computing approaches for predicting the trust of service-oriented environments are suggested to be very successful in the literature. In [17], Wang et al. propose several trust evaluation metrics and a formula for trust computation, with which a final trust value is computed. In addition, they propose a fuzzy logic based approach for determining reputation ranks that particularly differentiate the service periods of new service providers and

old (long-existing) ones. The aim is to provide incentives to new service providers and penalize those old service providers with poor service quality.

Other machine learning methods such as neural networks, decision trees, support vector machines and other classifiers are suggested and applied for predicting the ordinal trust based on the QoS data [18], [19].

D. Uncertainty

It is reported that the uncertainty in a system can be due to several reasons [20]:

- The high complexity of the environment, which necessitates the adaptation of abstraction (granulation of information) for generalization purposes [21].
- The influence of human subjective judgement in the decision process or the involvement of human-machine interactions [22], [23], [24].
- Partially available information, [25] due to miss-recording or inaccurate measurements [26].

The trust terms in service-oriented environments come with uncertainty as they are derived from the *human subjective judgement* - one of the above reasons.

However, to the best of authors' knowledge, the application of FLRA for predicting the trust in service-oriented environments has been only studied in [13]. This approach applies FLRA to fit the ordinal data and does not extract the fuzziness overlooked in the data set.

There are other approaches in different engineering applications where FLRA has been applied. However, a manual fuzzification of the data (by adding spreads) with some predefined variables estimated by experts is applied.

Therefore, there is a necessity to propose a novel fuzzy regression approach for QoS based trust prediction in service-oriented environments.

III. FUZZY REGRESSION ANALYSIS

Regression analysis [27] is a statistical technique for modeling and investigating the relationship between two or more variables. For example, here regression analysis can be used to build up a model that represent the trust of the delivered service as a function of a set of advertised QoS values. Then this model can be used to predict the trust value of the new delivered service with a new set of advertised QoS values.

A. Classical Regression Analysis

In the classical regression method, a set of parameters of an unknown function $f(x, \omega)$ can be estimated by making measurements of the function with uncertain at any point x_i:

$$y_i = f(x_i, \omega) + \epsilon_i, \tag{1}$$

where the uncertain ϵ_i is independent of x and is distributed according to a known density $p_\omega(\epsilon)$. Based on the observed data sample $S = \{(x_i, y_i) | i = 1, 2, \ldots, n\}$, the likelihood is given by

$$P(S|\omega) = \sum_{i=1}^{n} \ln p_\omega(y_i - f(x_i, \omega)). \tag{2}$$

Assuming that the uncertain is normally distributed with mean 0 and variance δ, the likelihood is given by

$$P(S|\omega) = -\frac{1}{2\delta^2} \sum_{i=1}^{n} (y_i - f(x_i, \omega))^2 - n \ln(\sqrt{2\pi}\delta) \tag{3}$$

Maximizing the likelihood in Eq. (3) is equivalent to minimizing

$$E(\omega) = \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i, \omega))^2, \tag{4}$$

which is in fact the same as the estimation by the method of least squares. Namely, the regression line is estimated so that the sum of squares of the deviations between the observations and the regression line is minimized.

Classical regression analysis [27] is a statistical technique for modeling and investigating the relationship for quantitative or homogeneous data. However, for qualitative heterogeneous data, an application of fuzzy regression analysis is more promising. In such systems, uncertainty arises usually not due to randomness but due to the phenomenon of fuzziness [20].

B. Fuzzy Regression Model Parameters

Prior to presenting the detailed fuzzy regression model, some definitions about parameters of the model should be introduced firstly in this section.

1) Membership Function: The fuzzy number mentioned in this paper is $\widetilde{A}(\alpha, C)$ with the following *membership function* [28],

$$\mu_{\widetilde{A}}(x) = \begin{cases} \begin{cases} 1 - \frac{|x - \alpha|}{C}, & |x - \alpha| \leq C, \\ 0, & \text{otherwise}, \end{cases} & C > 0, \\ \begin{cases} 1, & x = \alpha, \\ 0, & \text{otherwise}, \end{cases} & C = 0, \end{cases} \tag{5}$$

According to Eq. (5), it is easy to prove that $\lambda \widetilde{A}$ is fuzzy number $\widetilde{A}(\lambda\alpha, |\lambda|C)$ and $\widetilde{A}_1 + \widetilde{A}_2$ is fuzzy number $\widetilde{A}(\alpha_1 + \alpha_2, C_1 + C_2)$. So if

$$\widetilde{T}_i^* = \widetilde{A}_0 + q_{i1}\widetilde{A}_1 + \ldots + q_{in}\widetilde{A}_n, \tag{6}$$

then $\widetilde{T}_i^*$ is the fuzzy number

$$\widetilde{T}_i^*(\alpha_0 + \sum_{j=1}^{n} q_{ij}\alpha_j, C_0 + \sum_{j=1}^{n} |q_{ij}|C_j). \tag{7}$$

2) Goodness-of-fit:

Definition 1: Let A, B be the fuzzy sets in real space $\mathbf{R}$, then

$$h = \bigvee_{x \in \mathbf{R}} \{\mu_{\widetilde{A}}(x) \wedge \mu_{\widetilde{B}}(x)\} \tag{8}$$

is the *goodness-of-fit* from A to B.

In fact, the goodness-of-fit is defined as the inner product here. According to Definition 1, the goodness-of-fit from fuzzy number $\widetilde{A}(\alpha_1, C_1)$ to fuzzy number $\widetilde{B}(\alpha_2, C_2)$ is

$$h = \begin{cases} 1 - \frac{|\alpha_1 - \alpha_2|}{C_1 + C_2}, & |\alpha_1 - \alpha_2| \leq C_1 + C_2, \\ 0, & \text{otherwise}, \end{cases} \tag{9}$$

Hence, the goodness-of-fit h_i from $\widetilde{T}_i(T_i, e_i)$ to Eq. (7) $\widetilde{T}_i^*(\alpha_0 + \sum_{j=1}^n q_{ij}\alpha_j, C_0 + \sum_{j=1}^n |q_{ij}|C_j)$ is

$$h_i = \begin{cases} 1 - \dfrac{|T_i - (\alpha_0 + \sum_{j=1}^n q_{ij}\alpha_j)|}{C_0 + \sum_{j=1}^n |q_{ij}|C_j + e_i}, & |T_i - (\alpha_0 + \sum_{j=1}^n q_{ij}\alpha_j)| \\ & \leq C_0 + \sum_{j=1}^n |q_{ij}|C_j + e_i, \\ 0 & \text{otherwise.} \end{cases}$$

3) Fuzziness:

Definition 2: Let $A(\alpha, C)$ be the fuzzy number, then the *fuzziness* [28] of A is

$$S_A = \frac{1}{2}C. \tag{10}$$

The fuzziness measures how fuzzy or vague the fuzzy set is or how clear it is not. According to Definition 2, the fuzziness of $\widetilde{T}_i^*$ in Eq. (6) is

$$S_{\widetilde{T}_i^*} = \frac{1}{2}(C_0 + \sum_{j=1}^n |q_{ij}|C_j). \tag{11}$$

C. Fuzzy Linear Regression Analysis (FLRA)

In the FLRA model, we are interested in finding a fuzzy function $\widetilde{y}$ in the form given below Eq. (12) which fits a finite number of numeric input-fuzzy output data $\big((x_{1i}, x_{2i}, \cdots, x_{(n-1)i}, x_{ni}), \widetilde{y}_i^*\big)$ $i = 1, \cdots, k$ with a minimum uncertain [24].

$$\widetilde{y}_i = \tag{12}$$
$$\widetilde{A}_n x_{ni} + \widetilde{A}_{n-1} x_{(n-1)i} + \cdots + \widetilde{A}_j x_{ji} + \cdots + \widetilde{A}_1 x_{1i} + \widetilde{A}_0$$

In a more general situation, both inputs and outputs are treated as fuzzy numbers. In this case, we are looking for a function such as the one given by Eq. (13), to fit the fuzzy input and fuzzy output data $\big((\widetilde{x}_{1i}, \widetilde{x}_{2i}, \cdots, \widetilde{x}_{(n-1)i}, \widetilde{x}_{ni}), \widetilde{y}_i^*\big)$.

$$\widetilde{y}_i = \tag{13}$$
$$\widetilde{A}_n \widetilde{x}_{ni} + \widetilde{A}_{n-1} \widetilde{x}_{(n-1)i} + \cdots + \widetilde{A}_j \widetilde{x}_{ji} + \cdots + \widetilde{A}_1 \widetilde{x}_{1i} + \widetilde{A}_0$$

where in Eqs (12) and (13), $\widetilde{A}_j$ $(j = 0, \cdots, n)$ is the fuzzy coefficient (parameter) of the regression model. The parameters are optimized such that the differences between the observed outputs $\widetilde{y}_i^*$ and estimated ones $\widetilde{y}_i$ are as small as possible. All $\widetilde{y}_i$ $(i = 1, \cdots, k)$, $\widetilde{A}_j$ $(j = 0, \cdots, n)$ and $\widetilde{x}_{ji}$ are fuzzy numbers and the operations used there are treated as the multiplication and addition of fuzzy numbers.

The first FLR model, proposed by Tanaka et al. [24], was based on the possibility theory and solved by Linear Programming (LP). Dimond [29] proposed a FR model based on the Least Square (LS) approach. According to the definition of the error, FR models are classified in two categories: probabilistic with LP, and LS approaches. In the former approach, the aim is to minimize the overall fuzziness by minimizing the total spread of the fuzzy coefficients while the estimated outputs and the observed ones are within a certain h-level of confidence [30]. The term h expresses the fitness between the estimated

fuzzy outputs and the observed ones. On the other hand, in the LS-based model, a similarity measure between observed and estimated outputs, is used as the measurement for the fitness of the model.

In fuzzy mathematics, LP is a well-known technique used in probabilistic approaches to minimize the fuzziness of the coefficients. In the LP based approaches with additional observation data, two constraints are added to LP [20]. However, having extra constraints is not an issue in practice since the data sets for fuzzy regression are small and also, LP problems even with thousands of variables and constraints can be solved in a few seconds. The goal of LP based approaches is to cover the spreads of all the observations (up to an h-level). The goal of LS based approaches is to find a model which has the most similar estimated outputs to the observed ones based on the chosen similarity measure.

For example, here regression analysis can be used to build up a model that represents the trust of the delivered service as a function of a set of advertised QoS values. Then, this model can be used to predict the trust of the new delivered service with a new set of advertised QoS values in the forthcoming transaction.

IV. FUZZY REGRESSION BASED TRUST PREDICTION

In this paper, we investigate the application of FLRA for the trust prediction based on a set of advertised QoS values. The main goal in this paper is to predict the trust rating of a web-service, which could be represented by linguistic expressions such as platinum, gold, silver and bronze. So, a common approach is to seek the relationship between the independent attributes (the QoS attributes) and dependent attribute (the trust attribute). For doing so, a numeric form of the trust attribute is considered to construct the model – even if the QoS data comes with some imprecision or uncertainty. Hence, initially the linguistic expressions of trust are all considered to be ordinal data. Then, after they are quantized (defuzzified), classical approaches can be applied [10]. Clearly, the quantization causes some important information to be overlooked or neglected [20]. In environments with uncertain data, the best option is to not defuzzify the data but to take the uncertainty into consideration in the level of inference [10]. However, the final results could be defuzzified for the presentation purposes only if it becomes necessary [10].

A linear regression model in the form of

$$Y = a_0 + a_1 x_1 + \cdots + a_n x_n,$$

where n is the number of input variables, is applied for QoS based trust estimation for each web-service. However, we have added an extra term to fuzzify the results. A triangular fuzzy value, like $\widetilde{B}_i$ which derives from

$$(\sum_{i=1}^n x_i) \cdot (b_1, b_2, b_3),$$

where $b_1 \leq b_2 \leq b_3$ is added to the above linear regression model given in Eq. (14).

$$\widetilde{Y}_i = a_0 + a_1 x_1 + \cdots + a_9 x_9 + \widetilde{B}_i \qquad (14)$$

The coefficients $a_0, a_1, \cdots, a_n$ are estimated by statistical linear regression. The motivation of applying the model given in Eq. (14) is due to the fact that this model does not have spread increasing problem – an issue exists in the original models given in Eqs (12) and (13). A model with spread increasing problem would increases the uncertainty of the trust of a service by the increase in the magnitude of the input QoS variables.

To derive the best values of b_1, b_2 and b_3, tabu-harmony search are applied [20]. In order to measure the goodness of the individual models in tabu-harmony search, we need to count how many predictions of the model are correct. However, the models have fuzzy estimation of the trust value which is hard to be compared with the actual trust value which are ordinal values. One way to tackle this issue is to defuzzify the estimation and then assign it to a class of trust which has the smallest distance. Basically by doing so, the fuzziness is taken into consideration for inference and the quantization takes place only in the final stage for the sake of presentation [10]. In this paper, we have defined the same margin for all classes by assigning a class for each estimation. For example, if the margin is 0.2, then in our example any estimations in the bands of

$$(-\infty, 1.2], \ (1.2, 2.2], \ (2.2, 3.2], \ \text{and} \ (3.2, +\infty).$$

We introduce b_4 to derive the best margin. Since the value of b_4 depends on the selection of b_1, b_2 and b_3, we optimize all the four variables in one phase.

In contrast to the other methods, the trust will be estimated in terms of a fuzzy number. In fact, trust is naturally an inexact term. However, due to the abstraction, it is usually converted to a traditional classification problem and the fuzziness is overlooked.

V. EXPERIMENTS

The most comprehensive data sets for web-services classification used by [19], [31], [32], include ten crisp independent attributes and one linguistic expression which represents the trust level of the web-service. To the best of our knowledge, in all studies in the literature, the trust terms (being platinum, gold, silver and bronze) are quantized to 1 to 4, respectively.

In the experiment, the web-services data set used for the simulation purposes is taken from [33]. The data set contains 365 real web service implementations that exist on the Web today. These services were collected using the Web Service Crawler Engine (WSCE). The majority of the Web services were obtained from public sources on the Web including Universal Description, Discovery, and Integration (UDDI) registries, search engines, and service portals.[1] From the list of attributes given in Table I, we have excluded *WSRF* due to its high dominance [19].

The settings of experiments in this section are as follows.

[1]For more information visit:
http://www.uoguelph.ca/~qmahmoud/qws/index.html#Service_Classification_

- The introduced data set with the properties given in Table I is first normalized and then we perform 10-fold cross validation as used in [19].
- tabu-harmony search a population size of 100 for 2000 iterations is used.
- The migration and crossover fractions are set to 0.7 and 0.05, respectively.
- The lower and upper bounds of b_1, b_2, b_3 and b_4 are set to -1 and 1, respectively.

The comparison of the method and the results obtained from Support Vector Machines (SVM), decision tree and linear regression are given in Table II. We can observe the following results.

- The accuracy of the proposed method based on fuzzy linear regression is about 80%, which is superior to the other methods compared in this paper.
- The standard deviation is estimated to be 0.07, which reveals the robustness of the approach on different runs.
- Both standard deviation and the performance of our proposed fuzzy linear regression based approach is better than the one which applies traditional linear regression.
- Moreover, as discussed earlier, fuzzy linear regression based approach is able to re-generate the overlooked uncertainty in the trust of service.
- The standard deviations and the training performances of the other methods are missing in Table II, as they are not reported in [19].

As an example, we have estimated the trust of two web services from the obtained model during the training. The trust values are in the form of fuzzy triangular numbers as follows:

$$(1.9139, 3.8692, 4.6155), \ (1.5714, 3.4027, 4.1017).$$

Although after defuzzification both would be classified as *gold*, Figure 1 illustrates that these two fuzzy numbers reveal that the first service is slightly better than the second one. Moreover, there is an uncertainty that the quality of both services depicted in Figure 1 might be silver, as the left spreads of both fuzzy numbers are very wide.
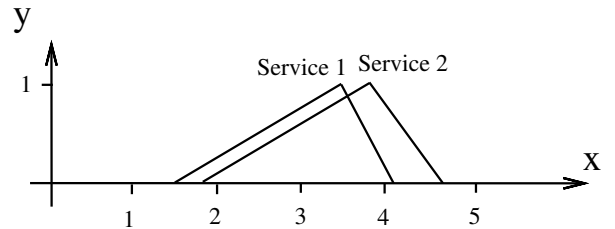

Fig. 1. The comparison of the trust of two web services before defuzzification

From Figure 1, one may argue that with the statistical linear regression or other exact methods, the trust of service 2 would be estimated to be better than service 1. However, the exact trust of service 1 might be better than an alternative service 2, although service 2 could be better.

TABLE I
WEB-SERVICE DATASET; ATTRIBUTES AND DESCRIPTIONS [33]

ID	Attribute name	Description	Units
x_1	Response Time	Time taken to send a request and receive a response	ms
x_2	Availability	Number of successful invocations/total invocations	%
x_3	Throughput	Total Number of invocations for a given period of time	$\frac{invokes}{second}$
x_4	Successability	Number of response / number of request messages	%
x_5	Reliability	Ratio of the number of error messages to total messages	%
x_6	Compliance	The extent to which a documentation follows specification	%
x_7	Best Practices	The extent to which a Web service follows WS-I Basic Profile	%
x_8	Latency	Time taken for the server to process a given request	ms
x_9	Documentation	Measure of documentation (i.e. description tags) in WSDL	%
x_{10}	WSRF	Web Service Relevancy Function	%
y	Service Classification	Levels representing service offering qualities (1 through 4) (trust rating)	Type

TABLE II
ACCURACY OF THE CLASSIFIERS AFTER REMOVING WSRF

Classifier	Average performance on testing sets (%)	Standard deviation of the testing sets errors	Average performance on training sets (%)	Standard deviation of the training sets errors
Fuzzy linear regression (This work)	79.46	0.0756	83.36	0.0060
Decision Trees (CART) [19]	78.61	-	-	-
Linear regression (This work)	74.17	0.0816	79.91	0.0095
SVM [19]	60.55	-	-	-
Decision trees (J48) [19]	67.77	-	-	-

As another example, we assume the trust of two web services given in Figure 2. In case of using traditional (exact) methods, it is deduced that service 2 is superior to service 1. However, as the left spread of service 2 goes all the way to the bronze quality, it may result in selecting service 1 over service 2.
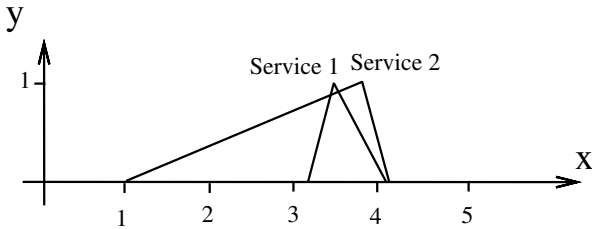

Fig. 2. An example where a traditional method and an inexact method may vary in the result

VI. CONCLUSIONS

In this paper, we have proposed an approach based on fuzzy linear regression to extract the qualitative information which may have been overlooked in the original data set. The extracted qualitative data is then used for modeling the data. The proposed approach is applied for the trust prediction of the delivered web-services based on a set of advertised QoS values. The method not only provides more detailed information of trust but it also has promising results. Although in order to quantitatively compare the performance of the method, the results are defuzzified, one could use the fuzzy results as they provide more information.

For future work, one may consider having different weights for each of the QoS variables. By doing so the number of variables for optimization increases, but the obtained model will be more generalized.

REFERENCES

[1] M. P. Papazoglou, P. Traverso, S. Dustdar, and F. Leymann, "Service-oriented computing: a research roadmap," *Int. J. Cooperative Inf. Syst.*, vol. 17, no. 2, pp. 223–255, 2008.

[2] L. Li, Y. Wang, and E.-P. Lim, "Trust-oriented composite service selection and discovery," in *Proceedings of the International Conference on Service-Oriented Computing* (ICSOC/ServiceWave 2009), pp. 50–67, 2009.

[3] L. Li, Y. Wang, and E.-P. Lim, "Trust-oriented composite service selection with qos constraints," *Journal of Universal Computer Science (J. UCS)*, vol. 16, no. 13, pp. 1720–1744, 2010.

[4] L. Li and Y. Wang, "Subjective trust inference in composite services," in *Proceedings of the AAAI Conference on Artificial Intelligence* (AAAI 2010), pp. 1377–1384, 2010.

[5] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decision Support Systems*, vol. 43, no. 2, pp. 618–644, 2007.

[6] Y. Wang and E.-P. Lim, "The evaluation of situational transaction trust in e-service environments," in *Proceedings of the IEEE International Conference on e-Business Engineering* (ICEBE 2008), pp. 265–272, 2008.

[7] G. Zacharia and P.Maes, "Trust management through reputation mechanisms," *Applied Artificial Intelligence*, vol. 14, no. 9, pp. 881–907, 2000.

[8] L. Xiong and L. Liu, "PeerTrust: Supporting reputation-based trust for peer-to-peer electronic communities," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 7, pp. 843–857, 2004.

[9] L.-H. Vu, M. Hauswirth, and K. Aberer, "QoS-based service selection and ranking with trust and reputation management," in *Proceedings of the International Conference on Cooperative Information Systems* (CoopIS 2005), pp. 466–483, 2005.

[10] W. Näther, "Regression with fuzzy random data," *Computational Statistics & Data Analysis*, vol. 51, no. 1, pp. 235–252, 2006.

[11] R. M. Dom, *Fuzzy regression model for the prediction of oral cancer susceptibility*. PhD thesis, University of Malaya, 2009.

[12] C. Kao and C.-L. Chyu, "Least-squares estimates in fuzzy regression analysis," *European Journal of Operational Research*, vol. 148, no. 2, pp. 426–435, 2003.

[13] L. Li, Y. Wang, and V. Varadharajan, "Fuzzy regression based trust prediction in service-oriented applications," in *Proceedings of the 6th International Conference on Autonomic and Trusted Computing* (ATC '09), (Berlin, Heidelberg), pp. 221–235, Springer-Verlag, 2009.

[14] *eBay*. http://www.eBay.com/.

[15] E. Damiani, S. D. C. di Vimercati, S. Paraboschi, P. Samarati, and F. Violante, "A reputation-based approach for choosing reliable resources in peer-to-peer networks," in *ACM Conference on Computer and Communications Security* (CCS 2002), pp. 207–216, 2002.

[16] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, "The eigentrust algorithm for reputation management in p2p networks," in *Proceedings of the international conference on World Wide Web* (WWW 2003), pp. 640–651, 2003.

[17] Y. Wang, K.-J. Lin, D. S. Wong, and V. Varadharajan, "Trust management towards service-oriented applications," *Service Oriented Computing and Applications*, vol. 3, no. 2, pp. 129–146, 2009.

[18] E. Al-Masri and Q. H. Mahmoud, "Discovering the best web service: A neural network-based solution," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pp. 4250 – 4255, 2009.

[19] R. Mohanty, V. Ravi, and M. R. Patra, "Web-services classification using intelligent techniques," *Expert Syst. Appl.*, vol. 37, no. 7, pp. 5484–5490, 2010.

[20] M. H. Mashinchi, M. A. Orgun, M. Mashinchi, and W. Pedrycz, "Tabu-harmony search based approach to fuzzy linear regression," *IEEE Transactions on Fuzzy Systems*, 2011. DOI:10.1109/TFUZZ.2011.2106791.

[21] A. Bargiela, W. Pedrycz, and T. Nakashima, "Multiple regression with fuzzy data," *Fuzzy Sets and Systems*, vol. 158, no. 19, pp. 2169–2188, 2007.

[22] R. A. Aliev, B. Fazlollahi, and R. M. Vahidov, "Genetic algorithms-based fuzzy regression analysis," *Soft Computing*, vol. 6, no. 6, pp. 470–475, 2002.

[23] B. Kim and R. R. Bishu, "Evaluation of fuzzy linear regression models by comparing membership functions," *Fuzzy Sets and Systems*, vol. 100, no. 1-3, pp. 343–352, 1998.

[24] H. Tanaka, S. Uejima, and A. Asai, "Linear regression analysis with fuzzy model," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 12, pp. 903–907, 1982.

[25] D. H. Hong and C. Hwang, "Support vector fuzzy regression machines," *Fuzzy Sets and Systems*, vol. 138, no. 2, pp. 271–281, 2003.

[26] D. T. Redden and W. H. Woodall, "Further examination of fuzzy linear regression," *Fuzzy Sets and Systems*, vol. 79, pp. 203–211, 1996.

[27] W. W. Hines, D. C. Montgomery, D. M. Goldsman, and C. M. Borror, *Probability and Statistics in Engineering*. John Wiley & Sons, Inc, 2003.

[28] A. T. Azar, *Fuzzy Systems*. Intech, 2010.

[29] P. Diamond, "Fuzzy least squares," *Information Sciences*, vol. 46, no. 3, pp. 141–157, 1988.

[30] J. Lu and R. Wang, "An enhanced fuzzy linear regression model with more flexible spreads," *Fuzzy Sets and Systems*, vol. 160, no. 17, pp. 2505–2523, 2009.

[31] E. Al-Masri and Q. H. Mahmoud, "Investigating web services on the world wide web," in *Proceeding of the 17th international conference on World Wide Web* (WWW '08), (New York, NY, USA), pp. 795–804, ACM, 2008.

[32] M. Alrifai, T. Risse, P. Dolog, and W. Nejdl, "Service-oriented computing — icsoc 2008 workshops," ch. A Scalable Approach for QoS-Based Web Service Selection, pp. 190–199, Berlin, Heidelberg: Springer-Verlag, 2009.

[33] E. Al-Masri and Q. H. Mahmoud, "Discovering the best web service," in *Proceedings of the 16th international conference on World Wide Web* (WWW '07), (New York, NY, USA), pp. 1257–1258, ACM, 2007.