A Web Service QoS Forecasting Approach Based on Multivariate Time Series

Pengcheng Zhang[1], Liyan Wang[1], Wenrui Li[2], Hareton Leung[3], Wei Song[4]
[1]College of Computer and Information, Hohai University, Nanjing, China, 211110
[2]School of Mathematics & Information Technology, Nanjing Xiaozhuang University, Nanjing, P.R.China 211147
[3]Department of Computing, Hong Kong Polytechnic University, HongKong, China
[4]School of Computer Science and Engineering, Nanjing University of Science and Technology
Email: pchzhang@hhu.edu.cn

***Abstract*—In order to accurately forecast Quality of Service (QoS) of different Web Services, this paper proposes a novel QoS forecasting approach called MulA-LMRBF (Multi-step forecasting with Advertisement and Levenberg-Marquardt improved Radial Basis Function) based on multivariate time series. Considering the correlation among different QoS attributes, we use phase-space reconstruction to map historical multivariate QoS data into a dynamic system, use Average Dimension (AD) to estimate the embedding dimension and delay time of reconstructed phase space. We also add the short-term QoS advertisement data of service provider to form a more comprehensive data set. Then, RBF (Radial Basis Function) neural network improved by the Levenberg-Marquardt (LM) algorithm is used to update the weight of the neural network dynamically, which improves the forecasting accuracy and realizes the dynamic multiple-step forecasting.The experimental results demonstrate that MulA-LMRBF is better than previous approaches in term of precision and is more suitable for multi-step forecasting.**

***Index Terms*—Quality of Service; Multivariate time series; Phase-space reconstruction; LM algorithm; RBF neural network; Dynamic multiple step forecasting**

I. Introduction

Service-Oriented Architecture (SOA) as a component model, leads the development of a new service-oriented Information Technology era. Web Service is currently the most popular technology to achieve SOA, and SOA combines the Web services with different functions of application through the interface, furthest achieved functional reuse and expansion [19]. Therefore, it is significant to choose a Web service that is suitable for the system and meets the demands of users.

With the rapid development Service Oriented Computing (SOC) technology, there are many similar Web services on the network, which increase the difficulty of service selection and combination [12]. In the process of choosing Web services to satisfy users requirements, non-functional attributes are often ignored. In recent years, QoS as a nonfunctional factor, has been paid more and more attention. Researchers try to provide reference about service selection and combination through QoS attributes [22]. Due to the changeable network environment and other factors, the real QoS values have the characteristics of nonlinearity, dynamics and variability in the time dimension. Consequently, it is important to provide service QoS information effectively and accurately before choosing suitable services [5].

At present, more and more researchers have proposed different kinds of Web service QoS forecasting approaches. Most of these approaches are based on Structural Equation [4], similarity measurement [15], [16], artificial intelligence [21], or time series model [11]. They provide technical supports to select high quality service or to avoid choosing failure services. Most of approaches use historical data to forecast the future value of the target QoS attributes.

However, current Web service QoS forecasting approaches mainly suffer from the following problems: *i)* The relationship among different QoS attributes is not well considered. QoS attributes include *response time*, *throughput*, *reliability*, *availability*, *price*, and so on [5], [21]. At present, many studies only consider a single attribute [27] regardless of the sample size and the forecast period. In addition, the complex relationship among different attributes cannot be described by accurate equation models, and fixed relationship often changes with different data samples. *ii)* The forecasting period is short, as a result, long-term or multi-step forecasting result is very bad. With dynamic and changeable characteristics of QoS, directly using historical data of the target attribute to forecast with multiple steps does not accurately reflect the future QoS attribute values. Therefore, the forecasting accuracy is very limited. Forecasting results for single step model is good, but these approaches are not suitable for multi-step/long-term forecasting. *iii)* Most of current approaches do not support online dynamic forecasting. QoS has dynamic and changeable characteristics [20]. The overall trend of QoS changes according to the service provider's different strategies [1]. In addition, network environment and other factors will also have an impact on QoS [27]. However, most of the fixed models are established based on historical data, without considering the updated data. Consequently, they are only suitable for static forecasting, and not suitable for online dynamic forecasting.

In order to solve these problems, this paper proposes a novel Web service QoS forecasting approach called MulA-LMRBF (Multiple-step forecasting with Advertisement and Levenberg-Marquardt improved Radial Basis Function) based on multivariate time series.

978-1-5386-0752-7/17 $31.00 © 2017 IEEE
DOI 10.1109/ICWS.2017.27

- For problem *i)*, considering the complex relationship among different QoS attributes, the time sequence of each QoS attribute of historical data is mapped to a power system by phase space reconstruction approach. In this way, the original multidimensional nonlinear system is approximately restored [6], and then the reconstruction system is used to express dynamic correlation information among multiple attributes.
- For problem *ii)*, given that the short-term QoS advertisement data is released based on the future trend of service, and it represents current QoS attribute values. Therefore, short-term advertisement data of QoS time series is added to the forecasting data set to realize multi-step forecasting.
- For problem *iii)*, RBF neural network model improved by LM (Levenberg-Marquardt) algorithm [3] is used to realize dynamic forecasting. In the improved model, the attribute values use the LM optimization algorithm in weight training, improve the efficiency of operation, and reduce the time overhead and space overhead [20] [13]. Consequently, in the process of forecasting, according to dynamic and changeable characteristics of QoS attributes, for each new collection sample, neural network parameter value is updated dynamically. In this way, dynamic forecasting is completely realized and has relatively accurate forecasting results.

The rest of this paper is organized as follows: Section 2 introduces related work of QoS forecasting. Section 3 gives background knowledge of this approach and the related theoretical basis. Section 4 introduces the proposed forecasting approach MulA-LMRBF. Section 5 presents the experimental design and result analysis. Section 6 gives conclusions and future work.

II. Related work

In recent years, driven by the development of service-oriented system, more and more people focus on web services QoS forecast technology, and emerged lots of forecasting models in the field. Existing forecasting approaches are mostly based on four different models or theories: Structural Equation [4], similarity measurement [15], [16], artificial intelligence [21] and time series [11].

The Structural Equation approach can analyze the properties of internal relations, but cannot forecast the concrete value [4], [8]. According to the trend of a single attribute history information, Le [8] et al. propose an approach which can only forecast the evolution of the attribute, and the prediction result of this approach is not accurate.

The similarity measurement approaches use QoS values of users who used the Web service to forecast other user's QoS [15], [16]. The earliest approach proposed by Shao et al [15], [16], uses collaborative filtering approach based on similarity of users to forecast unused service QoS. In general, this kind of approach is simple and fast, but the accuracy is not high, and can only be used for a single function and task.

The artificial intelligence approaches have strong adaptive ability and can forecast complex nonlinear QoS attribute values [26], [27], [10]. For example, Liu et al. [10] put forward a kind of BP (Back Propagation) neural network based Web service QoS prediction approach. However, the training speed and precision of prediction are not perfect, and most models have some disadvantages, such as slow training speed, no dynamic update mechanism, and poor timeliness.

The time series forecasting approaches mainly use time sequence of QoS attributes to construct model, and forecast the future QoS attribute values with time. Amin et al. [1] propose a forecast model which combines ARIMA and GARCH (Generalized AutoRegressive Conditional Heteroskedasticity) to predict reliability, and they find that QoS reliability has changeable characteristic on the time series. Whereafter, they come up with a novel approach to automatically predict QoS values based on linear and nonlinear time series. The approach uses the same default value for some model parameters, and consequently cannot offer high-precision prediction. Zhen et al. [23] propose a kind of QoS long-term forecast approach based on multivariate time series, which uses ARIMA and Holt-Winters to set up a long-term forecast model. But the approach is only designed for static forecasting, and no mechanisms supporting dynamic retraining.

In order to improve long-term forecasting precision and realize dynamic forecasting, this paper proposes a novel approach based on multivariate time series model called MulA-LMRBF.

III. Preliminaries

A. Phase space reconstruction

Nonlinear time series can be viewed as produced by certain nonlinear system. The phase space is one of the most powerful tools to describe movement and evolution of system [20]. It is a space used to represent all possible states of a system; each possible state of the system has a corresponding phase space point [6]. Phase space reconstruction is first put forward by Takens [18], who considers that the development of the random variables in the system are determined by the interaction with other variables, and the information about the relevant variable is implicit in the development of any other variables. The original system can be reconstructed by choosing appropriate embedding dimension [24].

In the process of multi-step forecasting of QoS attributes, each attribute data can be viewed as a univariate time series, and multiple QoS attribute samples form a multivariate time series. There is only one attribute in the univariate time series, which contains less nonlinear system information, and it is impossible to determine whether or not the original system of QoS time series would reconstructed in phase space reconstruction. However, multivariate time series contains more QoS information than the univariate time series, and consequently multivariate time series is more suitable for QoS dynamic forecasting with multi-step.

QoS time series are expressed as $X = \{x_1, x_2, ..., x_N\}$, where the number of samples is N, and $i = 1, 2, ..., N$. Each data sample of multivariate time series X represents the value of the QoS attribute at each point in time, expressed as $X_i = \left\{x_{i,1}, x_{i,2}, ..., x_{i,j}, ..., x_{i,M}\right\}$, where j represents the jth QoS attribute, X_i is the sample data at the ith time point and $x_{i,j}$ is the jth QoS attribute value of the ith time point data.

After phase space reconstruction, the new time series is defined as $X_i'=[x_{i,1}, ..., x_{i-(m_1-1)\tau_1,1}, x_{i,2}, ..., x_{i-(m_2-1)\tau_2,2}, ..., x_{i,j}, ..., x_{i-(m_j-1)\tau_j,j}, ..., x_{i,M}, ..., x_{i-(m_M-1)\tau_M,M}]$, where $i = 1, 2, ..., N$, m_j is the embedding dimension of the jth attribute, and τ_j is the time delay of the jth attribute. When the embedding dimensions and time delays are appropriate, the time series can satisfy nonlinear relationship, and the relationship can be expressed as $x'_{i+1} = f\left(x'_i\right)$.

B. Average displacement method

Average displacement (AD) is a kind of phase space reconstruction approach which simultaneously considers embedding dimension and time delay [6]. Its main idea is to introduce average displacement (AD) for each QoS attribute and give an embedding dimension m and time delay τ [9].

The average displacement is calculated as:

$$< S_m(\tau_j) >= \frac{1}{N} \sum_{t=1}^{n} \sqrt{\sum_{i=0}^{m-1} (x_{t-i\tau_j,j} - x_{t,j})^2} \quad (1)$$

For the embedding dimension m of the jth QoS attribute, the corresponding $< S_m(\tau_j) >$ is calculated. The original average displacement principle is that when the growth slope of the $< S_m(\tau_j) >$ is reduced to 40% of the initial value for the first time the corresponding point is the jth delay time. This paper adopts improved average displacement approach [9], namely with the increase of τ_j, the first peak point of τ_j is the delay time.

C. LM algorithm

LM (Levenberg-Marquardt) is a kind of widely used optimization approach [3], which is the compromise between Newton method and gradient descent method. The approach uses the approximate second order derivative to solve optimal value of the second-order function $f(W)$, and the speed is better than the gradient descent method. Let g be the gradient vector of the function, H is the Hessian matrix, then the optimal adjustment of the parameter vector W is $\Delta W = (H + \lambda I)^{-1} g$, where I is the unit matrix same as H, and λ is the regular coefficient to guarantee the positive definite of $H + \lambda I$.

D. RBF Neural Network

RBF proposed by Moody and Darken [17] is a feed-forward neural network based on the perceptual area of human cerebral cortex. Usually it includes input layer, hidden layer and output layer. The input layer is composed of perceptual neurons, and the input vector X_i is introduced into the neural network. The hidden layer maps the input vector from low dimension to high dimension to realize high dimensional curve fitting, and the hidden layer selects Gauss function $G(X)$ as its activation function. The output layer has a neuron, and the output of the hidden layer is linearly weighted as the final output value denoted as Y.

IV. MulA-LMRBF forecasting approach

The MulA-LMRBF architecture is shown in Fig.1. The first step is QoS data collection and pretreatment. The data includes complete *history data* and *short-term advertisement data*. It is clear that the history data refers to the values of the QoS attributes in the past time period. The advertisement data is the future QoS attributes information published by the service provider. The history data preprocessing includes data noise processing, scale transformation and phase space reconstruction, and advertisement data preprocessing only includes scale transformation. Then integrated QoS data set is composed of them. The second step is to train the forecasting model with the RBF neural network improved by LM algorithm by optimising the neural network weight W between the hidden layer and output layer, to further update W, and to achieve efficient training effect. In the third step, for each newly collected sample, we can calculate the related variables of W for dynamic forecasting.

A. Data collection and pretreatment

For the collected complete QoS data, history data and advertisement data are preprocessed respectively, and the details are described in the following steps:

(1) History data noise processing

The QoS history data contains a lot of noise, which can evolve over time and cause computational loss. The time series model with phase space reconstruction is sensitive to data noise. Therefore, the nonlinear wavelet transform threshold denoising method [2] is used to process QoS history data.

(2) Scale Transformation

When the neural network input variables between 0 and 1, the network operation effect will be better. Therefore, history data and advertisement data are normalized by scale transformation, and the values are controlled between [0,1].

(3) Phase space reconstruction

To reconstruct the phase space of QoS history data.Multivariate time series of QoS history data is $Q = \{Q_1, Q_2, Q_3, ...Q_i, ..., Q_N\}$, where the ith data is $Q_i = \left\{q_{i,1}, q_{i,2}, ..., q_{i,j}, ..., q_{i,M}\right\}$. The embedding dimension of jth attribute is m_j and the delay time is τ_j, the reconstructed time series is expressed as $Q' = \left\{Q'_1, Q'_2, ..., Q'_i, ..., Q'_N\right\}$, where the ith data Q'_i represents the reconstructed time series of Q_i, expressed as follows:

$$\begin{aligned} Q_i' = [&q_{i,1}, q_{i-\tau_1,1}, ..., q_{i-(m_1-1)\tau_1,1}, \\ &q_{i,2}, q_{i-\tau_2,2}, ..., q_{i-(m_2-1)\tau_2,2}, \\ &... \\ &q_{i,j}, q_{i-\tau_j,j}, ..., q_{i-(m_j-1)\tau_j,j}, \\ &... \\ &q_{i,M}, q_{i-\tau_M,M}, ..., q_{i-(m_M-1)\tau_M,M}] \end{aligned} \quad (2)$$

(4) QoS Integrated Data Set

The processed QoS history data and the QoS advertisement data are combined as an integrated data set expressed as $X = \{X_1, X_2, ..., X_i, ..., X_N\}$, where $X_i = [Q'_i, A'_i]^T$. In which, $A_i =$

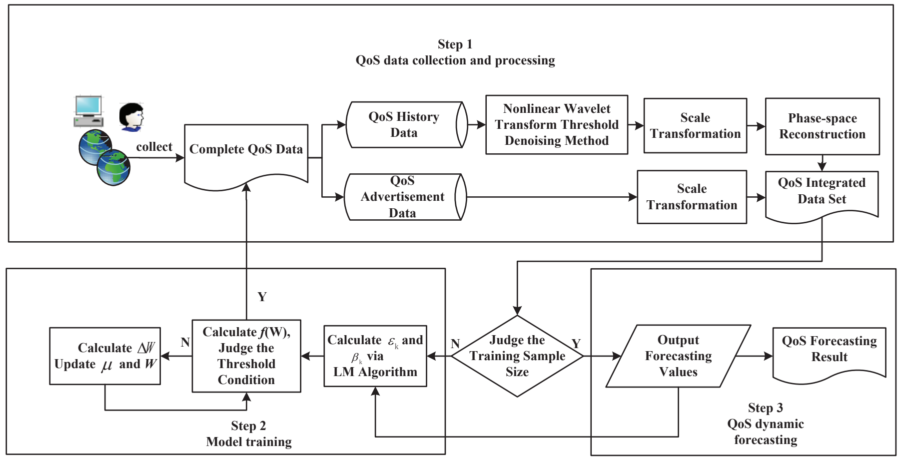


Fig. 1. MulA-LMRBF architecture overview

$\left\{a_{i,1}, a_{i,2}, ..., a_{i,j}, ..., a_{i,M}\right\}$, $A_i' = \left\{a_{i,1}', a_{i,2}', ..., a_{i,j}', ..., a_{i,M}'\right\}$, where $i = 1, 2, ..., N$, $j = 1, 2, ..., M$, N is the number of samples, M is the number of QoS attributes, Q_i' and A_i' represent processed history data and processed advertisement data, respectively.

B. MulA-LMRBF model training and dynamic forecasting

The RBF neural network improved by LM algorithm (LM-RBF) not only efficiently trains the network weight W, but also updates W for each input sample to realize dynamic forecasting in time. The input is X_i, output is Y_i, the function of network hidden layer is Gaussian [27]:

$$\varphi(X_i, C_k) = exp(-\frac{\|X_i - C_k\|^2}{2\sigma^2}) \tag{3}$$

where $i = 1, 2, ..., N$, $k = 1, 2, ..., L$, N represents the number of samples, L represents the number of hidden layer nodes, $\varphi(r)$ is the output function of hidden layer, C_k is the core of hidden layer, and σ is the expansion constant. The functions of σ are given as follows:

$$d_j = min(\left\|c_j - c_t\right\|), j \neq t \tag{4}$$

$$\sigma_j = \lambda d_j \tag{5}$$

where d_j is the minimum distance between the jth hidden node and all hidden nodes, λ is the overlap coefficient, and δ_j is the extended constant of the jth hidden node, and calculate by self organized selection method [14]. The function of RBF neural network output layer is

$$Y_i = \sum_{k=1}^{L} w_k \varphi(X_i, C_k) = \sum_{k=1}^{L} w_k exp(-\frac{\|X_i - C_k\|^2}{2\sigma_k^2}) \tag{6}$$

where $i = 1, 2, ..., N$, $W = [w_1, w_2, ..., w_L]^T$, Y_i is the output of RBF, w_k is the weight connecting the kth hidden node to the output node, and W is the weight vector between the hidden layer and output layer.

(1) The model training phase

The purpose of the model training is to satisfy the error function $f(W)$ as small as possible until the threshold condition is reached. T_i represents the ith actual value. When $f(W)> \delta$, weight W is update iteratively.

$$f(W) = \frac{1}{2N}\sum_{i=1}^{N} s_i^2 = \frac{1}{2N}\sum_{i=1}^{N}(Y_i - T_i)^2 \tag{7}$$

LM algorithm is used to train W, as shown in algorithm 1. In the training process, the iterative updating functions of W are

$$W = W + \Delta W \tag{8}$$

$$\Delta W = (J^T J + \mu I)^{-1} J^T S \tag{9}$$

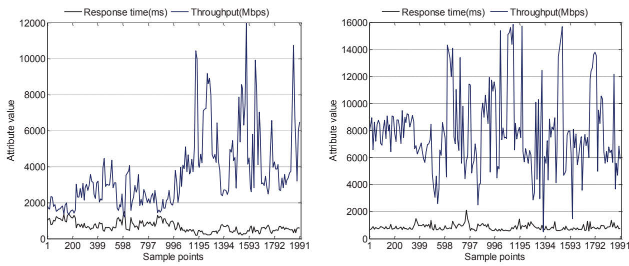
For adjustment of W, the approximate second derivative instead of the original complex calculation is used. In the Nth iteration, if $f(W)$ is greater than $f(W)$ of the $n-1$th iteration, the regularization coefficient $\mu = a * \mu$ $(\mu > 0, a > 1)$ is changed. Otherwise when $\mu = \mu\ /a$, W is updated [7]. The error matrix S of the sample forecasting values and the actual values is $S = [s_1, s_2, ..., s_N]$, and $s_i = Y_i - T_i$. The Jacobian matrix J of $f(W)$ is

$$J = \begin{bmatrix} \frac{\partial s_1}{\partial w_1} & \cdots & \frac{\partial s_1}{\partial w_L} \\ \frac{\partial s_2}{\partial w_1} & \cdots & \frac{\partial s_2}{\partial w_L} \\ \cdots & \cdots & \cdots \\ \frac{\partial s_N}{\partial w_1} & \cdots & \frac{\partial s_N}{\partial w_L} \end{bmatrix} = \begin{bmatrix} \varphi(X_1, C_1)...\varphi(X_1, C_L) \\ \varphi(X_2, C_1)...\varphi(X_2, C_L) \\ \cdots \\ \varphi(X_N, C_1)...\varphi(X_N, C_L) \end{bmatrix} \tag{10}$$

By simplification, J is expressed as a matrix of hidden layer function, denoted as Γ. ΔW can be calculated more efficiently. Firstly, $J^T J$ is calculated,

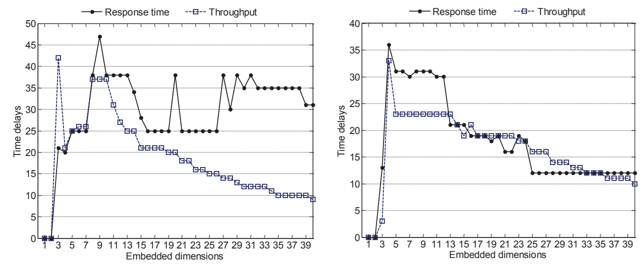
$$J^T J = \Gamma^T \Gamma = \sum_{k=1}^{N} \xi_k^T \xi_k \tag{11}$$

$$\xi_k = [\varphi(X_k, C_1)\ \varphi(X_k, C_2)\ ...\ \varphi(X_k, C_L)] \tag{12}$$



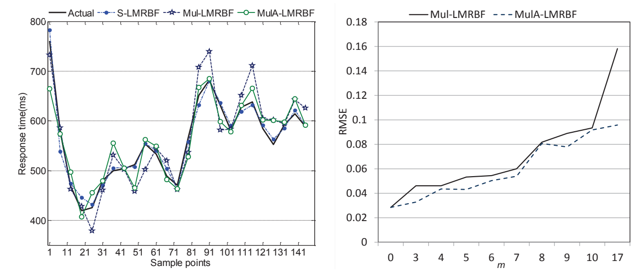
(a) RMB Instant Quotation (b) MobileCode

Fig. 2. Original data



(a) RMB Instant Quotation (b) MobileCode

Fig. 3. $m-\tau$ curve for processed history QoS data



(a) Forecasting results($step$=1 m=4) (b) RMSE($step$=1)

Fig. 4. Forecasting results and $RMSE$ for unvariate and multivariate time series for RMB

Then $J^T S$ is calculated,

$$J^T S = \Gamma^T S = \sum_{k=1}^{N} \xi_k^T s_k \tag{13}$$

where

$$\xi_k^T s_k = \beta_k = s_k[\varphi(X_k, C_1)\ \varphi(X_k, C_2)\ \ldots\ \varphi(X_k, C_L)]^T \tag{14}$$

When a sample is collected, ξ_k and β_k are calculated, finally W is updated. The algorithm is described in Algorithm 1.

Algorithm 1 The RBF neural network forecasting algorithm improved by LM algorithm

Require: In $f(W) = 0$, $\xi_k = 0$, $\beta_k = 0$ Randomly generated L, W and S;

Ensure: The updated weight W, $W = [w_1, w_2, ..., w_L]^T$.

$f(W1) = f(W)$; //$f(W1)$ is $f(W)$ in the last iteration
Calculate ξ_k and β_k;
if $count = TrainingSize$ **then**
  Calculate $f(W)$, W;
  while $f(W) > \delta$ **do**
    $f(W2) = f(W)$; //$f(W2)$ is $f(W)$ in the this iteration
    Calculate ΔW;
    if $f(W2) > f(W1)$ **then**
      $\mu = a * \mu$; //μ is the regularization coefficient
    else
      $\mu = \mu/a$;
      Update W, $W = W + \Delta W$;
    end if
    Calculate $f(W)$;
  end while
end if

(2) The QoS dynamic forecasting

To support the model dynamically forecast QoS attribute values, the idea of model training is extended to the forecasting phase, as shown in Algorithm 1. After training, when a new online data is collected, the corresponding ξ_k and β_k are calculated again. When the sample size reached training size, the error function $f(W)$ is also calculated. To judge the threshold condition, when it is not satisfied, according to the error function $f(W)$, μ and W are adjusted. With the collection of online data, all the parameters of the model are updated as soon as possible to adjust the requirements of dynamic and nonlinear QoS forecasting, which can improve forecasting accuracy.

V. Experimental evaluation

In this section, we conduct a series of experiments to validate MulA-LMRBF based on collected Web services data sets that contain multiple QoS attributes. These experiments are designed to investigate the following basic questions:

- REQ 1: Do *phase space reconstruction* and *advertisement* data influence QoS forecasting?
- REQ 2: Is multivariate time series more accurate than traditional univariate time series forecasting?
- REQ 3: Is MulA-LMRBF approach better than other traditional approaches?

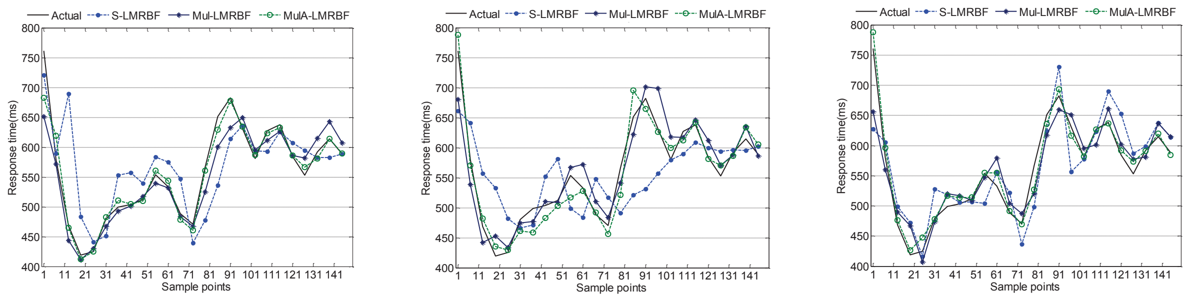
A. Data Collection and Setup

We conducted our experimental evaluation on a PC with Intel(R) Core(TM) i5-4200U CPU@1.60GHz and 4.00GB RAM, Windows 7, used Matlab7.11 to establish the forecasting model, and to achieve dynamic forecast.

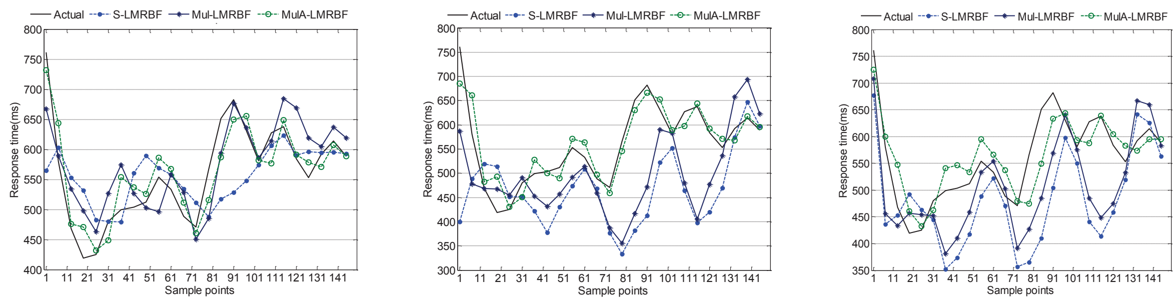
There are four parts in the experimental data. The first part is public dataset [1]. The second part is the data of different Web services from datatang [2]. The third part data is collected by ourselves. The fourth part is the simulated short-term QoS advertisement data. These data mainly include *response time* and *throughput*, using the two attributes to forecast future *response time*.

The collection time of the first part data is three months from 8:00 to 17:00 every day, which has a record each 15 minutes. In this way, 4 Web service QoS data sets with

[1] https://sourceforge.net/projects/qosmonitoring/files/
[2] http://more.datatang.com/data/



(a) Forecasting results when *step*=3, *m*=4 (b) Forecasting results when *step*=5, *m*=4 (c) Forecasting results when *step*=3, *m*=6



(d) Forecasting results when *step*=5, *m*=6 (e) Forecasting results when *step*=3, *m*=8 (f) Forecasting results when *step*=5, *m*=8

Fig. 5. Different forecasting results for unvariate and multivariate time series for RMB

8000 data points are collected. Fig.2(a) shows the example *RMB Instant Quotation* data curve. The second part is from 142 distributed computers of the QoS data of the 4532 real-world Web services. The third part comes from the data collected by the proxy server.The example *MobileCode* data is shown in Fig.2(b). According to the nonlinear and dynamic characteristics of QoS, the fourth part simulates advertisement data of response time. Compared with the real data, the fitting degree is controlled within 50% – 80%.

The measurement is based on the $RMSE$ (the Root Mean Square Error) to compare the forecasting results more intuitively:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i - T_i)^2} \tag{15}$$

where Y_i is the ith forecasted value, T_i is the ith actual value, and N represents the number of samples. $RMSE$ can represent the relative error of forecasting, which can reflect the stability of forecasting.

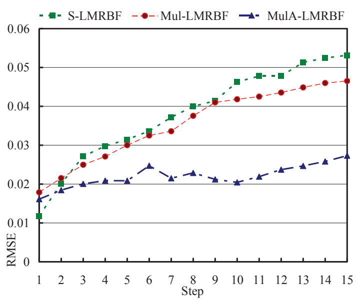
B. Experimental results

To address *REQ 1)*, we separately create three sets of experiments to indicate that *phase space reconstruction* and *advertisement* data have influences on the multi-step forecasting. i) LMRBF forecasting approach only based on *response time* (S-LMRBF). ii) LMRBF forecasting approach for multivariate time series with *phase space reconstruction* based on *response time* and *throughput* (Mul-LMRBF). iii) LMRBF forecasting approach for multivariate time series with *phase space reconstruction* and *advertisement* data based on *response time* and *throughput* (MulA-LMRBF).

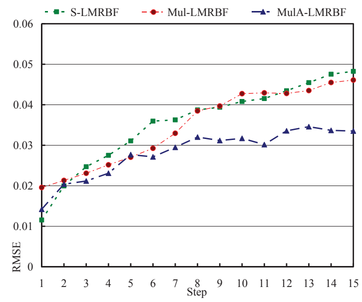
The average displacement approach was used to simultaneously consider m and τ, and to calculate every τ corresponding to each m in the range of [1,40]. Fig.3(a) and Fig.3(b) show $m-\tau$ curve of *RMB* data and *MobileCode* data, respectively. Considering the limited number of samples and the fluctuation of m and τ, for *RMB* data the range of [4,8] is selected and for *MobileCode* data the range of [5,10] is selected.

Taking *RMB* data as an example, 1000 data is used as training samples, and 300 data is used as test samples. Fig.4(a) gives single step forecasting results of 150 data when m=4, and it can be shown that the fitting degree of S-LMRBF forecasting results is high. Consequently, the information of QoS univariate time series is sufficient to realize the requirement of single step forecasting. Fig.4(b) shows the $RMSE$ value of 300 test samples for single step forecasting, when m is in the range of [0,17]. The forecasting error of MulA-LMRBF is smaller than Mul-LMRBF. When m=0, it represents the corresponding $RMSE$ value of S-LMRBF.When the forecasting step is shorter, univariate time series of QoS can be used to forecast future value.

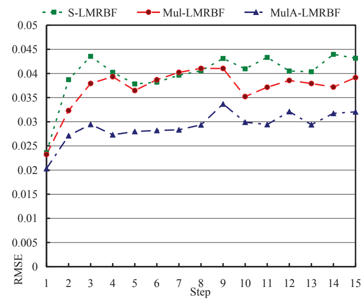
Fig.5 shows the comparison results of multi-step forecasting under different *phase space reconstruction*. When m is 4,6,8 respectively and forecasting response time with 3 and 5 steps, the results of 6 experiments show that the forecasting error increases with the increase of *step*, but MulA-LMRBF and Mul-LMRBF forecasting results are better than S-LMRBF forecasting results. The forecasting results depend on the parameters of phase space reconstruction. When reconstruction dimension m is in the stable range of $m-\tau$ curve, the forecasting results



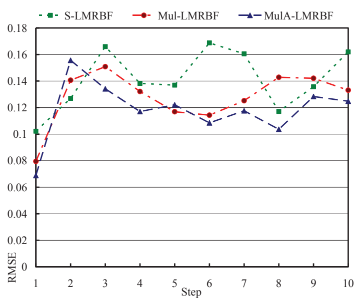
(a) The relationship between *step* and $RMSE$ for RMB, when m=4



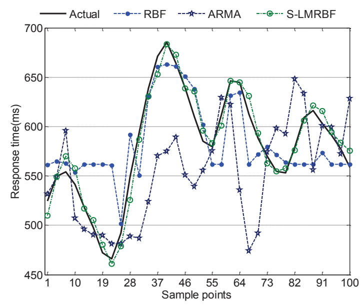
(b) The relationship between *step* and $RMSE$ for RMB, when m=6



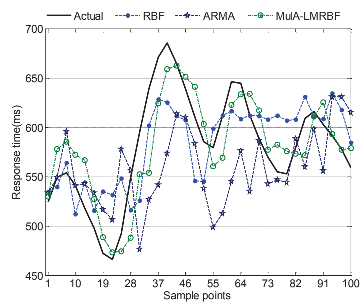
(c) The relationship between *step* and $RMSE$ for MobileCode, when m=5



(d) The relationship between *step* and average $RMSE$ for the second data, when m=4



(e) Comparison of forecast results for RMB, when m=6, *step*=1



(f) Comparison of forecast results for RMB, when m=6, *step*=5

Fig. 6. Forecasting results and $RMSE$ of response time

have high fitting degree. In addition, the fitting results of MulA-LMRBF is also better than Mul-LMRBF. Therefore, the *advertisement data* for QoS multi-step forecasting will further improve the forecasting accuracy. In summary, *phase space reconstruction* and adding *advertisement data* improve the accuracy of multi-step forecasting.

To address *REQ 2)*, we can observe the forecasting changes of $RMSE$ in S-LMRBF, Mul-LMRBF and MulA-LMRBF by increasing *step* size. The $RMSE$ is calculated using normalized data.It is shown that $RMSE$ of MulA-LMRBF is the smallest when $step \geqslant 1$, and RMSE of Mul-LMRBF and MulA-LMRBF multi-step forecasting are generally better than S-LMRBF. Therefore, in multi-step and dynamic forecasting, multivariate time series is more accurate than univariate time series forecasting.

Fig.6 represents the results for RMB 300 test samples that compare the $RMSE$ of 3 kind of experiments. As shown in Fig 6(a), when $step \geqslant 3$, and m=4, $RMSE$ of Mul-LMRBF is always less than S-LMRBF. In Fig 6(b), when m=6 and *step* in [3,9], Mul-LMRBF is better than S-LMRBF.

Fig.6(c) shows $RMSE$ of MobileCode 300 test samples, the parameters of *phase space reconstruction* are selected from the stable range of the $m - \tau$ curve. Taking m=5 as an example, $RMSE$ increases with the increase of the *step*.

258 complete Web services QoS data from the second part is selected for testing. Then we analyze the $RMSE$ value of m in the range of [3,10]. Fig.6(d) shows the average $RMSE$ of 258 groups of experiments when m=4. When *step* is in the range of [3,7], MulA-LMRBF is better than Mul-LMRBF and S-LMRBF. Compared with S-LMRBF, the average $RMSE$ of MulA-LMRBF decreases by 19% in general, and average $RMSE$ of Mul-LMRBF decreases by 16% compared with S-LMRBF.

All the above experiments fully show that multivariate time series is more suitable for multi-step forecasting.

Web Service	RMB Instant Quotation			
Method	RBF	ARMA	S-LMRBF	MulA-LMRBF
Step=1	0.0827	0.1193	0.0115	0.0142
Step=3	0.0955	0.1473	0.0247	0.0211
Step=5	0.1159	0.2062	0.0311	0.0276

TABLE I
COMPARISON OF $RMSE$ FOR RMB DATA FORECASTING RESULTS

Web Service	MobileCode			
Method	RBF	ARMA	S-LMRBF	MulA-LMRBF
Step=1	0.1004	0.1431	0.0236	0.0203
Step=3	0.1592	0.1942	0.0435	0.0294
Step=5	0.1274	0.2251	0.0378	0.0280

TABLE II
COMPARISON OF $RMSE$ FOR MOBILECODE DATA FORECASTING RESULTS

To address *REQ 3)*, we compare MulA-LMRBF model, traditional RBF neural network model, ARMA model, and

Web Service	258 Web services (average)			
Method	RBF	ARMA	S-LMRBF	MulA-LMRBF
Step=1	0.1208	0.1597	0.1021	0.0688
Step=3	0.1735	0.1605	0.1659	0.1369
Step=5	0.2059	0.2172	0.1369	0.1220

TABLE III
Comparison of $RMSE$ for 258 Web services data average forecasting results

the *step* is set as 1, 3 and 5 respectively. $RMSE$ is used to analyze forecasting results.

Taking RMB data set as an example, firstly, SLMRBF single step forecasting is compared with the traditional RBF neural network and ARMA model. Then, Mul-LMRBF results of m=6 and $step$ = 5 are compared with RBF and ARMA. When step>1, RBF and ARMA both use response time and throughput as history data. Fig.6(e) and Fig.6(f) show different forecasting values of S-LMRBF and MulA-LMRBF respectively. Table 1 shows the $RMSE$ of RMB data forecasting results. After comprehensively analysing forecasting results, it is concluded that MulA-LMRBF forecasting results are better than the results of RBF and ARMA.

Table 2 gives the $RMSE$ comparison of the 300 forecasting values for MobileCode data, when the *step* is set as 1, 3, 5 respectively, and m=6. Table 3 gives the average $RMSE$ of complete data of 258 Web services in second part. By comparing the experimental results, we can conclude that MulA-LMRBF and S-LMRBF have the lowest $RMSE$ value, and MulA-LMRBF is better than S-LMRBF when *step* is multiple. Therefore, it is suggested that the accuracy of MulA-LMRBF forecasting approach is better than the traditional RBF and ARMA approaches, and MulA-LMRBF with dynamic and multi-step forecasting has high forecasting accuracy.

VI. Conclusions and suggestions for future work

In this paper, a novel QoS forecast approach MulA-LMRBF based on multivariate time series is proposed to address limitations of traditional forecasting and thereby it provides dynamic forecasting with more accurate multi-step forecasting results.

For future work, how to choose optimal parameters of reconstruction is an important problem which needs further study. We also plan to further forecast more QoS attributes [25], and also build more accurate attribute relationships.

VII. Acknowledgements

The work is supported by National Natural Science Foundation of China (No.61572171), the Key Technologies Research and Development Program of China (2015BAB07B00), and Fundamental Research Funds for the Central Universities (No.B15020191).

References

[1] A. Amin, A. Colman, and L. Grunske, "An Approach to Forecasting QoS Attributes of Web Services Based on ARIMA and GARCH Models," in *IEEE International Conference on Web Services*, 2012, pp. 74–81.

[2] A. Azzalini, M. Farge, and S. Kai, "Nonlinear wavelet thresholding: A recursive method to determine the optimal denoising threshold," *Applied and Computational Harmonic Analysis*, vol. 18, no. 2, pp. 177–185, 2005.

[3] J. Bilski, J. Smolag, and J. M. Zurada, *Parallel Approach to the Levenberg-Marquardt Learning Algorithm for Feedforward Neural Networks*. Springer International Publishing, 2015.

[4] E. Brier, C. Clavier, and F. Olivier, "Correlation Power Analysis with a Leakage Model," in *Cryptographic Hardware and Embedded Systems - CHES 2004: International Workshop Cambridge, Ma, Usa, August 11-13, 2004. Proceedings*, 2004.

[5] L. Grunske, "Specification patterns for probabilistic quality properties," in *International Conference on Software Engineering*, 2008, pp. 31–40.

[6] M. Han, *Prediction theory and method of chaotic time series*. China Water Power Press, 2007.

[7] S. S. Haykin, "Neural networks and learning machines," *Pearson Schweiz Ag*, 2009.

[8] T. H. Le, J. Cldire, C. Canovas, B. Robisson, C. Servire, and J. L. Lacoume, *A Proposition for Correlation Power Analysis Enhancement*. Springer Berlin Heidelberg, 2006.

[9] J. Lin, "Modification of Average Displacement Method for Selection of Time-delay in Phase Space Reconstruction of Speech Signals," *Journal of National University of Defense Technology*, 1999.

[10] K. F. Liu, H. Wang, and X. U. Zuo-Ping, "A Web Service Selection Mechanism Based on QoS Prediction," *Computer Technology and Development*, 2007.

[11] H. Ltkepohl, *New Introduction to Multiple Time Series Analysis*. Springer, 2005.

[12] B. M. A. Madi, Q. Z. Sheng, L. Yao, Y. Qin, and X. Wang, "PLMwsp: Probabilistic Latent Model for Web Service QoS Prediction," in *IEEE International Conference on Web Services*, 2016, pp. 623–630.

[13] Z. X. Ming and N. G. Liang, "An Improved RBF Network On-Line Learning Algorithm," in *International Symposium on Information Science and Engineering*, 2009, pp. 547–552.

[14] T. Mu and A. K. Nandi, "Detection of breast cancer using v-svm and rbf networks with self organized selection of centers," in *Medical Applications of Signal Processing, 2005. the Iee International Seminar on*, 2005, pp. 47–52.

[15] L. Shao, J. Zhang, Y. Wei, and J. Zhao, "Personalized QoS Prediction for Web Services via Collaborative Filtering," in *IEEE International Conference on Web Services*, 2007, pp. 439–446.

[16] L. Shao, L. Zhou, J. Zhao, B. Xie, and H. Mei, "Web Service QoS Forcasting Approach," *Journal of software*, no. 8, pp. 2062–2073, 2009.

[17] J. Sun, "An Original RBF Network Learning Algorithm," *Chinese Journal of Computers*, 2003.

[18] F. Takens, *Detecting strange attractors in turbulence*. Springer Berlin Heidelberg, 1981.

[19] V. K. Tiwari, A. K. Gautam, and S. Diwaker, "Web service discovery and integration with QoS parameter using SOA based repository," in *Computer Engineering and Applications*, 2015, pp. 1014–1009.

[20] X. Wang and M. Han, "Improved extreme learning machine for multivariate time series online sequential prediction ," *Engineering Applications of Artificial Intelligence*, vol. 40, pp. 28–36, 2015.

[21] X. Wang, J. Zhu, and Y. Shen, "Network-Aware QoS Prediction for Service Composition Using Geolocation," *IEEE Transactions on Services Computing*, vol. PP, no. 99, pp. 630–643, 2015.

[22] Y. Ye, S. Squartini, and F. Piazza, "Online sequential extreme learning machine in nonstationary environments," *Neurocomputing*, vol. 116, no. 10, pp. 94–101, 2012.

[23] Z. Ye, S. Mistry, A. Bouguettaya, and H. Dong, "Long-term QoS-aware cloud service composition using multivariate time series analysis," *IEEE Transactions on Services Computing*, vol. 9, no. 3, pp. 382–393, 2016.

[24] Y. Yin and P. Shang, "Forecasting traffic time series with multivariate predicting method," *Applied Mathematics and Computation*, vol. 291, pp. 266–278, 2016.

[25] P. Zhang, Q. Han, W. Li, H. Leung, and W. Song, "A novel QoS prediction approach for cloud service based on bayesian networks model," in *IEEE ICMS*, 2016, pp. 111–118.

[26] P. Zhang, Y. Sun, H. Leung, M. Xu, and W. Li, "A novel approach for QoS prediction based on bayesian combinational model," *China Communications*, vol. 13, no. 11, pp. 269–280, 2016.

[27] P. Zhang, Y. Sun, W. Li, W. Song, and H. Leung, "A combinational QoS-prediction approach based on RBF neural network," in *IEEE International Conference on Services Computing*, 2016, pp. 577–584.