Prediction of Website Response Time based on Support Vector Machine

Xiaoming Zhang[1], Cuixia Feng[1,2], Guang Wang[1,2]
[1]Department of Computer, Beijing Institute of Petrochemical Technology, Beijing, 102617, China
[2]College of Information Engineering, Beijing University of Chemical Technology, Beijing, China

Abstract—**Website response time is one of the most important performance parameter of website. It can be used to assess website performance to forecast the status of website. Large amounts of data are applied by a distributed monitoring system that monitoring a university website response time. Support vector machine with information granulation is studied to predict the response time. It can predict accurately the range of ultimate response time, the relative accuracy of the forecast average response time can reach 96.2%.**

Keywords-Website; Network measurement; Support vector machine; monitoring; Response time

I. Introduction

Response time is one of the most important performance parameter of the website. According to the response time to assess the performance of the server, not only can avoid thoughtless or bottlenecks impacting on server performance, but also can accurate assess the performance of web servers, while the response time is also conform to the user experience. When users apply the web system, the waiting time is response time. It usually depends on many factors such as the throughput, CPU utilization and network speed. How to accurate predict the response time is crucial for optimization website quality of service and improving the system performance.

The response time is an important indicator of the website performance, and it is also the most intuitive quality evaluation criteria. About the relationship of website performance analysis, a Nile program was used in [1] as a test, with Load-Runner as a performance test tool explored the relationship between website response time and the number of concurrent users. The relationship could be nearly linear usually. However, when load exceeds the scope that resources can bear, the relationship is not linear. Meanwhile, the exponential relationship was found between response time and the amount of user request before the peak of the load [2-3]. While a hospital management system under Load-Runner in [4] for performance testing, provided a quantitative analysis method based on the K-means clustering response time. About the prediction of website response time, Doyle built a model in [5] by CPU, disk I/O utilization to predict the response time of the server. But the model is only applicable to single, static pages of customer requests. Zheng [6] analyzed website conditions and predicted the performance dynamically by using historical data time series of Web services. Then, according to the characteristic pattern of data, the appropriate mathematical models are selected to predict to improve the accuracy of the forecast. About social networking website, Chen [7] measured the response time of 1289 websites, and accurately analyzed the website content and structure to predict the website response time within five minutes. The accuracy rate could reach to 86%. Besides, a distributed electronic monitoring system in [8] was developed to enable managers to monitor the form of the website and web services from remotely, but the system cannot be applied to more complex and workable website.

In summary, the above results have two shortcomings on webpage response time. The first is that the response time data is commonly accessed by using simulation tools, such as Load Runner, Silk performer and Visual studio. The monitoring response time data is too ideal to consider the reality of the network, non-linear and complexity. Secondly, because of single point measuring, the prediction is not confident and accurate. Meanwhile, the response time of the website is often non-linear, time-varying and complex sequence with interference, so the accuracy of generally prediction model is not high enough.

For these reasons, a kind of distributed measuring system is designed and applied to measure the appointed website. The source data are collected from 24 positions to monitor a university website for 84 days. Then, combining the information granulation with support vector machine, the response time of the website is analyzed and predicted.

II. Related Technologies for Prediction

A. Granulation computing

Information granulation is one of the computational intelligence approaches. The information granules form the complex information entities. Generally, there are three kinds of models for the information granulation: rough set theory, fuzzy set theory and the theory of quotient space model. The typical fuzzy particles are often in trapezoidal, triangular, Gaussian and parabolic type. Because of single cell data, the triangular fuzzy particle is chose here. The function of fuzzy membership is always shown as:

978-1-4799-5835-1/14/$31.00 ©2014 IEEE

$$I(x,a,m,b)=\begin{cases}0, x<a\\ \frac{x-a}{m-a}, a\le x\le m\\ \frac{b-x}{b-m}, m<x\le b\\ 0, x>b\end{cases} \quad (1)$$

B. Description of support vector machine (SVM)

SVM is related to statistical learning theory. SVM is now regarded as an important example of kernel methods, one of the key areas in machine learning. The basic idea of SVM is to define an optimal hyperplane firstly, and then it converts the problem how to find the optimal linear hyperplane. The optimal hyperplane is not only to ensure that separating the two kinds without error, but also require the largest classification interval. The former guarantees the empirical risk minimum value, while the latter makes the range minimum and thereby minimize the real risks.

Set the sample set of linear separable have n samples (x_i, y_i), i=1,2,…,n, x∈Rd, y∈{-1,1} as category label. Then, in high dimensional space, the hyperplane that separate the two kinds of samples without fault satisfy as following:

$$g(x)=w\cdot x-b=0 \quad (2)$$

Through normalizing the vector coefficient w, all samples are made to satisfy |g(x)|≥1. All samples separated without fault will satisfy as:

$$y_i(w\cdot x_i-b)-1\ge 0 \quad (3)$$

In high dimensional feature space, the distance is designed between the classification interval $2/\|w\|$.To make the interval biggest, let the following be the smallest:

$$\phi(w)=\frac{1}{2}\|w\|^2=\frac{1}{2}w^Tw=\frac{1}{2}(w.w) \quad (4)$$

Therefore, the optimal hyperplane is under the restriction of condition (3), and the Φ(w) will extreme value of (4).

Because the composition of information granulation and SVM are popular adopted in prediction of system status [10-11], the technology is also applied here for website operation monitoring.

III. DESIGN OF DISTRIBUTED WEBSITE MEASURMENT SYSTEM

Website measurement is used for website defect detection, performance measurement, website statistics and performance forecasting. It can be divided into single point and multi-point monitoring. Single point monitoring method is simple, but the accuracy is not high enough. Multipoint monitoring is distributed geographically with many measuring points. Therefore, the monitoring accuracy is much higher.

Meanwhile, website performance measurement can be divided into internal monitoring and external monitoring. Internal monitoring is deployed programs and scripts on the server monitoring system. It can monitor an internal error such as webpage change by webpage information hiding technology [9]. However, it cannot reflect the client's visiting states. External monitoring is adopt to measure the website performance by simulating the behavior of end users, including the availability of server memory, CPU utilization, disk reads and writes, and processes and website response time. However, this approach cannot describe the webpage’s inner features.

After simulation with OPNET tool, the distributed website measurement system is carried out by collecting a number of data from the cooperated partner. Each record includes seven data of the maximum time, minimum time, average time, DNS resolution time, connection time, server computing time and download time. The distributed website measurement system is described, as shown in Figure 1.

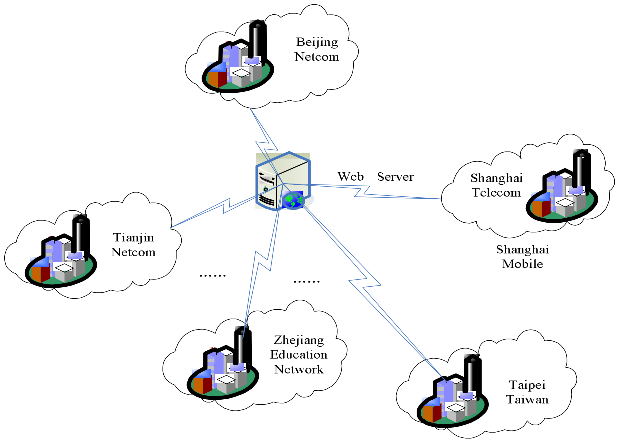

Figure 1. Description of distributed website measurement system

IV. WEBSITE RESPONSE TIME

Response time means that the external website visitors from began to send requests to receive response from the target website. The necessary condition for the target website must be reachable. Otherwise, the response time is infinite theoretically. Here, the response time is divided into four parts: DNS resolution time, connection time, server computation time, download time. The Response time is shown in Figure 2.

A. DNS resolution time T_{DNS}

T_{DNS} means time of converting website domain name to an IP address. By making one DNS communication, browser firstly sends DNS requests time t_q, and receives the effective DNS response time tr. Then,

$$T_{DNS}=t_r-t_q \quad (5)$$

DNS resolution time really depends on the performance of DNS server and network status from user to the DNS server. In addition, DNS records also may be cached on the DNS servers that Internet service provider access to, this depends TTL value of DNS records.

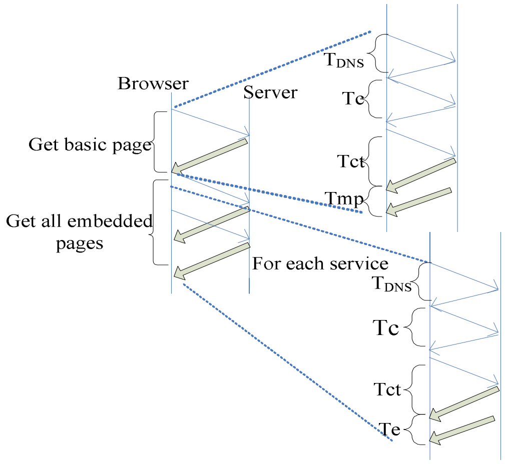

Figure 2. Process of obtaining a webpage

B. *Connection time Tc*

Tc is the browser to establish a TCP connection with the Web server. Make browser sends SYN time at t_S, the browser receives SYN ACK at t_{SA}, then,

$$T_C = t_S - t_{SA} \tag{6}$$

Establishing a connection reflects the ability that users can access the Web server quickly.

C. *Server computation time T_{ct}*

Server computation time includes computing static and dynamic files types.

$$T_{ct} = t_{cf} - t_{cb} \tag{7}$$

where, t_{cb} is the server start receives the connection request, and t_{cf} is the server accomplish computation.

D. *Download time*

Download time includes the main page download time T_{mp} and embedded file download time T_e. The browser sends the request to the main page at t_q, and receives the last message of the main web page t_l, then

$$T_{mp} = t_l - t_q \tag{8}$$

The download time for embedded file T_e is that the user sees the embedded documents of web page. The browser can download multiple embedded file parallel. The download time is mutually overlapping. The browser sends DNS request or connection request for the first embedded file at t_f , receives the last message of embedded document at t_l, then

$$T_e = t_l - t_f \tag{9}$$

The download time depends on the bandwidth from monitoring points to website server, it does not mean the users real download time.

The total response time of T_t is equal to the sum of each phase of the response time:

$$T_t = T_{DNS} + T_C + T_{ct} + T_{mp} + T_e \tag{10}$$

V. DESIGN OF PREDITION MODEL

The prediction model is processed by methods of information granulation and support vector machine, as shown in Figure 3.

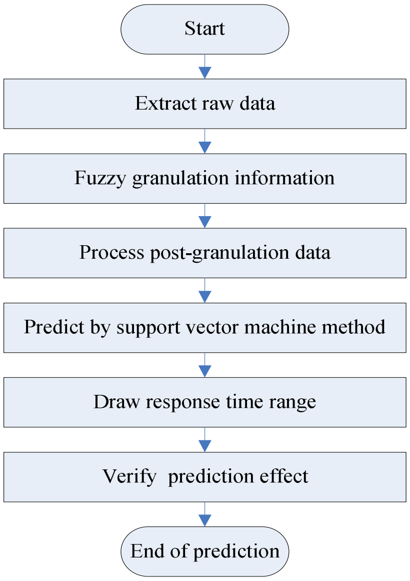
Start
Extract raw data
Fuzzy granulation information
Process post-granulation data
Predict by support vector machine method
Draw response time range
Verify prediction effect
End of prediction

Figure 3. Description of prediction process model

A number of response time data from the 24 monitoring regions can be obtained for the web server. Then, combining two technologies of the information granulation and support vector machine, the LIBSVM toolkit developed by the National Taiwan University is adopt to predict the response time of website.

In order to obtain the ideal prediction accuracy, it is required to adjust two parameters of kernel function g and the penalty parameter c by using SVM predict. The cross-validation approach is applied to finding the best parameters of c and g under no test set labels.

VI. EXPERIMENTAL ANALYSIS

Firstly, the input data is trained by triangular fuzzy particle that belongs to fuzzy set theory and support vector machine model. It can obtain the optimal solution which is the optimal parameter of support vector machines by cross-validation. Then, the original response time data is converted by 24-hour time as a window, and finally three fuzzy particles of the website response time is extracted ultimately. Fuzzy particles which have three parameters of Low, R and Up, represent the minimum, average and maximum of response time respectively. The original sequence diagram of website response time is drawn in Figure 4, while the fuzzy information granulation diagram is as shown in Figure 5.

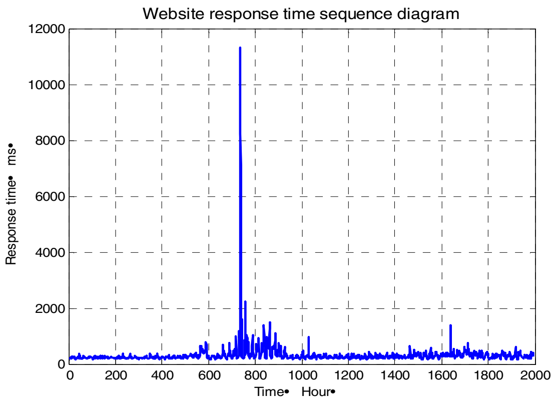

Figure 4. Sequence of website response time

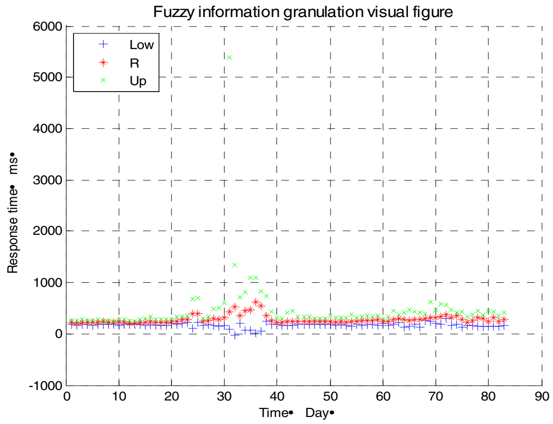

Figure 5. Fuzzy information granulation of website response time

Choosing optimal parameters of Low, R and Up respectively after granulation, parameter c and g are selected when the average error is the smallest. In order to avoid over-learning situation occurring, if more than one set of corresponding value appears, the smallest c corresponds to g is chosen. If there are more than one g corresponding with one c, the first search is used as the best combination. Therefore, for the optimum parameter of Low, the following results are obtained:

Conditions: c = 256, g = 1.41421;
Prediction error = 2.41132;
Corresponding regression coefficient (RC) = 0.999436;
Predictive value for the Low =168.6875.

The predict response time corresponding to the minimum is shown in Figure 6， and the prediction error is shown in Figure 7. For value of R, the prediction results are as follows:

Conditions: c=256, g=0.353553;
MSE=46091.7, RC=0.357447;
Prediction results = 324.6906.

Similarly, the predict response time corresponding to the average and the prediction error are shown in Figure 8 and Figure 9 independently.

Similarly, for the value of Up:
Condition: c=256, g=0.25;
MSE=1064.42, RC=0.618938;
Prediction results = 492.5728.

The predict response time corresponding to the Maximum is shown in Figure 10, and the prediction error is shown in Figure 11.

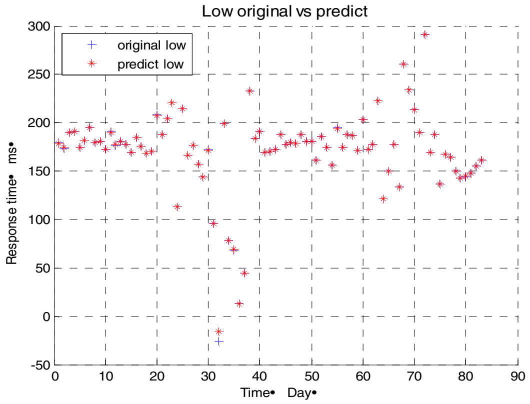

Figure 6. Minimum of original values and predicted values

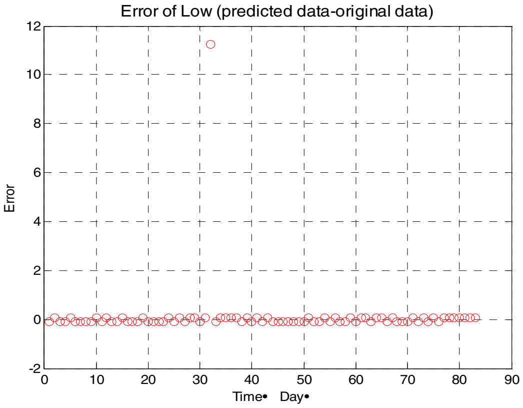

Figure 7.Error of original values and predicted values of minimum

The actual scope of website response time is [160.55, 311.9375, 568.17] representing the minimum, average and maximum. The predict scope is [168.6875, 324.6906, 492.5728]. It shows that the error of predicted results and actual results is very small.

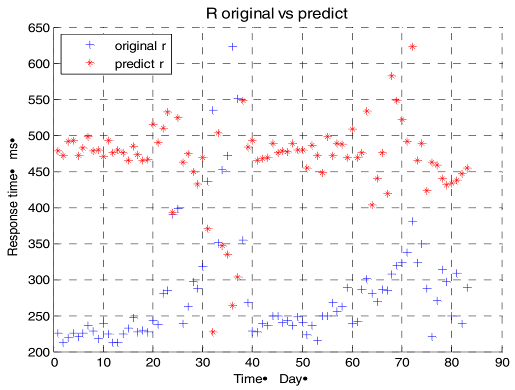

Figure 8. Average of original values and predicted values

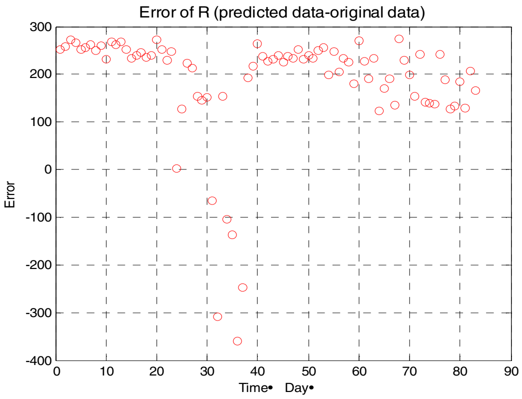

Figure 9. Error of original values and predicted values of average

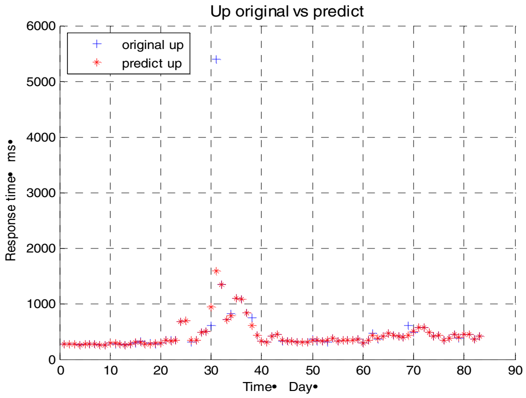

Figure 10. Maximum of original values and predicted values

The forecasting comparison charts by single-point and distributed measuring approaches are shown in Figure 12, Figure 13 and Figure 14.

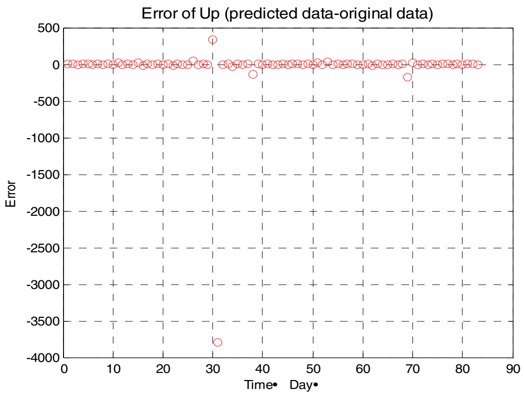

Figure 11. Error of original values and predicted values of maximum

As shown in Figure 12 to Figure 14, there are 24 single-point measuring values colored as green, yellow and red independently. There exist tow purple results in each of the three figures. The left purple value represents actual response time, and the right purple value represents predicted value by the distributed monitoring system.

From the above comparison of the three groups, the prediction error by single point measurement data is large, and the accuracy rate is low. However, the prediction accuracy by multi-monitoring data is high. The predicting relative accuracy rate of the minimum can reach 95%, the mean relative accuracy rate can reach 96.2%, and the maximum relative accuracy rate can reach 87%. Therefore, the prediction effect is satisfying.

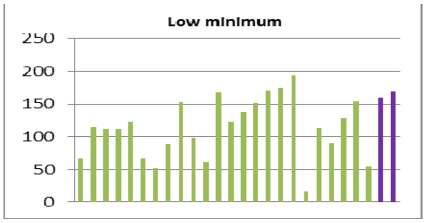

Figure 12. Minimum prediction by single-point and distributed

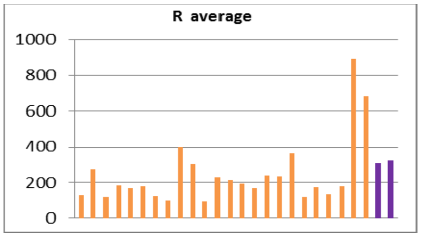

Figure 13. Average prediction by single-point and distributed

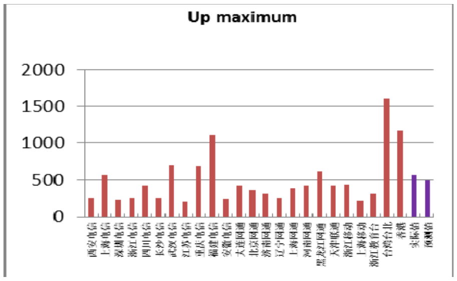

Figure 14. Maximum prediction by single-point and distributed

VII. CONCLUSION

A kind of distributed measuring system is designed and implemented to predict the response time of website. Approaches combining fuzzy information granulation and support vector machine are adopted successfully. The prediction method is easy to use, long-term forecasting, and it forecast accuracy is high.

REFERENCES

[1] S. Liang.S.M.Li. Study of the response time about Web applications. Journal of Computer Research and Development.2003,40(7):1076-1080

[2] X.H.Xu,T.T. Xu. The Web server performance evaluation method based on response time. Journal of Chinese Computer Systems..2013.(1)

[3] T.T.Xu. The Web server performance evaluation model based on response time.Hangzhou dianzi university.2010.12.

[4] Y.Hua,X.Jiang,Z.Y.Liu,et al. Response time analysis method based on clustering. Computer Applications and Software.2012.8.

[5] R.P.Doyle, J.s.Chase, O.M.Asad, et al. Model-based resource provisioning in a web service utility. Proc. 4th Int. Conf. on USE-NIX Symposium on Internet Techologies and systems,2003.

[6] X.X.Zheng, J.F.Zhao, Z.W.Cheng, et al. A dynamic prediction method of WebService response time. Journal of Chinese Computer Systems.2011.8, (8).

[7] C.Liang, S.Hiremagalore, A.Stavrou, et al. Predicting network response times using social information. Proc. Int. Conf. on Advances in Social Networks Analysis and Mining, 2011.

[8] K.W. Frank, K.W. Cheong.C.Dickson,et al. Developing a distributed e-monitoring system for enterprise website and web services: an experience report with free libraries and tools. Proc. IEEE Int. Conf. on Web Services,2007.

[9] X.M.Zhang, G.Q.Zhao, P.F.Niu. A novel approach of secret hiding in webpage by bit grouping technology. Journal of Software, 2012,4(11)

[10] G.F.He, M.Yang, X.D.Gu,et al. A novel active website fingerprinting attack against Tor anonymous System. Proc. IEEE 18th Int. Conf. on Computer Supported Cooperative Work in Design, 2014.

[11] J.Yi, J.Peng, T.F. Li. Integration of fuzzy information granulation and support vector machine for prediction alumina concentration. Proc. 11th Int. Conf. on Cognitive Informatics & Cognitive Computing ,2012