

# Internet Traffic Prediction by W-Boost: Classification and Regression\*

Hanghang Tong<sup>1</sup>, Chongrong Li<sup>2</sup>, Jingrui He<sup>1</sup>, and Yang Chen<sup>1</sup>

<sup>1</sup> Department of Automation, Tsinghua University, Beijing 100084, China  
{walkstar98, hejingrui98}@mails.tsinghua.edu.cn

<sup>2</sup> Network Research Center of Tsinghua University, Beijing 100084, China  
licr@cernet.edu.cn

**Abstract.** Internet traffic prediction plays a fundamental role in network design, management, control, and optimization. The self-similar and non-linear nature of network traffic makes highly accurate prediction difficult. In this paper, we proposed a new boosting scheme, namely W-Boost, for traffic prediction from two perspectives: classification and regression. To capture the non-linearity of the traffic while introducing low complexity into the algorithm, ‘stump’ and piece-wise-constant function are adopted as weak learners for classification and regression, respectively. Furthermore, a new weight update scheme is proposed to take the advantage of the correlation information within the traffic for both models. Experimental results on real network traffic which exhibits both self-similarity and non-linearity demonstrate the effectiveness of the proposed W-Boost.

## 1 Introduction

Internet traffic prediction plays a fundamental role in network design, management, control, and optimization [12]. Essentially, the statistics of network traffic itself determines the predictability of network traffic [2], [12]. Two of the most important discoveries of the statistics of Internet traffic over the last ten years are that Internet traffic exhibits self-similarity (in many situations, also referred as long-range dependence) and non-linearity. Since Will E. Leland’s initiative work in 1993, many researchers have dedicated themselves to proving that Internet traffic is self-similar [10]. On the other hand, Hansegawa et al in [6] demonstrated that Internet traffic is non-linear by using surrogate method [16]. The discovery of self-similarity and non-linearity of network traffic has brought challenges to traffic prediction [12].

In the past several decades, many methods have been proposed for network traffic prediction. To deal with the self-similar nature of network traffic, the authors in [15] proposed using FARIMA since FARIMA is a behavior model for self-similar time series [4]; the authors in [19] proposed predicting in wavelet domain since wavelet is a natural way to describe the multi-scale characteristic of self-similarity. While these methods do improve the performance of prediction for self-similar time series, they are both time-consuming. To deal with the non-linear nature of network traffic, Arti-

---

\* This work is supported by National Fundamental Research Develop (973) under the contract 2003CB314805.

ficial Neural Network (ANN) is probably the most popular method. ANN can capture any kind of relationship between the output and the input theoretically [6], [9], [11], however, it might suffer from over-fitting [5]. Another kind of prediction method for non-linear time series is support vector regression (SVR) [11] which is based on structural risk minimization. However, the selection of suitable kernel functions and optimal parameters might be very difficult [6].

In our previous work [17], [18], we have introduced boosting technique into traffic prediction by considering it as a regression problem. The initial experimental results demonstrated the prediction performance by Feed-Forward Neural Network (FFNN) can be largely improved by boosting technique. However, some very important issues leave to be solved: 1) Which model, classification or regression, is more suitable for traffic prediction? 2) How to design weak learners? 3) How to update weight distribution? In this paper, all these aspects are investigated under a new boosting scheme, namely W-Boost. Firstly, network traffic prediction is modeled from two perspectives: both classification and regression are considered. To capture the non-linearity of the traffic while introducing low complexity into the algorithm, ‘stump’ and piecewise-constant function are adopted as weak learners for classification and regression, respectively. Furthermore, a new weight update scheme is proposed for both models, which aims to maximally take the advantage of the correlation information within the traffic. Experimental results on real network traffic which exhibits both self-similarity and non-linearity demonstrate the effectiveness of W-Boost.

The rest of this paper is organized as follows: in Section 2, the proposed W-Boost for traffic prediction is presented in detail. Experimental results are given in Section 3; finally, we conclude the paper in Section 4.

## 2 W-Boost for Traffic Prediction

As a general method for improving the accuracy of some weak learning algorithms, boosting has been shown to be very effective for classification [1], [8], [14]. When applied to traffic prediction, there are mainly two classes of methods: 1) modifying some specific steps of the existing boosting algorithms for classification so that they are suitable for a regression problem [1], [8]; 2) modeling traffic prediction as a classification problem by introducing an additional variable [13], [14]. Both methods will be investigated in the proposed W-Boost: on the whole, W-Boost shares the similar flow-chart of Ada-Boost as well as Real-AdaBoost. In this section, after giving out the problem definition, we will discuss the details of two key points of W-Boost, which make it different from Both Ada-Boost and Real-AdaBoost: weak learner design and weight update scheme.

### 2.1 Problem Definition

To predict network traffic, it is generally denoted as a time series:  $X = (x_i : i = 0, 1, 2, \dots)$ . The prediction problem can be defined as follows [2]: given the current and past observed values  $X_i = (x_{i-p+1}, \dots, x_{i-1}, x_i)$ , predict the future value  $x_{i+q}$ , where  $p$  is the length of history data used for prediction and  $q$  is the prediction step.

**Prediction as Regression:** In practice, the traffic prediction problem can be considered as a classical regression problem under the assumption that  $x_i \in \zeta_2$  ( $i = 1, 2, \dots$ ) [2]. That is, the prediction of  $x_{t+q}$  can be written as:

$$\hat{x}_{t+q} = \underset{y \in \zeta_2}{\operatorname{argmin}} E\{(y - x_{t+q})^2\}. \quad (2)$$

**Prediction as Classification:** As in [13], traffic prediction can also be modeled as a classification problem by introducing an additional reference variable  $S$ :

$$\hat{x}_{t+q} = \inf_s \{s : P(Y^* = 1 \mid \{x_{t-p}, \dots, x_t\}, S) \geq 1/2\}. \quad (3)$$

Where  $S$  is the reference variable, and  $Y = \{0, 1\}$  is the output defined on  $(\{x_{t-p}, \dots, x_t\}, S)$  (For details, refer to [13]).

## 2.2 Weak Learner Design

Like Real-AdaBoost [3], in W-Boost, a weak learner is trained for each feature  $x_{t-i}$  ( $i = 0, \dots, p$ ) and the one with the minimum training error is selected in each iteration. A good weak learner design should consider two factors: 1) to capture the non-linearity within network traffic, the weak learner itself should be non-linear; 2) on the other hand, to achieve good generation performance of the final combined learner, such weak learner should not be too ‘complex’. Basted on the above observation, the weak learner in W-Boost is designed as follows:

**Weak Learner in Classification:** In this case, “stumps” is adopted as the weak classifier. “Stumps” are single-split trees with only two terminal nodes [3].

**Weak Learner in Regression:** In this case, the piece-wise-constant function (PWC) [16] is adopted as the weak regressor.

## 2.3 Weight Update Scheme

How to exploit the correlation structure is the key problem in traffic prediction [12]. In [17], we proposed using PCA as a preprocessing step to take advantage of self-similar nature of traffic while avoiding the disadvantage of self-similarity. In W-Boost, by using each feature to training a weak learner, we could make use of the correlation structure of traffic in a more sophisticated way, which is motivated by our previous work in image classification [7]. Its key point is that, in W-Boost, a set of weights are kept on all examples for each feature component to generate the weak learner. Let  $w(i), w(j)$  ( $i, j = 0, \dots, p; i \neq j$ ) denote the present sets of weights for feature component  $x_{t-i}$  and  $x_{t-j}$ , respectively. Let  $w'(i), w'(j)$  denote the updated sets of weights. In W-Boost, they are updated as follows:

1. Suppose  $x_{t-i}$  ( $i = 0, \dots, p$ ) is selected to generate the weak learner in the current iteration. Update  $w(i)$  as Ada-Boost (Real-AdaBoost):  $w(i) \xrightarrow[\text{Real-AdaBoost}]{\text{Ada-Boost}} w'(i)$ ;
2. For all  $x_{t-j}$  ( $j = 0, \dots, p; j \neq i$ ), update  $w(j)$  as Eq. 4:

$$w'(j) = (1 - \alpha_j)w'(i) + \alpha_j w(j). \quad (4)$$

where  $\alpha_{ij} \in [0,1]$  is the parameter, indicating the dependence extent between  $x_{t-i}$  and  $x_{t-j}$ : if they are independent,  $\alpha_{ij} = 1$ ; else if they are totally dependent,  $\alpha_{ij} = 0$ . In [7],  $\alpha_{ij}$  is estimated by the Kullback-Leibler distance between  $x_{t-i}$  and  $x_{t-j}$ . For traffic prediction, it can also be approximated by the auto-correlation function (ACF) of the time-series  $X_i = (x_{i-p+1}, \dots, x_{i-1}, x_i)$ , i.e.  $\alpha_{ij} \approx 1 - ACF(|i - j|)$ .

### 3 Experimental Results

The network traffic that we use is the JPEG and MPEG version of the “Star Wars” video which is widely used to examine the performance of the network traffic prediction algorithms [9]. In our experiment, we divide the original traffic into some traces with equal length 1000. Then we make use of variance-time [10] and surrogate method [16] to test self-similarity and non-linearity of a given trace, respectively. Those traces which exhibit both self-similarity and non-linearity are selected to examine the performance of BBF-PT. Furthermore, each trace is normalized to  $[0,1]$  for comparison simplicity.

There are a set of parameters and operations that need to be set in W-Boost:

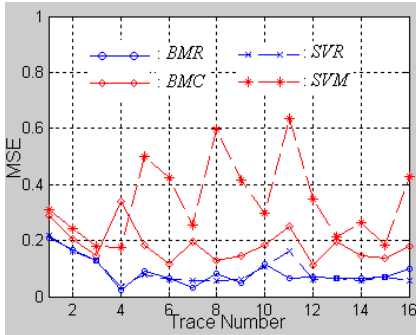
- ◆ The length of history data  $p$  used for predicting is set to be 10;
- ◆ At the current stage, we are concerned with one-step prediction so  $q$  is 1;
- ◆ We used the first 50% data in each trace to compose the training set;
- ◆ The length of the reference variable  $S$  in classification model is set to be 20;
- ◆ For both classification and regression models, the maximum iteration number and the bin number are determined by cross-validation using the training data;
- ◆ The mean-square-error (MSE) is used to evaluate different methods.

First, both classification model by W-Boost (CMW) and regression model by W-Boost (RMW) are evaluated. The results are compared with those by Support Vector Machine (SVM) and Support Vector Regression (SVR), respectively. It can be seen from Figure 1 that 1) for both CMW and RMW, W-Boost results in better or comparable performance with those by SVM and SVR, respectively; 2) both SVR and RMW outperform SVM and CMW, indicating that for traffic prediction, regression might be more suitable. Then, the weak learner design and weight update scheme in W-Boost are evaluated. In this case, we only present the result by regression model for the limited space. The result are compared with that by what is proposed in our previous work [17], where Feed-forward Neural Network is adopted as the basic weak regressor and Principle Component Analysis is used as a preprocess step (FFNN+PCA). From Figure 2, it can be seen that W-Boost can further boost the prediction performance. Finally, two prediction examples are presented in Figure 3.

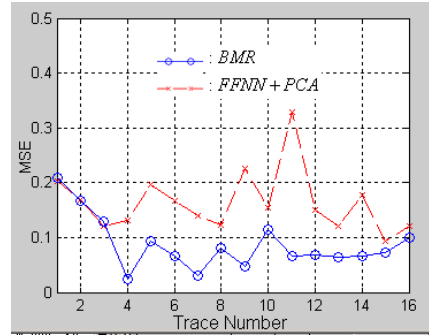
### 4 Conclusions

In this paper, we have proposed a new boosting version, namely W-Boost for self-similar and non-linear network traffic prediction. On the whole, W-Boost shares the similar flow-chart of both Ada-Boost and Real-AdaBoost. W-Boost supports both regression and classification methods for traffic prediction. To capture the non-

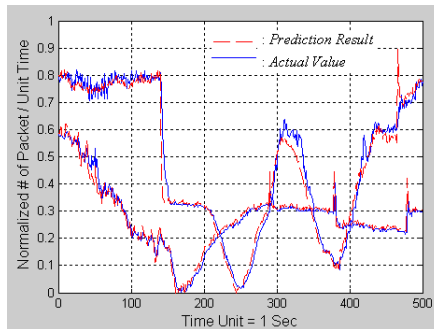
linearity of the traffic while introducing low complexity into the algorithm, “stumps” and piece-wise-constant function are adopted as weak learners for classification and regression, respectively. Furthermore, a new weight update scheme is proposed for both models, which aims to maximally take the advantage of the correlation information within the traffic. Experimental results demonstrate that 1) the regression model is more effective for traffic prediction; and 2) both the proposed weaker learner and weight update scheme are effective.



**Fig. 1.** Evaluation of different models



**Fig. 2.** Evaluation of weak learner design and weight update scheme in W-Boost



**Fig. 3.** Prediction example using regression by W-Boost

## References

1. Bone, R., Assaad, M., Crucianu, M.: Boosting Recurrent Neural Networks for Time Series Prediction. Proceedings of the International Conference on Artificial Neural Networks and Genetic Algorithms (2003)
2. Brockwell, P., Davis, R.: Time Series: Theory and Methods. Springer-Verlag, New York, 2nd edition (1991)
3. Friedman, J., Hastie, T., Tibshirani, R.: Additive Logistic Regression: A Statistical View of Boosting. The Annual of Statistics, **28** (2000) 337-374
4. Gripenberg, G., Norros, I.: On the Prediction of Fractional Brownian Motion. Journal of Applied Probability, (1996) 400-410

5. Hall, J., Mars, P.: The Limitations of Artificial Neural Networks for Traffic Prediction. IEE Proceedings on Communications (2000) 114-118
6. Hansegawa, M., Wu, G., Mizuno, M.: Applications of Nonlinear Prediction Methods to the Internet Traffic. The 2001 IEEE International Symposium on Circuits and Systems, (2001) 169-172
7. He, J., Li, M., Zhang, H.J., Zhang, C.: W-Boost and Its Application to Web Image Classification. IEEE Int. Conf. on Pattern Recognition (2004) 148-151
8. Kegl, B.: Robust Regression by Boosting the Median. COLT/Kernel, (2003) 258-272
9. Khotanzad, P., Sadek, N.: Multi-Scale High-Speed Network Traffic Prediction Using Combination of Neural Network. IJCNN (2003) 1071-1075
10. Leland, W.E., Taqqu, M.S., Willinger, W., Wilson, D.: On the Self-Similar Nature of Ethernet Traffic. IEEE/ACM Tran. on Networking (1994) 1-15
11. Muller, K.: Predicting Time Series with Support Vector Machines. Proceedings of the International Conference on Artificial Neural Network (1997) 999-1004
12. Ostring, S., Sirisena, H.: The Influence of Long-rang Dependence on Traffic Prediction. IEEE ICC, (2001) 1000-1005
13. Ridgeway, D., Madigan, D., Richardson, T.: Boosting Methodology for Regression Problem. Proc. 7th Int. Workshop on Artificial Intelligence and Statistics (1999) 152-161
14. Schapire, R.E.: The Boosting Approach to Machine Learning: an Overview. MSRI Workshop on Nonlinear Estimation and Classification (2002)
15. Shu, Y., Jin, Z., Zhang, L., Wang, L.: Traffic Prediction Using FARIMA Models. IEEE ICC (1999) 891-895
16. Small, M., Yu, D., Harrison, R.G.: Surrogate Test for Pseudoperiodic Time Series Data. Physical Review Letter (2001)
17. Tong, H., Li, C., He, J.: A Boosting-Based Framework for Self-Similar and Non-linear Internet Traffic Prediction. ISNN (2004) 931-936
18. Tong, H., Li, C., He, J.: Boosting Feed-Forward Neural Network for Internet Traffic Prediction. ICMLC (2004)
19. Wang, X., Shan, X.: A Wavelet-based Method to Predict Internet traffic. IEEE International Conference on Communications, Circuits and Systems and West Sino Expositions (2002) 690-694