A Novel Network Traffic Prediction Method Based on RMPM

To cite this article: Jianjun Wu et al 2019 J. Phys.: Conf. Ser. 1267 012067

View the article online for updates and enhancements.
A Novel Network Traffic Prediction Method Based on RMPM

Jianjun Wu¹, a, Weijun Gong¹, b and Zhen Shang¹, c

¹College of Information Technology and Communication, Hexi University, Zhangye, 734000, China.

awujj@hxu.edu.cn; bgongwj@hxu.edu.cn; cshangzh@hxu.edu.cn

Abstract. The network traffic prediction plays the key role in congestion control and bandwidth allocation. A variety of traditional learning models such as artificial neural networks (ANN) have been applied in prediction. To avoid the drawbacks of traditional models for prediction, a novel robust minimax probability machine (RMPM)-based traffic prediction method is proposed in this paper. The prediction performance is tested on two different types of traffic data, Ethernet data flow and MPEG4 video flow, at the timescale 1. The experiments demonstrate that the proposed method attains satisfactory performance in prediction accuracy. Therefore, the proposed method can be used for congestion control or bandwidth allocation, to meet the user QOS requirements.

1. Introduction
The network data streams exhibit uncertainty and time-varying, and it is difficult to capture the rules of change. Through network traffic prediction and other means, it can predict the size of data traffic in the future, which is of great practical significance for network planning and design, performance analysis and network management [1].

In order to improve the prediction accuracy, a nonlinear model such as artificial neural network (ANN) is commonly used for network traffic prediction [2]. However, the artificial neural network has problems such as overfitting, too large sample size and too small local values, which limits its application. In addition to neural networks, other machine learning models are also used for network traffic prediction. For example, Hong Fei proposed a multi-scale network traffic prediction model based on wavelet, which can achieve multi-scale accurate prediction [3]. Zhang Yinglu proposed a network traffic prediction model based on support vector machine, and support vector machine relied on genetic algorithm to optimize the parameters. The experimental results show that the prediction accuracy is higher than that of traditional neural network [4]. In addition, Dai Yue et. al. researched the chaotic characteristics of network traffic sequence, and then realized network traffic prediction by means of phase space reconstruction theory [5].

The minimax probability machine (hereinafter referred to as MPM) is a new learning model that has emerged in recent years. It achieves classification [6] and regression prediction by minimizing the upper bound of the generation error [7] and other functions. Since the minimum and maximum probability machine has good characteristics, the model has been successfully used in medical diagnosis and face recognition [8]. However, some parameters in the MPM, such as the mean vector and the covariance matrix, are estimated by learning samples. When the number of samples is too small, the estimation of the mean vector and the covariance matrix will bring a large error, which will affect the accuracy of the classification. To this end, C. H. Hoi and Lanckriet proposed a robust
minimax probability machine, named RMPM to solve this problem [9-10].

This paper is based on the good characteristics of RMPM and using it for network traffic prediction. Firstly, we give an accurate description of the network traffic prediction problem. Then the basic principles of MPM and RMPM are briefly reviewed. Finally, the prediction methods proposed in this paper are used to predict the two types of network traffic such as Ethernet data stream Ethernet and MPEG4 video stream. The prediction results verify the effectiveness of the proposed method.

2. Network traffic prediction problem
Suppose there is a piece of network traffic data \( s(n) \), it can be seen as a time series. Based on past several observations \( s(n-1), s(n-2),...,s(n-d) \) of the time series, and it can predict the current value \( s(n) \) with a predictive model. The corresponding mathematical formula is

\[
s(n) = f[s(n-1), s(n-2),...,s(n-d)]
\]  

The parameter \( d \) is called the embedding dimension, and a function \( f \) is a mapping from a \( d \)-dimensional space to a one dimensional space.

Assume a learning sample \( x_n = (s(n-1), s(n-2),...,s(n-d))^T \) ( \( n = 1,2,\cdots,N \) ) and corresponding target value \( t_n = s(n) \), Eq. (1) can be abbreviated as

\[
t_n = f(x_n), \quad n = 1,2,\cdots,N
\]  

It can be seen from Eq. (2) that the main task of the prediction model is to obtain \( N \) learning samples based on the \( d \) dimension and the corresponding target value estimation function \( f(x) \).

3. Robust MiniMax Probability Machine
Given \( N \) learning samples \( x_1, x_2,\cdots,x_N \) and their corresponding target values \( t_1, t_2,\cdots,t_N \), Now write the function \( f(x) \) to be solved

\[
f(x) = w^T x + b = w^{(1)} x^{(1)} + w^{(2)} x^{(2)} + \cdots + w^{(d)} x^{(d)} + b
\]  

Where \( w = (w^{(1)}, w^{(2)},\cdots,w^{(d)})^T \) is the weight vector. The vector \( x = (x^{(1)}, x^{(2)},\cdots,x^{(d)})^T \) is the total of \( n \) learning samples. The parameter \( b \) is a constant. In addition, the total number of \( t_1,t_2,\cdots,t_N \) is \( t \), It's a one-dimensional variable.

In order to get \( w \) and \( b \) in Eq. (3), we let

\[
\min \left[ \max P(\|f(x) - t\| \geq \varepsilon) \right] \quad \text{or} \quad \min \left[ \sup P(\|f(x) - t\| \geq \varepsilon) \right]
\]  

The minimax decision rules is used in Above Eq.(4). Now rewrite it into the following equivalent form, it can clarify its physical meaning better.

\[
\max \left[ \min P(\|f(x) - t\| \leq \varepsilon) \right] \quad \text{or} \quad \max \left[ \inf P(\|f(x) - t\| \leq \varepsilon) \right]
\]  

Suppose there is a tube with a radius of \( \varepsilon \) (referred to as \( \varepsilon \) tubes), it is obtained by shifting the learning samples by \( \varepsilon \) in the positive and negative directions of the \( t \)-axis. According to \( \|f(x) - t\| \leq \varepsilon \), \( t - \varepsilon \leq f(x) \leq t + \varepsilon \) is obtained (for the \( n \)th sample, there is \( t_n - \varepsilon \leq f(x_n) \leq t_n + \varepsilon \)). If \( P_\varepsilon = P(\|f(x) - t\| \leq \varepsilon) \), then the physical meaning of \( P_\varepsilon \) is the probability that the function \( f(x) \) is located in the \( \varepsilon \)-tube. The physical meaning of \( \inf P(\|f(x) - t\| \leq \varepsilon) \) in Eq.(5) is the lower bound of the
probability that \( f(x) \) is located in the \( \epsilon \)-tube, and the whole Eq. (5) is to maximize the lower bound, which indicates that the function \( f(x) \) is as far as possible in the \( \epsilon \)-tube.

If the variable \( \alpha \) is introduced, the Eq. (5) can also be rewritten into the following constraint optimization form

\[
\begin{align*}
\max & \quad \alpha \\
\text{s.t.} & \quad \inf P(\|f(x) - t\| \leq \epsilon) \geq \alpha
\end{align*}
\]

(6)

The physical meaning of \( \alpha \) in the above equation is the same as the physical meaning of \( \inf P(\|f(x) - t\| \leq \epsilon) \), so the parameter \( \alpha \) is the lower bound of \( P_\epsilon \), and satisfying \( \alpha \leq P_\epsilon \).

It is very difficult to solve Eq. (5) or Eq. (6) directly. Now we will convert \( f(x) \) to a classification problem and then solve it \cite{4}. Specifically, it is utilized

\[
\begin{align*}
\min & \quad \|R_{xx}^{1/2}a\|_2 + \|R_{yy}^{1/2}a\|_2 \\
\text{s.t.} & \quad a^T(\tilde{\mu}_x - \tilde{\mu}_y) = 1
\end{align*}
\]

(7)

and

\[
b = a^T\mu_x - \kappa\sqrt{a^TR_{xx}a} = a^T\mu_y + \kappa\sqrt{a^TR_{yy}a}
\]

(8)

Firstly we calculate the parameters \( a = (a^{(0)}, a^{(1)}, \ldots, a^{(d-1)})^T \) and \( b \), and then substituting the parameters \( a \) and \( b \) into Eq. (9)

\[
\begin{align*}
\omega^{(i)} & = -\alpha^{(i+1)}/\alpha^{(i)} \\
b^{(i)} & = b/\alpha^{(i)}
\end{align*}
\]

(9)

To get \( w \) and \( b \). Where \( \tilde{\mu}_x, \tilde{\mu}_y, R_{xx}, R_{yy} \) in Eq. (7) and Eq. (8) is the mean and covariance matrix of the two types of samples respectively.

In order to improve the robustness of estimates \( \mu_x, \mu_y, R_{xx} \) and \( R_{yy} \) based on the learning samples, it is assumed that the range of values of these parameters is known, and the value is allowed to fluctuate within the range of values. So define the set

\[
S_x = \left\{ (\mu_x, R_{xx}) : (\mu_x - \tilde{\mu}_x)^T R_{xx}^{-1} (\mu_x - \tilde{\mu}_x) \leq \nu^2, \|R_{xx} - \Sigma_{xx}\|_{FRO} \leq \rho \right\}
\]

(10)

\[
S_y = \left\{ (\mu_y, R_{yy}) : (\mu_y - \tilde{\mu}_y)^T R_{yy}^{-1} (\mu_y - \tilde{\mu}_y) \leq \nu^2, \|R_{yy} - \Sigma_{yy}\|_{FRO} \leq \rho \right\}
\]

(11)

In the Eq. (10) and Eq. (11) above, \( \|\cdot\|_{FRO} \) represents the Frobenius norm, It satisfies the

\[
\|A\|_{FRO}^2 = Tr\left(A^T A\right), \nu \geq 0 \quad \text{and} \quad \rho \geq 0 \text{ are two constants.}
\]

4. Experiments

The two network data streams used for prediction in the experiment are Ethernet and MPEG4. Among them, Ethernet is the test data collected by the research institute named Bellcore Morristontown, and MPEG4 is the video data stream of the medium-quality version of "Star Wars 4". The data stream above is first processed as the number of bytes passed per unit time, and then aggregated data having a
time scale of 1 second and 5 seconds, respectively, is formed, and only aggregated data of a scale of 1 second is used herein.

4.1. Ethernet network traffic data prediction
Before predicting Ethernet network traffic, the shilling embedded dimension $d$ is equal to 3, and then 200 learning samples and 150 test samples are constructed based on the traffic data. When using RMPM prediction, the parameter $\varepsilon$ contained in RMPM is taken as 0.2, and the mean square error (MSE) is used to measure the prediction performance.

The predicted results are shown in Figure 1. Figure 1 is divided into upper and lower subgraphs. The solid line in the above figure represents the true value of Ethernet network traffic, and the dotted line represents the predicted value of network traffic. The solid line in the figure below represents the prediction error. As can be seen from Figure 1, the predicted curve of the Ethernet network traffic is well matched to the actual curve. The calculation results show that the mean square error (MSE) of this prediction is $5.307 \times 10^{-3}$.

![Figure 1](image1.png)

**Figure 1.** True value Ethernet network traffic data, the predicted value and the prediction error

4.2. MPEG4 network traffic data prediction
When using RMPM to predict MPEG4 network traffic, the parameter value is the same as the previous experiment, that is $d = 3$. The number of learning samples and the number of test samples are still 200 and 150 respectively, but the parameter $\varepsilon$ is changed to 0.4. The predicted results are shown in Figure 2. Similar to Figure 1, the Figure 2 is also divided into upper and lower subgraphs. The solid line in the upper graph represents the true value of MPEG4 network traffic, and the dashed line represents the predicted value of network traffic. The solid line in the following figure represents the prediction error. It can also be seen from Figure 2 that the prediction curve of MPEG4 network traffic matches well with the actual curve. The calculation results show that the mean square error of this prediction is...
3.182×10³, which means that the prediction accuracy of MPEG4 network traffic is higher than that of Ethernet.

Figure 2. True value, predicted value and prediction error of MPEG4 network traffic data

5. Conclusions
Network traffic prediction is a challenging task, and traditional learning models have achieved some results in prediction. In order to further improve the prediction accuracy, this paper proposes a new network traffic prediction method based on the robust minimax probability machine (RMPM), and uses this method to predict the network traffic data such as Ethernet and MPEG4. The prediction results verify the superiority of the method. In the next step, we will focus on the local self-similarity research of the network traffic, in order to construct a more accurate network traffic prediction method based on the essential characteristics of network traffic data.

References
[1] Paxson, V. and Floyd, S, 1995, Wide-Area Traffic: the Failure of Poisson Modeling, IEEE/ACM Transactions on Networking, vol. 3, pp.226-244.
[2] Xin Yao, Manfred Fischer and Gavin Brown, 2001 Neural network ensembles and their application to traffic flow prediction in telecommunication networks, IJCNN’ 2001, vol.1 pp.693-698.
[3] Hong Fei, and Wu Zhimei, 2006, Multiscale Network Traffic Prediction Model Based on Wavelet, Journal of Computer Science, pp. 88-91.
[4] Zhang Yinglu, 2008, Network traffic forecasting based on genetic algorithm optimization support vector machine, Computer Science, Vol. 35, pp.177-179.
[5] Dai Yue, 2008, Research on chaotic characteristics of network traffic and prediction algorithm of network traffic. Jiangnan University, Master's Degree Thesis.
[6] G. R. G. Lanckriet, L. E. Ghaoui, C. Bhattacharyya, and M. I. Jordan, 2002, Minimax probability machines. in Advances in Neural Information Processing Systems (NIPS) 14, T. G. Dietterich, S. Becker, and Z. Ghahramani editors, Cambridge, MA: MIT Press.

[7] T. R. Strohmann and G. Z, 2002, Grudic, A Formulation for minimax probability machine regression, In Advances in Neural Information Processing Systems (NIPS) 14, T. G. Dietterich, S. Becker, and Z. Ghahramani editors, Cambridge, MA: MIT Press.

[8] K. Z. Huang, H. Q. Yang, and I. King et al, 2004, Biased minimax probability machine for medical diagnosis, Tech. Rep., Dep. Computer Science and Engineering, The Chinese University of Hong Kong.

[9] C. H. Hoi and M. R. Lyu, 2004, Robust face recognition using minimax probability machine, Tech. Rep., Dep. Computer Science and Engineering, The Chinese University of Hong Kong.

[10] G. R. G. Lanckriet, L. E. Ghaoui, C. Bhattacharyya, and M. I. Jordan, 2002, A robust minimax approach to classification, Journal of Machine Learning Research, Vol.3, pp.555-582.