

Combined Prediction Model of Network Traffic Based on Kohonen Neural Network

Xingle Tang, Wensheng Sun, Jinsong Yao

School of Communication Engineering, Hangzhou Dianzi University, Hangzhou
Email: 62785154@qq.com

Received: Oct. 22nd, 2014; revised: Nov. 16th, 2014; accepted: Nov. 25th, 2014

Copyright © 2014 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Considering that the original prediction model whose accuracy is low, and which highly depends on the training data and can't well described the characteristics of network traffic, we proposed a mixed traffic prediction model. The model is based on the Kohonen neural network features, that is, quickly learning rate, highly classification accuracy and strongly anti-noise. By wavelet transforming, we decompose the network traffic into high frequency part and the low frequency part, and the high frequency part is dealt by using Kohonen neural network prediction model, the low frequency part by using autoregressive AR model to predict by using Matlab to simulat. Through the experiment we conclude this combination prediction model can improve the prediction accuracy on multiple time scales and the nonlinear changing network traffic.

Keywords

Wavelet Transform, Neural Network, Traffic Prediction, Self-Organizing Mapping

基于Kohonen神经网络的组合式流量预测模型

唐幸乐, 孙文胜, 姚劲松

杭州电子科技大学通信工程学院, 杭州
Email: 62785154@qq.com

收稿日期: 2014年10月22日; 修回日期: 2014年11月16日; 录用日期: 2014年11月25日

摘要

该文针对传统预测模型预测精度低、对训练数据依赖程度高以及不能很好的刻画网络流量特征等不足，提出了一个混合的流量预测模型。该模型根据Kohonen神经网络的学习速率快、分类精度高、抗噪声能力强等特性，通过小波变换将网络流量分解为高频部分和低频部分，高频部分采用Kohonen神经网络进行预测，低频部分采用自回归AR模型进行预测，并采用Matlab进行仿真，通过实验得，这种组合预测模型可以提高对非线性、多时间尺度变化的网络流量的预测精度。

关键词

小波变换，神经网络，流量预测，自组织映射

1. 引言

随着网络的迅猛发展，网络结构越来越复杂，网络出现拥塞的可能性也越来越大[1]，如果能对网络流量提前预测，对流量进行疏通，就可以有效避免因拥塞而产生的时延等问题，因此建立一个有效的网络流量预测模型，成为了国内外研究的热点。传统的网络流量预测模型只处理平稳过程和非平稳过程，但真实的网络环境，呈现出明显的多尺度性和自相似性，导致预测误差较大，如 Matkov 模型、Poisson 模型、AR 模型等。文献[2]使用 BP 神经网络进行流量预测，达到了良好的拟合效果，也有较高的预测精度，但 BP 神经网络不能很好的处理多尺度特性的数据[2]。Kohonen 神经网络是一种自组织网络，采用无监督学习算法进行网络训练，在样本分类、排序和样本检测方面有着广泛应用，基于 Kohonen 神经网络的流量预测模型，不仅可以保证预测的精度，同时可以提高网络的认知特性[3]。鉴于此，本文提出一种新型的基于 Kohonen 网络流量的组合式预测模型，对网络流量进行预测。

2. 网络流量的小波分解

小波变换 $X(t)$ 是在一组小波基上的投影，分别是时间平移后的近似信号和尺度(频率)上缩放后的细节信号。小波基包括一个由原型小波通过规范平移和伸缩得到的函数族 $\psi(t)$ 和一个尺度函数族 $\varnothing(t)$ ，这两个函数集是[4]：

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k) \quad j, k \in Z \quad (1)$$

$$\varnothing_{j,k}(t) = 2^{-j/2} \varnothing(2^{-j}t - k) \quad j, k \in Z \quad (2)$$

W_j 表示尺度函数基 $\varnothing_{j,k}(t)$ 生成的子空间， V_j 表示小波函数基 $\psi_{j,k}(t)$ 生成的子空间，依次正交分解 $V_j = V_{j-1} \oplus W_{j-1}, \forall j \in Z$ ，可得到式 3：

$$V_j = V_{j-1} \oplus W_{j-1} = V_{j-2} \oplus W_{j-2} \oplus W_{j-3} = \dots = V_{j-n} \oplus W_{j-n} \oplus \dots \oplus W_{j-2} \oplus W_{j-1} \quad (3)$$

其中 n 为分解层数。从而可以得到近似信号和细节信号：

$$A_j(t) = \sum a_x(j, k) \varnothing_{j,k}(t) \quad (4)$$

$$D_j(t) = \sum d_x(j, k) \psi_{j,k}(t) \quad (5)$$

式中， $D_j(t)$ 为第 j 层的细节信号；系数 $d_x(j, k)$ 表示通过信号 $X(t)$ 和 $\psi_{j,k}(t)$ 的内积。因此，网络流量信号可以通过其逼近信号以及相应的细节信号无损地表示出来，即

$$X(k) = A_j(t) + \sum D_j(t) \tag{6}$$

3. 改进的流量预测模型

3.1. 基于 Kohonen 神经网络的组合式预测模型建立

传统小波神经网络预测模型，有着一些缺陷，存在过拟合现象。过拟合现象表示，当训练能力到达一个极限时，继续提高训练能力，反而使预测精度下降[5]。因此，提出了基于 Kohonen 神经网络并辅以小波 BP 神经网络的预测模型，新模型引入最小二乘法对输入样本进行处理，然后将处理过的样本进行小波分解，并将得到的高频序列和低频序列，利用 Kohonen 网络和自回归 AR 网络进行分开预测，最后用 BP 网络进行拟合，得到最后的流量预测。新流量预测模型如图 1 所示。

3.2. Kohonen 神经网络分类预测模型

Kohonen 神经网络是一种自组织，无监督的自适应映射模型。这种映射是对输入模式 x 在输出层中找出所对应的类别，即最优匹配的神经元 C 。Kohonen 分类预测模型先将标准样本进行分类，并对分类的结果分别进行 BP 网络预测，预测模型如图 2 所示。

该模型将 Kohonen 神经网络，BP 神经网络有机结合起来，先通过已有的流量数据训练 Kohonen 网络，并进行有效的分类，再对前一天的流量输入找到所属的类别，并用相应的 BP 网络预测后一天的流量。

3.3. Kohonen 分类训练算法

Kohonen 训练数据是来自事先测量的流量数据，在本文中是历时 15 天每隔 1 小时的网络流量。

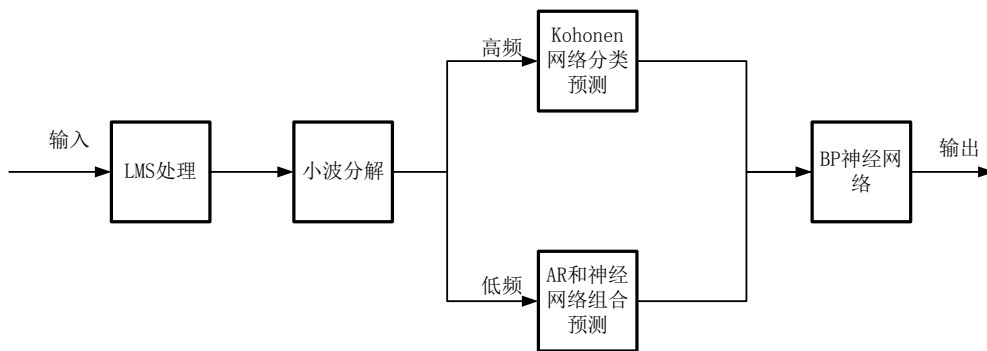


Figure 1. Network prediction model

图 1. 网络预测模型

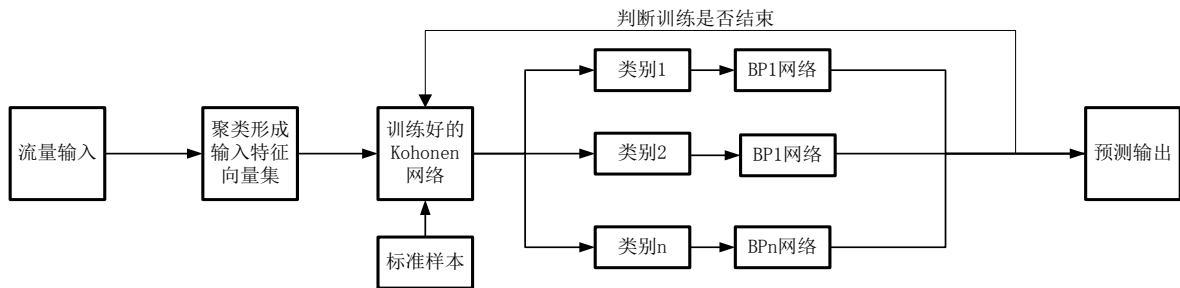


Figure 2. The classification and prediction of Kohonen neural network

图 2. Kohonen 神经网络分类及预测

算法步骤如下：

- 1) 确定 Kohonen 神经网络结构；
- 2) 随机选取权重矢量 m_j 初值为较小值，其中 $m_j = (\mu_{1j}, \mu_{2j}, \dots, \mu_{nj})$, $j = 1, 2, \dots, 16$ ；
- 3) 对于输入特征向量集 x^p ($p = 1, 2, \dots, 60$)，选取其中之一作为 Kohonen 神经网络的输入；
- 4) 计算输入向量 x^p 与权重矢量 m_j 的匹配程度，并用欧氏距离表示

$$d_j = \sqrt{\sum_{i=1}^{60} (x_i^p - u_{ij})^2}, \quad j = 1, 2, \dots, 16 \quad (7)$$

d_j 就是该输入向量的获胜神经元；

- 5) 根据下式修正获胜神经元 c 所连接的权向量与其几何邻域 $N_c(t)$ 内的神经元

$$\begin{cases} m_j(t+1) = m_j(t) + \alpha(t)(x - m_j(t)), & j \in N_c(t) \\ m_j(t+1) = m_j(t), & j \notin N_c(t) \end{cases} \quad (8)$$

其中， $\alpha(t)$ 表示可变学习速度；

- 6) 反复进行式 3~式 6 的步骤，直到输出层获胜神经元 c 保持稳定为止；

3.4. Kohonen 分类预测模型算法

- 1) 对输入信号进行聚类形成输入特征向量集，聚类分析每一样本为 $u_{ii} = \{x_{i1}, x_{i2}, \dots, x_{i24}, i, \max, \min\}$ ，以数据包为单位构造特征向量。其中 u_{ii} 为第 i 层细节信号第 i 天的样本集， $x_{i1}, x_{i2}, \dots, x_{i24}$ 为第 i 天的 1, 2, ..., 24 小时的流量数据， i 表示第 i 层细节信号， \max 表示这一天流量最大值， \min 表示这一天流量最小值；

- 2) 对特征向量归一化，公式如下：

$$x_{ik} = \frac{x_{ik} - \min}{\max - \min}, \quad i = 1, 2, 3, 4; k = 1, 2, 3, \dots, 24 \quad (9)$$

- 3) 将特征向量输入到 Kohonen 神经网络进行分类；

- 4) 根据分类的结果，进行相应的 BP 神经网络的预测。BP 神经网络的训练数据来自标准样本事先分类的网络流量。若聚类后某类别还有 n 个相似的天，用这 n 个相似的样本信号 $X_{ii} = \{x_{i1}, x_{i2}, \dots, x_{i24}\}$ 来预测第 i 层第 $i+1$ 天 $X_{i(i+1)} = \{x_{(i+1)1}, x_{(i+1)2}, \dots, x_{(i+1)24}\}$ 的数据，样本构成如下：

样本 1：输入 $\{x_{i1}, x_{i2}, \dots, x_{i24}\}$ 输出 $x_{(i+1)1}$

样本 2：输入 $\{x_{i2}, x_{i3}, \dots, x_{(i+1)1}\}$ 输出 $x_{(i+1)2}$

...

样本 24：输入 $\{x_{i24}, x_{(i+1)1}, \dots, x_{(i+1)23}\}$ 输出 $x_{(i+1)24}$

训练样本不要选择太多，过多会导致神经网络过渡学习，影响网络的预测效果；

- 5) 判定是否达到预测天数，若未达到，则继续预测，若达到，则预测结束，输出预测结果；

4. 仿真结果与分析

本文主要想通过前 11 天的流量数据，来预测后 4 天的流量，利用 Matlab 工具进行仿真分析，网络流量数据是来自杭州电子科技大学的某实验室路由器历时 15 天每隔 1 小时的历史数据，如图 3 所示。

首先利用小波分解，由于 4 层分解已经能满足本实验的要求，故将原始流量 X 分解成[6]： $X = D1 + D2 + D3 + D4 + A4$ ，如图 4 所示。

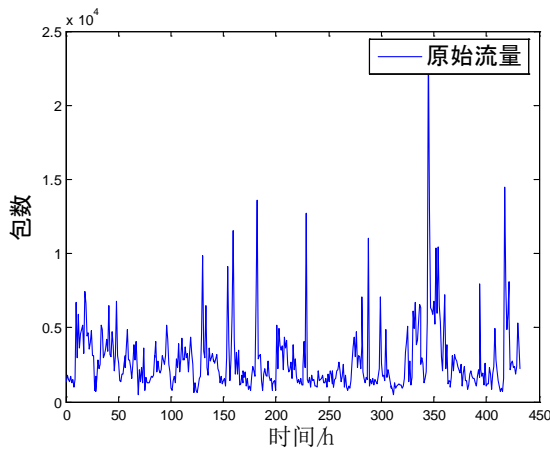


Figure 3. The original data
图 3. 原始数据

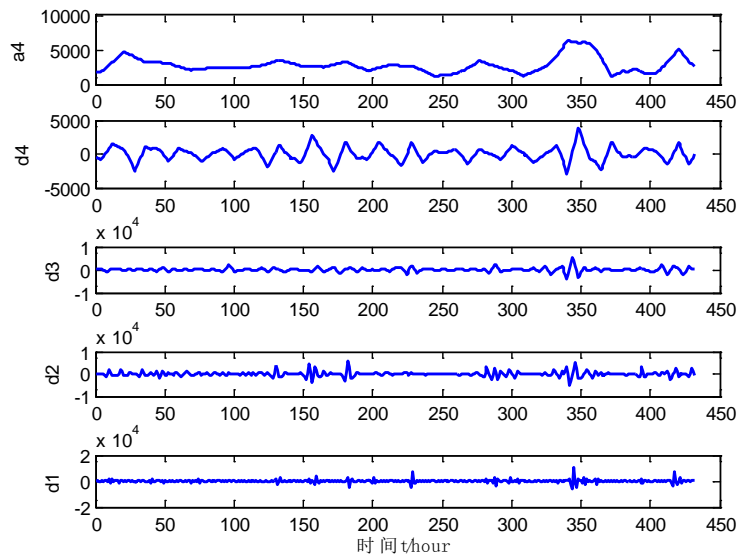


Figure 4. Four layer wavelet decomposition signal figures
图 4. 小波分解 4 层细节信号图

其次，对近似信号 A4 进行 AR 模型和 BP 神经网络组合模型预测。第一步，先确定 AR 模型阶数。本文采用 AIC、BIC 准则对 AR 模型进行估计，得出 AR 模型最佳预测阶数为 37 阶[7]。AR 模型参数如下：

$$AR = [-1.189, 0.0546, 0.05028, 0.01359, -0.02574, 0.03234, 0.04957, 0.01302, -0.03648, 0.0228, \\ -0.01926, 0.03582, 0.09604, -0.02604, -0.006172, -0.006831, -0.03916, -0.02708, -0.02063, \\ 0.03503, 0.01705, -0.0226, -0.01048, -0.02193, 0.02011, 0.01779, 0.01958, -0.01218, -0.02148, \\ 0.0189, -0.004451, 0.008942, 0.01338, -0.003202, 0.009283, -0.02904, -0.003816]$$

第二步，确定 BP 网络是 $1 \times 25 \times 1$ 三层结构，隐藏层采用 tansig 算法，输出层采用 purelin 算法，训练函数是 trainlm 算法，且目标误差是 0.001。

然后，将小波分解的细节信号进行 Kohonen 分类，利用 Matlab 神经网络工具箱，对 4 层细节信号进

行分类, 由于实验样本量不大, 仅有 15 天流量数据, 既 60 组样本, 因此设计网络的竞争层为 4×4 的结构即可。将第 1 层细节信号的编类为 1, 2, ..., 15 以此类推, 第 4 层细节信号的编类为 46, 47, ..., 60, 由网络的输出结果可以看到 60 组数据分成了 8 组, 如表 1 所示。

最后得到的组合式模型预测后与实际流量输入曲线, 如图 5 所示。

从图 5 所示的预测结果可以看出, 基于 Kohonen 神经网络的预测模型对后 4 天的流量有较好的预测。为了衡量该模型与传统的预测方法之间的优劣, 采用相对均方差(RMSE)作为评判标准:

$$RMSE = \frac{\sum(x_i - \hat{x}_i)}{\sum x_i^2},$$

值越小则越准确。从表 2 的结果可知, 基于 Kohonen 神经网络的预测模型比传统预测模型有更高的预测精度。

5. 结束语

本文将 Kohonen 神经网络用于网络流量预测, 提出了基于 Kohonen 神经网络的组合式流量预测模型。

Table 1. Sample clustering results
表 1. 样本聚类结果

| 聚类类别 | 该聚类的样本编号 |
|------|--|
| 1 | 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 |
| 2 | 16, 17, 18, 19, 23, 24, 25, 27 |
| 3 | 20, 21, 22, 26, 28, 29, 30 |
| 4 | 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 43, 44, 45 |
| 5 | 42 |
| 6 | 46, 47, 48, 49, 50, 53, 54, 58, 59, 60 |
| 7 | 51, 52 |
| 8 | 56, 57 |

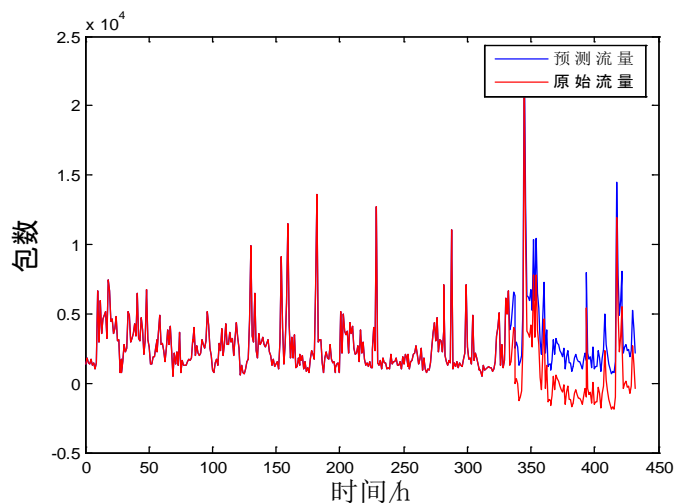


Figure 5. Traffic prediction by combined model of network based on Kohonen neural network
图 5. 基于 kohonen 神经网络的组合式流量预测

Table 2. Compared with traditional model by RMSE
表 2. 与传统预测模型 RMSE 对比

| 预测模型 | RMSE |
|--------------------|-------|
| AR模型 | 0.673 |
| 小波BP神经网络模型 | 0.438 |
| 基于Kohonen神经网络的预测模型 | 0.397 |

该模型综合运用神经网络，小波分解，AR 线性预测等理论方法进行建模，根据前 11 天历史流量进行后 4 天的流量预测，结果表明，这种组合预测模型可以提高对非线性、多时间尺度变化的网络流量的预测精度。

参考文献 (References)

- [1] 邢婷 (2010) 认知网络中一种基于神经网络预测模型的负载均衡算法. 北京交通大学, 北京, 1-10.
- [2] 谭骏, 陈兴蜀, 杜敏, 等 (2012) 基于自适应 BP 神经网络的网络流量识别算法. *电子科技大学学报*, **41**, 580-584.
- [3] Wang, P. and Liu, Y. (2008) Network traffic prediction based on improved BP wavelet neural network in wireless communications. *Networking and Mobile Computing*, Beijing, 1-5.
- [4] 王祥 (2012) 小波分析与神经网络的网络流量预测模型. *无线电工程*, **6**, 8-11.
- [5] 麻书钦 (2013) 基于 Kohonen 神经网络算法的网络入侵聚类算法的测试研究. *中国测试*, **4**, 113-116.
- [6] 马力, 张高明, 苟娟迎 (2012) 一种基于小波变换的校园网流量预测方法研究. 第一届中国互联网学术会议论文集, 北京, 1-10.
- [7] 郑文翰, 邹金慧 (2012) 小波分解与 AR 结合预测太阳黑子数的研究. *信息技术*, **11**, 98-102.