

# 一种新的 P2P 网络流量预测模型<sup>\*</sup>

韩志杰<sup>1</sup> 王汝传<sup>2</sup>

(苏州大学计算机学院 苏州 215006)<sup>1</sup> (南京邮电大学计算机科学与技术系 南京 210003)<sup>2</sup>

**摘要** 当前,对等计算(Peer to Peer)引起了广泛的关注,其典型应用有文件共享,即时通信等等,为了更好地使用 P2P 资源,合理规划 P2P 网络资源,建立 P2P 流量预测模型具有十分重要的理论意义和现实价值。基于小波神经网络提出了一个 P2P 网络流量预测模型,预测模型共分为输入层、隐含层、输出层,通过使用小波代替传统神经网络的 Sigmoid 作为神经网络的隐节点激励函数,在 Matlab6.5 仿真平台中实现,结果显示能够有效提高 P2P 网络流量预测精度。

**关键词** 神经网络,小波,P2P,网络流量

## Novel Peer to Peer Network Traffic Prediction Algorithm

HAN Zhi-jie<sup>1</sup> WANG Ru-chuan<sup>2</sup>

(School of Computer Science & Technology, Soochow University, Suzhou 215006, China)<sup>1</sup>

(Department of Computer Science and Technology, Nanjing University of Posts & Telecommunications, Nanjing 210003, China)<sup>2</sup>

**Abstract** Recently, there has been a growing interest in the potential use of Peer to Peer computing (P2P) in many applications such as file sharing, instant communication. Therefore, to realize their potential, there is a need of a P2P traffic prediction algorithm that facilitates the deployment of a network traffic that is optimized in terms of network bandwidth. This paper focuses on developing a novel P2P traffic prediction using wavelet neural network. The architecture has three layers: input layer, hidden layer, and output layer. By using the wavelet analysis taken place on the Sigmoid as excitation function of neural network hidden node, the proposed algorithm is implemented in matlab 6.5 and the simulation result has been proved is viable.

**Keywords** Neural network, Wavelet, P2P, Network traffic

## 1 引言

目前对等计算技术(Peer to Peer Computing, 简称 P2P)已经受到学术界和产业界的双重关注,财富杂志更将 P2P 列为影响 Internet 未来的四项科技之一,与网格计算技术(Grid Computing)一并成为分布式计算技术领域的两大研究热点。P2P 的思想改变 Internet 原来的 C/S 计算(Client/Server Computing)或是 B/S 计算(Browser/Server Computing)这样不对称的计算模式,每个节点地位对等,可以同时成为服务的使用者和提供者,这为大规模的信息共享、直接通信和协同工作提供了灵活的、可扩展的计算平台。

P2P 应用日益广泛也带来了一些负面问题,如(1)大量耗费网络带宽资源,据统计<sup>[1]</sup>某些 P2P 网络流在 Internet 上的流量有时占到 50%以上,容易造成其他应用资源紧张;(2)安全性问题,有些 P2P 软件可以穿透现有防火墙和安全代理,从内部打开一个企业网络安全防护的漏洞,使得各类病毒可以轻易进入企业,同时也可能造成个人或企业私密泄漏。研究和分析 P2P 网络流量特征,预测 P2P 网络流量,合理规划 P2P 网络流量具有十分重要的理论意义和现实价值。

文献[2,3]提出并验证 P2P 网络流量具有固定连接性,表现出明显的非线性变化特征,文献[1]表明 P2P 网络节点的拓扑连接、在线时长、节点流量及平均带宽均具有重尾分布特征,但均不服从幂率分布,文献[4]测量了采用多点下载技术的 Edonkey 网络对于 Internet 骨干网汇聚流量的影响,测量结果的拟合模型表明,由 Edonkey 流量主宰的汇聚流量,其长相关性消失,此发现直接质疑了传统自相似流量模型的适用性。

目前尚未有针对 P2P 网络流量的预测模型,而传统的网络流量预测模型分为线性预测模型<sup>[5]</sup>和非线性预测模型。线性预测模型显然不适于 P2P 网络流量预测,神经网络作为非线性预测模型能够有效自适应处理 P2P 网络流量的非线性变化特征,且不需要数学模型。本文将小波分析与神经网络的方法结合在了一起,建立了小波神经网络 P2P 流量预测模型(WNNP2PEA, Wavelet Neural Network P2P Estimation Algorithm),经过学习,能够自适应处理 P2P 流量的非线性变化特征,具有良好的预测效果,并结合具体的 P2P 应用场景, eMule 文件共享系统,采用爬行法采集 P2P 网络流量<sup>[2,3]</sup>,并在网中进行 P2P 网络流量采样,在 Matlab6.5 仿真实现了

<sup>\*</sup> 本课题得到国家自然科学基金(60573141 和 60773041),江苏省自然科学基金(BK2005146),江苏省高技术研究计划(BG2005038, BG2006001),国家高科技 863 项目(2006AA01Z219, 2006AA01Z439, 2007AA01Z404, 2007AA01Z478),2006 江苏省软件专项,南京市高科技项目(2007 软资 127),现代通信国家重点实验室基金(9140C1101010603),江苏省计算机信息处理技术重点实验室基金(kjs050001、kjs06006)资助和江苏省高校自然科学基金研究计划(07KJB520083)资助。韩志杰 博士研究生,研究方向为计算机软件理论、计算机网格、信息安全与对等计算;王汝传 教授,博士生导师,主要研究方向为计算机软件、计算机网络和网格、信息安全、无线传感器网络、移动代理和虚拟现实技术等。

小波神经网络 P2P 网络流量预测模型 WNNP2PEA, 通过与实际 P2P 网络流量相比较, 仿真预测结果误差在允许范围之内, 有效提高了预测精度。

本文组织如下, 第 2 节介绍了小波神经网络及相关工作; 第 3 节提出了 P2P 网络环境下的小波神经网络流量预测模型 WNNP2PEA; 第 4 节仿真验证了小波神经网络 P2P 流量预测模型的性能, 通过采用爬行法在网关采集 P2P 网络流量, 在 Matlab6.5 仿真平台对小波神经网络流量预测模型进行了仿真, 并针对仿真结果进行了比较分析; 最后总结全文。

## 2 相关工作

目前小波神经网络作为一种预测模型, 已经成功应用在电力消耗<sup>[6]</sup>, 经济建模<sup>[7]</sup>, 天气预报<sup>[8]</sup>等方面, 在网络流量预测方面, 文献<sup>[9]</sup>将小波和神经网络叠加作为网络流量预测模型, 但只是小波放在神经网络之前, 提前用小波对网络流量进行除噪, 然后利用单纯的神经网络进行流量预测, 神经网络学习周期长, 易陷入局部次优, 且无法发挥小波多尺度分析的特点, 难以有效提高网络流量预测的效能。

由于目前无法建立有效的数学模型来描述 P2P 网络流量特征, 本文建立基于小波神经网络的 P2P 网络流量预测模型, 在本文预测模型中, 直接以小波函数代替常规神经网络的 Sigmoid 作为神经网络的隐节点激励函数, 以小波的尺度和平移参数作为神经网络的权值和阈值参数, 构成一个前馈型神经网络。利用小波的多尺度分析特性, 缩短了小波神经网络的学习时间, 使神经网络避免陷入局部次优, 从而提高神经网络的预测性能。

## 3 小波神经网络 P2P 流量预测模型

### 3.1 P2P 流量预测模型 WNNP2PEA

小波神经网络 P2P 流量预测模型 WNNP2PEA 采用三层的小波神经网络结构, 即输入层, 隐含层, 输出层。如图 1 所示。

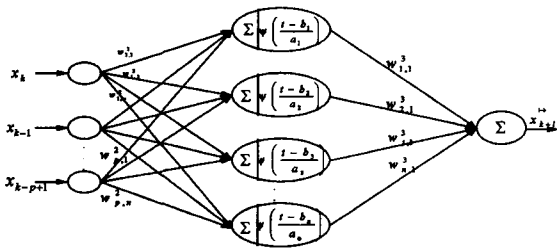


图 1 小波神经网络 P2P 网络流量预测模型

其中在输入层中有  $p$  个输入, 即一次输入含有  $p$  个元素的输入序列, 这  $p$  个元素是  $p$  个网络流量值,  $p$  表示预测流量与它之前的相关步数。隐含层包含  $n$  个神经元。输出层有 1 个神经元, 输出为第  $k+l$  个网络流量的预测值。  $w_{ij}^m$  表示从  $m-1$  层的神经元  $i$  到第  $m$  层的神经元  $j$  之间的权值,  $a_{jk}^m$  表示第  $m$  层神经元  $j$  的第  $k$  次输入,  $\psi_m$  表示第  $m$  层的转移函数,  $b_{jk}^m$  表示第  $m$  层的相应输出, 即

$$a_{jk}^m = \sum_i w_{ij}^m b_{ik}^{m-1}, b_{jk}^m = \psi_m(a_{jk}^m) \quad (1)$$

针对 P2P 网络流量的特征, 采用的三层网络结构,  $b_{jk}^m = x_{k-j+1}$ ,  $b_{jk}^m = x_{k+l}$ , 隐含层的转移函数采用非正交复 Morlet 小波, 这次采用非正交复 Morlet 小波因为满足 Mercy 条件<sup>[11]</sup>和再生性<sup>[12]</sup>, 能够有效降低神经网络的学习速度, 避免陷入

局部次优。

即

$$\Psi(t) = \cos(1.75t) e^{-\frac{t^2}{2}} \quad (2)$$

则有

$$a_{jk}^2 = \sum_{i=1}^p w_{ij}^2 x_{k-j+1}, b_{jk}^2 = \psi_2\left(\frac{a_{jk}^2 - b_i}{a_j}\right) \quad (3)$$

$$x_{k+l} = \sum_{j=1}^n b_{jk}^2 w_j^3 = \sum_{j=1}^n \psi_2\left(\frac{a_{jk}^2 - b_i}{a_j}\right) w_j^3 \quad (4)$$

用  $\theta$  表示上式中所有参数的集合, 网络的输入为  $p$  个网络流量  $x_k, x_{k-1}, \dots, x_{k-p+1}$  为元素的一个流量序列, 输出为第  $k+l$  个网络流量的预测值  $x_{k+l}$ , 把预测值的均方差函数作为目标函数  $C(\theta)$ 。

$$C(\theta) = \frac{1}{2} (x_{k+l} - \hat{x}_{k+l})^2 \quad (5)$$

为了使上述误差最小, 得出网络模型的最优参数, 本文采用共轭梯度下降法计算误差函数的最小值。为了方便, 令

$$\xi_j = \frac{a_{jk}^2 - b_i}{a_j}, e = x_{k+l} - \hat{x}_{k+l} \quad (6)$$

### 3.2 WNNP2PEA 学习算法

Step1 选用合适的小波函数, 这里选用 Morlet 母小波, 即

$$\Psi(t) = \cos(1.75t) e^{-\frac{t^2}{2}} \quad (7)$$

Step2 初始化参数  $\theta$ 。

Step3 输入学习样本得到相应的预测值。

Step4 计算  $C(\theta)$  的梯度

$$h(a_i) = \frac{\partial C}{\partial a_i} = - \sum_{k=1}^N e_k w_i \xi_i a_i^{-1} \{1.75 \sin(1.75 \xi_i) + \xi_i \cos(1.75 \xi_i)\} e^{-\xi_i^2/2} \quad (8)$$

$$h(b_i) = \frac{\partial C}{\partial b_i} = - \sum_{k=1}^N e_k w_i \xi_i a_i^{-1} \{1.75 \sin(1.75 \xi_i) + \xi_i \cos(1.75 \xi_i)\} e^{-\xi_i^2/2} \quad (9)$$

$$h(w_i) = \frac{\partial C}{\partial w_i} = \sum_{k=1}^N e_k \cos(1.75 \xi_i) e^{-\xi_i^2/2} \quad (10)$$

Step5 采用梯度最速下降法, 来对参数  $w_i, a_i, b_i$  进行调整, 调整方式为:

$$\begin{aligned} w_i(m+1) &= w_i(m) - ah(w_i) \\ a_i(m+1) &= a_i(m) - ah(a_i) \\ b_i(m+1) &= b_i(m) - ah(b_i) \end{aligned} \quad (11)$$

Step6 误差函数的绝对值小于预先设定的某个正值时, 停止网络的学习; 否则返回(3)步骤, 重复上述算法。

## 4 仿真比较

### 4.1 仿真说明

仿真的主要目的是比较小波叠加神经网络流量预测模型和小波神经网络 P2P 流量预测模型用来预测真实 P2P 网络流量的性能, 共分为训练时间比较和预测效果比较两个方面, 采用了 Matlab 6.5 平台进行比较仿真验证, 具体仿真步骤如下所示:

(1) 利用爬行法进行 P2P 网络流量采集, 在网关上对 Emule 网络流量数据进行采样, 采集是每隔 5m 采集一个数据, 共采集 2 次, 第一次采集了 88 组数据作为两种仿真模型的训练样本, 第二次采集 256 组数据作为预测数据。

(2) 在 Matlab6.5 中, 分别实现小波叠加神经网络流量预测模型和小波神经网络 P2P 流量预测模型。

(3)把步骤(1)中所采集得到的第一次 P2P 网络流量训练样本序列化为矩阵,作为样本分别进行小波叠加神经网络流量预测模型和小波神经网络 P2P 流量预测模型的学习训练。

(4)把步骤(1)中所采集得到的第二次 P2P 网络流量序列化为矩阵代入小波神经网络流量预测模型和小波神经网络 P2P 流量预测模型进行流量预测仿真。

(5)将仿真结果利用 Matlab 绘图工具绘制成图。

其中 Matlab6.5 小波神经网络 P2P 流量预测模型仿真采用了 8-10-1 结构,即输入层 8 个神经元,10 个隐含神经元,1 个输出神经元。模型参数的设定为相关步数  $T=8$ ,训练学习率  $lr=0.01$ ,动量系数  $mc=0.95$ ,训练的最大步数  $epochs=150000$ ,训练目标  $goal=0.07$ 。

### 4.2 训练时间仿真结果分析

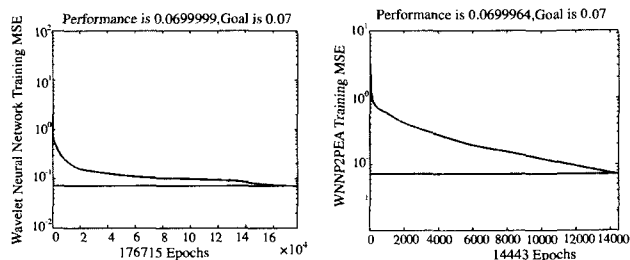


图 2 小波叠加神经网络学习时间 图 3 WNNP2PEA 学习时间

如图 2,图 3 所示,当训练目标  $goal=0.07$  时,小波叠加神经网络需要 176715 个训练步数才能达到目标,而 WNNP2PEA 只需要 14443 个训练步数就可以达到目标,约为小波叠加神经网络的十分之一,这是因为 WNNP2PEA 将小波作为常规神经网络的 Sigmoid 作为神经网络的隐节点激励函数,有效降低了神经网络的学习训练时间。

### 4.3 预测效果仿真结果分析

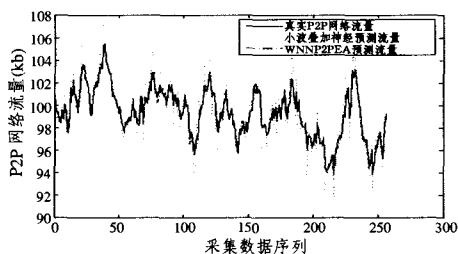


图 4 预测模型仿真结果

从图 4 中可以看出,与小波叠加神经网络相比, WNNP2PEA 预测流量曲线对实际 P2P 网络流量曲线具有良好的逼近性能,从图 5 和图 6 的比较分析,小波叠加神经网络的预测误差范围是  $(-2.5, 3.4)$ ,而 WNNP2PEA 的误差范围是  $(-0.8, 0.8)$ ,预测误差更小,说明 WNNP2PEA 的预测性能

更优, WNNP2PEA 利用小波的多尺度分析特性,避免陷入局部次优,能够有效提高神经网络的性能,从而提高了其预测效果。

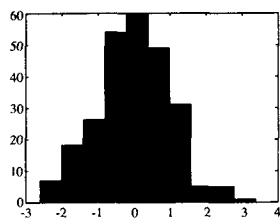


图 5 小波叠加神经网络预测误差统计图

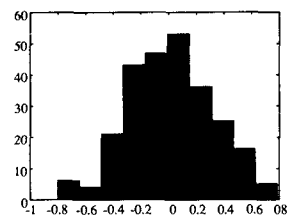


图 6 WNNP2PEA 预测误差统计图

**结束语** 本文提出了一种基于小波神经元的网络流量预测方法 WNNP2PEA,并在 Matlab6.5 仿真平台中实现在网中进行流量预测仿真实验,并与小波叠加神经网络作了比较分析,仿真结果表明适当地选择训练数据样本和恰当地选择网络的结构,可以使 WNNP2PEA 达到较好的预测精度。本文下一步的工作根据 P2P 网络流量特征,结合其他混合 P2P 流量,提出一个简单有效的实时在线预测模型。

### 参考文献

- [1] Sen S, Wang J. Analyzing Peer to Peer Traffic Across Large Networks, ACM / IEEE Transactions on Networking, 2004, 12 (2):137-150
- [2] M R, Foster I. Mapping the Gnutella network. IEEE Internet-Computing, Jan. 2002, 6:50-57
- [3] Saroui S, Gummadi P K, Gribble S D. A Measurement Study of Peer to Peer File Sharing Systems//Multimedia Computing and Networking 2002 (MMCN '02)
- [4] Azzouna B, Guillemin F N. Impact of Peer to Peer Applications on Wide Area Network Traffic: An Experimental Approach// IEEE Globecom 2004, Dallas, USA, Nov - Dec, 2004
- [5] Yu Guoqiang, Zhang Changshui. Switching ARIMA model based forecasting for traffic flow[C]//ICASSP'04. vol 2:429-432
- [6] 张大海,毕研秋,等.小波神经网络及其在电力负荷预测中应用概述//Proceeding of the EPSA. 2004(8)
- [7] 胡俊胜,肖东荣,夏景明.基于小波神经网络的经济预测研究.统计与决策,2005(3)
- [8] 金龙,秦伟良,姚华栋.多步预测的小波神经网络预报模型[J].大气科学,2000(1)
- [9] 冯海亮,陈涤,林青家,等.一种基于神经网络的网络流量组合预测模型.计算机应用,2006,26(9)
- [10] Vapnik V N. Statistical Learning Theory[M]. New York, Springer-Verlag, 2000
- [11] 蒋刚,肖建,宋昌林,等.一种非正交复小波核函数及其非线性参数辨识应用.系统仿真学报,2006,18(9)