

组合神经网络的网络流量预测研究

章 治

(江西科技师范学院 现代教育技术中心, 江西 南昌 330013)

摘 要: 提出一种组合神经网络的网络流量预测模型. 首先采用 SMOF 网络对网络流量数据进行聚类, 然后采用 Elman 网络对聚类后的流量数据进行训练并预测, 同时采用遗传算法对 Elman 网络的网络结构进行优化, 提高网络流量预测精度. 仿真结果表明, 组合神经网络加快了网络流量预测速度, 提高了网络流量预测精度, 克服了单一预测模型不足, 为网络流量预测提供了新的思路, 具有很好的应用前景.

关键词: 网络流量; 遗传算法; 神经网络; 预测

中图分类号: TP391

文献标识码: A

文章编号: 1000-7180(2012)03-0098-04

Network Traffic Prediction Based on Combination Neural Network

ZHANG Zhi

(Technique Center of Modern Education, Jiangxi Science & Technology Normal Universit, Nanchang 330013, China)

Abstract: This paper proposes a combination network traffic prediction model based on neural networks. Firstly, SMOF network is used to cluster the network traffic data, and then used the clustering Elman network to train and predict, and Elman network is optimized by genetic algorithm to improve the network traffic prediction accuracy. Simulation results indicate that the combined neural network prediction model accelerates the prediction speed and improve the network traffic prediction accuracy, overcoming the single forecasting model insufficient, it has the very good application prospects.

Key words: network traffic; genetic algorithm; neural network; prediction

1 引言

网络流量主要有传统统计方和智能预测方法两大类预测算法, 统计方法主要包括回归预测法、多元统计方法和时间序列模型法^[1-2]. 智能预测法包括支持向量机和神经网络法等. 传统统计方法基于线性建模, 难以准确描述网络流量通与影响历子间的非线性关系, 预测结果不是很可靠. 神经网络预测精度比较高, 但收敛速度慢、泛化能力差, 易陷入局部极小^[3]. 随着网络规模增大, 网络流量具有随机性、周期性和突发性变化特点, 当前大多数预测系统仍然是由单一的预测方法或凭网络管理人员的经验进行预测, 预测结果与实际需要有一定差距^[4-5].

为了克服当前网络流量存在的上述缺陷, 本文利用组合模型思想, 采用本文采用 SMOF 网络和

Elman 神经网络对网络流量进行预测, 并将遗传算法引入神经网络参数优化中, 对神经网络模型进行优化改进, 从而取得了较高的网络流量预测精度.

2 网络流量预测原理

网络流量在某一时刻的总流量 $X(t)$ 可用下述模型描述为:

$$X(t) = N(t) + W(t) + U(t) \quad (1)$$

其中, $N(t)$ 表示网络流量在某个时刻的基本正常流量; $W(t)$ 表示上网人行为对网络流量的影响; $U(t)$ 表示周末和节假日对网络流量的影响.

从式(1)可知, 网络流量复杂多变的非线性系统, 采用单一的模型无法全面的反映网络流量的变化规律, 难以获得比较高的预测精度^[10]. 因此本文采用组合模型对其进行预测, 以提高网络流量的预

测精度.

3 组合神经网络的网络流量预测模型构建

3.1 网络流量数据预处理

网络流量具有突发性和随机性,最大值与最小值差异较大,对预测模型的训练过程产生干扰,为此,在进行网络流量训练前,对网络流量数据进行归一化处理,将所有数据归一化到 $[0.1, 0.9]$,具体归一化公式为:

$$x' = 0.8 \left(\frac{x}{x_{\max} - x_{\min}} \right) + \left(0.9 - 0.8 \frac{x_{\max}}{x_{\max} - x_{\min}} \right) \quad (2)$$

其中, x' 表示网络流量归一化后的值, x 表示归一化之前的网络流量实际值, $x_{\max}$ 是网络流量样本中的最大值, $x_{\min}$ 表示网络流量样本中的最小值. 可通过下面反归一化公式可以将模型的预测值转换为真实值.

$$x = \frac{(y - 0.9)}{0.8} \Delta + x_{\max} \quad (3)$$

3.2 采用自组织特征神经网络对流量聚类

自组织特征映射 (Self-Organizing Feature Map, SOFM) 网络是由芬兰学者 Kohonen 提出的人工神经网络, 是一种由全连接的神经元阵列无教师示教的聚类方法. SOFM 网络模拟了大脑神经细胞对外界信息刺激的反应, 其可以对各个模式中所含的特征进行捕捉, 同时对其进行自组织, 最后在网络的竞争层将聚类结果表示出来, 具有很强的抗干扰能力. SOFM 网络一般包括输入层和竞争层两部分, 其基本结构如图 1 所示.

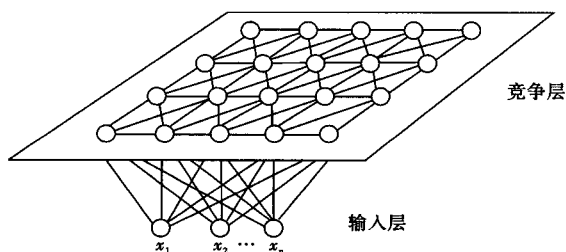


图1 SOFM网络的基本结构

SOFM 网络分为训练和工作两个阶段. 训练阶段包括选择最佳匹配神经元、网络权值自适应变化, 以及网络学习速度的更新过程, 确定输入和输出之间的关系. 工作阶段对输入进行自动分类, 因此本文首先采用 SOFM 网络对网络流量样本进行聚类处理.

将某时刻的网络流量数据作为 SOFM 网络一个输入向量, 将某段时间的网络流数据作为输入样本集对 SOFM 网络进行训练. 训练完成后, 每一聚类的特征矢量对应一条网络流量特征曲线. 由于 SOFM 网络仅根据输入向量与权值向量之间欧氏距离进行分类, 因此可以反映网络流量变化总体特征, 从而准确描述网络流量变化的趋势.

3.3 Elman 神经网络进行网络流量预测

Elman 神经网络是一种动态反馈型网络, 系统结构通常包括输入层、隐层输出层 3 个层次, 其还有一个有特殊的联系单元, 即连接层. 具体结果如图 2 所示. Elman 神经网络在进行预测时, 隐层输出通过连接层的延迟和存储连接到隐层, 作为隐层的输入, 通过该连接方式使其对网络流量历史状态的数据具有敏感性, 同时由于内部反馈网络的加入使处理动态信息的能力增加, 可以实现动态建模目的.

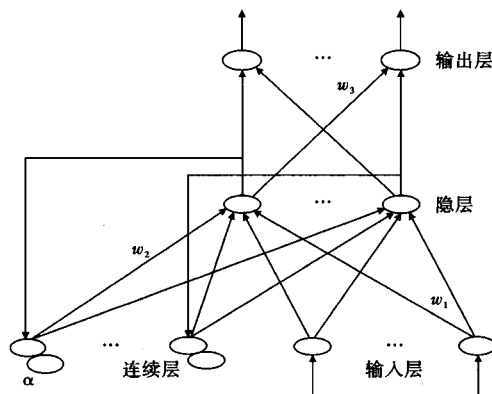


图2 Elman神经网络的结构

由于传统前向神经网络只包括输入层、隐层和输出层 3 个层次, 是一种静态神经网络, 对网络流量的预测只与当前网络的输入相关, 而与之之前网络的输入不相关, 不具有记忆能力. 但是网络流量变化具有周期性和自相似性, 为了弥补前向神经网络的不足, 传统的方法就是在输入层内包含大量的历史流量数据, 计算量增大, 易陷入局部最优, 因此不能提高网络流量的预测精度, 反而使预测模型的性能降低. Elman 神经网络可以对隐层输出进行反馈, 将隐层当前时刻的状态保存下来, 在下一时刻可以再次反馈给隐层, 这样使网络学习具有动态记忆性, 因此可以提高网络流量的预测精度.

3.4 遗传算法优化优化 Elman 神经网络参数

在 Elman 神经网络学习过程, 存在对网络初始权值比较敏感, 隐层节点数难确定等难题, 但是这些对模型的预测性相当的重要. 本文利用遗传算法

的全局搜索能力,确定网络初始权值和隐层节点数.采用遗传算法对 Elman 神经网络进行优化时,首先给定网络连接权和隐节点数的取值范围,即参数的解空间,然后确定编码方式和适应度函数,本文编码方式为实数编码,最后确定交叉概率和变异概率,种群个数,最大进化代数. Elman 神经网络参数的优化步骤具体如下:

(1)采用实数编码方法对初始权值和隐层节点数进行编码,因此遗传算法的个体由权值、阈值和隐层节点数组成.

(2)在解空间,通过采用随机方式产生初始种群.

(3)将训练样本输入到 Elman 神经网络,计算群体中每一个体的适应度值 $f(i)$.

$$f(i) = \frac{1}{1 + E(i)} \quad (4)$$

其中, $i=1,2,\dots,m$, m 表示种群个数, $E(i)$ 表示网络流量拟和的绝对误差平均值.

(4)根据适应度值,对种群中的个体进行选择、交叉和变异等遗传操作,产生新的种群.

(5)不断的重复步骤(3)和(4),并保留每一代中最好的个体,即最好参数解.

(6)对终止条件进行判断,如果满足,就结束,输入最优遗传算法找到的最优个体,即 Elman 神经网络参数的最优初始权值和隐层节点数.

3.5 组合神经网络的网络流量预测模型

通过对 SOFM 网络、Elman 网络和遗传算法分析和研究,基于组合神经网络的网络流量预测过程如下:

(1)通过一定的设备对网络流量进行收集.

(2)对原始网络流量数据进行归一化处理,消除数据之间差异.

(3)采用 SOFM 网络对网络流量数据进行分类,在有效的减少了数据量同时,不丢失原始数据信息,有利于模型的建立.

(4)将分类的信息输入到 Elman 神经网络进行学习.

(5)采用遗传算法对 Elman 神经网络初始权值和隐层节点数进行优化.

(6)利用最优参数建立最优预测模型并对网络流进行预测,如果预测精度达不到要求,就转步骤继续优化网络参数,直到满足预测精度要求为止.组合神经网络的网络流量预测模型如图 3 所示.

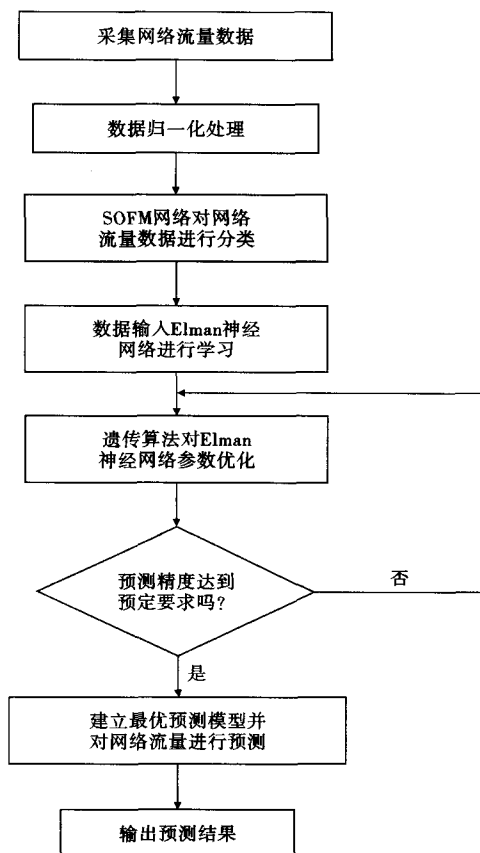


图3 组合神经网络的网络流量预测模型

4 仿真实例

4.1 网络流量数据

本文选择对某服务器的网络流量进行采集,其采集到 280 个数据,其如图 4 所示.由于网络流量变化范围比较大,因此首先对其进行归一化处理.

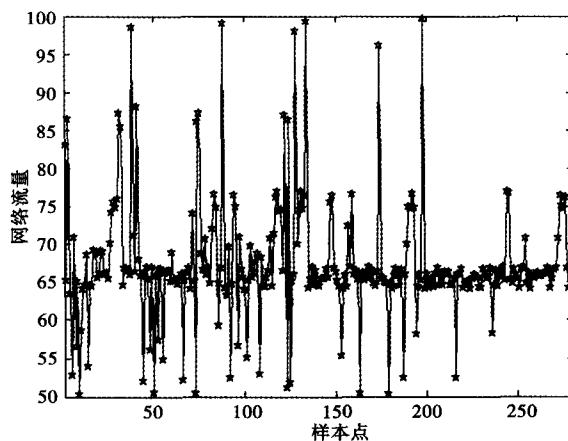


图4 网络流量原始数据

4.2 网络流量数据

将数据分成两部分,训练样本和测试样本,采用

BP神经网络、基本 Elman 网络和本文组合神经网络对训练样本进行学习,对测试样本进行预测,其预测结果如表 1、表 2 所示。

表 1 各种模型的网络流量预测结果

样本	实测值	BP 网络	Elman 网络	组合网络
1	66.28	65.33	71.95	67.67
2	65.66	71.67	69.60	66.63
3	65.78	72.05	75.24	69.34
4	66.14	74.51	68.35	67.37
5	66.18	70.30	72.73	69.76
6	66.90	68.45	75.47	65.91
7	66.25	74.40	68.26	66.07
8	64.20	71.91	71.01	67.26
9	65.42	69.29	70.91	65.10
10	66.65	74.98	70.22	67.64
11	66.90	66.36	70.67	66.90
12	70.72	71.71	73.50	70.14
13	76.52	83.11	86.22	80.35
14	75.04	76.69	79.24	75.74
15	74.63	79.09	79.69	77.64
16	76.16	84.74	83.56	78.71
17	66.86	73.40	76.74	70.48
18	64.17	68.26	65.35	63.32
19	66.29	75.21	74.78	68.88
20	64.42	68.23	66.46	64.69

表 2 三种预测模型的整体预测性能比较

模型	训练速度	绝对平均误差	均方根误差
BP 网络	100 s	5.075	5.790
Elman 网络	50 s	5.439	6.087
组合网络	10 s	1.7135	2.156

4.3 结果与分析

三种网络流量预测的预测结果和绝对误差分布分别如图 5 和图 6 所示。从图 5 和图 6,以及表 1 和表 2 可知,本文提出的组合神经网络的预测精度最快,预测误差最小,同时预测速度最快,仿真结果表明组合神经网络的网络流量预测模型是可靠性和合理性,在网络流量预测等非线性领域具有很好的应用价值。

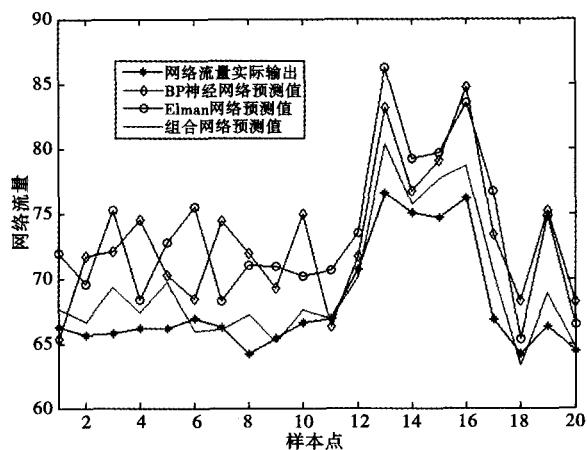


图 5 三种模型对预测样本的预测结果图

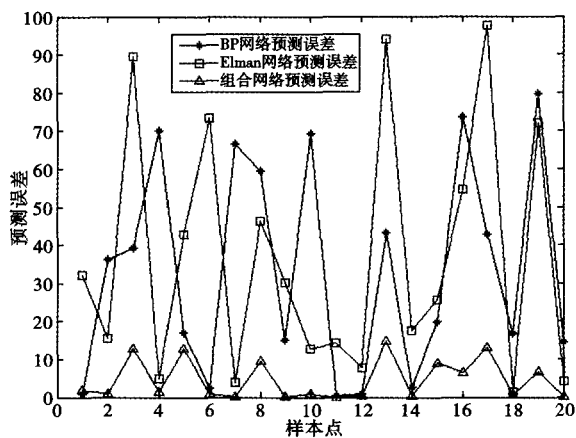


图 6 三种预测模型的预测样本预测误差

5 结束语

准确的网络流量预测是网络管理的基础,由于网络流量具有周期性、非线性等复杂变化规律,单一的预测模型的预测精度比较低。为提高网络流的预测精度,本文根据组合模型的原理,提出一种基于遗传算法的 SMOF 网络和 Elman 网络的网络流量组合模型。仿真结果表明,相对于单一的网络流量预测模型,组合模型不仅加快了网络流量预测的实时性,同时提高了网络流量的预测精度,结果说明组合模型利用各单一模型的优点,更加能够准确反映网络流的变化规律,在网络流量预测中具有广泛的应用前景。

参考文献:

- [1] 刘嘉,金志刚,薛飞,等. 基于 FARIMA 过程的网络业务预报与应用[J]. 电子与信息学报, 2001, 23(4): 403-407.

(下转第 105 页)

检测系统存在的比较高的漏报率以及高的误报率,同时也存在入侵检测的数据存在维数大、冗余度高等缺陷,本文基于均值聚类和遗传算法的优点的基础上,提出了一种新的基于聚类算法和遗传算法相结合的入侵检测方法模型.算法对聚类的中心采用二进制编码,将网络的正常行为和非正常行为分为不同的类,把每个点到它们之间的各自的聚类中心的欧几里得距离的综合来作为相似度量,然后采用粒子群优化算法,有效的降低网络拓扑路径长度,通过优化算法来寻找聚类的中心.仿真实验结果证明了本文提出的改进的网络异常检测方法,与较传统网络入侵检测系统模型相比,具有更好的入侵识别率和检测率,同时提高了算法的执行效率.

参考文献:

- [1] 段智彬,孙恩昌,张延华,等. 基于 ARMA 模型的网络流量预测[J]. 中国电子科学研究院学报, 2009,4(4): 352-356.
- [2] 王俊松,高志伟. 基于 RBF 神经网络的网络流量建模及预测[J]. 计算机工程与应用, 2008,44(13):7-9.
- [3] 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[J]. 计算机应用, 2006,26(9): 2206-2208.
- [4] Lahiri S K, Ghanta K C. Prediction of pressure drop

of slurry flow in pipeline by hybrid support vector regression and genetic algorithm model [J]. Chinese Journal of Chemical Engineering, 2008,16(6): 841-848.

- [5] Eslam Pourbasheer, Siavash Riahi, Mohammad Reza Ganjali, et al. Application of genetic algorithm-support vector machine (GA-SVM) for prediction of BK-channels activity[J]. European Journal of Medicinal Chemistry, 2009, 44(12):5023-5028.
- [6] Huang Cheng-Lung, Wang Chieh-Jen. A GA-based feature selection and parameters optimization for support vector machines [J]. Expert Systems with Applications, 2006,31(2):231-240.
- [7] Lin Shih-Wei, Ying Kuo-Ching, Chen Shih-Chieh, et al. Particle swarm optimization for parameter determination and feature selection of support vector machines [J]. Expert Systems with Applications, 2008,35(4): 1817-1824.

作者简介:

牛红惠 女,(1972—),硕士,讲师.研究方向为神经网络算法与应用、计算机网络.

徐甜 女,(1968—),硕士,副教授.研究方向为数据挖掘算法与应用、人工智能.

(上接第 101 页)

- [2] 谭晓玲,许勇,张凌,等. 网络流量短期预测方法的研究与应用[J]. 计算机工程与设计, 2006, 27(8): 1341-1345.
- [3] 邹月,陈启愉. 基于改进 BP 小波网络的网络流量预测研究[J]. 计算机仿真, 27(6):187-193.
- [4] 陈振伟,郭拯危. 小波神经网络预测模型的仿真实现[J]. 计算机仿真, 2008, 25(6): 147-150.
- [5] 薛可,李增智,刘浏,等. 基于 ARIMA 模型的网络流量

预测[J]. 微电子学与计算机, 2004,(07):84-87.

- [6] 邹月,陈启愉. 基于改进 BP 小波网络的网络流量预测研究[J]. 计算机仿真, 27(6):187-193.

作者简介:

章治男, (1977—),副教授,硕士.研究方向为现代教育技术和计算机网络.