

基于窗口的网络流量组合预测模型研究

岳玲玉, 郑明春

(山东师范大学 管理科学与工程学院, 山东 济南 250014)

摘要:在通信网络技术发展的过程中, 针对网络流量进行建模和预测的研究一直备受人们关注。为了更好地对网络流量进行建模和预测, 有效提高网络的运行速度和利用率, 加强网络管理建设, 文中提出了网络流量组合预测模型, 该模型由三次指数平滑模型和基于 BP 神经网络模型两个子模型组合而成。首先介绍了组合预测模型的预测机理, 然后对三次指数平滑模型和基于 BP 神经网络模型两种子模型进行了详细介绍, 最后运用实例进行了仿真实验。实验结果表明组合预测模型预测误差稳定在 2% 以内, 取得了比较好的预测效果。

关键词:组合预测; 指数平滑模型; BP 神经网络

中图分类号:TP393

文献标识码:A

文章编号:1673-629X(2012)04-0111-04

Study on Window-Based Network Traffic Combined Prediction Model

YUE Ling-yu, ZHENG Ming-chun

(Dept. of Management and Economics, Shandong Normal University, Jinan 250014, China)

Abstract: In the development process of communication network technology, the network traffic modeling and prediction has always been concerned. In order to better model and predict network traffic, effectively improve the network speed and utilization, strengthen the construction of network management, a combined prediction model is studied in this paper, which includes three exponential smoothing model and BP neural network model. First, introduce the mechanism of the combined prediction model and then the two sub-models are described in detail. At last conduct a simulation experiment with an example. Results show that the combined forecast model prediction error stability is within 2%, and achieve good prediction.

Key words: combined prediction; three exponential smoothing model; BP neural network

0 引言

随着网络技术的迅速发展, 网络上的业务和应用也越来越丰富, 与此相对应对信息网络的管理也提出了更高的要求, 其中网络流量预测是业务管理的关键问题^[1]。同时网络流量预测对于设计新一代网络协议、有效地进行网络管理、设计高性能路由器算法和评价网络安全和检测网络异常都具有重要意义。

网络流量特征的分析是建立网络流量预测模型的基础和前提, 传统的有泊松模型和马尔可夫模型。Le-land 和 Wilson^[2-5]等人通过对网络上真实数据流量的分析指出, 网络流量具有称之为自相似的特性, 近年来已广泛把自相似特性用在网络性能和特征分析中。

出于网络流量控制的需要, 该领域的学者从各个

方面提出了很多关于网络流量的预测模型, 较早期的主要有自回归滑动平均模型、自回归模型、滑动平均模型和历史平均模型等。这些线性预测模型考虑因素都较为简单, 具有(相对而言)计算简便、易于实时更新数据、大规模应用的优点^[6]。但这些传统模型通常假设网络流量满足线性关系, 而实际网络流量数据包含很多影响因素并且各因素之间是非线性的。传统模型虽然能对未来时刻的网络流量进行一定的预测, 但要更加灵活、高效地进行流量预测还需要引入人工智能等一些方法。近些年来, 人们广泛利用人工智能方法来进行流量预测比如专家系统、模糊推理、模糊神经网络等, 其中应用最多的是人工神经网络^[7]。文中就对指数平滑这种传统模型与 BP 神经网络模型进行组合应用, 并用基于窗口的流量实例进行了仿真实验分析。

1 组合预测模型简介

1.1 组合预测模型的预测原理

组合预测模型的依据是: 某网络未来时刻的网络

收稿日期: 2011-08-31; 修回日期: 2011-12-01

基金项目: 山东省自然科学基金(Y2008G16)

作者简介: 岳玲玉(1987-), 女, 山东临沂人, 硕士研究生, 研究领域为网络信息系统; 郑明春, 教授, 研究方向为 Internet 服务质量、流量控制和拥塞控制。

流量与该网络的历史流量数据是相关的。历史数据对未来时刻的影响以及两者之间的非线性关系分别用三次指数平滑模型和 BP 神经网络模型来进行描述,然后将这两种子模型通过某种形式组合起来。假设组合预测模型的基本原理可用下面的表达式描述:

$$x(t+1) = \alpha x_1(t+1) + \beta x_2(t+1)$$

式中 $x_1(t+1)$ 、 $x_2(t+1)$ 分别为三次指数平滑模型和 BP 神经网络模型的预测输出结果; α 、 β 分别为两个子模型在组合模型中的权重比例。通过改变 α 、 β 的取值,可以调节两个子模型在组合模型中的权重。当一个子模型权重系数取 0 时,实际相当于舍弃该模型^[8]。

1.2 两子模型描述

1.2.1 三次指数平滑预测模型

指数平滑法实质上是将历史数据进行加权平均作为未来时刻的预测结果,它认为数据的重要程度按时间上的远近呈非线性递减。即近期数据影响价值大,权数亦大;远期数据影响价值小,权数亦小^[9]。它具有计算简单、样本要求量较少、适应性较强、结果较稳定等优点,适合对网络流量进行预测分析。

设 n 个原始数据组成一时间序列为: $y_1, y_2, y_3, \dots, y_t$, 三次指数平滑模型为:

$$y_{(t+r)} = a_t + b_t T + c_t T^2 \quad (1)$$

$$a_t = 3S_t^{(1)} - 3S_t^{(2)} + S_t^{(3)} \quad (2)$$

$$b_t = \frac{a}{2(1-a)^2} [(6-5a)S_t^{(1)} - 2(5-4a)S_t^{(2)} + (4-3a)S_t^{(3)}] \quad (3)$$

$$c_t = \frac{a^2}{2(1-a)^2} [S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)}] \quad (4)$$

$$S_t^{(1)} = ay_t + (1-a)S_{t-1}^{(1)} \quad (5)$$

$$S_t^{(2)} = aS_t^{(1)} + (1-a)S_{t-1}^{(2)} \quad (6)$$

$$S_t^{(3)} = aS_t^{(2)} + (1-a)S_{t-1}^{(3)} \quad (7)$$

式中: $S_t^{(1)}$ 为一次指数平滑值, $S_t^{(2)}$ 为二次指数平滑值, $S_t^{(3)}$ 为三次指数平滑值, a 为加权系数,且 $0 < a < 1$, a 实际上可理解为新旧数据权重的一个分配比例, a 值越大,则新数据在计算结果中占权重越大。

1.2.2 BP 神经网络模型

BP 神经网络模型是一种适用于非线性的模型识别和分类预测问题的人工神经网络^[10,11],文中的另一个子模型就是建立在 BP 神经网络基础之上的。

BP 神经网络预测模型的建立,主要考虑两个因素:神经网络层数和每层中神经元个数。

1) 神经网络层数。

对神经网络有一个非常重要的定理,即对于任何在闭区间内的一个连续函数都可以用单隐层的网络逼

近,因而一个 3 层网络就可以完成任意的 n 维到 m 维的映射,只要隐层中有足够多的单元可用。因此文中选取三层的 BP 神经网络就足够了^[12]。

2) 神经网络各层单元个数。

(1) 确定输入输出节点数。

输入、输出的节点数的确定是与具体应用紧密相关的。而流量预测的实质是,已知 m 个历史流量数据 $X_{t+1}, X_{t+2}, \dots, X_{t+m}$, 通过网络最终得到未来时刻流量数据 X_{t+m+1} 的预测值,所以 m 为输入节点数,输出节点数为 1。文中的模型是首先将样本序列分组,每 $m+1$ 个值为一组,前 m 个值作为网络的输入,后一个值作为输出节点的期望值。此模型取 m 值为 4,因此输入节点数为 4,输出节点为 1。

(2) 计算隐层节点数。

隐含层神经元数目既不能过多也不能过少,过多或过少都存在一些问题。因此在实际设计时,先根据一些常用的经验公式试探确定隐含层神经元个数,然后使用不同神经元个数对网络进行训练找出最优神经元数。隐含层神经元个数可通过以下公式确定: $n_1 = \sqrt{m+n} + a$, 其中 m 为输入神经元数, n 为输出神经元数, a 为 1~10 之间的常数。 $n_2 = \log_2 n$, 其中 n 为输入层神经元数^[9]。

1.3 模型的动态变化机制

由于网络流量组合预测模型在实际的网络系统中需要连续不断进行预测,但实际网络的流量规律会随时间而发生变化,这时根据历史数据得出的模型有可能变得不适用,从而导致流量预测的误差变大。因此,需要引进组合预测模型的动态变化机制,以反映最新网络流量的特征规律。文中组合预测模型的动态变化机制分为两层,第一层分别更新各子模型系数;第二层更新各子模型在组合预测模型的权重比例。1.2.1 已经介绍了三次指数平滑模型的动态更新方法,此处不再赘述。BP 神经网络模型在训练网络的过程本身就是不断地调整各种参数的过程。两个子模型的预测输出值是作为组合模型的两个输入自变量,因此两个子模型权重比例的变化,可以按照两个子模型的预测误差进行动态变化。

2 仿真实验与结果分析

2.1 实验工具及数据来源

文中利用 Matlab 这一工具进行仿真实验,神经网络工具箱中提供了丰富的网络学习和训练函数,其中包括了大多数的神经网络算法^[13]。数据来源于奥克兰大学,此组数据取 5 毫秒为一个时隙 TS (time slot) 对应一个数据,每个时隙点是进行测量的采样点,一共

3081 个数据,网络流量图如图 1 所示:

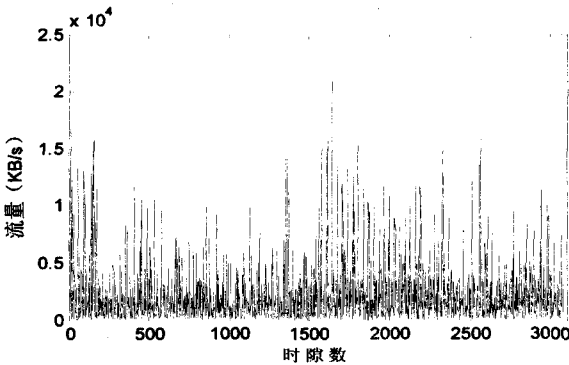


图 1 基于时隙的网络流量图

由于流量瞬时变化比较大,对于模型的预测及分析有影响,因此文中引进窗口(time window)的概念,每个窗口包含若干时隙,文中以 20 个时隙为一个窗口,每个窗口所对应的值为所有时隙数的平均值,基于窗口的流量图如图 2 所示。文中的实验就是基于窗口的流量进行预测。

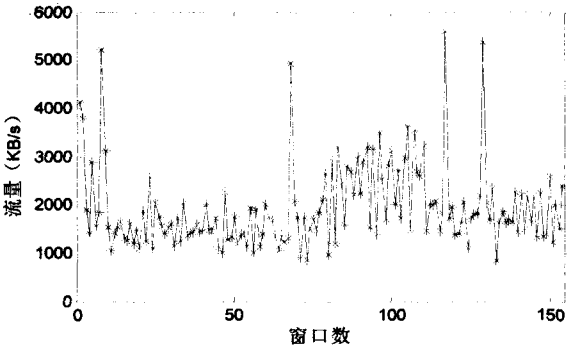


图 2 基于窗口的网络流量图

2.2 组合模型的实验结果

2.2.1 三次指数平滑预测模型

根据公式(1)~(7)采用指数平滑模型对基于窗口的流量进行预测,预测的初始值取前三个起始数据值的算术平均值,对 α 值进行反复实验不断更新,流量预测如图 3 所示:

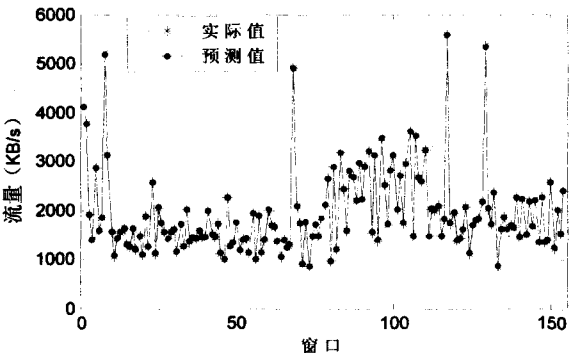


图 3 基于三次指数平滑的流量预测

2.2.2 BP 神经网络模型的流量预测

基于上述建模原理,采用三层网络结构,文中通过多次试算得出当隐层节点数为 7 时,网络收敛较好。

选择的隐层传递函数为 tansig,输出层的传递函数为 purelin,训练方法为 trainlm,最大训练次数为 1 000,训练目标为 10^{-6} 。

图 4 是实际值与预测值的比较图:

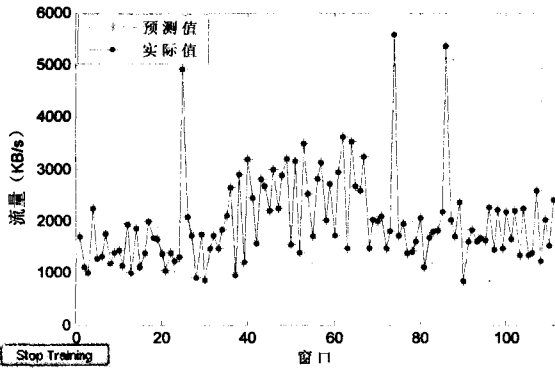


图 4 基于 BP 神经网络的流量预测

2.2.3 预测误差及权重系数

三次指数平滑模型和 BP 神经网络模型的预测误差如图 5 所示,设计组合模型权重系数调整的目的是:根据各个子模型的预测精度来自动调整各模型的权重比例,使得预测精度高的子模型的权重要大于其他子模型。从图中可以看出 BP 神经网络模型的预测精度总体上高于三次指数平滑模型。图 6 中两模型权重系数的变化也体现了这一过程,结合图 5 和图 6 可以看出模型的动态变化可以使精度高的模型权重更大,从而确保组合模型预测的可靠性。

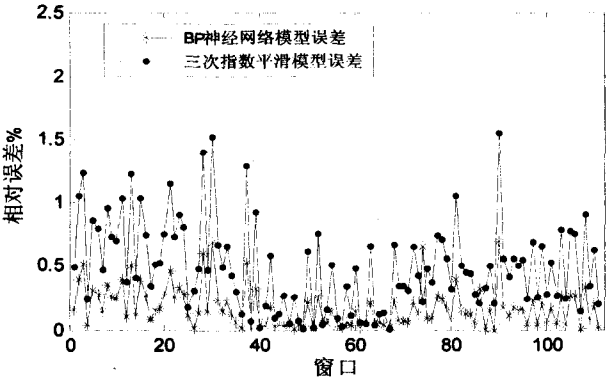


图 5 各子模型的预测误差变化曲线

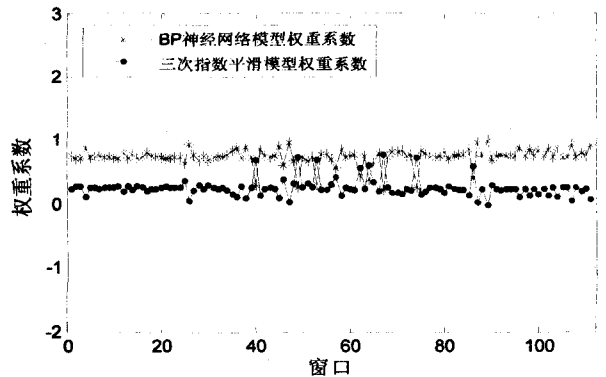


图 6 各子模型的权重比例变化曲线

3 结束语

为了更好地对网络流量进行建模和预测,文中提出了网络流量组合预测模型。该模型包含三次指数平滑模型和基于 BP 神经网络的预测模型,文中首先介绍了组合预测模型的预测原理、各个子模型实现细节以及模型系数的动态更新,最后仿真实验表明该组合预测能够达到较高的预测精度。

虽然预测模型预测结果精确合理,但是仍然存在一些不足之处,比如组合模型比较简单,只包含两个子模型,且两个子模型是一种简单的线性关系。因此还需要针对这些问题开展进一步的研究工作。

参考文献:

- [1] 董燕,孙恩昌,孙艳华,等.基于傅里叶模型的最佳网络流量预测[J].计算机应用研究,2010(4):1419-1421.
- [2] Paxon V, Floyd S. Wide Area Traffic: The Failure of Poisson Modeling[J]. IEEE/ACM Transactions on Networking, 1995(3):226-244.
- [3] Leland W E, Taqqu M S, Willinger W, et al. On the Self-Similar Nature of Ethernet Traffic(Extended Version)[J]. IEEE/ACM Transactions on Networking, 1994, 2(1):1-15.
- [4] Beran J, Sherman R, Taqqu M S, et al. Long-range dependence in variable-bit-rate video traffic[J]. IEEE Transaction on

Communications, 1995, 43(234):1566-1579.

- [5] Crovella M E, Bestavros A. Self-similarity in world wide web traffic - evidence and possible causes [C]//Proceedings of ACM Sigmetrics 96. [s. l.]:[s. n.], 1996:160-169.
- [6] 张帆. 城市交通网络流量预测方法比较[J]. 科技进步与对策, 2004(11):57-59.
- [7] Hippert H S, Pedreira C E, Souza R C. Neural networks for short-term load forecasting: a review and evaluation[J]. IEEE Transactions on Power Systems, 2001, 16(1):44-55.
- [8] 孙立光, 李瑞敏, 董岫, 等. 短时交通流组合预测模型研究[J]. 武汉理工大学学报(交通科学与工程版), 2010, 10(5):874-876.
- [9] 张丽霞, 张利红, 丰景春. 工程总承包企业违约风险分析及预测研究[J]. 科技管理研究, 2005, 25(9):209-211.
- [10] 程雪平, 林国龙. 基于三次指数平滑法的集装箱吞吐量预测[J]. 网络安全技术与应用, 2010(5):59-61.
- [11] 李歧强, 顾永如, 钱积新. 神经网络的具有自适应动量和步长的伪牛顿算法[J]. 信息与控制, 1998, 27(2):146-151.
- [12] Narendra K, Parthasarathy K. Identification and Control of Dynamical Systems Using Neural Networks[J]. IEEE Trans on Neural Networks, 1990, 1(1):4-27.
- [13] 蒋良孝, 李超群. 基于 BP 神经网络的函数逼近方法及其 MATLAB 实现[J]. 微型机与应用, 2004(1):52-53.

(上接第 110 页)

最终得到的汉化程序和俄文程序运行结果完全一致, 汉化程序完全可以取代俄文程序运行。

4 结束语

“某型导弹测试软件汉化研究”实现了引俄装备软件系统的全面汉化,形成了一套性能稳定、双语启动、操作方便的导弹测试系统。这一研究成果可以使操作人员更容易理解系统出现的提示信息,熟练地进行测试、分析故障和排除故障,从而克服保障设备操作环境不是母语的缺陷,操作人员只需要具备相应的专业知识就可以正确地执行操作,大量减少俄文环境下培训的时间,改变之前因为操作人员不熟悉俄语,依靠机械记忆位置、死记硬背流程来掌握保障装备的现状,可以降低对俄语的培训要求,缩短上岗培训的时间。该成果的研制具有广泛的实用价值和推广前景,对于我们国家的军队建设具有重大的经济和军事意义。

参考文献:

- [1] Nie Guangshu, Liu Hui. A Study of Aviation Weapon Equipment Maintenance Based on the Semi-Markov Model[J]. International Journal of Plant Engineering and Management, 2007, 12(2):101-106.

- [2] 彭荣木, 崔少辉, 陈胜, 等. 某型导弹测试系统的测试接口结构研究[J]. 兵工自动化, 2008, 27(6):22-24.
- [3] 王滨玲, 罗垂林. 俄文软件汉化技术[J]. 微计算机应用, 1995, 17(3):33-33.
- [4] Yan Shenggang, Yang Daoqing, Zhao Songzheng. The Development of a Knowledgeable Monitoring System for Weapon Equipment[J]. International Journal of Plant Engineering and Management, 1998(1):18-23.
- [5] 王家华, 程顺顺. 软件汉化技术的原理和应用[J]. 电脑知识与技术, 2011, 7(4):832-833.
- [6] 阮晓芬, 罗贤全. 嵌入式 Linux 系统的汉化[J]. 兵工自动化, 2006, 25(1):24-24.
- [7] 周镭, 李人贤. 软件汉化方法[J]. 哈尔滨师范大学自然科学学报, 1999(6):73-75.
- [8] 汤子瀛. 计算机操作系统[M]. 修订版. 西安:西安电子科技大学出版社, 2001.
- [9] Ehrenfeucht A, Haussler D, Rozenberg G, et al. On dos languages and dos mappings[J]. Semigroup Forum, 1993(1):123-148.
- [10] 小号. 浅谈 DOS 配置文件: Config. sys & Autoexec. bat[J]. 新潮电子, 1995(11):33-34.
- [11] 董占山. Windows 软件汉化的方法[J]. 电脑编程技巧与维护, 1995(5):62-64.