

基于多分类支持向量机的网络流量预测方法

王洪智¹, 刘震², 李东辉³

1. 大连交通大学网络信息中心, 大连 116028

2. 大连交通大学软件学院, 大连 116028

3. 大连交通大学电气信息学院, 大连 116028

摘要 针对支持向量机网络流量预测误差较大的问题, 提出一种基于多分类支持向量机的网络流量预测方法。该方法在网络流量数据训练阶段通过数据编码, 使多分类支持向量机的输出逼近编码值, 在预测阶段通过数据解码, 将多分类支持向量机的输出转换为实际的流量预测结果, 从而有效地降低了预测误差。实验结果显示, 该方法的预测结果与实际采集的网络流量数据具有相同的变化趋势; 在同等实验条件下, 该方法预测结果的均方根误差为 0.487, 而单一支持向量机方法、BP 神经网络方法预测结果的均方根误差分别为 1.0954 和 2.3642, 表明基于多分类支持向量机的网络流量预测方法具有更高的准确性。

关键词 网络流量预测; 多分类支持向量机; 编码; 解码

中图分类号 TP393

文献标志码 A

doi 10.3981/j.issn.1000-7857.2014.17.009

A Network Traffic Prediction Method Based on Multi-class Support Vector Machine

WANG Hongzhi¹, LIU Zhen², LI Donghui³

1. Network & Information Center of Dalian Jiaotong University, Dalian 116028, China

2. Software Institute of Dalian Jiaotong University, Dalian 116028, China

3. Electricity and Information School of Dalian Jiaotong University, Dalian 116028, China

Abstract Support vector machines (SVMs) have been used for network traffic prediction, but there often exist large prediction errors. This paper presents a network traffic prediction method based on multi-class support vector machine. Through establishing an encoding method in the training phase, the training output is binary encoded, and is established by one to one correspondence with the training input, training the multi-class support vector machine. By constructing a decoding method in the prediction phase, the prediction output is binary decoded, approaching the real value, realizing network traffic prediction and reducing the prediction errors. In simulation experiments, comparison of the real network traffic data and prediction results shows that they have the same evolution trends. Under the same condition, the accuracy of single SVM algorithm is about two times that of BP algorithm, while the accuracy of the proposed algorithm is about two times that of single SVM algorithm. The experiment results show that the proposed method has higher prediction accuracy in contrast with existing algorithms such as those of single SVM and BP neural network.

Keywords network traffic prediction; multi-class support vector machine; encoding; decoding

随着互联网的快速发展, 网络规模和网络复杂程度越来越大, 因此对于网络的优化、管理及安全提出了更高的要求。通过建立网络流量模型, 预测网络负载, 及时控制或调

整, 能够大幅提高网络性能和服务质量, 网络流量预测对于大规模网络规划、设计与管理显得越来越重要。网络流量时间序列是一个非线性、多时间尺度变化的动力系统, 真实网

收稿日期: 2013-06-03; 修回日期: 2014-04-22

作者简介: 王洪智, 高级工程师, 研究方向为网络与信息技术, 电子信箱: 1064776788@qq.com

引用格式: 王洪智, 刘震, 李东辉. 基于多分类支持向量机的网络流量预测方法[J]. 科技导报, 2014, 32(17): 60-63.

络环境中网络流量呈现出明显的多尺度特性,如分形、长程相关、自相似性、突发性等。传统的网络流量模型,如Markov模型、Poisson模型、自回归(AR)模型、自回归滑动平均模型、差分自回归求和滑动平均模型等,已经不能有效地刻画网络流量的这些特性。神经网络模型是基于经验风险最小原则,容易陷入局部最优从而导致泛化能力减弱,限制了其在实际网络流量预测中的应用。因此寻求新的模型方法成为当前网络流量行为研究的一个热点^[1]。

支持向量机(SVM)从1995年提出以来,以其小样本、全局最优解和使用简单等优势,已被广泛应用于预测领域。林楠等^[2]将相空间重构和支持向量机相结合,提出一种基于SVM的网络流量预测模型。Chen^[3]通过遗传算法优化神经网络结构和权重,实现Web流量的预测。白志中^[4]建立SVM回归预测模型用于网络流量预测。柴佳林等^[5]基于Net Flow协议,按照优化设计的组网方案进行数据流捕获,设计IP网络流量采集系统,提出一种基于支持向量机的网络流量预测方法。张颖璐^[6]将遗传算法优化参数支持向量机用于两组实际网络流量的预测。闵洁等^[7]将最小二乘支持向量机用于实际网络流量预测。曾勃炜等^[8]以融合蚁群算法优化支持向量机用于提高网络流量预测精度。Tan等^[9]将模拟退火支持回归机用于移动网络的流量预测。Li等^[10]构建冷却负载模型及与之相关的最小均方支持向量机用于提高空调冷却负载预测的精度。Wang等^[11]采用支持向量机估计动态系统模糊模型参数,实现移动网络的流量预测。Asif等^[12]采用V支持向量回归实现智能交通系统交通参数预测。Wei等^[13]通过支持向量回归实现网络流量预测。上述方法可分为两大类,一类采用单一支持向量机进行预测,其特点是将训练数据和预测数据归一化,但因实际数据的变化幅度比较大,归一化后预测稍有偏差,恢复时存在较大的预测误差;另一类采用支持向量机与其他方法结合实现预测,但多种方法融合的复杂度较高,计算工作量较大。多分类支持向量机^[8,14]主要用于解决识别中的多分类问题,由于其训练个数少,是进行网络流量预测的较好选择,但如何把多分类支持向量机用于网络流量预测,目前尚未见到相关文献。因此,本文提出一种应用多分类支持向量机的网络流量预测方法。

1 预测模型

多分类支持向量机网络流量预测模型如图1所示,主要包括网络流量数据、多个支持向量机、数据编码和数据解码输出4个部分。网络流量数据作为预测模型的输入,在训练阶段,假设用于训练的网络流量数据为 $Z = \{x_1, \dots, x_i, \dots, x_{N+1}\}$,具有 $N+1$ 个样本数据,其中 $x_i \in R^+$ ($i=1,2,\dots,N+1$), R^+ 为正实数; Z 中的相邻样本数据均为1个时间间隔(即网络流量数据单位采样时间)。训练输出数据为 $Y = \{x_1, x_2, \dots, x_N\}$,输入数据为 $X = \{x_2, x_3, \dots, x_{N+1}\}$,训练输出数据在训练阶段作为多分类支持向量机的参考,指导每个支持向量机学习,并对应于

多分类支持向量机的输出。

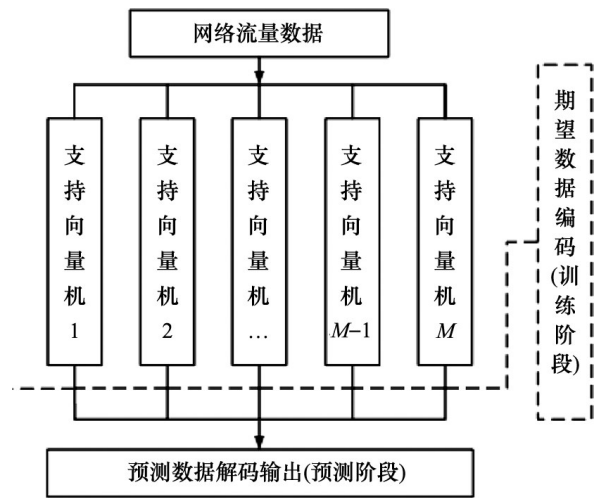


图1 多分类支持向量机网络流量预测模型

Fig. 1 Network traffic prediction model of multi-class SVM

但训练输出数据是实数,多分类支持向量机输出是一组二进制数,因此需要对训练输出数据进行编码。训练输出数据编码在训练阶段至关重要,其主要作用是将多分类支持向量机的输出向每个编码值逼近,从而确定每个支持向量机的权值和阈值。在训练阶段结束后保存每个支持向量机的权值和阈值,进入预测阶段,用新的一组网络流量数据进行预测,这组数据在多分类支持向量机的输出为一组二进制数,而希望输出实数,因此需要对这组二进制数进行解码,预测数据解码输出就是将多分类支持向量机的输出转化为实际的网络流量预测结果。

表1所示为采用9个支持向量机预测512个等级的数据编码表,其中“0”表示支持向量机的输出为-1,“1”表示支持向量机的输出为+1,-1和+1为每个支持向量机的输出值。表中第一行“000000000”表示9个支持向量机输出均为-1,对应于编码数据0(即将二进制数转换为十进制数),同理第二行“000000001”对应于编码数据1,……,最后一行“111111111”对应于编码数据511。

多分类支持向量机预测模型中的每个支持向量机分类器 $f(x)$ 的表述如式(1)所示。设训练集 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$,其中, n 为训练样本个数; $x_i \in R^l$ ($i=1,2,\dots,n$)为第 i 个训练样本, R^l 为实向量, l 为维数; $y_i \in \{+1, -1\}$ 为训练样本 x_i 的类标签。通过支持向量机分类器 $f(x)$ 可对训练集 D 进行分类,即

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^n y_i \alpha_i^* K(x_i, x) + b^* \right\} \quad (1)$$

式中,sgn为符号函数;当训练集 D 线性不可分时, $K(x_i, x)$ 为核函数;当训练集 D 线性可分时, $K(x_i, x)$ 为线性函数; α_i^* 和 b^* 为优化后的最佳解。

表1 多分类支持向量机数据编码

Table 1 Data encoding and decoding table of nine SVMs

编码数据	SVM1	SVM 2	SVM 3	SVM 4	SVM5	SVM 6	SVM7	SVM8	SVM9
0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	1
2	0	0	0	0	0	0	0	1	0
3	0	0	0	0	0	0	0	1	1
...
508	1	1	1	1	1	1	1	0	0
509	1	1	1	1	1	1	1	0	1
510	1	1	1	1	1	1	1	1	0
511	1	1	1	1	1	1	1	1	1

2 预测方法

2.1 数据编码与解码

在进行网络流量预测时,把网络流量数据分成两部分,一部分用于训练,另一部分用于预测,以检验预测方法的准确性。在训练阶段,需要对训练输出数据进行编码,使多分类支持向量机的输出逼近编码值,从而确定各支持向量机的参数。数据编码的具体方法:设网络流量数据的变化范围为 $[d_{\min}, d_{\max}]$,其中 d_{\min} 为网络流量数据的最小值, d_{\max} 为网络流量数据的最大值,通过式(2)对数据进行线性变换并四舍五入后取整,使其变化范围满足 $0 \leq d_i^* \leq 2^M - 1$,于是,在训练阶段 d_i^* 为训练输出数据,在预测阶段即为最终预测数据。正整数 M 为二进制的量化等级,也就是多分类支持向量机预测模型中支持向量机的个数。

$$d_i^* = \frac{x_i - d_{\min}}{d_{\max} - d_{\min}} \cdot 2^M \cdot \eta \quad (2)$$

式中, x_i 为网络流量数据; η 为比例因子,为大于0的正实数。

在网络流量的预测阶段,多分类支持向量机的输出数据为二进制数,需要通过解码将输出数据转换为实际的预测值,解码原理与编码过程相反。设多分类支持向量机的输出数据为 u ,则经过解码后的实际预测数据为

$$y = \frac{u}{2^M - 1} (d_{\max} - d_{\min}) + d_{\min} \quad (3)$$

经过解码后,输出 y 即为实际预测输出数据。

2.2 网络流量预测步骤

1) 根据网络流量数据的变化范围 $[d_{\min}, d_{\max}]$,确定支持向量机个数 M ,设定各个支持向量机的初始权值和阈值。

2) 将训练数据通过式(2)变换为 d_i^* ,准备送入 M 个支持向量机的输入端。

3) 对每个 d_i^* 按照表1进行数据编码,为训练各个支持向量机提供标识。

4) 把每个 d_i^* 依次送入多分类支持向量机进行训练,待所有训练数据都训练结束后,保存各个支持向量机的最终权值

和阈值。

5) 将新采集到的网络流量数据,送入多分类支持向量机进行预测。

6) 对 M 个支持向量机的输出,根据式(3)进行数据解码,得到预测结果。

3 预测实验及分析

预测实验采用的仿真环境为 Matlab 2009,计算机配置为 2.2 GHz CPU、2 GB 内存、Windows XP 操作系统,数据通过在线采样得到。在 2012 年 10 月 15 日 8:00 至 10 月 22 日 21:00,对某高校校园网总出口流量进行监测采样,采样时间间隔为 1 h。将采样数据分为两部分,前部分数据用于训练,后部分数据用于预测。选择 5 个时间间隔的连续流量数据作为输入,选择支持向量机个数为 $M=9$, $[d_{\min}, d_{\max}] = [5, 324]$, $\eta = 1.5772$ 。实际流量数据如图 2 所示,预测流量数据如图 3 所示。通过对图 2 和图 3 同时段的分析对比可以看出,预测数据与实际采集的网络流量数据具有相同的变化趋势,8 个主峰基本相同,因此预测结果体现了网络流量的实际变化规律。图 2 中的尖峰显示预测误差增大,如 30~40 h 的小尖峰以及 85 h 附近的尖峰,这主要与数据的快速变化和学习机的自适应预测能力有关。

为了进一步验证多分类支持向量机预测方法的优越性,分别对照单一支持向量机方法和 BP 神经网络方法,其中单一支持向量机数据集的延迟时间和嵌入维均为 10, BP 神经网络输入神经元个数为 12, 1 个隐含层中隐含的神经元个数为 8, 输出 1 个神经元。与对比方法一样,多分类支持向量机网络流量预测结果最后归一化至 $[0, 1]$, 以观测值和预测值的均方根误差 (MSE, minimum squared error) 作为最终衡量。实验结果显示,多分类支持向量机预测方法的 MSE 为 0.487, 单一支持向量机方法的 MSE 为 1.0954, BP 神经网络方法的 MSE 为 2.3642, 说明运用多分类支持向量机方法对网络流量的预测具有更高的准确性。

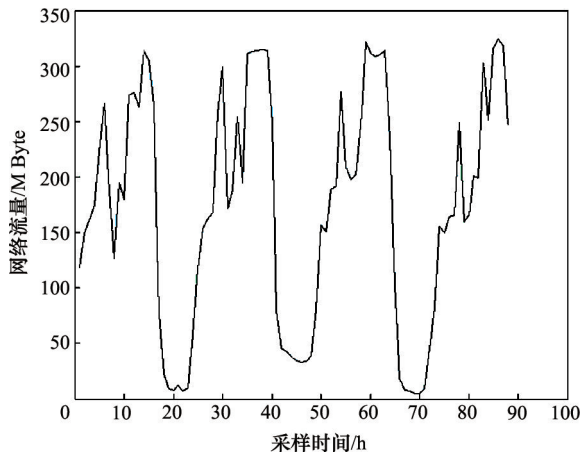


图2 实际流量数据

Fig. 2 Actual traffic data

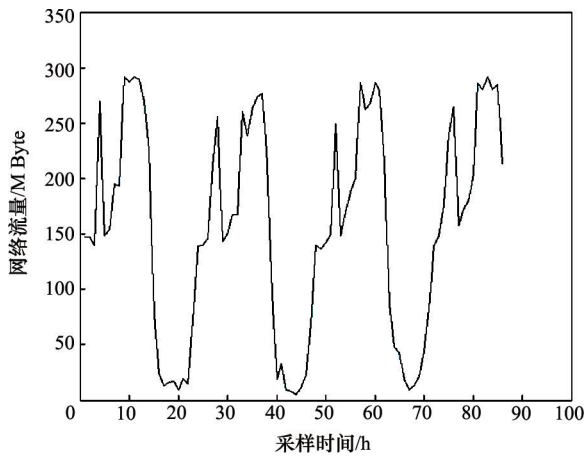


图3 预测流量数据

Fig. 3 Predicted traffic data

4 结论

针对目前采用支持向量机进行网络流量预测存在归一化后预测有偏差、恢复时带来较大预测误差的问题,本文提出多分类支持向量机网络流量预测方法,在训练阶段建立数据编码方法,在预测阶段建立数据解码方法,有效地提高了预测精度。实验结果表明,与单一支持向量机预测方法和BP神经网络预测方法相比,本文方法具有更高的预测精度。

参考文献(References)

- [1] 林楠,李翠霞. SVM在非线性网络流量预测中的应用研究[J]. 计算机仿真, 2011, 28(5): 259-262.
Lin Nan, Li Cuixia. Study on nonlinear network traffic prediction based on support vector machine[J]. Computer Simulation, 2011, 28(5): 259-262.
- [2] Chen M. Short-term forecasting model of web traffic based on genetic algorithm and neural network[C]//2011 2nd International Conference on Artificial Intelligence. Las Vegas, Nevada, USA: Management Science

and Electronic Commerce, 2011: 623-626.

- [3] 白志中. 采用支持向量机的网络流量预测研究[J]. 计算机与信息技术, 2009, 193(10): 45-49.
Bai Zhizhong. Research on network traffic prediction using support vector machine[J]. Computer & Information Technology, 2009, 193(10): 45-49.
- [4] 柴佳林, 简银, 刘兴伟. 一种基于支持向量机的IP网络流量预测方法[J]. 西华大学学报: 自然科学版, 2010, 29(1): 54-57.
Chai Jiailin, Jian Yin, Liu Xingwei. Prediction of ip network traffic based on support vector machine[J]. Journal of Xihua University: Natural Science Edition, 2010, 29(1): 54-57.
- [5] 张颖璐. 基于遗传算法优化支持向量机的网络流量预测[J]. 计算机科学, 2008, 35(5): 177-181.
Zhang Yinglu. Internet traffic forecasting based on support vector machine optimized by genetic algorithm[J]. Computer Science, 2008, 35(5): 177-181.
- [6] 闵洁, 李潇. 基于最小二乘支持向量机的网络流量预测[J]. 科技创新导报, 2010, 3(1): 206-207.
Min Jie, Li Xiao. Internet traffic forecasting based on least squares support vector machine[J]. Science and Technology Innovation Herald, 2010, 3(1): 206-207.
- [7] 曾勃伟, 徐知海, 付爱英, 等. 融合蚁群算法与支持向量机的网络流量预测[J]. 南昌大学学报, 2011, 35(4): 406-408.
Zeng Qingwei, Xu Zhihai, Fu Aiyong, et al. Prediction of network traffic by combination of ant colony optimization and support vector machine [J]. Journal of Nanchang University, 2011, 35(4): 406-408.
- [8] 苟博, 黄贤武. 支持向量机多类分类方法[J]. 数据采集与处理, 2006, 21(3): 334-339.
Gou Bo, Huang Xianwu. Svm multi-class classification on[J]. Journal of Data Acquisition & Processing, 2006, 21(3): 334-339.
- [9] Tan Y, Qin X, Jia Z, et al. An improved method of traffic forecasting based on Tariff- SASVR[C]//2009 Fifth International Conference on Natural Computation. Tianjin, China: IEEE Computer Society, 2009: 463-467.
- [10] Li X, Lu J, Ding L, et al. Building cooling load forecasting model based on LS-SVM[C]//2009 Asia-Pacific Conference on Information Processing. Shenzhen, China: Information Processing APCIP, 2009: 55-58.
- [11] Wang J, Peng Y, Peng X. Mobile communication traffic forecast based on a new fuzzy model[C]//IEEE International Instrumentation and Measurement Technology Conference I2MTC'09 IEEE. Singapore: Instrumentation and Measurement Technology Conference, 2009: 872-877.
- [12] Asif M T, Dauwels J, Goh C Y, et al. Unsupervised learning based performance analysis of n-support vector regression for speed prediction of a large road network[C]//2012 15th International IEEE Conference on Intelligent Transportation Systems. Anchorage, USA: Intelligent Transportation Systems (ITSC), 2012: 983-988.
- [13] Wei X. Supporting vector-machine prediction of network traffic[C]//2011 International Conference on Electrical and Control Engineering. Yichang, China: Electrical and Control Engineering (ICECE), 2011: 3203-3206.
- [14] Wang R, Kwong S, Chen D. A new method for multi-class support vector machines by training least number of classifiers[C]//Proceedings of the 2011 International Conference on Machine Learning and Cybernetics. Melbourne, Australia: Machine Learning and Cybernetics (ICMLC), 2011: 648-659.

(责任编辑 韩星明)