



基于自回归模型与神经网络模型的车流量预测对比

张婷婷^{1,2}, 张武雄^{1,3}, 裴冬^{1,2}, 赵铖^{1,2}, 俞涵^{1,2}

(1. 中国科学院上海微系统与信息技术研究所, 上海 200050;

2. 上海科技大学, 上海 201210; 3. 上海无线通信研究中心, 上海 201210)

摘要: 车流量建模是车联网 (vehicular Ad Hoc network, VANET) 路由、多媒体接入协议、无线算法设计的基础。准确的车流量模型将对智能交通系统 (intelligent transportation system, ITS) 实时调度和车联网的信息安全起到十分重要的作用。基于上海市的交通流量数据, 利用自回归 (auto regressive, AR) 模型与神经网络模型 (back-propagation, BP) 对车流量实测数据进行了仿真对比, 给出了相应的预测结果。研究发现, 两个模型均能有效地对数据进行跟踪与预测, 但对不同时段数据预测的准确性有所不同。研究结果将为未来智能交通应用、车联网的理论研究等提供有力依据。

关键词: 车联网; 智能交通; 自回归模型; 神经网络模型; 交通流量预测

中图分类号: TN929.53

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2016049

Comparison of traffic flow prediction based on AR model and BP model

ZHANG Tingting^{1,2}, ZHANG Wuxiong^{1,3}, PEI Dong^{1,2}, ZHAO Cheng^{1,2}, YU Han^{1,2}

1. Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Science, Shanghai 200050, China

2. Shanghai Tech University, Shanghai 201210, China

3. Shanghai Research Center for Wireless Communications, Shanghai 201210, China

Abstract: Traffic flow modeling plays an important role in routing, MAC algorithm and protocol designs in vehicular Ad Hoc networks (VANET). An accurate traffic flow model is crucial to traffic management of an intelligent transportation system (ITS) and information safety in a VANET. Based on Shanghai's traffic flow data, the performance of the two different models was compared using auto regressive (AR) model and back-propagation (BP) network model, and the corresponding prediction result was given. Research finds that both of the two models can efficiently predict the traffic data, but they have different prediction accuracy for the data of different periods. The research result will provide support for future research on ITS and VANET.

Key words: VANET, ITS, AR model, BP network model, traffic flow predicting



1 引言

汽车已成为人们出行的重要交通工具。《2014 年国民经济和社会发展统计公报》统计数据显示,2014 年末,全国民用汽车保有量达 15 447 万辆(包括三轮汽车和低速货车),较 2013 年末增长了 12.4%,其中私人汽车保有量达 12 584 万辆,增长了 15.5%;民用轿车保有量达 8 307 万辆,增长了 16.6%,其中私人轿车 7 590 万辆,增长了 18.4%^[1]。《2014 年交通运输行业发展统计公报》统计数据显示,2014 年末,我国公路总里程达 446.39 万公里,较 2013 年末增加了 10.77 万公里;公路密度为 46.50 公里/百平方公里,提高了 1.12 公里/百平方公里,其中全国高速公路里程达 11.19 万公里,较 2013 年末增加了 0.75 万公里^[2]。中国公路和车辆的迅速增长给智能交通系统(ITS)和车联网(VANET)带来了无比广阔的市场前景。

ITS 是未来交通系统的发展方向,它将先进的信息技术、数据通信传输技术、电子传感技术、控制技术以及计算机技术等集成运用于整个地面交通管理系统,从而构成实时、准确、高效的综合交通运输管理系统。ITS 可以有效地利用已有的交通设施,减少交通带来的环境污染,保证交通安全,提高运输效率等,正越来越多地受到世界各国的重视。

VANET 主要应用于高速公路或者行驶车道上关于车辆安全信息和行驶环境信息发布的数据控制和数据通信。作为辅助 ITS 的一项重要功能,VANET 通过实时发布车辆的行驶环境信息,针对未来车流量进行预测,以避免道路堵塞。车流量信息的准确预测对于减少意外事故、缓解交通压力、提高道路网络的畅通能力、减少环境污染等,均起着重要作用。

对交通流量的建模,已有人做了部分研究,例如,参考文献[3]主要讲述了高速公路中诱导交通量变化的参数及模型设定,提出了使用曲线模型预测诱增交通量,它主要针对交通量的增加做了预测,但并未对同一地区不同时间段的交通量做出实时预测;参考文献[4,5]重点对时间序列模型进行了分析,虽然参考文献[5]提到了基于 AR 模型的车流量分析,但并未提及与 BP 模型所做预测准确性的对比;参考文献[6]重点阐述了 AR 模型的原理和参数估计的 MATLAB 仿真,但并未对实测的交通量进行验证。与这些研究不同,本文采集了上海市大量的交通流量数据,并利用 AR 模型和 BP 模型分别进行了建模,利用 MATLAB 进行

了仿真对比,结果发现,两种模型对数据的预测各有优缺点。本文所述之预测方法与结果将对车联网的研究与智能交通的应用提供支持。

目前道路车流量检测手段主要有 3 种:埋入地下的线圈检测、超声波检测和图像处理技术检测车辆的计算机处理系统。本文所用数据来源于 2012 年上海市交通信息中心所测的真实数据,主要手段是结合埋入地下的线圈和视频监控系统。对获得的数据进行了整理,得到了每半个小时通过的车流量数据。

对车流量的预测方法有很多,主要分为线性预测和非线性预测两大类。其中线性预测以 ARIMA 自回归积分滑动平均模型 (autoregressive integrated moving average model) 最为著名,它包括移动平均(MA)过程、自回归(auto regressive, AR)过程、自回归移动平均 (ARMA) 过程以及 ARIMA 过程。本文重点讨论自回归过程 AR 模型的预测方法,非线性部分主要基于 BP 网络,文章基于真实数据对两种方法进行了讨论和比较,以期获得更精准的预测方法。

2 AR 与 BP 网络预测

AR 模型是一种线性预测,通过已知的 N 个点可以预测出未来 P 个点的值;AR 模型能准确分析出各个变量间的相关性,检验回归方程拟合程度,提高预测效果,对于统一标准的数据其输出结果唯一^[7]。BP 模型是一种非线性预测模型,通过对已知数据不同程度的学习,对数据进行预测;BP 模型具有很好的容错性和顽健性,具有非线性映射特性,同时可以集成定性数据和定量数据^[8]。

2.1 AR 时间序列模型

ARIMA 自回归积分滑动平均模型是 20 世纪 70 年代初由 Box 和 Jenkins 提出的著名的时间序列预测方法^[9],故又称之为 Box-Jenkins 模型。ARIMA 模型是指将非平稳时间序列转化为平稳时间序列,然后因变量仅对它的滞后值以及随机误差项的现值和滞后值进行回归所建立的模型。使用 $ARIMA(p, d, q)$ 描述差分自回归移动平均模型,其中 AR 是自回归, p 为自回归项, MA 为移动平均, q 为移动平均项数, d 为时间序列成为平稳时所做的差分次数。ARIMA 模型根据原序列是否平稳以及回归中所含部分的不同,包括 MA 过程、AR 过程、ARMA 过程以及 ARIMA 过程。

AR 模型是描述动态模型的一种常见模型结构。其数学表达式可以表述为^[9]:

$$y(k) = \sum_{i=1}^n a_i(k)y(k-i) + e(k) \quad (1)$$

其中, $e(k)$ 为预测的误差, 其均值 $E[e(k)] = 0$, 方差 $Var[e(k)] = \sigma^2$ 。将式(1)改写成完整的形式, 如式(2)所示:

$$y(k) = h_{k-1}^T \theta(k) + e(k) \quad (2)$$

h_{k-1}^T 和 $\theta(k)$ 如下所示:

$$h_{k-1}^T = [y(k-1) y(k-2) \cdots y(k-n)]^T$$

$$\theta(k) = [a_1(k) \cdots a_n(k)] \quad (3)$$

其中, AR 模型参数估计是对 $\theta(k)$ 这 n 个参数的估计。参数估计的方法主要有最小二乘法、解 Yule-Walker 方程法、LUD 法、BSMF 等。其中最小二乘法估计简单且参数估计无偏、精度高。本文采用最小二乘法对 $\theta(k)$ 进行参数估计, 如式(4)所示, 然后选取合适的模型适用性检验标准, 如 Akiake 信息检验标准 FPE (final prediction error)、AIC (an information criterion)、BIC (bayesian information criterion) 准则等^[9], 再将 $\hat{\theta}(k)$ 代回到式(1)中, 即可完成对后续车流量的预测工作。

$$\hat{\theta}(k) = (h_{k-1}^T h_{k-1}^T)^{-1} (h_{k-1}^T y(k)) \quad (4)$$

上述使用 AR 模型进行预测的方法步骤可以总结为: 先对样本进行预处理, 对非平稳时间序列进行差分、消除趋势项等操作, 变为平稳时间序列; 其次进行相应的参数估计, 对模型进行检验, 得到时间序列的初步模型; 最后进行预测。

2.2 BP 神经网络模型

Rumelhart、McClelland 等人^[8]在 1982 年成立了 PDP 小组, 研究并行分布信息处理方法, 探索人类认知的结构, 1985 年发展了 BP 网络 (back-propagation network)^[9]。BP 网络核心思想主要集中在以下两方面: 正向传播过程, 给出输入信息, 经过输入层到隐含层逐层处理并计算每个单元的实际输出值; 反向传播过程, 若在输出层未能得到期望输出值, 则逐层递归地计算实际输出与期望输出之差, 即误差, 以便据此差值调节权重, 使误差减少。这两个过程的反复运用使误差信号逐渐减少, 直至误差达到要求时, 网络的学习过程才结束。BP 网络结构模型如图 1 所示。

同其他人工神经网络类似, 神经元是组成神经网络的重要组成部分, 其网络结构中神经元模型如图 2 所示。根据 BP 算法的需求, 必须采用处处可导的激活函数, S 型函数是最为常用的, 可将神经元的网络输入表述如下:

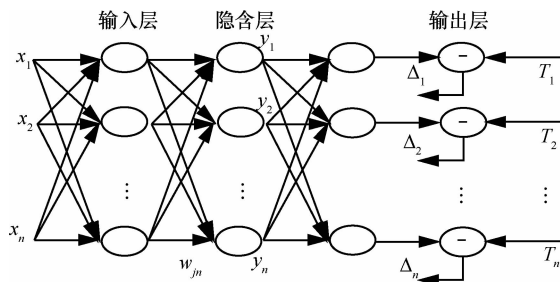


图1 BP神经网络结构模型

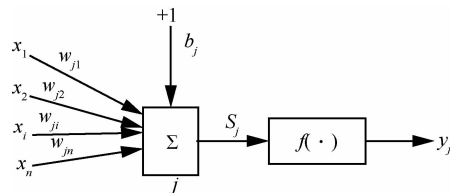


图2 每一个神经元的模型

$$net_j = x_1 w_{j1} + x_2 w_{j2} + \cdots + x_i w_{ji} + \cdots + x_n w_{jn} \quad (5)$$

其中, x_1, x_2, \cdots, x_n 是图 1 中该神经元的输入, $w_{j1}, w_{j2}, \cdots, w_{jn}$ 是对应输入的权重。因此, 神经元对应的输出是:

$$y_j = f(net) = \frac{1}{1 + e^{-net}} \quad (6)$$

简单来说, 上述两个核心思想可理解为 BP 网络分为 3 层: 输入层、隐含层和输出层。每个层之间、每两个节点之间都有一个权值 w , 每一个节点都有一个阈值 b , 并且还有一个激活函数 $f(\cdot)$, 将神经元组合起来, 便可得到整个 BP 网络的模型, 如图 1 所示。

综上所述, 可将 BP 网络的构建分为样本选取、网络结构设计 (即确定网络层数的各项参数)、网络拓扑设计 (主要是网络输出层的激活函数)、网络初始化、网络训练和网络预测 6 部分。其中网络结构设计对 BP 网络的预测精准性起着至关重要的作用, 而网络层数的选择主要有如下两个依据。

(1) Kolmogorov 定理

给定任意连续函数 $f: U^n \rightarrow R^m$, $f(X) = Y$, 此处 U 是闭单位区间 $[0, 1]$, f 可以精确地用一个三层前向网络实现, 此网络的第一层 (即输入层) 有 n 个神经元, 中间层有 $2n+1$ 个神经元, 第三层 (即输出层) 有个 m 神经元。

(2) BP 定理

给定任意 $\varepsilon > 0$ 和任意函数 $f: [0, 1]^n \rightarrow R^m$, 存在一个三层 BP 网络, 它可在任意 ε 平方误差精度内逼近 f 。

根据 Kolmogorov 定理和 BP 定理, 选取 $5 \times 11 \times 1$ 的网络结构, 激活函数为 transig 函数, 学习速率为 0.7, 标准误差为 0.01, 开展 BP 模型的车流量预测 MATLAB 仿真。



3 MATLAB 仿真结果

对 AR 模型和 BP 模型进行仿真,得到如图 3 所示的结果,BP 预测输出和 AR 预测输出这两条线都能很好地与期望输出相匹配,即 AR 模型和 BP 模型都可有效预测实际车流量的数值。

表 1 描述了 AR 模型的误差和 BP 模型与真实车流量数据的误差,根据平均误差的计算式(7),可以计算得到 AR 模型的平均误差为 0.008、BP 模型的平均误差为 0.118 6。使用式(8)可以计算得到其均方误差,AR 模型均方误差为

0.014 8、BP 模型的均方误差为 0.005 6。

$$E(y)=\frac{y(k+1)+\cdots,y(k+i)+\cdots,y(k+n)}{n} \tag{7}$$

$$D(y)=\frac{(y(k+1)-M)^2+\cdots,(y(k+i)-M)^2+\cdots+(y(k+n)-M)^2}{n} \tag{8}$$

从表 1 可以看出,10:00-16:00 这个时间段,AR 模型明显优于 BP 模型,这是因为这个时间段的车流量数据具有一定的线性相关性,车流量变化的数量级不是特别明显;18:00-21:00 时间段,BP 模型的误差明显小于 AR 模型,这是因为此时公路上的车流量已经不如白天的车流量

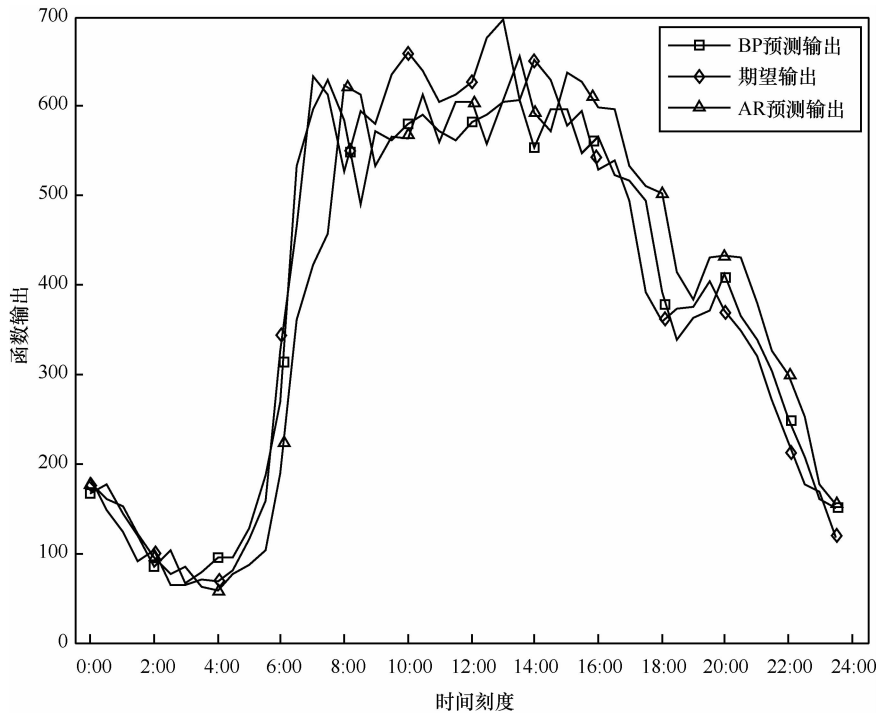


图 3 基于 BP-AR 模型的车流量预测

表 1 基于 BP 模型和 AR 模型的误差比较

时间	BP	AR	时间	BP	AR	时间	BP	AR	时间	BP	AR
0:00	0.16	0.02	6:00	0.16	0.19	12:00	0.10	0.01	18:00	0.09	0.28
0:30	0.11	0.14	6:30	0.15	0.11	12:30	0.16	0.11	18:30	0.03	0.15
1:00	0.12	0.02	7:00	0.07	0.09	13:00	0.17	0.10	19:00	0.11	0.03
1:30	0.20	0.02	7:30	0.03	0.28	13:30	0.03	0.06	19:30	0.01	0.15
2:00	0.20	0.04	8:00	0.09	0.02	14:00	0.18	0.02	20:00	0.04	0.07
2:30	0.38	0.25	8:30	0.23	0.16	14:30	0.07	0.12	20:30	0.04	0.16
3:00	0.05	0.31	9:00	0.03	0.10	15:00	0.03	0.02	21:00	0.09	0.09
3:30	0.10	0.05	9:30	0.14	0.03	15:30	0.06	0.08	21:30	0.09	0.02
4:00	0.28	0.18	10:00	0.16	0.11	16:00	0.02	0.01	22:00	0.09	0.11
4:30	0.15	0.11	10:30	0.10	0.07	16:30	0.04	0.13	22:30	0.10	0.12
5:00	0.12	0.08	11:00	0.09	0.12	17:00	0.19	0.01	23:00	0.12	0.01
5:30	0.19	0.11	11:30	0.11	0.002	17:30	0.07	0.03	23:30	0.18	0.09

数据那样有很强的相关性;其他时间段两者的预测结果不相上下,这可能跟夜晚车流量变化较大有关。

从以上分析可知,预测数据的准确性与选择的模型密切相关,但总体而言,AR模型和BP模型都具备对车流量进行预测的能力。

4 结束语

AR模型与BP模型都能很好地预测未来车流量的变化情况,但从引入的平均误差而言,AR模型具有更好的预测能力;从均方误差而言,BP模型的误差相对更集中。因此,针对不同的车流量数据特点,需权衡选用合适的模型。本文所述研究与仿真结果将为今后VANET的车流量预测提供理论和结果依据。

参考文献:

- [1] 中华人民共和国国家统计局. 2014年国民经济和社会发展统计公报 [R/OL]. (2015-02-26)[2015-02-26]. http://www.stats.gov.cn/tjsj/zxfb/201502/t20150226_685799.html.
National Bureau of Statistics of China. Statistical Communiqué of the People's Republic of China on the 2014 National Economic and Social Development[R/OL]. (2015-02-26)[2015-02-26]. http://www.stats.gov.cn/tjsj/zxfb/201502/t20150226_685799.html.
- [2] 中华人民共和国交通运输部. 2014年交通运输行业发展统计公报 [R/OL]. (2015-04-30)[2015-04-30]. http://www.moc.gov.cn/zfxgk/bnssj/zhghs/201504/t20150430_1810598.html.
Ministry of Transport of the People's Republic of China. Statistical bulletin of transportation industry development in 2014[R/OL]. (2015-04-30)[2015-04-30]. http://www.moc.gov.cn/zfxgk/bnssj/zhghs/201504/t20150430_1810598.html.
- [3] 向前忠. 生长曲线模型在高速公路诱增交通量预测中的应用[J]. 公路交通技术, 2007(2):161-163.
XIANG Q Z. Application of growth curves model in expressway induced traffic forecast [J]. Technology of Highway and Transport, 2007(2):161-163.
- [4] BOX G E P, JENKINS G M, REINSEL G C. Time series analysis: forecasting and control [J]. Journal of Marketing Research, 1994, 14(2):556-569.
- [5] WANG Y, WU X. An adaptive glucose prediction method using auto-regressive (AR) model and Kalman filter[C]//IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), January 5-7, 2012, Hong Kong, China. New Jersey: IEEE Press, 2012:556-569.
- [6] 陈国强, 赵俊伟, 黄俊杰, 等. 基于Matlab的AR模型参数估计[J]. 工具技术, 2005, 39(4):39-40.
CHEN G Q, ZHAO J W, HUANG J J, et al. Matlab-based parameter estimation of AR model [J]. Tool Engineering, 2005, 39(4):39-40.
- [7] JUVA I, SUSITAIVAL R, PEUHKURI M, et al. Traffic characterization for traffic engineering purposes: analysis of funet data[J]. Next Generation Internet Networks, 2005:404-411.
- [8] RUMELHART D E, MCCLELLAND J L. The PDP research group parallel distributed processing [J]. IEEE, 1998 (1): 443-453.
- [9] FORBES G J, HALL F L. The applicability of catastrophe theory in modelling freeway traffic operations [J]. Transportation Research Part A: General, 1990, 24(5):335-344.

[作者简介]



张婷婷(1991-),女,中国科学院上海微系统与信息技术研究所硕士生,主要研究方向为车联网关键技术。



张武雄(1985-),男,博士,中国科学院上海微系统与信息技术研究所、上海无线通信研究中心助理研究员,一直从事车联网/新一代宽带无线通信网络关键技术的研究,主持国家自然科学基金面上项目“基于车流量与3D信道建模的车联网连接性理论研究”课题、上海市科学技术委员会自然科学基金关于车联网理论研究课题、负责“面向南水北调工程安全的传感器网络技术研发”重大专项子任务的关键技术研究,参与国家重大专项、国家自然科学基金重点项目、上海市科学技术委员会/上海市经济和信息化委员会重点项目十余项。目前主要从事蜂窝无线移动通信标准、车联网体系架构及组网技术、多网异构协作关键技术研究工作。发表学术论文24篇,其中SCI/EI索引19篇,申报国内发明专利5项、国际PCT专利1项。

裴冬(1989-),男,中国科学院上海微系统与信息技术研究所硕士生,主要研究方向为编码调制技术。

赵铖(1991-),男,中国科学院上海微系统与信息技术研究所硕士生,主要研究方向为Wi-Fi网络下的干扰消除。

俞涵(1992-),女,中国科学院上海微系统与信息技术研究所硕士生,主要研究方向为下一代移动通信技术。