

短时车流量组合预测模型

张凯¹ 卢邹颖¹

摘要

随着道路车量不断增多,由交通异常事件造成的非正常拥堵情况严重影响了出行者的出行效率和路网的整体运行水平.因此,需要准确及时地检测出非正常拥堵情况,通过诱导、疏通等方式改善拥堵状况.对车流量的准确预测是检测非正常拥堵的有效方法.根据交通流量的不确定性和非线性的特点,将改进的BP神经网络模型和ARIMA模型进行组合,建立组合预测模型.实验结果表明,组合模型的预测结果比单个模型的预测结果理想,且达到较高的预测精度.

关键词

车流量预测;BP神经网络;差分自回归移动平均模型;组合预测

中图分类号 U491

文献标志码 A

0 引言

当前,我国机动车拥有量和道路交通量急剧增加,常常导致道路交通拥堵,而由交通事故等造成的非正常拥堵所造成的危害巨大,如不及时处理,容易导致拥堵的恶性循环.已有的非正常拥堵检测方法大多集中在对短时车流量的预测问题上.国外学者对短时交通流预测的研究主要有历史均值法^[1]、Kalaman 滤波法^[2]、非参数回归法^[3]和神经元网络法^[4]等.其中,历史均值法算法简单,但无法研究交通流动态性、非确定性和非线性特性;Kalaman 滤波法对交通流预测精度较高,但受线性估计模型限制,无法适应交通流的非线性变化;非参数回归法适合非线性的动态系统,并符合交通流量的非线性特性,需要大量的历史数据,建立序列内部联系,用来预测当前时间之后的交通流信息,但存在预测速度慢,参数调整需要试凑的缺点;神经网络在交通预测中的应用,解决了传统方法需要建立精确模型的困扰,但也存在训练速度慢、易陷入局部极小点的不足.在国内,对交通流预测的研究大多是在对以上各种方法的改进基础之上的.其中,曹征^[5]利用小波变换对短时交通流进行了预测;郇洪江等^[6]利用BP神经网络对交通流量进行了预测.这些方法大多是以单一模型来做预测,预测精度不高,对交通流的不确定性的考虑还有所欠缺.

短时车流量变化具有不确定性和非线性特点,而时间序列和神经网络技术正好能较好地解决交通流预测的这些特性所带来的预测难题.本文采用基于遗传算法^[7]优化的BP神经网络模型和ARIMA模型(差分自回归移动平均模型)^[8]组合的方法对交通流进行预测,改进的BP网络模型解决了传统BP网络收敛速度慢,容易陷入局部极小点的缺点.将训练成功的网络用于预测实验,结果表明此方法具有较高的预测精度,对交通流预测和非正常拥堵的检测的实际应用有很好的参考价值.

1 模型建立

1.1 BP神经网络

BP神经网络是一种按误差逆传播算法训练的多层前馈网络,其基本思想是:向逐层计算得到的网络输出与原先给定的期望输出进行误差计算,若误差不满足要求,则将误差延网络反向逐层传播,修改连接权重,重复正向计算和反向传播的过程,直到满足误差精度.

收稿日期 2012-03-22

作者简介

张凯,男,博士,教授,研究方向为智能检测. zkark@nuist.edu.cn

¹ 南京信息工程大学 信息与控制学院,南京, 210044

由于传统的 BP 神经网络采用随机初始化网络权值、阈值,收敛速度较慢,易陷入局部极小点,而遗传算法对求解的寻优性能好,全局搜索性强的特点正好能弥补 BP 神经网络的不足。

1.2 遗传算法

遗传算法是一种具有内在的隐并行性和更好的全局寻优能力的迭代算法,采用概率化的寻优方法,能自动获取和指导优化的搜索空间,自适应地调整搜索方向,不需要确定的规则。其算法的核心思想是:模拟自然选择和遗传发生的繁殖、交配和突变现象,从初始种群出发,根据适应度函数计算出适应度值,通过选择、交叉和变异这 3 个操作,获得更适应环境的个体,经过多次迭代,使种群往搜索控件中的最优解区域靠近,最终收敛到满足要求的最优解。

1.3 遗传算法优化 BP 神经网络权值模型

BP 神经网络算法具有寻优精确,但寻优速度慢的特点,而遗传算法具有全局、快速搜索最优解的性能,因此,把遗传算法和 BP 神经网络结合起来,可以达到全局、高效、精确的寻优效果。

BP 神经网络中,初始权值和阈值都影响最终的解,通过遗传算法优化 BP 神经网络的权值,具体步骤如下:

- 1) 给定神经网络的输入、输出样本集;
- 2) 将网络权值(阈值)进行编码,随机生成一组分布,对应一组神经网络的权值(阈值);
- 3) 输入训练样本,经 BP 算法得到对应权值的输出;
- 4) 设定网络目标函数,将其转换为适应度,对每一个输出进行评估;
- 5) 选择适应度大的个体,直接遗传给下一代;
- 6) 利用选定的交叉、变异及相关算法、参数对当前群体进行进化,得到新一代种群;
- 7) 返回 3),直到性能满足要求,得到优化权值;
- 8) 取整个操作中的最优个体作为 BP 神经网络的初始权(阈)值,对网络进行训练,得到全局最优解。

误差函数和适应度函数分别为记为

$$E = \frac{1}{2} \sum_{k=1}^m (t_k - y_k)^2, \quad (1)$$

$$f = \frac{1}{E}. \quad (2)$$

其中, E 为输出误差, $t_k, y_k (k = 1, 2, 3, \dots, m)$ 分别为网络的期望输出和实际输出, f 为适应度函数。

1.4 ARIMA 模型

ARIMA 模型是著名的时间序列预测方法,其中 ARIMA(p, d, q) 为差分自回归移动平均模型, p 为自回归阶数, q 为移动平均阶数, d 为时间序列平滑次数。其通用表达式为

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + u_t - \varphi_1 u_{t-1} - \varphi_2 u_{t-2} - \dots - \varphi_q u_{t-q}, \quad (3)$$

其中, z_t 为样本值, $\phi_1, \phi_2, \dots, \phi_p$ 为自回归系数, $\varphi_1, \varphi_2, \dots, \varphi_q$ 为滑动平均系数, u_t 为服从独立正态分布 $N(0, \delta^2)$ 的白噪声序列。

通过对检验数据序列的平稳性检验,再通过自相关函数(ACF)和偏自相关函数(PACF)确定 p 和 q ,并进行误差检验,然后用通过检验的模型进行预测。

1.5 组合预测模型的建立

本文采用基于信息融合思想的非线性组合预测算法。首先选用改进的 BP 网络和 ARIMA 2 种方法分别对样本进行预测,再使用 BP 神经网络把单个模型的同一时刻预测值作为网络输入,把相应时刻的实测值作为期望输出来训练神经网络,将完成训练的组合模型用于最终预测。组合预测模型结构如图 1 所示。

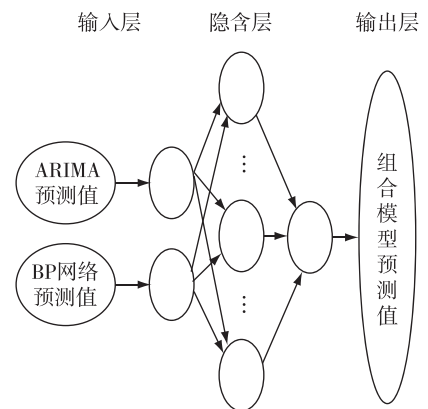


图 1 组合预测模型结构

Fig. 1 Structure of integrated forecasting model

2 模型应用

根据组合模型算法,采用南京市某路段检测器采集的 2011 年 5 月 16 日全天的交通流量数据,选取其中 7:40—16:00 时间段的数据,数据间隔为 5 min,一共 100 个数据。先用改进的 BP 神经网络和 ARIMA 模型算法分别对样本数据进行预测,用前 80 个数据作为建模数据,预测后 20 个数据,再用预测

所得数据的前 15 个数据作为组合模型的训练样本输入,对应时刻的实测值作为期望输出来训练建立组合预测模型.用剩余的 5 个预测数据作为训练后网络的输入,输出组合模型的预测值.

2.1 改进的 BP 神经网络预测

用 Matlab 7.1 中的神经网络工具箱函数和遗传算法工具箱函数建立由遗传算法优化 BP 神经网络权值的模型.把数据中 7:45—14:20 的 80 个数据作为训练样本,后 20 个数据作为测试样本.输入数据为 $t-20$ 、 $t-15$ 、 $t-10$ 、 $t-5$ 和 t 时刻的 5 个流量数据,输出为 $t+5$ 时刻的流量数据,测试样本选 14:00—15:55 的 25 个数据,划分方式同训练样本的输入数据,即连续的 5 个时刻为一个测试输入,一共 20 个.整个神经网络输入层节点数为 5,输出层节点数为 1,隐含层节点数为 9,目标精度为 0.001,学习率为 0.05,最大训练周期为 5 000.遗传算法初始种群大小为 50,最大进化代数为 100.图 2 和 3 分别为改进前后的训练效果.

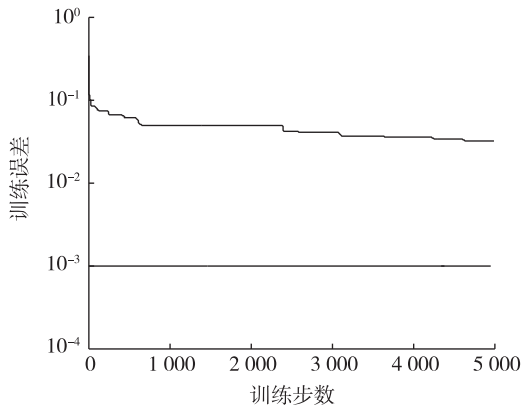


图 2 BP 神经网络训练效果

Fig. 2 Training result of BP neural network

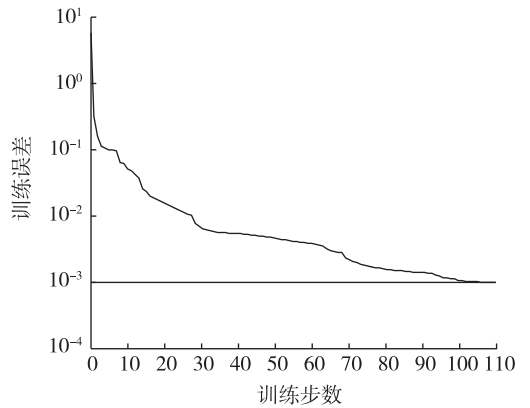


图 3 改进的 BP 网络训练效果

Fig. 3 Training result of improved BP neural network

由图 2、3 可知,优化前 BP 神经网络训练到 5 000 步未达到精度要求,经优化后的 BP 网络用了 109 步就完成了训练,可见改进的 BP 神经网络模型性能较好.用训练完成后的 BP 神经网络对后 20 个时间段进行预测,得到预测结果如表 1 所示.

经计算,用遗传算法优化后的 BP 神经网络模型预测出的 20 个值的平均误差率为 6.3%.精度不算太高的原因可能是因为训练样本数量小而造成的.

2.2 ARIMA 模型预测

根据 ARIMA 模型思想,分别对原始数据进行平稳化处理,再通过分析时间序列的自相关函数和偏自相关函数以及 BIC 准则,最终确定模型参数.利用统计学软件 SPSS 18.0 对交通流量进行预测.

采用 7:45—14:20 的 80 个实测数据来预测 14:25 的交通流量,再用 7:50—14:25 的 80 个实测数据来预测 14:30 的交通流量,以此滚动预测,得出 14:25—16:00 的 20 个预测值.利用前 80 个数据经

表 1 改进的 BP 网络预测值与实测值比较

Table 1 Comparison between forecasting result of improved BP neural network and measured values

北京时间	实测值/辆	预测值/辆	误差率	北京时间	实测值/辆	预测值/辆	误差率
14:25	109	101	0.032	15:15	94	99	0.053
14:30	100	105	0.047	15:20	101	104	0.030
14:35	95	92	0.032	15:25	95	103	0.084
14:40	107	102	0.047	15:30	106	101	0.047
14:45	96	103	0.073	15:35	110	103	0.064
14:50	109	96	0.119	15:40	112	99	0.116
14:55	112	102	0.089	15:45	101	94	0.069
15:00	99	97	0.020	15:50	108	102	0.056
15:05	101	98	0.030	15:55	99	95	0.040
15:10	97	103	0.062	16:00	107	96	0.103

过一次差分处理和一次平滑处理之后的时间序列模型得到了拟合度较好的模型,记为 $ARIMA(0,1,1)$. 图 4 和 5 分别为原时间序列和经过一次差分后的时间序列.

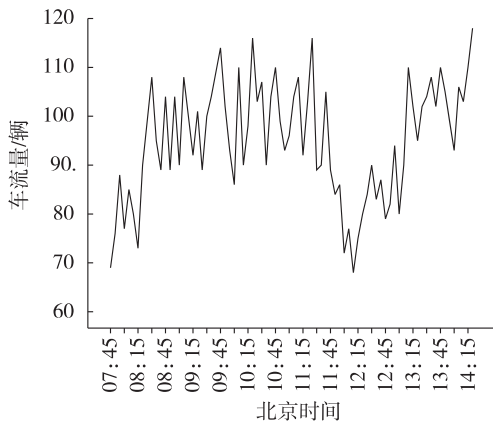


图 4 原时间序列

Fig. 4 Original time sequence

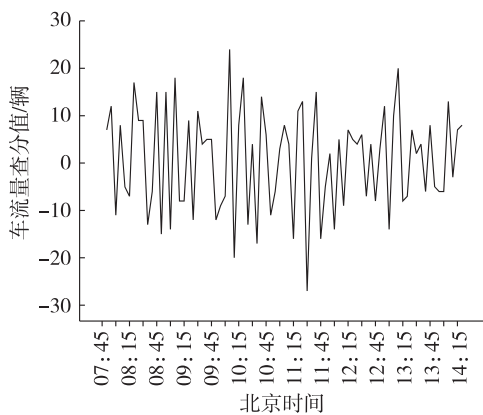


图 5 差分后的时间序列

Fig. 5 Differential time sequence

由图 4 可见该时间序列为非平稳序列,而图 5 中经过一次差分后的时间序列始终在零点附近波动,没有明显趋势或周期,基本可以视为平稳序列.

图 6 和 7 分别为车流量差分序列的自相关和偏自相关. 可见图 6 和 7 都为拖尾序列,可看做 $ARMA(p,q)$ 模型. 根据其拖尾步长,初步定 $p=2,3, q=1, 2$, 由此构成 $ARIMA(2,1,1), ARIMA(2,1,2), ARIMA(3,1,1), ARIMA(3,1,2)$ 4 个模型,再用平稳的 R^2 (模型解释的总变异所占比例的估计值), 正态化的贝叶斯信息准则 (BIC) 值和 MAPE (平均百分比绝对误差) 来估计模型的精度指标. 表 2 为 4 个模型的参数估计结果.

从表 2 可以看出, $ARIMA(2,1,1)$ 的 BIC 值最小,但是由于其调整后样本的决定系数平稳的 R^2 较

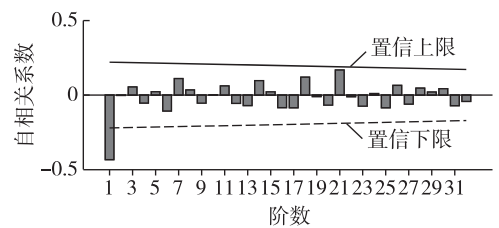


图 6 原序列一阶差分的自相关

Fig. 6 Auto correlation of first difference for original serial

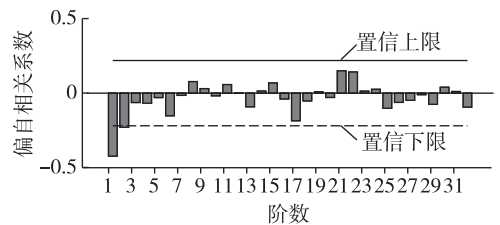


图 7 原序列一阶差分的偏自相关

Fig. 7 Partial auto correlation of first difference for original serial

表 2 4 个模型的估计参数结果对比

Table 2 Comparison of estimated parameters results of four models

模型	平稳的 R^2	正态化的 BIC	$E_{MAP}/\%$
$ARIMA(2,1,1)$	0.253	4.836	8.866
$ARIMA(2,1,2)$	0.249	4.910	8.769
$ARIMA(3,1,1)$	0.264	4.890	8.739
$ARIMA(3,1,2)$	0.265	4.958	8.679

低,且 MAPE 值较大,故精度不高. 由于 $ARIMA(3,1,1)$ 模型的 BIC 值相对较低,且平稳的 R^2 和 MAPE 值在各模型中处于较好的水平,故选取 $ARIMA(3,1,1)$ 模型作为车流量的预测模型.

图 8 和 9 分别为用 $ARIMA(3,1,1)$ 模型的残差序列的自相关和偏相关.

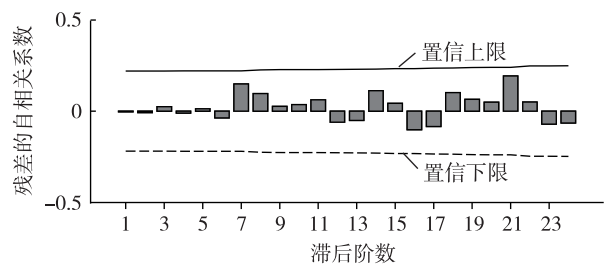


图 8 $ARIMA(3,1,1)$ 模型残差的自相关函数

Fig. 8 Auto correlation function of residuals of $ARIMA(3,1,1)$

由图 8、9 可见, $ARIMA(3,1,1)$ 模型的车流量残差序列的自相关系数和偏自相关系数都落在置信区

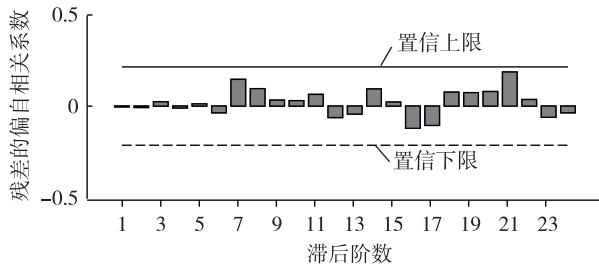


图9 ARIMA(3,1,1)模型残差的偏自相关函数
Fig.9 Partial auto correlation function of residuals of ARIMA(3,1,1)

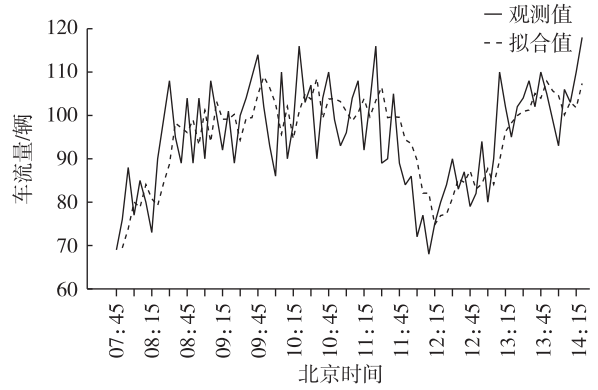


图10 ARIMA(3,1,1)模型对车流量的拟合
Fig.10 Traffic volume fitting result of ARIMA(3,1,1) model

间内,且没有明显的趋势,基本上可视残差序列为纯随机序列,模型检验通过。

将通过检验的模型用作预测,预测对象为14:25—16:00的20个车流量数据.由ARIMA(3,1,1)模型预测的14:25的预测值为103.47,实际车流量取整数103.

图10为模型对车流量拟合曲线,可见模型的拟合曲线与实测值有较高的近似度,可以推断该模型是合理的.然后利用滚动预测,得到14:25—16:00的20个预测结果,如表3所示.

经计算,ARIMA(3,1,1)模型预测出20个值的平均误差率为5.9%,基本符合预测精度.

2.3 组合模型预测

将2种单独模型预测出来的20个预测值中的前15个数据作为输入样本,对应时刻的实测值作为期望输出,训练建立组合预测模型,模型输入层为2个神经元,经多次试验,得出隐含层神经元数为5,输出层为1个神经元.图11为组合模型的训练效果.

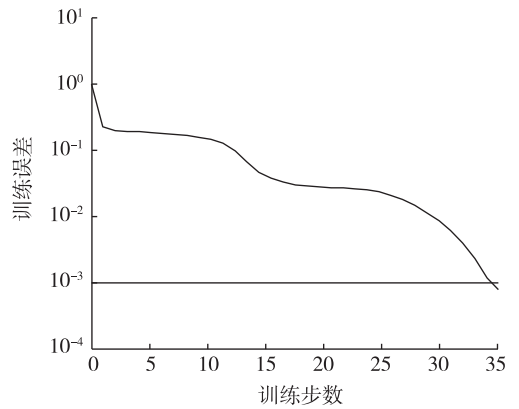


图11 组合模型训练
Fig.11 Training diagram of integrated model

用训练后的组合模型进行预测,把单独模型预测出的后5个数据作为模型的输入样本,得到对应时刻的组合预测结果,如表4所示.该过程通过Matlab 7.1实现.

表3 ARIMA(3,1,1)模型的预测值与实测值的比较

Table 3 Comparison between forecasting data of ARIMA(3,1,1) model and measured data

北京时间	实测值/辆	预测值/辆	误差率	北京时间	实测值/辆	预测值/辆	误差率
14:25	109	103	0.055	15:15	94	100	0.064
14:30	100	105	0.050	15:20	101	98	0.030
14:35	95	103	0.084	15:25	95	99	0.042
14:40	107	100	0.065	15:30	106	99	0.066
14:45	96	102	0.063	15:35	110	102	0.073
14:50	109	101	0.073	15:40	112	106	0.054
14:55	112	104	0.071	15:45	101	109	0.079
15:00	99	107	0.081	15:50	108	106	0.019
15:05	101	105	0.040	15:55	99	106	0.071
15:10	97	103	0.062	16:00	107	103	0.037

表 4 3 种模型预测值与实测值对比

Table 4 Comparison between forecasting results of three models and measured values 辆

北京时间	实测值	预测值		
		ARIMA 模型	改进后的 BP 模型	组合模型
15:40	112	106	99	108
15:45	101	109	94	105
15:50	108	106	102	106
15:55	99	106	95	98
16:00	107	103	96	104

2.4 结果分析

用 APE(百分比绝对误差)和 MAPE(平均百分比绝对误差)指标来评估 3 种算法的预测性能:

$$E_{AP} = \frac{|y_i - \hat{y}_i|}{y_i} \times 100\%, \quad (4)$$

$$E_{MAP} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100\%, \quad (5)$$

其中, y_i 是交通流量实测值, \hat{y}_i 是交通流量预测值, n 为预测数据的数量.

图 12 为 3 种预测方法的百分比绝对误差的对比,表 5 列出了 3 种预测方法的平均百分比绝对误差对比结果.

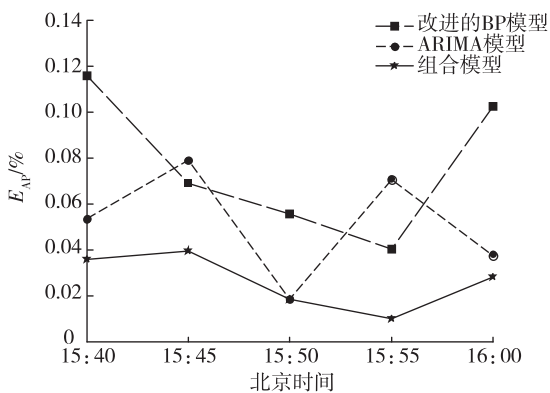


图 12 3 种预测方法的百分比绝对误差对比

Fig. 12 APE comparison of three forecasting methods

表 5 3 种预测方法的 MPAE 对比

Table 5 MAPE comparison of three forecasting methods

预测方法	$E_{MAP}/\%$
改进的 BP 模型	7.683
ARIMA 模型	5.188
组合模型	2.640

由图 12 和表 5 可知,组合模型的预测误差比单个模型预测误差小,且 MAPE 值小于 5%,达到了较

高的预测精度. BP 神经网络的预测误差相对较大,可能是由于 BP 神经网络的训练样本数量较少,影响了模型的训练效果,导致预测模型精度相对较低,而 ARIMA 模型的预测精度趋于平稳状态,波动幅度较小,预测带有一定的滞后性. 组合模型的总体预测效果都优于单个模型,较小的平均百分比绝对误差说明该组合模型对于车流量预测有较好的实用性.

3 结语

车流量的准确预测是非正常拥堵检测的关键因素,选择合理的预测方法又是提高车流量预测准确率的前提. 本文采用改进的 BP 神经网络和 ARIMA 模型的组合预测的方法建立了车流量的短时预测模型,符合车流量不确定性和非线性特性,并利用遗传算法优化 BP 神经网络,解决了网络训练速度慢、易陷入局部最小的缺点. 对实测数据的实验结果表明,组合模型预测效果优于单个模型,该组合预测模型对交通流预测的实际应用有较好的参考价值.

参考文献

References

- [1] Stephanedes V J, Michalopoulos P G, Plum R A. Improved estimation of traffic flow for real-time control[J]. Transportation Research Record, 1981, 79(5): 28-39
- [2] Smith B L, Demetsky M J. Traffic flow forecasting: Comparison of modeling approaches[J]. Journal of Transportation Engineering, 1997, 123(4): 261-265
- [3] Okutani I, Stephanedes Y J. Dynamic prediction of traffic volume through Kalman filtering theory [J]. Transportation Research. Part B: Methodological, 1984, 18(1): 1-11
- [4] Hobeika A G, Chang Kyun Kim. Traffic-flow-prediction systems based on upstream traffic [C]//Proceedings of Vehicle Navigation and Information System Conference, 1994: 345-350
- [5] 曹征. 基于小波变换的交通流短时预测模型研究 [D]. 北京: 北京交通大学机械与电子工程学院, 2010
CAO Zheng. Study of the forecasting model of short-term traffic flow based on wavelet transform [D]. Beijing: School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University, 2010
- [6] 郇洪江, 宫宇生, 胡斌. 改进的 BP 神经网络在交通流量预测中应用[J]. 微电子学与计算机, 2010, 27(1): 106-108
HUAN Hongjiang, GONG Ningsheng, HU Bin. Application of modified BP neural network in traffic flow forecasts [J]. Microelectronics & Computer, 2010, 27(1): 106-108
- [7] 苑士义, 撒力. 基本遗传算法设计及改进[J]. 微计算机信息, 2011, 27(12): 133-134
YUAN Shiyi, SA Li. The basic design and improvement of genetic algorithm [J]. Microcomputer Information, 2011,

27(1):133-134

- [8] 裴武,陈凤,程立勤. 交通量时间序列 ARIMA 预测技术研究[J]. 山西科技,2009(1):75-76

PEI Wu, CHEN Feng, CHENG Liqin. Research of ARIMA forecasting technology for traffic volume time series [J]. Shanxi Science and Technology, 2009(1):75-76

Integrated forecasting model for short-term traffic flow

ZHANG Kai¹ LU Zouying¹

1 School of Information and Control, Nanjing University of Information Science & Technology, Nanjing 210044

Abstract With the continuous increase of road vehicles, occasional congestion caused by traffic accidents seriously affect the commuting efficiency of traveler and the overall operation level of road network. Real-time and exact forecasting of short-term traffic flow volume is the key point to intelligent traffic system and precondition to solve the congestion situation by route guidance and clearing. According to the uncertain and non-linear features of traffic volume, a model integrated of the improved BP neural network and autoregressive integrated moving average (ARIMA) model is established to forecast the short-term traffic flow. The case application result shows that the combined model has an advantage over the single models in forecasting performance and forecasting accuracy.

Key words traffic flow forecasting; BP neural network; ARIMA model; integrated forecasting