

基于CNN-SVR混合深度学习模型的 短时交通流预测

罗文慧,董宝田*,王泽胜

(北京交通大学 交通运输学院, 北京 100044)

摘要: 精准且快速的短时交通流预测是智能交通发展的重要组成部分.本文针对当前交通流预测模型不能充分提取交通流数据的时空特征、预测性能容易受到外界干扰因素影响的问题,提出一种基于深度学习的短时交通流预测模型.该模型结合卷积神经网络(Convolutional Neural Network, CNN)与支持向量回归分类器(Support Vector Regression, SVR)的特点:在网络底层应用CNN进行交通流特征提取,并将提取结果输入到SVR回归模型中进行流量预测.为验证模型的有效性,取G103国道的实际交通流量数据进行试验.结果表明,提出的预测模型与传统的预测模型相比具有更高的预测精度,预测性能提高了11%,是一种有效的交通流预测模型.

关键词: 智能交通;交通流预测;卷积神经网络;交通流;支持向量回归;深度学习

Short-term Traffic Flow Prediction Based on CNN-SVR Hybrid Deep Learning Model

LUO Wen-hui, DONG Bao-tian, WANG Ze-sheng

(School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China)

Abstract: It is very important for intelligent transportation development to realize accurate and fast traffic forecast. However, dominant models for short-term traffic flow forecasting can't extract spatial-temporal characteristics of traffic flow data amply. Moreover, these models are susceptible to outside factors. To resolve these problems, an innovative model based on deep learning is proposed in this paper. Convolutional Neural Network(CNN) and Support Vector Regression(SVR) classifier are combined in this model: feature learning of traffic flow is carried out by using CNN in underlying network, then the extracted results are transmitted to SVR model as input to predict traffic flow. To verify the validity of the proposed model, experiments are conducted on actual traffic flow data of China national highway 103(G103). Experimental results show that the proposed model has higher prediction accuracy than the traditional prediction model, and the prediction performance is improved by 11%, which is an effective traffic flow forecasting model.

Keywords: intelligent transportation; traffic flow prediction; convolutional neural network; traffic flow; support vector regression; deep learning

0 引言

实现对未来交通状态快速和精准的判断是智

能交通控制与诱导、道路网络规划的基础.短时交通流预测作为国内外智能交通系统的研究热点,

收稿日期:2017-06-22

修回日期:2017-08-09

录用日期:2017-08-28

作者简介:罗文慧(1983-),女,新疆乌鲁木齐人,博士生.

*通信作者 btdong@bjtu.edu.cn

30多年来取得了丰富研究成果,主要包括基于时间序列的统计模型、非参数预测模型和混合模型等。随着大数据时代的到来,深度学习近几年来发展迅速,交通科学研究者也将深度学习理论应用于解决交通问题:Huang W.等^[1]提出由DBN和顶部多任务回归层的深度学习模型;Lv Y.等^[2]应用自动编码器深层架构模型预测交通流特征;Yang H. F.等^[3]提出新型Levenberg-Marquardt堆叠自编码器深度学习模型,应用Taguchi方法优化结构,逐层提取特征并无监督学习交通流特征;Wu Y.等^[4]用一维卷积网络捕获交通流空间特征,用两个长期短时记忆法挖掘交通流短时周期性,设计特征级融合的深层架构预测短时交通流;Jia Y.等^[5]建立DBN模型预测短时交通速度,采用贪婪无监督方法训练模型并通过标记数据进行微调;Gaetano Fusco等^[6]在对比了季节性自回归、贝叶斯模型和神经网络等预测模型在交通拥挤与正常状态下的性能表现的基础上,提出了一种基于交通状态的交通流速度预测混合模型;Wang J. Y.等^[7]提出单隐层卷积神经网络结合误差反馈的交通流速度预测模型。

以上模型将深度学习理论应用到了交通领域,并取得了很好的成效。但是,在有限的计算条件下,这些模型不能充分提取交通流的时空特征,这对预测精度有一定的影响。另一方面,交通流因易受外界因素的影响,包括天气情况、交通事故、交通管制等,这使得交通流经常出现异常状态。已有的方法在交通流异常状态下,预测精度往往有明显的下降。本文利用卷积神经网络模型权值共享的特点,在降低模型复杂度、减轻过拟合、降低计算量的同时,有效地提取交通流数据的时空特征,降低外界干扰因素的影响程度,抽象出交通流的本质特性,最后以此作为输入向量,输入到顶层的支持向量回归模型进行交通流的预测,保证了模型的预测精度。若没有特殊的说明,下文以CNNs为模型的简写形式。

1 基于CNNs的短时交通流预测模型网络的构造

1.1 模型网络结构

考虑到梯度扩散与计算量的问题,本文设计

的卷积神经网络预测模型采用单隐含层结构。隐含层包含一个卷积层(conv layer)和一个子采样层(mean pooling),卷积层和子采样层对应相连。在子采样层之后,将子采样层的所有节点展开成一条特征向量,并与输出节点间进行全连接(full connection),输入到SVR中进行交通流量的最后预测。具体的网络模型结构如图1所示,图中 w 表示卷积核,pool表示子采样因子。

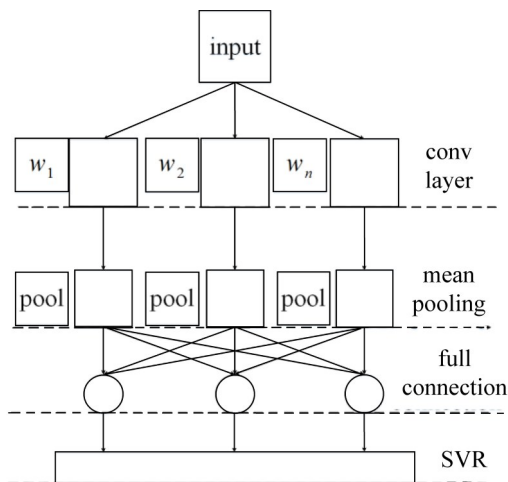


图1 CNNs 模型结构图

Fig. 1 Structure diagram of CNNs model

1.2 基于时空特征的输入矩阵

对于某个监测点某一时刻的交通流量可以从时间和空间两个方面解释:在时间上,交通流遵循一定的变化规律,下一时段交通流量可看作上一时段交通流量的延续;在空间上,交通流受上下游交通状态的影响而呈现出一定的相关性,下游路段的交通流量可由上游路段交通流量估计得到^[8]。为了综合考虑时间与空间维度上的各种因素对交通流的影响,本文将原始交通流量数据处理成带有时空信息的交通流矩阵,处理过程如下:

若以 $x_{p,t}$ 表示数据采集点 p 在 t 时刻的交通流量,则由所有 n 个数据采集点组成的 t 时刻的数据流量 A 为

$$A = \begin{pmatrix} x_{1,t} \\ \vdots \\ x_{p,t} \end{pmatrix} \quad (1)$$

式中: $p=1,2,\dots,n$ 。

为了使模型更加精确,我们采用1 h预测模型,采样间隔为5 min,因此模型的输入矩阵 B 可以表示为

$$B = \begin{pmatrix} x_{1,t-12} & \cdots & x_{1,t} \\ \vdots & & \vdots \\ x_{p,t-12} & \cdots & x_{p,t} \end{pmatrix} \quad (2)$$

1.3 网络模型的构造

(1) 卷积层的构造.

卷积层是通过卷积来计算的,包含多个卷积面,用 $h_{1,a}$ 和 $w^{1,a}$ 分别表示此层第 a 个卷积面和卷积核,激励函数为ReLU,则有

$$h_{1,a} = \text{ReLU}(x * w^{1,a} + b^{1,a}) \quad (3)$$

假设 x 和 $w^{1,a}$ 的大小分别为 $M \times N$ 和 $m \times n$,且 $M \geq m, N \geq n$,则它们的卷积 $C = x * w$ 的所有元素为

$$C_{ij} = \sum_{s=1}^m \sum_{t=1}^n x_{i+m-s, j+n-t} \cdot w_{st}^{1,a}, \quad (4)$$

$$1 \leq i \leq M - m + 1; 1 \leq j \leq N - n + 1$$

卷积层 H_1 由所有的卷积面 $h_{1,a}$ 构成.

(2) 下采样层的构造.

如果对 $h_{1,a}$ 进行不重叠分块,每块的大小为 $\lambda \times \tau$,则其中的第 ij 个块可以表示为 $G_{\lambda\tau}^A(i,j) = (a_{st})_{\lambda \times \tau}$,其中 $(i-1) \cdot \lambda + 1 \leq s \leq i \cdot \lambda, (j-1) \cdot \tau + 1 \leq t \leq j \cdot \tau$ 对 $G_{\lambda\tau}^A(i,j)$ 的下采样为平均池化:

$$\text{down}(G_{\lambda\tau}^A(i,j)) = \frac{1}{\lambda \times \tau} \sum_{s=(i-1) \cdot \lambda + 1}^{i \cdot \lambda} \sum_{t=(j-1) \cdot \tau + 1}^{j \cdot \tau} a_{st} \quad (5)$$

此层是对卷基层的下采样计算,用 $h_{2,a}$ 表示此层的第 a 个采样面,则有

$$h_{2,a} = g(\beta_2 \text{down}_{\lambda_2 \tau_2}(h_{1,a}) + \gamma_2) \quad (6)$$

式中:权值 β_2 取值为1;偏置 γ_2 取值为0矩阵; $g(\cdot)$ 为恒等线性函数 $g(x) = x$.

下采样层 H_2 由所有的下采样面 $h_{2,a}$ 组成.

(3) 全连接层的构造.

此层为普通的前馈网络,激活函数为ReLU,则有

$$H_3 = \text{ReLU}(w^3 H_2 + b^3) \quad (7)$$

(4) 输出层的构造.

基于最小化间隔损失的SVR在某些情况下优于基于交叉熵损失的softmax^[9-10],所以本文利用线性 ε -SVR代替CNN常用的softmax作为模型的输出.

给定训练数据集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$,对于样本 (x_i, y_i) 根据模型的输出 $f(x_i)$ 与真实值 y_i 之间的差别来计算损失,当 $|f(x_i) - y_i| > \varepsilon$ 时,计算损失^[11-12]. ε -SVR的目标函数为

$$\begin{cases} \min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \\ \text{s.t.} \quad f(x_i) - y_i \leq \varepsilon + \xi_i, \\ y_i - f(x_i) \leq \varepsilon + \xi_i^*, \\ \xi_i \geq 0, \xi_i^* \geq 0, i = 1, 2, \dots, N \end{cases} \quad (8)$$

式中: C 为惩罚系数; ξ 和 ξ_i^* 为松弛变量.

引入拉格朗日乘子, $\mu_i \geq 0, \hat{\mu}_i \geq 0, \alpha_i \geq 0, \hat{\alpha}_i \geq 0$,定义拉格朗日函数为

$$L(w, b, \alpha, \hat{\alpha}, \mu, \hat{\mu}, \xi, \hat{\xi}) = \frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^N (\xi_i + \hat{\xi}_i) - \sum_{i=1}^N \mu_i \xi_i - \sum_{i=1}^N \hat{\mu}_i \hat{\xi}_i + \sum_{i=1}^N \alpha_i (f(x_i) - y_i - \varepsilon - \xi_i) + \sum_{i=1}^N \hat{\alpha}_i (y_i - f(x_i) - \varepsilon - \hat{\xi}_i) \quad (9)$$

根据拉格朗日对偶性可得

$$w = \sum_{i=1}^N (\hat{\alpha}_i - \alpha_i) x_i \quad (10)$$

假设最终解为 $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)$,在 $\hat{\alpha}^* = (\hat{\alpha}_1^*, \hat{\alpha}_2^*, \dots, \hat{\alpha}_N^*)^T$ 中找出 α^* 的某个分量 $C > \alpha_j^* > 0$,则有

$$b^* = y_i + \varepsilon - \sum_{i=1}^N (\hat{\alpha}_i^* - \alpha_i^*) x_i^T x_j \quad (11)$$

则模型输出的预测值为

$$f(x) = \sum_{i=1}^N (\hat{\alpha}_i^* - \alpha_i^*) x_i^T x + b^* \quad (12)$$

2 网络训练

本文的网络训练方式可以分为信号正向传播和误差反向传播两个阶段.误差反向传播时从顶层的SVR开始,目标函数 L 为式(8)^[9-10],SVR的输入为 H_3 .由于易受引言中提到的外界因素的影响,交通流会体现出不同的状态,因此在特征提取时,需要考虑常态与异常两种状态,然而SGD(Stochastic Gradient Descent)对所有的参数更新使用同样的学习率,不能满足要求.所以本文使用优化后的Adam(Adaptive Moment Estimation)梯度下降算法,Adam利用梯度的一阶矩估计和二阶矩估计动态调整每个参数的学习率,经过偏置校正后,每一次迭代学习率都有一个确定范围,使得参数比较平稳,这有利于充分提取交通流特征^[13].

3 实验

3.1 数据描述

本文实验数据来自北京市交通管理部门,路

段为国道G103京塘线,路线长度为33.39 km,车道数为6车道,共设置13个观测站如图2所示。

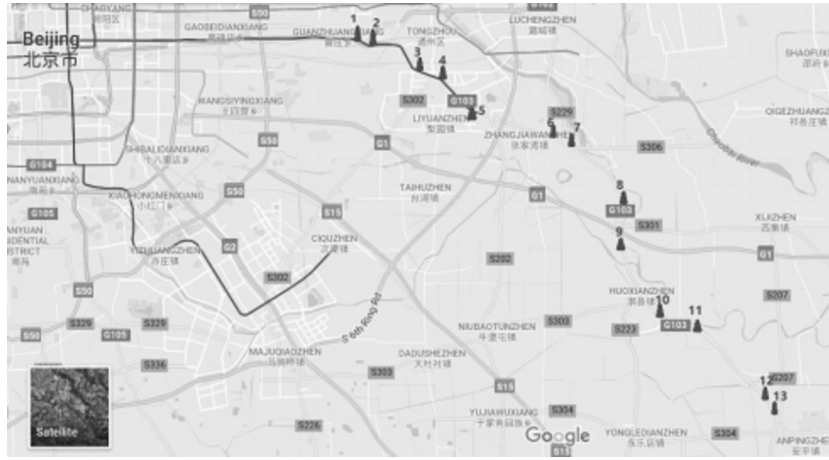


图2 数据采集设备位置图

Fig. 2 Location map of data collection devices

本文采集的数据为前12个观测站的2016年4~5月两个月的车流数据,数据采集间隔为5 min,以4月1~30日、5月6~16日数据作为训练数据,5月17~26日的的数据作为测试数据集,共有12 144个训练样本和2 760个测试样本,以此作为模型的输入。

3.2 数据的预处理

样本数量级的差异会导致量级较大的属性占据主导地位,从而使得目标函数值很大程度上依赖该属性,另一方面,数量级的差异会导致迭代收敛的速度减慢。然而,所有依赖于样本距离的算法对于数据的数量级都非常敏感^[14],所以本文在验证模型性能之前对样本数据进行数据的标准化处理,本文选取经典的Z-score标准化方法:

给定数据集 $D=\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$,其中 $x_i=(x_{i,1}, x_{i,2}, \dots, x_{i,d})$, $i=1, 2, \dots, N$, d 为 x_i 对应属性的个数,对于每个属性 $x_{i,j}$, $j=1, 2, \dots, d$,先计算该属性的标准值 u_j 和标准差 σ_j :

$$u_j = \frac{1}{N} \sum_{i=1}^N x_{i,j} \quad (13)$$

$$\sigma_j = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_{i,j} - u_j)^2} \quad (14)$$

然后计算标准化后的属性值:

$$\hat{x}_{i,j} = \frac{x_{i,j} - u_j}{\sigma_j} \quad (15)$$

标准化后的所有属性的均值为0,标准差为1。

3.3 模型参数的确定和对比方案

本文中CNNs模型的参数为经验数据,卷积层

数为1,下采样层数为1、卷积核的大小为 2×12 、卷积核的个数为16个,CNNs模型训练的迭代次数为10 000,学习率为 $1e-1$,batch-size为100,线性SVR惩罚因子 $C=0.1$, $\epsilon=0.1$ 。

为了证实模型的有效性,本文选取了单个的SVR和CNN深度学习模型与CNNs做性能比较,其中SVR、CNN参数值与CNNs中的一致。实验取值为50次计算结果的平均值。本文的实验平台为2.5GHz,4核,Inter(R)Core(TM)i7-6500U,8G内存,项目在python的IDE PyCharm中实施。

3.4 结果与分析

(1) 直观效果对比与分析。

本文对编号为G103L226110112的观测站5月17日、5月1日、5月23日的交通流实际测量值和所选模型预测值进行对比。5月1日代表节假日,5月17日天气微风、晴代表正常状态,5月23日为雷阵雨天气代表异常天气。图3分别为5月17日、5月1日、5月23日实际测量值(黑色实线)和模型预测值(浅黑色虚线)的对比情况,横轴标记时间,纵轴标记每5 min的车流量。如图3所示,5月1日交通流量相比正常状态较大,但波动较小,异常值较少,5月23日相对正常状态流量相当,但波动较大,异常值较多。从各个模型的预测情况来看,SVR预测值较为平稳,在正常状态下预测较为准确,但在高峰时段和异常状态下表现较差;CNN在高峰时段和异常状态下拟合较好但在正常状态时,表现不太稳定,时高时低;CNNs综合了两个模型的优点,不仅在正常状

态下保持较高的准确度,而且在高峰时段和异常状态
态下仍能够抓取异常峰值的特征,保持较高的预测
精度度,较为真实地反映交通流地变化规律.

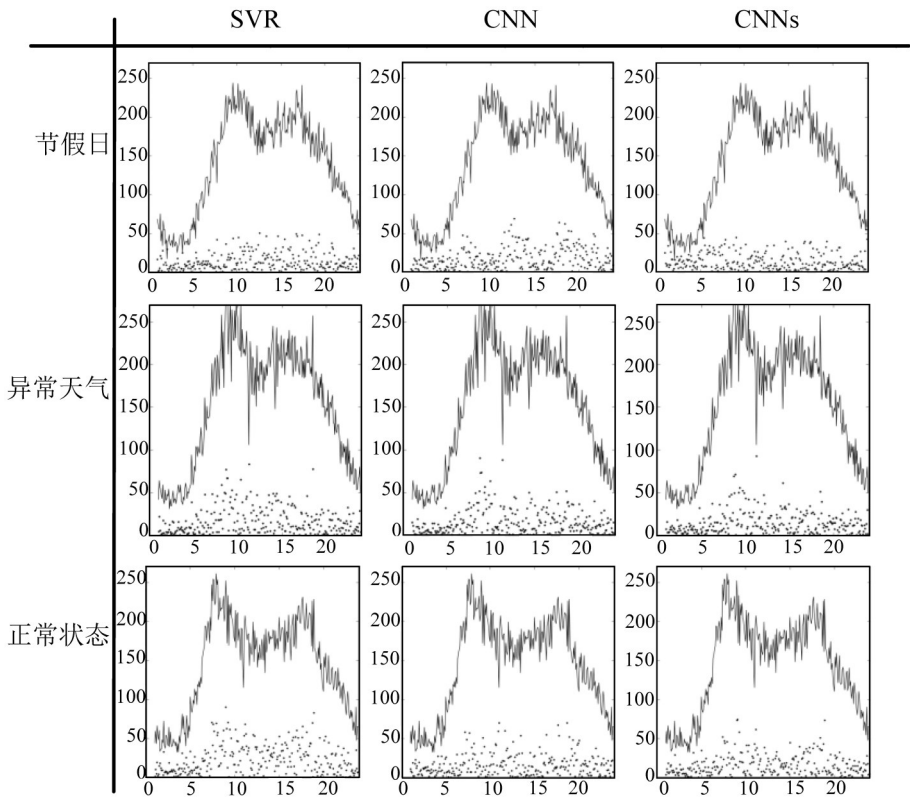


图3 CNNs与其他2种模型在不同交通状态下的预测结果对比

Fig. 3 Comparison of result between CNNs and the two other models under multi-condition

为了验证CNNs的适用性,本文另提取5月25
日交通流数据,并对全路段12个采集点的交通流
量进行了预测.图4中,(a)为12个监测点交通流量
实际测量图,(b)为CNNs模型预测图,x轴为时间,y
轴标记12个监测点,z轴为交通流量.如图4所示,

CNNs较为贴切地反映了1天288个时间点上整段
路线的交通流变化情况,同时也显示出了交通流
随着时间的变化在各个观测点的传递情况,从预
测结果来看,本文提出的CNNs适用于整条路段.

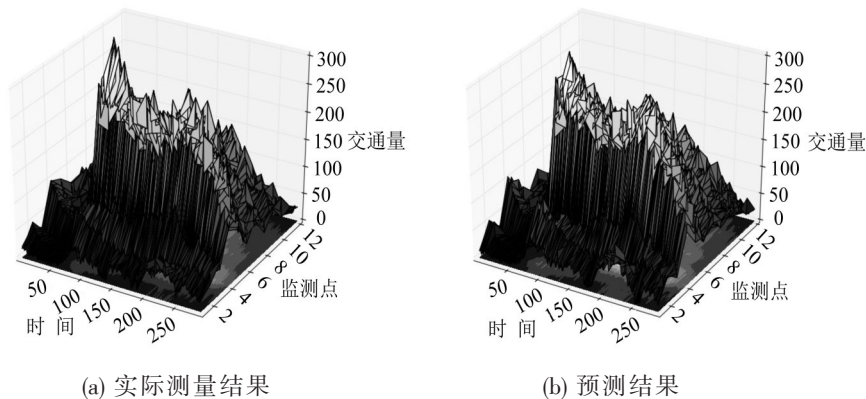


图4 CNNs在多个样本上的预测结果

Fig. 4 The predicted results of CNNs on multi-samples

(2) 客观参数对比与分析.

Mean Absolute Error(MAE)能很好地反映预测

值误差的实际情况,Root Mean Squared Error
(RMSE)可以评价数据与预测模型适应性.所以,本

文根据提出模型的监督机制和数据集连续的特性,采用平均绝对误差 e_{MAE} 、均方根误差 e_{RMSE} 两个误差评价指标对模型性能进行分析,公式如下:

$$e_{MAE} = \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{n} \tag{16}$$

$$e_{RMSE} = \sqrt{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n}} \tag{17}$$

式中: y_i 为实际观测值; \hat{y}_i 为预测值; n 为样本数量.

为了定量验证模型的适用性,本文利用3种模型分别计算了2016年4月、5月两月中所有特殊天气、节假日和正常天气的 e_{MAE} 和 e_{RMSE} 值如表1所示,从1表中可以定量的看出,CNNs在不同状态下,以及综合测试集上保持了较低的值,在测试集上,相比SVR模型MAE减小了19%,相比CNN模型MAE减小了28%.

表1 CNNs和其他两种模型的MAE、RMSE比较

Table 1 The comparison of MAE and RMSE between CNNs and the two other models

	MAE			RMSE		
	SVR	CNN	CNNs	SVR	CNN	CNNs
正常状态	9.3	12.1	8.4	12.6	13.3	10.2
节假日	10.8	13.8	8.7	13.0	14.6	11.1
异常天气	12.3	11.3	8.1	14.2	12.1	9.8
测试集	10.2	11.5	8.2	13.5	12.9	10.0

(3) 残差分析.

为了验证模型鲁棒性,本文分别对3种模型在节假日、特殊天气、正常情况的残差进行了对比,结果如图5所示,其中,横轴标记时间,纵轴标记交通流量,曲线为当天的实际流量值,散点为对应时刻的残差值.从图5中我们可以看出,SVR的残差

在交通流波动较小的情况下较小,而在交通流在短时间内变大或变小的情况下增大;CNN的残差在特殊情况和正常状态下无明显的变化,但整体上残差的值较大;相比而言,CNNs的残差在正常状态、高峰时段和交通流波动较大的情况下都能保持较小的值.

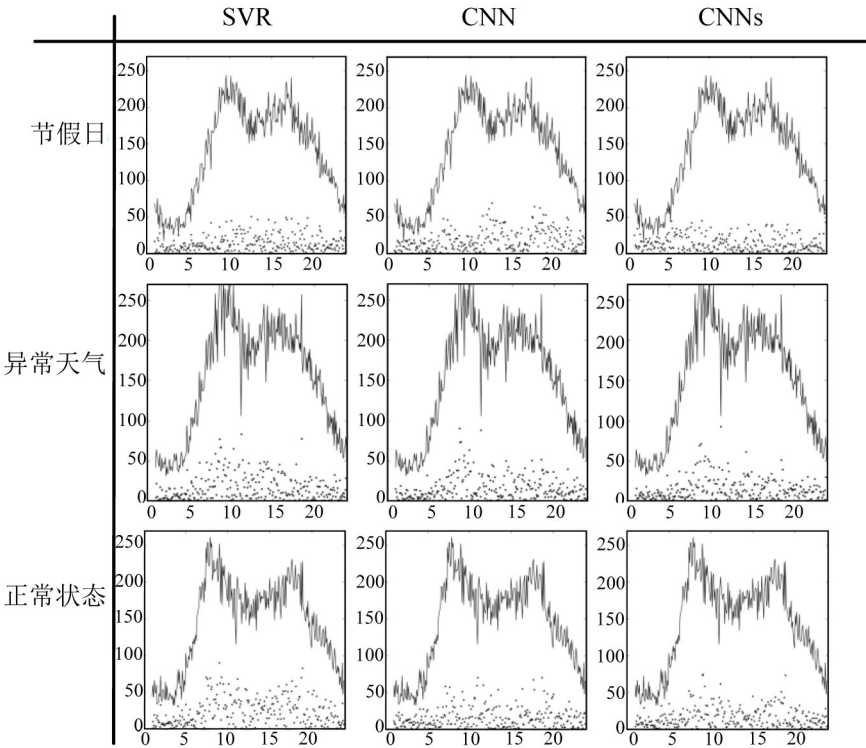


图5 CNNs与其他两种模型在不同交通状态下的残差对比

Fig. 5 Comparison of the residuals between CNNs and the two other models under multi-condition

4 结 论

本文提出了一种应用卷积神经网络模型结合支持向量回归的短时交通流预测方法.设计了单隐层CNN模型加SVR的混合模型,将带有时空信息的交通流数据处理成适应于卷积神经网络输入的形式;应用SVR的目标函数来训练整个网络;采用了优化后的Adam算法代替SGD随机梯度下降算法保证交通流时空特征的充分提取;最后利用G103国道交通流数据进行模型性能分析.实验结果表明,本文提出的预测模型有较高的预测精度,预测值能够与真实值保持高度的一致,而且受外界干扰因素影响较小,是一种有效的交通流量预测方法.

本文此次试验路段为高速路段,未考虑交通网络的复杂性,下一步的研究将结合复杂交通网络相关知识,对城市路段进行研究,从而增强预测模型的泛化能力,进一步提高算法的适用性.

参考文献:

- [1] HUANG W, SONG G, HONG H, et al. Deep architecture for traffic flow prediction: Deep belief networks with multitask learning[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(5): 2191-2201.
- [2] LV Y, DUAN Y, KANG W, et al. Traffic flow prediction with big data: A deep learning approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(2): 865-873.
- [3] YANG H F, DILLON T S, CHEN Y P. Optimized structure of the traffic flow forecasting model with a deep learning approach[J]. IEEE Transactions on Neural Networks & Learning Systems, in press, doi:10.1109/TNNLS.2016.2574840.
- [4] WU Y, TAN H. Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework[J/OL]. (2016-12-3) [2017-3-10]. <https://arxiv.org/abs/1612.01022>.
- [5] JIA Y, WU J, DU Y. Traffic speed prediction using deep learning method[C]//2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). Piscataway, NJ, USA: IEEE, 2016: 1217-1222.
- [6] FUSCO G, COLOMBARONI C, ISAENKO C. Short-term speed predictions exploiting big data on large urban road networks[J]. Transportation Research, Part C: Emerging Technologies, 2016(73): 183-201.
- [7] WANG J Y, GU Q, WU J J, et al. Traffic speed prediction and congestion source exploration: A Deep Learning Method[C]// Proceedings- IEEE International Conference on Data Mining, ICDM. Barcelona, Spain: Institute of Electrical and Electronics Engineers Inc., 2017: 499-508.
- [8] LOLLI F, GUMBERINI R, REGATTIERI A, et al. Single-hidden layer neural networks for forecasting intermittent demand[J]. Production Economics, 2017 (183): 116-128.
- [9] TANG Y. Deep learning using linear support vector machines[J/OL]. (2015-2-21) [2017-3-10]. <https://arxiv.org/abs/1612.01022>.
- [10] LI W, HU J, CHEN B H. A deep quasi-linear kernel composition method for support vector machines[C]// Proceedings of the International Joint Conference on Neural Networks. Vancouver, Canada: Institute of Electrical and Electronics Engineers Inc., 2016: 1639-1645.
- [11] 傅贵, 韩国强, 逯峰, 等. 基于支持向量机回归的短时交通流预测模型[J]. 华南理工大学学报(自然科学版). 2013, 41(9): 71-76. [FU G, HAN G Q, LU F, et al. Short-term traffic flow prediction based on support vector machine[J]. Journal of South China University of Technology (Natural Science Edition), 2013, 41(9): 71-76.]
- [12] 曾绍华. 支持向量回归机算法理论研究与应用[D]. 重庆: 重庆大学, 2006. [ZENG S H. The theory research of algorithm on support vector regression and application[D]. Chongqing: Chongqing University, 2006.]
- [13] HUANG H, HUANG X, LI R, et al. Sound quality prediction of vehicle noise using deep belief networks[J]. Applied Acoustics, 2016(113): 149-161.
- [14] YU D, DENG L. Deep learning and its applications to signal and information processing[exploratory DSP][J]. IEEE Signal Processing Magazine. 2011, 28(1): 145-154.