

基于维度加权的残差 LSTM 短期交通流量预测

李月龙^{1a,3}, 唐德华^{1a,2,3}, 姜桂圆², 肖志涛^{1b,3}, 耿磊^{1b,3}, 张芳^{1b,3}, 吴骏^{1b,3}

(1. 天津工业大学 a. 计算机科学与技术学院; b. 电子与信息工程学院, 天津 300387;

2. 南洋理工大学 计算机科学与工程学院, 新加坡 639798;

3. 天津市光电检测技术与系统重点实验室, 天津 300387)

摘 要: 基于神经网络的交通流量预测由于嵌入了部分手工设计的特征, 使得提取的网络特征功能单一, 存在适应性及鲁棒性差、数据局部特征刻画不准确等问题。为此, 提出基于残差长短期记忆网络(LSTM)的交通流量预测方法, 利用集成学习思想将空间分布的数据端到端训练到残差 LSTM 网络中, 同时在每个 LSTM 单元后引入维度加权单元, 显式建模特征维度之间的相互依赖关系。实验结果表明, 该方法能实现短期交通流量数据的自适应建模分析。

关键词: 智能交通; 短期交通流量预测; 残差连接; 长短期记忆网络; 维度加权

中文引用格式: 李月龙, 唐德华, 姜桂圆, 等. 基于维度加权的残差 LSTM 短期交通流量预测[J]. 计算机工程, 2019, 45(6): 1-5.

英文引用格式: LI Yuelong, TANG Dehua, JIANG Guiyuan, et al. Short term traffic flow forecasting based on dimension weighted residual LSTM[J]. Computer Engineering, 2019, 45(6): 1-5.

Short Term Traffic Flow Forecasting Based on Dimension Weighted Residual LSTM

LI Yuelong^{1a,3}, TANG Dehua^{1a,2,3}, JIANG Guiyuan², XIAO Zhitao^{1b,3},
GENG Lei^{1b,3}, ZHANG Fang^{1b,3}, WU Jun^{1b,3}

(1a. School of Computer Science and Technology; 1b. School of Electronics and Information Engineering,
Tianjin Polytechnic University, Tianjin 300387, China; 2. School of Computer Science and Engineering,

Nanyang Technological University, Singapore 639798, Singapore;

3. Tianjin Key Laboratory of Optoelectronic Detection Technology and Systems, Tianjin 300387, China)

【Abstract】 Due to the fact that the current neural network-based traffic flow forecasting method embeds part of the manually designed features, the feature extracted by the network has single function, bad adaptability, poor robustness and inaccurate characterization of the local features. Therefore, this paper proposes a traffic flow forecasting method based on residual Long Short Term Memory (LSTM). The model uses the idea of ensemble learning to train the spatially distributed data into a residual LSTM network. In addition, dimension weighted units are introduced after each LSTM unit to present the interdependencies between different dimensions of modeling features. Experimental results show that the method can realize adaptive modeling analysis of short term traffic flow data.

【Key words】 intelligent transportation; short term traffic flow forecasting; residual connection; Long Short Term Memory (LSTM) network; dimension weighted

DOI: 10.19678/j.issn.1000-3428.0052120

0 概述

随着社会经济的发展, 城市机动车数量急剧增加, 引入智能技术进行交通优化管理, 能够在较低成本投入下有效缓解交通拥堵。短期交通流量预测技

术是交通智能决策整体体系中的核心环节, 能实时对潜在拥挤状况进行预测, 以引导人们选择更合适的出行路线, 缓解交通压力^[1-3]。由于短期交通流量预测问题受天气变化、交通事故等外部因素的影响, 同时受限于错综复杂的实际道路交通网络, 因此进

基金项目: 国家自然科学基金(61771340, 61302127, 61601325); 天津市自然科学基金(18JCYBJC15300)。

作者简介: 李月龙(1982—), 男, 副教授、博士, 主研方向为智能交通、机器学习、模式识别; 唐德华, 硕士研究生; 姜桂圆, 研究员、博士; 肖志涛, 教授、博士; 耿磊、张芳、吴骏, 副教授、博士。

收稿日期: 2018-07-16 **修回日期:** 2018-08-31 **E-mail:** yuelong_li@126.com

行准确预测的难度较高^[4-6]。

在交通流量预测过程中,时空相关性是必须考虑的因素。时间相关性是指在相同的时间间隔下,当前的交通流量和过去的交通流量与时间跨度(即时间域)的相关性,而空间相关性指的是目标路段与其他路段的交通流量之间(即空间域)的相关性^[7]。对于强调瞬时、高精度的局部短期交通流量预测问题,如何有效建模时空相关性,对最终结果具有决定性影响。

当前主流预测方法主要基于深度神经网络架构实现。文献[8]提出深度置信网络(Deep Belief Network, DBN)来处理多任务分配问题,使用深度置信网络提取交通流量数据的时间特征,同时利用多任务回归层提取交通流量数据的空间特征。文献[9]对大数据下的交通流量预测进行综述,并从交通系统网络的角度构建堆叠式自动编码器(Stacked Automatic Encoder, SAE)模型来学习通用交通流量特征。文献[10]将有监督的深度学习理论扩展到大规模的交通网络分析中,使用深度受限的玻尔兹曼机来建模数据的空间相关性,同时使用循环神经网络(Recurrent Neural Network, RNN)来建模数据的时间相关性。文献[7]提出将不同路段间的关系矩阵融合到长短期记忆(Long Short Term Memory, LSTM)网络中,利用 2D-ODC-LSTM 网络的时空相关性进行建模。

本文提出一个新的端到端网络模型,其所有参数均为自主学习,不需要人工干预。该网络模型选用具有集成性质的残差 LSTM^[11]作为网络主框架,同时在每个 LSTM 单元后引入一个维度加权单元来显式建模信息在网络传递过程中特征维度间的依赖关系。

1 短期交通流量预测方法

1.1 网络结构

考虑到不同路段的历史交通流量数据本质上是分布有差异的时间序列数据,并且不同路段在空间上的相关性最终会反映在交通流量统计数据的时间相关性上。因此,本文利用集成学习思想,设计一个具有集成性质的神经网络,通过将空间分布的所有数据都输入到网络中进行训练,使该网络模型能够自主建模交通流量数据的时空相关性。同时,模型对不同时间、路段产生的所有车流量数据具有强大的泛化能力。

基于集成性质的残差 LSTM 网络框架有利于时间依赖性描述,同时具有较好的泛化能力。为显式

建模信息在网络传递过程中特征维度上的依赖关系,增强模型表现力,在每个 LSTM 单元后引入一个维度加权单元。维度加权网络结构如图 1 所示,其中, X_0 为在 LSTM 网络前输入的特征向量, t 为不同的时刻, n 为 LSTM 的输出维度, X_1 为时间序列经过每个 LSTM 层得到的输出特征, X_2 为原始输入特征与权重相乘后得到重标定后的特征, GAP 表示全局平均池化层, ReLU 代表非线性整流函数, FC 表示全连接层。

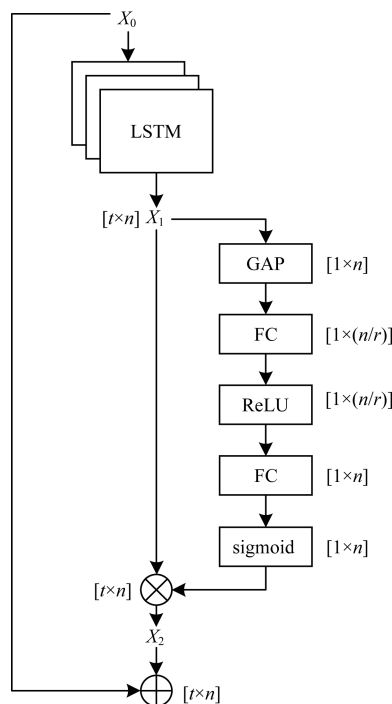


图 1 维度加权网络结构

注意力机制可以通过隐藏层计算输入信号在信息传递过程中的分类分布,以此对信息进行重新过滤。传统注意力机制一般是通过门限函数(如 softmax、sigmoid)来调节信息在模型中的传递。本文提出的维度加权单元是一个轻量级门控机制。

维度加权单元是一个可以安插在任意 LSTM 层后的计算单元,能够显式建模特征维度间的依赖关系。通过引入少量参数来学习并表示信息在传递过程中每个特征维度的权重,然后依据该权重来增强对当前任务有用的特征并抑制其他无用的特征。

1.2 预测过程

在每个 LSTM 后连接一个维度加权单元,同时在每个 LSTM 和维度加权单元后进行残差连接。残差连接的 LSTM 同时要求隐藏单元的维度是相同的,将隐藏单元的输出维度设为超参数 n 。

时间序列经过每个 LSTM 层得到的输出为一个 $[t \times n]$ 的二维特征向量 $\mathbf{x}_1$ 。

$$\mathbf{x}_1 = [\mathbf{x}_1^1, \mathbf{x}_1^2, \dots, \mathbf{x}_1^t] \quad (1)$$

其中, t 为不同时刻, $\mathbf{x}_1^t$ 有 n 个维度, n 为 LSTM 的输出维度。

为解决时间序列在特征维度上的依赖问题,首先考虑输出特征在每个特征维度上的信息。LSTM 单元会主动“遗忘”和“更新”时间维度上的信息,但是 LSTM 并不会在其隐藏单元的特征维度之间做出相应计算。因此,每个 LSTM 的输出特征并没有利用到其隐藏单元的特征维度之间的相关信息。针对该问题,首先将时间上的全局特征沿着时间维度进行特征压缩。将 $1 \sim t$ 时刻的每一维压缩成一个实数,该实数表示特征在时间维度上响应的全局分布。原来的 $[t \times n]$ 维 $\mathbf{x}_1^t$ 转化为 $[1 \times n]$ 维的向量 $\tilde{\mathbf{x}}_1$ 。

$$\tilde{\mathbf{x}}_1 = \frac{1}{t} \sum_{i=1}^t \mathbf{x}_1^i \quad (2)$$

该操作能够有效提取主要特征,同时增大感受野,减少计算量。

为对信息传递过程中各特征维度的依赖关系进行建模,选择 2 个全连接层,引入少数参数进行学习,以达到使用一个灵活且非互斥的函数描述每个特征维度重要程度的目的。第 1 个全连接层的特征维度为 $\frac{1}{r}$, r 为引入的超参数,全连接层的激活函数为非线性整流函数(ReLU)。第 2 个全连接层恢复其维度,全连接层的激活函数为 sigmoid,如式(3)所示。

$$\hat{\mathbf{x}}_1 = \sigma(w_2 \delta(w_1 \tilde{\mathbf{x}}_1)) \quad (3)$$

其中, δ 代表 ReLU 非线性整流函数, $w_1 \in \mathbb{R}^{\frac{n}{r}}$, $w_2 \in \mathbb{R}^n$, σ 代表 sigmoid 激活函数。

该操作既可以引入非线性函数,拟合特征维度间复杂的依赖性,又可以减少参数,降低计算量,限制模型复杂度。 $\tilde{\mathbf{x}}_1$ 通过 sigmoid 函数后可以获得 $0 \sim 1$ 之间的归一化权重。原始的输入特征与权重相乘后便得到重标定的特征,如式(4)所示。

$$\mathbf{x}_2 = \mathbf{x}_1 \cdot \hat{\mathbf{x}}_1 \quad (4)$$

基于此结构,为短期交通流量预测建立维度加权的残差 LSTM 网络。考虑到短期交通流量的预测时间延迟情况,本文测试了多个时间步长的数据来检验模型的鲁棒性。在此本文使用之前的历史观测数据, $\{\dots, \mathbf{x}_{t-2}, \mathbf{x}_{t-1}, \mathbf{x}_t\}$ 预测了 1 个 ~ 3 个时间步长的短期交通流量 $\{\tilde{\mathbf{x}}_{t+1}, \tilde{\mathbf{x}}_{t+2}, \tilde{\mathbf{x}}_{t+3}\}$ 。

模型训练是确定模型参数的过程。在本文模型中,隐藏层维度、迭代次数、网络层数、维度加权单元中降维的倍数均为超参数。这些超参数经人为设定,而其他参数 θ (包括模型的所有权重和偏差)是通过梯度下降算法训练得到。将整个数据集分为 3 个部分:训练集,验证集和测试集,目的是为通过训练集的数据进行训练来确定 θ ,再通过计算验证集中的损失函数来确定超参数,最后评估模型在测试集中的性能。通过梯度下降训练来调整 θ 以解决如式(5)所示的优化问题。

$$\theta = \underset{\theta}{\operatorname{argmin}} L(\mathbf{x}, \tilde{\mathbf{x}}) \quad (5)$$

其中, $\mathbf{x}$ 为历史的短期交通流量数据, $\tilde{\mathbf{x}}$ 为其对应一个时间步长的短期交通流量数据, L 为损失函数。损失函数如式(6)所示。

$$L(\mathbf{x}, \tilde{\mathbf{x}}) = \frac{\sum_{i=0}^n (\mathbf{x}_{t+1} - \tilde{\mathbf{x}}_{t+1})^2}{2n} \quad (6)$$

其中, $\mathbf{x}_{t+1}$ 为 $t+1$ 时刻的真实短期交通流量值, $\tilde{\mathbf{x}}_{t+1}$ 为网络在 $t+1$ 时刻的预测值。本文使用梯度下降方法来优化损失函数,使其值达到最小。

2 实验结果与分析

本文方法在配备英特尔 i7 3.4 GHz CPU, 32 GB 内存和 NVIDIA 1080 GPU 的实验平台上进行验证。首先将交通流量预测结果与原始流量数据进行比较,然后分别与其他预测方法进行对比实验,最后讨论维度加权单元中激活函数的选择和超参数的取值对实验结果的影响,从而验证提出模型的准确率。实验中模型将预测未来 15 min、30 min 和 45 min 的交通流量。

2.1 数据集

实验基于英国高速公路交通流量数据集^[12]。在该数据集中,交通流量使用多来源数据组合计算,包括自动车牌识别摄像机、车载 GPS 和内置在路面上的感应线圈等。数据集中的交通流量是每个路段 15 min 内的总交通流量。本文利用 2015 年 3 月 1 日—2015 年 3 月 31 日 M25 高速公路的 65 条路段,共 193 440 条有效交通流量数据来验证本文提出网络模型的有效性。同时,将原始数据集分为 3 个子集:2015 年 3 月 1 日—2015 年 3 月 18 日的数据作为训练集,2015 年 3 月 19 日—2015 年 3 月 24 日的数据作为验证集,2015 年 3 月 25 日—2015 年 3 月 31 日的数据作为测试集。

2.2 实验评价指标

评价交通流量短期预测模型的指标通常为平均相对误差 (Mean Relative Error, MRE)^[7]。该评价指标的定义如下:

$$MRE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{\varphi}_i - \varphi_i}{\varphi_i} \right| \quad (7)$$

其中, $\hat{\varphi}_i$ 为预测值, φ_i 为真实值。

2.3 结果分析

在该模型中每个 LSTM 有 500 个隐藏单元, 同时每个 LSTM 设置了 dropout 为 0.5 来减少过拟合, 网络使用 Adam 优化器, 迭代次数为 500 次。

2.3.1 短期交通流量预测及精度分析

当模型预测 15 min 内的交通流量时, 从 2015 年 3 月 25 日—2015 年 3 月 31 日的 65 条路段的交通流量预测结果和真实交通数据如图 2 所示。对比结果表明, 预测的交通流量与观测的交通流量具有相似的交通模式, 预测结果接近真实数据, 其 MRE 为 0.899%。根据预测结果验证了本文方法在实际应用中对于交通流量预测的有效性及其可靠性。

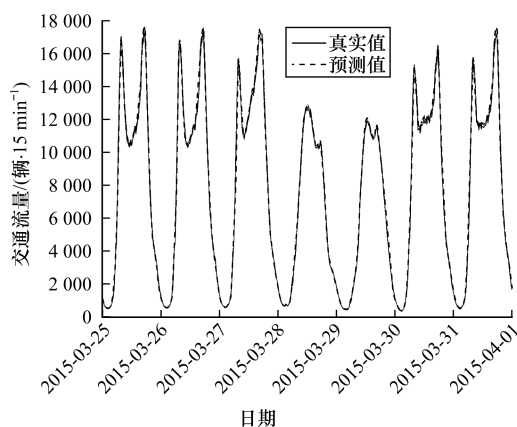


图2 预测交通数据与真实交通数据对比结果

为验证本文提出模型效果, 与支持向量机 (Support Vector Machine, SVM)^[13]、梯度提升回归树 (Gradient Boosting Regression Tree, GBRT)^[14]、差分整合移动平均自回归 (Autoregressive Integrated Moving Average, ARIMA)、DBN^[8]、RNN^[15]、堆叠 (Stacked) LSTM^[16]、2D-ODC-LSTM^[7] 模型进行对比实验。其中, Res-LSTM 为没有维度加权单元的模型, Res-OW-LSTM 为含有维度加权单元的模型, Stacked-OW-LSTM 为含有维度加权单元的 Stacked LSTM 模型。表 1 表示使用不同方法对 65 个路段进行交通流量预测的实验结果。

表1 不同模型对路段交通流量预测的 MRE %

模型	预测时间为 15 min	预测时间为 30 min	预测时间为 45 min
SVM	5.675	15.767	36.597
GBRT	5.234	15.154	33.381
ARIMA	7.139	16.431	32.645
DBN	5.618	15.392	29.176
RNN	3.302	14.021	27.559
Stacked LSTM(6 层)	4.133	13.174	26.501
Stacked-OW-LSTM(6 层)	3.562	12.667	26.219
Stacked LSTM(3 层)	2.963	11.718	24.634
Stacked-OW-LSTM(3 层)	2.451	10.995	24.317
2D-ODC-LSTM	1.811	8.717	21.842
Res-LSTM(6 层)	1.568	6.972	19.406
Res-OW-LSTM(6 层)	0.899	5.247	18.628

由表 1 可知, Res-OW-LSTM 的性能优于上述比较的所有模型, 特别是在预测时间较长的情况下, 优势更为明显。对比 2D-ODC-LSTM, Res-LSTM 在减少手工特征的嵌入后, 残差连接方式提供的集成性质使网络能够具有更强的泛化能力, 预测结果也更准确。在 LSTM 的层数为 6 时, Res-LSTM 平均相对误差为 1.568%, 大约为同等参数下 Stacked LSTM 平均相对误差的 1/3, 证明了残差连接的方式能够有效减轻“梯度消失”的情况, 同时增强了网络模型泛化能力。对于 Stacked LSTM 来说, 当 LSTM 的层数从 3 增加到 6 时, Stacked LSTM 的反向传播过程会受到阻碍, 会出现“梯度消失”的情况导致其预测精度下降。同时, 从表 1 中可以观察到, 无论是 Res-LSTM 还是 Stacked LSTM, 含有维度加权单元的网络模型预测误差均会比不含维度加权单元的网络模型预测误差低, 证明了维度加权单元的有效性。

当预测时间不超过 15 min 时, RNN 预测结果相对准确, 但是随着预测时间增长, 其误差会大幅增加, 表明 RNN 在长时间序列问题上通常会失效。SVM、GBRT 作为传统机器学习方法, 与其他深度学习方法 (如 DBN、RNN 和 LSTM) 相比, 对交通流量预测的泛化性不强。此外, ARIMA 具有更明显的预测误差, 表明经典参数化方法对于大规模交通数据的适应性较差。

2.3.2 超参数选择及精度分析

本文分析维度加权单元中激活函数的选择和超参数的取值对实验结果的影响。表 2 为选择不同参数时的实验结果。

表2 不同参数对路段交通流量预测的 MRE %

模型	预测时间 为 15 min	预测时间 为 30 min	预测时间 为 45 min
Res-LSTM	1.568	6.972	19.406
Res-OW-LSTM($\tanh$, $r=16$)	1.478	6.744	19.693
Res-OW-LSTM(ReLU, $r=8$)	1.133	6.402	19.531
Res-OW-LSTM(ReLU, $r=16$)	0.899	5.247	18.628
Res-OW-LSTM(ReLU, $r=32$)	1.363	6.460	19.566
Res-OW-LSTM(ReLU, $r=64$)	1.479	5.935	18.853
Res-OW-LSTM(ReLU, $r=128$)	1.561	6.859	20.301
Res-OW-LSTM(ReLU, $r=256$)	1.055	6.349	18.890
Res-OW-LSTM(ReLU, $r=320$)	1.208	6.419	19.631
Res-OW-LSTM(ReLU, $r=500$)	1.423	6.572	19.738

由表2可知,在维度加权单元中,激活函数为 $\tanh$ 时,其结果并没明显变化,但是激活函数为ReLU时,其预测结果有明显提高。当LSTM的隐藏单元为500、维度加权结构中的特征维度 r 为16时取得了最好的预测结果,MRE相对于基础Res-LSTM分别降低了0.669%、1.725%、0.778%。针对维度加权单元中激活函数的选择问题,本文认为 $\tanh$ 激活函数在RNN神经元内部的循环过程中会不断增强特征,但是当其在层间进行反向传播时,由于 $\tanh$ 函数在 $[-1,1]$ 间收敛,因此容易出现梯度消失的情况,从而无法完成深层网络训练。而ReLU激活函数在层间传递不会出现该现象。同时,根据实验结果可得出,维度加权单元中超参数的变化并不明显,相对于没有维度加权单元的Res-LSTM来说误差都有降低,证明了维度加权单元的有效性。

3 结束语

短期交通流量预测是实现交通智能化的一个关键问题之一。本文提出基于维度加权的残差LSTM短期交通流量预测方法。通过设计一个端到端的网络,将空间分布的交通流量数据同时训练到一个残差LSTM网络中,使网络能自主建模交通流数据的时空依赖关系。同时,在每个LSTM单元后引入一个维度加权单元,显式建模信息在网络传递过程中特征维度依赖关系。实验结果表明,本文提出的网络模型对不同路段的数据均具有较强的泛化能力,能够鲁棒解决短期交通流量预测问题。下一步可将图卷积神经网络用于交通流量预测,使模型具有更强的可解释性。

参考文献

- [1] HUANG Wenhao, SONG Guojie, HONG Haikun, 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] 贾显超,陈旭梅,弓晋丽,等.基于混沌理论的短期交通流量多步预测[J].交通信息与安全,2013,31(6):27-32.

- [3] 曹政才,韩丁富,王永吉.面向城市交通网络的一种新型动态路径寻优方法[J].电子学报,2012,40(10):2062-2067.
- [4] 韩志聪,樊彦国,吴会胜,等.基于GA-SVR模型的短期交通流量预测方法研究[J].公路交通科技,2017,34(1):130-136.
- [5] DU Shengdong, LI Tianrui, GONG Xun. Traffic flow forecasting based on hybrid deep learning framework[C]//Proceedings of International Conference on Intelligent Systems and Knowledge Engineering. Washington D. C., USA:IEEE Press,2018:1-6.
- [6] JEONG Y S, BYON Y J, CASTRO-NETO M M, et al. Supervised weighting-online learning algorithm for short-term traffic flow prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(4):1700-1707.
- [7] ZHAO Zheng, CHEN Weihai, WU Xingming, et al. LSTM network; a deep learning approach for short-term traffic forecast[J]. IET Intelligent Transport Systems, 2017, 11(2):68-75.
- [8] HUANG Wenhao, SONG Guojie, HONG Haikun, 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.
- [9] LÜ Yisheng, DUAN Yanjie, KANG Wenwen, et al. Traffic flow prediction with big data; a deep learning approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(2):865-873.
- [10] MA Xiaolei, YU Haiyang, WANG Yunpeng, et al. Large-scale transportation network congestion evolution prediction using deep learning theory[J]. PLoS One, 2015, 10(3):1-17.
- [11] PRAKASH A, HASAN S A, LEE K, et al. Neural paraphrase generation with stacked residual LSTM networks[EB/OL]. [2018-03-20]. <https://www.aclweb.org/anthology/C16-1275>.
- [12] Live traffic information from highways England[EB/OL]. (2015-12-10) [2018-03-20]. <https://data.gov.uk/dataset/live-traffic-information-from-the-highways-agency-road-network>.
- [13] PENG Zhenrui, GAO Pu, MENG Jianjun. Traffic forecasting using least squares support vector machines[J]. Transportmetrica, 2009, 5(3):193-213.
- [14] ZHANG Yanru, HAGHANI A. A gradient boosting method to improve travel time prediction[J]. Transportation Research Part C: Emerging Technologies, 2015, 58:308-324.
- [15] LINGRAS P, SHARMA S, ZHONG M. Prediction of recreational travel using genetically designed regression and time-delay neural network models[EB/OL]. [2018-03-20]. https://www.researchgate.net/publication/241761639_Prediction_of_Recreational_Travel_Using_Genetically_Designed_Regression_and_Time-Delay_Neural_Network_Models.
- [16] DUAN Yanjie, LÜ Yisheng, WANG Feiyue. Travel time prediction with LSTM neural network[C]//Proceedings of the 19th International Conference on Intelligent Transportation Systems. Washington D. C., USA: IEEE Press, 2016:1053-1058.