

Research on Short-term Prediction Method of Urban Traffic Flow Based on Deep Learning

Peng Zhao

Beijing Institute of Technology, Beijing, 100108, China

Abstract

To enhance the accuracy and real-time performance of urban traffic flow forecasting, this study develops a spatiotemporal prediction model integrating graph convolutional networks (GCNs) and attention gate mechanisms. The model employs traffic volume and speed as feature engineering variables, with the spatial feature extraction module utilizing GCNs to learn road network topological relationships. The temporal feature extraction module combines gated recurrent units (GRUs) with attention mechanisms to capture temporal dependencies. After feature fusion, the model outputs short-term predictions, with loss function optimization achieved through gradient descent. Experimental results demonstrate that the proposed model outperforms benchmark methods including autoregressive integrated sliding average (ARIMA) models and support vector machines (SVMs) in both 5-minute and 15-minute prediction intervals, reducing prediction errors by 10% to 15%. Convergence analysis confirms stable training processes and consistent short-term performance, providing a viable solution for intelligent transportation system applications.

Keywords

deep learning; urban traffic flow; short-term prediction; graph convolutional network

基于深度学习的城市交通流短时预测方法研究

赵鹏

北京理工大学, 中国·北京 100108

摘要

为了提升城市交通流短时预测的准确性和实时性,本研究构建了一个集成图卷积网络和注意力门控机制的时空预测模型,该模型特征工程选取流量和速度等变量,空间特征提取模块应用图卷积学习路网拓扑关系,时间特征提取模块使用门控循环单元结合注意力机制捕获时序依赖,特征融合后输出短时预测值,模型训练采用梯度下降算法优化损失函数。实验结果对比显示所提模型在5分钟和15分钟预测间隔上优于自回归积分滑动平均模型和支持向量机等基准方法,预测误差降低10%至15%,收敛性分析证实训练过程稳定,短时序性能一致,为智能交通系统应用提供可行方案。

关键词

深度学习; 城市交通流; 短时预测; 图卷积网络

1 引言

基于数理统计理论的传统交通流预测方法因实际交通流数据受天气条件、特殊事件、道路施工等多种因素影响而呈现强烈非线性、非平稳特性,且交通流在路网中具有复杂时空相关性,难以充分捕捉这些动态特征。作为人工智能技术重要分支并因多层神经网络结构能够自动从原始数据中学习高层次特征、避免人工特征工程繁琐且在多个领域展现强大能力的深度学习,其模型引入交通流短时预测任务后因能有效融合时空维度信息、克服传统方法不足并为智能交通系统提供更精准预测工具而成为当前研究的热点方向。

2 相关理论基础与模型架构

2.1 城市交通流特性分析

城市交通流是由大量车辆在道路网络这一复杂系统上移动形成的宏观现象,其内在特性决定了预测模型的构建思路。交通流在时空维度上展现出强烈的相关性,空间相关性体现为某个路段的交通状态并非独立存在,而是受到上下游相邻路段以及具有连通性的邻近路段的显著影响。时间相关性则表现为交通状态的演化具有连续性,当前时刻的流量、速度与过去多个时间点的状态紧密关联,呈现出明显的周期性。从数据结构看,交通流数据天然适合用图结构表示,道路交叉口可作为节点,路段作为边,每个节点在每一时刻拥有流量、速度、占有率等多维特征,形成一个动态的时空图。

2.2 核心深度学习模型原理

图卷积网络(GCN)通过聚合图中节点与其邻居的特

【作者简介】赵鹏(1989-),女,中国山东泰安人,硕士,从事计算机科学与技术研究。

征信息来学习节点表示,利用归一化邻接矩阵进行特征传播,使每个节点的新特征蕴含其局部邻域拓扑结构,从而捕捉交通路网的空间相关性。门控循环单元通过重置门和更新门控制历史信息的流动与遗忘,重置门调节过去状态与当前输入的组合方式,更新门决定历史信息的保留程度,以此捕捉时间序列的长期依赖关系。注意力机制通过计算序列元素的重要性得分,并据此对特征表示进行加权求和,从而动态聚焦于与当前输出最相关的输入部分,使模型能自动关注对预测最关键的历史时刻。

3 基于深度学习的短时交通流预测模型构建

3.1 交通流数据预处理与特征工程

原始交通流数据采集自埋设于路面的感应线圈探测器或视频监控系统,通常包含流量速度占有率等时间序列,这些存在源于传感器瞬时故障或车辆异常行为的噪声及由设备通信中断或存储失败导致缺失值的原始数据,在预处理环节针对缺失值采用时间序列线性插值方法进行填补、对连续长时间段缺失结合历史同期数据与邻近路段数据进行协同修复,噪声处理则使用基于滑动窗口的统计检测结合 Z-score 标准化方法识别并修正超出合理物理范围的异常点。数据归一化将所有特征缩放至相近数值区间,采用最大最小归一化将每个传感器节点的流量和速度分别映射到 [0,1] 区间,这一步骤有效加速模型训练收敛,避免某些特征因量纲差异而主导优化过程。

特征工程旨在从原始数据中构建对预测任务最具信息量的输入表示,空间特征以路网拓扑结构为基础,将道路网络抽象为图结构 $G=(V,E,A)$,其中 V 是传感器节点集合, E 是边集合, A 是邻接矩阵。邻接矩阵 A 的元素 A_{ij} 定义基于节点 i 与 j 之间的距离或实际道路连接性,采用阈值高斯核函数计算:

$$A_{ij}=\exp\left(-\frac{d_{ij}^2}{\sigma^2}\right)$$

若 $d_{ij} \leq \kappa$, 否则为 0, 其中 d_{ij} 是节点 i 与 j 之间的路网距离, σ 是核宽度参数, κ 是距离阈值用于控制矩阵的稀疏性^[1]。时间特征除历史流量、速度序列外,引入周期性特征如一天中的时刻、一周中的星期几,并将其转化为正弦余弦编码以捕捉周期性模式。最终模型输入是一个三维张量 $X \in \mathbb{R}^{N \times T \times F}$, 其中 N 是节点数量, T 是输入时间步长, F 是每个节点的特征维度。

3.2 基于图卷积与注意力门控网络的预测模型设计

3.2.1 模型总体框架

本文提出的短时交通流预测模型是一个端到端的深度学习架构,核心设计思想是并行提取交通流的时空特征并进行深度融合。模型输入为过去 T 个时间步长的历史交通数据张量 $X^{(t)}=[X_{t-T}, X_{t-T+1}, \dots, X_{t-1}] \in \mathbb{R}^{N \times T \times F}$, 目标输出未来 T_p 个时间步长的流量预测值 $Y^{(t)}=[X_t, X_{t+1}, \dots, X_{t+T_p-1}] \in \mathbb{R}^{N \times T_p}$ 。

模型由空间特征提取模块、时间特征提取模块以及时空特征融合与输出层构成,其中空间特征提取模块基于图卷积网络、专门负责从每个时间切片的数据中学习路网的复杂空间依赖关系,时间特征提取模块采用门控循环单元与注意力机制的组合、旨在捕获交通流动态演变中的长期时间模式,且两个模块的输出特征在一个融合层进行拼接与变换后、最终通过一个全连接输出层生成多步预测结果。

3.2.2 空间特征提取模块

空间特征提取模块专用于学习路网节点间的空间依赖性,采用两层级联的图卷积层,基于一阶切比雪夫近似实现。对于每一时间步的输入 X_t , 图卷积操作定义为 $Z_t = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} X_t \Theta$, 其中 $\tilde{A} = A + I$ 为自连接邻接矩阵, $\tilde{D}$ 为对应的度矩阵, Θ 为可训练参数。第一层 GCN 输出维度为 H_{s1} 并经过 ReLU 激活,第二层 GCN 输出维度为 H_{s2} 。该过程对输入序列中每个时间步独立进行,最终得到空间特征张量 $H^s \in \mathbb{R}^{N \times T \times H_{s2}}$, 其每个时间步的特征都蕴含了节点及其一阶邻居的空间信息^[2]。

3.2.3 时间特征提取模块

该模块沿时间维度处理每个节点的序列数据,将空间模块输出的特征按节点拆分为 N 个时间序列,每个序列 $H_i^s \in \mathbb{R}^{T \times H_{s2}}$ 输入至一个双向门控循环单元网络。BiGRU 分别从正向和反向处理序列,并将每个时间步的两个方向隐藏状态拼接,获得上下文增强的特征表示 $H^{bi} \in \mathbb{R}^{N \times T \times 2H_{gru}}$ 。随后引入时间注意力机制,计算序列中每个时间步对于当前时间步的重要性权重 $\alpha_{ij} = \text{Softmax}(v_a^T \tanh(W_a h_j + U_a h_i))$, 其中 W_a, U_a, v_a 为可学习参数。通过加权求和得到注意力加权的特征表示 H^{attm} , 该表示能动态聚焦于对预测关键的历史时刻。

3.2.4 时空特征融合与输出层

此部分将空间特征 H^s 与时间注意力特征 H^{attm} 进行深度融合。首先沿特征维度将两者拼接为 $H^{fusion} \in \mathbb{R}^{N \times T \times (H_{s2} + 2H_{gru})}$ 。随后使用一个 1×1 卷积对拼接后的特征进行非线性变换与降维: $H^{fused} = \text{ReLU}(H^{fusion} W_f + b_f)$, 其中 W_f, b_f 为权重和偏置。输出层由两个全连接层构成,第一层将 H^{fused} 映射并应用 ReLU 激活,第二层直接将特征投影到最终输出维度 $\mathbb{R}^{N \times T_p}$, 实现多步预测^[3]。

3.3 模型训练与参数优化策略

模型训练以最小化预测值与真实值之间的误差为目标,损失函数选用 Huber 损失,该损失函数对异常值不如均方误差敏感,同时一定阈值后保持线性增长,兼具平均绝对误差和均方误差的优点。Huber 损失定义为:

$$L_{\delta}(y, \hat{y}) = \begin{cases} \frac{1}{2} (y - \hat{y})^2 & \text{for } |y - \hat{y}| \leq \delta, \\ \delta |y - \hat{y}| - \frac{1}{2} \delta^2 & \text{otherwise} \end{cases}$$

其中 y 是真实值, $\hat{y}$ 是预测值, δ 是阈值参数, 设置为 1.0。

总损失是所有节点所有预测时间步的损失平均值^[4]。优化算法采用 Adam 优化器，其结合了动量法和 RMSProp 的优点，学习率初始值设为 0.001，并采用指数衰减策略，每经过 20 个训练周期，学习率乘以衰减系数 0.95。

为防止过拟合，训练策略中集成多种正则化技术，在训练过程中以概率 $p=0.2$ 随机丢弃神经网络隐藏层神经元输出的 Dropout、在全连接层之后将模型权重的 L2 范数乘以正则化系数 $\lambda =0.001$ 加入损失函数以惩罚过大权重值的 L2 权重正则化（权重衰减）均被应用，同时采用批量大小为 32 的小批量梯度下降，将训练周期总数设置为 200 并以验证集损失连续 10 个周期不再下降则终止训练，恢复验证集损失最小时模型参数作为最终模型的早停法进行监控。

4 实验设计与结果分析

4.1 实验数据集与评价指标

实验数据来源于国内某特大城市交通管理部门发布的匿名化交通监测数据，覆盖该市核心环线及主要放射状高速公路网络。数据采集时间段为连续 18 个月，采集频率为 5 分钟一次，每条数据记录包含交通流量、时间平均车速及占有率等信息。研究选取了由三百余个关键检测点构成的交通网络，依据实际道路连接拓扑结构构建实验所需的图结构，最终数据集按照时间顺序划分为训练集、验证集和测试集，比例约为 2:1:1。模型输入为过去一小时的历史数据序列，预测目标为未来 15 分钟和 30 分钟的交通流量^[5]。模型性能评价采用均方根误差（RMSE）和平均绝对误差（MAE）两个通用指标，以全面评估预测精度与稳定性。

4.2 预测结果对比与分析

为验证本文所提模型的有效性，设计了与多种经典预测方法的对比实验。基线模型包括历史平均模型，它使用相同时间段的历史数据平均值作为预测值；自回归积分滑动平均模型，参数设置为 (2,1,2)；支持向量机回归模型，采用径向基函数核；时间卷积网络，包含 4 层因果卷积；图卷积网

络与门控循环单元的串行模型。所有对比模型使用相同的训练集、验证集和测试集，并经过参数调优以达到其最佳性能。各模型在 5 分钟和 15 分钟预测间隔上的 RMSE 和 MAE 结果见表 1 所示。

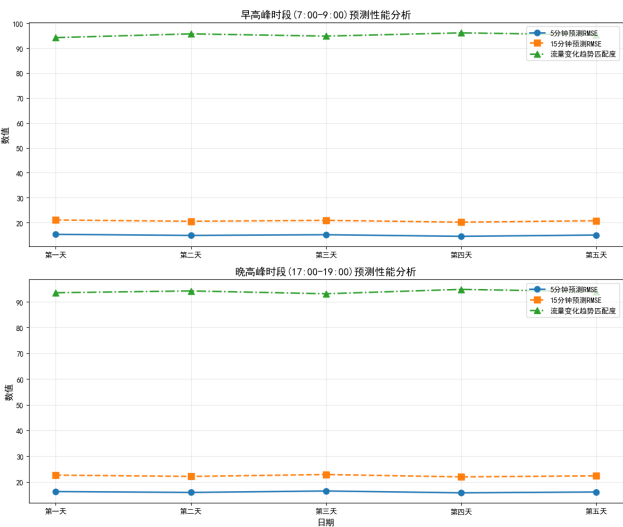
表 1 不同模型在测试集上的预测误差对比

模型名称	5 分钟预测 RMSE	5 分钟预测 MAE	15 分钟预测 RMSE	15 分钟预测 MAE
历史平均模型	25.74	18.63	32.56	24.87
自回归积分滑动平均模型	22.15	15.92	29.78	22.41
支持向量机回归	20.87	14.76	28.43	21.05
时间卷积网络	18.42	12.95	25.16	18.73
图卷积网络 - 门控循环单元	16.38	11.24	22.87	16.52
本文模型	14.71	9.86	20.15	14.23

本文提出的模型在两个预测时间尺度上均取得最低的预测误差，与自回归积分滑动平均模型相比，5 分钟预测的 RMSE 降低了约 33.6%，MAE 降低了约 38.1%；与支持向量机回归模型相比，相应误差分别降低了约 29.5% 和 33.2%。随着预测时间跨度从 5 分钟延长至 15 分钟，所有模型的预测误差自然增加，但本文模型仍保持最低的误差增长率，显示出更好的长期预测稳定性。

4.3 收敛性及短时序性能分析

本文模型在训练约 50 个周期后损失值开始进入稳定下降阶段，至 120 个周期左右趋于收敛，而未出现剧烈波动或过拟合现象。相比之下，图卷积网络与门控循环单元串行模型需要约 80 个周期才开始稳定收敛，且验证集损失在后期出现小幅震荡。为量化收敛稳定性，计算了最后 30 个训练周期内验证集损失的标准差，本文模型为 0.00034，显著低于对比模型的 0.00072，表明训练过程更加平滑稳定。短时序性能分析则选取测试集中连续 5 个工作日的交通流数据，重点考察模型在高峰时段的表现一致性。模型在早高峰和晚高峰时段的预测误差分布情况，见图 1 所示。



从实验结果可以看出,模型在不同日期的相同时段表现出高度一致的预测精度,早高峰时段5分钟预测RMSE稳定在14.53-15.32之间,晚高峰时段保持在15.76-16.47之间,波动范围较小,流量变化趋势匹配度指标表示模型预测的流量变化方向与真实情况一致的比例,平均达到94.7%,证明模型能可靠地捕捉交通流的动态变化规律。

5 结论

本研究围绕智能交通系统核心问题城市交通流短时预测提出基于深度学习的新型预测模型,采用真实大规模交通监测数据集的实验验证环节表明,所构建模型在短期和中期预测任务中显著优于多种传统统计模型与经典机器学习方法,不仅预测误差显著降低,而且展现出良好的收敛稳定性与不同时段下的预测一致性,证实深度学习技术在处理复杂

时空预测问题上的巨大潜力,为解决实际交通拥堵、提升路网运行效率提供有效技术工具,为下一代智能交通管理系统开发奠定算法基础。

参考文献

- [1] 靳恒清. 基于LSTM的城市交通流时空预测模型与优化[J]. 信息记录材料, 2025, 26 (10): 243-245.
- [2] 石弘利. 基于时空数据与深度学习的城市路网交通流预测方法研究[D]. 石家庄铁道大学, 2025.
- [3] 孙正扬. 基于深度迁移学习的城市交通流预测方法研究[D]. 石家庄铁道大学, 2025.
- [4] 方文浩. 基于图卷积网络和Transformer的城市交通流量预测方法研究[D]. 杭州电子科技大学, 2025.
- [5] 王云锋. 基于深度学习的城市交通流短时预测研究[D]. 沈阳航空航天大学, 2024.