

DOI: 10.13382/j.jemi.B2205273

基于改进 VMD-GAT-GRU 的交通流量组合预测模型*

殷礼胜 吴洋洋

(合肥工业大学电气与自动化工程学院 合肥 230009)

摘要:针对短时交通流时间序列非平稳性、空间相关性和时间依赖性的特点,为提高短时交通流预测模型的预测精度和收敛速度,该文提出了一种基于改进的变分模态分解(VMD)、图注意力(GAT)网络和门控循环单元(GRU)网络的交通流量组合预测模型。首先,利用互信息熵(MI)改进的变分模态分解算法,将交通流量时间序列分解成一系列调幅调频信号子序列,降低了时序信号的非平稳性,提高后续预测模型的预测精度;然后,将其输入图注意力网络,捕捉路网邻近节点的交通流量对中心预测节点交通流量不同程度的影响,从而实现交通流量序列的空间相关性建模,进一步提高模型预测精度;接着,将交通流量分量子序列分别送入门控循环单元网络,捕捉其时间依赖性,并使用改进的 RMSPRop 优化算法迭代寻优,在提升优化算法收敛速度的同时提高了模型的预测精度;最后,结合各分量子序列的预测值,作为预测模型的最终输出。实验采用 RTMC 系统交通数据,结果表明,该文提出的改进 VMD-GAT-GRU 时空融合组合预测模型相较于 LSTM、GCN 和 GAT 基准模型,平均绝对误差(MAE)分别降低 9.35、4.12、4.09,平均绝对百分比误差(MAPE)分别降低 16.42%、7.32%、8.1%,优化算法的收敛速度和组合模型的预测精度均得到有效提升。

关键词:交通流量预测;变分模态分解;互信息熵;图注意力网络;门控循环单元网络

中图分类号: U491.1 **文献标识码:** A **国家标准学科分类代码:** 580.2099

Traffic flow combination prediction model based on improved VMD-GAT-GRU

Yin Lisheng Wu Yangyang

(School of Electrical Engineering and Automation, Hefei University of Technology, Hefei 230009, China)

Abstract: For the characteristics of non-stationarity, spatial correlation and temporal dependence of short-term traffic flow time series, this paper proposes a combined prediction model of traffic flow based on improved variational mode decomposition (VMD), graph attention (GAT) network and gated recurrent unit (GRU) network to improve its prediction accuracy and convergence speed. First, the variable mode decomposition algorithm improved by mutual information entropy (MI) is used to decompose the traffic flow time series into a series of amplitude modulation and frequency modulation signal sub-sequences, which reduces the non-stationarity of the time series signal and improves the prediction accuracy of the model. Then, they are sent to the graph attention network to capture the traffic flow of adjacent nodes of the road network to different degrees on the traffic flow of the central prediction node, so as to realize the spatial correlation modeling and further improve the prediction accuracy of the combined model. Next, the traffic flow component sub-sequences are sent to the gated recurrent unit network separately to capture the temporal dependence of the traffic flow sequence, and use the improved RMSPRop optimization algorithm to iteratively search for optimization, which not only improves the convergence speed of the optimization algorithm, but also improves the prediction accuracy of the model. Finally, the prediction values of each component sub-sequences are combined as the final output of the prediction model. The experiment used traffic data from the RTMC system, the results show that compared with LSTM, GCN and GAT baseline models, the mean absolute error (MAE) is reduced by 9.35, 4.12 and 4.09, respectively, and the mean absolute percentage error (MAPE) is reduced by 16.42%, 7.32%, and 8.1%, respectively. The convergence speed of the optimization algorithm and the prediction accuracy of the combined model are effectively improved.

Keywords: traffic flow prediction; variational mode decomposition; mutual information entropy; graph attention network; gated recurrent unit

收稿日期: 2022-03-16 Received Date: 2022-03-16

* 基金项目: 国家自然科学基金(62073114, 6207022417)、安徽省自然科学基金(JZ2021AKZR0344)项目资助

0 引言

随着城市化和数字化发展,构建智能交通系统,发展自动驾驶和车路协同的出行服务,推广公路智能管理、交通信号联动^[1]成为“十四五”规划重点领域。而准确、及时的短时交通流预测是实现智能交通系统管控的基础之一。短时交通流是指在一段较短时间内(一般取 5 min 左右)通过某一道路断面的车辆数量,是短时交通参数的基础参数之一,常常称为交通流量。交通流量具有复杂的非平稳性、空间相关性和时间依赖性等特点,针对这些特点,各专家学者提出了诸多分析预测模型。

张晓晗等^[2]先利用经验模态分解对交通流量数据进行去噪,降低其非平稳性,再使用长短期记忆神经网络挖掘到交通流量时间序列的长期信息,捕捉到交通流量时序的时间依赖性,提升了预测精度,但忽略了交通路网的空间相关性对预测节点交通流量的影响;殷礼胜等^[3]利用非线性主成分分析和改进灰色组合模型,通过将多维交通流量序列进行时空降维,考虑了附近路段的空间相关性影响,再使用相空间重构的改进灰色模型,提高建模的精确度,但未能充分考虑到交通流量序列的长期时间依赖性,模型表达能力有限;Guo 等^[4]利用 3D 卷积神经网络和重校准块,使用真实数据集验证了模型在不同交通预测任务下的性能,考虑了交通流量序列的时空相关性,但并未考虑原始交通流量的非平稳性,且数据管理和模型应用是基于规则的栅格数据,忽略了一般交通路网基于不规则的图结构对模型预测精度的影响。

基于交通流时间序列的复杂非平稳性、空间相关性和时间依赖性的特点,以上研究,或运用循环神经网络等方法捕捉交通流量序列的时间依赖性;或通过降低交通流时序的非平稳性,再运用统计学习方法进行预测;或使用卷积神经网络方法模拟捕捉各节点之间交通流量的空间相关性。各模型均不同程度地提高了预测精度,但未考虑将降低交通流时间序列非平稳性和捕捉时间序列的空间相关性以及时间依赖性相结合的方法。综上所述,本文尝试提出一种基于改进的变分模态分解、图注意力路网以及门控循环单元网络相结合的交通流量组合预测模型。首先,使用互信息熵改进的变分模态分解算法将交通流量序列分解为一系列相对平稳的高低频分子序列,最大程度保留原始时间序列的信息,通过对原始时间序列数据的分解处理,降低其一般情况下具有的复杂非平稳性,以利于后续模型训练的收敛速度和预测精度的提高;然后,将选取的相关交通流节点的各分量序列作为图注意力网络的特征输入,利用图注意力网络可通过学习到的注意力系数对特征输入进行加权,从而对相关节点的各分量序列经不同程度的加权后输出,以捕捉

交通路网邻近节点交通流量对中心预测节点交通流量的有区别影响,实现对其空间相关性的考虑;最后,将各节点加权后的分量序列输入门控循环单元网络,实现对其时间依赖性的考虑,并通过改进的 RMSProp 优化算法迭代训练,将各分量预测值叠加求和作为交通流组合预测模型的最终预测输出。通过使用真实数据集对本文组合模型以及 LSTM、GCN 和 GAT 基准模型进行实验对比,以验证优化算法收敛速度的有效提升及组合模型预测精度的有效提高。

1 基于互信息熵改进的交通流量 VMD 分解模型

1.1 互信息熵原理

互信息 (mutual information, MI)^[5]是香农信息熵 (information entropy) 理论中用于度量随机变量之间统计依赖性的方法。设一组交通流量时间序列 $X = (x_1, x_2, \dots, x_n)$, 其边缘信息熵 $H(X)$ 定义为:

$$H(X) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (1)$$

式中: x_i 为 X 的交通流量数据, $p(x_i)$ 为其边缘概率密度函数。

再设另一组交通流量时间序列 $Y = (y_1, y_2, \dots, y_n)$, 则 X 与 Y 的联合信息熵 $H(X, Y)$ 定义为:

$$H(X, Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(x, y) \quad (2)$$

式中: $p(x, y)$ 为 X 与 Y 的联合概率密度函数。则 X 与 Y 的互信息熵 $I(X; Y)$ 定义为:

$$I(X; Y) = H(X) + H(Y) - H(X, Y) \quad (3)$$

1.2 交通流量 VMD 分解原理

变分模态分解 (variational mode decomposition, VMD)^[6-7]是一种新的自适应非平稳信号模态变分提取方法,可实现固有模态分量 (VIMF) 的有效分离、信号的频域划分,将给定信号有效分解为多个频率不同且相对平稳的子序列,从而有效降低具有复杂非线性时间序列的非平稳性。

通过使用二次惩罚项和 Lagrange 乘子,交通流量时间序列分解可视为求解以下无约束变分问题:

$$L(u_k, \omega_k, \lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \|f(t) - \sum_k u_k(t)\|_2^2 + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle \quad (4)$$

式中, $u_k(t) = A_k(t) \cos(\phi_k(t))$, 为交通流量分解后第 k 个模态分量,其中, $A_k(t)$ 、 $\phi_k(t)$ 分别为相应的幅值和相位, $\{u_k\} = \{u_1, u_2, \dots, u_K\}$; $\omega_k(t) \approx \phi_k'(t)$, 为交通流

量分解后的第 k 个模态分量的中心频率, $\{\omega_k\} = \{\omega_1, \omega_2, \dots, \omega_K\}$; $\delta(t)$ 为单位冲激函数; $*$ 为卷积运算符; ∂_t 为对时间 t 的偏微分; α 为二次惩罚项, 用于降低时间序列中高斯噪声干扰, 其权重由贝叶斯先验推导而来, 与噪声水平成反比; λ 为 Lagrange 乘子, 用于对约束的严格执行。

1.3 基于互信息熵改进的交通流量 VMD 分解算法

由于 VMD 分解算法对含有强烈噪声的信号敏感的特性, 使得一部分分量序列的中心频率 ω_k 可能出现在噪声频域附近, 而为达到信号降噪的效果, 有学者或直接滤除分解后的高噪分量^[8], 或筛选并保留包含原信号信息较多的分量^[9-10], 此类做法虽然可以有效降低原始信号序列的非平稳性, 但重构后的信号序列一定程度上会出现失真, 并且含噪分量中也可能含有有用信息^[11-12]。

考虑到上述问题, 为降低直接舍弃交通流量的含噪分量对最终预测模型的预测精度带来的影响, 本文尝试用一种基于互信息熵改进的 VMD 交通流量分解算法, 其基本思想是利用互信息熵来有效度量相邻两个交通流量序列间的关联性, 由互信息原理, 各相邻 VIMF 分量的互信息熵值会随着从低频到高频出现极小值, 以此作为交通流量子序列的低频分量和高频分量分界点, 保留低频分量子序列, 并对高频分量子序列进行阈值降噪, 抑制噪声干扰, 以最大程度减少原交通流量时间序列中有用信息的损失, 进而提高模型的预测精度。

算法 1: 基于互信息熵改进的交通流量 VMD 分解算法

步骤 1) 初始化交通流量各模态 $\{u_k^1\}$ 的幅值 $A_k(t) \geq 0$, 相位 $\phi_k(t)$ 非递减: $\phi_k'(t) \geq 0$; 中心频率 $\omega_k(t) \approx \phi_k'(t)$, 拉格朗日乘子 $\{\lambda_k^1\}$, 模态个数 K 以及 $n = 0$ 。

步骤 2) $n = n + 1$, 进入迭代内循环步骤 3) ~ 5)。 n 为搜索迭代次数。

步骤 3) 利用式(5)、(6)交替方向乘子迭代算法、傅里叶等距变换优化得到频域内的交通流量各模态分量 $\hat{u}_k^{n+1}(\omega)$ 和中心频率 ω_k^{n+1} , 搜寻增广拉格朗日目标函数式(4)的“鞍点”, 即此函数的海森矩阵在梯度为 0 的位置上的特征值有正有负时:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \hat{\lambda}(\omega)/2}{1 + 2\alpha(\omega - \omega_k)^2} \quad (5)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \quad (6)$$

式中: $\hat{u}_k^{n+1}(\omega)$ 、 $\hat{f}(\omega)$ 、 $\hat{u}_i(\omega)$ 和 $\hat{\lambda}(\omega)$ 分别为 $u_k^{n+1}(t)$ 、 $f(t)$ 、 $u_i(t)$ 和 $\lambda(t)$ 的傅里叶变换; ω_k^{n+1} 为对应模态子序

列功率谱的重心。

步骤 4) 至频域内交通流量序列 $\hat{f}(\omega)$ 分解个数为 K , 停止内循环, 由式(7)更新 Lagrange 乘子 $\hat{\lambda}$ 。

$$\hat{\lambda}^{n+1}(\omega) \leftarrow \hat{\lambda}^n(\omega) + \tau(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega)) \quad (7)$$

式中: τ 为噪声容限, 满足交通流量时间序列分解的保真度要求。

步骤 5) 对以上相邻两次迭代得到的第 k 个交通流量模态 u_k^{n+1} 和 u_k^n , 若不满足式(8)条件:

$$\sum_k \|u_k^{n+1} - u_k^n\|_2^2 / \|u_k^n\|_2^2 < \varepsilon \quad (8)$$

返回步骤 2), 继续循环; 否则停止迭代循环, 得到各模态 $\hat{u}_k^{n+1}(\omega)$ 和最佳中心频率 ω_k^{n+1} ; 式中 ε 为给定精度。

步骤 6) 对频域内的 $\hat{u}_k^{n+1}(\omega)$ 进行傅里叶逆变换, 得到的实部即为交通流量分量子序列 $vimf_k$ 。

步骤 7) 由式(9)计算第 k 个分量子序列 $vimf_k$ 的边缘信息熵:

$$H(vimf_k) = - \sum_{i=1}^K p(vimf_i) \log_2 p(vimf_i) \quad (9)$$

式中: 各分量个数之和 $\sum_k = K$ 。

步骤 8) 由式(10)计算相邻分量子序列 $vimf_k$ 、 $vimf_{k+1}$ 的互信息:

$$I(vimf_k, vimf_{k+1}) = H(vimf_k) + H(vimf_{k+1}) - H(vimf_k, vimf_{k+1}) \quad (10)$$

步骤 9) 由目标函数式(11)确定交通流量 VIMF 低频分量 $\{vimf_1, vimf_2, \dots, vimf_k\}$ 和 高频分量 $\{vimf_{k+1}, vimf_{k+2}, \dots, vimf_K\}$:

$$I(vimf_k, vimf_{k+1}) \downarrow \& I(vimf_k, vimf_{k+1}) \uparrow \\ k = \text{first}(\arg \min_{1 \leq k \leq K-1} [I(vimf_k, vimf_{k+1})]) \quad (11)$$

步骤 10) 保留分解后的交通流时间序列低频分量, 对高频分量由式(12)进行阈值降噪:

$$T = \sigma \sqrt{2 \ln N} \quad (12)$$

$$\hat{\sigma} = \frac{\text{median}(|vimf_k|)}{0.675} \quad (13)$$

式中: N 为交通流量时间序列长度; σ 为噪声分量标准差, 由 $\hat{\sigma}$ 近似计算, median 为取中位数操作。

各分量均不重构, 输出 K 个交通流量分量子序列 $\{vimf_1, vimf_2, \dots, vimf_K\}$ 。

2 基于图注意力网络的交通流量空间相关性模型

交通路网结构一般抽象为图结构数据类型, 呈现无

规则性,且各邻近节点对中心预测节点的关联程度并不相同,因此经典卷积神经网络(convolutional neural network, CNN)及相应的扩展 CNN 就难以建模并提取其中的不规则结构信息。

考虑到上述问题,本文将图注意力网络(graph attention network, GAT)^[13]模型引入到交通流量预测问题,将上一节分解后的交通流量节点的各分量序列作为图注意力网络的特征输入,利用图注意力网络可通过学习到的注意力系数对特征输入进行加权,从而对相关节点的各分量序列经不同程度的加权后输出,以捕捉交通路网邻近节点交通流量对中心预测节点交通流量的有区别影响,实现对交通路网进行空间特征提取。

使用图注意力机制后,最终的一阶邻域节点 j 对中心预测节点 i 的注意力系数 α_{ij} 定义为:

$$\alpha_{ij} = \frac{\exp(\sigma(\vec{a}^T [Wx_i \parallel Wx_j]))}{\sum_{m \in n_i} \exp(\sigma(\vec{a}^T [Wx_i \parallel Wx_m]))} \quad (14)$$

式中: \exp 为自然底数; \vec{a}^T 、 W 分别为可学习的权重向量和矩阵; \parallel 为拼接操作; $\sigma(\cdot)$ 为LeakyReLU激活函数。 $\sigma(\cdot)$ 方法如下:

$$\sigma(\cdot) = \max(w^T x, 0) = \begin{cases} w^T x, & w^T x > 0 \\ 0.01w^T x, & \text{其他} \end{cases} \quad (15)$$

经过图注意力网络有区别地聚合一阶邻近节点交通流量对中心预测节点交通流量的关联程度后的各分量时间序列输出为:

$$\begin{bmatrix} \alpha_{ii}^1 & \alpha_{ii}^2 & \cdots & \alpha_{ii}^K \\ \alpha_{ij_1}^1 & \alpha_{ij_1}^2 & \cdots & \alpha_{ij_1}^K \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{ij_m}^1 & \alpha_{ij_m}^2 & \cdots & \alpha_{ij_m}^K \end{bmatrix} \begin{bmatrix} vimf_i^1 & vimf_{j_1}^1 & \cdots & vimf_{j_m}^1 \\ vimf_i^2 & vimf_{j_1}^2 & \cdots & vimf_{j_m}^2 \\ \vdots & \vdots & \ddots & \vdots \\ vimf_i^K & vimf_{j_1}^K & \cdots & vimf_{j_m}^K \end{bmatrix} \quad (16)$$

式中: K 为前一节基于互信息熵改进的交通流量VMD分解算法输出的模态分量个数; α_{ii}^k 为中心预测节点 i 的第 k 个分量的自注意力系数; α_{ij}^k 为一阶邻域节点 j 的第 k 个分量对中心预测节点 i 的注意力系数; $\{vimf_i^1, vimf_i^2, \dots, vimf_i^K\}$ 和 $\{vimf_{j_1}^1, vimf_{j_1}^2, \dots, vimf_{j_1}^K\}$ 分别为由上一节分解得到的中心预测节点及其一阶邻域节点的分量序列。

基于图注意力网络的交通流量空间相关性模型如图1所示,假设图中 x_1 为中心预测节点,其周围的 $x_2 \sim x_9$ 节点均为 x_1 的邻域节点。在这些邻域节点中,与 x_1 直接相连的节点,如 x_2, x_4, x_6, x_7 和 x_8 ,称为 x_1 的相邻节点,也称为 x_1 的一阶邻域节点;而 x_3, x_5 和 x_9 节点不与 x_1 直接相连,但与 x_1 的一阶邻域节点直接相连,称为 x_1 的二阶邻域节点。通过一阶邻域节点 x_2, x_4, x_6, x_7, x_8 及中心预测节点 x_1 自身,计算其分别对 x_1 的注意力系数 α_{12} 、

α_{14} 、 α_{16} 、 α_{17} 、 α_{18} 及 α_{11} ,来捕捉一阶邻域节点交通流量对中心预测节点交通流量的有区别影响,而不考虑二阶邻域节点 x_3, x_5 和 x_9 的影响。

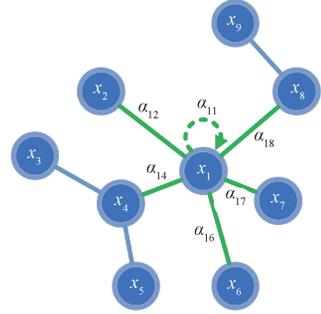


图1 交通流量空间相关性模型

Fig. 1 Spatial correlation model of traffic flow

3 基于门控循环单元网络的交通流量时间依赖性模型

3.1 门控循环单元网络原理

GRU^[14-15]是长短时记忆网络(long short term memory, LSTM)^[16]的改进模型,它将LSTM的内部状态向量和输出向量合并为隐藏状态向量,再结合重置门、更新门和候选隐藏向量,获取下一时刻的状态向量,模型复杂度更低,训练收敛速度更快。

经过前两节基于互信息熵改进的交通流量VMD分解模型对交通流量时间序列非平稳性的降低和阈值降噪,以及基于图注意力网络的空间相关性模型对邻近交通流量节点的有区别加权输出后,将各分量序列输入到基于GRU网络的交通流量时间依赖性模型,实现各模态分量的预测。交通流量时间依赖性模型如图2所示。

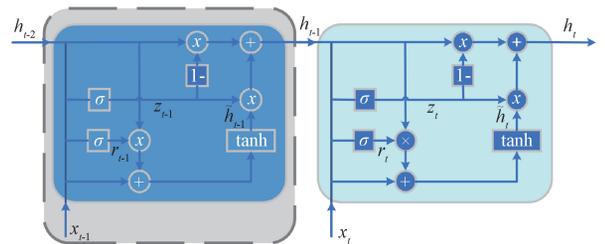


图2 交通流量时间依赖性模型

Fig. 2 Time dependence model of traffic flow

3.2 改进的RMSProp交通流量优化算法

RMSProp(root mean square propagation, RMSProp)算法^[17]使用小批量随机梯度按元素平方的指数加权移动平均来改变学习率,使得自变量每个数据元素的学习率

在迭代过程中重新调整,从而更快逼近最优解,有效解决了随机梯度下降(stochastic gradient descent, SGD)^[18]算法和动量法让其每个数据元素在相同时间步都使用同一学习率来自我迭代而可能导致的在后续时间步中迭代缓慢甚至发散的问题。而面对复杂且大量的交通流量多时间序列数据时,RMSProp 优化算法虽然可有效降低训练集的训练误差,算法的优化能力有所保证,但其收敛速度和效果依然会受到影响,并且在测试集中仍可能导致过拟合等问题,从而无法得到良好的泛化能力。

考虑到上述问题,本文尝试基于 RMSProp 算法,基本思想是保证其在一般非凸集的学习过程中有界性和收敛性^[19-20]的前提下,对其进行两点改进:一是对交通流量数据的小批量随机梯度 \mathbf{g}_t 增加一个参数惩罚项 $\lambda \mathbf{x}_{t-1}$,通过正则化策略减小测试集误差;二是对状态变量 \mathbf{s}_t 做偏差修正,以消除迭代初期各时间步小批量随机梯度权值之和较小而对后续迭代收敛速度的影响。以上两点改进,既可使得优化算法的泛化能力更好,不易过拟合,并且收敛速度更快,又有利于提升最终预测模型的预测精度。改进的 RMSProp 交通流量优化算法效果将在后续实验部分验证和分析。

算法 2:改进的 RMSProp 交通流量优化算法

步骤 1) 初始化时间步 $t = 0$, 学习率 η_0 , 状态变量 $\mathbf{s}_0 = \mathbf{0}$;

步骤 2) 随机初始化交通流量序列 \mathbf{x}_0 , 状态变量超参数 $\gamma \in (\gamma^*, 1)$, 惩罚参数 $\lambda \in (0, 1)$, 实常数 $\varepsilon \in (0, 1)$;

步骤 3) $t = t + 1$, 进入迭代循环步骤 4) ~ 8);

步骤 4) 由式(17)计算更新交通流量数据的小批量随机梯度 \mathbf{g}_t :

$$\mathbf{g}_t = \frac{1}{|\mathbf{b}|} \sum_{i \in \mathbf{b}_t} \nabla f_i(\mathbf{x}_{t-1}) + \lambda \mathbf{x}_{t-1} \quad (17)$$

式中: \mathbf{b}_t 为交通流量小批量样本,由训练集样本索引 i 组成; $|\mathbf{b}|$ 为交通流量小批量样本个数; $\nabla f_i(\mathbf{x}_{t-1})$ 为目标函数的随机梯度 $\nabla f(\mathbf{x}_{t-1})$ 的无偏估计; λ 越大,正则化惩罚越大。

步骤 5) 将步骤 4) 更新的小批量随机梯度 \mathbf{g}_t 逐元素平方,并由式(18)计算更新状态变量 \mathbf{s}_t :

$$\mathbf{s}_t = \gamma \mathbf{s}_{t-1} + (1 - \gamma) \mathbf{g}_t \odot \mathbf{g}_t \quad (18)$$

式中: \odot 为逐元素乘积运算符。

步骤 6: 由式(19)对状态变量 \mathbf{s}_t 进行偏差修正:

$$\hat{\mathbf{s}}_t = \frac{\mathbf{s}_t}{1 - \gamma_t} \quad (19)$$

步骤 7) 由式(20)调整目标函数 $f(\mathbf{x}_{t-1})$ 中每个交通流量数据元素的学习率:

$$\frac{\eta}{\sqrt{\hat{\mathbf{s}}_t + \varepsilon}} \quad (20)$$

式中:除法、开方和乘法均为逐元素运算; ε 为维持数值稳定性添加的实常数。

步骤 8) 由式(21)更新 t 时刻的交通流量序列 \mathbf{x}_t :

$$\mathbf{x}_t = \mathbf{x}_{t-1} - \frac{\eta}{\sqrt{\hat{\mathbf{s}}_t + \varepsilon}} \odot \mathbf{g}_t \quad (21)$$

循环迭代,直至最大迭代次数,并输出最优解。

4 基于改进 VMD-GAT-GRU 的交通流量组合预测模型

随着交通流量预测任务的日益复杂化,使得传统的单一预测模型表达能力有限。本文从交通流量时间序列的非平稳性、空间相关性及时间依赖性方面,设 \mathbf{x}_i 为中心预测节点交通流量输入, $\mathbf{x}_{j1} \sim \mathbf{x}_{jm}$ 为一阶邻域节点交通流量输入,经过改进 VMD 分解、空间相关性建模和时间依赖性建模等,得到最终交通流量预测输出 \hat{y}_{t+1} 。构建的交通流量组合预测模型结构如图 3 所示。

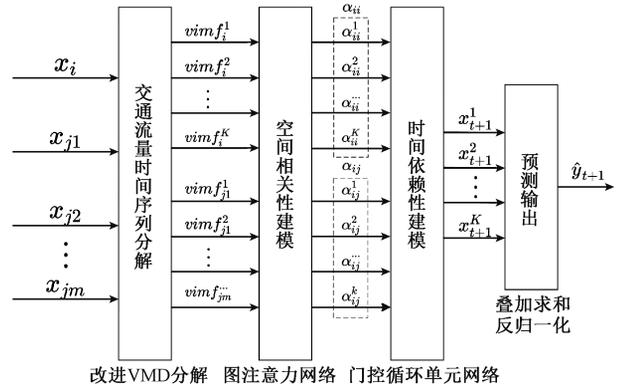


图 3 交通流量组合预测模型结构

Fig. 3 Structure diagram of the traffic flow combination prediction model

5 基于改进 VMD-GAT-GRU 的交通流量组合预测算法

交通流量时间序列具有非平稳性、空间相关性和时间依赖性等特点,使用单一的预测算法难以做到较好的预测性能。考虑到变分模态分解在非平稳数据处理方面的优势、图注意力网络对交通流量在不规则空间结构中的相关性的有效建模以及门控循环单元网络对交通流时间依赖性捕捉的优势,最终在保证收敛速度的情况下表现出良好的预测精度。本文的交通流量组合预测算法具体应用步骤如下:

步骤 1) 将交通流量预测过程形式化为用中心预测节点 i 及其 m 个一阶邻域节点 j 共 $m + 1$ 个节点的 T 个历史时间戳交通流量数据 $\mathbf{x}_{in} \in \mathbb{R}^{N \times T}: (\mathbf{x}_{i-T+1}, \mathbf{x}_{i-T+2}, \dots, \mathbf{x}_i)$ 作

为输入,通过组合预测模型学习到的一系列函数 $f(\cdot)$,以对未来 T' 步的交通流量进行预测:

$$[(\mathbf{x}_{t-T+1}, \mathbf{x}_{t-T+2}, \dots, \mathbf{x}_t)] \xrightarrow{f(\cdot)} (\mathbf{x}_{t+1}, \mathbf{x}_{t+2}, \dots, \mathbf{x}_{t+T'}) \quad (22)$$

步骤2)使用基于互信息熵改进的交通流量变分模态分解算法1将 i, j 节点的交通流量数据分解为一系列模态分量子序列 $\{vimf_i^1, vimf_i^2, \dots, vimf_i^K\}$ 、 $\{vimf_j^1, vimf_j^2, \dots, vimf_j^K\}$,实现平稳性处理。各交通流量模态分量子序列均不重构;

步骤3)将由步骤2)得到的共 $m+1$ 个节点的 $vimf$ 分量子序列,作为基于图注意力网络的交通流量空间相关性模型的输入,与由此计算得到的注意力系数 α_{ij}^k 为元素构成的注意力系数矩阵加权后得到序列输出,实现交通流量时间序列空间相关性建模;

步骤4)将步骤3)的经注意力系数矩阵加权后的共 $m+1$ 个节点的 $vimf$ 分量子序列,分别作为基于门控循环单元网络的交通流量时间依赖性模型的输入,由改进的RMSProp交通流量优化算法2迭代训练得到训练集的最优数值解,有效防止预测模型过拟合,得到各分量的预测输出 x_{t+1}^k ,实现交通流量时间序列时间依赖性建模;

步骤5)将经过步骤4)时间依赖性模型的分量预测输出进行叠加求和、反归一化处理,得到交通流量组合预测模型最终预测输出 \hat{y}_{t+1} :

$$\hat{y}_{t+1} = \frac{1}{(m+1) \cdot K} \sum_{k=1}^K x_{t+1}^k \quad (23)$$

改进的VMD-GAT-GRU时空融合交通流量组合预测算法流程如图4所示。

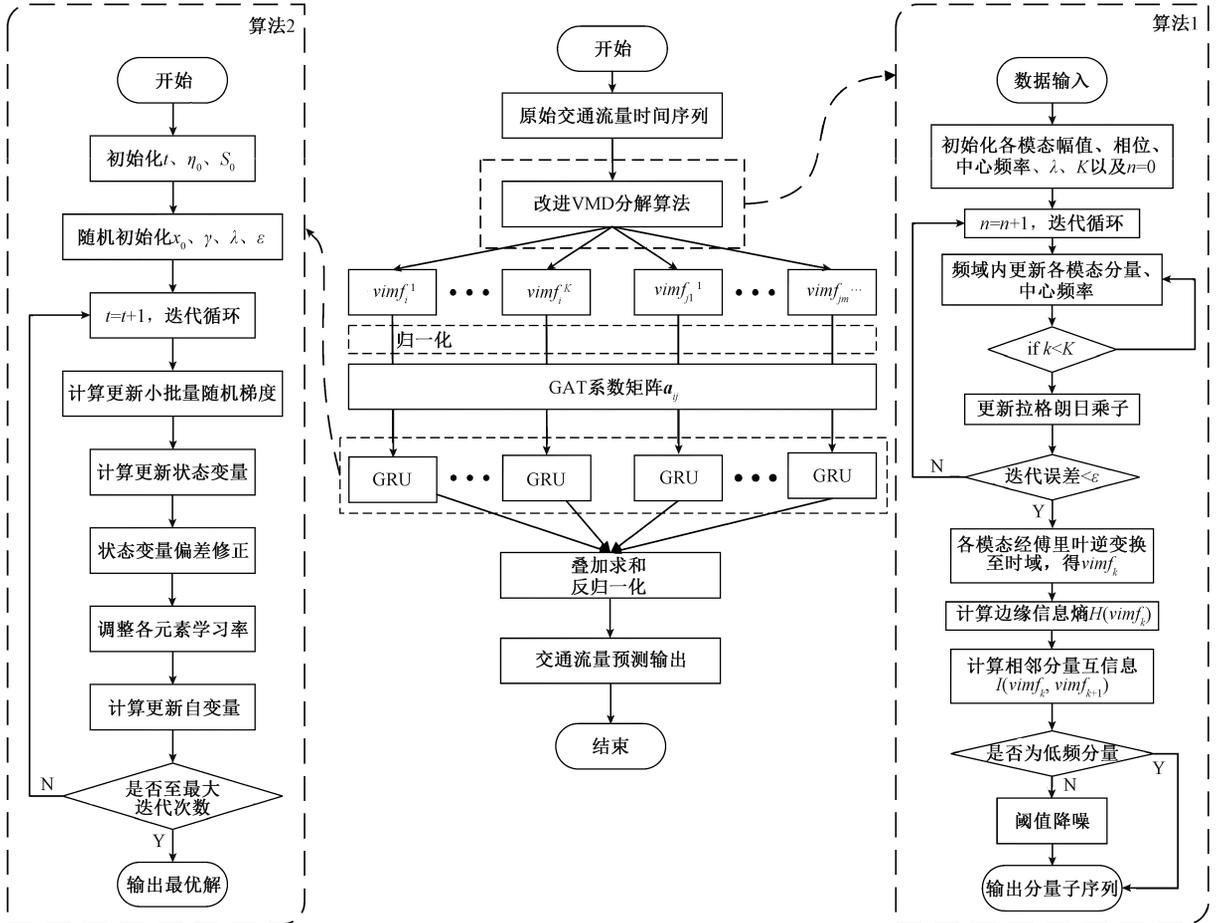


图4 交通流量组合预测算法流程

Fig. 4 Flowchart of traffic flow combination prediction algorithm

6 模型应用及实验分析

6.1 RTMC 交通数据集及预处理

本文实验数据集使用由美国明尼苏达州交通部门 (minnesota department of transportation, MnDOT) 下的分

区交通管理中心 (regional transportation management, RTMC) 收集到的交通数据。RTMC 通过安装在双子城高速公路网上的约 7 450 个感应环路探测器和 Wavetronix 雷达探测器, 收集到自 2000 年起包括每 30 s、每小时、每天的流量数据, 每 30 s 的速度以及占有率数据等。

本文选取的交通节点区域 I-35W&TH 62 位置如图 5

所示,包含 1 个中心预测节点和 4 个邻域节点。其中, S50 监测点为中心预测节点, S49、S329 及两个 Exit 监测点为一阶邻域节点。

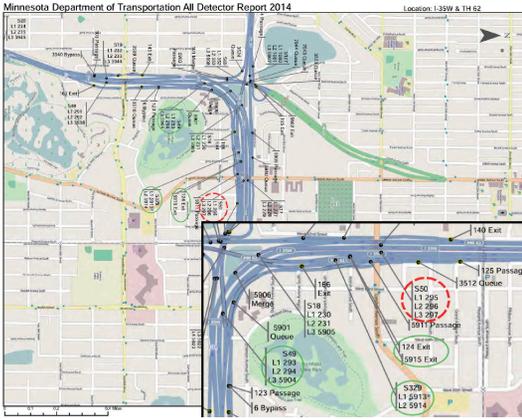


图 5 I-35W&TH 62 区域监测点分布图

Fig. 5 Distribution of monitoring points in I-35W&TH 62 area

采样时间为 2021 年 4 月 1 日~2021 年 5 月 30 日的 8:00~20:00 时段,交通流量时间序列采样间隔重新聚合为 5 min 每条数据,聚合后每个节点共 17 280 条数据。划分前 45 d 数据为训练集样本,后 15 d 数据为测试集样本。在训练集中使用宽度为 $T + 1$ 的滑动窗口来产生训练样本,其中 T 为历史时间戳数据长度,第 $T + 1$ 个时间戳数据为标签值,滑动窗口每次移动一个时间单位(5 min)。

由于交通流量数据在不同时间和节点上的幅值可能差异较大,因此在训练模型前将数据进行归一化操作,以使模型训练更加稳定有效。特征归一化计算方法如下:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (24)$$

通过式(24)将流量数据线性映射到目标范围 $[0, 1]$,使得交通流时间序列在同一幅值尺度。由于后续在预测阶段需要在训练过程中已归一化的数据重新返回到数据原始值,因此还需通过式(25)进行反归一化计算:

$$x = x' \times (\max(x) - \min(x)) + \min(x) \quad (25)$$

式中: x 为数据原始值, $\max(x)$ 、 $\min(x)$ 分别为原始数据的最大值和最小值, x' 为数据归一化后的值。

6.2 性能指标

为评估不同交通流量预测模型的预测性能,本文采用平均绝对误差(mean absolute error, MAE)、平均绝对百分比误差(mean absolute percentage error, MAPE)和均方根误差(root mean squared error, RMSE)作为性能指标。定义分别如下:

1) 平均绝对误差(mean absolute error, MAE), 绝对误差的平均值。

$$MAE = \frac{1}{N} \sum_{i=1}^N |x_i - \hat{x}_i| \quad (26)$$

2) 平均绝对百分比误差(mean absolute percentage error, MAPE), 预测误差占真实值的百分比。

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{x_i - \hat{x}_i}{x_i} \right| \times 100\% \quad (27)$$

3) 均方根误差(root mean squared error, RMSE), 预测值与真实值的均方根差。

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2} \quad (28)$$

式中: N 为交通流量测试集样本数量, x_i 和 \hat{x}_i 分别为测试集第 i 个样本的真实值和预测值。MAE、MAPE、RMSE 指标的数值越小,表示预测结果越好。

6.3 模型训练及实验结果分析

本文实验使用基于 Python 3.6 的 Anaconda 环境及 PyTorch 框架构建所需实验模型并进行训练和预测,并结合 MATLAB 实现可视化。主要超参数设置中,改进的交通流量 VMD 分解算法的惩罚因子 α , 根据文献[8]给出的范围($200 \leq \alpha \leq 2000$)及本文的多次实验,设置为 2000 时效果良好;噪声容限参数 τ , 可根据前期研究取经验值 0.05;产生训练集的历史时间戳步长 T , 根据数据集大小,取 $T = 7$ 时即可产生足量的训练样本;对于神经网络的隐藏层神经元个数,一般根据数据集大小在 $[16, 32, 64, 128, 256]$ 中选取,再根据多次实验的预测效果,将 GAT 网络的隐藏层神经元个数设置为 128, GRU 网络的隐藏层神经元个数设置为 64;网络中的批量处理 BatchSize, 考虑到数据集产生的训练样本数量可适配为参考值 128;改进 RMSProp 交通流量优化算法最大训练迭代次数 Epoch, 经多次训练实验,设置为 50 时即可收敛稳定且不会浪费计算机资源。利用改进的 VMD 交通流量分解算法将交通流量时间序列初步分解为 5 个不同频率和幅值的较平稳的分量子序列,由低频至高频的归一化的各模态分量子序列 VIMF 如图 6 所示。

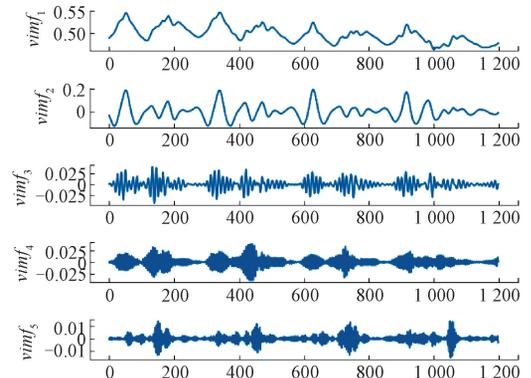


图 6 改进 VMD 分解的交通流量子序列

Fig. 6 Traffic flow subsequence of improved VMD decomposition

然后计算各分量序列间的互信息熵值,如表 1 所示,由表可知 $vimf_2$ 和 $vimf_3$ 分量间的互信息熵值 $I(vimf_2, vimf_3)$ 减小,而 $vimf_3$ 和 $vimf_4$ 分量的互信息熵值 $I(vimf_3, vimf_4)$ 开始增大,则出现极小值点,因此将 $vimf_3$ 分量作为低频分量和高频分量的分界点。保留低频分量 $vimf_1$ 和 $vimf_2$,并对高频分量 $vimf_3$ 至 $vimf_5$ 进行阈值降噪,最大程度保留原始交通流时间序列的信息。

表 1 各 $vimf$ 分量间互信息熵值
Table 1 Mutual information entropy between $vimf$ components

$vimf_{k,k+1}$	$vimf_{1,2}$	$vimf_{2,3}$	$vimf_{3,4}$	$vimf_{4,5}$
$I(vimf_{k,k+1})$	1.025	1.018	1.225	1.371

则最终的改进 VMD 部分分解效果如图 7 所示,其中实线部分为原始交通流量时间序列,点划线和虚线部分分别为 $vimf_1$ 和 $vimf_2$ 分量序列。由图 7 可知,基于互信息熵改进的交通流量 VMD 分解有效降低了原始交通流量序列的非平稳性,有利于后续模型迭代训练速度及预测精确度的提升。

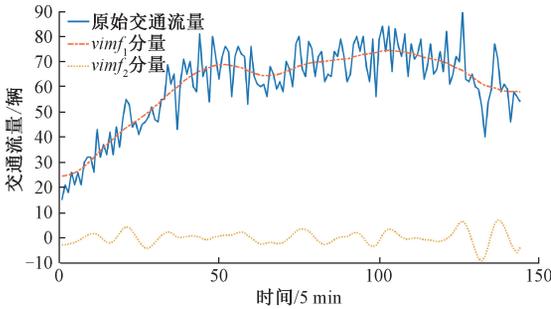


图 7 改进 VMD 部分分解效果图

Fig. 7 Improved VMD partial exploded rendering

由图 8 模型算法的训练迭代曲线可知,模型的训练损失随着迭代次数的增加不断减小,收敛性均可得到保证。其中,使用常规的 RMSProp 算法在约 10 次迭代搜索后达到全局最优解,但是在测试集中略微出现了过拟合现象;而使用改进的 RMSProp 算法在经过约 7 次迭代后即可搜索达到全局最优解,收敛速度和收敛效果相较于常规的 RMSProp 算法均有所提升,同时在测试集中表现良好,无过拟合现象。

通过时间依赖性模型得到中心预测节点经分解及阈值降噪后的各 VIMF 分量序列的各分量预测值;同时通过空间相关性模型得到中心预测节点及其一阶邻域节点经分解及阈值降噪后的各 VIMF 分量序列的注意力系数,与各分量加权后进一步训练得到重构的预测中心节点交通流量序列。各 VIMF 分量的预测结果如图 9 所示,其中实线部分为交通流量原始数值,点划线部分为交

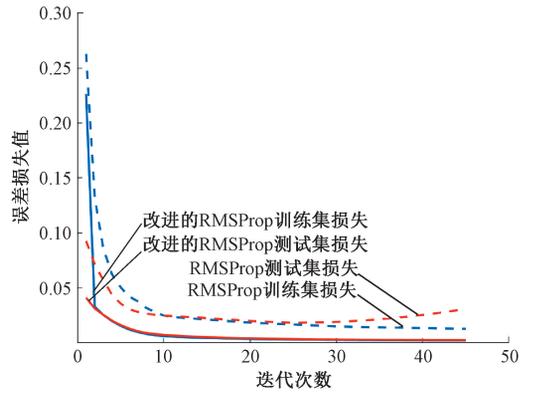


图 8 交通流量优化算法训练和测试 loss 曲线
Fig. 8 Traffic flow optimization algorithm trains and tests loss curves

通流量预测数值。

为综合验证本文所提出组合预测模型的性能,再分别使用原始交通流量数据训练集和测试集在 LSTM、GCN 和 GAT 这 3 种基准模型上进行实验。各模型的预测性能指标对比如表 2、图 10 所示。由表 2 可知,本文的组合预测模型相较于 LSTM、GCN 和 GAT,MAE 分别降低了 9.35、4.12、4.09,MAPE 分别降低了 16.42%、7.32%、8.1%,RMSE 分别降低了 9.97、4.6、4.81。

表 2 各交通流量预测模型性能指标对比
Table 2 Comparison of performance indicators of each traffic flow prediction model

模型	MAE	MAPE/%	RMSE
LSTM	11.72	20.45	12.84
GCN	6.49	11.35	7.47
GAT	6.46	12.13	7.68
Improved-VMD-GAT-GRU	2.47	4.03	2.87

结合性能指标对比图 10 及实验预测结果对比图 11 可以分析:

1) 未考虑交通流量时间序列非平稳性、各节点空间相关性的经典 LSTM 模型的 3 项性能指标相较于其他 3 种模型均表现一般;

2) 未考虑交通流量时间序列非平稳性的 GCN 和 GAT 模型的 3 项性能指标相近,相较于早些提出的 LSTM 网络均提升约 100%。结合分析 1) 表明,考虑到交通流量节点的空间相关性对提高模型预测精度的有效性;

3) 考虑到交通流量时间序列非平稳性、空间相关性和时间依赖性 3 方面的改进的 VMD-GAT-GRU 组合预测模型的 3 项性能指标较前 3 种基准模型有明显提升。

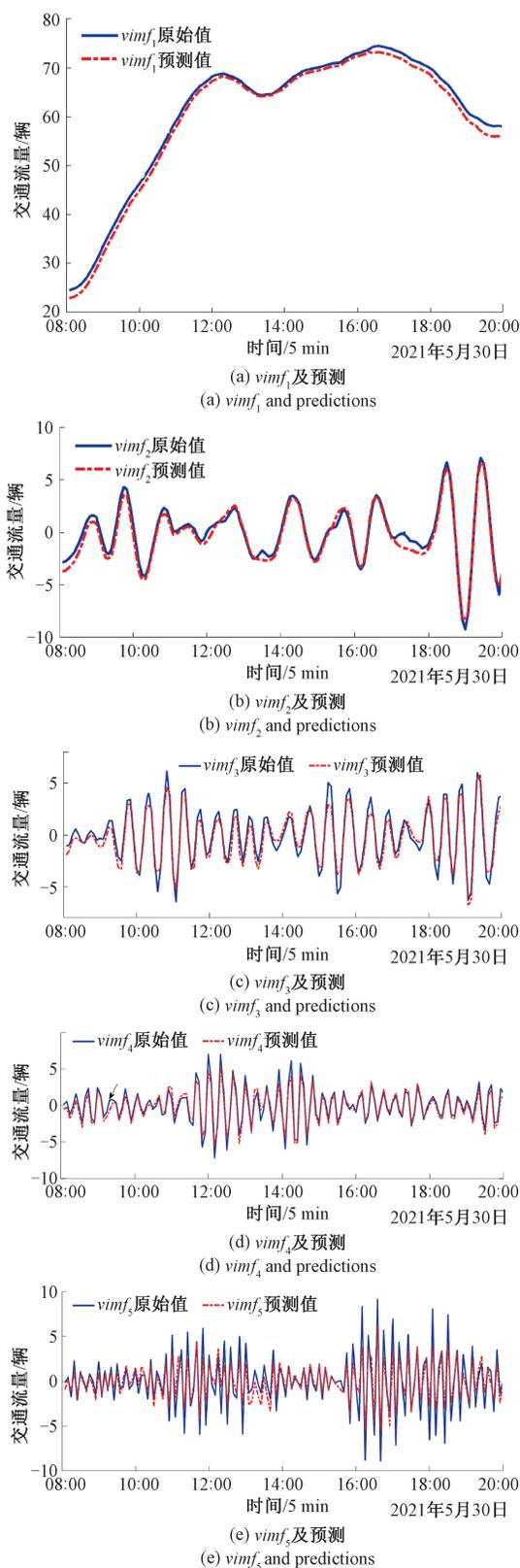


图 9 各 VIMF 分量及预测结果

Fig. 9 VIMF components and prediction results

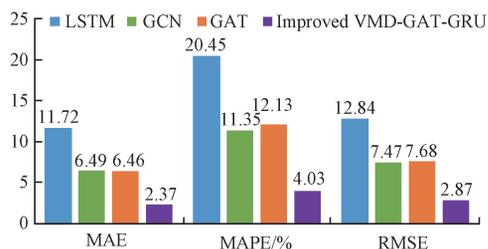


图 10 各交通流量预测模型性能指标对比

Fig. 10 Comparison of performance indicators of each traffic flow prediction model

各模型的预测效果对比如图 11 所示,实线部分为真实交通流量时间序列。本文所选的 8:00~20:00 预测时段的交通流量均为不平稳状态,例如在 10:00~12:00 时段,交通流量由最低值约 40 辆/5 min,波动至峰值约 80 辆/5 min,对交通流量未经平稳性处理的 LSTM、GCN 和 GAT 模型,均不能稳定跟踪预测交通流量的剧烈波动,而本文提出的组合预测模型依然表现出良好的预测效果。

7 结 论

本文基于交通流量时间序列的复杂非平稳性、空间相关性和时间依赖性的特点,从提高交通流量预测模型的预测精度和收敛速度角度,提出了改进的 VMD-GAT-GRU 交通流量组合预测模型以用于复杂的交通流量预测。实验结果表明,基于互信息熵改进的交通流量变分模态分解算法有效降低了交通流量时间序列的非平稳性,提高了模型预测精度,且捕捉到时间序列的空间相关性以及时间依赖性特点的组合模型相较于已有的基准模型 LSTM、GCN 和 GAT,预测精度均有一定程度提升,同时基于改进的 RMSProp 交通流量优化算法在测试集所表现出的泛化能力良好,收敛效果和收敛速度均得到有效提升。但本文尚未考虑到交通系统及道路中的动态突发因素,如车道数量、交通事故以及极端天气等对交通流量预测精度的影响,因而此类问题将是下一步将要考虑研究的内容。

参考文献

[1] 新华社. 中华人民共和国经济和社会发展第十四个五年规划和 2035 年远景目标纲要[R]. 北京, 2021. Xinhua News Agency. Outline of the 14th five-year plan for economic and social development of the People's Republic of China and the vision for 2035 [R]. Beijing, 2021.

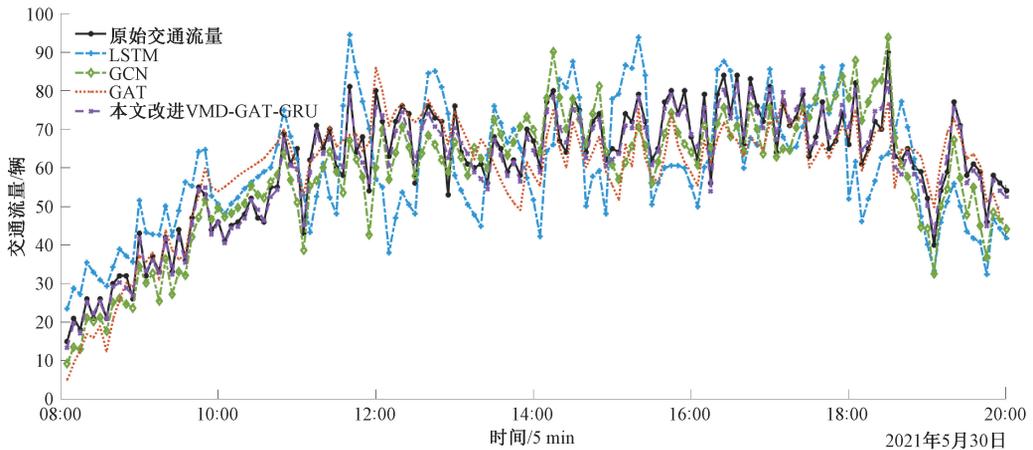


图 11 各交通流量模型预测效果对比

Fig. 11 Comparison of prediction effects of traffic flow models

- [2] 张晓晗, 冯爱民. 基于经验模态分解和长短期记忆神经网络的短期交通流量预测[J]. 计算机应用, 2021, 41 (1): 225-230.
ZHANG X H, FENG AI M. Short-term traffic flow prediction based on empirical mode decomposition and long short-term memory neural network [J]. Journal of Computer Applications, 2021, 41 (1): 225-230.
- [3] 殷礼胜, 高贺, 魏帅康, 等. 基于多维时空的 NPCA-PSR-IGM(1,1)组合模型的短时交通流预测[J]. 电子与信息学报, 2021, 43 (4): 1035-1041.
YIN L SH, GAO H, WEI SH K, et al. Short-term traffic flow prediction based on NPCA-PSR-IGM (1, 1) combined model of multi-dimensional space-time [J]. Journal of Electronics & Information Technology, 2021, 43 (4): 1035-1041.
- [4] GUO SH N, LIN Y F, LI SH J, et al. Deep spatial-temporal 3D convolutional neural networks for traffic data forecasting [J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20 (10): 3913-3926.
- [5] BARNUM H, BARRETT J, CLARK L O, et al. Entropy and information causality in general probabilistic theories [J]. New Journal of Physics, 2010, 12: 129401.
- [6] DRAGOMIRETSKIY K, ZOSSO D. Variational mode decomposition [J]. IEEE Transactions on Signal Processing, 2014, 62 (3): 531-544.
- [7] 吕宏政, 陈仁文, 张祥, 等. 基于 VMD 交叉样本熵的旋翼桨叶故障诊断方法[J]. 电子测量技术, 2019, 42 (9): 107-111.
LYU H ZH, CHEN R W, ZHANG X, et al. Faultdiagnosis method of rotor blade based on VMD and cross-sample entropy [J]. Electronic Measurement Technology, 2019, 42 (9): 107-111.
- [8] 赵昕海, 张术臣, 李志深, 等. 基于 VMD 的故障特征信号提取方法[J]. 振动、测试与诊断, 2018, 38 (1): 11-19, 202.
ZHAO X H, ZHANG SH CH, LI ZH SH, et al. Fault feature signal extraction method based on VMD [J]. Journal of Vibration, Measurement & Diagnosis, 2018, 38 (1): 11-19, 202.
- [9] 李帅永, 韩明秀, 文井辉. 基于 VMD-SVD 自优化的管道微泄漏信号增强方法[J]. 电子测量与仪器学报, 2021, 35 (12): 68-78.
LI SH Y, HAN M X, WEN J H. Weak signal enhancement based on self-optimizing VMD-SVD for leak location in water-supply pipeline [J]. Journal of Electronic Measurement and Instrumentation, 2021, 35 (12): 68-78.
- [10] 李帅永, 毛维培, 程振华, 等. 基于 VMD 和 K-SVD 字典学习的供水管道泄漏振动信号压缩感知方法[J]. 仪器仪表学报, 2020, 41 (3): 49-60.
LI SH Y, MAO W P, CHENG ZH H, et al. Compressed sensing method for leakage vibration signal in water-supply pipelines based on VMD and K-SVD dictionary learning [J]. Chinese Journal of Scientific Instrument, 2020, 41 (3): 49-60.
- [11] 梁喆, 彭苏萍, 郑晶. 基于 EMD 和互信息熵的微震信号自适应去噪[J]. 计算机工程与应用, 2014, 50 (4): 7-11, 32.
LIANG ZH, PENG S P, ZHENG J. Self-adaptive denoising for microseismic signal based on EMD and mutual information entropy [J]. Computer Engineering and Applications, 2014, 50 (4): 7-11, 32.
- [12] SHI G, QIN CH J, TAO J F, et al. A VMD-EWT-LSTM- based multi-step prediction approach for shield

- tunneling machine cutterhead torque [J]. Knowledge-Based Systems, 2021, 228: 107213.
- [13] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks [C]. International Conference on Learning Representations, ICLR 2018, Vancouver, Canada, 2018.
- [14] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. ArXiv Preprint, 2014, arXiv:1412.3555.
- [15] SHAHID F, ZAMEER A, MUNEEB M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM [J]. Chaos, Solitons, and Fractals, 2020, 140: 110212.
- [16] GREFF K, SRIVASTAVA R K, KOUTNÍK J, et al. LSTM: A search space odyssey [J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28: 2222-2232.
- [17] POSTALCIOGLU S. Performance analysis of different optimizers for deep learning-based image recognition [J]. International Journal of Pattern Recognition and Artificial Intelligence, 2020, 34:2051003:1-2051003:12.
- [18] ZOU D F, YUAN C, ZHOU D R, et al. Gradient descent optimizes over-parameterized deep ReLU networks[J]. Machine Learning, 2019, 109: 467-492.
- [19] XU D P, ZHANG SH D, ZHANG H SH, et al. Convergence of the RMSProp deep learning method with penalty for nonconvex optimization [J]. Neural Networks: The Official Journal of the International Neural Network Society, 2021, 139: 17-23.

- [20] SHI N CH, LI D W, HONG M Y, et al. RMSprop converges with proper hyper-parameter [C]. International Conference on Learning Representations, ICLR 2021, Vienna, Austria, 2021.

作者简介



殷礼胜, 1994 年于安徽大学获得学士学位, 2004 年于广西大学获得硕士学位, 2007 年于重庆大学获得博士学位, 现为合肥工业大学副教授, 主要研究方向为通信信道建模、现代智能算法理论及其应用、复杂系统建模、交通流量预测等。

E-mail: yls20000@163.com

Yin Lisheng received his B. Sc. degree from Anhui University in 1994, M. Sc. degree from Guangxi University in 2004, and Ph. D. degree from Chongqing University in 2007, respectively. Now he is an associate professor in Hefei University of Technology. His main research interests include communication channel modeling, modern intelligent algorithm theory and its application, complex system modeling, traffic flow prediction, etc.



吴洋洋 (通信作者), 2019 年于合肥学院获得学士学位, 现为合肥工业大学硕士研究生, 主要研究方向为复杂系统建模、交通流量预测。

E-mail: angkory@163.com

Wu Yangyang (Corresponding author) received his B. Sc. degree from Hefei University in 2019. Now he is a M. Sc. candidate in Hefei University of Technology. His main research interests include complex system modeling and traffic flow prediction.