doi: 10.7690/bgzdh.2026.10.007

基于混合深度学习的短时交通流预测方法

魏晗颖，赵梓涵，迟宗正
（大连理工大学开发区校区软件学院，辽宁 大连 116620）

摘要：针对目前短时交通流存在考虑特征不充分导致预测精度低的问题，提出一种基于混合深度学习的短时交通流预测方法。基于道路图模型，对交通流预测中空间、时间和语义相关性进行分析。模型包含输入层、卷积层、递归层和输出层的交通流预测，并基于时间、周期性和事件特征提取历史交流数据时空、语义关系。通过消融实验验证所提方法有效。与 SVM、CNN、LSTM、CNN-LSTM、T-GCN 等模型相比，所提模型综合性能最优，预测误差降低 5%左右。结果表明，该模型能为智能交通发展提供一定借鉴，具有广泛的应用前景。

关键词：智能交通；交通流预测；深度学习；卷积神经网络；图神经网络
中图分类号：TP183　　**文献标志码：**A

Short-term Traffic Flow Forecasting Method Based on Hybrid Deep Learning

Wei Hanying, Zhao Zihan, Chi Zongzheng
(*School of Software*, *Development Zone Campus*, *Dalian University of Technology*, *Dalian* 116620, *China*)

Abstract: In order to solve the problem of low prediction accuracy caused by insufficient consideration of the characteristics of short-term traffic flow, a short-term traffic flow prediction method based on hybrid deep learning is proposed. Based on the road graph model, the spatial, temporal and semantic correlations in traffic flow forecasting are analyzed. The model consists of input layer, convolution layer, recursive layer and output layer, and extracts the spatio-temporal and semantic relationship of historical communication data based on time, periodicity and event characteristics. The effectiveness of the proposed method is verified by ablation experiments. Compared with SVM, CNN, LSTM, CNN-LSTM and T-GCN, the proposed model has the best comprehensive performance, and the prediction error is reduced by about 5%. The results show that the model can provide a reference for the development of intelligent transportation and has a wide application prospect.

Keywords: intelligent transportation; traffic flow prediction; deep learning; convolutional neural network; graph neural network

0　引言

交通流预测[1-2]是智能交通系统[3]中的一个关键问题，可用来分析城市道路网络上的交通状况（如速度、流量和密度），挖掘交通模式，从而提高交通运行效率和人们出行的安全性。

由于复杂的空间、时间和语义相关性，交通流预测是一项具有挑战性的任务。文献[4]提出一种融合多元时序和模式挖掘（multi-variate time series and pattern mining，MTSPM）的终端区交通流预测模型。该模型考虑了交通、天气等特征，并利用深度学习模型学习特征间相关性。文献[5]提出了一种基于多图融合的交通流预测模型。该模型利用数据驱动的时间图生成方法用于捕获节点的时间模式，并基于图卷积网络（graph convolutional network，GCN）提取时空特征。文献[6]提出了一种基于注意力机制的图模型预测交通流。该模型采用注意力机制替代传统的局部卷积核结构，可以有效提取长距离时空依赖关系。上述方法基于传统机器学习或深度学习模型有效预测了交通流，然而预测性能仍有巨大提升空间。如部分方法建立了基于道路网络空间建模的 GCN 或其变体网络[5-9]。这些方法只考虑道路之间的拓扑关系来构建图，而忽略了所有其他可以测量道路之间相关性的语义因素（例如，历史规律、局部功能和道路类型等）。

为进一步提升短时交通流预测性能，该研究提出了一种基于混合深度学习的短时交通流预测方法。该方法充分考虑了道路空间、时间和语义相关性与周期和事件等全局特征，并基于多图卷积网络实现道路网络中的交通流预测。

1　问题陈述

交通流预测方法中涉及的相关定义及概念。

定义 1（道路图）参考图论中相关概念，将道路

收稿日期：2025-02-04；**修回日期：**2025-03-02
基金项目：2024 年辽宁省大学生创新创业训练计划项目（20241014110271）
第一作者：魏晗颖（2002—），女，福建人。
通信作者：迟宗正（1984—），男，山东人，博士。

图定义为一个加权图 G：

$$G=(V, E)。\tag{1}$$

式中：$V=(v_1, v_2, \cdots, v_i, \cdots)$为道路图 G 中的所有节点(道路)构成的集合，$v_i \in V$ 为第 i 个节点(道路)；E 为道路图 G 中边构成的集合，如 $e_{ij} \in E$ 表示节点(道路)v_i 和 v_j 之间的相关性。需注意，e_{ij} 越大表示 v_i 和 v_j 越相关。不失一般性，道路图中节点均代指道路，为简化描述，之后均以节点表示。

该研究中构建的道路图包括道路网络拓扑图 G_r、权重拓扑图 G_w、交通模式图 G_p 和交通功能图 G_f 4 个子图。通过设置多个子图多维度编码节点之间的相关性，从而有助于模型学习交通流数据包含的特征，提升预测性能。一般来说，笔者将较大的权重分配给来自不同方面的相关性较强的道路之间的边。

定义 2(道路网络拓扑图)道路网络的拓扑图可以表示为有向图 G_r：

$$G_r=(V, E, W_r)。\tag{2}$$

式中 W_r 为道路图中节点之间权重构成的集合，如 $w_r(i,j) \in W_r$ 表示从节点 v_i 到 v_j 最小链路的倒数。拓扑图中，越相邻的节点将有更高的权重；因此，由节点之间权重构成的相邻矩阵 $\boldsymbol{X}_r$：

$$\boldsymbol{X}_r=\begin{pmatrix} 0 & w_r(1,2) & \cdots & w_r(1,N) \\ w_r(2,1) & 0 & \cdots & w_r(2,N) \\ \vdots & \vdots & & \vdots \\ w_r(N,1) & w_r(N,2) & \cdots & 0 \end{pmatrix}。\tag{3}$$

定义 3(加权拓扑图)拓扑图 G_r 只考虑从节点 v_i 到 v_j 的中间链路的数量。这些链路的长度也可能影响 v_i 到 v_j 的相关性。为此，引入加权拓扑图 G_w 表征链路的长度对节点相关性的影响。

$$G_w=(V, E, W_w)。\tag{4}$$

式中 W_w 为加权拓扑图 G_w 中的权重集合。对于任意的边 $e_{ij} \in E$，其在加权拓扑图 G_w 中的权重 $w_w(i,j) \in W_w$ 可计算为：

$$w_w(i,j)=M_e\{L(v_k) \mid v_k \in V\} / (\min L(v_k))。\tag{5}$$

式中：$M_e\{L(v_k) \mid v_k \in V\}$ 为路径图中所有节点路径长度的中位数；$M_e\{\bullet\}$ 为中位数函数；$L\{\bullet\}$ 为路径长度计算函数。权拓扑图 G_w 的相邻矩阵 $\boldsymbol{X}_w$：

$$\boldsymbol{X}_w=\begin{pmatrix} 0 & w_w(1,2) & \cdots & w_w(1,N) \\ w_w(2,1) & 0 & \cdots & w_w(2,N) \\ \vdots & \vdots & & \vdots \\ w_w(N,1) & w_w(N,2) & \cdots & 0 \end{pmatrix}。\tag{6}$$

定义 4(交通模式图)交通模式相似的道路在地理空间上可能不一定很近。利用每条道路的历史交通流数据，引入交通模式图 G_p，通过测量每条道路对的历史交通条件模式的相似性判断道路间的相关性。

$$G_p=(V, E, W_p)。\tag{7}$$

式中 W_p 为交通模式图 G_p 中的权重集合，对于任意权重 $w_p(i,j) \in W_p$ 用来度量节点 v_i 和节点 v_j 的历史交通模式。对于任意节点 $v_i \in V$，令 v_i 的交通模式定义为该道路的平均历史每周交通流状况 s_{v_i}，$s_{v_i}[t]$ 定义为 v_i 的历史交通状况数据上的一周的第 t 个时隙的平均交通状况。接着，对于任意 2 个节点 v_i 和 v_j，使用动态时间扭曲[10](dynamic time warping，DTW)计算 s_{v_i} 和 s_{v_j} 之间的距离。最后，将距离测量转换为相似性测量，该测量即为交通模式图 G_p 中的权重 $w_p(i,j) \in W_p$：

$$w_p(i,j)=e^{-\alpha \mathrm{DTW}(s_{v_i}, s_{v_j})}。\tag{8}$$

式中：$\mathrm{DTW}(s_{v_i}, s_{v_j})$ 为 s_{v_i} 和 s_{v_j} 之间的 DTW 距离；α 为控制距离的衰减率参数，该衰减率应根据相关交通条件的范围进行设置。同理，交通模式图 G_p 中的相邻矩阵 $\boldsymbol{X}_p$ 可定义如下：

$$\boldsymbol{X}_p=\begin{pmatrix} 0 & w_p(1,2) & \cdots & w_p(1,N) \\ w_p(2,1) & 0 & \cdots & w_p(2,N) \\ \vdots & \vdots & & \vdots \\ w_p(N,1) & w_p(N,2) & \cdots & 0 \end{pmatrix}。\tag{9}$$

定义 5(交通功能图)具有相似功能的城市地区的道路通常具有相似的交通模式，例如，工业区可能在工作日的高峰时段交通流量很大，而市中心地区可能在周末交通流量很大。为此，引入交通功能图 G_f 描述具有相似功能的道路的交通模式。

$$G_f=(V, E, W_f)。\tag{10}$$

式中 W_f 为加权拓扑图 G_f 中的权重集合，对于任意权重 $w_f(i,j) \in W_f$ 用来度量节点 v_i 的局部区域和节点 v_j 的局部区域的功能的相似性。采用兴趣点[11](point of interest，POI)分布衡量城市区域的功能。对于任意节点 $v_i \in V$，从 v_i 周围的 POI 中可计算住宅、工作(如公司、办公楼等)、商业(如商场、商店等)、餐厅、学校、交通(如火车站、地铁站等)、娱乐(如剧院、酒吧等)和风景(如公园、湖泊等)8 类 POI 的密度，形成特征向量 $\boldsymbol{p}_{v_i}$，$p_{v_i}[j]$ 表示节点 v_i 周围的 POI 类别 j 的密度，则有：

$$p_{v_i}[j]=\frac{m_j^i}{m^i}\log\frac{M}{M_j}。\tag{11}$$

式中：m_j^i 为节点 v_i 周围的类别 j 的 POI 的数量；m^i 为 v_i 周围的 POI 总数；M_j 为 POI 数据集中的类别 j 中的 POI 数量；M 为 POI 信息集中的 POI 总数量。同时，给定任意 2 个节点 v_i 和 v_j，使用余弦相似度测量 $\boldsymbol{p}_{v_i}$ 和 p_{v_j} 之间的相似度，该相似度可视为权重 $w_f(i,j)\in W_f$。接着，加权拓扑图 G_f 的相邻矩阵 $\boldsymbol{X}_f$ 为：

$$\boldsymbol{X}_f=\begin{pmatrix}0 & w_f(1,2) & \cdots & w_f(1,N)\\ w_f(2,1) & 0 & \cdots & w_f(2,N)\\ \vdots & \vdots & & \vdots\\ w_f(N,1) & w_f(N,2) & \cdots & 0\end{pmatrix}。\tag{12}$$

定义 6（交通流样本）令时间 t 时道路网上的交通流信号（例如速度和体积）表示为 N 维向量 $\boldsymbol{x}_t$，其中 N 为道路数量。交通流样本 $\boldsymbol{s}_t$ 可定义为：

$$\boldsymbol{s}_t=(\boldsymbol{s}_{\text{int}},\boldsymbol{s}_{\text{out}})。\tag{13}$$

式中：$\boldsymbol{s}_{\text{int}}$ 为输入，且有 $\boldsymbol{s}_{\text{int}}=[\boldsymbol{x}_{t-W-1},\cdots,\boldsymbol{x}_{t-1},\boldsymbol{x}_t]$ 为 W 个历史交通流序列数据；$\boldsymbol{s}_{\text{out}}$ 为输出，且有 $\boldsymbol{s}_{\text{out}}=x_{t+h}$ 为时间 $t+h$ 时的交通流。

基于上述定义，交通流预测问题可认为是在已建立的道路图 G 中，从大量的交通流样本中学习规律，以获得将当前时间 t 的 W 个历史交通流数据并映射到时间 $t+h$ 的未来交通流的函数：

$$\tilde{\boldsymbol{x}}_{t+h}=f\left(G_r,G_w,G_p,G_f;\left(\boldsymbol{x}_{t-W-1},\cdots,\boldsymbol{x}_{t-1},\boldsymbol{x}_t\right)\right)。\tag{14}$$

式中：G_r、G_w、G_p、G_f 为道路图 G 中道路网络拓扑图、权重拓扑图、交通模式图和交通功能图；$f(\bullet)$ 为交通流预测函数。

2　交通流预测网络

基于深度学习的交通流预测网络结构如图 1 所示。该网络由输入层、卷积层、递归层和输出层 4 层组成：首先，给定交通流样本 s_t，输入层使用滑动窗口处理历史输入交通流序列数据 s_{int} 以获得分段序列数据；其次，对于每个分段，卷积层使用 4 个单独的 GCN 分别对道路图 G 中道路网络拓扑图 G_r、权重拓扑图 G_w、交通模式图 G_p 和交通功能图 G_f 执行卷积运算，并将结果融合以获得融合特征矩阵；接着，递归层应用改进门限循环单元-循环神经网络（gated recurrent unit-recurrent neural network，GRU-RNN）处理从分段序列数据中获得的融合特征矩阵序列；最后，输出层将特征矩阵与全局特征融合，通过全连接层得到最终预测结果。

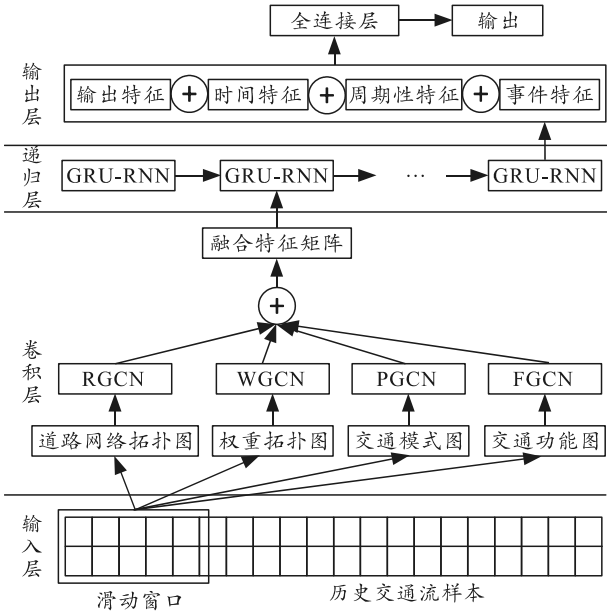


图 1　基于深度学习的交通流预测网络结构

2.1　输入层

对于交通流样本 $\boldsymbol{s}_t$，输入交通流序列数据 $\boldsymbol{s}_{\text{int}}$ 为 $N\times W$ 维矩阵，其中 N 为道路的数量，W 为历史交通流序列的数量。所提交通流预测网络中输入层使用窗口大小为 w、步长为 d 的滑动窗口处理 $\boldsymbol{s}_{\text{int}}$。当处理完成后，可获得一个具有 K 个片段的序列 $\left(\boldsymbol{s}_t^1,\boldsymbol{s}_t^2,\cdots,\boldsymbol{s}_t^K\right)$，同时每个片段都是一个 $N\times w$ 的矩阵。通过滑动窗口将输入交通流序列数据 $\boldsymbol{s}_{\text{int}}$ 分割成多个片段可有效增加样本多样性。

2.2　卷积层

卷积运算能够很好地捕捉局部相关性并保持平移不变性，是捕捉空间相关性的有效方法。然而，标准 CNN 只能在欧几里得空间（例如，图像、具有规则网格的空间等）中使用，因此传统卷积操作不能适应道路网络的复杂拓扑结构。由于道路之间的相关性是以图的形式定义的，为此该研究应用 GCN 对道路网络进行卷积运算，以捕捉道路网络中各节点之间的相互作用。

卷积层中，道路网络拓扑图 G_r、权重拓扑图 G_w、交通模式图 G_p 和交通功能图 G_f 都被输入到一个单独的 GCN 中。GCN 的传播规则为：

$$\left.\begin{aligned}&H^{(l+1)}=\sigma\left(\tilde{\boldsymbol{D}}^{-1/2}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{D}}^{-1/2}\boldsymbol{H}^{(l)}\boldsymbol{W}^{(l)}\right)\\&\tilde{\boldsymbol{X}}=\boldsymbol{X}+\boldsymbol{I}_N\\&\tilde{\boldsymbol{D}}[i,j]=\sum\nolimits_j\tilde{\boldsymbol{X}}[i,j]\end{aligned}\right\}。\tag{15}$$

式中：$\tilde{\boldsymbol{X}}$ 为道路图中包含自连接的广义相邻矩阵；

$\boldsymbol{I}_N$ 为单位矩阵；$\boldsymbol{X}$ 为道路图中相邻矩阵，可以为道路网络拓扑图中相邻矩阵 $\boldsymbol{X}_r$、权重拓扑图中相邻矩阵 $\boldsymbol{X}_w$、交通模式图中相邻矩阵 $\boldsymbol{X}_p$ 或交通功能图中相邻矩阵 $\boldsymbol{X}_f$；$\tilde{\boldsymbol{D}}$ 为对角矩阵；$\boldsymbol{W}^{(l)}$为第 l 层的可训练权重矩阵；$\boldsymbol{H}^{(l)}$为第 l 层的输入特征矩阵；GCN 的原始输入特征矩阵为 $\boldsymbol{H}^{(0)}$，其维度为 $N\times w$ 的分段矩阵；$\sigma(\bullet)$为 ReLU 激活函数。GCN 中传播规则可以认为是傅里叶域中的频谱滤波器。每个片段将输入到 4 个 GCN（即道路网络拓扑图卷积网络 GCN_r、权重拓扑图卷积网络 GCN_w、交通模式图卷积网络 GCN_p 和交通功能图卷积网络 GCN_f，以及相应的道路图（即 G_r、G_w、G_p 和 G_f），并产生 4 个特征矩阵，表示为道路网络拓扑图特征矩阵 $\boldsymbol{H}_r$、权重拓扑图特征矩阵 $\boldsymbol{H}_w$、交通模式图特征矩阵 $\boldsymbol{H}_p$ 和交通功能图特征矩阵 $\boldsymbol{H}_f$。

接下来，基于参数矩阵的融合方法将特征矩阵 $\boldsymbol{H}_r$、$\boldsymbol{H}_w$、$\boldsymbol{H}_p$ 和 $\boldsymbol{H}_f$ 融合为矩阵 $\boldsymbol{Y}$：

$$\left.\begin{aligned}&\boldsymbol{Y}=\boldsymbol{W}_r'\bullet\boldsymbol{H}_r+\boldsymbol{W}_w'\bullet\boldsymbol{H}_w+\boldsymbol{W}_p'\bullet\boldsymbol{H}_p+\boldsymbol{W}_f'\bullet\boldsymbol{H}_f\\&(\boldsymbol{W}_r',\boldsymbol{W}_w',\boldsymbol{W}_p',\boldsymbol{W}_f')=\varphi(\boldsymbol{W}_r,\boldsymbol{W}_w,\boldsymbol{W}_p,\boldsymbol{W}_f)\end{aligned}\right\}。\tag{16}$$

式中：•为逐元素乘积符号；$\boldsymbol{W}_r$、$\boldsymbol{W}_w$、$\boldsymbol{W}_p$ 和 $\boldsymbol{W}_f$ 分别为调整特征矩阵 $\boldsymbol{H}_r$、$\boldsymbol{H}_w$、$\boldsymbol{H}_p$ 和 $\boldsymbol{H}_f$ 的权重可学习参数；$\varphi(\bullet)$为 Softmax 函数；$\boldsymbol{W}_r'$、$\boldsymbol{W}_w'$、$\boldsymbol{W}_p'$ 和 $\boldsymbol{W}_f'$ 为经 Softmax 函数归一化后的权重。

2.3　递归层

GRU 是循环神经网络（recurrent neural network，RNN）的一个变体，其中前一个单元的输出是当前单元输入的一部分，如图 2 所示。

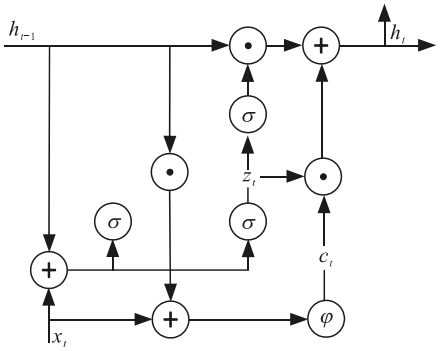


图 2　经典 GRU 模型架构

这种机制允许信息逐步传递，因此，RNN 和 GRU 能够捕获时间相关性。随着模型深度加深，传统 RNN 和 GRU 很容易产生梯度消失或爆炸、训练时间过长的问题[12]。为改善上述问题，笔者提出了一种改进的 GRU-RNN 网络，用于高效学习交通流序列数据中的时间相关性。

图 2 为经典 GRU 模型架构，其由复位门和更新门组成。复位门 $\boldsymbol{r}_t$ 用于决定是否考虑全部或部分前一隐藏状态特征 $\boldsymbol{h}_{t-1}$。更新门 $\boldsymbol{z}_t$ 用于根据候选层 $\boldsymbol{c}_t$ 的状态决定更新多少当前隐藏状态特征 $\boldsymbol{h}_t$。对于时间序列样本集 $\boldsymbol{x}_t$，GRU 的计算为：

$$\left.\begin{aligned}&\boldsymbol{z}_t=\sigma\left(\boldsymbol{W}_{xz}\boldsymbol{x}_t+\boldsymbol{U}_{hz}\boldsymbol{h}_{t-1}+\boldsymbol{b}_z\right)\\&\boldsymbol{r}_t=\sigma\left(\boldsymbol{W}_{xr}\boldsymbol{x}_t+\boldsymbol{U}_{hr}\boldsymbol{h}_{t-1}+\boldsymbol{b}_r\right)\\&\boldsymbol{c}_t=\varphi\left(\boldsymbol{W}_{xc}\boldsymbol{x}_t+\boldsymbol{u}_{hc}(\boldsymbol{r}_t\odot\boldsymbol{h}_{t-1})+\boldsymbol{b}_c\right)\\&\boldsymbol{h}_t=(1-\boldsymbol{z}_t)\odot\boldsymbol{h}_{t-1}+\boldsymbol{z}_t\odot\boldsymbol{c}_t\end{aligned}\right\}。\tag{17}$$

式中：$\boldsymbol{W}_{xz}$、$\boldsymbol{W}_{xr}$、$\boldsymbol{W}_{xc}$ 分别为连接输入层和更新门、复位门和候选层的权重；$\boldsymbol{U}_{hz}$、$\boldsymbol{U}_{hr}$ 和 $\boldsymbol{u}_{hc}$ 为当前时间 t 和前一时间 t-1 之间的自连接权重；$\boldsymbol{b}_z$、$\boldsymbol{b}_r$ 和 $\boldsymbol{b}_c$ 分别为 GRU 单元中更新门、复位门和候选层的偏差；$\sigma(\bullet)$为激活函数；$\varphi(\bullet)$为候补激活函数。

与传统 GRU 不同，改进 GRU-RNN 中每个门只使用先前的隐藏状态和偏置计算特征，这大大减少了训练参数的总数，如图 3 所示。改进 GRU-RNN 的门变量可表示为：

$$\left.\begin{aligned}&\boldsymbol{z}_t=\sigma(\boldsymbol{W}_{hz}\boldsymbol{h}_{t-1}+\boldsymbol{b}_z)\\&\boldsymbol{r}_t=\sigma(\boldsymbol{W}_{hr}\boldsymbol{h}_{t-1}+\boldsymbol{b}_r)\end{aligned}\right\}。\tag{18}$$

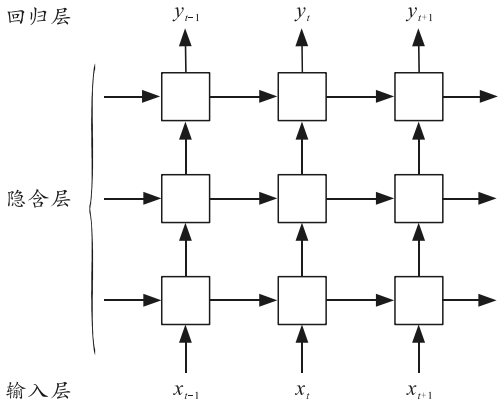


图 3　改进 GRU-RNN 结构

在实际应用中，多参数时间序列信号往往表现出高度的非线性和非平稳特性。为了进行准确可靠的映射，将上述 GRU-RNN 层叠加以构建一个叠加 GRU-RNN。通过在顶部引入回归层，叠加 GRU-RNN 可提升预测性能。

$$\hat{\boldsymbol{y}}_t=\sigma(\boldsymbol{W}_{yh}\boldsymbol{h}_t+\boldsymbol{b}_y)。\tag{19}$$

式中：$\hat{\boldsymbol{y}}_t$ 为回归层之后的预测输出；$\boldsymbol{W}_{yh}$ 为回归层的权重；b_y 为回归层的偏差。

改进 GRU-RNN 使用具有动量的随机梯度下降法（stochastic gradient descent，SGD），通过最小化实际输出和预测输出之间的 MSE 调整权重和偏差：

$$\left.\begin{aligned} E &= \sum_{t=1}^{T}(\boldsymbol{y}_t-\hat{\boldsymbol{y}}_t)^2/2T \\ \boldsymbol{\theta}_{q+1} &= \boldsymbol{\theta}_q-\eta\bullet(\partial E/\partial\boldsymbol{\theta}_q)+\beta\bullet(\boldsymbol{\theta}_q-\boldsymbol{\theta}_{q-1}) \end{aligned}\right\}。\quad (20)$$

式中：E 为 MSE 函数；$\boldsymbol{\theta}_q$ 为第 q 次迭代时的参数集，该参数集由不同权重和偏差组成；$\boldsymbol{y}_t$ 为实际输出；$\hat{\boldsymbol{y}}_t$ 为预测输出；T 为训练样本的总数；η 为学习率。传统训练过程中，学习率和动量在训练过程中都保持不变，这不足以有效地调整 GRU-RNN 的所有权重和偏差。尽管通过实时调整学习率在一定程度上提高了训练性能，但仍然可能导致收敛缓慢和过拟合。

为了克服传统训练策略的局限性，进一步提高训练性能，笔者开发了一种新的改进自适应学习训练算法，该算法融合自适应梯度(adaptive gradient，AdaGrad)和可调动量，从而在训练过程中实现不同梯度下的自适应学习率和动量。自适应学习原理为：

$$\left.\begin{aligned} \boldsymbol{\theta}_{q+1} &= \boldsymbol{\theta}_q-\eta_q\bullet(\partial E_q/\partial\boldsymbol{\theta}_q)+\beta_q\bullet(\boldsymbol{\theta}_q-\boldsymbol{\theta}_{q-1}) \\ \eta_q &= \eta_0\Big/\sqrt{\sum_{i=1}^{q}g_i^2+\lambda} \\ g_q &= \partial E_q/\partial\boldsymbol{\theta}_q \\ \beta_q &= e^{-\kappa-\|g_q\|} \end{aligned}\right\}。\quad (21)$$

式中：E_q 为第 q 次迭代时的 MSE 值；η_q 为第 q 次迭代时的学习率；g_q 为第 q 次迭代时的梯度；β_q 为第 q 次迭代时的动量；λ 为固定参数，设置为 1×10^{-8}；κ为可调节动量项的初始系数；η_0 为初始学习率。

通过以上分析，给定划分为 K 个交通流片段 $(\boldsymbol{s}_t^1,\boldsymbol{s}_t^2,\cdots,\boldsymbol{s}_t^K)$ 的交通条件样本 $\boldsymbol{s}_t$，通过卷积层产生一系列特征矩阵 $(\boldsymbol{y}_t^1,\boldsymbol{y}_t^2,\cdots,\boldsymbol{y}_t^K)$，将这些特征矩阵按时间顺序输入到递归层中。最后将一个隐藏状态 $\boldsymbol{h}_t^{z\kappa}$ 的输出特征矩阵作为递归层的输出 $\boldsymbol{R}_t^z$：

$$\boldsymbol{R}_t^z=\boldsymbol{h}_t^{z\kappa}。\quad (22)$$

2.4　输出层

递归层输出特征 $\boldsymbol{R}_t^z$ 的编码信息包含了道路网络拓扑结构的空间相关性、道路间语义相关性和紧密度的时间相关性。还有其他全局特征会影响交通状况(例如，时间和事件)；因此，在输出最终特征矩阵之前，还提取了时间、周期性、事件等全局特征。

定义 7(时间特征)给定交通流样本 $\boldsymbol{s}_t$，将会影响城市交通流量一天中的时间、一周中的某一天和节假日事件(即假日或正常日)视为时间特征，并提取未来时间 $t+h$ 的时间特征以预测交通状况。令交通流样本 $\boldsymbol{s}_t$ 中时间特征矩阵表示为 $\boldsymbol{T}_z^t$。

定义 8(周期性特征)城市交通流具有较强的周期性(例如，工作日期间高峰时段的交通状况通常具有相似性)。考虑到输入层中滑动窗口 W 不能设置太长，递归层只能捕获短期周期性。这是因为过长的滑动窗口使得模型难以收敛，且过长的序列数据的梯度消失也削弱了周期性的影响。为此，给定交通流样本 $\boldsymbol{s}_t$，将最后一天 $t-n^d$ 和最后一周 $t-n^w$ 的同一时间的交通状况视为长期周期性特征，其中 n^d 和 n^w 分别为一天和一周的持续时间。此外，还考虑了以 $t-n^d$ 和 $t-n^w$ 为中心的交通流信号小窗口，从而避免因交通状况通常不是严格的周期性而带来的干扰问题。例如，尽管几乎所有的工作日都有早高峰，但交通流量高峰可能出现在每天早上的不同时段。同时，令交通流样本 $\boldsymbol{s}_t$ 中周期性特征矩阵表示为 $\boldsymbol{P}_z^t$，其为 $N\times(2\times w^p)$维矩阵。其中，w^p 为小窗口中的交通流序列的数量。

定义 9(事件特征)在某些情况下，事件(如交通事故和节假日)会导致某些道路区域周围的交通模式在短时间内急剧变化。这种情况的出现非常罕见，仅考虑以有限地交通流数据几乎不可能捕捉到规律性。为了解决这个问题，利用一些事件发布网站(例如，基于位置的社交网络和公共设施的官方网站)中共享的具有的位置和时间属性的文本信息来识别事件，并基于深度神经网络捕捉其对未来交通模式的影响。具体而言，给定交通流样本 $\boldsymbol{s}_t$，通过聚合在每条节点 v_i 附近的输入序列 $\boldsymbol{s}_{\text{int}}$ 的周期内发布的文本片段，以形成先验知识 $D(t,\ i)$，然后使用 TextCNN[13]通过预训练的单词嵌入层、卷积层、最大时间池化层和全连接层从 $D(t,\ i)$中提取特征向量。之后，将所有提取的特征向量组合起来，形成道路网络的事件特征矩阵，表示为 $\boldsymbol{E}_z^t$。

进一步，将融合输出特征 $\boldsymbol{R}_t^z$、时间特征矩阵 $\boldsymbol{T}_z^t$、周期性特征矩阵 $\boldsymbol{P}_z^t$ 和事件特征矩阵 $\boldsymbol{E}_z^t$ 级联形成融合特征矩阵 $\boldsymbol{Z}_t$：

$$\boldsymbol{Z}_t=\boldsymbol{R}_t^z\oplus\boldsymbol{T}_z^t\oplus\boldsymbol{P}_z^t\oplus\boldsymbol{E}_z^t。\quad (23)$$

式中⊕为级联操作符。

接着，在 $\boldsymbol{Z}_t$ 上堆叠 2 个完全连接层，其中第一层可以视为嵌入层，第二层用于确保输出矩阵 $\boldsymbol{Z}_t$ 具有与 $\boldsymbol{X}_t$ 具有相同的形状。请注意，全局特征是可选的，可以根据数据可用性选择其中的任何一个形

成最终的 $\boldsymbol{Z}_t$。最后，使用 Sigmoid 作为 $\boldsymbol{Z}_t$ 上的激活函数输出时间 $t+h$ 处的预测交通流状况。需注意，模型训练的目标函数为最小化预测交通流状况和真实交通流状况之间的均方根误差(root mean square error，RMSE)。

3　实验与分析

3.1　数据集和实验设置

为验证所提模型的有效性，基于真实世界的交通流量数据集进行测试。该数据集由某市综合交通研究中心采集。采集时使用环形探测器对某市主城区约 600 条道路进行采样，采样时间段为 2018 年 1 月 1 日至 2018 年 3 月 31 日，采样频率为 15 min/次。环形探测器主要记录道路上车辆速度，共包含约 8 640 条记录。同时，基于百度地图 API 收集该市 POI。考虑到百度地图 API 不允许一次查询一个区域内的所有 POI，只接受一个关键词和一个中心位置作为输入，并返回中心位置周围满足该关键词的预定义数量的 POI。通过提供不同的 POI 类别作为关键字多次查询中心位置周围的 POI，然后将中心位置移动到下一个附近位置并重复查询，最终生成道路网络 POI 集合。

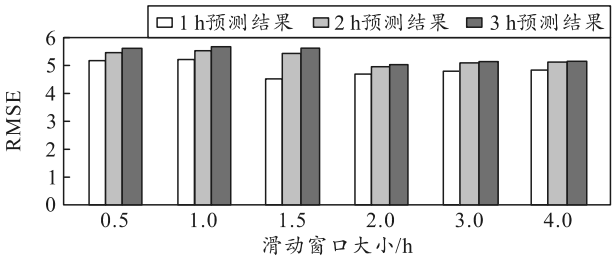
实验时，使用线性插值方法填充交通速度数据的缺失值。同时，在输入到模型中之前，使用最小-最大归一化方法将交通速度值缩放到[0, 1]的范围内。训练完成后，将预测的交通速度值重新缩放到正常值，以与地面实值进行比较。

实验时参数：交通流样本的输入序列长度设置为 96，DTW 距离的衰减率设置为 0.1，投影 POI 的正交距离设置为 1 km，周期性特征的小窗口中的交通流分割数量设置为 4。事件特征提取网络 TextCNN 中单词嵌入的维度设置为 100，输出的特征向量的维度设置为 50。交通流网络中初始学习率设置为 0.001，将批量大小设置为 32，将训练时间设置为 2 000，可调节动量项的初始系数设置为 0.1。

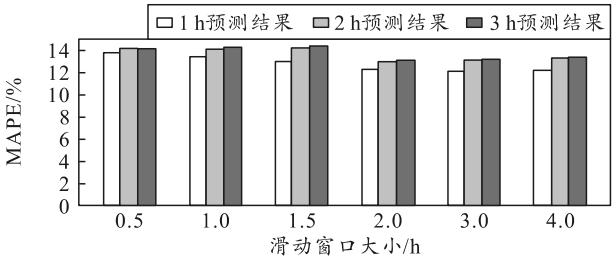
此外，实验基于 RMSE 和平均绝对百分比误差(mean absolute percent error，MAPE)验证了预测方法的性能。

3.2　消融实验

实验研究了滑动窗口取值对交通流预测性能的影响，将滑动窗口取值从 0.5 h 逐步增加到 4 h，不同滑动窗口对预测性能性能结果如图 4 所示。



(a) RMSE 指标结果

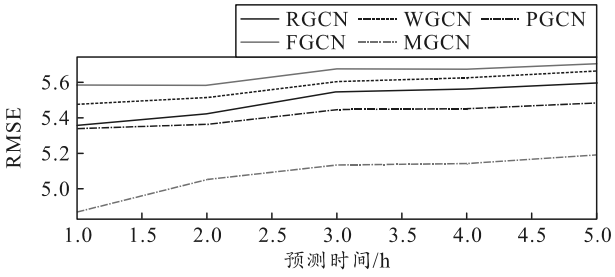


(b) MAPE 指标结果

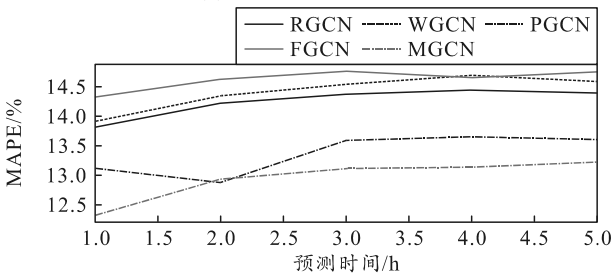
图 4　不同滑动窗口对预测性能性能结果

其中横轴为滑动窗口取值，纵轴分别为 RMSE 和 MAPE，实验时分别对比了预测 1、2 和 3 h 的交通流情况。可以看出：随着滑动窗口增加时，RMSE 和 MAPE 均先呈下降趋势，然后逐渐增加。实验结果与理论分析一致，即滑动窗口取值不能太大也不能太小，需要满足精度与训练时间之间的平衡。最终，取滑动窗口的值为 2 h。

对 4 个单独的网络拓扑图卷积神经网络(RGCN)、权重拓扑图卷积神经网络(WGCN)、交通模式图卷积神经网络(PGCN)、交通功能图卷积神经网络(FGCN)以及所提融合上述 4 个 GCN 的混合图卷积神经网络(MGCN)进行消融实验，结果如图 5 所示。



(a) RMSE 对比结果

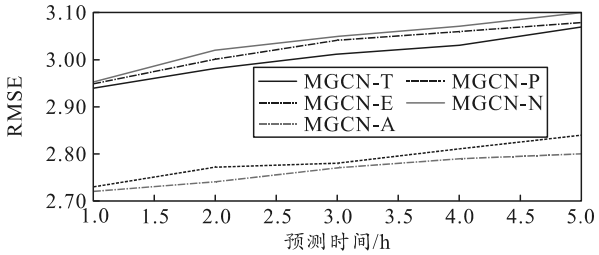


(b) MAPE 对比结果

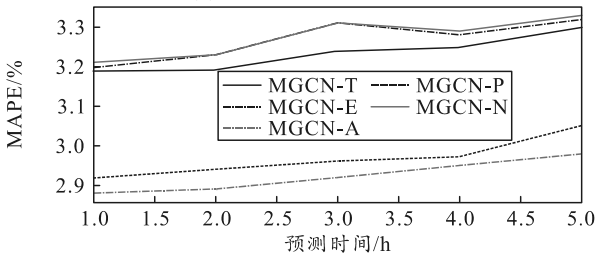
图 5　不同 GCN 下消融实验结果

可以看出：除了预测 2 h 的 MAPE 指标外，所提混合图卷积神经网络(MGCN)整体性能最优，表明所提模型可有效预测短时交通流。此外，PGCN 性能仅次于所提混合图卷积神经网络，表明历史交通模式对未来交通流的预测具有较强的影响。实验结果验证了所提模型的有效性及较高的预测性能。

对全局特征，即时间特征、周期性特征和事件特征进行消融实验。实验时输入为 MGCN，输入特征主要有单独考虑时间特征(MGCN−T)、周期性特征(MGCN−P)、事件特征(MGCN−E)、所有特征(MGCN−A)以及不考虑任何特征(MGCN−N)等几种情况。全局特征消融实验对比结果如图 6 所示。



(a) RMSE 对比结果



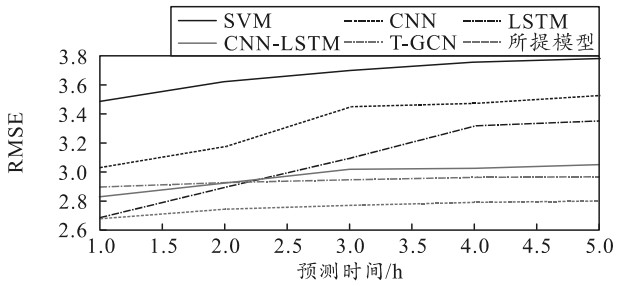
(b) MAPE 对比结果

图 6　全局特征消融实验对比结果

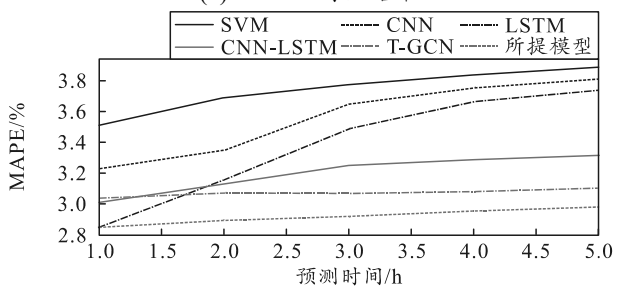
可以看出：所提 MGCN 在所有特征下的整体性能最好，不考虑任何特征时的整体性能最差。实验结果验证了全局特征的有效性。表明融合多个特征可有效提升模型性能，增强模型学习能力。单独考虑周期性特征优于其他单独特征。这表明，在所有全局特征中，周期性特征是最有效的。实验结果符合实际，表明城市交通流通常具有很强的周期性。此外，单独考虑事件特征仅优于不考虑任何特征，与其他单独特征比无明显优势。这表明事件特征的偶然性对未来交通流的影响非常有限。

3.3　交叉对比实验

为了验证所提方法综合性能，笔者对比了支持向量机(support vector machine，SVM)、卷积神经网络(convolutional neural networks，CNN)、长短期记忆(long short−term memory，LSTM)、卷积神经网络−长短期记忆(convolutional neural networks−long short term memory，CNN−LSTM)、时间图卷积网络(time graph convolutional network，T−GCN)与所提模型性能。不同模型综合性能对比结果如图 7 所示。



(a) RMSE 对比结果



(b) MAPE 对比结果

图 7　不同模型综合性能对比结果

首先，基于深度学习的方法(如 CNN、LSTM、CNN−LSTM、T−GCN 和所提模型)比传统方法(SVM)性能更优。分析原因：与传统机器学习方法相比，基于深度学习的方法可以更好地捕捉非线性时空相关性；其次，基于递归深度学习的方法(如 LSTM、CNN−LSTM、T−GCN 和所提模型)优于 CNN。分析原因：交通流具有很强的时间相关性，CNN 无法有效捕捉数据间时间相关性，因此性能较低、最终，所提方法综合性能最优，将预测误差降低了 5%左右。分析原因：所提模型融合多种道路图，并综合考虑了各种相关性(即空间、时间和语义相关性)和全局特征(即时间、周期和事件特征)，因此可有效学习各种类型道路的交通流，从而获得较优的预测结果。

4　结论

笔者提出了一种混合深度学习的短时交通流预测方法。该方法通过融合道路网络拓扑图、权重拓扑图、交通模式图和交通功能图神经网络提取道路数据特征，通过叠加递归神经网络学习数据特征之间的相关性，并基于时间、周期性、事件等全局特征对交通流进行预测。该模型可为短时交通流预测技术的发展提供一定借鉴作用。

(下转第 96 页)