

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2022.0230

# 基于图对比学习的稳健交通流量预测

刘伟\*, 贾素玲

(北京航空航天大学 经济管理学院, 北京 100191)

**摘要:** 作为智能交通系统的核心技术, 稳健的交通流量预测是一个长期存在但具有挑战的任务。不论是工业界还是学术界, 现有模型需要大量训练数据、易受数据噪声影响而表现不稳健, 是限制该领域发展与实际应用落地的重要因素。而在学术界, 图对比学习可以通过数据增强与对比损失来降低数据需求量, 同时提升模型抵抗数据噪声的能力。提出一种交通流量图对比学习 (TFGCL) 框架, 用于稳健的交通流量预测。TFGCL 框架有 3 个创新点: 针对交通流量图 (TFG) 数据的独特时空特性, TFGCL 框架从时间和空间 2 个角度出发, 提出 3 种 TFG 数据增强方法。针对 TFG 数据中语义相似的假负样本, 提出一个过滤策略使 TFGCL 框架能够免受其干扰, 从而学习到高质量的表征。TFGCL 框架通过联合交通流量预测任务和图对比学习任务进行同时训练。在 3 个真实交通数据集上与 8 个基线模型进行对比实验, 结果表明: TFGCL 框架的预测性能更为稳健, 较最优基线模型最高提升 6.24%, TFGCL 框架的稳健性尤其体现在数据缺失较为明显的数据集和长时交通流量预测任务中。

**关键词:** 交通流量预测; 交通流量图; 数据噪声; 图对比学习; 预测稳健性

**中图分类号:** TP399

**文献标志码:** A

**文章编号:** 1001-5965(2024)01-0122-12

交通流量预测是智能交通系统 (intelligent transportation system, ITS) 的一个基本问题, 对交通管理和公共安全具有重要意义。一个稳健的交通流量预测模型对于现实世界中的很多 ITS 应用都非常有益。例如, 准确的车流量预测可以支撑交通管理部门更好地理解和管理交通拥堵, 准确的人流量预测可以帮助大型活动组织者维持人群安全。

交通流量预测旨在基于历史交通流量观测数据, 设计预测模型, 对未来的交通流量趋势进行准确的预测。这是一个典型的时空数据预测问题, 挑战主要来自于时间和空间 2 个方面: ①时间, 某个道路历史不同时刻的流量会对该道路未来不同时刻的流量产生影响; ②空间, 道路之间通过路网连接, 不同道路的交通状况会产生相互影响。因此, 有效捕捉时空相关性对解决交通流量预测问题至关重要。

得益于深度学习技术的发展, 现在的交通流量预测模型可以充分挖掘交通流量数据的依赖关系, 包括利用循环神经网络 (recurrent neural networks, RNN)<sup>[1-2]</sup> 或时域卷积网络 (temporal convolutional networks, TCN)<sup>[3]</sup> 捕捉时间依赖, 利用卷积神经网络 (convolutional neural networks, CNN)<sup>[4]</sup> 或图神经网络 (graph neural networks, GNN)<sup>[5-6]</sup> 捕捉空间依赖, 从而预测未来的交通流量。然而, 已有的交通流量预测模型通常需要大规模的数据支撑才能取得较好的模型效果, 并且模型效果很容易受到数据噪声的干扰<sup>[7]</sup>, 也就是说, 基于深度学习的交通流量预测模型的稳健性依赖于数据质量。与此同时, 交通预测领域的公开数据集规模通常较小, 例如只有几个月, 这大大限制了最终可用于训练的样本量; 另一方面, 交通流量传感器的读数通常存在误差, 甚至有时会因为信号中断产生数据丢失。因此, 现

收稿日期: 2022-04-07; 录用日期: 2022-05-14; 网络出版时间: 2022-05-31 14:22

网络出版地址: link.cnki.net/urlid/11.2625.V.20220530.2025.004

\* 通信作者. E-mail: wayne@buaa.edu.cn

**引用格式:** 刘伟, 贾素玲. 基于图对比学习的稳健交通流量预测 [J]. 北京航空航天大学学报, 2024, 50 (1): 122-133.

LIU W, JIA S L. Robust traffic flow prediction based on graph contrastive learning [J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50 (1): 122-133 (in Chinese).

实中的交通数据质量较低。为了缓解交通流量预测中数据质量与模型稳健性之间的矛盾,一个直观的想法是对交通流量数据进行数据增强,一方面可以扩展训练样本量,让模型获得足够的训练数据,另一方面可以通过引入对比学习损失,让模型能够学到数据中的扰动不变性,从而产生噪声不敏感的、高质量的交通流量数据表征,用于稳健的交通流量预测。

目前交通对比学习方面的研究较少,由于交通流量数据通常定义在交通路网上,而交通路网可以抽象为图结构(graph structure),因此,本文可以借鉴图对比学习<sup>[8]</sup>方法进行交通对比学习。现有的图对比学习方法主要针对静态图数据(即不随时间变化的图数据),其中不乏一些在无监督任务上取得了卓越性能的研究<sup>[8-10]</sup>。这些方法的共同理念是对于每个输入图,采用随机数据增强方法生成2个相似视图,然后输入到图编码器中得到2个视图的高级图表征,最后在表征空间中,最大限度地拉近具有类似语义信息的图表征(正样本对),同时最小化具有不相关语义信息的图表征(负样本对),从而让模型学习到数据中的扰动不变性。通常情况下,对于一个输入图(也称为锚点),正样本对就是通过图数据增强生成的该锚点的2个视图,负样本对则由该锚点的视图与其他图的视图构成。通过这种数据增强与对比学习的方式,可以得到更加稳健的图表征,用于多种下游任务,例如节点分类,链接预测等。

然而,现有的图对比学习方法无法直接应用于随时间和空间变化的交通流量数据中,主要有以下几点原因:

1) 2个阶段的训练适用性差。图对比学习通常包含2个阶段,首先利用对比损失预训练一个表征编码器,然后利用编码器输出的表征做输入训练线性回归模型做预测,或者在预训练的编码器后加一个未训练的解码器,最后微调整个模型进行预测。然而,文献[8]表明,这种方法比较适用于图分类或图节点分类任务,在预测任务上表现并不好。

2) 数据增强未充分研究。数据增强的好坏直接决定了对比学习是否有效<sup>[11]</sup>,好的数据增强可以帮助模型学习到稳健的数据表征,排除数据噪声的影响。然而,专门针对随时间和空间变化的交通流量数据增强技术还未被充分研究。

3) 语义相似样本被忽略。在图对比学习方法中,通常把同一批次不同样本的数据增强结果看作是负样本对。这忽略了交通流量数据中的语义相似性,例如交通流量的邻近性和周期性<sup>[12]</sup>。而具有

语义相似的交通流量数据样本不应该被当作负样本对,在表征空间中推远他们之间的距离。

为了解决训练适用性数据噪声语义相似样本等挑战,本文提出一种交通流量图对比学习(traffic flow graph contrastive learning, TFGCL)框架。TFGCL框架针对交通流量数据独特的时空特性,对现有的图对比学习范式进行了三大调整:①本文将原始的交通流量预测损失和对比学习损失相耦合,以联合训练的方式来提高模型的性能;②本文从空间域和时间域2个角度对交通流量数据进行扰动,达到数据增强的目的,从而构建样本对用于图对比学习;③本文设计了一个策略用于过滤假负样本对(语义相似度高的负样本对),避免TFGCL框架被假负样本误导。

## 1 相关工作

历经多年不断的研究和实践,交通流量预测任务已经取得了重大的进展和丰富的成果,主要分为2个方面:基于物理模型的方法和数据驱动的方法。基于物理模型的方法通常基于某一理论假设(例如运动波理论<sup>[13]</sup>、车辆跟驰理论<sup>[14]</sup>、排队论<sup>[15]</sup>),将交通系统表示为耦合的微分方程组,然后通过微分方程驱动的交通系统仿真实现交通流量预测。这种方法保证了仿真结果可以表示全域上的交通动态变化,而不仅仅是基于历史观测数据的拟合。但是,这样的模型通常对交通系统做了很强的假设,导致模型可能无法捕捉到现实世界中复杂的人类行为及其他不确定因素<sup>[16]</sup>。

数据驱动方法在早期主要利用时间序列模型进行交通流量预测,包括差分整合移动平均自回归(autoregressive integrated moving average, ARIMA)模型和向量自回归(vector auto-regression, VAR)<sup>[17]</sup>模型。这些方法通常基于平稳性假设,因此,很难捕捉到动态、复杂的交通流量数据中的非线性相关性。近些年来,得益于深度学习强大的表征能力,基于深度学习的交通预测广受关注,并且已经形成了相对标准的模式。对于时间相关性的建模,常用的方法有RNN、一维CNN(1D-CNN)及TCN。由于RNN在序列数据建模方面的强大能力,其常被用于交通数据的时间相关性建模,如时间图卷积网络(temporal graph convolutional network, TGCN)<sup>[18]</sup>、交通图卷积长短期记忆神经网络(traffic graph convolutional long short-term memory neural network, TGCLSTM)<sup>[19]</sup>、扩散卷积递归神经网络(diffusion convolutional recurrent neural network, DCRNN)<sup>[11]</sup>、自适应图卷积递归网络(adaptive graph convolutional recurrent network,

AGCRN)<sup>[2]</sup>等。RNN在建模中依赖于前一个时间步骤的结果,而1D-CNN不受此限制,因此可以实现并行计算,在时间相关性建模方面效率更高,相关研究包括时空图卷积网络(spatio-temporal graph convolutional networks, ST-GCN)<sup>[20]</sup>、基于注意力机制的时空图卷积网络(attention based spatial-temporal graph convolutional networks, ASTGCN)<sup>[21]</sup>等。有些模型引入了TCN,其每个隐藏层的感受野可以指数级地扩张,这样扩张的时间卷积模块可以捕捉更长的时间相关性,相关研究如Graph-WaveNet<sup>[3]</sup>、基于图神经网络的多元时间序列预测(multivariate time series forecasting with graph neural networks, MTGNN)<sup>[22]</sup>等。对于空间相关性的建模,常用的方法有CNN和GNN。由于传统的CNN模型只能对欧式空间中的数据进行建模,研究人员一般将城市地图按照经纬度划分为标准的网格结构,从而将不同区域的交通状况转换成“图像”形式,然后用CNN来学习不同区域之间的空间相关性,如时空残差网络(spatial-temporal residual network, ST-ResNet)<sup>[12]</sup>、时空动态网络(spatial-temporal dynamic network, STDN)<sup>[23]</sup>等。但是,真实的城市交通网络(如路网、地铁网)更类似于图的结构,而GNN适合这种结构,所以目前的研究大多采用GNN直接对交通数据进行建模。GNN模型的一个关键问题是图邻接矩阵的构建,现有研究包括根据路网结构预定义邻接矩阵的方法,如DCRNN<sup>[1]</sup>、STGCN<sup>[20]</sup>、TGCN<sup>[18]</sup>,以及从数据中自适应学习空间相关性来定义邻接矩阵的方法,如Graph-WaveNet<sup>[3]</sup>、MTGNN<sup>[22]</sup>、层次图卷积网络(hierarchical graph convolution networks, HGCN)<sup>[24]</sup>等。

基于深度学习的交通预测模型通常需要大量且高质量的数据做支撑,然而,在实际应用中,由于传感器采集误差等外部因素,通常导致采集到的数据集中存在很多缺失值和异常值<sup>[25]</sup>,而且能够获取的数据量通常也是有限的,比如几天或者几个月。因此,仍然需要构建稳健的时空预测模型对这类低质量数据进行良好的预测。本文针对存在误差和缺失的低质量交通流量预测问题展开研究,以加强学术界在这方面的研究。

## 2 问题定义

### 2.1 交通流量图

交通流量图(traffic flow graph, TFG)描述了一段时间内交通路网(或传感器网络)上的交通流量状况,如图1所示。其中,交通路网由图 $G=(V,E,A)$ 定义, $V$ 为节点集合,每个节点代表一条路段,且

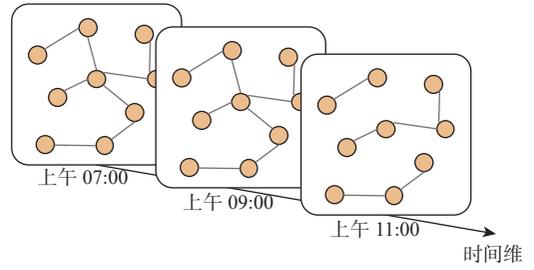


图1 TFG数据示例

Fig. 1 Illustration of traffic flow graph

$|V|=N$ 为路段总数, $E$ 为边的集合,描述节点之间的连通性, $A \in \mathbf{R}^{N \times N}$ 为图 $G$ 邻接矩阵;每一时刻 $t$ 的交通流量状况就是定义在图 $G$ 上的特征矩阵 $\mathbf{X}^t \in \mathbf{R}^{N \times F}$ , $F$ 为特征维度,除了流量之外通常还包含其他的辅助信息(如天气等),一段时间内的交通流量状况为 $\mathbf{X}^{(t-S):t} \in \mathbf{R}^{S \times N \times F}$ 。那么交通流量图可以写为 $G = [\mathbf{X}^{(t-S):t}; G]$ ,其时间相关性体现在 $\mathbf{X}^{(t-S):t}$ 中,而空间相关性由 $G$ 表示。

### 2.2 交通流量预测

交通流量预测任务就是学习一个函数 $f$ ,基于历史 $S$ 个时刻的TFG数据预测未来 $T$ 个时刻的交通流量变化,形式化为

$$[\mathbf{X}^{(t-S):t}; G] \xrightarrow{f(\cdot)} \mathbf{Y}^{t:(t+T)} \quad (1)$$

式中: $\mathbf{X}^{(t-S):t} \in \mathbf{R}^{S \times N \times F}$ 包含了时刻 $[t-S, t)$ 的输入特征; $\mathbf{Y}^{t:(t+T)} \in \mathbf{R}^{T \times N \times F}$ 为未来 $T$ 步的预测结果。

## 3 交通流量图对比学习框架

本文提出一个TFGCL框架,引入图对比学习来提升交通流量预测模型的稳健性。TFGCL框架通过图对比学习使得模型学到的交通流量数据表征既能够用于区分不同的交通流量模式,有助于提升模型的性能,又能够抵御数据噪声的干扰,能够增强模型的稳健性。

### 3.1 图对比学习

图对比学习的目标是学习一个图编码器,其能够有效地从输入数据中抽取图/节点表征。一个典型的图对比学习框架(如GraphCL<sup>[8]</sup>)的工作原理如下:对于每个输入图,采用随机数据增强方法生成2个相似视图,然后输入到图编码器和图读出函数中,分别得到2个视图的高级图表征,接着用一个非线性变换(称为投影头)将图表征映射到新的隐空间,在该空间中计算对比损失。

模型训练的时候,同一批次会有 $M$ 个图作为输入进行上述操作,因此得到 $2M$ 个图表征。不妨用 $\mathbf{z}_{n,i}$ 、 $\mathbf{z}_{n,j}$ 表示同一批次内第 $n$ 个图的2个相似视图产生的图表征,用 $\phi(\mathbf{z}_{n,i}, \mathbf{z}_{n,j}) = \mathbf{z}_{n,i}^T \mathbf{z}_{n,j} / (\|\mathbf{z}_{n,i}\| \|\mathbf{z}_{n,j}\|)$ 表示其之间的余弦相似度。那么,GraphCL中的对比

损失可以写为

$$\mathcal{L}_n = -\ln \frac{\exp(\phi(\mathbf{z}_{n,i}, \mathbf{z}_{n,j})/\tau)}{\sum_{n'=1, n' \neq n}^M \exp(\phi(\mathbf{z}_{n,i}, \mathbf{z}_{n',j})/\tau)} \quad (2)$$

式中:  $\tau$  为温度参数, 同一批次内的第  $n$  个样本共有  $2M-1$  个负样本。

为了将预训练的模型用于下游任务, 例如图分类, 通常在图编码器之后, 使用交叉熵损失训练一个线性分类器, 该图编码器可以被冻结 (对应线性评估) 或解冻 (对应微调), 此外, 在模型推理时一般会舍弃投影头。

### 3.2 交通流量图对比学习

本文 TFGCL 框架如图 2 所示, 其中  $G'$  所缺失的边和蓝色的结点表示对 TFG 数据中图结构和特征的扰动。首先, 利用 TFG 数据增强模块 (详见第 3.2.1 节) 将输入数据  $G = [X^{(t-S):t}; G]$  转化为其相似视图  $G'$ , 然后, 利用时空编码器将  $G$  和  $G'$  映射到表征空间得到  $H, H' \in \mathbf{R}^{S \times N \times D}$ , 其中  $D$  为表征空间的维度。表征被输入到 2 个分支中:

1) 预测分支。该分支将表征  $H$  输入到时空解码器中, 用于未来的预测交通流量, 预测结果  $\hat{Y}^{t:(t+T)}$  将用于与真实值  $Y^{t:(t+T)}$  计算交通流量预测损失。

2) 对比分支。该分支以  $H$  和  $H'$  作为输入, 利用平均函数在时间和空间维度做池化, 得到表征  $\mathbf{g}, \mathbf{g}' \in \mathbf{R}^D$ , 利用映射头将图表征映射到新空间中得到  $\mathbf{z}, \mathbf{z}' \in \mathbf{R}^D$ , 其中, 映射头由 2 个线性变换组成, 第

1 个线性变换之后带有线性整流 (rectified linear unit, ReLU) 激活函数, 通过对比损失函数来最大化正样本对 ( $\mathbf{z}$  和  $\mathbf{z}'$ ) 之间的相似度, 最小化负样本对 (过滤掉假负样本后) 之间的相似度。得到两部分损失之后, 通过对  $L_p$  和  $L_c$  两部分损失进行加权求和得到最终的损失  $L$ 。

#### 3.2.1 交通流量数据增强

数据增强是对比学习的重要组成部分。其主要负责构建语义上相似的样本, 影响着所学表征的质量<sup>[11]</sup>。目前针对图的数据增强技术主要考虑空间关系 (例如边扰动、子图采样), 但是没有考虑 TFG 数据独有的时间关系。因此, 本文提出 3 种专门针对 TFG 数据的增强方法, 分别从图结构 (边)、图特征 (节点)、时间域来对输入数据进行扰动, 使得模型学到的表征能够抵御数据噪声的干扰。注意在数据增强的时候只关注 TFG 数据中的交通流量信息, 其他辅助信息不做考虑, 因此本节  $X^{(t-S):t} \in \mathbf{R}^{S \times N}$ 。下面对数据增强方法给出具体的描述:

1) 边遮挡。图数据增强中的边扰动是指对无权图的边进行删除或增加。然而, TFG 数据中的图是带权重的, 在增加边的时候很难去给定一个好的权重, 这使得传统的边扰动方法难以完全迁移到 TFG 数据中。因此, 本文对边扰动进行修正来满足 TFG 数据, 即通过遮挡邻接矩阵  $A$  中的一部分带权边来扰动图结构。增强后的结果  $A'$  表示为

$$A'_{ij} = \begin{cases} A_{ij} & M_{ij} > \alpha_1 \\ \mathbf{0} & M_{ij} \leq \alpha_1 \end{cases} \quad (3)$$

式中:  $M$  采样自均匀分布  $U(0, 1)$ ;  $\alpha_1$  为可调节的参数, 表示边遮挡比例, 例如  $\alpha_1 = 0.1$  表示会有 10% 的边会被遮挡。针对同一批次的所有样本, 本文采用同样的增强邻接矩阵  $A'$ 。

2) 节点特征遮挡。受到环境因素的影响, TFG 数据采集过程中通常会出现数据丢失的现象。为了增强模型的抵御数据丢失现象的能力, 本文通过对 TFG 数据的节点特征进行随机遮挡来模拟该现象。通过节点特征遮挡得到的增强 TFG 数据  $R_{ij}^{(t-S):t}$  表示为

$$R_{ij}^{(t-S):t} = \begin{cases} X_{ij}^{(t-S):t} & M_{ij} > \alpha_2 \\ \mathbf{0} & M_{ij} \leq \alpha_2 \end{cases} \quad (4)$$

式中:  $M \sim U(0, 1)$ ;  $\alpha_2$  为可调节的参数, 表示节点遮挡比例。

3) 时间插值。TFG 数据天然地随时间连续变化, 但是负责采集数据的传感器只能给出离散的值 (例如 5 min 间隔的读数)。受此启发, 本文沿着时间轴移动数据, 以利用 2 个连续时间步之间的中间状态。具体实现为

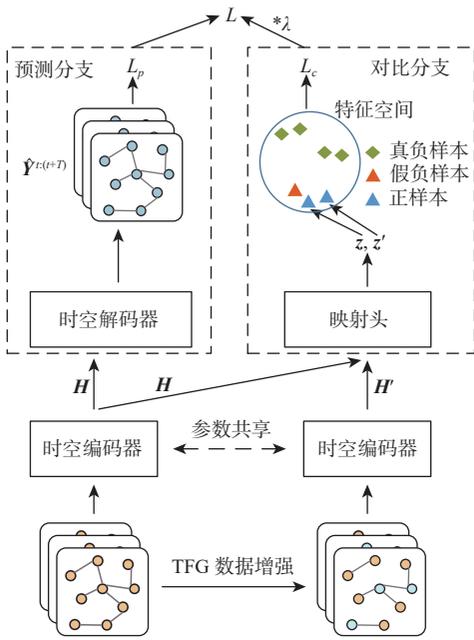


图 2 TFGCL 框架

Fig. 2 TFGCL architecture

$$\mathbf{R}^{(t-S):t} = \beta \mathbf{X}^{(t-S):t} + (1-\beta) \mathbf{X}^{(t-S+1):(t+1)} \quad (5)$$

式中:  $\beta \sim U(\alpha_3, 1)$ , 每一个样本都有其 $\beta$ ;  $\alpha_3$ 为可调节的参数, 表示在插值结果的组成中原始数据所占的比例不少于 $\alpha_3$ , 例如 $\alpha_3 = 0.5$ 表示每一个插值结果都有不少于0.5的成分来自原始数据, 其余的来自下一时刻的数据。这种数据增强方式类似于混淆增强的思想<sup>[26]</sup>, 区别在于本文在连续的时间域上进行加权平均来保证插值的精度。

注意, 节点特征遮挡和时间插值都是对节点特征进行干扰。为了保证插值结果的可靠性, 实验中先对节点特征进行插值, 再对其进行特征遮挡。

### 3.2.2 时空编/解码器

由于TFG数据具有复杂的时间和空间依赖, 本文考虑同时对时间依赖和空间依赖进行建模, 保证时空编码器能够有效提取时空依赖。为此, 本文提出一个融合图卷积网络(graph convolutional networks, GCN)<sup>[27]</sup>和Transformer<sup>[28]</sup>的时空编码器, 在保证Transformer建模动态时间信息的同时, 引入GCN捕捉空间依赖, 如图3所示, GCN为图卷积网络, Concat代表在时间维度上拼接, Att表示注意力机制。由于每个时间步的TFG数据都分别使用了不同的GCN, 因此, 可以捕捉到不同时间步动态的空间依赖性。

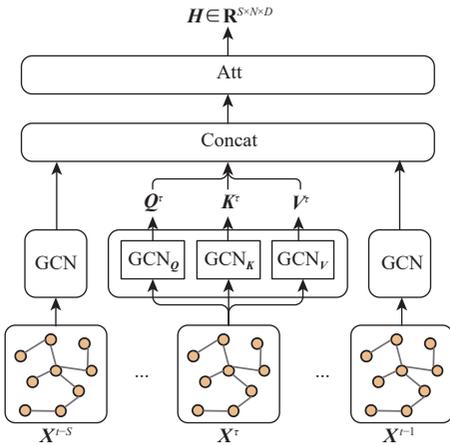


图3 时空编码器。

Fig. 3 Spatio-temporal Encoder

Transformer中最重要的结构是自注意力机制, 而自注意力机制的输入: 查询( $\mathbf{Q}$ )、键值( $\mathbf{K}$ )和值( $\mathbf{V}$ ), 通常由输入数据 $\mathbf{X}$ 经过线性变换得到, 线性变换不能很好地捕捉TFG数据中的空间依赖, 因此, 本文提出使用GCN替换线性变换。具体来说, 针对每一时刻的TFG数据 $\mathbf{X}^t \in \mathbf{R}^{N \times F}$ , 都有:

$$\begin{cases} \mathbf{Q}^t = \tilde{\mathbf{A}} \mathbf{X}^t \mathbf{W}_Q^t \\ \mathbf{K}^t = \tilde{\mathbf{A}} \mathbf{X}^t \mathbf{W}_K^t \\ \mathbf{V}^t = \tilde{\mathbf{A}} \mathbf{X}^t \mathbf{W}_V^t \end{cases} \quad (6)$$

式中:  $\tilde{\mathbf{A}} \in \mathbf{R}^{N \times N}$ 为带自环的正则化邻接矩阵;  $\mathbf{Q}^t, \mathbf{K}^t, \mathbf{V}^t \in \mathbf{R}^{N \times D}$ 为GCN的输出结果;  $\mathbf{W}_Q^t, \mathbf{W}_K^t, \mathbf{W}_V^t \in \mathbf{R}^{F \times D}$ 为GCN中待学习的参数矩阵。具体来说,  $\tilde{\mathbf{A}} = \mathbf{I}_n + \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$ ,  $\mathbf{I}_n$ 为单位矩阵,  $\mathbf{A}$ 为交通流量图的邻接矩阵;  $\mathbf{D} = \text{Diag}(\text{rowsum}(\mathbf{A}))$ 为邻接矩阵 $\mathbf{A}$ 的度矩阵,  $\text{rowsum}$ 表示行求和,  $\text{Diag}$ 表示构造对角阵。从公式可以看出, 针对某时刻的交通流量图 $\mathbf{X}^t \in \mathbf{R}^{N \times F}$ , 首先对其右乘一个参数矩阵进行特征的线性变换, 并对其左乘了一个带自环的正则化邻接矩阵, 使得流量图中每个节点都接收了来自一阶邻居的信息, 从而捕捉了局部的空间相关性, 当堆叠多个这样的模块之后, 就可以捕捉更长距离的空间相关性。为了便于加入时间维度后的计算, 将GCN输出结果的维度合并为 $d_{\text{model}} = N \times D$ , 即 $\mathbf{Q}^t, \mathbf{K}^t, \mathbf{V}^t \in \mathbf{R}^{d_{\text{model}}}$ 。

将输入数据 $\mathbf{X}^{(t-S):t}$ 分别通过GCN得到 $\mathbf{Q}^{(t-S):t}$ ,  $\mathbf{K}^{(t-S):t}$ ,  $\mathbf{V}^{(t-S):t} \in \mathbf{R}^{S \times d_{\text{model}}}$ , 用于计算注意力系数:

$$\text{Att}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax} \left( \frac{\mathbf{Q} \mathbf{K}^T}{\sqrt{d_{\text{model}}}} \right) \mathbf{V} \quad (7)$$

式中:  $\text{Att}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) \in \mathbf{R}^{S \times d_{\text{model}}}$ , 为了简化公式, 本文省略了3个输入的上标。最后, 将表征的后二维展开, 得到图表征 $\mathbf{H}^{(t-S):t} \in \mathbf{R}^{S \times N \times D}$ , 用于交通预测任务和对比学习任务。为了简化书写, 省略其上标, 用 $\mathbf{H}$ 表示学到的数据表征, 即 $\mathbf{H} = \mathbf{H}^{(t-S):t}$ 。

在原始的Transformer中, 解码器输出时间步长度与输入时间步保持一致, 然后提取键值( $\mathbf{K}$ )和值( $\mathbf{V}$ )用于解码器来完成序列-序列(Seq2Seq)的任务, 如机器翻译。由于交通预测任务也是一个标准的Seq2Seq任务, 所以时空编码器和时空解码器的交互过程与原始版本的Transformer一致。

### 3.2.3 双任务联合训练

为了保证交通流量预测的准确性和稳健性, 本文对流量预测任务和对比学习任务同时优化。流量预测分支使用平均绝对值误差(mean absolute error, MAE)作为损失函数, 定义为

$$\mathcal{L}_p = \frac{1}{M} \sum_{i=1}^M |\hat{\mathbf{Y}}_i^{t:(t+T)} - \mathbf{Y}_i^{t:(t+T)}| \quad (8)$$

式中:  $M$ 为一个批次数据内的样本总数。

在对比学习分支中, 传统损失函数没有考虑假负样本的情况, 这不利于对比学习模型有效学习数据表征。假负样本, 即与正样本相似度高的正样本被错标为负样本。这在TFG数据中很常见, 因为TFG数据中每个时刻的样本都与其他时刻的样本有关系, 例如邻近性和周期性<sup>[12]</sup>。举例来说, 早上07:00和07:30的交通状况变化不会很大, 周一早上07:00和周二早上07:00的交通状况也较为相

似, 这周和上周相同时间的交通状况也会呈现出类似的现象。因此, 这些语义相似的样本应该具有相似的特征, 在表征空间中被拉近, 而不是像传统方法一样被当作负样本推远。

为此, 本文提出一个简单的策略用于筛掉假负样本。该策略利用“当天时间”这个输入信息筛掉假负样本, 不将其纳入损失函数的计算当中。具体来说, 设定一个可调节的假负样本阈值  $\alpha_4$ , 用  $t_i$  代表输入数据的采样时间, 对于锚点数据及其采样时间  $t$ , 筛掉假负样本之后的负样本集合定义为

$$\tilde{Z} = \{z_i \mid |t - t_i| > \alpha_4\} \quad (9)$$

于是, 可以得到对比分支的损失函数为

$$L_c = \frac{1}{M} \sum_{i=1}^M -\ln \frac{\exp(\phi(z_i, z'_i)/\tau)}{\sum_{z_j \in \tilde{Z}} \exp(\phi(z_i, z_j)/\tau)} \quad (10)$$

式中:  $\phi(\cdot, \cdot)$  为余弦相似度。如果设定  $\alpha_3 = 0$ , 那么公式 (10) 就会退化为公式 (2) 中的传统对比学习损失。

最终的损失函数定义如下:

$$L = L_p + \lambda L_c \quad (11)$$

式中:  $\lambda$  用于在模型的 2 个分支之间进行权衡。

## 4 实验结果与分析

### 4.1 实验平台

本文所用实验平台的操作系统为 Ubuntu 16.04, 配置一个 64 GB 内存、2 个 CPU (Intel Xeon E5-2 680 v2) 和 1 个 GPU (NVIDIA GeForce GTX 1 080 Ti)。实验使用 Python 3.6.10 和机器学习库 PyTorch 1.4.0 来构建模型, 并利用 CUDA 9.2 加速模型训练。

### 4.2 数据集

为了评估本文 TFGCL 框架, 本文在 3 个真实交通数据集上进行了实验, 包括 METR-LA<sup>[1]</sup>、PEMS04 和 PEMS08<sup>[29]</sup>。

METR-LA 数据集记录了美国洛杉矶高速公路上 207 个传感器的交通流量数据。PEMS04 和 PEMS08 记录了美国加州 2 个不同地区的交通数据, 这些数据由加州交通局测量系统 (Caltrans Performance Measurement System, PEMS) 收集<sup>[30]</sup>, 原始数据中包含采样率为 30 s 的加州高速公路数据, 包含交通流量、平均速度和平均占用率数据, 本文着重对交通流量进行实验分析, 首先将交通流量数据以 5 min 的时间窗口进行汇总, 每天产生 288 个数据点, 然后使用 Z-score 对数据进行标准化, 并按照 7:1:2 的比例划分为训练、验证和测试集。

此外, 为了构建多步预测任务, 本文用 12 步的

历史数据来预测未来的 12 步。邻接矩阵由路网距离构造<sup>[31]</sup>, 具体为

$$A_{ij} = \begin{cases} \exp\left(-\frac{\text{dist}(v_i, v_j)^2}{\sigma^2}\right) & \text{dist}(v_i, v_j) \leq k \\ 0 & \text{其他} \end{cases} \quad (12)$$

式中:  $A_{ij}$  为邻接矩阵的元素;  $\text{dist}(v_i, v_j)$  为传感器  $v_i$  到传感器  $v_j$  之间的路网距离;  $\sigma$  为所有距离的标准差;  $k$  为阈值。所有数据集的统计信息如表 1 所示。

表 1 数据统计信息

Table 1 Statistical information on data

数据集	节点个数	边个数	时间步个数	数据缺失率/%
METR-LA	207	1 515	34 272	8.11
PEMS04	307	340	16 992	3.18
PEMS08	170	295	17 856	0.70

### 4.3 基准模型与评估指标

本文在 3 个评估指标下对比了 8 个基准模型, 评估指标包括平均绝对误差 (mean absolute error, MAE)、均方误差 (root mean square error, RMSE) 和平均绝对百分比误差 (mean absolute percentage error, MAPE), 对比的基准模型如下:

1) 传统模型: 历史平均 (historical average, HA) 模型。将交通流按季节建模, 用历史季节的加权平均值预测。

2) ARIMA 模型: 一种经典的时间序列预测模型。

3) 时空深度学习模型: 门控循环单元 (gated recurrent unit, GRU) 模型。RNN 的变种, 以 GRU 作为隐藏单元的编码器-解码器框架, 适用于时间序列预测。

4) 扩散卷积循环神经网络 (diffusion convolution recurrent neural network, DCRNN)<sup>[1]</sup> 模型: 用扩散图卷积替换 GRU 中的矩阵相乘以捕捉空间依赖关系, 并构建编码器-解码器框架用于时空预测。

5) 时空图卷积网络 (spatial-temporal graphconvolutional network, STGCN)<sup>[20]</sup> 模型: 基于卷积网络的模型, 结合图卷积和一维卷积来分别捕捉空间和时间的关系。

6) 图小波网络 (graph wavenet, GWNet)<sup>[3]</sup> 模型: 在图卷积中引入自适应邻接矩阵捕捉动态的空间依赖, 并使用一维因果卷积捕捉时间相关性。

7) AGCRN<sup>[2]</sup> 模型: 借用了图的思想, 通过自适应模块增强了传统的图卷积, 并将其组合到 RNN 中以捕捉时空相关性。

8) 多变量时间序列图神经网络 (multivariate time series forecasting with graph neural networks,

MTGNN)<sup>[22]</sup>模型:在学习空间表征的同时优化图结构以学到更好的空间表征,再结合TCN提取时空表征进行多维时间序列预测。

#### 4.4 实现细节与超参数

文中涉及的所有机器学习基准模型都用Pytorch实现。在TFGCL框架中,对于预测分支,本文堆叠了2层改进后的Transformer分别作为时空编码器和时空解码器,其中时空表征的维度 $D=128$ ,本文还用4个注意力头来稳定模型学习效果,其实现机制与原始版本的Transformer一致。最后,改进Transformer的前向传播中隐藏单元数为256。对于对比学习分支,映射头的线性变化中第1层的

输出维度为256,对比损失中的温度参数 $\tau=0.1$ ,本文从 $\alpha_1=\{0.05,0.1,0.15,0.2,0.25\}$ , $\alpha_2=\{0.01,0.03,0.05,0.07,0.09\}$ , $\alpha_3=\{0.1,0.3,0.5,0.7,0.9\}$ ,以及 $\alpha_4=\{0\text{ min},30\text{ min},60\text{ min},120\text{ min}\}$ 中根据模型在验证集上的效果来选定数据增强的阈值。模型损失函数中的平衡系数通过验证集的效果从 $\lambda=\{0.01,0.05,0.1,0.5,1.0\}$ 中确定。本文采用随机梯度下降优化器,模型的最大训练轮数为100,批次大小为128。对于所有的实验,本文在不同的随机种子上重复5次,并报告了评估指标的平均值。

#### 4.5 测试结果分析

表2为不同模型的实验结果。从表2中可以发现:

表2 不同模型在数据集METR-LA、PEMS04和PEMS08上的性能比较

Table 2 Performance comparison of different models on the METR-LA, PEMS04 and PEMS08 datasets

数据集	模型	MAE			MAPE			RMSE		
		15 min	30 min	60 min	15 min	30 min	60 min	15 min	30 min	60 min
METR-LA	HA	4.145	4.145	4.145	0.129	0.129	0.129	7.772	7.772	7.772
	ARIMA	3.853	4.925	6.915	0.093	0.137	0.174	6.475	9.608	14.539
	GRU	3.879	3.971	4.101	0.109	0.113	0.120	7.468	7.746	8.231
	DCRNN	2.709	3.125	3.599	0.070	0.085	0.105	5.254	6.386	7.620
	STGCN	2.793	3.167	3.540	0.073	0.087	0.102	5.455	6.514	7.525
	GWNet	2.804	3.206	3.587	0.074	0.091	0.108	5.491	6.530	7.421
	AGCRN	2.836	3.209	3.574	0.075	0.089	0.102	5.507	6.535	7.424
	MTGNN	<u>2.660</u>	<u>3.038</u>	<u>3.467</u>	<u>0.069</u>	<u>0.083</u>	<u>0.099</u>	<u>5.114</u>	<u>6.132</u>	<u>7.217</u>
	TFGCL框架	<b>2.517</b>	<b>2.827</b>	<b>3.245</b>	<b>0.066</b>	<b>0.079</b>	<b>0.093</b>	<b>4.875</b>	<b>5.806</b>	<b>6.763</b>
	PEMS04	HA	24.520	24.520	24.520	0.167	0.167	0.167	39.838	39.838
ARIMA		21.822	24.681	28.869	0.146	0.170	0.221	35.521	41.218	53.266
GRU		22.441	22.506	22.583	0.155	0.156	0.157	36.286	36.342	36.447
DCRNN		19.581	21.467	24.864	0.133	0.147	0.175	31.125	34.067	39.228
STGCN		19.149	20.133	21.567	0.133	0.140	0.150	30.301	31.886	34.200
GWNet		<u>17.692</u>	<u>18.574</u>	19.956	0.125	0.131	0.149	<u>28.516</u>	<u>29.888</u>	<u>31.848</u>
AGCRN		18.132	18.834	<u>19.851</u>	<u>0.124</u>	<u>0.130</u>	<u>0.139</u>	29.221	30.464	31.965
MTGNN		17.929	18.760	20.135	0.133	0.140	0.150	28.837	30.296	32.510
TFGCL框架		<b>17.377</b>	<b>18.093</b>	<b>19.290</b>	<b>0.122</b>	<b>0.127</b>	<b>0.135</b>	<b>28.271</b>	<b>29.080</b>	<b>30.668</b>
PEMS08		HA	21.193	21.193	21.193	0.138	0.138	0.138	36.644	36.644
	ARIMA	20.992	24.998	28.953	0.140	0.168	0.179	36.570	39.545	48.436
	GRU	19.992	20.126	20.461	0.124	0.125	0.127	32.276	32.569	33.200
	DCRNN	15.139	16.619	19.345	0.098	0.107	0.125	23.476	25.982	30.058
	STGCN	15.166	16.188	17.819	0.100	0.106	0.116	23.615	25.401	27.818
	GWNet	<u>13.486</u>	<u>14.349</u>	<u>15.672</u>	<u>0.091</u>	<u>0.093</u>	<u>0.108</u>	<u>21.615</u>	<u>23.375</u>	<u>25.855</u>
	AGCRN	14.146	14.962	16.427	0.092	0.097	0.121	22.241	24.055	26.557
	MTGNN	14.001	14.883	16.583	0.097	0.101	0.116	21.968	23.624	26.128
	TFGCL框架	<b>13.277</b>	<b>14.057</b>	<b>15.221</b>	<b>0.090</b>	<b>0.091</b>	<b>0.105</b>	<b>21.280</b>	<b>22.900</b>	<b>25.112</b>

注:其中最优结果用粗体表示,次优结果用下划线表示。

1) TFGCL框架在不同数据缺失率的数据集上表现均为最优。数据缺失率在3个数据集上存在差异(见表1),本文TFGCL框架在所有数据集上表

现均为最优,但次优的模型却不同如图4所示,模型性能排序越小越好,例如在PEMS08上是GWNet,在METR-LA上却是MTGNN,同时GWNet在METR-

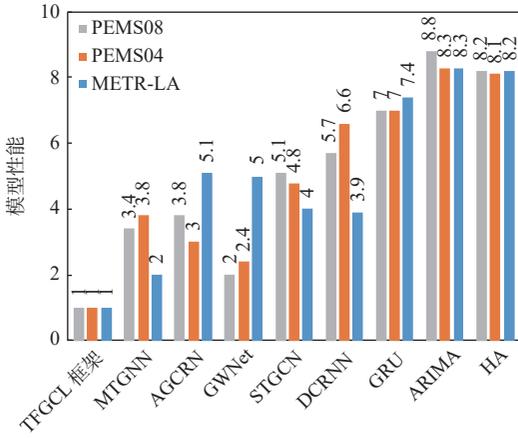


图 4 模型性能在不同数据集的平均排序对比

Fig. 4 Comparison of model performances of average ranking on different datasets

LA 上表现排名第 5。这是因为 TFGCL 框架在训练过程中通过数据增强和对比学习已经适应了数据扰动情况,即使在数据缺失的情况下也能学到高质量的表征,从而做出更稳健的、受数据扰动影响更小的预测。

2) TFGCL 框架在缺失率更大的数据集上提升更多。数据缺失率在不同数据集的排序为 METR-LA>PEMS04>PEMS08(见表 1),而如表 3 所示 TFGCL 框架相较于最优基线模型的提升排序却是 METR-LA>PEMS04>PEMS08。这种现象一方面说明已有的交通流量预测模型对于数据缺失较为敏感,另一方面也验证了本文引入的对比学习策略的确能够有效应对这种状况。

表 3 TFGCL 框架较最优基线模型的提升

Table 3 The improvement of TFGCL compared to optimal baselines model

数据集	MAE	MAPE	RMSE	平均得分
METR-LA	6.24	4.54	5.42	5.40
PEMS04	2.57	2.03	2.42	2.34
PEMS08	2.15	2.12	2.15	2.14

3) TFGCL 框架在长时预测上的提升更大。如图 5 所示,TFGCL 框架在长时预测上相较于最优基线模型的提升通常更大,因为 TFGCL 框架在短时预测和长时预测上的性能平衡地更好,所以针对不同长度的时间步的预测更加稳健,而其他基线模型在长时预测时效果损失地较快、也较为明显。

本节 3 个发现说明了本文 TFGCL 框架在真实数据集的交通流量预测问题上表现地更为稳健,不论是不同数据缺失率还是不同预测时长,TFGCL 框架相较于基线模型都有明显的提升。

### 4.6 模型稳健性分析

为进一步评估 TFGCL 框架的稳健性,设计实

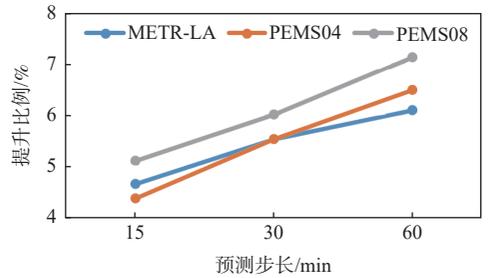


图 5 TFGCL 框架在不同预测步长下较最优基线模型的提升 Fig. 5 Improvement of TFGCL compared to optimal baseline model on different prediction horizon

验如下:对标准化后的训练数据随机添加高斯噪声,以交通流量图中每个节点为单位产生数据扰动,其中,高斯噪声服从标准正太分布,扰动比例为 10%、20%、30%、40%、50%。此外,对被扰动的数据按照 10% 的比例构造数据缺失,相当于产生更大程度的扰动,本文采用数据遮挡(置 0)实现数据缺失。以不同模型的多步预测性能取平均值作为泛化性能指标,在同等数据扰动比例的情况下,模型的泛化性能越好,则代表模型越稳健。本文仅给出以 RMSE 作为评估指标不同模型在 PEMS08 数据集上的实验结果,因为其他评估指标和其他数据集上的结果与之类似。

实验结果如图 6 所示,从中有 3 个点发现:①随着数据扰动率逐渐越大,模型性能普遍出现下滑趋势,因为数据扰动会给模型带来干扰导致性能下降;②时空深度学习模型性能下滑速度比传统模型快,这说明时空深度学习模型容易受到数据扰动影响;③ TFGCL 框架的性能下滑在所有时空深度学习模型中最小,也就是 TFGCL 框架的稳健性较强,因为本文专门设计了数据增强模块对输入数据进行扰动,并利用对比学习损失函数赋予模型抵抗这种干扰的能力,从而达到稳健的交通流量预测。

### 4.7 模型消融实验

本文提出 3 种数据增强方法和 1 个假负样本过

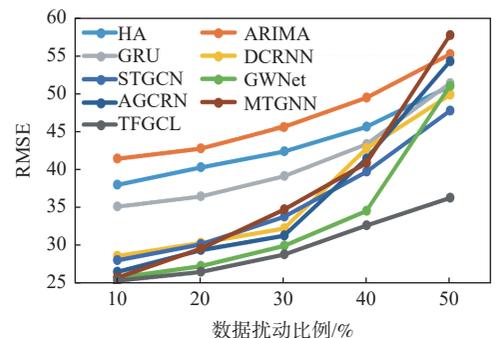


图 6 不同模型在 PEMS08 数据集上的稳健性比较 Fig. 6 Robustness comparison of different models on PEMS08 dataset

滤策略来赋予 TFGCL 框架抵抗数据噪声干扰的能力,提升模型的稳健性。下面对这些模块的有效性进行验证。为此,引入 TFGCL 框架的 4 个变种模型,分别是:① w/o EM,代表没有边遮挡模块的 TFGCL 框架;② w/o NM,代表没有节点属性遮挡模块的 TFGCL 框架;③ w/o TI,代表没有时间插值模块的 TFGCL 框架;④ w/o FN,代表没有假负样本过滤策略的 TFGCL 框架。由于实验结果类似,这里只给模型以 RMSE 为评估指标在验证集上的平均结果。

实验结果如图 7 所示。首先, w/o NM 与 w/o FN 相较于 TFGCL 框架的性能下降较大,验证了节点

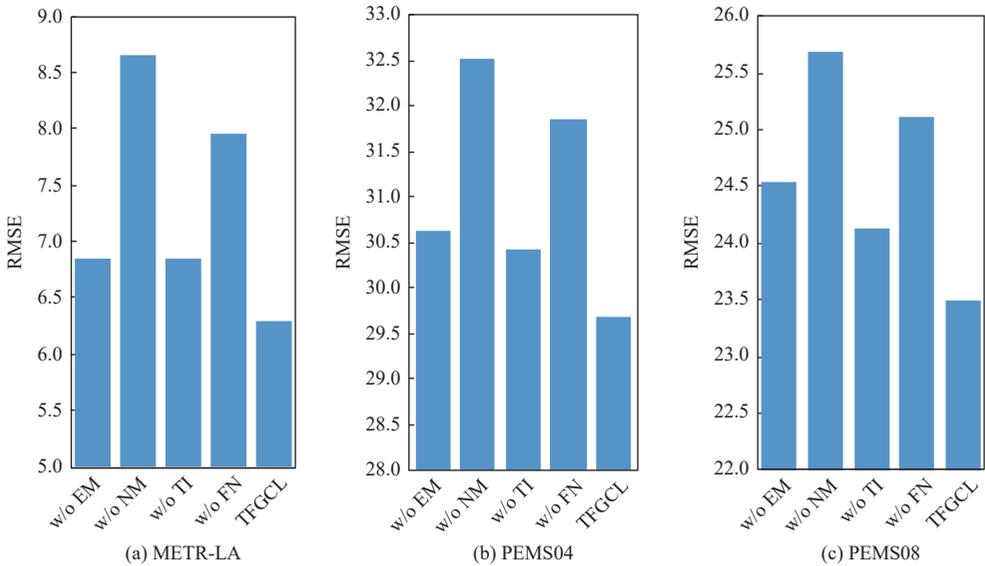


图 7 不同数据增强方法与策略的消融实验

Fig. 7 Ablation study on different data augmentation methods and strategies

其次, w/o EM 与 w/o TI 相较于 TFGCL 框架的性能下降较小,说明边遮挡和时间插值的有效性较弱。因为边遮挡是基于全局随机采样方法的,可能会遮挡掉一些权重较低的对模型来说不重要的边,因此局限了该方法的有效性,观察边遮挡在 3 个数据集上的有效性变化可以发现:在边较多的数据集中,边遮挡的有效性越弱,因为其中有大量低权重的边,而在边较少的数据集中,边遮挡的有效性逐渐提升,因为较大的概率遮挡掉重要的边,迫使模型在没有改变的情况下也能做预测,从而使得模型的预测更稳健。时间插值相当于一种借助相邻数据进行数据补全方法,该方法相当于对数据进行平滑操作,并不会引入很强的数据增强结构,因此其有效性较为有限。

总的来说,这些变种模型性能排序为: w/o NM < w/o FN < w/o EM < w/o TI,也就是说,所提出数据增强方法和策略的有效性排名为:节点属性遮挡 > 假负样本过滤 > 边遮挡 > 时间插值。

属性遮挡和假负样本过滤策略的有效性。节点属性遮挡会将所选节点一段时间内的属性全部置 0,模拟了交通流量数据采集过程中通常会出现的数据丢失现象,这直接要求模型必须在部分数据缺失的情况下做出准确的预测,因此,该数据增强方法的有效性最强。假负样本过滤策略排除掉了那些语义相似却被认为是负样本对的样本,避免了模型受到错误的指导,朝着错误的方向优化参数,因为交通流量数据中邻近性和周期性的存在使得传统的负样本对生成方法很容易产生较多的假负样本对,因此该策略的有效性较强,这也说明了时空数据对比学习中排除假负样本的重要性。

#### 4.8 计算效率对比

本文对比了不同模型在所有数据集上的运行效率。为了公平起见,主要对比性能较优的时空深度学习方法。计算了这些模型每个轮次 (epoch) 的平均训练时长 (train) 以及在验证集上的测试消耗时长 (eval),以此作为运行效率进行对比,实验结果如表 4 所示。

从表 4 可以发现所有模型的训练时间基本上与测试时间成正比,训练时间越长的模型测试时间通常越长; TFGCL 框架的运行效率总体优于其他模型,因为 TFGCL 框架采用了效率较高的 Transformer 来捕捉全局的时间相关性,而基于 RNN 架构的 DCRNN、AGCRN 需要迭代计算、耗时较长,基于 CNN 的 STGCN、GWNet 等在捕捉全局相关性时需要叠加较多的卷积层导致效率有所下降。虽然 GWNet 在 PEMS08 数据集上的评估时长与本文模型 TFGCL 框架接近,但 TFGCL 框架的训练时长较短,且在该数据集上,TFGCL 框架的性能优于 GWNet

表 4 不同模型运行效率比较  
Table 4 Operational efficiency comparison of different models

s

模型	train			eval		
	METR_LA	PEMS04	PEMS08	METR_LA	PEMS04	PEMS08
DCRNN	382.9	274.9	262.3	28.6	21.2	19.2
STGCN	40.3	8.7	20.3	4.7	3.8	2
GWNet	33.7	25.4	15	1.2	1	0.5
AGCRN	37.4	25	18.6	2.1	1.4	1
MTGNN	34	24.4	15.7	1.5	1.2	0.7
TFGCL框架	24.2	17.4	11.6	0.8	0.6	0.4

(见表 2)。

综合表 2 与表 4 来看, TFGCL 框架不仅能够取得稳健的交通预测性能, 而且可以达到较低的运行耗时。因此, 在实际应用中, TFGCL 框架的实用性更强。

## 5 结 论

1) TFGCL 框架针对 TFG 数据的独特时空特性, 对现有图对比学习方法进行改进使其能够赋能交通流量预测问题。

2) 分别从时间和空间 2 个方面设计了 3 种 TFG 数据增强方法用于对比学习, 并提出 1 个过滤策略用于过滤对比学习过程中的假负样本, 这 2 个方面的结合使得模型能够学到高质量的表征, 用于更稳健的交通流量预测。

3) 使用 3 个真实交通数据集, 与 8 种模型进行比较, 发现 TFGCL 框架的性能更加稳健, 尤其是在数据缺失较为明显的数据集上和长时交通流量预测任务中。

为使本文 TFGCL 框架能处理各种类型的时空预测任务, 仍需要优化数据预处理过程和实验过程的各项参数。

### 参考文献 (References)

- [1] LI Y G, YU R, SHAHABI C, et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting[EB/OL]. (2018-02-22)[2021-12-20]. <https://arxiv.org/abs/1707.01926.pdf>.
- [2] BAI L, YAO L N, LI C, et al. Adaptive graph convolutional recurrent network for traffic forecasting[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: ACM, 2020: 17804-17815.
- [3] WU Z H, PAN S R, LONG G D, et al. Graph wavenet for deep spatial-temporal graph modeling[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence. New York: ACM, 2019: 1907-1913.
- [4] TANG X F, YAO H X, SUN Y W, et al. Joint modeling of local and global temporal dynamics for multivariate time series forecasting with missing values[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(4): 5956-5963.
- [5] 李佩钰. 一种基于小波和神经网络的短时交通流量预测[J]. 计算机技术与发展, 2020, 30(1): 135-139.  
Li P Y. Short-term traffic flow prediction based on wavelet and neural network[J]. Computer Technology and Development, 2020, 30(1): 135-139(in Chinese).
- [6] TIAN C Y, CHAN W K V. Spatial-temporal attention wavenet: A deep learning framework for traffic prediction considering spatial-temporal dependencies[J]. IET Intelligent Transport Systems, 2021, 15(4): 549-561.
- [7] 冯宁, 郭晟楠, 宋超, 等. 面向交通流量预测的多组件时空图卷积网络[J]. 软件学报, 2019, 30(3): 759-769.  
FENG N, GUO S N, SONG C, et al. Multi-component spatial-temporal graph convolution networks for traffic flow forecasting[J]. Journal of Software, 2019, 30(3): 759-769(in Chinese).
- [8] YOU Y N, CHEN T L, SUI Y D, et al. Graph contrastive learning with augmentations[EB/OL]. (2021-04-03)[2021-12-21]. <https://arxiv.org/abs/2010.13902.pdf>.
- [9] XU D K, CHENG W, LUO D S, et al. InfoGCL: Information-aware graph contrastive learning[EB/OL]. (2021-10-28)[2021-12-23]. <https://arxiv.org/abs/2110.15438.pdf>.
- [10] ZHU Y Q, XU Y C, YU F, et al. Graph contrastive learning with adaptive augmentation[C]// Proceedings of the Web Conference 2021. New York: ACM, 2021: 2069-2080.
- [11] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]// Proceedings of the 37th International Conference on Machine Learning. New York: ACM, 2020: 1597-1607.
- [12] ZHANG J B, ZHENG Y, QI D K. Deep spatio-temporal residual networks for citywide crowd flows prediction[C]// Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. New York: ACM, 2017: 1655-1661.
- [13] DAGANZO C F, LAVAL J A. Moving bottlenecks: A numerical method that converges in flows[J]. Transportation Research Part B: Methodological, 2005, 39(9): 855-863.
- [14] 田钧方, 朱陈强, 贾宁, 等. 基于轨迹数据的车辆跟驰行为分析与建模综述[J]. 交通运输系统工程与信息, 2021, 21(5): 148-159.  
TIAN J F, ZHU C Q, JIA N, et al. Review of car-following behavior analysis and modeling based on trajectory data[J]. Journal of Transportation Systems Engineering and Information Technology, 2021, 21(5): 148-159 (in Chinese).
- [15] SAEED N, LARSEN O I. Application of queuing methodology to analyze congestion: A case study of the Manila International Con-

- tainer Terminal, Philippines[J]. *Case Studies on Transport Policy*, 2016, 4(2): 143-149.
- [16] MO Z B, SHI R Y, DI X A. A physics-informed deep learning paradigm for car-following models[J]. *Transportation Research Part C: Emerging Technologies*, 2021, 130: 103240.
- [17] JORG B, HAMILTON J D. *Time series analysis*[J]. *Contemporary Sociology*, 1995.
- [18] ZHAO L, SONG Y J, ZHANG C, et al. T-GCN: A temporal graph convolutional network for traffic prediction[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 21(9): 3848-3858.
- [19] CUI Z Y, HENRICKSON K, KE R M, et al. Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 21(11): 4883-4894.
- [20] YU B, YIN H T, ZHU Z X. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting[C]// *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*. California: International Joint Conferences on Artificial Intelligence Organization, 2018: 3634-3640.
- [21] GUO S N, LIN Y F, FENG N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33(1): 922-929.
- [22] WU Z H, PAN S R, LONG G D, et al. Connecting the dots: Multivariate time series forecasting with graph neural networks[C]// *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York: ACM, 2020: 753-763.
- [23] YAO H X, TANG X F, WEI H A, et al. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33(1): 5668-5675.
- [24] GUO K, HU Y L, SUN Y F, et al. Hierarchical graph convolution network for traffic forecasting[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 35(1): 151-159.
- [25] YI X, ZHENG Y, ZHANG J, et al. ST-MVL: Filling missing values in Geo-sensory time series data[C]// *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. New York: AAAI Press, 2016: 2704-2710.
- [26] ZHANG H Y, CISSE M, DAUPHIN Y N, et al. Mixup: Beyond empirical risk minimization[EB/OL]. (2018-04-27)[2021-12-25]. <https://arxiv.org/abs/1710.09412.pdf>.
- [27] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[EB/OL]. (2017-02-22)[2021-12-26]. <https://arxiv.org/abs/1609.02907.pdf>.
- [28] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// *Proceedings of the 31st International Conference on Neural Information Processing Systems*. New York: ACM, 2017: 6000-6010.
- [29] SONG C, LIN Y F, GUO S N, et al. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(1): 914-921.
- [30] CHEN C, PETTY K, SKABARDONIS A, et al. Freeway performance measurement system: Mining loop detector data[J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2001, 1748(1): 96-102.
- [31] SHUMAN D I, NARANG S K, FROSSARD P, et al. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains[J]. *IEEE Signal Processing Magazine*, 2013, 30(3): 83-98.

# Robust traffic flow prediction based on graph contrastive learning

LIU Wei\*, JIA Suling

(School of Economics and Management, Beihang University, Beijing 100191, China)

**Abstract:** Robust traffic flow prediction, as the core technology of Intelligent Transportation Systems, is a long-standing but challenging task. The fact that current models need a lot of training data and are susceptible to noise disturbance is a major factor that is restricting the growth of this subject. In academia, graph contrastive learning can alleviate the data-demanding issue and improve the model's ability to resist data noise through data augmentation and contrastive learning. Therefore, this paper proposes a Traffic Flow prediction framework that incorporates graph contrast learning (TFGCL) for robust traffic flow prediction. The framework has three contributions: First of all, given the unique spatio-temporal characteristics of traffic flow graph (TFG) data, TFGCL proposes three TFG data augmentation methods from the perspective of time and space. Secondly, in order to learn high-quality representations, this work also suggests a filtering method to shield the model from harsh negative samples with identical semantics. Finally, TFGCL jointly trains the traffic flow prediction task and the graph contrastive learning task. Extensive experiments with 8 baselines on 3 real traffic datasets show that the prediction performance of the TFGCL framework is more robust (an improvement of 6.24% compared to the best baseline), especially in datasets with obvious data missing and long-term traffic flow forecasting tasks.

**Keywords:** traffic flow prediction; traffic flow graph; noise perturbation; graph contrastive learning; robustness of prediction