

学校代码：10255

学 号：2201825

東華大學

基于图卷积神经网络的交通流预测方法研究

**Research on Traffic Flow Prediction Method Based on
Graph Convolutional Neural Network**

学科专业： 控制科学与工程

论文作者： 吴万鹏

指导老师： 周武能教授

答辩日期： 2023 年 1 月



基于图卷积神经网络的交通流预测方法研究

摘要

随着社会的发展,汽车的数量不断增加,交通拥堵逐渐成为阻碍现代城市发展的主要问题之一。交通流数据记载了一段时间内,通过道路不同断面的交通实体数,对其进行实时准确的预测对于改善交通拥堵问题起着重要作用,但由于交通系统的强动态性和复杂性,很难得到一个相对准确的预测值。因此,研究交通流预测理论和方法对于更好地发挥城市交通路网的性能具有重要意义。

现今,循环神经网络(Recurrent Neural Networks, RNNs)凭借其记忆功能,迅速在时间序列预测方法中占据了重要比重,也是捕获交通流数据时间特性的基本方法之一。图卷积神经网络(Graph Convolutional Network, GCN)出现后,凭借其处理拓扑网络结构的优势,被广泛应用于捕获交通流数据的空间特性。RNNs 与 GCN 的结合也因此成为综合时空层面进行交通流预测的常用手段,但对长时预测的需求、邻接关系的缺失等给交通流预测任务带来了新的挑战,现有模型无法深入挖掘交通流时间序列间复杂的空间相关性和动态的时间相关性,往往达不到需要的预测精度。基于此,本文综合利用图卷积、一维卷积、注意力机制等深度学习方法构建交通流预测模型并进行实验验证,实验结果表明最终提出的交通流预测模型取得了优于基准模型的预测效果。

本文的主要研究工作和创新点如下。

1) 针对传统交通流预测方法无法深入捕获交通数据时空特性的问题,提出了一种基于图卷积与序列到序列结构 (Sequence To Sequence, Seq2seq)的交通流预测模型。首先利用基于时序相似性的图卷积计算各交通流序列间的相关度以建模空间依赖,然后利用由门控循环单元(Gated Recurrent Unit, GRU)组成的 Seq2seq 模型

建模时间依赖。最后针对数据集存在较多误差数据的情况将改进的损失函数用以模型训练，以改善模型的过拟合问题，增加模型的鲁棒性。选取数据集进行实验，验证该模型对于交通流预测具有改良效果。

2) 针对循环神经网络不能并行计算与传统图卷积无法充分利用交通数据空间信息的问题，提出了一种基于 k-shape 聚类与时空图卷积神经网络的交通流预测模型。首先利用 k-shape 聚类算法对交通流时间序列进行聚类以计算语义邻接关系，进而将语义邻接关系与空间邻接关系融合得到复合邻接关系并应用注意力机制得到最终的邻接矩阵。在得到邻接矩阵后，将其与交通数据一同输入堆叠的时空卷积模块，通过门控时间卷积与图卷积捕获交通数据的时空特征，进而得到最终的预测结果。选取数据集进行实验，并和已知的交通流预测方法进行对比，验证该模型对于交通流预测具有改良效果。

3) 固定的邻接矩阵无法捕获交通序列间的动态相关性，自注意力机制在感知长期依赖方面表现出突出的优势。基于此，提出了一种基于卷积自注意力机制与动态图卷积网络的交通流预测模型，首先将基于节点相似性生成的动态拉普拉斯矩阵用于图卷积，以深度挖掘节点之间的空间相关性；进而通过卷积神经网络捕获局部时间相关性，通过自注意机制捕获长期时间相关性。此外，为改善图融合和深层网络的过平滑问题，通过特殊的组合方式引入了自回归模块，增强了模型对输入数据非周期变化的感知。也就是说，在交通流预测中，本模型不仅考虑了多个时间序列之间的动态关系，还考虑了单个时间序列的长期、短期、周期性和非周期性变化。为本模块选取四个经典数据集进行模拟实验，并和已知的交通流预测方法进行对比，验证了该模型对于交通流预测具有改良效果。

关键词：交通流预测；动态图卷积；自注意力机制；时间序列聚类；自回归。



RESEARCH ON TRAFFIC FLOW PREDICTION METHOD BASED ON GRAPH CONVOLUTIONAL NEURAL NETWORK

ABSTRACT

With the development of society, the number of cars continues to increase, and traffic congestion has gradually become the main problem hindering urban development. Traffic flow data records the number of traffic entities passing through a certain section of the road for a period of time, and its real-time and accurate prediction plays an important role in improving traffic congestion. However, due to the strong dynamics and complexity of the traffic system, it is difficult to get a relatively accurate forecast value. Therefore, it is of great significance to study the theory and method of traffic flow forecasting to better utilize the performance of urban traffic road network.

Nowadays, recurrent neural networks (RNNs) quickly occupy an important proportion in time series forecasting methods by virtue of its memory function, and it is also one of the basic methods to capture the temporal characteristics of traffic flow data. After the emergence of graph convolutional network (GCN), it has been widely used to capture the spatial characteristics of traffic flow data by virtue of its advantages in processing topology. The combination of RNNs and GCN has therefore become a common method for traffic flow prediction at the comprehensive spatio-temporal level. However, the demand for long-term prediction and the lack of adjacency relationships have brought new challenges to the task of traffic flow prediction, existing models cannot dig deep into traffic flow. The complex spatial correlation and dynamic temporal correlation between stream time series often fail to achieve the required prediction accuracy. This thesis comprehensively uses deep learning methods such as graph convolution, one-dimensional convolution, autoregressive, and attention mechanism to build a traffic flow prediction model and conduct experimental verification. The experimental results show that the final traffic flow prediction model has achieved better than the benchmark the predictive performance of the model.

The main research work of this thesis are as follows:

1) Aiming at the problem that traditional traffic flow prediction methods cannot deeply capture the spatiotemporal characteristics of traffic data, a traffic flow prediction model based on graph convolution and sequence to sequence (Seq2seq) is proposed. First, the correlation between traffic flow sequences is calculated by graph convolution

based on temporal similarity to model spatial dependencies, and then a sequence-to-sequence model composed of gated recurrent units (GRUs) is used to model temporal dependencies. And for the case that there are more error data in the data set, the improved loss function is used for model training to improve the over-fitting problem of the model and increase the robustness of the model. The data set is selected for experiments to verify that the model has an improved effect on traffic flow prediction.

2) Aiming at the problem that the recurrent neural network cannot be calculated in parallel and the traditional graph convolution cannot make full use of the spatial information of traffic data, a traffic flow prediction model based on k-shape clustering and spatio-temporal convolutional neural network is proposed. First, the k-shape clustering algorithm is used to cluster the traffic flow time series to calculate the semantic adjacency relationship, and then the semantic adjacency relationship and the spatial adjacency relationship are fused to obtain a composite adjacency relationship, and the final adjacency matrix is obtained by applying the attention mechanism. After the adjacency matrix is obtained, it is input together with the traffic data into the stacked spatio-temporal convolution module, and the spatio-temporal characteristics of the traffic data are captured through gated temporal convolution and graph convolution, and then the final prediction result is obtained. The data set is selected for experiments, and compared with known traffic flow prediction methods, it is verified that the model has an improved effect on traffic flow prediction.

3) The fixed adjacency matrix cannot capture the dynamic correlation between traffic sequences, and the self-attention mechanism shows outstanding advantages in perceiving long-term dependencies. Based on this, a traffic flow prediction model based on convolutional self-attention mechanism and dynamic graph convolution is proposed. Firstly, the dynamic laplacian matrix generated based on node similarity is used in graph convolution to deeply mine the relationship between nodes. The spatial correlation between them; then the local temporal correlation is captured by the convolutional neural network, and the long-term correlation is captured by the self-attention mechanism. In addition, in order to improve the over-smoothing problem of graph fusion and deep network, an autoregressive module is introduced through a special combination method, which enhances the model's perception of aperiodic changes in input data. That is to say, in traffic flow forecasting, this model not only considers the dynamic relationship between multiple time series, but also considers the long-term, short-term, periodic and non-periodic changes of a single time series. For this module, four classic data sets are selected for simulation experiments, and compared with known traffic flow prediction methods, it is verified that the model has an improved effect on traffic flow prediction.

KEY WORDS: traffic flow prediction; dynamic graph convolution; self-attention mechanism; autoregressive.

目 录

第 1 章 绪论.....	1
1.1 课题研究背景及意义.....	1
1.2 国内外研究现状.....	2
1.2.1 时间依赖建模.....	3
1.2.2 空间依赖建模.....	5
1.2.3 图结构学习.....	6
1.3 论文研究内容及章节安排	7
1.3.1 论文研究内容.....	7
1.3.2 论文章节安排.....	8
第 2 章 交通流预测基本方法与原理	9
2.1 交通流预测参数.....	9
2.2 交通流特征.....	10
2.2.1 时间特征.....	10
2.2.2 空间特征.....	10
2.3 交通流预测概述.....	11
2.4 基于深度学习的单节点交通流时序预测方法	12
2.4.1 门控循环神经网络.....	12
2.4.2 序列到序列模型.....	13
2.4.3 一维卷积神经网络.....	13
2.5 本文中应用的其他深度学习方法	14
2.5.1 图卷积神经网络.....	14
2.5.2 注意力机制.....	15
2.6 本章小结.....	16
第 3 章 基于 GCN 与 Seq2seq 结构的交通流预测模型	17
3.1 引言.....	17
3.2 基于 GCN 与 Seq2Seq 模型的交通流预测模型构建.....	18
3.2.1 模型结构.....	18
3.2.2 基于空间相关性的 GCN 模块.....	20
3.2.3 改进的序列到序列模型.....	20
3.2.4 损失函数.....	21
3.3 基于 GCN 与 Seq2seq 模型的交通流预测方法实验及分析	23
3.3.1 实验数据.....	23
3.3.2 数据预处理.....	23
3.3.3 基准模型与实验参数.....	24

3.3.4 实验结果与分析	26
3.3.5 消融实验	27
3.4 本章小结	28
第 4 章 基于 k-shape 与时空图卷积网络的交通流预测模型	29
4.1 引言	29
4.2 基于 k-shape 与时空图卷积的交通流预测模型构建	30
4.2.1 模型框架	30
4.2.2 复合邻接关系构建	31
4.2.3 时空预测	34
4.3 交通流预测模型 KSTGCN 实验及分析	37
4.3.1 实验数据	37
4.3.2 数据预处理	38
4.3.3 基准模型与实验参数	38
4.3.4 实验结果与分析	40
4.3.5 消融实验	42
4.4 本章小结	43
第 5 章 基于卷积自注意力机制与动态图卷积神经网络的交通流预测模型	45
5.1 引言	45
5.2 基于卷积自注意力机制和动态 GCN 的交通流预测模型构建 ..	46
5.2.1 模型结构	46
5.2.2 邻接关系学习	48
5.2.3 时空预测网络	49
5.2.4 自回归模块	51
5.3 交通流预测模型 SDGCN 实验及分析	51
5.3.1 实验数据	51
5.3.2 数据预处理	52
5.3.3 基准模型与实验参数	52
5.3.4 实验结果与分析	53
5.3.5 消融实验	55
5.4 本章小结	57
第 6 章 总结与展望	59
6.1 总结	59
6.2 展望	60
参考文献	61



第1章 绪论

随着经济社会的发展,汽车的数量不断增加,大城市的交通拥堵问题愈加严重。实时准确的交通流预测在改善交通拥堵问题中起着重要作用,交通流量的预测问题近年来也因此成为了相关领域的研究热点。本章首先在节 1.1 中介绍了本课题的研究背景和意义,然后在节 1.2 中介绍了交通流预测的国内外研究现状,最后在节 1.3 中给出本文的研究内容与章节安排。

1.1 课题研究背景及意义

社会和国民经济的不断进步,使得汽车的数量不断增加,这虽然给人民的生活带来了极大的便利,但也使得城市中的交通拥堵问题愈加严重。交通拥堵除了增加了居民的出行时间成本外,在生产损失、能源成本、安全成本、污染造成的健康风险成本等方面也有较为严重的影响,控制交通拥堵以提高社会发展质量、改善人民的生活水平在当下显得非常迫切。交通流量预测是对给定区域未来特定时间间隔内的交通流量进行估计,实时准确的交通流预测可以给相关部门制定缓解交通拥堵的政策方案提供参考,但由于交通系统的强动态性和复杂性^[1],很难得到一个相对准确的预测值。因此,研究交通流预测理论和方法对于更好地发挥城市交通路网的性能具有重要意义。

交通流预测的研究已有几十年的历史,交通流预测算法也在研究人员的努力下,不断更新换代。在研究的早期,相关的交通流预测方法主要聚焦在时间层面的建模且预测方法主要分为以下三大类。

1) 第一类是基于传统统计的方法,包括基于卡尔曼滤波(Kalman Filter, KF)模型的方法^[2]、基于历史平均(Historical Average, HA)模型的方法^[3]和基于差分整合移动平均自回归(Autoregressive Integrated Moving Average, ARIMA)模型的方法^[4-5]。然而由于交通数据的随机性和非线性特性,基于传统统计方法的预测手段很难提供准确的预测结果。

2) 第二类是基于机器学习的预测方法,包括基于支持向量机回归(Support Vector Regression, SVR)模型的方法^[6-7]、基于贝叶斯网络的方法^[8]和基于K近邻算法(K-nearest Neighbors Algorithm, KNN)的方法^[9]。但机器学习的表达能力有限,基于机器学习的方法难以在大规模交通路网数据集上取得精准的预测结果。

3) 第三类是基于深度学习的预测方法, 包括被广泛用于感知时间序列变化的循环神经网络(例如, 长短期记忆(Long Short-Term Memory, LSTM^[10])和GRU^[11-12]), 以及基于编码解码结构的 Seq2seq^[13]模型。但传统浅层的深度学习网络难以在当前复杂的交通情况下取得准确的预测结果, 且早期的预测方法也缺乏对空间依赖的建模, 在预测精度上还有提升的空间。

随着深度学习及相关硬件的发展, 计算能力不再是制约交通流预测方法发展的主要因素, 基于深度学习的交通流预测方法在目前已成为最新研究关注的重点。且在后续的研究中, 研究人员意识到交通节点间的相互影响, 在许多研究中综合利用处理空间依赖的 CNN^[14-15] 和捕获时间依赖的 RNN 来预测交通流量, 深度时空卷积网络(DSTCN^[1])就是一个使用 CNN 探索空间关系, 使用 RNN 建模时间信息的代表研究。在图卷积网络(GCN)出现后, 由于其在处理拓扑关系方面的优势, 逐渐取代了 CNN, T-GCN^[16] 就是一个结合 GCN 和 GRU 进行交通流预测且取得了良好的预测效果的例子。

以上的模型虽然可以达到较高的预测精度, 但仍有一些问题需要解决。首先, 有效地对交通数据的动态空间性进行建模仍然是一个挑战。图卷积通常被用于捕获空间相关性, 传统的 GCN 使用预定义的固定邻接矩阵来表征交通节点之间的邻接关系, 但是除了相邻节点之间存在空间相关性之外, 城市中的一些远程节点间也可能存在一定相关性, 即它们会共享一个相似的“时间”模式。其次, 现有方法学习时间依赖性的有效性较低。RNN 存在梯度消失的问题, 以至于其不能很好地捕捉时间间隔较远的交通数据间的依赖性, 虽然门控 CNN 在一定程度上改善了这个问题, 但它仍需要堆叠多层网络来增强长时预测能力。

为了解决这些问题, 基于现有深度学习算法、机器学习算法和优化算法, 提出具有创新性的交通流预测方法, 使交通流预测的效率和精度获得提升是现代工业背景下交通流预测研究的必行之路。

1.2 国内外研究现状

交通流预测利用各种传感器收集到的时空数据来预测未来的交通状况, 同时涉及空间和时间特征, 故可以看作是一种时空预测任务, 近年来的研究也主要集中在从这两个方面提高预测性能。如图 1-1 所示, 在时间特征提取方面, 主要分为基于传统统计理论的方法、基于机器学习的方法和基于深度学习的方法^[2-10]; 在空间特征提取方面, 近年来对交通网络的建模方法可分为基于欧氏空间的方法和基于非欧氏空间的方法。由于基于非欧氏空间的方法主要利用图卷积建模空间依赖, 但传统预定义的邻接矩阵不能充分表达交通节点的空间联系, 因此图结构学习也成为了研究人员关注的热点。

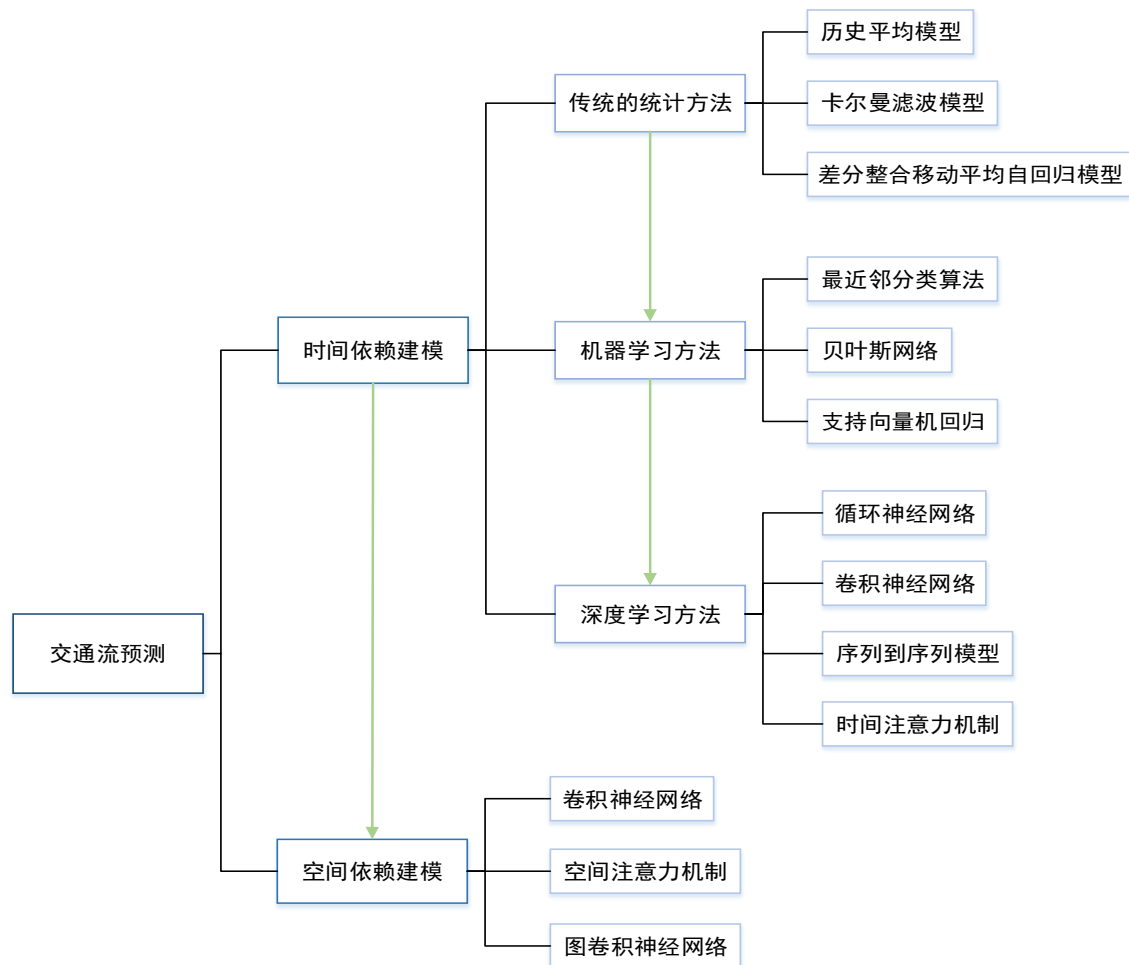


图 1-1 交通流预测方法分类

Fig.1-1 Classification of traffic flow prediction methods

1.2.1 时间依赖建模

对于交通流时间依赖建模的研究已有几十年的历史，早期对于交通流的研究多基于传统的统计方法且集中于时间依赖层面的建模。在计算机硬件得到发展后，各类机器学习方法与深度学习方法被并应用于交通流预测，且鉴于深度学习较强的表达能力，目前基于深度学习的方法已成为研究人员关注的重点。

在基于传统统计理论的方法中，基于 $KF^{[2]}$ 的方法使用包含了噪声和其他误差的一系列随时间变化的观测数据，产生对未知变量的估计；此类方法难以处理交通数据的随机性与周期性，在交通流预测领域的应用存在较大的局限性。 $HA^{[3]}$ 用一组最近的历史数据，来预测未来一期或多期的数据；这种方法简单易行，在时间序列没有季节性、周期性的情况下，具有良好预测效果，但在具有明显周期性的时序数据上表现乏力。基于 $ARIMA$ 的方法^[4-5]在 $ARMA$ 的基础上

增加了差分的操作，其基本思想是自回归，即当前点的值由前面几个临近点的值回归而来；ARIMA 模型对每个序列分别进行拟合，在处理大规模时间序列时效率较低，并且由于每个序列分别独立拟合，无法共享不同序列间存在的相似规律。

在基于机器学习的方法中，基于 SVR 模型的方法^[6-7]通过调整回归平面，让训练数据到该平面的距离之和最近，进而对未来的交通数据进行预测；但 SVR 不适合大型数据集，在交通数据集误差数据较多或单个数据点的特征维度大于数据的样本量时，SVR 表现不佳。基于贝叶斯网络^[8]的预测方法通过数理统计分析相邻道路间的邻接信息以提取交通数据的空间特征并判断当前预测点的未来趋势；但受限于机器学习的表达能力，贝叶斯网络对于大规模数据集难以达到良好的建模效果。基于 KNN 的方法^[9]通过欧几里得距离、曼哈顿距离等距离度量关系通过与待预测点相近的数值进行回归预测；KNN 不需要数据输入假定，且数据中的异常值对 KNN 模型影响较小，但其计算复杂度与空间复杂度较高。

随着人工智能技术的发展及相关硬件的完善，深度学习越来越多地被用于交通流预测。Dougher 首先将神经网络应用于交通流预测任务，Polson 等人建立了用于交通流预测的深度学习架构^[16]。在用于捕获时间序列依赖的 RNN 和用于提取空间特征的 CNN 出现后，相关研究逐渐将卷积神经网络和循环神经网络结合起来进行交通流预测^[17-18]，但是单纯结合两者的模型在获取高维交通数据的时空依赖性方面还有待改进。

在捕获时间依赖方面，RNNs 能用其内部向量保存之前输入的序列信息，在交通流预测领域应用广泛。但 RNNs 仍存在一些弊端，首先，基于 RNNs 的方法不能并行处理时间序列，从而增加了训练时间。其次，当 RNNs 与图卷积网络结合时，更容易遇到梯度消失的问题^[19-20]。

针对 RNNs 的上述局限，CNN、注意力机制等深度学习方法在一些研究中被用于建模时间依赖。CNN 能够并行的处理时间序列数据，因此在近年受到研究人员的青睐。其中扩展卷积网络是 CNN 的一种特殊形式，由空洞的 CNN 组成，它不必像 RNN 那样按顺序处理，并且梯度反向传播不会一直沿时间维度累积，从而避免了梯度弥散的问题。因上述优点，越来越多的研究将扩展卷积用于交通流预测，STFGNN^[21] 和 GraphWavenet^[22] 使用了并行门控扩张 CNN，通过门控机制和随层数加深而递增的膨胀因子从不同层面捕获交通数据的时间特性。但长期依赖的捕获仍需要堆叠大量扩展卷积层以增加 CNN 的感受野，如何更好地捕获长期依赖仍是一个需要深入研究的领域。



注意力机制具有学习时间信号间依赖性的能力, 被广泛引入到自动编码、深度卷积和生成网络中, 并取得了良好的效果。在时间序列预测研究中, 注意力机制已被证明是提取数据特征的一种有效技术。例如, Wu 等人^[23]提出了一种用于时序预测的多组件注意方法, 其中双向 LSTM 模型用于捕获季节性趋势和时序特征, CNN 模型用于捕获残差单元的局部趋势特征, 并引入注意力模块使得多组分流数据在最终的预测中可以连接高度相关的历史信息。ASTGCN^[24]和 DGCN^[25]使用了注意力机制, 直接不同时间点交通数据间的相关联系, 从而提升模型对长期序列的感知能力。谷歌公司在 transformer^[26]中提出利用自注意力机制取代传统的循环神经网络, 并在自然语言处理领域取得了成功的应用。相对其它时序预测方法自注意力机制更擅长捕捉时间序列数据的内部相关性, 同样 CNN 的内在结构也很难比像自注意机制^[27-29]这样的自回归模型更好, 自注意力机制在长期依赖捕获方面表现出突出的优势。

1.2.2 空间依赖建模

在捕获空间依赖方面, 一些研究^[30-34]基于欧几里得空间将交通网络建模为规则的二维网格, 由于 CNN 具有稳定不变的特性, 能够很好地提取图像的空间特征, 故可以通过 CNN 处理二维交通网络数据以捕获空间依赖性。但拓扑结构不具备图片的平移不变性, 因此传统的卷积方法不再适用, 基于非欧几里德空间建模的交通流预测研究逐渐成为热点。

在利用欧几里得空间进行交通流预测的研究中, Conv-LSTM^[30]模型首先将交通流信息映射到一维向量空间, 然后将一维向量组合成交通信息矩阵并通过 CNN 结合 LSTM 的方式提取交通数据的时空特征; ST-ResNet^[31]模型将城市中的轨迹信息按照不同的地理位置划分成代表不同区域的网格地图, 统计各个区域入口的流入流出量, 并将其转化为二通道矩阵, 采用类似图像处理的方式处理交通数据; Zhang 等人^[1]在将实际地理信息划分为网格地图的基础上进一步把网格地图建模为加权有向图, 经过图嵌入操作后将其输入 CNN 进行空间特征学习。RSTN^[32]通过残差网络将 CNN、LSTM 和 Conv-LSTM 模块结合起来, 以捕获时空依赖关系, 将传统的交通预测问题视为每个时间区间内出行密度的残差函数。MGSTC^[33]通过多个时空卷积分支探索不同层面的时空相关性, 并将时空特征与外部因素动态结合来提升模型预测性能。

在使用非欧几里德空间建模的交通流预测研究中, 常使用图卷积来捕捉空间相关性。图卷积主要有两种类型, 一种是基于谱的方法^[34], 另一种是基于空间的方法^[35]。STGCN^[36]为充分利用空间信息, 采用通用图结构对交通网络进行建模, 将交通网络中的站点建模为节点, 图中的节点包含在不同时间测得的一

系列观测数据，而后利用图卷积和门控线性单元分别提取空间和时间特征；Chai 等人^[37]使用多图卷积网络来预测交通流量，基于交通网络中站点之间的距离、交通交互和相关性建立多个图并进行图融合，进而通过图卷积和编码解码网络学习时空特征。

传统图卷积中所用的邻接矩阵大多为固定的取值，由现实世界中的邻接关系决定。但预定义的固定邻接矩阵并不能很好地表示交通时间序列之间的空间关系。首先，交通网络在不断变化。其次，除了空间上相邻的节点之间的直接相关性之外，距离较远的节点也可能由于相似的流量模式而具有语义相关性。一些研究已经注意到了这一点并对其进行了改进，ASTGCN 使用注意力机制^[24]结合 GCN 来进一步捕获空间相关性，STFGNN^[21]、Graph WaveNet^[22]和 MTGNN^[38]使用图结构学习来挖掘节点之间的空间关系。

1.2.3 图结构学习

目前的图结构学习主要通过计算时间序列间的相似度来获取邻接矩阵，可以分为静态图和动态图两类。静态图构建方法旨在利用已有知识在预处理期间构建图结构，早期的图构建方法仅利用道路之间的连接关系得到一个二值邻接矩阵^[39-41]，为进一步改进图结构，一些研究引入了一些度量时间序列静态相似度的方法，如 DTW^[42]、皮尔逊系数法等^[43]，根据节点间的时间相似度得到静态图。

动态图的应用则是为了进一步提高邻接矩阵的表示能力，由于交通网络的动态特性，静态图在描述复杂的道路网络时有天然的不足，有研究引入基于二值邻接矩阵的注意力机制，使节点的影响权重动态变化，进而适应路网的动态变化。但是，除了相邻节点之外，一些较远的交通节点也可能存在一定程度的相关性，所以对动态图结构学习需要有更加深入的研究。现阶段动态图的实现主要依靠节点嵌入^[44-45]和动态时间序列相似性度量^[21]两类方法，通过节点嵌入获取自适应邻接矩阵可实现性能提升，其中邻接矩阵的参数可以通过训练过程学习；动态相似性度量则通过实时的计算交通序列间的相似性并更新邻接矩阵来捕获交通数据的动态联系。

时间序列相似性度量方法也被广泛应用于图结构学习。现有的相关性测量方法大致可以分为以下三类。一是基于欧氏距离^[46]，对于长度相同的序列，计算每两个点之间的距离，然后求和，距离越小，相似度越高；但是该方法无法区分形状相似性，无法反映时序趋势动态范围变化的相似性。二是基于模式距离，该类方法将时间序列根据变化趋势划分为不同的子集，每个子集代表一类变化模式。第三种基于形状距离，如 DTW^[42]，DTW 先将时间序列从形状上对



齐, 然后通过动态规划计算一序列上的点到另一序列上各个点的距离之和并取距离和的最小值来衡量序列的相似性, 但利用 DTW 算法生成邻接图具有较高的时间复杂度。

总的来说, 目前的交通流预测方法亟待优化的方向如下。第一, 深度建模时空依赖; 第二, 基于静态邻接矩阵的图卷积方法在挖掘空间联系上的作用有限; 第三, 长时依赖的捕获仍是一个需要深入研究的方向。

1.3 论文研究内容及章节安排

1.3.1 论文研究内容

基于上述对本课题国内外研究现状的分析, 本文的主要研究内容如下。

1) 针对由传感器测得的交通流数据误差较多和现今的交通流预测方法不能有效捕获交通数据间复杂时空特性的问题, 本文提出了基于 GCN 与 Seq2seq 结构的交通流预测模型, 首先利用皮尔森系数计算交通流序列间的相似性并结合图卷积捕获空间特性, 然后利用由 GRU 组成的序列到序列模型捕获时间特性。将改进的损失函数用于模型训练, 选取真实世界的交通数据集进行实验, 并和基础的交通流预测方法进行对比, 验证该模型具有改良效果。

2) 针对循环神经网络不能并行训练与传统图卷积无法充分利用交通路网空间信息的问题, 提出了一种基于 k-shape 聚类与时空图卷积神经网络的交通流预测模型。本文将聚类思想用于交通流预测, 首先利用 k-shape 聚类算法对交通流时间序列进行聚类以计算语义邻接关系, 并将语义邻接关系与空间邻接关系进行融合得到复合邻接关系, 继而在复合邻接关系上应用了注意力机制得到最终的邻接矩阵。最后将邻接矩阵与交通数据输入堆叠的时空卷积模块, 通过门控一维卷积与图卷积捕获交通数据的时空特征, 得到最终的预测结果。选取数据集进行实验, 并和已知的交通流预测方法对比, 验证该模型的预测性能。

3) 针对固定的邻接矩阵无法捕获交通序列间的动态相关性, 及传统预测模型长时预测能力不足的问题, 本章提出了一种基于卷积自注意力机制与动态图卷积神经网络的交通流预测模型。首先对节点进行空间嵌入, 通过设计的相似性计算方法计算节点间的相似性以生成用于图卷积的邻接矩阵, 从而深入挖掘节点之间的动态空间相关性; 其次通过卷积神经网络捕获局部时间相关性, 并通过自注意机制捕获长期相关性; 最后为了改善图融合和深度网络的过平滑问题, 通过特殊的组合方式增加了 AR 模块, 增加了对非周期序列的预测。为本模块选取四个经典数据集进行模拟实验, 并和已知的交通流预测方法对比, 验证该模型对于交通流预测具有改良效果。

1.3.2 论文章节安排

本文在结构上主要分为六章，论文的章节关系如图 1-2 所示，每章的主要内容如下。

第 1 章：首先阐述了本课题的研究背景及意义；然后介绍了交通流预测方法的发展进程和国内外研究现状；最后点明本文的主要研究内容和章节安排。

第 2 章：先简要介绍了交通流预测原理，然后阐述了交通流预测的数学表述，最后介绍了之后提出的交通流预测方法中主要应用的各类算法，以便于更清晰地理解后续章节内容。

第 3 章：提出基于 GCN 与 Seq2seq 结构的交通流预测模型，有效的捕获了交通流的时空特性，并采用改进的损失函数训练模型，提高了预测模型的鲁棒性。

第 4 章：提出了一种基于 k-shape 聚类与时空卷积神经网络的交通流预测模型。利用 k-shape 聚类算法、注意力机制及 GCN 捕获交通数据的空间特征，利用门控时间卷积捕获交通数据的时间特征，通过实验验证了模型的性能。

第 5 章：提出了一种基于卷积自注意力机制与动态图卷积的交通流预测模型。通过动态图卷积建模交通数据的动态空间联系，通过一维卷积、自注意力机制及 AR 建模交通数据的时间联系，通过实验验证了本章模型具有良好的预测性能。

第 6 章：对本论文的研究内容进行总结，并对可以拓展研究的方向进行展望。

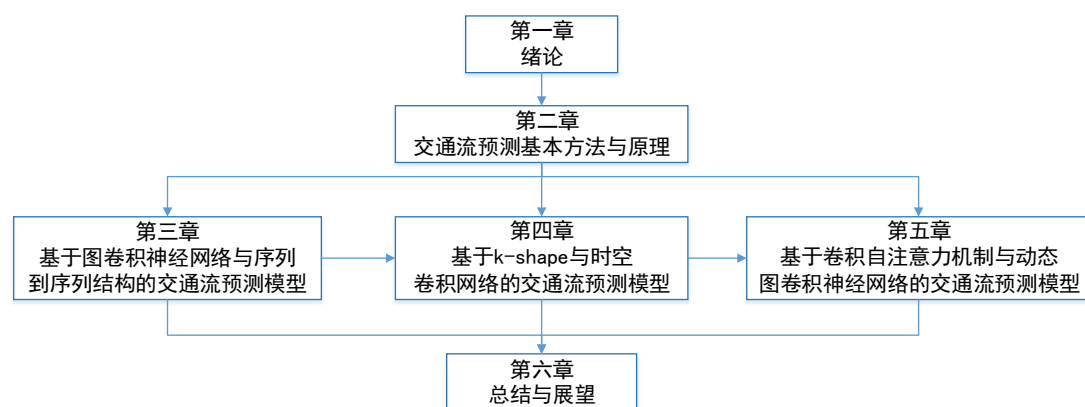


图 1-2 论文章节关系

Fig. 1-2 The thesis chapter relationship



第2章 交通流预测基本方法与原理

随着社会的发展,交通拥堵成为了阻碍现代城市进一步发展的主要问题之一。借助物联网、人工智能等新一代信息技术开展精准的交通流量预测,可以有效解决交通拥堵问题,缓解环境污染和气候变化。然而,交通数据的强时空相关性使得准确的交通预测任务成为一项挑战,研究交通流预测理论和方法以提高交通流预测精度,对于更好地发挥城市道路网的性能具有重要意义。

本章首先介绍了交通流预测的参数及交通流的时空特征,并对交通流预测任务进行数学描述;随后介绍了基于深度学习的单节点交通流预测方法与本文用到的其他深度学习方法,为后续构建改进预测模型打下基础。

2.1 交通流预测参数

交通流包括平均交通流量、占用率、交通流速度、车辆间距和交通密度等参数。在实际的交通流预测场景中,主要基于交通流的三个基本特征参数,综合分析交通网络的时空特征,进而对交通流进行特征判定与预测。因此,对宏观交通流的研究主要涉及的参数为平均交通流量、占用率和交通流速度。

1) 平均交通流量,表示在一个采样时间内通过道路特定位置的车辆数量。交通流量是交通情况最直观的参数表示,通常由线圈传感器测量得到。计算方法如式(2-1):

$$Q = \frac{N}{T} \quad (2-1)$$

式(2-1)中 N 为采样时间内通过道路特定位置的车辆总数, T 为选取的传感器采样时间跨度, Q 为计算得到的平均交通流量。

2) 交通流速度,表示车辆整体行驶速度的快慢,通常用行驶至指定道路断面车辆的平均速度作为速度标准。计算方法如式(2-2):

$$\bar{V}_t = \frac{1}{N} \sum_{i=1}^N V_i \quad (2-2)$$

式(2-2)中 N 为在给定时间内通过传感器放置位置的车辆数, V_i 为 t 时刻车辆 i 的瞬时速度, \bar{V}_t 表示计算得到的交通流平均速度。

3) 占用率,表示交通流的疏密程度,分为时间占用率与空间占用率两类。时间占有率表示测量时间内有车辆通过线圈传感器的时间占比;空间占用率则代表道路单位长度上车辆的数量,但空间占有率不易测量,通常用时间占有率代表交通流占有率。时间占有率的计算如式(2-3):

$$V^o = \frac{\sum_{i=1}^N t_i}{T} \quad (2-3)$$

式(2-3)中 T 为传感器的采样时间跨度, t_i 为第 i 车通过该路段的时间, N 为采样时间内通过该路段的车辆总数, V^o 表示计算得到的道路占用率。

2.2 交通流特征

2.2.1 时间特征

图2-1为深圳市龙岗区坂田街道五和稼先路口十日的交通流量数据,从图中可以看出同一交通节点的交通数据在多天内呈现出周期性非线性变化规律,但又在节假日与工作日具有不同的交通模式,且因人类社会活动的规律性,早上和傍晚为出行高峰,在夜间交通流量达到低峰。且因为传感器的测量误差或突发事件的影响,存在明显偏离周期曲线的数据,在历史数据曲线中表现为明显的毛刺。此外,天气、交通管制等影响因素也进一步增强了交通流数据的非周期性。

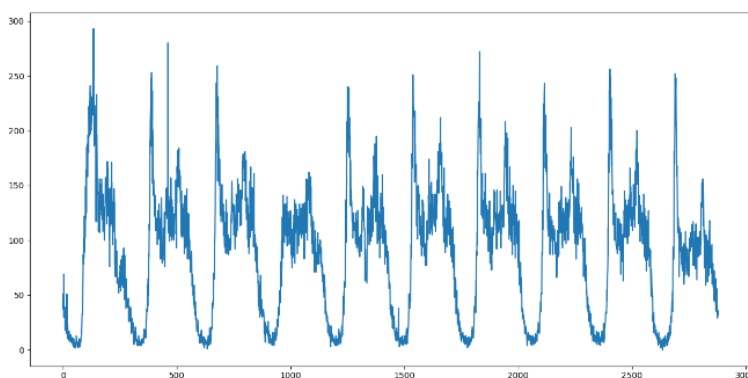


图 2-1 五和稼先路口历史流量

Fig. 2-1 Historical traffic flow at Wuhe-Jiaxian intersection

2.2.2 空间特征

交通流数据从设置在现实世界不同空间位置的传感器测量得到,单个传感器测得的交通流序列具有时间自相关性,故可以采用时间序列的处理方法对其进行处理;多个传感器的测量节点又构成一个复杂的连接系统,同方向上游的交通节点的交通情况对下游具有明显的影响,相邻节点间会具有较强的空间联系;但较远的节点间也可能因为具有相似的社会功能而具有相似的交通模式,如高速公路出入口、商场和医院等。因此,交通流量存在空间相关性,这是交通流量数据与其他时间序列数据最明显的区别。

2.3 交通流预测概述

交通预测任务可以表述为一个多变量的具有辅助先验知识的时间序列预测问题。道路网络表示为图 $G = (V, E)$ ，其中 V 是一组 N 个交通节点， E 是一组表示节点与节点间邻接关系的边。根据节点间的邻接关系，可得到预先确定的静态邻接矩阵来表征交通道路节点之间的空间交互如图 2-2 所示，其中 1 代表两节点相邻接，0 代表两节点不邻接。

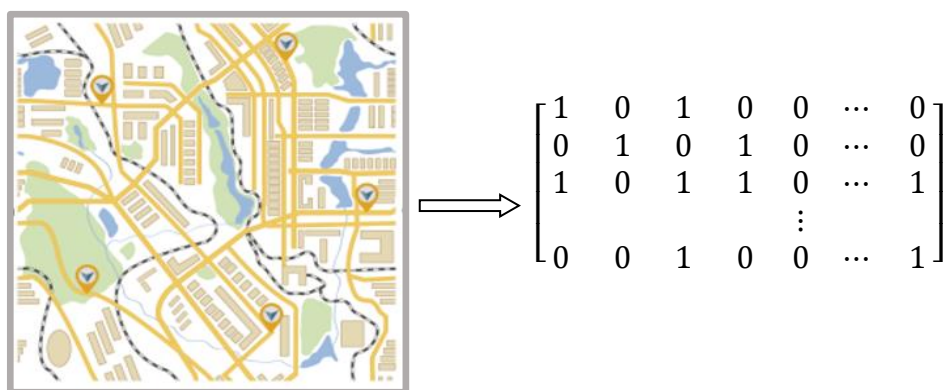


图 2-2 交通路网邻接矩阵

Fig. 2-2 Adjacency matrix of traffic road network

在交通网络中，通过 N 个不同传感器获取的交通流时间序列可表示为： $X = [X^1, X^2, \dots, X^N]$ 。在此，我们定义 $X_{in} = R^{N \times d \times T}$ 为模型的输入， $X_{out} = R^{N \times d \times T}$ 为模型的期望输出， $Y \in R^{N \times d \times T}$ 为模型的实际输出，其中 T 是输入的步长。那么第 i 节点的输入矩阵可以表示为 $X_{in}^i = [X_{t-T}^i, X_{t-T+1}^i, \dots, X_t^i]$ ，第 i 个节点的输出矩阵可以表示为 $X_{out}^i = [X_{t+1}^i, X_{t+2}^i, \dots, X_{t+T}^i]$ 。给定在交通网络 G 上观察到的张量 X 和空间邻接关系，交通流预测的目标是学习历史观察数据 X_{in} 到未来的交通量 X_{out} 间的映射函数，如图 2-3 所示。在完成映射函数的学习后，可通过其对历史数据 X_{in} 进行预测进而得到模型的实际输出 Y 。

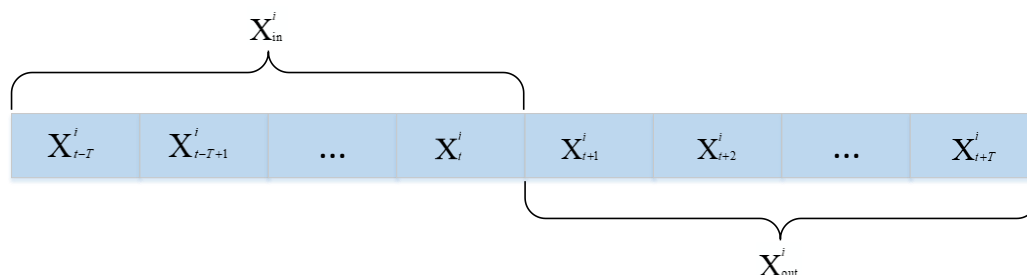


图 2-3 时间序列预测示意图

Fig. 2-3 Adjacency handling process

2.4 基于深度学习的单节点交通流时序预测方法

深度学习模型在机器视觉、自然语言处理等研究领域取得了成功应用后，也逐渐被引入到交通流时间序列预测问题中。不同时间序列通过共享一个深度学习模型，让模型能从多个序列中学到知识，从而提升了在大规模数据上的求解效率。

2.4.1 门控循环神经网络

与 LSTM 类似，GRU 也是循环神经网络的一种，虽然 LSTM 已被证明是避免梯度消失或梯度爆炸的可行选择，但其架构中有多个存储单元，具有更高的存储要求。GRU 具有与 LSTM 类似的调节自身内部信息流的门控机制，且在每个时间步公开状态，并计算现有状态与新计算状态的线性和，GRU 的框图如图 2-4 所示，在时刻 t 其内部状态 h_t 被线性建模为：

$$h_t = (1 - z_t)h_{t-1} + z_t\tilde{h}_t \quad (2-4)$$

式 (2-4) 中更新门 z_t 控制状态的更新， z_t 越接近 1，代表门控单元对上一时刻的值遗忘的越多，反之，状态的更新越小。 z_t 可由式 (2-5) 计算得到。

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \quad (2-5)$$

式 (2-5) 中 W_z 和 U_z 是要学习的权重矩阵，候选状态 \tilde{h}_t 的计算如下：

$$\tilde{h}_t = \tanh(W_h x_t + U_h (r_t \odot h_{t-1})) \quad (2-6)$$

式 (2-6) 中 r_t 是一组重置门，定义为：

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \quad (2-7)$$

这允许单元通过读取输入序列的第一个数据来忘记之前的状态。文献^[47]中概述了 GRU 网络和 LSTM 网络之间的一些异同，总的来说，两种模型中的任一模型仅在某些特殊任务中的表现优于另一模型，需要依据具体应用场景选用合适的模型。

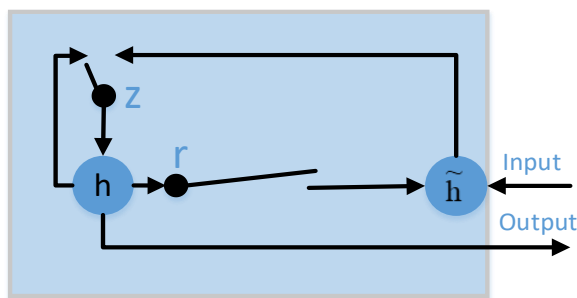


图 2-4 GRU 结构图

Fig. 2-4 GRU structure

2.4.2 序列到序列模型

Seq2Seq 模型是一种端到端的算法框架, 被广泛用于处理输入和输出序列长度可变的预测任务, 该模型最初被应用于语音识别、机器翻译^[48-50]和 PM2.5 浓度预测等时序任务中。

如图 2-5 所示, 基础的 Seq2seq 模型包含了用于处理输入的编码器、用于生成输出的解码器以及连接两者的中间状态向量三部分。其中编码器和解码器单元通常由 LSTM 或 GRU 组成, 编码器通过学习长度为 m 的时序输入 $X = [X^1, X^2, \dots, X^m]$, 并经过一系列编码单元的计算, 将输入编码成中间向量 S , 如式 (2-8)。

$$S = q(h_1, h_2, \dots, h_m) \quad (2-8)$$

式 (2-8) 中, h_i 为第 i 个编码单元的中间状态, 编码后 S 被传递给解码器进行解码, 在解码过程中, 解码器再通过对状态向量 S 的学习来得到输出, 可以将输入序列映射到可变长度的输出序列 $Y = [Y^1, Y^2, \dots, Y^n]$ 。

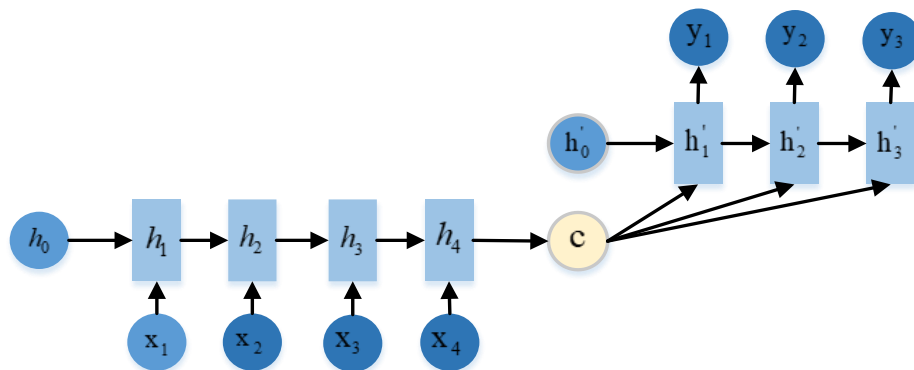


图 2-5 Seq2seq 结构图

Fig. 2-5 Seq2seq structure

2.4.3 一维卷积神经网络

在过去十年中, CNN 在机器视觉和图像处理领域大放异彩^[52], 在一系列基准实验和实际任务中取得了先进的水平。相比传统意义上的二维 CNN, 一维 CNN 在宽度或者高度方向上进行卷积窗口的滑动并相乘求和。近年的研究中, 一维卷积在处理诸如机器翻译^[51-52]、时序预测等时间序列任务中表现出了显著前景。2018 年, 白等人^[53]表明, 对于多序列建模任务, 如使用当前最佳实践的扩张卷积通常能表现出基于循环神经网络的模型同等水平或更加优异的性能。

一维 CNN 属于前馈神经网络, 由多个用于执行特征提取任务的卷积层及一个用于组合高级特征并预测所需输出的输出层组成。图 2-6 显示了一个基础的一维 CNN 架构示例, 图中每个卷积层都有用于池化操作的可学习卷积核, 这些卷

卷积核将一组权重与输入进行卷积并在其上应用非线性激活函数以从输入数据中提取高级特征，随后将卷积层的输出馈送到池化层中，池化层减小了核提取的特征空间大小，同时强调了每个卷积层学习到的主要特征。随着输入通过卷积阶段进行，网络会学习到更多针对于特定问题的特征。

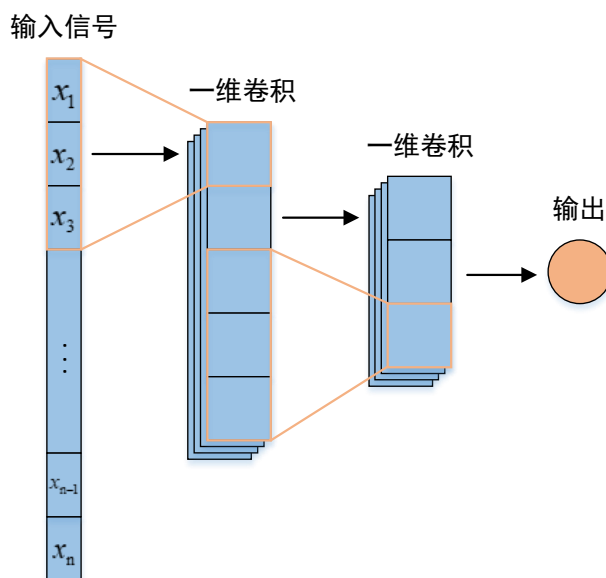


图 2-6 具有两个卷积层的一维 CNN 架构

Fig. 2-6 1D CNN architecture with two convolutional layers

2.5 本文中应用的其他深度学习方法

2.5.1 图卷积神经网络

CNN 的卷积模式与图片的平移不变性相匹配，CNN 也因此图像处理等任务中被广泛应用，但拓扑图网络不具有类似于图片的平移不变性，不能直接利用 CNN 对拓扑图进行卷积。针对 CNN 的上述局限，通常使用邻接矩阵来表示交通节点的连通关系，通过谱图理论实现在拓扑图上的卷积操作，GCN 将传统的卷积操作从结构化数据推广到图，它们能够捕获隐藏在图中的非结构化模式，通过在节点之间交换信息来学习节点表示。

如本章的节 2.3 所述，道路网络可表示为图 $G = (V, E)$ ，其中 V 是一组 N 个交通节点， E 是连接相关节点的一组边，每个节点包含了由传感器测得的交通流时间序列。用邻接矩阵 A 代表图 G ，则 $A \in R^{N \times N}$ ，其中 N 是传感器放置节点的数量；在完成邻接矩阵的构建后，可使用 GCN 对站点之间的空间依赖关系进行建模，其计算公式如式 (2-9)：

$$H^{(l+1)} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} w^{(l)} \right) \quad (2-9)$$



式 (2-9) 中, $\sigma(\cdot)$ 表示激活函数, $w^{(l)}$ 是第 l 层的权重参数; $\tilde{A} = A + I$, 即图网络的邻接矩阵加上自连接矩阵, \tilde{D} 是 \tilde{A} 的度矩阵, 其中 \tilde{D}_{ii} 是对邻接矩阵进行归一化处理的结果且 $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$, $\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$; $H^{(l)} \in R^{N \times D}$ 表示第 l 层图卷积的输入信号, 且 $H^{(0)} = X$ 。

对于谱图卷积, 可将图上的频域卷积定义为信号 $x \in R^N$ 与其卷积核 g_θ 在傅里叶域的乘积, 其中 $g_\theta = \text{diag}(\theta)$, $\theta \in R^N$, 卷积操作如式 (2-10) 所示:

$$g_\theta * x = U g_\theta U^\top x \quad (2-10)$$

式 (2-10) 中 U 为对拉普拉斯矩阵 L 的特征向量进行归一化后的结果, $U^\top x$ 是对 x 的图傅里叶变换, L 的计算公式如式 (2-11):

$$L = I_N - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} = U \Lambda U^\top \quad (2-11)$$

式 (2-11) 中 Λ 为 L 的特征值对角矩阵, g_θ 为 L 的特征值函数评估方程式。但式 (2-11) 的计算复杂度达到了 $O(N^2)$, 且对于大型图, 计算 L 的特征分解也可能会带来较高的复杂度。为了规避这个问题, Hammond 等人提议^[54] 函数 $g_\theta(\Lambda)$ 可以通过切比雪夫 (Chebyshev) 多项式 $T_k(x)$ 的 K 阶展开来很好地近似:

$$g_{\theta'}(\Lambda) \approx \sum_{k=0}^K \theta'_k T_k(\tilde{\Lambda}) \quad (2-12)$$

式 (2-12) 中 $\theta'_k \in R^K$ 是 Chebyshev 系数的参数向量, $\tilde{\Lambda} = \frac{2}{\lambda_{\max}} \Lambda - I_N$, λ_{\max} 表示 L 的最大特征值, 则递归的 Chebyshev 多项式被定义为 $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$ 。

将其代入对信号 x 与滤波器 $g_{\theta'}$ 卷积的定义, 现在有:

$$g_{\theta'} * x \approx \sum_{k=0}^K \theta'_k T_k(\tilde{L}) x \quad (2-13)$$

式 (2-13) 中 $\tilde{L} = \frac{2}{\lambda_{\max}} L - I_N$, $(U \Lambda U^\top)^k = U \Lambda^k U^\top$, 此图卷积是 K 阶局部化的, 因为式 (2-13) 是拉普拉斯算子中的 K 阶多项式, 即它仅取决于中心节点的 K 阶邻域。经过近似后的式 (2-13) 仅有取决于边数量的线性复杂度, 迪弗拉德等人^[34]使用这种 K 阶局部化卷积在图上定义卷积神经网络。

2.5.2 注意力机制

注意力机制提供了通用的方法来建模值集合与其查询目标间的依赖关系, 能够自适应地为集合中的每个值分配一个由查询目标和与其相关联的键值确定的权重, 已在自然语言处理和语音识别等任务中得到广泛应用。SAnD^[55]和 DSANet^[56] 等研究将注意力机制用于多元时间序列预测并取得了良好效果, ASTGCN^[24]和 DGCN^[25]等研究也验证了注意力机制在多元交通数据预测任务中的有效性。

自注意力是注意力机制的一种特殊实现，完全依赖自注意力机制的编码解码模型在一系列时间序列建模任务中取得了卓越的性能^[26]。它通过自注意力机制计算输入序列不同时刻数据的相关表示并进行序列转换。由于每个时间点数据的表示直接由其自身与所有其他时间点的数据计算得到，这导致了一个有效的全局感受野，自注意力机制也因此提供了一种较 RNN 与 CNN 更灵活的机制来模拟交通数据的复杂和动态的长期变化模式。

2.6 本章小结

本章首先介绍了交通流预测的参数与时空特征，对交通流数据的时空特征进行描述并给出了数学表达。然后介绍了本文中用到的 GRU、一维 CNN、Seq2seq、GCN、注意力机制等深度学习方法，为后几章改进的交通流预测模型打下理论基础。



第3章 基于 GCN 与 Seq2seq 结构的交通流预测模型

本章在节 3.1 中给出引言。在节 3.2 中进行模型的构建与分析，首先给出了模型结构，然后介绍了基于时序相关性的 GCN 与基于 GRU 的 Seq2seq 模型在交通流预测中的应用，最后针对数据集离群点较多的特点，将改进的损失函数用于模型训练。在节 3.3 中进行实验分析，介绍了实验数据集与实验参数，展开与基线模型间的对比实验及消融实验，验证本章所提出预测模型的优越性。最后在节 3.4 中给出本章小结。

3.1 引言

随着人工智能技术的发展和相关硬件的升级，深度学习方法被越来越多地用于交通流预测。Dougherty 首先将神经网络应用于交通流预测任务，Polso 和 Sokolov 建立了用于交通预测的深度学习架构^[16]。如今，图卷积网络和循环神经网络逐渐结合起来预测交通流量^[17-18]，但简单结合它们的模型在获取高维交通数据的时空依赖性方面需要改进。此外，许多交通流量预测研究使用均方误差作为损失函数来训练模型，然而当交通数据有很多噪声时，模型对噪声数据的学习会降低模型的性能。

基于此，本章提出了基于图卷积与序列到序列结构的交通流预测模型。首先通过基于皮尔森系数法的图卷积来捕获不同节点之间交通流数据的空间相关性，从而得到影响权重不同的交通流时间序列作为编码器输入；然后交通数据通过编码器编码得到中间向量，并在中间向量中加入了时间注意力机制，防止长序列编码造成的信息丢失；最后中间向量被解码器解码为预测结果。此外，为了减少噪声数据对模型训练的影响，本章将改进的损失函数用于模型训练。总的来说，在本章的模型中，基于 GRU 的 seq2seq 模型用于捕获输入序列之间的时间特征，基于序列相似度的图卷积用于捕获空间特征，还整合了时间特征信息，将改进的损失函数改善了网络的过拟合，进一步提高了预测精度。

实验选取的数据集为深圳龙岗区坂田街道的交通数据集，该数据集误差数据较多且具有明显的时空特征。将数据集根据功能划分为三个部分，一是利用用于模型训练的训练集，并在训练过程中将由改进损失函数计算得到的误差值用于误差反向传播；二是用于进行超参数调节的验证集；三是测试模型性能的测试集。之后将模型的预测误差与交通流预测领域内颇具代表性的 HA、ARIMA、SVR、LSTM、GRU 与 Seq2seq 模型进行对比，证明了本章提出的预测模型具有高于基准模型的预测性能。

3.2 基于 GCN 与 Seq2Seq 模型的交通流预测模型构建

3.2.1 模型结构

本文提出的模型结构如图 3-1 所示，预测流程伪代码如表 3-1 所示，本模型整体上分为图卷积与 Seq2seq 模型两个部分。输入数据先经用于捕获空间特征的 GCN 处理，之后输入编码解码结构，由 GRU 单元组成的 Seq2seq 模块捕获输入数据的时间特征。此外，模型融合时间特征信息，并将改进的损失函数用以模型训练以提升模型的预测性能。

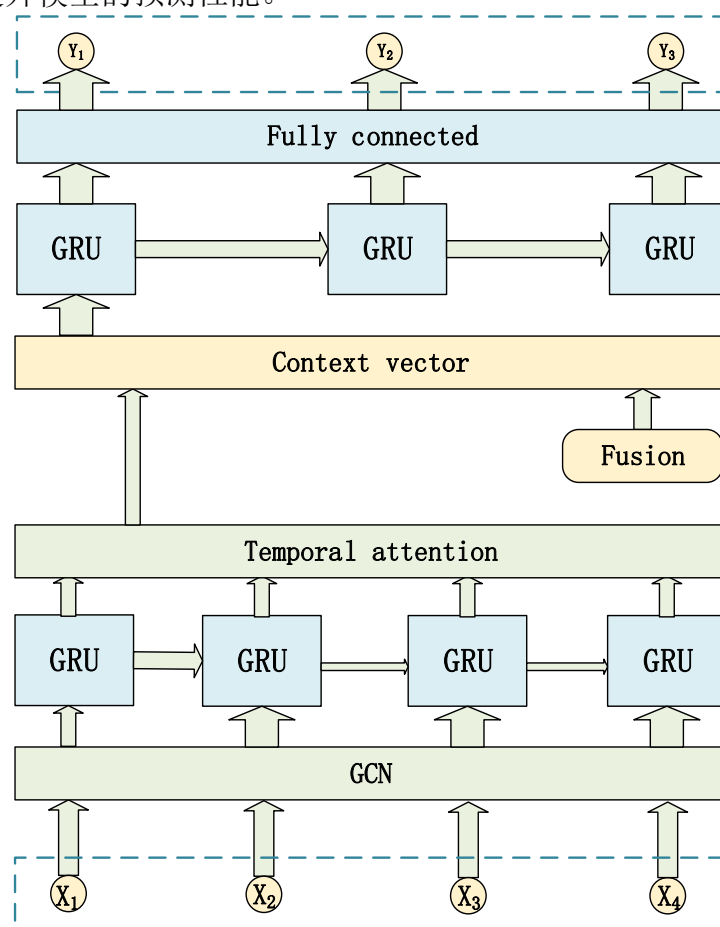


图 3-1 模型结构

Fig. 3-1 The model structure

模型的训练过程如图 3-2 所示，在首轮训练前对模型的参数进行初始化，之后输入数据经模型得到预测输出。基于预测值与实际值，利用节 3.2.4 中改进的损失函数计算模型预测结果的总误差，通过误差反向传播以优化网络参数。当迭代参数达到设定的参数上限时结束训练，并保存网络参数，否则继续迭代以上训练过程。



表 3-1 预测流程伪代码

Table 3-1 Forecast process pseudocode

输入: $X \in R^{N \times L}$, $P \in R^{N \times N}$, $E \in R^{N \times L}$ 。 N 代表交通节点数, L 代表输入数据的步长, X 代表交通流数据, P 代表交通节点间的距离集合, E 代表时间特征嵌入。

输出: $Y \in R^{N \times T}$ 。 T 代表输出数据的步长。

1. 根据式(3-1)计算输入 X 中任意两节点间的节点相关度 $\gamma_{A,B}$;
2. 根据任意 $\gamma_{A,B}$ 计算邻接矩阵 A ;
3. 将输入数据 X 与邻接矩阵 A 代入式(3-3)进行图卷积操作;
4. 将经过图卷积得到的数据 x 通过式(3-4)编码得到 c, s ;
5. 通过式(3-9)解码得到输出 Y 。

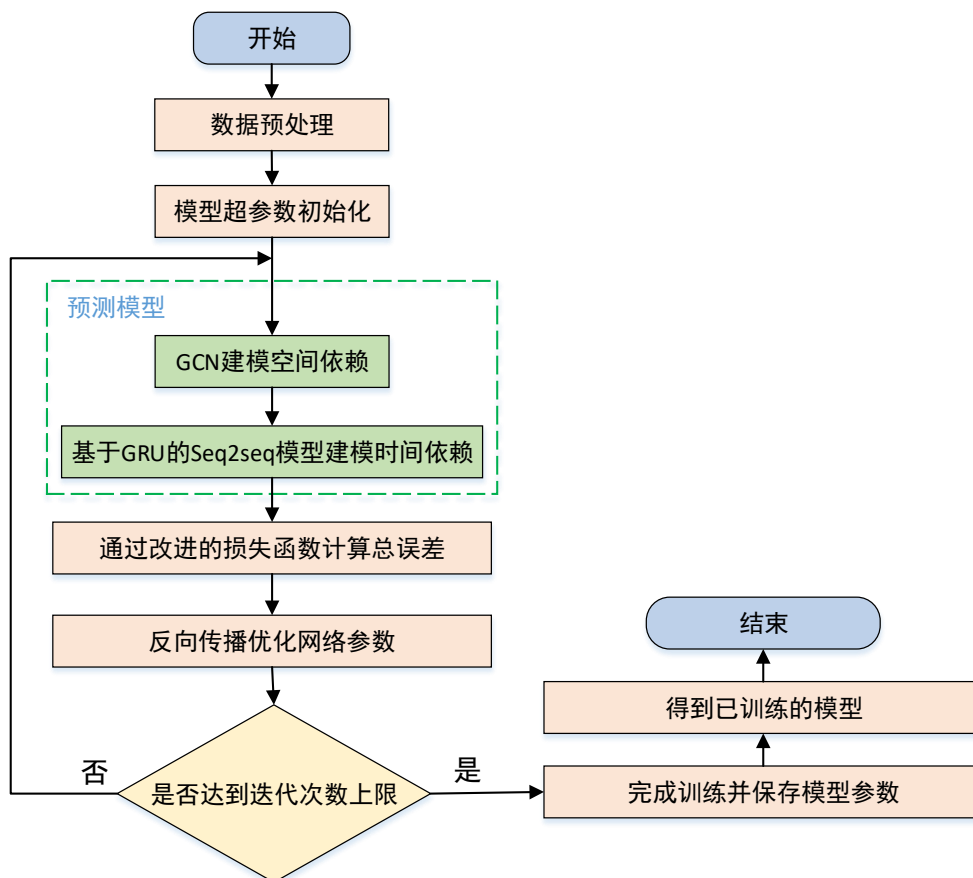


图 3-2 模型训练流程

Fig. 3-2 The model training process

3.2.2 基于空间相关性的 GCN 模块

空间中不同节点的数据不是完全独立的，而是具有一定的空间相关性。为了捕捉其空间相关性，引入皮尔森系数法，用以计算每个节点对预测节点的影响。影响越大，节点的权重就越大。假设目标交叉口为 A 且一个相邻的交叉口为 B ，我们计算两个时间序列的相似度如式（3-1）：

$$\gamma_{A,B} = \left| \frac{\sum_{n=1}^T (x_n^A - \bar{x}^A)(x_n^B - \bar{x}^B)}{P_{A,B} \sqrt{\sum_{n=1}^T (x_n^A - \bar{x}^A)^2} \sqrt{\sum_{n=1}^T (x_n^B - \bar{x}^B)^2}} \right| \quad (3-1)$$

式（3-1）中 x^A 与 x^B 代表输入数据 X 中的两个交通流时间序列， $\gamma_{A,B}$ 代表两者的相似度， $P_{A,B}$ 表示两个节点之间的距离。 $\gamma_{A,B}$ 越接近 1，说明目标路段与相邻路段的相关度越大；当 $\gamma_{A,B}$ 等于 0 时，表示两节点不相关。任意交通节点间的相似度构成相似度集合 γ ，根据 γ 可以得到 GCN 中的邻接矩阵 A 。给定邻接矩阵 A ，GCN 的具体计算公式如式（3-2）：

$$H^{(l+1)} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right) \quad (3-2)$$

式（3-2）中各参数的意义同式（2-9）。本模型采用两层 GCN 进行交通流数据的空间特征提取，计算如式（3-3）：

$$f(X, A) = \text{relu} \left(\hat{A} \text{relu} \left(\hat{A} X W_0 \right) \right) W_1 \quad (3-3)$$

式（3-3）中 $\hat{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$ ， W_0 是从输入层到隐藏层的变换参数， W_1 是从隐藏层到输出层的变换参数。在此利用 relu 作为神经网络的激活函数， W_0 与 W_1 的取值在模型训练过程更新，通过图卷积，Seq2Seq 模型的输入数据中就包含了空间信息。

3.2.3 改进的序列到序列模型

Seq2Seq 模型的编码器可以提取输入信息之间的特定连接，而解码器可以将输入向量映射到可变长度的输出向量。GRU 作为循环神经网络的一种变体，被广泛应用于时序预测，且与 RNN 和 LSTM 相比，GRU 网络结构简单，预测效率更高。本章模型采用 GRU 作为 Seq2seq 模型的基本组成单元。

图 3-1 展示了本章中 Seq2seq 模型的框架结构，其编码器和解码器由 GRU 单元组成。经过 GCN 处理后的输入 x 和编码器初始状态 h_0 ，通过编码器得到中间向量 c 和编码器状态 s ，如式（3-4）：

$$c, s = \text{encoder}(x, h_0) \quad (3-4)$$



编码过程中 GRU 单元的中间状态 h_t 与前一时刻的隐藏层值 h_{t-1} 和当前时刻的输入 x_t 有关, 数学表示如式 (3-5):

$$h_t = f(h_{t-1}, x_t) \quad (3-5)$$

但传统的 seq2seq 结构将输入信息压缩成一个固定长度的中间向量, 使得单个中间向量不能完全表示输入信息, 而且较早输入的信息很可能随着信息的传递而丢失, 影响了时间序列预测的准确率。针对上述问题, 本模型在解码前引入注意力机制, 该机制会根据输入序列中需要关注的部分生成下一个输出, 从而提高模型的预测精度, 该过程步骤如式 (3-6) 至式 (3-8):

$$u_t^o = v_d^T \tanh(w_d * s_{t-1} + w_d' c + b_d) \quad (3-6)$$

$$\gamma_t^o = \frac{\exp(u_t^o)}{\sum_{j=1}^T \exp(u_t^j)} \quad (3-7)$$

$$c_t = \sum_{o=1}^T (\gamma_t^o * c) \quad (3-8)$$

式 (3-6) 至式 (3-8) 中 $w_d \in R^{v \times v}$, $w_d' \in R^{v \times v}$, $b_d \in R^v$ 三个参数可在网络训练过程中学习, T 为中间向量的长度, s_{t-1} 为上一时刻的编码器状态, c_t 是经注意力机制加权后的中间向量。解码过程如式 (3-9) 所示, Y_{t-1} , c_t 和特征融合部分的输入 E 拼接在一起, 用作解码器的输入。将经过时间注意力机制的中间向量作为解码器的初始状态, 然后解码器解码得到模型的输出 Y 。

$$Y_t = \text{decoder}([Y_{t-1}, c_t, E], s) \quad (3-9)$$

具体到解码过程中的每个 GRU 单元, 其输入由前一个单元的隐藏状态和前一个单元的真实交通流数据组成。计算公式如式 (3-10):

$$Y_t = f(h_{t-1}, Y_{t-1}) \quad (3-10)$$

3.2.4 损失函数

由传感器测量的交通流时间序列存在很多显著误差, 这些显著误差数据的存在会增大交通流序列的预测误差。传统的深度学习网络使用均方误差作为损失函数进行梯度更新, 当输入数据包含大量噪声时容易出现过拟合。为了减少模型训练过程中噪声数据对梯度更新的影响, 本模型将改进的损失函数用于模型训练, 损失函数如式 (3-11):

$$\rho(e) = \frac{\frac{1}{m} \sum_{i=1}^L (Y_i - X_i)^2}{2 + \alpha \sqrt{\frac{1}{L} \sum_{i=1}^L (Y_i - X_i)^2}} \quad (3-11)$$

式 (3-11) 中, Y 和 X 分别是模型的预测值与真实值, L 为步长, e 为预测误差值, 且 $e = Y_i - X_i$, α 为损失函数的调整参数, $\alpha \in R$ 。 α 值越大, 损失函数对测量误差导致梯度更新的抑制作用就越大。则损失函数 $\rho(e)$ 的导函数为:

$$\frac{d\rho(e)}{de} = \begin{cases} \frac{4e + \alpha e^2}{(2 + \alpha e)^2}, & e > 0 \\ \frac{4e - \alpha e^2}{(2 - \alpha e)^2}, & e \leq 0 \end{cases} \quad (3-12)$$

在 α 不同的取值下, 该导函数的值随着误差 e 的变化曲线如图 3-3 所示:

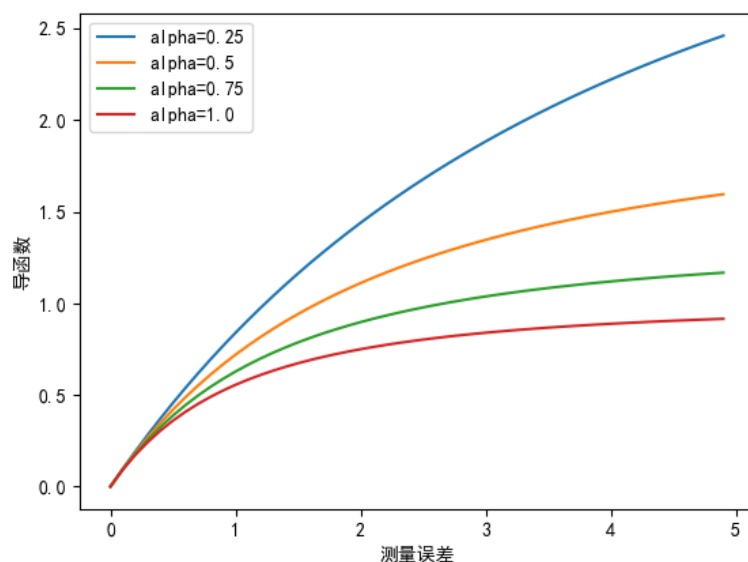


图 3-3 影响函数随测量误差的变化曲线

Fig. 3-3 Variation curve of influence function with measurement error

可知损失函数的导函数取值随着真实数据与误差数据间误差的不断增大而趋近于一个定值, 这表明测量误差无限增大时, 用于梯度反向传播的损失值不会随之快速增大, 从而抑制了测量误差对模型训练的影响。损失函数的二阶导函数如式 (3-13)。

$$\frac{d^2\rho(e)}{de^2} = \begin{cases} \frac{8}{(2 + \alpha e)^3}, & e > 0 \\ \frac{8}{(2 - \alpha e)^3}, & e < 0 \end{cases} \quad (3-13)$$

该损失函数的二阶导函数值恒为正值, 故本损失函数属于凸函数, 这有利于深度学习模型中可学习参数的迭代学习, 有效避免了模型在梯度下降过程中陷入局部极小点。



3.3 基于 GCN 与 Seq2seq 模型的交通流预测方法实验及分析

3.3.1 实验数据

本实验使用的交通数据是通过传感器获取的深圳市龙岗区坂田街道附近路口的交通流量数据，源自 2020 中国高校计算机大赛·华为云大数据挑战赛热身赛，该数据集共包括 6 个路口，有 4 个四叉路口和 2 个三叉路口，数据的时间间隔为 5 分钟，数据集中节点 3 及节点 6 的交通流量图如图 3-4 所示。本实验模型的训练集数据的收集日期为 2019.1.1-2019.1.14，验证集数据收集的收集日期为于 2019.1.14-2019.1.16，测试集的数据收集日期为于 2019.1.17-2019.1.20。

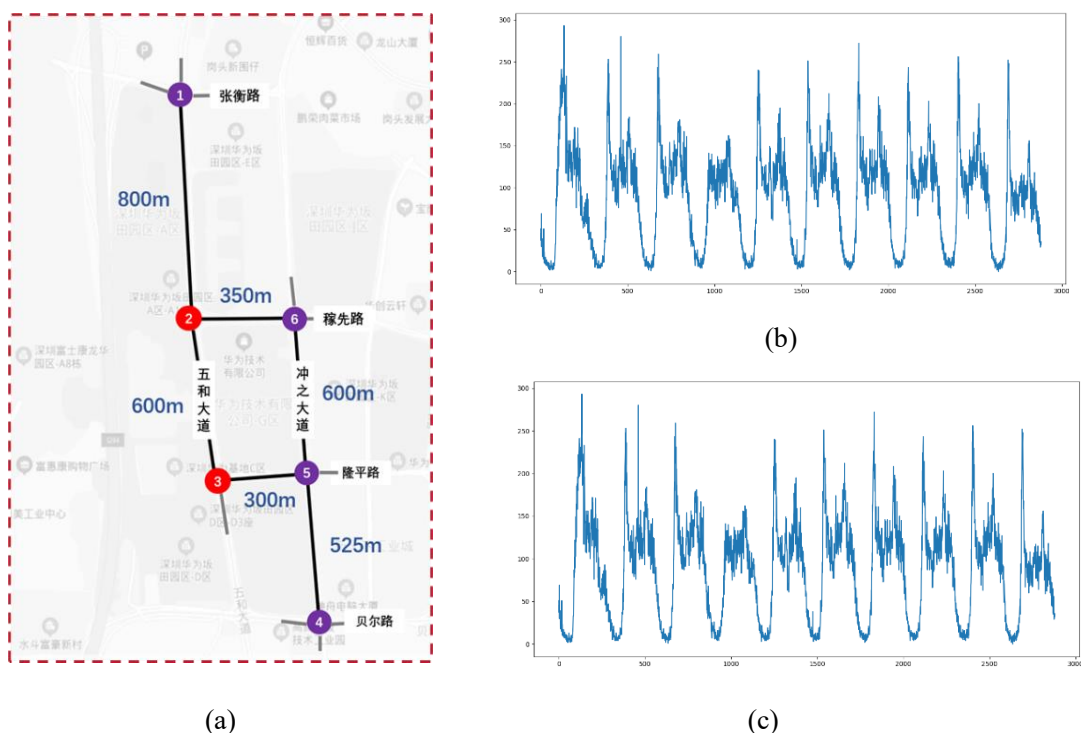


图 3-4 数据集地图与节点 3、节点 6 的交通流量图: (a) 数据集地图, (b) 节点 3 的交通流量图, (c) 节点 6 的交通流量图

Fig. 3-4 Dataset map and traffic flow map of node 3 and node 6: (a) dataset map, (b) traffic flow map of node 3, (c) traffic flow map of node 6

3.3.2 数据预处理

首先，为了加快网络的训练速度，需要对数据进行归一化处理。在本章中采用 Min-Max 归一化方法对原始交通流量数据进行转换，将其映射到最小值为 0、最大值为 1 的区间上。假设原始交通数据中最小值为 x_{min} ，最大值为 x_{max} ，那么归一化公式定义为式 (3-14)。

$$x = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (3-14)$$

其次，为了使模型的输入形式更适合时间序列预测，通过滑动窗口的方法得到一种便于输入模型的数据格式。图 3-5 为滑动窗口法处理数据的示意图， t_0 为开始时刻， t_n 为最后时刻，窗口的步长为 l 并沿时间方向滑动。

此外，交通数据变化受节假日等额外因素的影响。例如，在办公室附近，早上 9 点和晚上 10 点有明显的车流量高峰，周末车流量较少；天气和突发事件等对交通流量也有很大影响，将这些影响因素嵌入到网络中有助于提高预测精度。本实验在数据预处理的过程中提取节假日和早晚高峰信息 E ，并在解码器的输入数据中拼接 E 。

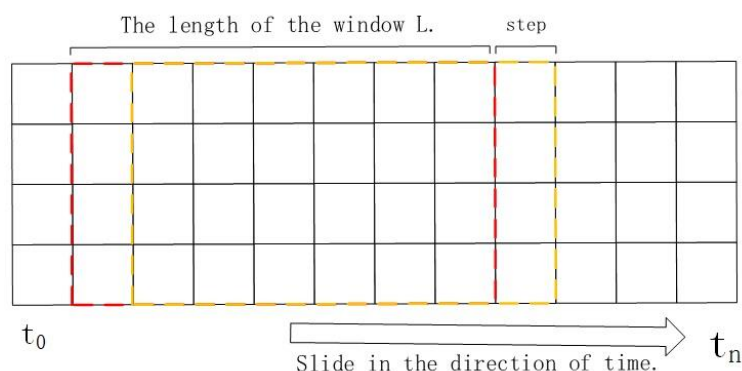


图 3-5 滑动窗口

Fig. 3-5 The sliding window

3.3.3 基准模型与实验参数

本实验将本章提出的基于 GCN 与 Seq2seq 结构的交通流预测模型的预测误差与以下基准模型进行对比：

1) HA^[3]: HA 是一种静态预测，其预测值由给定时间间隔内的交通流量加权得到。

2) ARIMA^[4]: 对单个时间序列的历史数据进行回归，处理相关性高且稳定的数据时较为有效。

3) LSTM^[10]: LSTM 是一种特殊的 RNN，LSTM 与一般 RNN 的区别在于更新中间状态的方式不同，被广泛用于处理时间序列数据。

4) GRU^[11]: GRU 是一种特殊的 LSTM，它的结构比 LSTM 更简单。GRU 也主要用于处理时间序列数据。



5) SVR^[6]: SVR 通过找到一个离训练数据距离之和最近的回归平面, 通过这个回归平面进行未来数据的预测。

6) Seq2seq^[13]: 输入序列经编码器编码, 经模型处理后通过解码器解码得到预测序列。

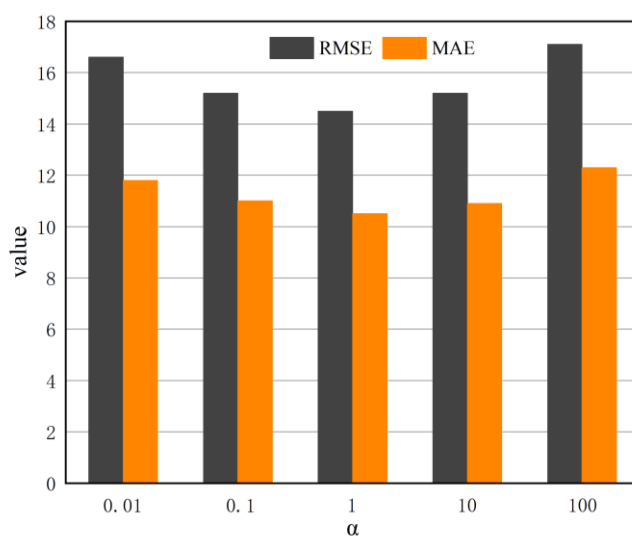
本章使用平均绝对值误差 (Mean Absolute Error, MAE) 和均方根误差 (Root Mean Square Error, RMSE) 来对本章模型与基准模型的预测性能进行评估, 这两者用以反映预测结果相对于真实数据的偏离程度, 因此 MAE 与 RMSE 值越小, 表明预测效果越好。式 (3-15) 和式 (3-16) 中 Y_i 和 X_i 是第 i 个时间步的预测值和真实值, L 是模型预测值的长度。

$$MAE = \frac{1}{L} \sum_{i=1}^L |Y_i - X_i| \quad (3-15)$$

$$RMSE = \sqrt{\frac{1}{L} \sum_{i=1}^L (Y_i - X_i)^2} \quad (3-16)$$

本模型输入具有三个维度, 前两个维度代表批次与交通节点, 最后一个维度代表输入长度, 设置模型的输入与输出步长均为 6, 使用改进的损失函数进行模型训练。为加快模型训练, 设置批训练量为 400, 学习率设为 0.00001。Seq2seq 模型的参数设置与本模型一致。ARIMA 的参数 p , d 和 q 分别设置为 5、1 和 0。SVR 采用线性内核, 其惩罚因子设置为 1.25; LSTM 模型和 GRU 模型都设置两个隐藏层, 节点数设置为 64; 实验模型使用 Pytorch 框架实现, 在具有 1 个 NVIDIA GeForce 2060 GPU (16 GB RAM) 的设备上进行模型训练。

损失函数中的超参数 α 决定了模型训练过程中对测量异常值的抑制能力, α 值过小会使得异常值对模型训练有较大影响, 而过大的 α 值会降低模型的训练速度, 并带来过平滑问题。因此在此比较了改进损失函数的参数 α 不同取值下模型的预测精度, 当 α 为 0.01、0.1、1、10 和 100 时, 预测模型的 RMSE 和 MAE 如图 3-6 所示。可以看出, 相对于 $\alpha=1$ 的情况, 模型的预测精度随着 α 值的增加而降低, 并且模型的预测精度逐渐低于采用标准损失函数训练的模型, 当 α 的值减小时, 模型的预测精度同样会降低, 但其预测效果优于标准损失函数所训练模型的预测效果, 可知 1 对于 α 来说是一个较为合适的取值。

图 3-6 测量误差随 α 取值的变化Fig. 3-6 Variation of measurement error with the value of α

3.3.4 实验结果与分析

将本文提出的模型获得的预测结果的 RMSE 和 MAE 值与基线实验进行比较, 实验结果如表 3-2 所示, 并绘制出 300 个步长预测值与真实值的曲线如图 3-7 所示。从实验结果可以看出, 对于本章的交通流预测任务, HA 的效果最差, 这是因为 HA 在具有明显周期性的时序数据上表现乏力; ARIMA 捕获数据非线性变化的能力较差, 因此也未取得好的预测效果; 基于深度学习的交通流时序预测方法普遍优于基于传统统计方法的交通流时序预测方法。本文基于 GCN 与 Seq2seq 的交通流预测模型综合考虑了交通数据的时间特性与空间特性, 其预测结果的 RMSE 和 MAE 值均低于基准模型, 因此可知本文的模型具有良好的预测性能。

表 3-2 实验结果

Table 3-2 Comparison with base model

模型	RMSE	MAE
HA	29.47	16.77
ARIMA	19.05	14.14
LSTM	16.97	12.48
GRU	16.83	12.64
Seq2seq	17.21	12.55
SVR	16.17	11.38
Our Model	14.56	10.56

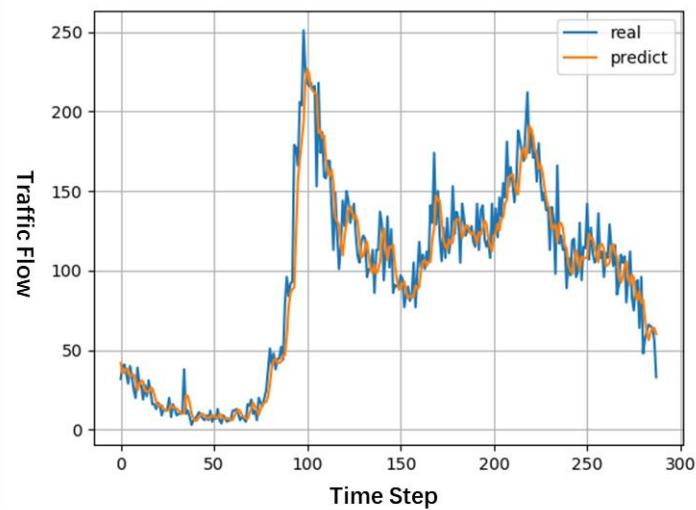


图 3-7 预测结果比较
Fig. 3-7 Forecast comparison

3.3.5 消融实验

本实验中，由空间相似度计算模块计算得到的节点间相似度取绝对值后如表 3-3，为验证模型中各模块的有效性，本章将图卷积、时间特征嵌入两模块移除并与标准模型进行比较。在表 3-4 中，Model-ns 代表去除图卷积的模型，Model-nf 代表去除时间特征嵌入的模型，这两者预测结果的 RMSE 和 MAE 值都比标准模型大，这是因为去除图卷积模块后模型丧失了对空间信息的捕获，变为了对每个序列独立拟合；而移除时间特征嵌入模块后，模型对非周期特征的感知能力变弱。因此，可以得出结论，图卷积模块与时间特征嵌入提高了模型的预测精度。

表 3-3 节点相似性计算结果
Table 3-3 Node similarity calculation result

	五和- 张衡	五和- 稼先	五和- 隆平	冲之- 贝尔	冲之- 隆平	冲之- 稼先
五和-张衡	1.0	0.212	0.146	0.086	0.751	0.324
五和-稼先	0.212	1.0	0.294	0.349	0.493	0.213
五和-隆平	0.146	0.294	1.0	0.452	0.657	0.144
冲之-贝尔	0.086	0.349	0.452	1.0	0.426	0.095
冲之-隆平	0.751	0.493	0.657	0.426	1.0	0.247
冲之-稼先	0.324	0.213	0.144	0.095	0.247	1.0

同样的，用标准损失函数（MSE）代替改进的损失函数进行模型训练，并与标准模型进行比较。从表 3-4 可以看出，Model-nl 代表了具有标准损失函数的模型，它的 RMSE 和 MAE 值都比标准模型大，这是因为测量误差数据对模型训练的影响变大，综上可知改进的损失函数提高了模型的预测精度。

表 3-4 模块评估结果

Table 3-4 Evaluation of different module

模型	RMSE	MAE
Model-ns	18.21	14.37
Model-nl	16.95	12.06
Model-nf	16.62	11.52
Model	14.46	10.56

3.4 本章小结

针对现有模型难以有效深入建模交通流数据时空关联性的问题，本章提出了基于 GCN 与 Seq2Seq 模型的交通流预测方法，通过皮尔森系数法计算交通节点间的空间联系，利用图卷积来捕获空间信息；为了解决传统 Seq2Seq 模型的中间向量不能完全表示输入信息的问题，增加了时间注意机制。此外，交通流序列中存在大量噪声，为了减少它们对模型训练的影响，提出以改进的损失函数来训练模型，这种改进不仅适用于本文提出的模型，也适用于其他具有大量噪声数据的深度学习场景。最后，将本章模型与基线模型进行比较，通过比较它们的 RMSE 和 MAE 数据，得出本文提出的模型具有更好预测性能的结论。



第4章 基于 k-shape 与时空图卷积网络的交通流预测模型

本章在节 4.1 中给出引言。在节 4.2 中进行模型的构建与分析，介绍了基于 k-shape 聚类算法的邻接矩阵构建方法及其在图卷积中的应用，并将图卷积与门控卷积结合用以捕获交通流时间序列的时空特征。在节 4.3 中进行实验分析，介绍了实验数据集与实验参数，展开与基线模型间的对比实验及消融实验，验证了本章所提出预测模型的优越性。节 4.4 中给出本章小结。

4.1 引言

现实世界中的交通路网错综复杂，各交通路段互相影响。本文在绪论中介绍了基于统计方法与基于深度学习的单节点交通流预测方法，这些方法单独对每个节点的交通数据进行回归，虽能较好捕获单个时间序列的时序特征，但若将其用于大型数据集还需考虑交通节点间的空间联系。目前的研究多使用 GCN 和 RNN 分别建模时间依赖和空间依赖来预测交通流量，T-GCN^[16] 就是一个结合 GCN 和 GRU 进行交通流预测且取得了良好的预测效果的例子。

虽然目前的组合式交通流预测模型可以达到较高的预测精度，但仍然存在一些问题。首先，有效模拟交通数据的空间联系仍然是一个挑战。图卷积通常用于捕获空间相关性，传统的 GCN 使用固定的预定义邻接矩阵来表征节点之间的邻接关系，但除了相邻节点之间存在空间相关性外，城市中一些偏远节点也可能存在一定的相关性，即共享一个相似的“时间”模式。ST-MGCN^[60]考虑到了综合利用交通连通性图与时序相似性图，但其采用了基于欧式距离的相似性度量方法，这对于数据波动性较大的交通序列难以达到精准的相似性度量结果。同样的，本文的第三章在建模空间相关性时采用的皮尔森系数法也属于基于欧式距离的相似性度量，可能无法最佳地衡量节点间的语义相关性。其次，RNNs 所存在的梯度消失与梯度下降问题使其难以距离较远的交通数据间的依赖关系，且因其不可并行训练的缺陷，极大的降低了基于 RNNs 的预测模型的计算效率。

基于此，本章构建了基于 k-shape 与时空图卷积的交通流预测模型 (KSTGCN)，k-shape 算法用于获取交通序列间的语义邻接关系，堆叠的时空卷积模块用以捕获交通数据的时空特征。并通过实验证明了本章提出的模型较基准模型有更小的预测误差。

4.2 基于 k-shape 与时空图卷积的交通流预测模型构建

4.2.1 模型框架

本部分构建的预测模型如图 4-1 所示，预测流程伪代码如表 4-1 所示。在原始数据中，多个传感器的测量数据作为节点在一个历史时间段内形成一个图结构，但各个交通流时间序列的时间特性不尽相同。为更好的建模交通数据的空间联系，在本模型中，首先进行复合邻接矩阵的构建，通过 K-shape 聚类算法对所有交通流时间序列进行划分，从而形成不同的时间序列类别，每一个类别中各时间序列都有相似的时间特性，继而根据聚类结果计算出语义邻接信息并和预定义的空间邻接信息进行融合得到复合邻接关系；然后，将这些图结构数据与复合邻接关系输入到时空预测（ASTGCN）模块，以预测未来的交通情况。

表 4-1 预测流程伪代码

Table 4-1 Forecast process pseudocode

输入： $X \in R^{N \times L}$, $A_k \in R^{N \times N}$, $k \in R$, $l_s \in R$, $l_t \in R$. N 代表交通节点数, L 代表输入数据的步长, k 代表聚类中心数, l_s 代表 STGCN 的堆叠层数, l_t 代表 TCN 的层数, X 代表交通流数据, A_k 代表交通节点间的预定义邻接关系。

输出： $Y \in R^{N \times T}$. T 代表输出数据的步长。

1. 利用表 4-1 的 k-shape 算法对输入 X 聚类;
2. 根据式(4-7)计算语义邻接关系 A_s ;
3. 根据式(4-8)计算计算复合邻接关系 A ;
4. 利用式(4-10)调整 A ;
5. For $i = 1$ to l_s do:
6. 利用(4-11)进行图卷积;
7. For $l = 1$ to l_t do:
8. 利用(4-12)进行时间卷积;
9. End for
10. 利用(4-13)进行门控融合;
11. End for
12. 通过输出层输出得到 Y 。

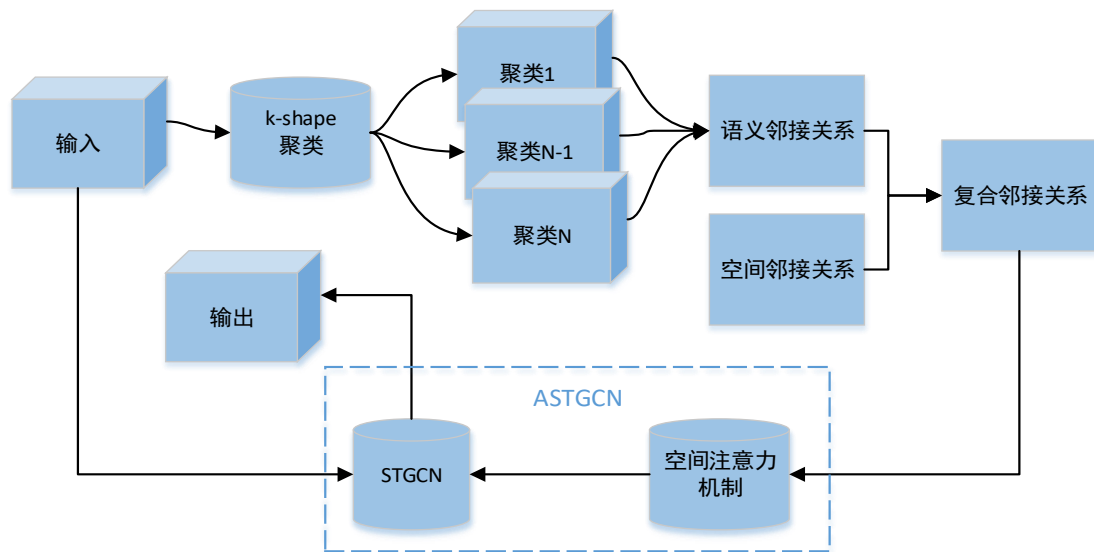


图 4-1 模型框架图

Fig. 4-1 The model structure

4.2.2 复合邻接关系构建

仅依靠现实空间中的节点相邻关系无法深度挖掘交通数据的空间相关性，通过聚类算法对各交通序列进行聚类得到交通节点间的语义邻接关系，进而综合利用语义邻接关系与预定义的空间邻接关系是一个可行的思路。本部分将介绍 K-shape 聚类算法的步骤与算法框架，并进一步介绍利用 K-shape 求取复合邻接矩阵的方法。

K 均值法 (k-means) 通常用于数据聚类^[57]，但对于时间序列聚类、K-means 仅利用序列间的欧几里得距离衡量相关程度，对波动性较大的时间序列的聚类精度不高。与 K-means 聚类方法相比，K-shape^[58]聚类方法在复杂度不变性、遮挡不变性、均匀缩放不变性等方面都有显著提升。本章采用 K-shape 聚类方法对测量节点进行分类，以反映各节点间的语义邻接关系，并根据形状相似性将得到的相似度矩阵与预定义的邻接关系进行融合得到复合邻接矩阵。其中时间序列形状相似度计算和时间序列形状提取是该聚类算法的核心步骤。

1) k-shape 时间序列聚类

捕获基于形状的时间序列相似性需要能够处理幅度和相位失真，DTW 先将时间序列从形状上对齐，然后通过动态规划计算序列上的点到其他序列上各个点的距离之和并取距离和的最小值来衡量序列的相似性，但 DTW 计算复杂度较高，为减小复杂度，K-shape 采用互相关方法计算不同时间序列间的相似度^[58]。互相关通过统计学方法度量两时间序列 $\vec{x} = (x_1, \dots, x_m)$ 与 $\vec{y} = (y_1, \dots, y_m)$ 间的相似性，为实现平移不变性， \vec{y} 不变，一步一步划动 \vec{x} ，并计算每一步 \vec{x} 与 \vec{y} 的内积，如式 (4-1) 所示。

$$\vec{x}_{(s)} = \begin{cases} (\overbrace{0, \dots, 0}^{|s|}, x_1, x_2, \dots, x_{m-s}), & s \geq 0 \\ (x_{1-s}, \dots, x_{m-1}, x_m, \underbrace{0, \dots, 0}_{|s|}), & s < 0 \end{cases} \quad (4-1)$$

定义 \vec{x} 与 \vec{y} 的互相关系数 $CC_w(\vec{x}, \vec{y})$ 如式(4-2)所示,

$$CC_w(\vec{x}, \vec{y}) = R_{w-m}(\vec{x}, \vec{y}), w \in \{1, 2, \dots, 2m-1\} \quad (4-2)$$

式(4-2)中的 $R_{w-m}(\vec{x}, \vec{y})$ 作为内积运算可以用式(4-3)计算:

$$R_k(\vec{x}, \vec{y}) = \begin{cases} \sum_{l=1}^{m-k} x_{l+k} \cdot y_l, & k \geq 0 \\ R_{-k}(\vec{y}, \vec{x}), & k < 0 \end{cases} \quad (4-3)$$

目标是找到 $CC_w(\vec{x}, \vec{y})$ 最大时 w 的值, 然后基于 w 的值, 对 \vec{x} 进行平移以对准 \vec{y} 得到 $\vec{x}_{(s)}$, 其中 $s = w - m$ 。在进行互相关度量时, K-shape 采用基于形状的距离测量方法, 代表相似度大小的 SBD 的计算公式如下:

$$SBD(\vec{x}, \vec{y}) = 1 - W_{\max} \left(\frac{CC_w(\vec{x}, \vec{y})}{\sqrt{R_0(\vec{x}, \vec{y}) \cdot R_0(\vec{y}, \vec{y})}} \right) \quad (4-4)$$

不同的时间序列曲线与坐标轴围成图形的重叠部分越多, 则形状越像、 $W_{\max} \left(\frac{CC_w(\vec{x}, \vec{y})}{\sqrt{R_0(\vec{x}, \vec{y}) \cdot R_0(\vec{y}, \vec{y})}} \right)$ 也越大, 对比所有可能位置的相似度值, 取最大的 $W_{\max} \left(\frac{CC_w(\vec{x}, \vec{y})}{\sqrt{R_0(\vec{x}, \vec{y}) \cdot R_0(\vec{y}, \vec{y})}} \right)$, 然后用 $1 - W_{\max} \left(\frac{CC_w(\vec{x}, \vec{y})}{\sqrt{R_0(\vec{x}, \vec{y}) \cdot R_0(\vec{y}, \vec{y})}} \right)$ 得到 SBD 相似度, 也就是说形状越相似, 距离 SBD 越小。归一化后的 CC_w 值在 $[-1, 1]$ 之间, 因此 SBD 的取值在 $[0, 2]$ 之间。

2) 时间序列形状提取

对于时间序列聚类, 选择代表聚类集合的质心式是必不可少的步骤。计算质心的目标是使得其与聚类集合内其余交通流时间序列的平方和距离最小, 可以用式(4-5)来描述:

$$\mu_k = \operatorname{argmax}_{\mu_k} \sum_{x_i \in P_k} \left(\frac{\max_w CC_w(x_i, \mu_k)}{\sqrt{R_0(x_i, x_i) \cdot R_0(\mu_k, \mu_k)}} \right)^2 \quad (4-5)$$

式(4-5)中, P_k 为经 K-shape 聚类后的单个聚类簇, μ_k 为质心向量, 对质心的调整需要计算聚类内时间序列的最佳偏移量, 在迭代过程中, 上轮迭代过程中的聚类中心可以作为新一轮聚类中心的参考。同时, 省略分母, 用方程 $\mu_k = \mu_k \cdot Q \cdot Q$ 归一化 μ_k . 其中 Q 定义为 $Q = I - \frac{1}{m}O$, I 和 O 分别是单位矩阵和酉矩阵。求和过程用 S 代替, 最终的方程表示如式(4-6):

$$\mu_k = \operatorname{argmax}_{\mu_k} \frac{\mu_k^T \cdot M \cdot \mu_k}{\mu_k^T \cdot \mu_k} \quad (4-6)$$



在式 (4-6) 中, $M = Q^T.S.Q$ 。这种新的基于质心的聚类算法从根本上来说是源自相互相关系数的 k-means 距离度量方法的变体^[58]。

3) 基于 K-shape 的交通流序列聚类算法框架

表 4-2 k-shape 算法伪代码

Table 4-2 k-shape algorithm pseudocode

输入: $X \in R^{N \times L}$, $k \in R$ 。X 为多变量时间序列, N 为交通节点数, L 为经滑动窗口法处理后的输入时间序列片段的长度; k 为聚类中心数。

输出: $I \in R^N$, $C \in R^{k \times L}$ 。I 为 N 个时间序列的聚类结果, C 为聚类中心对于每个时间步的分配结果。

```

1.  $Iter = 0$ ;
2.  $I' = []$ ;
3. While  $I \neq I'$  and  $Iter < 100$  do:
4.    $I' = I$ ;
5.   #分配步骤
6.   For  $j = 1$  to  $k$  do:
7.      $X' = []$ ;
8.     For  $i = 1$  to  $n$  do:
9.       If  $I(i) = j$  then:
10.         $X' = [X'; X(i)]$ ;
11.      End if
12.    End for
13.     $C(j) = \text{ShapeExtraction}(X', C(j))$ ;
14.  End for
15.  #调整步骤
16.  For  $i = 1$  to  $N$  do:
17.     $mindist = \infty$ ;
18.    For  $j = 1$  to  $k$  do:
19.       $[dist, x'] = \text{SBD}(C(j), x(i))$ ;
20.      If  $dist < mindist$  then:
21.         $mindist = dist$ ;
22.         $I(i) = j$ ;
23.      End if
24.    End for
25.  End for
26.   $Iter = Iter + 1$ ;
27. End while

```

时间序列形状相似度计算和时间序列形状提取是 K-shape 时间序列聚类算法的核心^[58]。K-shape 聚类是一种不断迭代调整的分区聚类算法, 在整个迭代过程

中,追求最小化聚类中心与集合内时间序列的距离平方和,K-shape 算法在迭代优化的过程主要执行以下两步:

(1) 分配步骤:将每个交通流时间序列与所有质心进行相似度的比较,从而将每个交通流序列分配到与其最近的聚类集合。

(2) 调整步骤:更新各聚类集合的质心以适应上一轮中集群成员的变化。

分配步骤依赖于时间序列的相似性的计算方法,调整步骤取决于聚类质心的计算方法,表 4-2 展示了 K-shape 算法的整个过程,表中ShapeExtraction代表本小节第二部分的时间序列形状提取过程,在该过程进行聚类质心的分配。

4) 邻接图融合

输入数据经过聚类,得到 n 个不同类别的交通流时间序列 I 。如式(4-7),对于语义邻接矩阵 A_s ,若两时间序列 i, j 属于同一类别,则 $A_{s(i,j)} = 1$,否则 $A_{s(i,j)} = 0$ 。

$$A_{s(i,j)} = \begin{cases} 1, & I_{(i)} = I_{(j)} \\ 0, & \text{otherwise} \end{cases} \quad (4-7)$$

若实验选用的数据集不包含预定义邻接信息,则将空间邻接关系 $A_{k(i,j)}$ 设为零矩阵,即认为任意节点在空间上不相邻,否则利用空间邻接信息计算 $A_{k(i,j)}$ 。如式(4-8),在计算复合邻接关系 $A_{(i,j)}$ 时,对于任意两节点,若它们在 $A_{k(i,j)}$ 与 $A_{s(i,j)}$ 中的任一矩阵中存在邻接关系,则认为两节点邻接,否则不邻接。将空间邻接信息与语义邻接信息进行融合,得到最终的复合邻接关系 $A_{(i,j)}$,之后作为图注意力卷积模块的邻接输入。

$$A_{(i,j)} = \begin{cases} 1, & A_{s(i,j)} = 1 \text{ or } A_{k(i,j)} = 1 \\ 0, & \text{otherwise} \end{cases} \quad (4-8)$$

4.2.3 时空预测

在经过复合邻接矩阵构建后,模型构建了图注意力机制用于建模交通数据的空间联系,构建了门控时间卷积模块用于建模时间联系。如图4-2所示,注意力机制被用于计算邻接强度,并通过以残差网络相连接的堆叠时空图卷积模块捕获交通数据的时空相关性。

1) 图注意力机制

GCN 将传统的卷积操作从结构化数据推广到图,它们能够捕获隐藏在图中的非结构化模式,但传统图卷积中各节点的邻接关系固定,且给各相邻节点赋予相等的权重。为深度建模交通流节点间的空间联系,本模型将注意力机制与图卷积进行结合,实时计算邻接节点间的相关性权重大小。

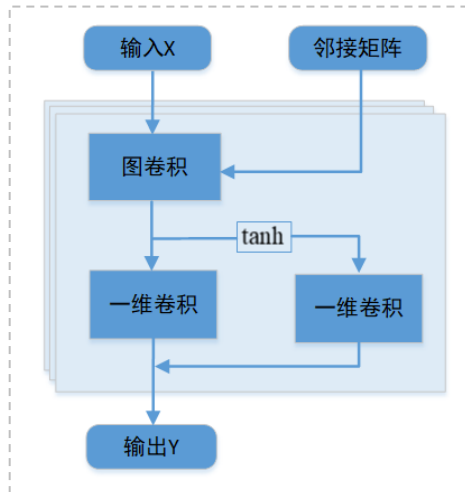


图 4-2 STGCN 模块结构

Fig. 4-2 The STGCN module structure

图 4-3 为注意力机制的图示，注意力机制旨在将查询值 Q ，通过不同的键值 k 映射到值 v ，并对值 v 的进行线性加权求和得到输出 O ，分配给每个值 v 的权重由相应的 k 和 Q 共同确定，每个权重代表了每个键值对之间的关系强度。

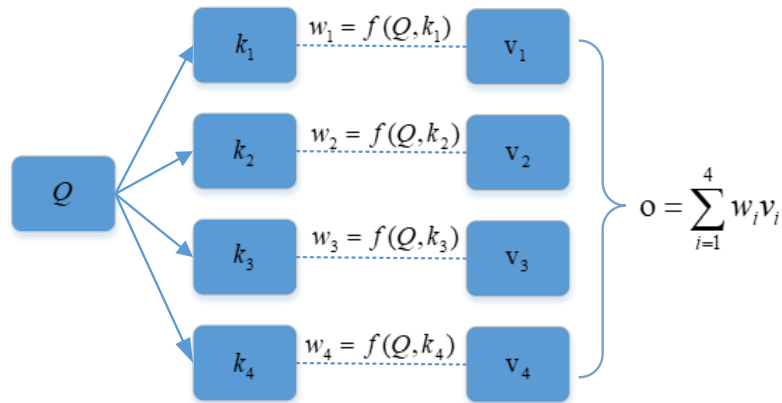


图 4-3 注意力机制

Fig.4-3 Attention mechanism

本模型利用注意力机制动态计算节点之间的空间相关强度，给定节点表示 $X^{N \times d}$ 作为输入，空间相关权重矩阵 S 计算如下：

$$S = \text{softmax} \left(\frac{XX^T}{\sqrt{d}} \right) \in \mathbb{R}^{N \times N} \quad (4-9)$$

$$S_{(i,j)} = \begin{cases} S_{(i,j)}, & S_{(i,j)} > \varepsilon \\ 0, & \text{otherwise} \end{cases} \quad (4-10)$$

式（4-10）中， S 的元素 S_{ij} 表示节点 i 和节点 j 之间的相关强度，值越大表示相关性强，值越小相关性越弱，在得到 S 后，通过设置阈值 ε 过滤掉相关性低的值。在空间相关权重矩阵 S_t 后，利用 S_t 通过逐元素点积运算的方式来调整图卷积中采用的邻接关系矩阵 A ，如式（4-11）所示。

$$x_t^{(l)} = \sigma((A \odot S_t) Z_t^{(l-1)} W^{(l)}) \quad (4-11)$$

式(4-11)中,对于 S_t ,模型捕获的空间动态联系取决于输入,这使得图卷积模块利用由输入 Z 的变化决定的邻接矩阵聚合邻居信息。最后,我们得到第 l 层时空图卷积中图卷积模块的输出 $x_t^{(l)}$ 。

2) 门控时间卷积

模型利用基于扩张因果卷积的时间卷积网络(Temporal Convolution Network, TCN)来捕捉交通节点的时间趋势。与RNN相比,扩张的因果卷积网络能够以非递归方式处理长时序列,这能够并行训练网络并缓解梯度爆炸问题;与一维CNN相比,扩展卷积具有扩张的因果卷积结构,更适合处理时间序列。在扩张的因果卷积中,通过向输入数据填充零来保证对当前时间步长的预测只涉及历史信息,图4-4显示了具有3层扩展卷积的CNN架构,它可以跳过一定的时间步进行卷积,从而增加感受野,达到捕捉长期依赖的目的。在数学上,用 f 对数据 x 进行扩张因果卷积运算可以表示为:

$$x * f(t) = \sum_0^{k-1} f(s) x(t - d \times s) \quad (4-12)$$

式(4-12)中, d 为扩展因子, k 为扩展卷积层数。门控机制被证明在控制时间卷积网络层的信息流方面非常强大,如图4-2,本模型中添加了门控机制来学习交通数据复杂的时间依赖性。给定输入 $x \in R^{N \times L}$,采用式(4-13)所示的门控机制。

$$Y^{(l)} = (\Theta_1 \star x + b_1) \odot \tanh(\Theta_2 \star x + b_2) \quad (4-13)$$

式(4-13)中 θ_1 、 θ_2 、 b_1 和 b_2 是模型参数, \odot 是元素点积, \tanh 是输出的激活函数, $Y^{(l)}$ 为第 l 层时空图卷积中门控时间卷积的输出。

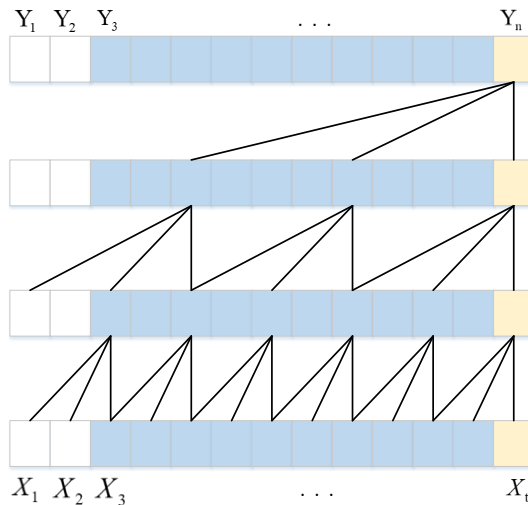


图 4-4 扩展卷积

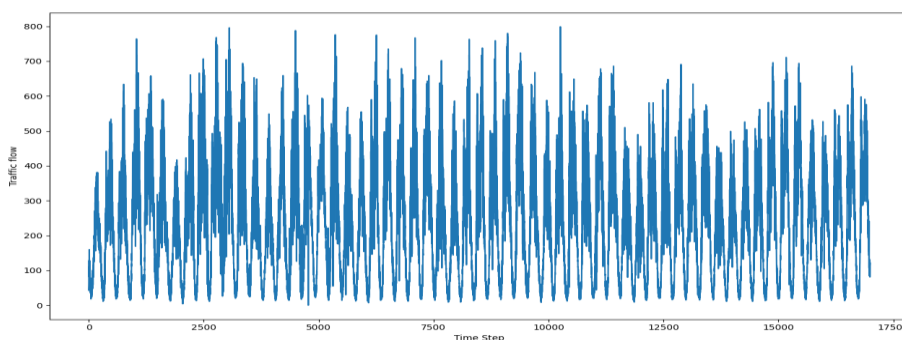
Fig. 4-4 Dilated convolution



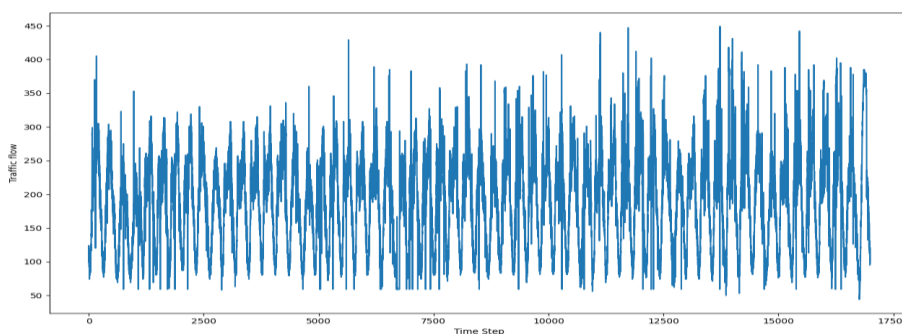
4.3 交通流预测模型 KSTGCN 实验及分析

4.3.1 实验数据

本章在四个数据集上对基于 k-shape 与时空卷积的交通流预测模型进行实验验证，这些数据集来自加利福尼亚州，由加州运输公司在高速公路上通过传感器收集，可在 <https://pems.dot.ca.gov> 下载。PeMS03 有 2018 年 9-12 月记录的 358 个路段的交通数据。PeMS04 有 2018 年 1-2 月记录的 307 个路段的交通数据。PeMS07 有 2017 年 5-9 月记录的 883 个路段的交通数据。PeMS08 有 2016 年 7-8 月收集的 170 个路段的交通数据。四个数据集中数据的采样时间间隔均为 5 分钟。数据集的具体信息如表 4-3 所示，交通流时间序列长度的前 60% 被用于模型训练，后 20% 用于模型测试，其余 20% 作为验证集用以超参数选取，对 PeMS04 数据集与 PeMS08 数据集中单个节点的交通流数据可视化如图 4-5 所示。



(a)



(b)

图 4-5 PeMS04 数据集与 PeMS08 数据集单节点数据可视化：(a) PeMS04 数据集，
(b) PeMS08 数据集

Fig.4-5 Single node data visualization of PeMS04 dataset and PeMS08 dataset: (a) PeMS04 dataset, (b) PeMS08 dataset

表 4-3 PeMS 数据集
Table 4-3 PeMS dataset

数据集	节点数	边数	间隔时间(min)	时间步
PeMS03	358	547	5	26208
PeMS04	307	340	5	16992
PeMS07	883	866	5	28224
PeMS08	170	295	5	17856

4.3.2 数据预处理

交通数据进行无监督方法聚类前通过零均值归一化方法处理数据以提高基于形状相似性聚类的准确性；输入深度学习模型前，基于 Min-Max 归一化方法将交通数据线性变换到[0, 1]范围内，以加快模型训练速度；在归一化后采用滑动窗口法生成模型的输入输出数据，本文输入输出数据的长度均为 12，即利用前一个小时的交通流数据预测后一个小时的交通流。

4.3.3 基准模型与实验参数

在本部分的实验中，将本章模型的预测效果与以下模型进行对比：

1) LSTM^[10]: LSTM 是一种特殊的 RNN，LSTM 与一般 RNN 的区别在于更新中间状态的方式不同，被广泛用于处理时间序列数据。

2) SVR^[6-7]: 从训练数据中选择使各样本到回归平面距离和最小的支持向量，根据这些训练样本的值进行回归分析从而预测未来的交通数据。

3) VAR: 矢量自回归能够捕获多个交通流时间序列之间的成对关系并对未来数据进行预测。

4) STGCN^[36]: 是基于 GNN 和 CNN 的模型，结合了图卷积和一维卷积。

5) ASTGCN^[24]: ASTGCN 设计了空间注意力机制和时间注意力机制并结合了卷积网络来模拟空间和时间的动态相关性。

本章使用 MAE 和 RMSE 来评估预测结果，MAE 与 RMSE 值越小，表明预测效果越好。式 (4-14) 与式 (4-15) 中 Y_i 和 X_i 是预测值和真实值， L 是模型预测值的步长。

$$MAE = \frac{1}{L} \sum_{i=1}^L |Y_i - X_i| \quad (4-14)$$

$$RMSE = \sqrt{\frac{1}{L} \sum_{i=1}^L (Y_i - X_i)^2} \quad (4-15)$$



本章模型中的 STGCN 模块的堆叠层数对预测结果影响较大, 若堆叠层数不足, 模型难以有效建模交通流数据的非线性联系, 堆叠层数过多则会带来过平滑问题。因此, 本章通过实验来确定 STGCN 模块的堆叠层数 l , 实验过程中以测试数据集上的交通流预测误差作为衡量当前参数选择的优劣标准。图 4-6 展示了不同的堆叠层数下模型在 PeMS08 数据集上的预测结果, 由图可知, 模型在 STGCN 模块堆叠层数为 3 时, 模型达到最佳的预测效果, 因此本模型采用堆叠三层 STGCN 模块的结构。

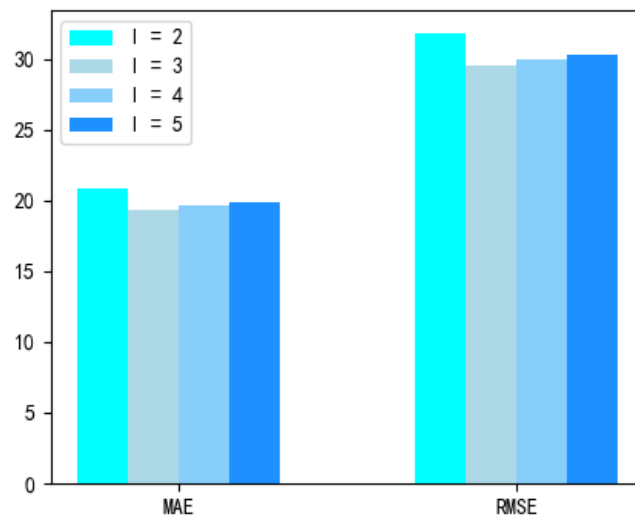


图 4-6 STGCN 堆叠层数选取

Fig. 4-6 STGCN stack layer selection

本实验 STGCN 模块设为 3 层, k-shape 聚类中心设为 7 个, 设置数据的批训练量为 64, 模型学习率为 0.001, 训练模型的迭代次数设为 150。实验模型使用在 Pytorch 框架上实现, 在具有 1 个 NVIDIA GeForce 2060 GPU (16 GB RAM) 的设备上进行模型训练, 具体实验参数如表 4-4。

表 4-4 实验参数

Table 4-4 Experimental parameters

名称	值
聚类中心数	7
STGCN 层数	3
TCN 层数	2
阈值 ε	0.9
学习率	0.001
批数据数目	64
迭代次数	150

4.3.4 实验结果与分析

表 4-5 预测结果的 MAE 值

Table 4-5 The MAE value of the predicted result

模型	MAE			
	PeMS03	PeMS04	PeMS07	PeMS08
VAR	21.08	23.75	101.2	22.32
SVR	22.02	28.66	32.92	23.25
LSTM	20.74	27.02	29.83	22.24
STGCN	18.27	22.16	27.35	18.01
ASTGCN	18.05	22.43	26.14	18.92
KSTGCN	15.79	20.33	22.48	16.51

表 4-6 预测结果的 RMSE 值

Table 4-6 The RMSE value of the predicted result

模型	RMSE			
	PeMS03	PeMS04	PeMS07	PeMS08
VAR	34.75	36.66	155.14	33.83
SVR	35.31	44.60	50.14	36.15
LSTM	33.89	40.86	45.57	33.62
STGCN	30.52	35.03	41.17	27.86
ASTGCN	30.05	34.86	40.12	28.70
KSTGCN	23.92	30.99	35.72	25.27

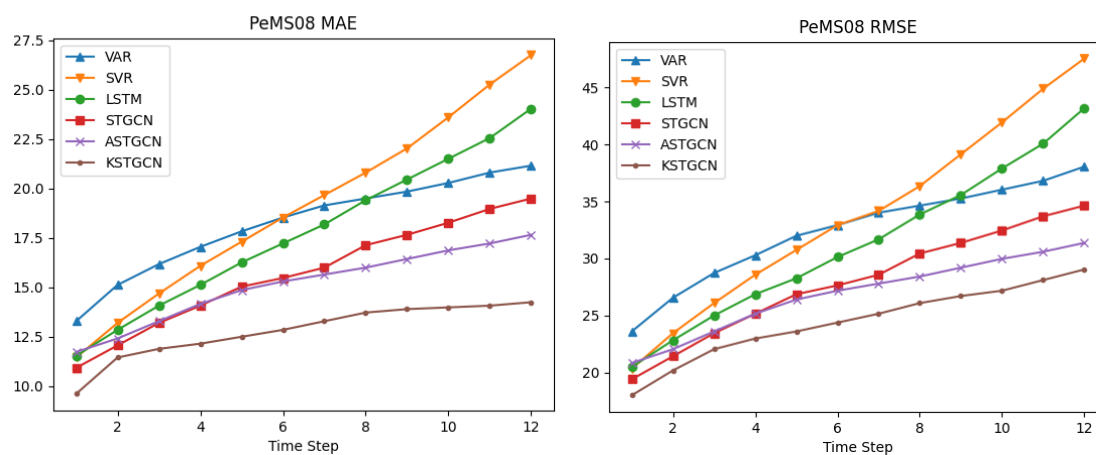


图 4-7 分时间步预测误差图

Fig. 4-7 Prediction error by time steps

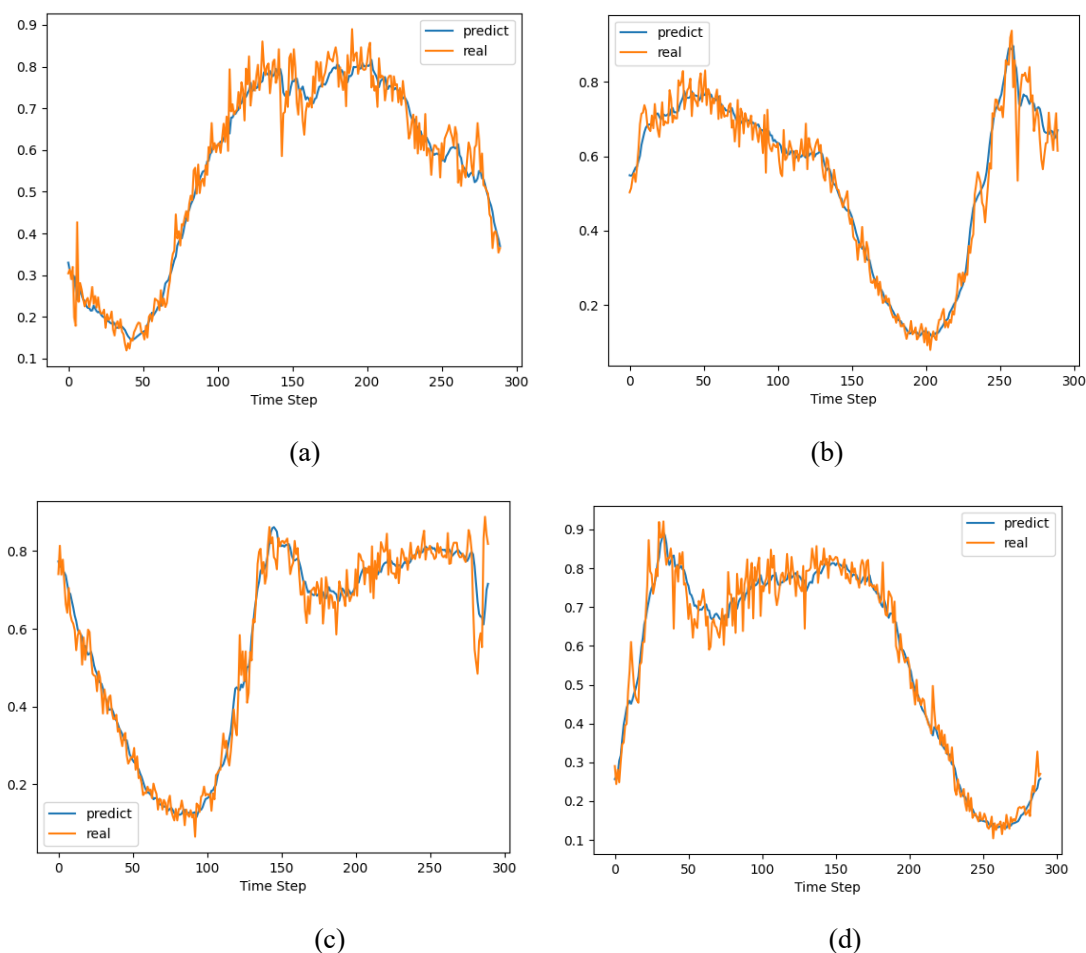


图 4-8 归一化后的预测值与真实值比较:(a)PeMS03 的节点 5, (b)PeMS03 的节点 20, (c)PeMS03 的节点 37, (d)PeMS03 的节点 46

Fig. 4-8 Comparison of normalized predicted value with real value: (a) node 5 of PeMS03, (b) node 20 of PeMS03, (c) node 37 of PeMS03, (d) node 46 of PeMS03

通过本章提出的模型获得的 RMSE 和 MAE 值与基线实验进行比较。从表 4-5、4-6 及图 4-7 可以看出, KSTGCN 预测结果的 RMSE 和 MAE 均优于基准模型, 因此可以得出本文的 KSTGCN 模型具有改良效果。SVR 和 LSTM 只考虑时间序列本身的依赖, 未考虑空间依赖, 所以预测精度最差。尽管 VAR 能够建模多个交通流序列之间的空间和时间相关性, 但受限于其表达能力, 捕捉非线性和动态时空相关性的能力较弱, 预测性能波动较大。STGCN 和 ASTGCN 考虑了时间序列间的空间依赖性, 并改进了捕获时间依赖性的方法, 从而提高了预测精度。本章的 KSTGCN 模型综合考虑了交通数据的空间邻接信息与语义邻接信息, 并设计了堆叠的时空图卷积模块以深入捕获交通流数据间的时空联系, 在四个数据集上都取得了最好的预测结果。最后可视化出归一化后预测值与真实值比较曲线如图 4-8。

4.3.5 消融实验

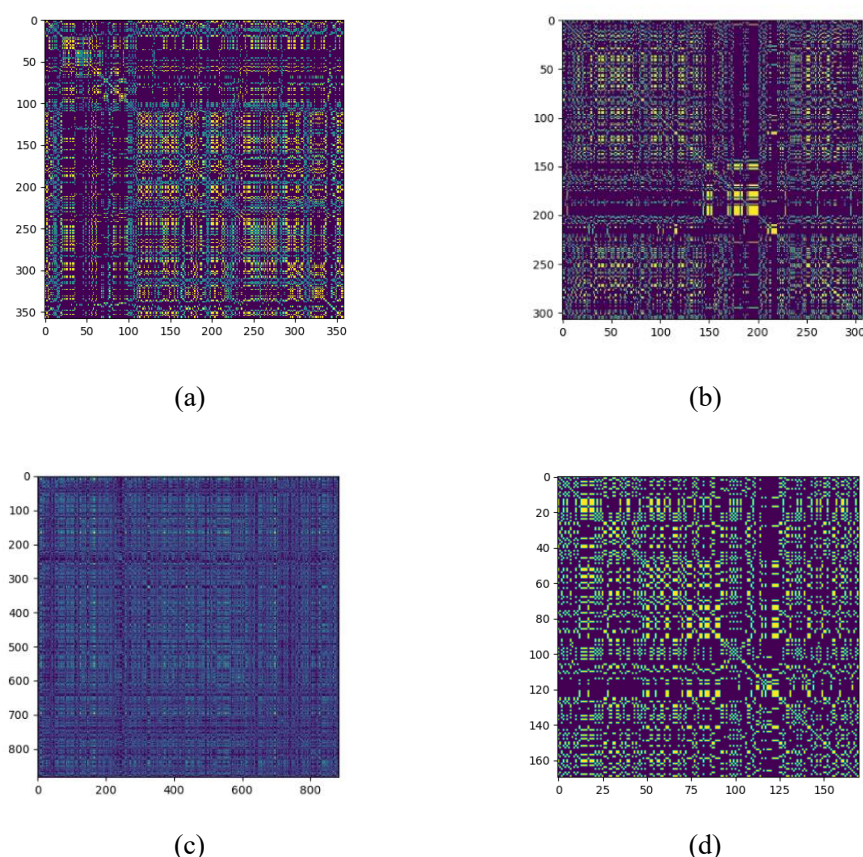


图 4-9 各数据集的语义邻接关系: (a)PeMS03, (b)PeMS04, (c)PeMS07, (d)PeMS08

Fig. 4-9 Semantic adjacency of each dataset:(a)PeMS03, (b) PeMS04,
(c) PeMS07, (d) PeMS08

在本部分中, 首先将使用 k-shape 聚类算法对交通流时间序列聚类得到的语义邻接关系可视化如图 4-9, 其中黄色代表纵横坐标轴上的两时间序列归属为一类, 紫色代表两时间序列不属于一类。之后将语义邻接关系移除, 仅利用空间邻接关系计算邻接矩阵, 在 PeMS08 数据集上开展实验比较预测结果, 如表 4-7 所示, KSTGCN-nk 代表移除语义邻接关系的模型, 在移除语义邻接关系后, 模型预测结果的 MAE 与 RMSE 值均增大, 这是因为仅利用空间邻接关系的模型不能深度利用交通流数据的空间信息, 由此可知综合利用语义邻接关系与空间邻接关系能够提升模型的预测精度。

然后, 将门控时间卷积模块移除并与标准模型 KSTGCN 进行比较。在表 4-7 中, KSTGCN-nc 代表移除门控时间卷积的模型, 它的预测效果显著差于 KSTGCN, 这是因为缺少时间卷积模块的模型无法很好地捕捉时间依赖关系, 可知门控时间卷积模块提高了模型的预测效果。



表 4-7 消融实验

Table 4-7 Ablation experiment

模型	MAE	RMSE
KSTGCN-nk	17.12	26.15
KSTGCN-nc	20.64	30.47
KSTGCN	16.51	25.27

4.4 本章小结

传统仅利用空间邻接关系的图卷积无法充分利用各交通流时间序列间的空间联系；基于循环神经网络的时序建模方法（RNN、LSTM、GRU 等）存在梯度消失与不可并行训练的问题。

针对以上问题，本章提出了基于 k-shape 与时空卷积图网络的交通流预测模型。采用 k-shape 对交通流时间序列进行聚类以得到语义邻接关系，将其与空间邻接关系融合生成复合邻接矩阵。将复合邻接关系与经过预处理的输入一同输入时空图卷积模块，在此模块中，注意力机制被用于计算各节点的影响权重，图卷积用于捕获空间特性，门控扩展卷积用于捕获时间特性，最终得到模型的预测输出。将本文提出的模型与基线模型比较，通过比较它们的 RMSE 和 MAE 数据，得出本章提出的模型具有更好的预测性能的结论。



第 5 章 基于卷积自注意力机制与动态图卷积神经网络的交通流预测模型

本章在节 5.1 中给出引言，在节 5.2 中给出基于卷积自注意力机制与动态图卷积神经网络的交通流预测模型（SDGCN），该模型设计了动态图卷积用于捕获交通流序列的动态空间相关性；设计了卷积自注意力机制用于捕获交通流序列间的局部与长期时间相关性；引入了自回归模块以感知时间序列的非线性变化。在节 5.3 进行实验分析，介绍了实验数据集及数据预处理步骤，展开与基线模型间的对比实验及消融实验，验证了本章所提出预测模型的优越性。

5.1 引言

谷歌公司在 `transformer`^[26]中提出利用自注意力机制取代传统的循环神经网络，并在自然语言处理领域取得了成功的应用，后来又被应用于时序预测、图像处理等领域，将自注意力机制在长时预测中的优势应用于交通流预测是一个可行的研究方向；节点嵌入以可学习的参数对节点进行表示，代表各个节点的参数会随着训练过程，向着损失函数最小的目标学习，以此构建出的邻接矩阵能更好的表征各个交通节点间动态变化的邻接关系。

虽然可以利用聚类算法捕获非直接连通的交通节点间的空间相关性，但基于时间序列形状相似性的方法计算复杂度较高，且捕获交通序列间相关性动态变化特征的能力较弱。通过堆叠扩展卷积层虽能一定程度上增加深层网络的感受野，增强模型的长时预测能力，但 CNN 的内在结构很难比像自注意力机制这样的自回归模型更好，自注意力机制在长期依赖捕获方面表现出突出的优势。

基于此，本章提出了基于卷积自注意力机制与动态图卷积神经网络的交通流预测模型（SDGCN）。首先，我们使用拉普拉斯矩阵学习的方法学习交通数据间的动态相关性，得到动态拉普拉斯矩阵，并通过图卷积捕获节点之间的空间相关性。然后，我们使用扩展卷积捕获交通数据局部时间相关性，使用自注意力机制捕获长期时间相关性。最后，考虑到交通流时间序列的非周期性变化及深层神经网络的过拟合问题，设计了特殊的组合方式引入线性自回归（AR）模块，进一步提高了模型的预测精度。选取加利福尼亚州高速公路 PeMS 数据集开展实验，经对比实验验证，本章提出的模型在四个 PeMS 数据集上均取得了优于基准模型的预测效果。

5.2 基于卷积自注意力机制和动态 GCN 的交通流预测模型构建

5.2.1 模型结构

如图 5-1，为了捕捉交通数据的动态性和复杂性，SDGCN 模型由邻接关系学习模块、时空预测网络与自回归模块（AR）组成。动态邻接矩阵模块用于表示节点之间的动态相关性，包括动态相似矩阵生成、图稀疏和拉普拉斯变换三部分，生成用于图卷积的动态拉普拉斯矩阵；时空预测网络由图卷积及卷积自注意力网络构成编码解码结构，图卷积用于提取空间特征，卷积自注意力从不同层次捕获流量依赖；自回归模块添加了自回归分量以进一步提高模型对输入数据非线性变化的感知。

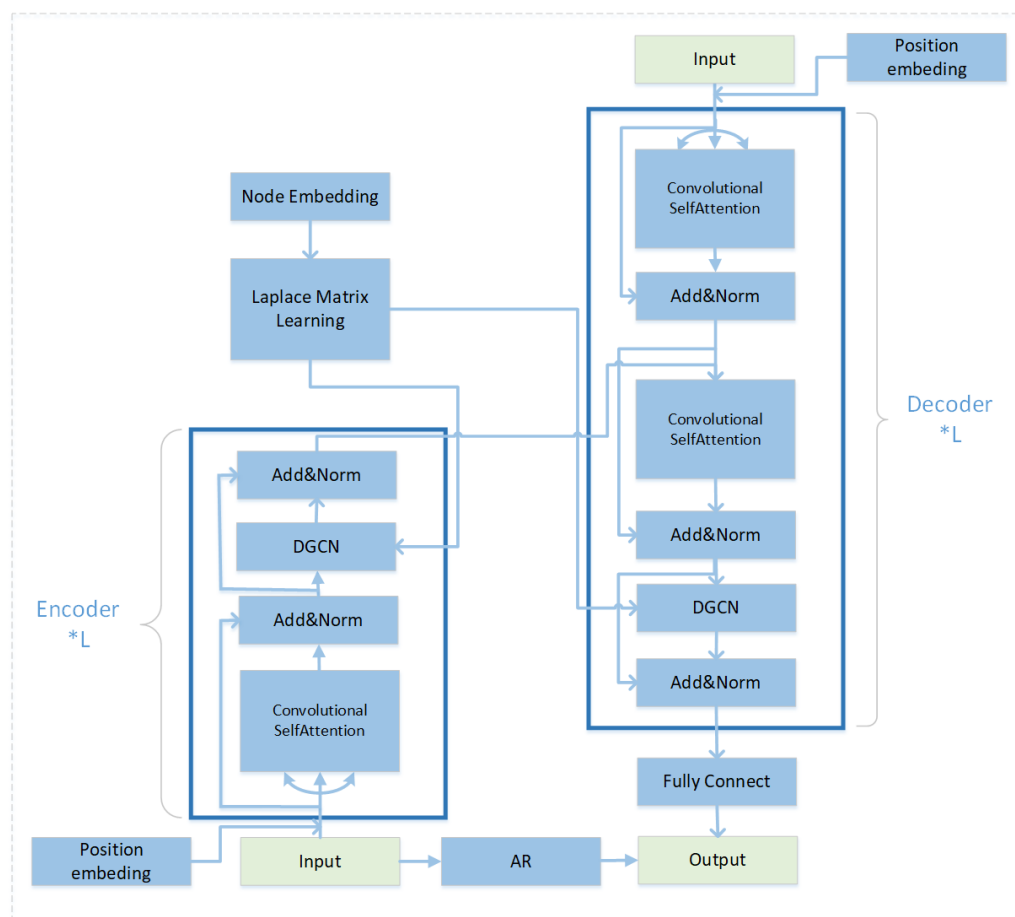


图 5-1 模型结构图

Fig. 5-1 The model structure

表 5-1 为预测流程伪代码，其中 LayerNorm 代表数据正则化，Sublayer 代表该层编码或解码结构，Output 为经残差连接后的输出。将数据输入模型后，分别通过编码解码结构与 AR 模块进行预测。在编码解码结构中，首先对输入数据进行位置编码并将其输入至卷积自注意力机制网络，卷积自注意机制中的



一维扩展卷积对短时时间模式进行学习，多头自注意力机制对长时时间模式进行学习，之后将其输入到动态图卷积模块进行动态空间关系的学习，以上步骤在编码器与解码器堆叠多次并通过残差网络进行连接，在解码器解码得到时空预测网络的输出。将编码解码结构的预测输出与 AR 模块的预测输出以特殊的方式进行融合，得到最终的预测输出。以预测输出与实际值计算得到的损失值将会用于进行神经网络的误差反向传播，在这个过程中位置嵌入向量也得到更新，经过多次迭代，损失值不再降低时结束模型训练。

表 5-1 预测算法伪代码

Table 5-1 Prediction Algorithm Pseudocode

输入： $X \in R^{N \times L}$ ， $l_e \in R$ ， $l_d \in R$ 。 N 代表交通节点数， L 代表输入数据的步长， l_e 代表编码器堆叠层数， l_d 代表解码器层数， X 代表交通流数据。

输出： $Y \in R^{N \times T}$ 。 T 代表输出数据的步长。

1. 利用式 (5-1) 至式 (5-7) 计算拉普拉斯矩阵 L ;
2. #编码阶段
3. 利用式 (5-9) 与 (5-10) 进行位置编码;
4. For $l = 1$ to l_e do:
 5. 利用表 5-2 的算法流程进行时间依赖建模;
 6. 利用式 (5-8) 捕获空间关系得到该层输出 y ;
 7. 通过 $Output = LayerNorm(y + Sublayer(y))$ 进行残差连接;
8. End for
9. #解码阶段
10. 利用式 (5-9) 与 (5-10) 进行位置编码;
11. For $l = 1$ to l_d do:
 12. 掩蔽未来时间步的信息并利用表 5-2 的算法流程进行时间依赖建模;
 13. 利用表 5-2 的算法流程进行时间依赖建模;
 14. 利用式 (5-8) 捕获空间关系得到该层输出 y ;
 15. 通过 $Output = LayerNorm(y + Sublayer(y))$ 进行残差连接;
16. End for
17. 利用式 (5-14) 计算自回归分量 Y_n ;
18. 利用式 (5-15) 融合自回归分量得到输出 Y 。

5.2.2 邻接关系学习

1) 动态邻接矩阵生成

仅依靠现实空间中的相邻关系无法深度挖掘空间相关性，因此引入动态矩阵模块。本模型使用该模块来学习时间序列之间的相似性，然后得到一个在训练过程中随着输入数据动态更新的生成图。动态相似度矩阵的生成过程如下：

$$N_1 = \tanh(\alpha w_1 E_1) \quad (5-1)$$

$$N_2 = \tanh(\alpha w_2 E_2) \quad (5-2)$$

$$A_l = \alpha (N_1 N_2^T - N_2 N_1^T) \quad (5-3)$$

式(5-1)至式(5-3)中， w_1, w_2 是模型参数， E_1, E_2 表示随机初始化的节点嵌入，在训练期间进行学习， α 是用于控制激活函数饱和率的超参数，然后通过动态更新机制得到交通序列间的相似度矩阵 A_l 。

2) 稀疏邻接矩阵

然而，即使某些边缘权重非常小，初始邻接矩阵也是密集的，这些密集的值可以被视为噪声，其在邻接图中的作用很小。因此，提出了一种稀疏手段，通过稀疏化进一步提高图的质量。稀疏邻接矩阵的过程如下：

$$A_{ls} = \text{relu}(\tanh(A_l)) \quad (5-4)$$

$$A_{(i,j)} = \begin{cases} A_{ls(i,j)}, & A_{ls(i,j)} > \varepsilon \\ 0, & \text{otherwise} \end{cases} \quad (5-5)$$

式(5-4)中 relu 和 \tanh 是激活函数；式(5-5)中 $A_{(i,j)}$ 是我们得到的邻接矩阵， ε 是设置的阈值，用来过滤掉较小的值，即相似度低的节点间的空间连接是不考虑的。

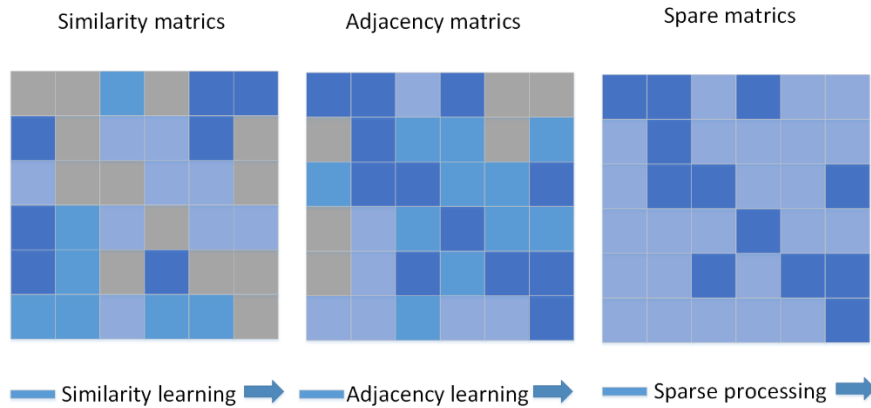


图 5-2 邻接关系处理过程

Fig. 5-2 Adjacency handling process



3) 拉普拉斯变换

得到稀疏邻接矩阵后，将数据从空间域转换到谱域。邻接矩阵被变换为拉普拉斯矩阵，进一步作为图卷积中拉普拉斯矩阵的输入。式（5-6）与式（5-7）是从邻接矩阵 A 到拉普拉斯矩阵 L 的变换过程，其中 I 是秩为 N 的单位矩阵。

$$D_{\text{res}} = \sum_j A_{ij}, i, j = 1, 2, \dots, N \quad (5-6)$$

$$L = I - D_{\text{res}}^{-\frac{1}{2}} A D_{\text{res}}^{-\frac{1}{2}} \quad (5-7)$$

5.2.3 时空预测网络

1) 动态图卷积

交通路网中不同节点的数据不是完全独立的，而是具有一定的空间联系，在本章的模型中同样利用 GCN 处理拓扑空间关系。本模型采用的图卷积类型为 Chebnet，将拉普拉斯矩阵学习模块生成的矩阵 L 作为图卷积的邻接输入进行预测，其公式如下：

$$g * X = \sum_{k=1}^n \theta_k T_k(L) X \quad (5-8)$$

式（5-8）中 X 为输入模型的交通数据， θ_k 是系数， T_k 是 k 阶的 Chebyshev 多项式， L 是拉普拉斯矩阵， $T_k(L)$ 代表第 k 阶 Chebyshev 多项式。

2) 卷积自注意力机制

在进行空间依赖的建模后，本章利用自注意力机制与卷积神经网络结合的方式进行时间依赖建模。自注意力机制将任意两个时间点间的距离减小到1，并从根本上解决了梯度扩散问题，从而为时间序列的长时间依赖捕获带来更好的性能^[26]。自注意机制的计算方式决定其没有输入序列的相对位置信息，需使用位置编码将相对位置信息注入到输入数据中^[26]。对于输入数据 X ，每个偶数时间步的位置嵌入向量定义为：

$$p_e(m, 2n) = \sin\left(m \exp\left[-2n \log(1000)/D\right]\right) \quad (5-9)$$

每个奇数时间步的位置嵌入向量定义为：

$$p_e(m, 2n+1) = \cos\left(m \exp\left[-2n \log(1000)/D\right]\right) \quad (5-10)$$

式（5-9）与式（5-10）中， p_e 是位置编码向量， D 为位置嵌入向量的维度数， m 为相对位置索引， n 为频率，经过位置编码，输入 X 与位置编码向量 P 拼接后的数据便包含了相对位置信息。

点积运算是自注意力机制的基本运算，其表达式如式（5-11）所示：

$$Y_c = \text{Attention}(Q, K, V) \equiv \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (5-11)$$

式(5-11)中 Q 、 K 和 V 是输入 X 的线性变换形式、 d 为比例因子。但将传统用于自然语言处理的自注意力机制直接用于交通流时间序列预测时难以达到理想的预测效果。首先，自注意力机制并不能有效地利用时间序列的序列信息，通常需要使用正余弦函数对其位置进行编码。其次，传统的自注意力机制使用线性变换来获得 Q 、 K 和 V ，这并不适合交通流时间序列预测场景^[59]。例如，如果对变化趋势不同但值接近的时间序列片段应用自注意力机制，可能会给序列中的数据对分配错误的相关强度，进而会得到一个可能不精准的序列表示，从而影响最终的预测性能。

CNN 可以提取时间序列数据的模式信息特征，并且在相邻数据点上取得较好的特征提取效果。基于此，本文在计算 Q 和 K 时，使用大于1的卷积核进行卷积运算，从而替代了原有自注意力机制中的线性变换。其结构如图5-3所示，这使注意力能够集中在相近时间点上，从而匹配更多相关的特征。此外，由于在解码器中预测时间 t 时刻的交通流时，不能使用 t 时刻及其之后的信息，我们在解码器中的卷积操作采用空洞卷积，并在计算自注意力权重时使用遮掩矩阵对未来时间的信息进行屏蔽，这样在预测当前时间的交通信息时，只能关注计算自注意力权重之前的交通信息，从而避免了信息泄露。

多头自注意力机制是自注意力机制的一种常见形式，它使用 h 个不同的线性变换投影 Q 、 K 和 V ，并将不同投影向量的计算得到的注意力结果相拼接得到多头自注意力机制的输出 Y_c ，如式(5-12)和(5-13)。通过多头的卷积自注意力机制，值接近且具有相似局部变化趋势的时间点可以实现正确匹配，其建模流程如表5-2。

$$MultiHead(Q, K, V) = Concat(h_1, h_2, \dots, h_n)W^O \quad (5-12)$$

$$h_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (5-13)$$

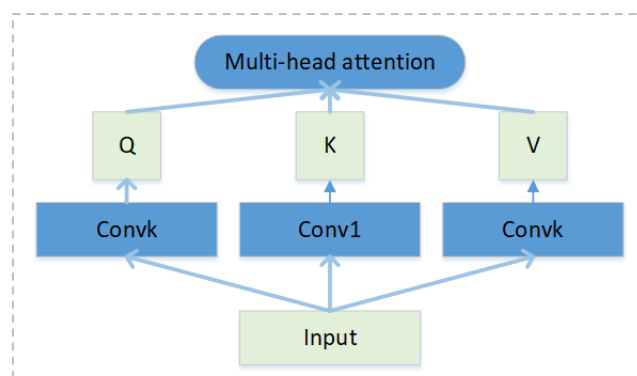


图 5-3 卷积自注意力机制结构

Fig.5-3 Convolutional self-attention mechanism architecture



表 5-2 多头卷积自注意力机制伪代码

Table 5-2 Pseudocode of multi-head convolutional self-attention mechanism

输入: $X \in R^{N \times L}$ 。 N 代表交通节点数, L 代表输入数据的步长, X 代表交通流数据。

输出: $Y_c \in R^{N \times L \times H}$ 。 H 代表多头注意力的数目。

1. 输入数据 X 经过卷积核为 k 的扩展卷积得到查询值 Q ;
2. 输入数据 X 经过卷积核为 1 的扩展卷积得到键值 K ;
3. 输入数据 X 经过卷积核为 k 的扩展卷积得到数值 V ;
4. 对 Q 、 K 、 V 进行多头转换;
5. 根据式(5-9)计算这一层的每个头的注意力值并进行拼接。

5.2.4 自回归模块

虽然深度网络可以很好地拟合时间序列的周期性, 但它对时间序列的非周期性并不敏感, 实际的交通时间序列又往往在短期内有很多非周期性的变化。为了更好地捕捉这些非周期性变化, 引入线性自回归 (AR) 模块如下:

$$Y_n = W \cdot X + b \equiv \left(\sum_{i=1}^L W_{i,z} \times X_{N,i} \right) + b \quad (5-14)$$

式 (5-14) 中 $W \in R^{L \times z}$ 是权重, $b \in R^z$ 是要学习的偏差。一般来说, AR 不适合长期预测。在本模型中, 我们不使用它来预测未来的所有时间步, 而只预测未来的 l 个时间步 ($l < T$)。然后, 将 AR 模块的预测结果 Y_n 与时空预测模块的预测结果 Y_p 通过以下方式融合:

$$Y_s = \gamma Y_p[:, :, l:] + (1 - \gamma) Y_n \quad (5-15)$$

式 (5-15) 中 γ 为融合系数, 模型充分利用了 AR 在短期预测中捕捉非周期性的优势, 从而增加了模型对短期变化的敏感性。

5.3 交通流预测模型 SDGCN 实验及分析

5.3.1 实验数据

本章在四个数据集上对基于卷积自注意力机制与动态图卷积的交通流预测模型进行了实验验证, 这些数据集均来自加利福尼亚州, 由加州运输公司在高速公路上通过传感器收集。四个数据集中数据的采样时间间隔均为 5 分钟, 数据划分方式同第 4 章, 具体信息如表 5-3 所示。

5-3 PeMS 数据集
Table 5-3 PeMS dataset

数据集	节点数	边数	间隔时间(min)	时间步
PeMS03	358	547	5	26208
PeMS04	307	340	5	16992
PeMS07	883	866	5	28224
PeMS08	170	295	5	17856

5.3.2 数据预处理

首先,为了加速模型训练,在将数据输入模型之前,基于 Min-Max 归一化方法将交通数据转换到[0, 1]范围内;在归一化后采用滑动窗口法生成模型的输入输出数据,本文输入输出数据的长度均为 12,即利用前一个小时的交通流数据预测后一个小时的交通流。

5.3.3 基准模型与实验参数

在本部分的实验中,将本章基于卷积自注意力机制与动态图卷积的交通流模型的预测效果与以下模型进行对比:

1) LSTM^[10]: LSTM 是一种特殊的 RNN, LSTM 与一般 RNN 的区别在于中间更新状态的方式不同。主要用于处理时间序列数据。

2) SVR^[6-7]: 通过模型训练,找到离各个数据点距离之和最近的回归平面并利用历史数据开展预测。

3) VAR: 矢量自回归是能够建模多个时间序列之间的依赖关系。

4) DCRNN^[60]: 该模型基于扩散 GCN 来捕获空间相关性,并使用编码器-解码器架构来捕获时间相关性。

5) STGCN^[36]: 基于 GNN 和 CNN 的模型,结合了图卷积和一维卷积。

6) STSGCN^[61]: 该模型使用空间图卷积和一维时间卷积来捕获局部时空依赖性。

7) AGCRN^[45]: 该模型提出了一种自适应图卷积模块,可以自主学习交通道路网络的邻接矩阵。

8) ASTGCN^[24]: 该模型设计了空间注意力和时间注意力机制来模拟空间和时间的动态联系。

9) STFGNN^[21]: 该模型使用时空融合图卷积获提交通网络的时空依赖。



本章使用第2章中的损失函数训练实验模型,使用MAE和RMSE来评估各模型的预测性能,预测结果的MAE与RMSE值越小,表明模型的预测效果越好。式(5-16)与式(5-17)中 Y_i 和 X_i 是预测值和真实值, L 是模型预测值的长度。

$$MAE = \frac{1}{L} \sum_{i=1}^L |Y_i - X_i| \quad (5-16)$$

$$RMSE = \sqrt{\frac{1}{L} \sum_{i=1}^L (Y_i - X_i)^2} \quad (5-17)$$

本章模型的输入有四个维度,第一个维度代表交通数据训练批量大小,第二个维度代表不同交通节点,第三个维度代表时间步长,最后一个维度代表不同的流量特征。编码器与解码器的堆叠层数设为3,卷积自注意力机制中的卷积核大小设为3,多头注意力数目设为8。AR模块的历史时间序列和预测时间序列采用相同的步长 $l=3$;融合系数 γ 取0.6。使用第二章中介绍的损失函数进行模型训练,初始学习率设置为0.0001,训练批次设置为150,具体的实验参数见表5-4。

表 5-4 实验参数

Table 5-4 Experimental parameters

参数名称	值
编码器与解码器层数	3
扩展卷积的卷积核大小	3
自回归融合系数	0.6
自回归历史步长	3
激活函数饱和率控制系数	0.05
邻接矩阵稀疏阈值	0.9
学习率	0.0001
批数据数目	64
迭代次数	150

5.3.4 实验结果与分析

通过本章提出的模型获得的RMSE和MAE值与基线实验进行了比较。表5-5与表5-6展示了利用本模型对未来一个小时交通流的预测结果的误差,从表可以看出,SDGCN模型的RMSE和MAE均优于基准模型。VAR和LSTM只考虑时间序列本身的依赖,未考虑空间依赖,所以预测精度一般。STGCN和ASTGCN考虑了时间序列间的空间依赖性,并改进了捕获时间依赖性的方法,从而提高了预测精度。STFGNN使用堆叠的扩张卷积层来捕获长期依赖关系,

但 CNN 的先天结构几乎无法胜过自注意模型等自回归模型^[32]。AGCRN 使用自适应邻接矩阵建模动态的空间关系并取得了良好的预测效果。本章的 SDGCN 模型综合考虑了交通流时间序列的动态空间依赖、长期依赖和短期依赖、周期性和非周期性规律,因此在四个数据集上都取得了最好的预测结果。图 5-4 展示了 PeMS08 分时间步的预测误差图,可以看出,相较于其他模型,本章的 SDGCN 模型在时间步靠后时预测误差增加的较为平缓,该模型在长时预测上具有突出优势,展现出稳定的长短时预测性能。

表 5-5 预测结果的 MAE 值

Table 5-5 The MAE value of the predicted result

模型	MAE			
	PeMS03	PeMS04	PeMS07	PeMS08
VAR	21.08	23.75	101.2	22.32
SVR	22.02	28.66	32.92	23.25
LSTM	20.74	27.02	29.83	22.24
DCRNN	17.51	21.19	24.64	16.83
STGCN	18.27	22.16	27.35	18.01
STSGCN	17.44	21.15	24.22	17.11
ASTGCN	18.05	22.43	26.14	18.92
AGCRN	15.94	19.81	21.10	15.96
STFGNN	16.81	19.89	22.14	16.73
SDGCN	15.04	19.36	21.89	15.72

表 5-6 预测结果的 RMSE 值

Table 5-6 The RMSE value of the predicted result

模型	RMSE			
	PeMS03	PeMS04	PeMS07	PeMS08
VAR	34.75	36.66	155.14	33.83
SVR	35.31	44.60	50.14	36.15
LSTM	33.89	40.86	45.57	33.62
DCRNN	29.16	33.41	37.85	26.29
STGCN	30.52	35.03	41.17	27.86
STSGCN	29.21	33.65	39.03	26.80
ASTGCN	30.05	34.86	40.12	28.70
AGCRN	28.12	32.27	35.21	25.24
STFGNN	28.72	31.94	35.96	26.31
SDGCN	22.78	29.51	34.02	24.07

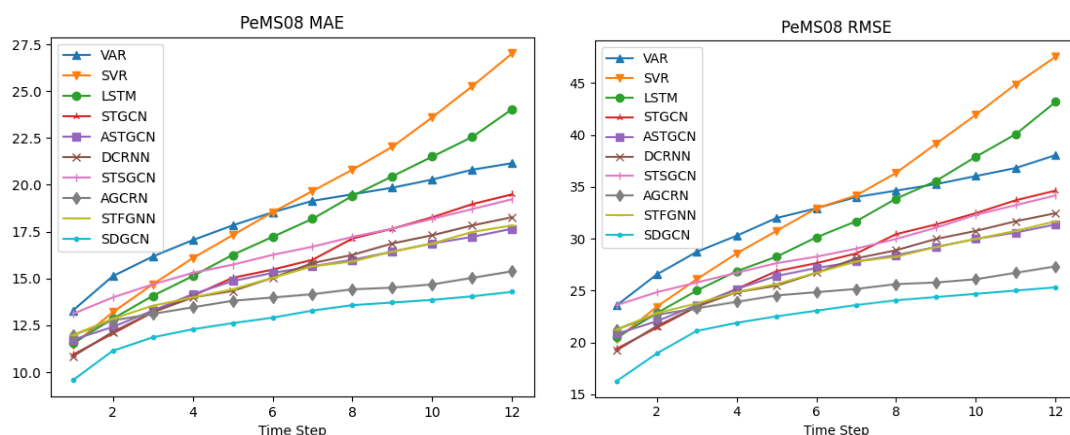


图 5-4 分时间步预测误差图

Fig.5-4 Prediction error plot by time steps

5.3.5 消融实验

将模型中的动态拉普拉斯矩阵替换为静态邻接矩阵得到 SDGCN-nd 模型，并在 PeMS08 数据集上与 SDGCN 进行对比。从表 5-7 可以看出，SDGCN-nd 的 RMSE 和 MAE 值都大于 SDGCN。这是因为预定义的静态邻接矩阵不适用于动态变化的交通网络。因此，可以得出结论，动态拉普拉斯矩阵模块提高了预测精度。我们还绘制了原始邻接矩阵和经过一定时间稀疏后的邻接矩阵，如图 5-5 所示，它们反映了 PeMS08 数据集中 160 个交通节点之间交通流数据变化模式的相似性。

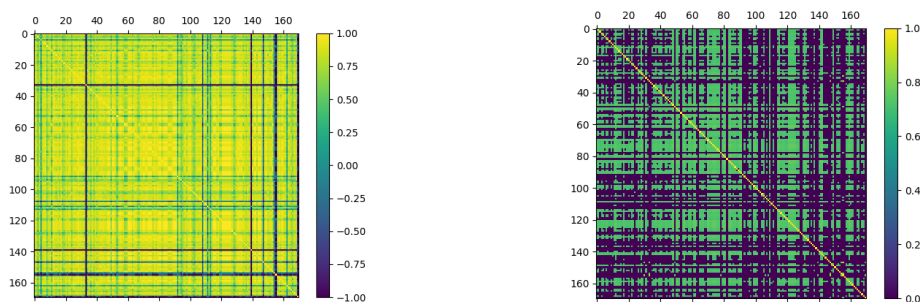


图 5-5 邻接矩阵处理可视化

Fig. 5-5 Adjacency handling process

卷积自注意力机制模块被替换为普通的自注意力机制并与 SDGCN 进行比较。从表 5-7 可以看出，DGTCN-ns 代表的是采用普通自注意力机制的模型，它的预测效果比 DGTCN 差。这是因为缺少自注意力模块的模型无法很好地捕捉长期依赖关系，可知自注意力机制模块提高了预测效果。

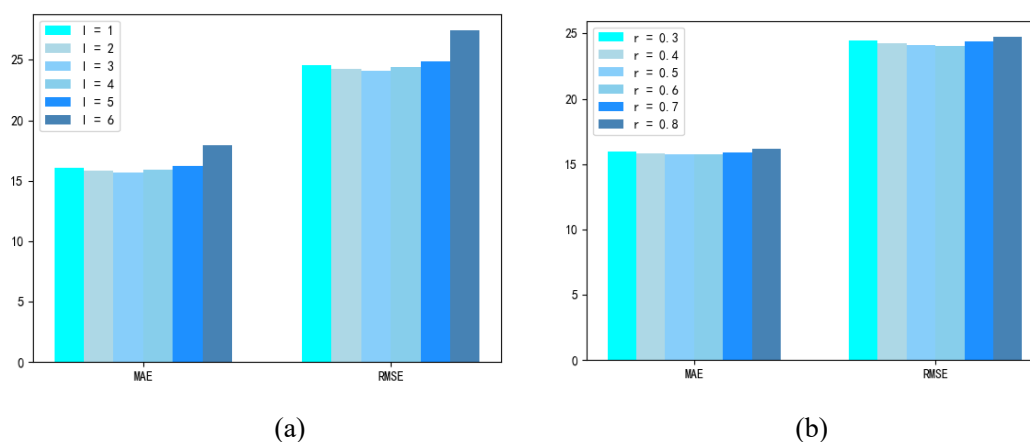
最后将 AR 模块移除与 SDGCN 模型进行了比较,并给出该模块的参数选取过程。从表 5-7 可以看出,SDGCN-na 代表移除 AR 模块的模型,它的 RMSE 和 MAE 值都比标准的 SDGCN 模型大。这是因为去除 AR 模块后,模型无法很好地捕捉到时间序列的非线性变化,故可以得出结论,AR 模块提高了模型的预测精度。

在验证 AR 模块的有效性后,对该模块的超参数选取进行验证。使用 AR 模块捕获时间序列的非周期性时,为便于参数的确定,我们设定输入该模块的历史时间序列 s 和待预测时间序列 z 采用相同的步长 l , 这种情况下,若 l 的取值太小(如 $l < 2$),AR 模块无法感知局部时间序列的变化趋势,进而降低模块的预测精度;而 l 的取值过大,AR 模块误差较大,进而使得输出融合后整个模型的预测误差变大。因此,本章通过实验探索模型预测精度随 l 值的变化,保持其余超参数不变,调整 l 的大小得到的预测精度如图 5-6 (a) 所示,可以看出 l 的最佳值为 3。在确定 l 的取值后,我们进一步验证 AR 模块的融合系数 γ , γ 的值越大,AR 模块的输出对预测结果的影响越大,但预测误差也可能变大。得到的模型在 γ 不同取值下的预测精度如图 5-6 (b) 所示,从图中可以看出,当 γ 设置为 0.6 时,模型的预测效果最好。

表 5-7 模块评估

Table 5-7 Module evaluation

模型	MAE	RMSE
SDGCN-nd	17.12	25.83
SDGCN-ns	16.25	24.66
SDGCN-na	16.20	24.32
SDGCN	15.72	24.07

图 5-6 调参实验: (a) 自回归步长 l 调节 (b) 自回归系数 γ 调节Fig. 5-6 Parameter adjustment experiment: (a) adjustment of autoregressive step size L (b) adjustment of autoregressive coefficient γ



5.4 本章小结

固定的邻接矩阵无法表征交通网络的动态联系，利用可学习的向量对节点进行表示并结合 GCN 可以动态建模空间依赖；RNNs 存在梯度消失的问题，基于 CNN 的时间序列建模方法需要堆叠多层空洞卷积来扩大感受野，基于任意点间点积运算的注意力机制在建模时间序列长期依赖方面具有特殊优势。

在本章中，为深入建模复杂且动态的交通路网关系，提出了基于卷积自注意力机制与动态图卷积的交通流预测模型。首先将动态拉普拉斯矩阵模块用于提取交通流序列之间的动态相关性，得到动态拉普拉斯矩阵后通过图卷积捕获空间依赖关系；然后利用 CNN 和自注意力机制分别提取长时和短时的时间相关性；为进一步提高模型的预测性能，还结合 AR 来提高深度网络对非周期数据的拟合。最后，将本文提出的模型与基线模型进行了比较，通过比较它们的 RMSE 和 MAE 数据，得出本文提出的模型具有更好的预测性能的结论。



第 6 章 总结与展望

本章首先在节 6.1 中阐述本篇文章的主要工作，然后在节 6.2 中对本文可继续研究的内容做出展望。

6.1 总结

交通流数据主要记载了一段时间内，通过道路不同断面的交通实体数。交通流有时间相关性，表现为流量数据随着时间变化呈现出周期性非线性变化的特点，且不同的时间段有着显著不同的特征。同时，城市路网错综复杂，各个路段构成复杂的连接系统，单个路段的交通流受其他路段不同程度的影响，交通流也因此具有较强的空间特性。目前的交通流预测方法虽已取得了良好的预测效果，但在深度建模交通数据的时空依赖性方面仍需深入研究。本文针对交通数据时空联系复杂多变的特点及现有研究中仍存在的问题，提出了三种改进的交通流预测模型并进行实验验证。具体的，本文工作有以下四个方面：

1) 阐述了交通流预测算法在国内外的研究现状和研究历程，给出了交通流预测的必要性、预测原理，详细说明了现有的交通流预测方法的不足之处和需改进的方向，为后续提出新的预测模型打下了良好的基础。

2) 针对现有模型不能深度建模交通数据时空依赖的问题，提出了基于 GCN 与 Seq2Seq 结构的交通流预测模型。通过皮尔森系数法与交通节点的距离建模交通流数据的空间联系，并通过 GCN 捕获相应的空间依赖；在基于 GRU 的 Seq2seq 模型中添加了时间注意力机制以应对长序列编码造成的信息丢失问题，并用其捕获时间依赖。此外，为降低交通流数据中误差数据对模型训练的影响，将改进的损失函数应用于模型训练。最后通过实验验证了模型性能。

3) 针对循环神经网络梯度消失及预定义邻接矩阵无法反映交通路网语义联系的问题，提出了一种基于 K-shape 聚类与时空卷积网络的交通流预测模型。首先通过 k-shape 算法对相似的时间序列进行聚类，得到复合邻接矩阵，并通过注意力机制及图卷积捕获节点之间的空间相关性；进而通过门控的时间卷积网络捕获时间相关性。最后，将本文提出的模型与基线模型进行了比较，通过比较他们的 RMSE 和 MAE 数据，得出本文提出的模型具有更好的预测性能的结论。

4) 针对静态邻接矩阵无法反映交通节点动态联系及传统模型捕获长时依赖能力较弱的问题，本文提出了基于卷积自注意力机制与动态图卷积的交通流预

测模型。将基于时间序列变化的相似性生成的动态拉普拉斯矩阵用于图卷积，以深度挖掘节点之间的空间相关性；通过卷积神经网络捕获局部时间相关性，并通过自注意机制捕获长期相关性。此外为了改善图融合和深度网络的过平滑问题，通过特殊的组合方式增加了 AR 模块，增加了对非周期序列的预测。也就是说，在交通流预测中，我们不仅考虑了多个时间序列之间的动态关系，还考虑了单个时间序列的长期和短期、周期性和非周期性的变化。最后，将本文提出的模型与基线模型进行了比较，通过比较他们的预测误差，得出本文提出的模型具有更好的预测性能的结论。

6.2 展望

本文针对现有的交通流预测算法的不足，提出的新的交通流预测算法在现实世界交通数据集上取得了优良的表现，但仍然有一定的局限性，需要在今后的研究中再接再厉，取得新的突破，具体的可进一步研究的内容和方向如下：

1) 自注意力机制增强了模型的长期预测能力，但增加了模型的复杂度，增加了预测时间，采用一定的稀疏方法对自注意力值的计算进行简化从而降低模型复杂度是一个需要深入研究的方向。

2) 本文采用的 PeMS 交通数据集不包含详细的交通节点位置、天气、交通事故等影响交通流的详细信息，在构建模型时也未将这些影响因素一一考虑进去，故收集更为详细的交通信息并将其考虑进模型构建将有助于提升模型的预测性能。



参考文献

- [1] Zhang J, Zheng Y, Qi D. Deep spatio-temporal residual networks for citywide crowd flows prediction [C]. Thirty-first AAAI Conference on Artificial Intelligence, San Francisco, America, February, 2017, 1655-1661.
- [2] Okutani I, Stephanedes Y. Dynamic prediction of traffic volume through kalman filtering theory [J]. Transportation Research Part B: Methodological, 18(1): 1-11, 1984.
- [3] Duan Y, Lv Y, Liu Y, Wang F. An efficient realization of deep learning for traffic data imputation [J]. Transportation Research Part C: Emerging Technologies, 72: 168-181, 2016.
- [4] Mehdi H, Pooranian Z, Vinueza Naranjo P. Cloud traffic prediction based on fuzzy ARIMA model with low dependence on historical data [J]. Transactions on Emerging Telecommunications Technologies, 33(3): 63-82, 2022.
- [5] 徐翠翠. 基于 ARIMA 的城市交叉路口交通流预测建模 [J]. 电子设计工程, 30(2): 20-23, 2022.
- [6] Philip A M, Ramadurai G, Vanajakshi L. Urban arterial travel time prediction using support vector regression [J]. Transportation in Developing Economies, 4(1): 1-8, 2018.
- [7] Zhang Y, Hou Z. Short term traffic flow prediction based on improved support vector machine [J]. Journal of Applied Science and Engineering, 21(1): 25-32, 2018.
- [8] 周燎, 张武雄, 杨秀梅. 改进贝叶斯动态线性模型在车流量预测中的应用 [J]. 电子设计工程, 26(19): 37-42, 2018.
- [9] Lin G, Lin A, Gu D. Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient [J]. Information Sciences, 608: 517-531, 2022.
- [10] Liu Y, Li D, Wan S. A long short term memory based model for greenhouse climate prediction [J]. International Journal of Intelligent Systems, 37(1): 135-151, 2022.
- [11] 牛哲文, 余泽远, 李波. 基于深度门控循环单元神经网络的短期风功率预测模型 [J]. 电力自动化设备, 38(5): 36-42, 2018.
- [12] 陈海文, 王守相, 王绍敏. 基于门控循环单元网络与模型融合的负荷聚合体预测方法 [J]. 电力系统自动化, 43(1): 93-102, 2019.

- [13] Mao J, Huang H, Lu W. Multi-precision traffic speed predictions via modified sequence to sequence model and spatial dependency evaluation method [J]. *Applied Soft Computing*, 130: 106-117, 2022.
- [14] Lin Z, Feng J, Lu Z. DeepSTN+: Context-aware spatial-temporal neural network for crowd flow prediction in metropolis [C]. *Proceedings of the AAAI Conference On Artificial Intelligence*, Honolulu, America, February, 2019, 1020-1027.
- [15] Yao H, Wu F, Ke J, Tang X, Jia Y, Lu S, Gong P, Ye J, Li Z. Deep multi-view spatial-temporal network for taxi demand prediction [C]. *Proceedings of the AAAI Conference on Artificial Intelligence*, New Orleans, America, February, 2018, 2588-2595.
- [16] Zhao L, Song Y, Zhang C. T-GCN: A temporal graph convolutional network for traffic prediction [J]. *IEEE Transactions on Intelligent Transportation Systems*, 21(9): 3848-3858, 2019.
- [17] Vlahogianni E, Matthew M, John J. Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach [J]. *Transportation Research Part C: Emerging Technologies*, 13(3): 211-234, 2005.
- [18] Ma X, Dai Z, J He, Wang Y, Wang Y. Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction [J]. *Sensors*, 17: 818-835, 2017.
- [19] Youngjoo S, Michael D, Pierre V, Xavier B. Structured sequence modeling with graph convolutional recurrent networks [C]. *International Conference on Neural Information Processing*, Siem Reap, Cambodia, December, 2018, 362-373.
- [20] 于家斌, 尚方方, 王小艺. 基于遗传算法改进的一阶滞后滤波和长短期记忆网络的蓝藻水华预测方法 [J]. *计算机应用*, 335(7): 295-299, 2018.
- [21] Li M, Zhu Z. Spatial-temporal fusion graph neural networks for traffic flow forecasting [C]. *Proceedings of the AAAI Conference on Artificial Intelligence*, Online, May, 2021, 4189-4196.
- [22] Wu Z, Pan S, Long G. Graph wavenet for deep spatial-temporal graph modeling [C]. *Twenty-Eighth International Joint Conference on Artificial Intelligence*, Macao, China, August, 2019, 1907-1913.
- [23] Wu K, Wu J, Feng L. An attention-based CNN-LSTM-BiLSTM model for short-term electric load forecasting in integrated energy system [J]. *International Transactions on Electrical Energy Systems*, 31(1): 263-277, 2021.



- [24] Guo S, Lin Y, Feng N. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting [C]. Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, America, February, 2019, 922-929.
- [25] Guo K, Hu Y, Qian Z. Dynamic graph convolution network for traffic forecasting based on latent network of laplace matrix estimation [J]. IEEE Transactions on Intelligent Transportation Systems, 99:1-10, 2020.
- [26] Vaswani A, Shazeer N, Parmar N. Attention is all you need [C]. Advances in Neural Information Processing Systems, Long Beach, America, 2017, 5998-6008.
- [27] 段大高, 梁少虎, 赵振东. 基于自注意力机制的中文标点符号预测模型 [J]. 计算机工程, 513(5): 297-303, 2020.
- [28] Xu C, Feng J, Zhao P. Long-and short-term self-attention network for sequential recommendation [J]. Neurocomputing, 423: 580-589, 2021.
- [29] 张义杰, 李培峰, 朱巧明. 基于自注意力机制的事件时序关系分类方法 [J]. 计算机科学, 46(8): 244-248, 2019.
- [30] Liu Y, Zheng H, Feng X, Chen Z. Short-term traffic flow prediction with conv-lstm [C]. IEEE 9th International Conference on Wireless Communications and Signal Processing, Nanjing, China, October, 2017, 103-109.
- [31] Zhang J, Zheng Y, Sun J, Qi D. Flow prediction in spatio-temporal networks based on multitask deep learning [J]. IEEE Transactions on Knowledge and Data Engineering, 32(3): 468-478, 2019.
- [32] Guo G, Zhang T. A residual spatio-temporal architecture for travel demand forecasting [J]. Transportation Research Part C: Emerging Technologies, 115: 239-253, 2020.
- [33] Chen C, Li K, Teo S, Zou X, Li K, Zeng Z. Citywide traffic flow prediction based on multiple gated spatio-temporal convolutional neural networks [J]. ACM Transactions on Knowledge Discovery from Data, 14: 1-23, 2020.
- [34] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering [C]. Advances in Neural Information Processing Systems, Barcelona, Spain, December, 2016, 3844-3852.
- [35] Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs [C]. Advances in Neural Information Processing Systems, Long Beach, America, December, 2017, 1024-1034.

- [36] Yu B, Yin H, Zhu Z. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting [C]. 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, July, 2018, 3634-3640.
- [37] Chai D, Wang L, Yang Q. Bike flow prediction with multi-graph convolutional networks [C]. Proceedings of the 26th ACM Sig-spatial International Conference on Advances in Geographic Information Systems, Seattle, America, November, 2018, 397-400.
- [38] Wu Z, Pan S, Long G. Connecting the Dots: Multivariate time series forecasting with graph neural networks [C]. The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Online, August, 2020, 753-763.
- [39] Huang R, Huang C, Yu Y, Liu G, Kong W. LSGCN: Long short-term traffic prediction with graph convolutional networks [C]. Twenty-Ninth International Joint Conference on Artificial Intelligence, Online, January, 2020, 2327-2333.
- [40] Cui Z, Henrickson K, Ke R. Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting [J]. IEEE Transactions on Intelligent Transportation Systems, 21(11): 4883-4894, 2019.
- [41] Geng X, Li Y, Wang L, Zhang L, Yang Q, Ye J, Liu Y. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting [C]. Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, America, February, 2019, 3656-3663.
- [42] Tormene P, Giorgino T, Quaglini S, Stefanelli M. Matching incomplete time series with dynamic time warping: An algorithm and an application to post-stroke rehabilitation [J]. Artificial Intelligence in Medicine, 45(1): 11-34, 2009.
- [43] 陈建国, 刘文霞, 肖永. 一种分布式光伏输出功率短期预测模型建模方法 [P]. CN104732296A, 2015.
- [44] Deng A, Hooi B. Graph neural network-based anomaly detection in multivariate time series [C]. Proceedings of the AAAI Conference on Artificial Intelligence, Online, February, 2021, 4027-4035.
- [45] Bai L, Yao L, Li C. Adaptive graph convolutional recurrent network for traffic forecasting [C]. Advances in Neural Information Processing Systems, Online, December, 2020, 17804-17815.
- [46] 郑日荣, 毛宗源, 罗欣贤. 基于欧氏距离和精英交叉的免疫算法研究[J]. 控制与决策, 20(2): 161-164, 2005.



- [47] Chung J, Gulcehre C, Cho K. Empirical evaluation of gated recurrent neural networks on sequence modeling [J]. ArXiv Preprint, 1412: 3555-3569, 2014.
- [48] Zhang Y, Li D, Wang Y. Abstract text summarization with a convolutional seq2seq model [J]. Applied Sciences, 9(8): 165-179, 2019.
- [49] You L, Han X, He Z. Improved sequence-to-sequence model for short-term vessel trajectory prediction using AIS data streams [J]. Computer Science, 47(9): 175-180, 2020.
- [50] Herbreteau S, Kervrann C. DCT2net: an interpretable shallow CNN for image denoising [J]. IEEE Transactions on Image Processing, 31: 4292-4305, 2022.
- [51] Stahlberg F. Neural machine translation: A review [J]. Journal of Artificial Intelligence Research, 69: 343-418, 2020.
- [52] Gehring J, Auli M, Grangier D, Dauphin Y. A convolutional encoder model for neural machine translation [C]. The 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, Canada, July, 2017, 123-135.
- [53] Bai S, Kolter J, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [J]. ArXiv Preprint, 1803: 1-12, 2018.
- [54] Hammond D, Vandergheynst P, Gribonval R. Wavelets on graphs via spectral graph theory [J]. Applied and Computational Harmonic Analysis, 30(2): 129-150, 2011.
- [55] Song H, Rajan D, Thiagarajan J, Spanias A. Attend and diagnose: Clinical time series analysis using attention models [C]. The Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, American, February, 2018, 4091-4098.
- [56] Huang S, Wang D, Wu X, Tang A. DSANET: Dual self-attention network for multivariate time series forecasting [C]. The 28th ACM International Conference on Information and Knowledge Management, Beijing, China, November, 2019, 2129-2132.
- [57] Zhang Y, Liu Y, Yu Z. Improving aggregated load forecasting using evidence accumulation k-shape clustering [C]. IEEE Power and Energy Society General Meeting, Online, August, 2020, 1147-1152.
- [58] Paparrizos J, Gravano L. K-shape: Efficient and accurate clustering of time series [C]. The 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Australia, May, 2015, 1855-1870.
- [59] Li S, Jin X, Xuan Y. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting [C]. Advances in Neural Information Processing Systems, Vancouver, Canada, December, 2019, 693-702.

- [60] Li Y, Yu R, Shahabi C, Liu Y. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting [J]. ArXiv Preprint, 1707: 1926-1941, 2017.
- [61] Song C, Lin Y, Guo S, Wan H. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting [C]. The AAAI Conference on Artificial Intelligence, New York, America, February, 2020, 914-921.