

文章编号:1671-1637(2024)02-0050-15

基于深度学习的短期交通流预测方法综述

崔建勋¹, 要 甲², 赵泊媛¹

(1. 哈尔滨工业大学 交通科学与工程学院, 黑龙江 哈尔滨 150090;

2. 大连理工大学 经济管理学院, 辽宁 大连 116024)

摘 要:针对基于深度学习的短期交通流预测问题,揭示了时空相关性建模本质,分析了建模过程中涉及的多尺度时空特性、异质性、动态性、非线性等特点,明确了基于深度学习进行短期交通流预测的核心挑战,阐述了短期交通流预测涉及的外部信息整合、多步预测与单步预测以及单体预测与集成预测等相关问题;按照网格化和拓扑化 2 种交通流数据组织方式,分别综述了当前最新的基于深度学习的短期交通流预测研究方向。研究结果表明:针对网格化交通流数据,当前研究主要包含了基于 2D 图像卷积神经网络、基于 2D 图像卷积神经网络与循环神经网络相结合、基于 3D 图像卷积神经网络 3 种预测建模方法;针对拓扑化交通流数据,当前研究主要包含了基于 1D 因果图像卷积与卷积图神经网络相结合、基于循环神经网络与卷积图神经网络相结合、基于自注意力与卷积图神经网络相结合、基于卷积图神经网络的时空同步学习 4 种预测建模方法;总体上,基于深度学习方法进行短期交通流预测相较于采用时间序列和经典机器学习方法获得了预测准确性上的极大提升;未来,针对物理理论、知识图谱与深度学习相结合,构建多时空数据挖掘大模型以及轻量化、可解释性、模型结构自动化搜索等维度的相关探索将成为重要研究方向。

关键词:智能交通;短期交通流预测;深度学习;卷积神经网络;图神经网络;循环神经网络

中图分类号:U491.1 **文献标志码:**A **DOI:**10.19818/j.cnki.1671-1637.2024.02.003

Review on short-term traffic flow prediction methods based on deep learning

CUI Jian-xun¹, YAO Jia², ZHAO Bo-yuan¹

(1. School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150090,

Heilongjiang, China; 2. School of Economics and Management, Dalian University of Technology,

Dalian 116024, Liaoning, China)

Abstract: For the short-term traffic flow prediction problem based on deep learning, the essence of spatiotemporal correlation modeling was revealed, the multi-scale spatiotemporal characteristics, heterogeneity, dynamics, nonlinearity, and other characteristics involved in the modeling process were analyzed, the core challenges were clarified, and the external information integration, multi-step prediction and single-step prediction, as well as individual prediction and integrated prediction were elaborated. The latest research directions were reviewed with two

收稿日期:2023-10-13

基金项目:国家自然科学基金项目(72371048);黑龙江省自然科学基金项目(LH2021E074)

作者简介:崔建勋(1982-),男,黑龙江绥化人,哈尔滨工业大学副教授,从事智能交通与自动驾驶研究。

引用格式:崔建勋,要 甲,赵泊媛.基于深度学习的短期交通流预测方法综述[J].交通运输工程学报,2024,24(2):50-64.

Citation:CUI Jian-xun, YAO Jia, ZHAO Bo-yuan. Review on short-term traffic flow prediction methods based on deep learning[J]. Journal of Traffic and Transportation Engineering, 2024, 24(2): 50-64.

organization methods of traffic flow data: grid and topology. Research results indicate that for gridded traffic flow data, current research mainly includes three prediction modeling methods: 2D image convolutional neural network, 2D image convolutional neural network combined with recurrent neural network, and 3D image convolutional neural network. For topological traffic flow data, current research mainly includes four prediction modeling methods: 1D causal image convolution combined with convolutional graph neural network, recurrent neural network combined with convolutional graph neural network, self-attention combined with convolutional graph neural network, and spatiotemporal synchronous learning of convolutional graph neural network. Overall, the short-term traffic flow prediction based on deep learning methods significantly improves the prediction accuracy compared to time series method and classical machine learning method. In the future, the combination of physics theory, knowledge graphs, and deep learning, the construction of large-scale models for multi-temporal and spatial data mining, as well as the lightweight, interpretability, and automated model structure search, will become important research directions. 1 tab, 8 figs, 81 refs.

Key words: intelligent transportation; short-term traffic flow prediction; deep learning; convolution neural network; graph neural network; recurrent neural network

Author resume: CUI Jian-xun(1982-), male, associate professor, PhD, cuijianxun@hit.edu.cn.

Foundation items: National Natural Science Foundation of China (72371048); Natural Science Foundation of Heilongjiang Province (LH2021E074)

0 引言

短期交通流预测是基于对历史交通流的观测数据以及一定的相应外部信息(天气、事件等)对未来短时期内的交通流进行预测的研究,包括了速度、流量、密度、拥堵情况等交通流指标的短期预测,是许多智能交通应用(如区域交通协调控制、路径诱导、出行路况信息提供、应急交通管控策略制定等)的基础核心科学问题。在道路交通领域,长期以来一直受到广泛关注和深入研究,具有近60年以上的研究历史^[1]。相较于交通规划问题中长期交通需求预测中,时间跨度往往为小时、天、月,甚至年,短期交通流预测更加有针对性地揭示交通流在微观尺度上(一般时间跨度不超过30 min)的变化情况。通常而言,交通流由于经常受到特殊事件、天气、驾驶人随机行为等因素的干扰,在短期内往往波动较大,呈现出相对于长期更强的非线性和复杂性,在预测难度上显著高于长期预测。

交通流预测理论发展的需求源于智能交通的应用,其发展的动力则来源于预测方法论的进步。总体而言,交通流预测大致经历了3代方法论的渐进式发展历程:(1)时间序列类模型(差分自回归移动平均(Auto-Regressive Integrated Moving Average, ARIMA)模型^[2]、广义自回归条件异方差(Generalized

Autoregressive Conditional Heteroskedasticity, GARCH)模型^[3]等)、滤波类模型^[4-5]以及二者的结合^[6];(2)传统机器学习类模型(K近邻^[7]、支持向量机(Support Vector Machine, SVM)^[8]、浅层神经网络^[9]、贝叶斯网络^[10]等);(3)深度学习类模型(卷积神经网络^[11]、循环神经网络^[12]、自注意力^[13]、图神经网络^[14]等)。这一发展历程,反映出研究人员对于短期交通预测问题认知由浅入深(从线性、平稳的时间序列预测到非线性、动态、异质、时空均相关的复杂序列预测)、数据样本逐步增加(从早期的定点交通观测到现今移动轨迹所覆盖的市域全局的观测)和预测模型数据建模能力不断增强(从起源于统计领域,基于线性、平稳假设的ARIMA,发展到浅层机器学习领域,支持非线性、高维变换的支持向量机、神经网络,再到可以具有巨大权重参数规模的深度神经网络)3条清晰的主线。

早期受限于对问题的认知深度、样本数据的规模以及预测方法论的建模能力的3重限制,最开始的短期交通流预测是从时间序列类模型开始的,包括自回归模型、滑动平均模型、自回归滑动平均(Auto-Regressive Moving Average, ARMA)模型、历史平均模型等。后期,随着人们认知的加深,发现交通流序列的不平稳性、非线性等特点,陆续又开始采用ARIMA、GARCH、卡尔曼滤波等进行预

测。这些模型总体上被归为“模型驱动”的预测研究,通常都有以下的不足,使得其预测性能效果不佳:很强的条件假设(时间序列的平稳性等),不适用于复杂动态的交通流预测;对每个观测的位置单元(某个区域、某个路段、某个站点等)进行单独的预测,缺乏空间相关性的考量。

20 世纪 80—90 年代机器学习理论得到了长足的发展和大规模的商业化应用,尤其是支持向量机在机器学习领域取得了辉煌的成就,一度占据了机器学习的王者地位。因此,相关理论,如 SVM、支持向量机回归(Support Vector Regression, SVR)等也被应用于道路交通预测问题的研究,此外,部分研究还采用了贝叶斯方法、K 近邻、浅层神经网络等进行相应的道路交通预测研究。这些模型总体上被归结为“数据驱动”的预测研究。尽管,传统的机器学习方法拥有稳固的理论基础,能够帮助我们有效理解预测的过程,但是在处理复杂、高度非线性交通数据时仍旧表现出预测性能不佳,并且需要非常细致的手动特征工程。

2016 年,深度学习掀起了新一波的 AI 浪潮,短期交通预测领域采用深度学习开展的研究也开始大量的增加。总体上而言,基于深度学习的短期交通预测模型由 2 个部分组成:时空特征学习器和预测器,前者负责从历史的交通流序列中提取高阶的时空特征,后者负责以提取到的高阶时空特征为输入,对未来交通流做出预测。预测所针对的数据类型一般表达为基于网格的和基于拓扑的,前者一般将观测区域进行等大小的网格划分,每个网格单元都具有相应的观测交通流特征,这种表达方式可以直接类比计算机视觉领域的图像数据结构,从而便于后期的卷积操作等;后者则以拓扑结构作为基础来组织交通流观测数据,好处是可以利用路网天然的连接关系捕获不同观测单元之间的空间相关性,但由于拓扑数据是非欧数据,需要用图卷积神经网络替代卷积神经网络。在对未来做出预测时,除了最基本的历史交通流序列的输入,有时也会引入关键性的外部信息(天气、节假日等)。预测的时长可以是当前时段的下一个时段(单步预测),也可以是未来的几个时段(多步预测)。所建立的预测模型可以是针对个体空间观测基本单位(网格划分的小区单元、观测站点、路段等)进行单体预测,也可以是对某个总体考察区域内的全部观测基本单位的集体预测。在建模过程中,通常考察的交通流特性包括了非线性(体现在深度网络的非线性激活函数输入到

输出的变换中)、动态性(不同时段的预测,时空相关性捕获模式不同)、异质性(不同的空间观测单元具有不同的时空相关性捕获模式或独立而非共享的预测器)、多尺度时间相关性(以时段、天、周、月等为周期的时间相关性)、多尺度空间相关性(局部空间相关性、大范围空间相关性、全局空间相关性)。基于深度学习的模型总体上也被归结为“数据驱动”的预测研究。

当前,由于交通大数据获取的充分和深度学习理论的相对成熟,采用深度神经网络作为短期交通流预测的核心方法论已经成为研究的热点和主流。自动编码器^[15]、各种类型的卷积神经网络(1D 因果卷积^[16]、2D 图像卷积^[17]、3D 图像卷积^[18]等)、循环神经网络(长短期记忆(Long Short-Term Memory, LSTM)人工神经网络、门控循环单元(Gate Recurrent Unit, GRU)^[19]等)、卷积与循环相结合的神经网络(卷积长短期记忆(Convolutional Long Short-Term Memory, ConvLSTM)网络^[20]、预测递归神经网络(Predictive Recurrent Neural Network, PredRNN)^[21]等)、图神经网络(图卷积网络(Graph Convolutional Network, GCN)^[22]、扩散卷积(Diffusion Convolution, DC)^[23]、图注意力网络(Graph Attention Network, GAT)^[24]等)、时空注意力机制^[25]、深度自注意力变换网络^[26]等神经网络领域的理论纷纷被用于短期交通流预测领域的研究。从当前学术文献公布的试验结果来看,基于深度学习理论的短期交通流预测相对于传统预测方法,在预测准确性上取得了质的提升。而深度学习不是一个结构确定、仅需调参的方法论模型(这一点与 ARIMA、支持向量机等具有很大不同),它是建立在基本神经网络架构、反向传播算法、梯度下降寻优、门控机制、卷积机制、记忆机制以及注意力机制等核心基础理论上的“架构可自由设计”的各种深度神经网络模型的集合,这就导致了研究者针对同一短期交通流预测问题,同样采用深度学习的方法论,其建立的深度模型结构可能很不相同。

短期交通流预测的研究历史中,有过很多代表性的文献综述^[27-30],系统阐述了时间序列、机器学习等方法在其研究中的应用,但截至目前,还没有针对基于深度学习进行短期交通流预测进行系统综述的研究公开发表。为此,本文系统梳理、总结了近年来基于深度学习进行短期交通流预测研究的最新成果,试图从短期交通流预测问题的本质和主要挑战出发,厘清不同深度神经网络模型

尝试解决的重点问题和独特研究思路,进一步将这些有代表性的研究思路进行归类、总结,以期为研究者提供有价值参考。

1 短期交通流预测问题

1.1 短期交通流预测问题的形式化表达

短期交通流预测问题可以表述为:将历史若干时段的交通流观测构成的数据序列输入到某个预测模型之中,输出对未来若干时段内的交通流所构成序列的预测。该预测问题的数学形式化表达为

$$X_p = F(X_h) \quad (1)$$

式中: $F(\cdot)$ 为用于短期交通流预测的模型; X_h 、 X_p 分别为历史观测的交通流序列和预测的未来交通流序列。

假定当前时段为 t ,预测采用的历史时段数为 τ_h ,预测未来的时段数为 τ_p ,则 $X_h = \{X_{t-\tau_h+1}, X_{t-\tau_h+2}, \dots, X_t\}$, $X_p = \{X_{t+1}, X_{t+2}, \dots, X_{t+\tau_p}\}$ 。

依据式(1),短期交通流预测问题可以进一步表述为:基于历史交通流观测数据,寻找一个“好的”函数 $F(\cdot)$,从而能够在给定历史时段状态数据输入的情况下,对未来做出“好的”预测。这里的预测“好坏”本质上是对预测函数 $F(\cdot)$ 的性能评价,一般包括均方误差(Mean Square Error, MSE)、绝对平均误差(Mean Absolute Error, MAE)和绝对平均百分比误差(Mean Absolute Percentage Error, MAPE)等多种方式^[18]。在这个意义上,短期交通流预测研究的历史可以看作是采用不同预测函数对观测数据进行拟合的发展史。早期是时间序列类模型(ARIMA等),然后是各种滤波类模型,再发展到 K 近邻等传统机器学习模型,直到现今的深度学习类模型。

在基于深度学习进行短期交通流预测的研究中,被寻找的预测函数 $F(\cdot)$ 是一个深度的神经网络,具有很强的拟合和泛化(即预测)能力,相对于传统机器学习模型和时间序列模型而言,能够更好的描述交通流中蕴含的复杂时空相关性。深度神经网络的拟合能力显著增强,在机器学习领域面临着过拟合的风险,但大规模的交通流时空数据又使得这种风险极大的降低,因此,在交通流预测领域中,深度学习与交通大数据的结合也是当前时代发展的必然。

1.2 短期交通流预测研究的本质

当前,短期交通流预测研究普遍假定:观测区域内,任何一个位置单元(网格区域、路段或其它空间观测单元)未来的交通流都可以基于其自身历史的交通流数据和与其具有空间相关性的其他位置单元

的历史交通流数据而预测得到。时间的维度上,位置单元未来的交通流与自身历史的交通流是相关的,这被称为时间相关性。以往熟知的时间序列分析理论(ARIMA、GARCH等)的根本假设都是基于这种时间相关性。这种相关性十分直观,一个位置单元未来的交通流是由其过去的交通流演变而来,相关是情理之中的事情。空间的维度上,位置单元未来的交通流与其周边位置单元历史的交通流是相关的,这被称为空间相关性。由于路网是连通的,即便是网格化后的路网,不同的网格单元之间也是有道路连接的,自然就会有交通流的相互传播,那么,不同的位置单元之间由于交通流的相互传播,必然使得它们之间的交通流存在着高度的相关性。这种时间相关性和空间相关性并不是相互独立的2个维度,而往往是融合在一起的时空相关性,这是由道路交通的特点决定的,因此,寻找预测函数 $F(\cdot)$ 的过程,也就是试图对这种时空相关性进行建模的过程,如果建立的模型 $F(\cdot)$ 能够有效捕获这种时空相关性,那么自然其预测性能就会较好。短期交通流预测研究的本质即是对这种时空相关性的建模(捕获),也就是某个位置单元未来待预测的交通流与其自身的历史交通流相关(即时间相关性),同时和与其邻近的位置单元的交通流相关(即空间相关性),因而该目标位置单元及其邻近位置单元历史的交通流观测数据被作为模型 $F(\cdot)$ 的输入,而未来该目标位置单元的交通流作为模型 $F(\cdot)$ 的输出,从而就通过模型 $F(\cdot)$ 将历史与未来、目标与邻近之间的时空交通流关联起来。然而,交通时空状态预测中的这种时空相关性是十分复杂的^[31-32],这种复杂性表现为多个方面,主要为时空相关性中存在的非线性、动态性、异质性、多尺度性等。

1.2.1 非线性

道路网上的交通流是由自主的人,在不确定的环境下(异常事件、不确定的需求等),驾驶的车辆运动所构成的集合现象。交通流在路网,尤其是城市路网中的传播是高度非线性的,这就导致了不同位置单元之间以及同一位置单元不同时段之间的交通流相关性也是高度非线性的。由于非线性激活函数的存在,深度神经网络本质上恰好是一个高度非线性的函数,拟合能力非常强大,这也是现在深度神经网络被普遍用于进行短期交通流预测的重要原因。

1.2.2 动态性

交通流的时间相关性和空间相关性往往都随时

间呈现出高度的动态性。这种动态性往往并不是相关联的“量”随着时间的改变而造成的,而是关联的“模式”随着时间改变而造成的。对于一个功能为居住小区的特定位置单元,早高峰期间,其人流流出量大致是上升的趋势,但是在晚高峰期间,其流出量大致呈现下降的趋势,这在本质上就要求不同时段范围内,基于历史流量预测未来流量的模型(关联模式)具有动态性。相反,如果采用不具有动态性的关联模式的预测模型,那么很难做出不同时段都精准的预测。此外,对于不同位置单元之间的空间相关性而言,同样是动态的,一天中不同时段内,2个位置单元之间的关联程度和方式也不尽相同,在一个住宅区(记为A)和一个商业区(记为B)之间,早高峰更多是从A到B的出行,晚高峰更多是从B到A的出行,这就决定了二者之间空间关联模式是随着时间而发生变化的,为了更精准地捕获二者之间的空间相关性带来的影响,需要更好地考虑这种动态性。

1.2.3 异质性

在空间上而言,不同的位置单元不能简单将其考虑为同质的,而在建模的过程中施加相同的交通流时间、空间相关性捕获模式。具体来说,一个位置单元的功能不同(居民区和商业区),它所呈现的交通流之间的时空相关模式就不同^[33]。居民区早高峰以出流为主,入流较少;而商业区则正好情形相反。此外,不同位置单元,由于地理位置不同,与周边位置单元之间的关联结构不同(路网连通程度、平均出行时间等),它们之间的空间关联模式也具有强烈的异质性,即同样是考虑周边位置单元对中心位置单元的交通流影响,不同中心位置单元之间具有明显的不同。建模过程中,对这种异质性的捕获和恰当描述,是提高预测准确性的重要因素。

1.2.4 多尺度性

短期交通流预测中,时间维度的相关性和空间维度的相关性都是多尺度的。在时间维度上,与未来时段交通流相关的历史交通流包含了同一天中与之相邻的若干历史时段(时段尺度)、每一历史天中与之相同的时段(天尺度)、每一历史周中与之相同的时段(周尺度)、每一历史月中与之相同的时段(月尺度)等;在空间维度上,一般会根据城市交通出行方式的发达程度确定不同的空间相关性捕获范围(市域范围、大范围、中范围、小范围等)。具体在实现方法上,时间维度上的多尺度相关性,通常是基于

多尺度的历史数据输入,然后在模型中进行融合得到综合表征,进一步对未来时段的交通流做出预测;在空间的维度上,则可以利用不同的卷积核大小,限定卷积操作的范围,从而实现不同尺度空间相关性的捕获。

1.3 短期交通流预测研究的其它相关问题

1.3.1 外部信息

位置单元未来的交通流固然直接和它自身及其周边位置单元的历史交通流相关联,但不可否认的是,一些外部信息同样对其构成了重要影响:(1)待预测时段的时间特性(所在的天是否是周末、节假日等);(2)历史时段和待预测时段的天气特征(风速、温度、降雨等);(3)特殊的交通事件(交通拥堵、演唱会、足球比赛等)。为了更好地对未来交通流做出预测,需要进一步整合这些相关的重要的外部信息,由此,式(1)所描述的一般性短期交通流预测问题需要扩展为

$$X_p = F(X_h, E) \quad (2)$$

式中: E 为与预测相关的历史和未来时段内外部信息相关特征。

通常情况下,外部信息的整合方式主要包括2种:全连接网络嵌入^[34-35]和知识图谱嵌入^[36]。外部信息可以被作为任意观测时段和位置单元所对应的一种特殊的交通流特征,从数据结构上 X_h 和 E 两者是一致的,只是代表的含义不同。采用全连接网络嵌入的方式进行外部信息整合一般需要先对外部信息进行编码(如对不同天气采用独热码进行编码),然后将代表外部信息的编码向量再输入到一个小型的全连接神经网络中,将其嵌入到与 X_h 数据结构相一致的隐空间,然后再将二者拼接起来,共同作为代表历史信息的输入。这种整合方式的好处是直观、有效,但是缺点是对于外部信息之间的相关性考察不足(将不同外部信息简单的进行拼接不能反映出它们之间的复杂相关性),为了进一步刻画外部信息之间的这种相关性,有的研究采用了知识图谱这种相对更为复杂的外部信息特征嵌入方式,进行外部信息到深度预测模型的整合^[37]。这种方式的优点是捕获外部信息与交通流信息之间的语义相关性提供了可能,但缺点是计算相对复杂,同时知识图谱网络与交通网络本质上是异质的,二者之间的融合存在一定的挑战。

1.3.2 单体预测与集体预测

以往的研究中,对于未来时段的交通流预测可以是针对每个单独的位置单元建立单体预测模型,

从而进行个体级别的预测^[38-39],也可以建立集体预测模型,在集体层面一次性的预测所有位置单元未来的交通流^[40-41]。这2种预测的差别主要在于预测范围的不同。单体预测是针对单个交通流数据采集位置单元(如一个特定交通检测站、一条特定路段、一个特定网格化的城市单元等)独立进行预测,而集体预测更多是关注市域范围内的多位置单元(如多条路、多个检测站、多个网格化城市单元等)进行同步预测。

单体预测模型为

$$X_p(i) = F(X_h) \quad i \in [1, N] \quad (3)$$

式中: $X_p(i)$ 为未来待预测交通流序列中位置单元*i*的状态; N 为位置单元总数。

集体预测模型为

$$\{X_p(i)\}_{i=1, \dots, N} = F(X_h) \quad (4)$$

1.3.3 单步预测与多步预测

交通流预测可以只预测当前时段的下一个时段^[42-43],也可以预测当前时段后的若干个时段,前者称为单步预测,后者称为多步预测^[44-45]。从预测的难度而言,多步预测显著高于单步预测,预测的步数越多越容易导致更大的误差,但从实用性来讲,预测步数越多,越对下游的交通管控、诱导、应急决策等任务制定起到更好的支撑作用。

单步预测模型为

$$X_{t+1} = F(X_h) \quad (5)$$

多步预测模型为

$$X_{t+1, t+\tau_p} = F(X_h) \quad (6)$$

式中: $X_{t+1, t+\tau_p}$ 为*t*+1到*t*+ τ_p 时段的多步预测的交通流序列。

2 基于深度学习的短期交通流预测一般性框架

一般而言,基于深度学习进行短期交通流预测的模型都包含3个基本组件,分别为输入特征嵌入器、时空特征萃取器和交通流预测器,如图1所示,其中, $\hat{X}_p$ 为所预测的未来交通流序列。输入特征嵌入器负责对原始数据的数据特征进行维度的变换。一般情况下,输入特征嵌入器都会是一个小型的全连接神经网络(一般2层以内);时空特征萃取器是整个深度模型的核心,用于对交通流预测中的复杂时空相关性进行高阶特征萃取;交通流预测器以时空特征萃取器所得的高阶时空表征为输入,最后输出对未来交通流的预测。如果是单步预测,交通流预测器一般都会是一个小型的全连接神经网络或

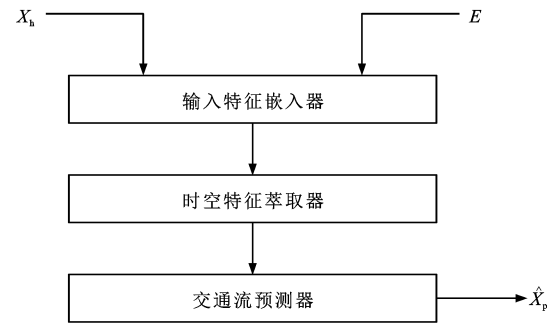


图1 基于深度学习的短期交通流预测一般框架

Fig. 1 General framework of short-term traffic flow prediction based on deep learning

卷积核大小为 1×1 的卷积神经网络;如果是多步预测,交通流预测器可以是一个全连接神经网络,也可以是一个基于循环神经网络单元(LSTM、GRU等)的解码器结构,依次对未来多个时段进行预测。

本文从交通流数据组织的2种不同方式(即网格化和拓扑化)出发,分别对各种基于深度学习方法进行短期交通流预测的研究进行综述,相关文献的总体概况如表1所示。

3 基于深度学习的网格化短期交通流预测

将反映道路交通流的指标数据按照网格化的位置单元进行汇总和统计,是以往短期交通流预测研究中常见的方式。研究区域被划分为若干等大小的矩形网格单元,基于每个网格单元,可以分时段统计其中的交通流指标(入流量、出流量、平均出行速度、交通密度等)。将研究区域进行网格化,然后在特定时段内,在每个网格上观测多种类型的交通流指标(流量、速度、密度等),最终形成的数据刚好可以类比为计算机图像处理领域的像素图:每个网格单元相当于一个像素点,其上的交通流指标观测相当于多个通道的取值。在将特定时段内研究区域的网格化交通流数据类比为图像之后,就可以很方便地引入图像处理领域非常成熟且取得巨大成功的深度卷积神经网络,从而对观测数据进行建模。图像卷积的本质是通过对目标像素位置周边一定范围内的特征进行加权融合,从而生成目标像素位置新的特征。卷积操作的这种方式,正好在本质上实现了像素(网格单元)之间的空间相关性建模,而卷积核大小的不同,则对应了空间相关性建模中的多尺度特性。按照时空相关性建模方法的不同,下面对基于网格化交通流数据的3种常用深度学习建模理论分别进行综述。

表 1 基于深度学习的短期交通流预测文献概况

Table 1 Overview of literatures on short-term traffic flow prediction based on deep learning

数据组织方式	时空相关性捕获方式	代表性文献	核心理念	优势	劣势
网格化数据	2D 图像卷积神经网络	[46]~[48]	将时空网格化的交通流数据转化为多通道图像数据,从而单纯采用卷积神经网络进行时空相关性特征学习	将时空相关性特征学习转化为单纯的空间特征学习,方法相对简单、高效	由于将时间维度的信息转化为图像通道的信息,导致对时间相关性捕获欠佳
	2D 图像卷积与循环神经网络相结合	[21],[25],[49]~[54]	采用图像卷积网络捕获空间相关性,采用循环神经网络捕获时间相关性	将复杂的时空相关性捕获分解为时间相关性和空间相关性的单独捕获,降低了问题复杂度,模型思路简洁、清晰	时空相关性的分离式建模导致无法捕获时空同步相关性,此外,循环神经网络的训练和推理效率较低
	3D 图像卷积神经网络	[18],[55]~[60]	将历史网格化交通流数据视为“视频流”型数据,采用可以在时空维度上进行滑动的 3D 卷积核进行时空相关性特征学习	能够一定程度上捕获到时空同步的相关性,且模型仅采用卷积计算,具有并行计算的优势	3D 卷积相对于 2D 卷积的计算量明显上升,模型的训练和推理效率有所下降
拓扑化数据	1D 因果图像卷积与卷积图神经网络相结合	[61]~[65]	采用 1D 因果卷积捕获时间相关性,采用拓扑图卷积捕获空间相关性	采用 1D 因果卷积捕获时间相关性,相较于循环神经网络在训练和推理效率上明显提升	时空相关性的捕获采用了时空分离的方式,无法捕获到同步的时空相关性
	循环神经网络与卷积图神经网络相结合	[66]~[68]	采用循环神经网络捕获时间相关性,采用卷积图神经网络捕获空间相关性	采用循环神经网络和卷积图神经网络分别进行时间和空间相关性捕获,理论较为成熟	时空相关性的捕获采用了时空分离的方式,无法捕获到同步的时空相关性,此外,循环神经网络训练和推理效率较低
	自注意力与卷积图神经网络	[69]	自注意力机制负责从时间的维度捕获相关性,卷积图神经网络负责从空间维度捕获相关性	注意力机制在时间维度相关性捕获上,由于并行本质,计算高效	无法捕获时空同步相关性,容易忽略时间顺序对未来预测的影响,需要时间位置编码补充
	基于卷积图神经网络的时空同步学习	[70],[71]	将空间拓扑图拓展到时空拓扑图,从而统一采用拓扑图卷积捕获时空相关性	空间拓扑到时空拓扑变换,导致网络规模急剧增加,模型运算复杂度明显提升	能够捕获到时空同步的相关性

3.1 基于 2D 图像卷积神经网络的时空相关性建模

采用 2D 图像卷积神经网络对单个时段内的网格化交通流观测进行建模,只能够捕获空间相关性。为了进一步捕获不同时段内网格化交通流观测之间的时间相关性,基于 2D 图像卷积神经网络的时空相关性建模方法,是将历史多个时段(以 $t \sim t+n$ 时段为例)的交通流观测图像集成为一个通道更多的单一图像,如图 2 所示。其中,每个时段的交通流图像有 F 个通道(对应多个交通流指标),在时间的维度上将所有的图像进行拼接聚合,最后得到了一个更多通道的图像。进一步,在这个多通道图像上就整合了历史交通流的时间和空间维度信息,此时,采用一个 2D 深度卷积神经网络对其进行建模,则可以同时捕获空间相关性和时间相关性。得到时空高阶表征后,输入到交通流预测器就可以对未来交通流做出预测。

这一类建模方法的优点是仅采用 2D 图像卷积神经网络进行时空相关性建模,而 2D 图像卷积由于本质上可以并行计算,因此,能够采用图形处理器

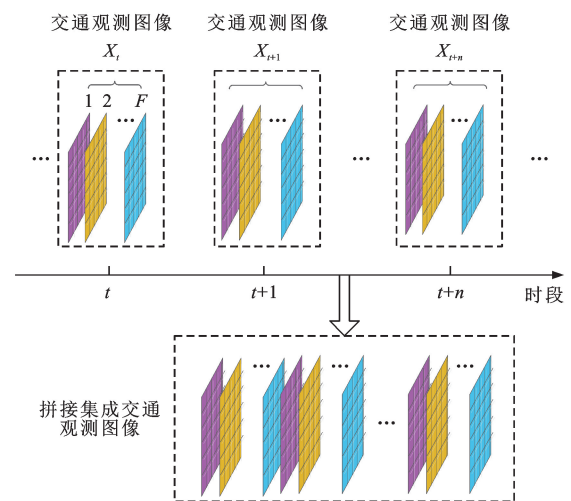


图 2 多个历史时段交通流图像拼接集成

Fig. 2 Concatenation aggregation of traffic flow images in multiple historical periods

(Graphics Processing Unit, GPU)等进行加速,计算效率很高;但是这类方法也具有明显的缺点,数据集成过程本质上是将时间维度的信息转换为图像的通道信息,而 2D 图像卷积在处理每个通道时是同

等对待的,然而实际上不同时间段的交通流对未来预测交通流的影响程度不可能是同等的。这种将时间维度的信息整合于空间维度信息之中的做法,尽管增加了问题处理的效率,却容易忽视时间维度建模的精度。

3.2 基于2D图像卷积与循环神经网络相结合的时空相关性建模

图3给出了基于2D图像卷积与循环神经网络相结合的时空相关性建模的一般性框架,其中, H_0 为初始化的隐藏状态, Y_t 为高阶的时空表征。在捕获交通流数据时空相关性时,每个时段的交通流图像首先通过一个卷积神经网络,从而在空间的维度上进行特征融合,然后将空间融合的新图像输入到一个循环神经网络中,循环的基本单元可以是LSTM、GRU等,如图3中的卷积神经网络(Convolutional Neural Network, CNN)和循环神经网络(Recurrent Neural Network, RNN)都可以是多层的。这样,最终输出的高阶时空表征 Y_t 中,每个位置单元的特征向量中就融合了与之相关的历史交通流数据中的时间信息和空间信息。进一步,将这个高阶时空表征作为输入,输入到交通流预测器中就可以对未来做出预测。

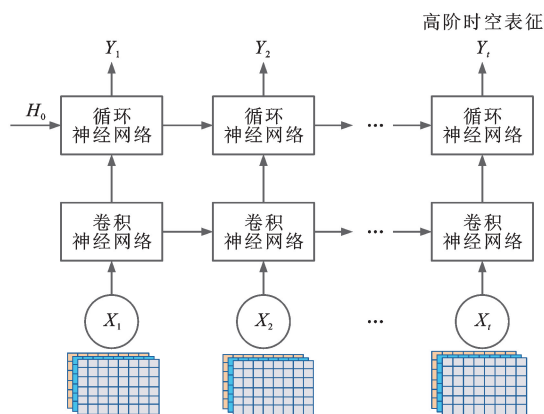


图3 基于2D图像卷积与循环神经网络的时空相关性建模
Fig.3 Spatiotemporal correlation modeling based on 2D image convolution and recurrent neural network

此外,文献[21]、[25]、[52]~[54]将卷积与循环神经网络进行了进一步的融合,从而形成能够同时捕获时空相关性的“新型循环体”(最早的研究是Shi等^[20]提出的ConvLSTM组件)。它们的原理与图3中的时空相关性捕获方式相同,区别在于将空间相关性捕获的卷积操作纳入到循环神经网络的循环体内部计算中(包括门控值和内部状态的计算)。

基于2D图像卷积与循环神经网络相结合的时空相关性建模方法是一种直观的建模方式,通过卷

积来捕获空间相关性,然后再通过循环来捕获时间相关性。由于有循环神经网络显式的来捕获时间维度的信息,因此,该方法克服了3.1节单纯基于2D图像卷积神经网络将时间维度信息变换为通道特征信息做法的缺点,但同时,由于该方法主体是基于循环神经网络的建模方式,其计算过程是串行的,计算效率较低,模型训练难度较大,因此,尽管理论上可以堆叠多层的循环单元(深度循环神经网络),但实际应用中往往不能太深,否则很难训练。

3.3 基于3D图像卷积的时空相关性建模

基于3D图像卷积的时空相关性建模最早是由Guo等^[18]于2019年提出的。2D图像卷积的建模方法中卷积核只能在2D平面空间中移动,因此,没有显式的捕获时间相关性。而2D图像卷积和循环神经网络的结合尽管显式考虑了时间和空间相关性建模,但是受限于循环神经网络的串行计算本质,其计算效率大大降低。交通流图像序列如果从计算机视觉角度来看,本质上就是一个视频(图像流),每个观测图像就是视频中的一帧。鉴于历史交通流观测序列本质上是与视频流相同,研究者试图采用卷积核可以在平面空间和时间3个维度上进行移动的3D图像卷积对交通流时空相关性进行建模^[55-60]。这种建模方式显式地捕获了时空相关性,又可以利用卷积运算的并行特点提高运算效率,具有很大的建模优势。

4 基于深度学习的拓扑化短期交通流预测

将反映道路交通流的指标数据按照拓扑化的位置单元为进行汇总和统计,是以往短期交通流预测研究中另外一种常见的方式。研究区域内有多个交通流观测站点,它们负责观测多个维度的交通流数据(速度、流量等),由于道路网络本身的拓扑化连接性质,这些观测站点也可以被视为是拓扑化连接起来的。如图4所示,左侧为道路网络拓扑图和观测站点,右侧为对应转换后观测站点之间的拓扑连接图。相

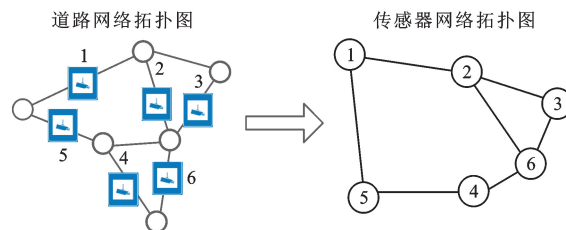


图4 交通观测拓扑化网络示例

Fig.4 Example of traffic observation topological networks

应的,每个时段内交通流观测数据不再是一张像素图像,而是一个多通道的拓扑图,如图 5 所示。

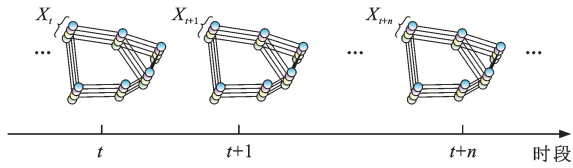


图 5 拓扑化道路交通时空观测数据示例

Fig. 5 Example of topological road traffic spatiotemporal observation data

拓扑化的交通流数据组织方式非常符合道路网络的天然特点,因为道路网络本质就是拓扑结构。拓扑化与网格化表征的交通流数据本质上是 2 种数据结构:拓扑图和像素图。深度学习领域对这 2 种数据结构研究的方法论也明显不同,针对像素图主流的深度学习方法是图像卷积神经网络,而针对拓扑图主流的深度学习方法是图神经网络。图神经网络是近几年深度学习领域研究的热点和前沿,它的核心是要在拓扑结构数据上进行深度学习。根据交通流预测核心问题,即时空相关性建模方式的不同,选取 4 种典型拓扑化短期交通流预测进行综述。

4.1 基于 1D 因果图像卷积与卷积图神经网络相结合的时空相关性建模

图 6 给出了基于 1D 因果图像卷积与卷积图神经网络相结合的时空相关性建模的一般性框架。思想核心在于:在时间的维度,采用 1D 因果图像卷积进行信息融合(时间相关性建模);在空间的维度上,采用卷积图神经网络(Convolutional Graph Neural Networks, ConvGNN)进行信息融合(空间相关性建模),进而将二者结合,从而完成时空相关性的集

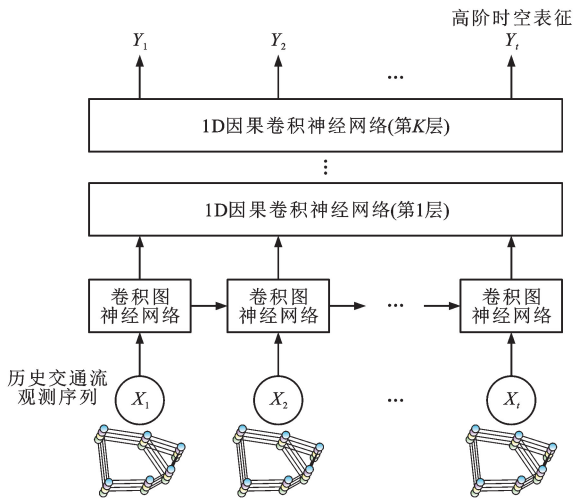


图 6 基于 1D 因果图像卷积与卷积图神经网络的时空相关性建模

Fig. 6 Spatiotemporal correlation modeling based on 1D causal image convolution and convolution graph neural network

成建模。进行空间相关性建模的卷积图神经网络主要分为 2 类:谱域图卷积网络和顶点域图卷积网络。谱域图卷积网络最有代表性的是切比雪夫多项式近似图卷积^[72]、图卷积网络^[73]以及扩散卷积神经网络^[74];而顶点域图卷积网络最有代表性的则是图注意力网络^[75]、基于图的归纳学习^[76]等。1D 因果图像卷积是一种循环神经网络的替代,它的优势是在时间的维度上共享卷积核,从而能够实现时间信息的加权融合(时间相关性捕获)。由于其本质上为卷积运算,因此是并行的,计算效率远高于循环神经网络。同时,有的研究^[65]还将 1D 因果卷积与膨胀卷积相结合,从而通过堆叠少数几层的网络就能够实现长期时间相关性的捕获。

4.2 基于循环神经网络与卷积图神经网络相结合的时空相关性建模

图 7 给出了基于循环神经网络与卷积图神经网络相结合的时空相关性建模的一般性框架。在每一个时间步上,首先采用卷积图神经网络(谱域图卷积或顶点域图卷积)对交通流的空间信息进行融合(空间相关性捕获),然后将空间融合后的新拓扑特征图输入到循环神经网络中(GRU、LSTM 等)进行时间维度的信息融合(时间相关性捕获)。最终输出时空高阶表征 Y_t ,并以此作为输入,采用交通流预测器对未来交通流做出预测。还可以进一步将分离的空间信息融合和时间信息融合整合到一起,构成一个能够同时捕获空间和时间相关性的新组件,也就是将图卷积的操作纳入到循环网络单元的内部计算之中。

如 3.2 节所述,由于时间相关性的捕获采用了循环神经网络的方式,其想法直观,但计算效率相对较低,如果网络堆叠较多层,训练和推理速度都会较慢。

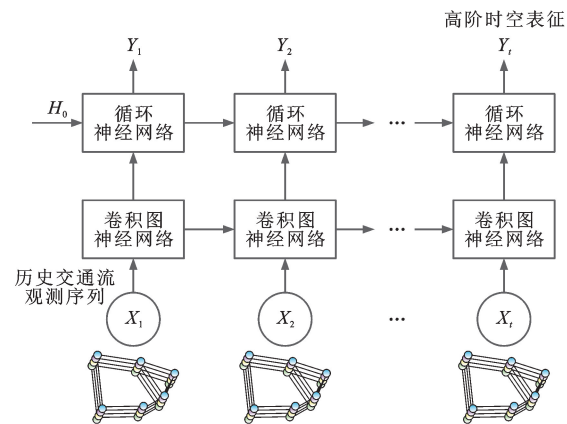


图 7 基于循环神经网络与卷积图神经网络的时空相关性建模

Fig. 7 Spatiotemporal correlation modeling based on recurrent neural network and convolutional graph neural network

4.3 基于自注意力与卷积图神经网络相结合的时空相关性建模

自注意力机制是近年来深度学习领域的重大突破,自从文献[77]发表以来,它被广泛应用于各种深度学习模型之中,见于自然语言处理领域最前沿的 GPT-3^[78]和视觉领域的视觉自注意力模型^[79]。自注意力良好的设计机制,可以很好地统一循环神经网络在时间维度的信息融合和卷积神经网络在空间维度的信息融合。在短期交通流预测领域主要有 2 种基于自注意力的研究思路:(1)结合自注意力和卷积图神经网络,前者负责从时间的维度捕获相关性,而后者从空间的维度捕获相关性;(2)时间和空间相关性建模都采用自注意力机制,即时间注意力与空间注意力机制融合。

自注意力机制无论是用于捕获空间相关性还是时间相关性上,单个注意力层就具有全局视野,这是它的运算本质决定的。相对来说,卷积图神经网络为了获取全局视野,往往要叠加很多个图卷积层才能够实现,因此,在全局空间相关性和时间相关性捕获方面,其独特的优势往往可能带来更好的预测效果。此外,自注意力机制本身的运算是并行的,非常利于 GPU 加速,因此相对于主体为循环神经网络的深度模型,其运算效率要高得多。但是,将自注意力机制用于时间和空间维度的信息聚合,很容易丢失时间和空间位置信息,而不同时间和空间位置的交通流信息对未来待预测的交通流影响是不等价的。在时间的维度上,可以借鉴 Transformer^[61]关于对时间序列每个分量位置进行编码的方式进行位置编码。在空间的维度上,Velikovi 等^[75]尝试用可学习的位置编码方式以及知识图谱的节点位置特征嵌入方式进行编码。尽管在空间的维度上进行了位置编码,但采用自注意力机制还是丢失了拓扑结构信息,而丢失了该信息也就违背了基于拓扑网络数据进行深度学习的本质诉求,因此,在实际中,空间相关性的捕获是基于卷积图神经网络和自注意力机制共同完成的^[69],前者可以理解为依赖拓扑结构的空信息融合,而后者可以理解为全局不依赖拓扑结构的空信息融合。

4.4 基于卷积图神经网络的时空同步学习相关性建模

针对拓扑化数据,4.1~4.3 节中的 3 种研究思路本质上都是将时间和空间相关性的捕获分离进行,而在真实世界中,交通流的时空相关性是并不总是独立的,还包括了同步的时空相关性^[70]。如图 8

所示,除了单纯的时间相关性和空间相关性以外,还存在着时空耦合的相关性。如果采用分离的时间相关性和空间相关性建模方式(t_1 、 t_2 两个时刻),则无法准确捕获这种时空耦合的相关性。

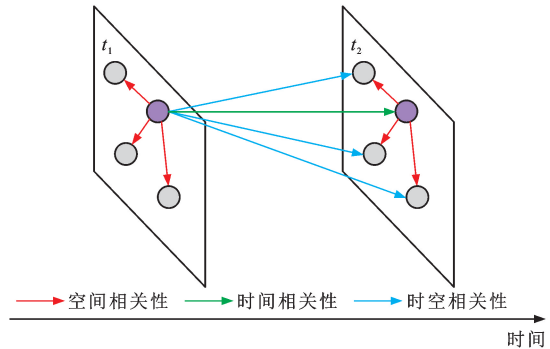


图 8 交通流序列中的各种相关性

Fig. 8 Various correlations in traffic flow sequence

在 3.1 节介绍的基于 2D 图像卷积进行网格化数据的时空相关性建模研究中,其核心是将不同时间段的交通流观测图像在通道的维度上进行拼接集成,从而构成一幅具有更多通道的图像,进而在这个多通道图像上进行 2D 图像卷积操作,就可以同时融合时间和空间维度的信息。相似地,可以将不同时间段的拓扑化状态观测图在特征的维度上进行拼接,从而构成一个特征维度更多的拓扑状态观测图,进而单纯基于卷积图神经网络就可以同时实现时空 2 个维度的信息聚合^[71]。此外,除了直接将不同时间段的拓扑特征图直接在特征的维度上进行拼接以外,还可以在不同时间段的拓扑图的顶点之间建立虚拟的连接,从而将原有的空间连接拓展到时空 2 个维度的连接关系,进而在更大规模的拓扑网络上采用拓扑图卷积实现时空相关性的同步建模^[70]。

5 研究展望

近年来,得益于交通大数据充分获取与深度学习方法论快速发展的双重驱动,基于深度学习进行短期交通流预测,在准确性上,相较于时间序列模型、传统机器学习等方法获得了较大的提升,但仍旧存在较多的研究方向有待进一步深入研究和探索。

(1)通过物理知识与深度学习相结合的方式,对短期交通流预测进行研究。单纯基于深度学习进行短期交通流预测,能够取得较高的预测精度,源于大量的高质量训练数据和深度学习模型的强大表征能力,然而在实际应用中,往往面临数据贫乏或者数据存在噪声的问题。在这种情况下,仅依靠数据驱动的模式很难得到准确可靠的预测结果。而在深度学

习模型中引入物理知识,即便在数据较少的情况下也能给出符合物理直观的预测,会极大增强预测的稳定性和准确度。

(2)建立交通时空数据的大模型,实现更为精准的短期交通流预测。短期交通流预测只是交通时空数据挖掘和建模的任务之一,除此之外还至少包含了异常检测、事件检测、数据填补等其他多项任务。当前人工智能、深度学习的发展已进入“大模型”时刻,前沿研究实践表明,建立支撑多任务的通用基础大模型,而非单一任务的深度学习模型,将会是未来发展的重要趋势和探索方向。

(3)基于深度学习进行短期交通流预测,不仅要关注预测的准确性,同时也应该关注模型轻量化、可解释性以及模型架构自动学习等相关问题。当前基于深度学习进行短期交通状态的研究主要还是在追求准确性的提升,而作为智能交通应用的核心基础科学问题,短期交通状态预测要支撑实践应用不能单一的追求准确性维度,未来应该在可解释性、轻量化(深度模型压缩)、自动化等多个维度进行提升。

(4)深度学习与知识图谱的融合,可以成为更加有效的短期交通流预测方法论。从方法论的角度来看,接下来第3代人工智能的发展将是基于数据与知识联合驱动的范式。单纯基于深度学习的预测建模,其可解释性较差,在对未来预测过程中究竟什么因素起到了重要影响,很难采用有效、直观的方式呈现出来,这就造成了模型可信性的问题。未来,将预测所涉及的外部因素、先验知识等抽象为知识图谱,然后将知识图谱与深度学习相结合进行预测预期是一个重要发展趋势。

6 结 语

(1)深度学习是继时间序列模型、卡尔曼滤波理论以及传统机器学习方法之后的新一代短期交通流预测理论,具有拟合和预测能力更强的特点,与大数据的结合使得短期交通流预测性能得到了极大提升,但同时也应该注意到,深度学习天然的“黑箱”特点,使得其相对于传统机器学习和时间序列类方法而言,往往不具有良好的可解释性,同时,调参复杂一直是基于深度学习的研究方法固有的缺点,有关模型架构、超参数自动化学习等来自自动化机器学习领域的研究方法,有望进一步拓展当前的短期交通流预测研究。

(2)基于深度学习进行短期交通流预测的关键在于对现实交通流中存在的物理现象的精准捕获,

而不在于各种深度学习方法本身。从深度学习的模型而言,1D因果卷积、循环神经网络和自注意力机制等都可以实现时间维度的特征提取和信息融合,除了对预测结果的准确性进行特定指标(MSE、均方根误差(Root Mean-Square Error, RMSE)等)的评价之外,很难评价其各自的优劣。但有价值的研究不应是简单的将各种可行的方法仅仅应用于交通流预测问题,而是要尽量阐明和揭示出深度学习的运算与实际物理现象的对应关系,亦即深度学习应用于短期交通需求预测成功的关键,并不仅仅在于预测准确性评价指标的高低,更应该偏重于是否对应物理现象和规律的揭示,这方面的研究恰恰需要更进一步在交通流特定规律上的深入研究,而不仅仅是将“时髦”的深度学习理论简单的应用于预测问题。

(3)短期交通流预测的目的是服务于下游管控、诱导、应急决策等应用,而下游策略的制定必将反过来影响交通流发生改变。当前的研究往往将短期交通流预测作为一个孤立的环节,基于历史交通预测未来交通,但从系统闭环的思维来看,需要将预测、决策与管控等进行一体化考量,同时将离线的预测转为在线的方式(即增量学习),从而才能进一步增加其现实的实用性,这方面的研究可能会成为来重要的拓展方向。

(4)本文仅针对图1中的主流监督式短期交通流预测范式进行了系统综述,但应该注意到,近年来有少数研究采用无监督的生成式对抗生成网络(Generative Adversarial Net, GAN)进行短期交通需求预测^[80-81]。从本质上而言,GAN提供了一种“生成式”架构,期望学到历史交通流分布的内在“规律”,然后按照“规律”生成符合相同分布的未来预测,但从所采用的组件而言,同样是基于CNN、LSTM等进行时空相关性的捕获。这些研究为交通流预测提供了同样重要的借鉴和参考。

参 考 文 献 :

References :

- [1] DU W D, ZHANG Q Y, CHEN Y P, et al. An urban short-term traffic flow prediction model based on wavelet neural network with improved whale optimization algorithm[J]. Sustainable Cities and Society, 2021, 69: 102858.
- [2] 史其信,郑为中.道路网短期交通流预测方法比较[J]. 交通运输工程学报, 2004, 4(4): 68-71, 83.
SHI Qi-xin, ZHENG Wei-zhong. Short-term traffic flow prediction methods comparison of road networks[J]. Journal of Traffic and Transportation Engineering, 2004, 4(4): 68-71, 83.

- (in Chinese)
- [3] 王晓全,邵春福,尹超英,等. 基于 ARIMA-GARCH-M 模型的短时交通流预测方法[J]. 北京交通大学学报, 2018, 42(4): 79-84.
WANG Xiao-quan, SHAO Chun-fu, YIN Chao-ying, et al. Short term traffic flow forecasting method based on ARIMA-GARCH-M model[J]. Journal of Beijing Jiaotong University, 2018, 42(4): 79-84. (in Chinese)
 - [4] 于 泉,姚宗含. 交通流预测的马尔科夫粒子滤波方法研究[J]. 交通运输系统工程与信息, 2019, 19(2): 209-215.
YU Quan, YAO Zong-han. Markov particle filter traffic flow prediction model [J]. Journal of Transportation Systems Engineering and Information Technology, 2019, 19(2): 209-215. (in Chinese)
 - [5] 白伟华,张传斌,张煊旂,等. 基于异常值识别卡尔曼滤波器的短期交通流预测[J]. 计算机应用研究, 2021, 38(3): 817-821.
BAI Wei-hua, ZHANG Chuan-bin, ZHANG Shuang-yi, et al. Outlier-identified Kalman filter for short-term traffic flow forecasting[J]. Application Research of Computers, 2021, 38(3): 817-821. (in Chinese)
 - [6] 杨高飞,徐 睿,秦 鸣,等. 基于 ARMA 和卡尔曼滤波的短时交通预测[J]. 郑州大学学报(工学版), 2017, 38(2): 36-40.
YANG Gao-fei, XU Rui, QIN Ming, et al. Short-term traffic volume forecasting based on ARMA and Kalman filter[J]. Journal of Zhengzhou University (Engineering Science), 2017, 38(2): 36-40. (in Chinese)
 - [7] 王 翔,陈小鸿,杨祥妹. 基于 K 最近邻算法的高速公路短时行程时间预测[J]. 中国公路学报, 2015, 28(1): 102-111.
WANG Xiang, CHEN Xiao-hong, YANG Xiang-mei. Short term prediction of expressway travel time based on K nearest neighbor algorithm[J]. China Journal of Highway and Transport, 2015, 28(1): 102-111. (in Chinese)
 - [8] LUO X L, LI D Y, ZHANG S R. Traffic flow prediction during the holidays based on DFT and SVR[J]. Journal of Sensors, 2019, 2019(10): 6461450.
 - [9] DONG C J, CUI A. Prediction models of short-term traffic flow based on neural network[J]. Advanced Materials Research, 2013, 671-674: 2908-2911.
 - [10] LIU Zhong-bo, YANG Zhao-sheng, GAO Peng. Research on the short-term traffic flow prediction method based on BP neural networks[C]//IEEE. 2012 World Automation Congress. New York: IEEE, 2012: 94214.
 - [11] MA Xiao-lei, DAI Zhuang, HE Zheng-bing, et al. Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction[J]. Sensors, 2017, 17(4): 818.
 - [12] 邵春福,薛 松,董春娇,等. 考虑时空相关性的网络交通流短期预测[J]. 北京交通大学学报, 2021, 45(4): 37-43.
SHAO Chun-fu, XUE Song, DONG Chun-jiao, et al. Short-term traffic flow prediction considering temporal-spatial correlation on road network[J]. Journal of Beijing Jiaotong University, 2021, 45(4): 37-43. (in Chinese)
 - [13] LIU Qing-chao, LIU Tao, CAI Ying-feng, et al. Explanatory prediction of traffic congestion propagation mode: a self-attention based approach[J]. Physica A: Statistical Mechanics and its Applications, 2021, 573: 125940.
 - [14] 陈喜群,周凌霄,曹 震. 基于图卷积网络的路网短时交通流预测研究[J]. 交通运输系统工程与信息, 2020, 20(4): 49-55.
CHEN Xi-qun, ZHOU Ling-xiao, CAO Zhen. Short-term network-wide traffic prediction based on graph convolutional network[J]. Journal of Transportation Systems Engineering and Information Technology, 2020, 20(4): 49-55. (in Chinese)
 - [15] LYU Yi-sheng, DUAN Yan-jie, KANG Wen-wen, et al. Traffic flow prediction with big data: a deep learning approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(2): 865-873.
 - [16] GE Liang, LI Si-yu, WANG Ya-qiang, et al. Global Spatial-temporal graph convolutional network for urban traffic speed prediction[J]. Applied Sciences, 2020, 10(4): 1509.
 - [17] LIN Hao-xing, JIA Wei-jia, SUN Yi-ping, et al. Spatial-temporal self-attention network for flow prediction[J]. arXiv, 2019, DOI: 10.48550/arXiv.1912.07663.
 - [18] GUO Sheng-nan, LIN You-fang, LI Shi-jie, et al. Deep spatial-temporal 3D convolutional neural networks for traffic data forecasting[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(10): 3913-3926.
 - [19] FU Rui, ZHANG Zuo, LI Li. Using LSTM and GRU neural network methods for traffic flow prediction[C]//IEEE. 31st Youth Academic Annual Conference of Chinese Association of Automation. New York: IEEE, 2016: 324-328.
 - [20] SHI Xing-jian, CHEN Zhuo-rong, WANG Hao, et al. Convolutional LSTM Network: a machine learning approach for precipitation nowcasting[C]//NIPS. 29th Annual Conference on Neural Information Processing Systems. San Diego: NIPS, 2015: 802-810.
 - [21] WANG Y, WU H, ZHANG J, et al. PredRNN: a recurrent neural network for spatiotemporal predictive learning [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(2): 2208-2225.
 - [22] ZHANG Zheng-chao, LI Meng, LIN Xi, et al. Multistep speed prediction on traffic networks: a deep learning approach considering spatio-temporal dependencies[J]. Transportation Research Part C: Emerging Technologies, 2019, 105: 297-322.
 - [23] ANDREOLETTI D, TROIA S, MUSUMECI F, et al. Network traffic prediction based on diffusion convolutional recurrent neural networks[C]//IEEE. 2019 INFOCOM IEEE Conference on Computer Communications Workshops. New York: IEEE, 2019: 246-251.
 - [24] JIN Cang-hong, TAO Ruan, WU De-xing, et al. HetGAT: a heterogeneous graph attention network for freeway traffic speed prediction[J]. Journal of Ambient Intelligence and Humanized Computing, 2021, DOI: 10.1007/s12652-020-02807-0.
 - [25] LIN Zhi-hui, LI Mao-mao, ZHENG Zhuo-bin, et al. Self-

- attention ConvLSTM for spatiotemporal prediction[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 11531-11538.
- [26] CAI L, JANOWICZ K, MAI G C, et al. Traffic transformer: capturing the continuity and periodicity of time series for traffic forecasting[J]. Transactions in GIS, 2020, 24(3): 736-755.
- [27] 刘 静, 关 伟. 交通流预测方法综述[J]. 公路交通科技, 2004, 21(3): 82-85.
LIU Jing, GUAN Wei. A summary of traffic flow forecasting methods[J]. Journal of Highway and Transportation Research and Development, 2004, 21(3): 82-85. (in Chinese)
- [28] 王 进, 史其信. 神经网络模型在短期交通流预测领域应用综述[J]. 河南科技大学学报(自然科学版), 2005, 26(2): 22-26.
WANG Jin, SHI Qi-xin. Review of application of neural network based models in short-term traffic flow forecasting[J]. Journal of Henan University of Science and Technology (Natural Science), 2005, 26(2): 22-26. (in Chinese)
- [29] 李振龙, 张利国, 钱海峰. 基于非参数回归的短时交通流预测研究综述[J]. 交通运输工程与信息学报, 2008, 6(4): 34-39.
LI Zhen-long, ZHANG Li-guo, QIAN Hai-feng. Review of the short-term traffic flow forecasting based on the non-parametric regression[J]. Journal of Transportation Engineering and Information, 2008, 6(4): 34-39. (in Chinese)
- [30] 郭 敏, 肖 翔, 蓝金辉. 道路交通流短时预测方法综述[J]. 自动化技术与应用, 2009, 28(6): 8-9, 16.
GUO Min, XIAO Xiang, LAN Jin-hui. A summary of the short-time traffic flow forecasting methods[J]. Techniques of Automation and Applications, 2009, 28(6): 8-9, 16. (in Chinese)
- [31] ZHU Jia-wei, HAN Xing, DENG Han-han, et al. KST-GCN: a knowledge-driven spatial-temporal graph convolutional network for traffic forecasting[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(9): 15055-15065.
- [32] CUI Zhi-yong, KE Rui-min, PU Zi-yuan, et al. Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values[J]. Transportation Research Part C: Emerging Technologies, 2020, 118: 102674.
- [33] PAN Zhe-yi, WANG Zhao-yuan, WANG Wei-feng, et al. Matrix factorization for spatio-temporal neural networks with applications to urban flow prediction[C]//ACM. Proceedings of the 28th ACM International Conference on Information and Knowledge Management. New York: ACM, 2019: 2683-2691.
- [34] 谷振宇, 陈 聪, 郑家佳, 等. 基于时空图卷积循环神经网络的交通流预测[J]. 控制与决策, 2022, 37(3): 645-653.
GU Zhen-yu, CHEN Cong, ZHENG Jia-jia, et al. Traffic flow prediction based on STG-CRNN[J]. Control and Decision, 2022, 37(3): 645-653. (in Chinese)
- [35] 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.
- [36] HUANG Wen-hao, SONG Guo-jie, HONG Hai-kun, et al. Deep architecture for traffic flow prediction: deep belief networks with multitask learning[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(5): 2191-2201.
- [37] HUANG Fei-hu, YI Pei-yu, WANG Jin-ce, et al. A dynamical spatial-temporal graph neural network for traffic demand prediction[J]. Information Sciences, 2022, 594: 286-304.
- [38] LI Wei, WANG Xin, ZHANG Yi-wen, et al. Traffic flow prediction over multi-sensor data correlation with graph convolution network[J]. Neurocomputing, 2021, 427: 50-63.
- [39] CHEN Li, ZHENG Lin-jiang, YANG Jie, et al. Short-term traffic flow prediction: from the perspective of traffic flow decomposition[J]. Neurocomputing, 2020, 413, 444-456.
- [40] 陈喜群, 曹 震, 沈楼涛, 等. 融合路段传输模型和深度学习的城市路网短时交通流状态预测[J]. 中国公路学报, 2021, 34(12): 203-216.
CHEN Xi-qun, CAO Zhen, SHEN Lou-tao, et al. Short-term traffic-state prediction of urban road networks based on the fusion of a link-transmission model and deep learning[J]. China Journal of Highway and Transport, 2021, 34(12): 203-216. (in Chinese)
- [41] 高华兵, 舒文迪, 刘 志. 基于深度学习的城市快速路交通流预测方法[J]. 浙江工业大学学报, 2022, 50(4): 406-412, 463.
GAO Hua-bing, SHU Wen-di, LIU Zhi. Urban expressway traffic flow prediction method based on deep learning[J]. Journal of Zhejiang University of Technology, 2022, 50(4): 406-412, 463. (in Chinese)
- [42] GUO Sheng-nan, LIN You-fang, WAN Huai-yu, et al. Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(11): 5415-5428.
- [43] YE Ji-hua, XUE Sheng-jun, JIANG Ai-wen. Attention-based spatio-temporal graph convolutional network considering external factors for multi-step traffic flow prediction[J]. Digital Communications and Networks, 2022, 8: 343-350.
- [44] XU Yao-bin, LIU Wei-tang, JIANG Zhong-yi, et al. MAF-GNN: multi-adaptive spatiotemporal-flow graph neural network for traffic speed forecasting[J]. arXiv, 2021, DOI: 10.48550/arXiv.2108.03594.
- [45] POLSON N G, SOKOLOV V O. Deep learning for short-term traffic flow prediction[J]. Transportation Research Part C: Emerging Technologies, 2017, 79: 1-17.
- [46] ZHANG Jun-bo, ZHENG Yu, SUN Jun-kai, et al. Flow prediction in spatio-temporal networks based on multitask deep learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 32(3): 468-478.
- [47] ZHANG Jun-bo, YU Zheng, QI De-kang, et al. Predicting citywide crowd flows using deep spatio-temporal residual networks[J]. Artificial Intelligence, 2018, 259: 147-166.
- [48] LIANG Y, OUYANG K, WANG Y, et al. Revisiting convolutional neural networks for citywide crowd flow

- analytics[C]//Springer. European Conference on Machine Learning and Knowledge Discovery in Databases. Berlin: Springer, 2020: 578-594.
- [49] 李磊,张青苗,赵军辉,等. 基于改进 CNN-LSTM 组合模型的分时段短时交通流预测[J]. 应用科学学报, 2021, 39(2): 185-198.
- LI Lei, ZHANG Qing-miao, ZHAO Jun-hui, et al. Short-term traffic flow prediction method of different periods based on improved CNN-LSTM[J]. Journal of Applied Sciences, 2021, 39(2): 185-198. (in Chinese)
- [50] KE Rui-min, LI Wan, CUI Zhi-yong, et al. Two-stream multi-channel convolutional neural network for multi-lane traffic speed prediction considering traffic volume impact[J]. Transportation Research Record, 2020, 2674(4): 459-470.
- [51] ZOU Z, HAO P, LIN L, et al. Deep convolutional mesh RNN for urban traffic passenger flows prediction[C]//IEEE. 2018 SmartWorld/UIC/ATC/ScalCom/CBDCoM/IoP/SCI. New York: IEEE, 2018: 1305-1310.
- [52] WANG Yun-bo, ZHANG Jian-jin, ZHU Hong-yu, et al. Memory in memory: a predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics[C]//IEEE. 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 9146-9154.
- [53] WANG Yun-bo, GAO Zhi-feng, LONG Ming-sheng, et al. PredRNN++: towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning[C]//IMLS. 35th International Conference on Machine Learning. San Diego: IMLS, 2018: 8122-8131.
- [54] ZHENG Hai-feng, LIN Feng, FENG Xin-xin, et al. A hybrid deep learning model with attention-based Conv-LSTM networks for short-term traffic flow prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(11): 6910-6920.
- [55] 彭博,唐聚,蔡晓禹,等. 基于 3DCNN-DNN 的高空视频交通状态预测[J]. 交通运输系统工程与信息, 2020, 20(3): 39-46.
- PENG Bo, TANG Ju, CAI Xiao-yu, et al. 3DCNN-DNN based traffic status prediction from aerial videos[J]. Journal of Transportation Systems Engineering and Information Technology, 2020, 20(3): 39-46. (in Chinese)
- [56] YU Feng, WEI Dan, ZHANG Shu-ting, et al. 3D CNN-based accurate prediction for large-scale traffic flow[C]//IEEE. The 4th International Conference on Intelligent Transportation Engineering (ICITE). New York: IEEE, 2019: 99-103.
- [57] CHEN Yi-bi, ZOU Xiao-feng, LI Ken-li, et al. Multiple local 3D CNNs for region-based prediction in smart cities[J]. Information Sciences, 2021, 542: 476-491.
- [58] ZHANG Shuai-chao, ZHOU Ling-xiao, CHEN Xi-qun, et al. Network-wide traffic speed forecasting: 3D convolutional neural network with ensemble empirical mode decomposition[J]. Computer-Aided Civil and Infrastructure Engineering, 2020, 35(10): 1132 - 1147.
- [59] UL ABIDEEN Z, SUN H L, YANG Z, et al. The deep 3D convolutional multi-branching spatial-temporal-based unit predicting citywide traffic flow[J]. Applied Sciences, 2020, 10(21): 7778.
- [60] CEN C, LI K L, TEO S G, et al. Exploiting spatio-temporal correlations with multiple 3D convolutional neural networks for citywide vehicle flow prediction[C]//IEEE. 2018 IEEE International Conference on Data Mining. New York: IEEE, 2018: 893-898.
- [61] GUO Sheng-nan, LIN You-fang, FENG Ning, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[C]//AAAI. The Thirty-Third AAAI Conference on Artificial Intelligence. Washington DC: AAAI, 2019: 922-929.
- [62] CHEN Xu, ZHANG Yuan-xing, DU Lun, et al. TSSRGCN: temporal spectral spatial retrieval graph convolutional network for traffic flow forecasting[C]//IEEE. 2020 IEEE International Conference on Data Mining. New York: IEEE, 2020: 954-959.
- [63] WU Zong-han, PAN Shi-rui, LONG Guo-dong, et al. Graph WaveNet for deep spatial-temporal graph modeling[C]//AAAI. Proceedings of Twenty-Eighth International Joint Conference on Artificial Intelligence IJCAI-19. Washington DC: AAAI, 2019: 1907-1913.
- [64] BAI Jian-dong, ZHU Jia-wei, SONG Yu-jiao, et al. A3T-GCN: attention temporal graph convolutional network for traffic forecasting[J]. ISPRS International Journal of Geo-Information, 2021, 10(7): 485-496.
- [65] TIAN Yan-ling, ZHANG Qie-shi, REN Zi-liang, et al. Multi-scale dilated convolution network based depth estimation in intelligent transportation systems[J]. IEEE Access, 2019, 7(99): 185179-185188.
- [66] LYU Ming-qi, HONG Zhao-xiong, CHEN Ling, et al. Temporal multi-graph convolutional network for traffic flow prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(6): 3337-3348.
- [67] ZHANG Jian-ni, SHI Xing-jian, XIE Jun-juan, et al. GaAN: gated attention networks for learning on large and spatiotemporal graphs[J]. arXiv, 2018, DOI: 10.48550/arXiv.1803.07294.
- [68] CUI Zhi-yong, LIN Long-fei, PU Zi-yuan, et al. Graph Markov network for traffic forecasting with missing data[J]. Transportation Research Part C: Emerging Technologies, 2020, 117: 102671.
- [69] PARK C, LEE C, BAHNG H, et al. ST-GRAT: a novel spatio-temporal graph attention networks for accurately forecasting dynamically changing road speed[C]//ACM. 9th ACM International Conference on Information and Knowledge Management. New York: ACM, 2020: 1215-1224.
- [70] SONG Chao, LIN You-fang, GUO Sheng-nan, et al. Spatial-temporal synchronous graph convolutional networks: a new framework for spatial-temporal network data forecasting[C]//

- AAAI. The Thirty-Fourth AAAI Conference on Artificial Intelligence. Washington DC: AAAI, 2020: 914-921.
- [71] SUN Jun-kai, ZHANG Jun-bo, LI Qiao-fei, et al. Predicting citywide crowd flows in irregular regions using multi-view graph convolutional networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(5): 2348-2359.
- [72] TANG Shan-shan, LI Bo, YU Hai-jun. ChebNet: efficient and stable constructions of deep neural networks with rectified power units using Chebyshev approximations[J]. arXiv, 2019, DOI:10.48550/arXiv.1911.05467.
- [73] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[C]//ICLR. 5th International Conference on Learning Representations. Washington DC: ICLR, 2017: 1-14.
- [74] LI Y G, YU R, SHAHABI C, et al. Diffusion convolutional recurrent neural network: data-driven traffic forecasting[C]//ICLR. 6th International Conference on Learning Representations. Washington DC: ICLR, 2018: 1-16.
- [75] VELIKOVI P, CUCURULL G, CASANOVA A, et al. Graph attention networks[C]//ICLR. 6th International Conference on Learning Representations. Washington DC: ICLR, 2018: 1-12.
- [76] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs[C]//NIPS. 31st Conference on Neural Information Processing Systems. San Diego: NIPS, 2017: 1-11.
- [77] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. arXiv, 2017, DOI: 10.48550/arXiv.1706.03762.
- [78] BROWN T B, MANN B, RYDER N, et al. Language models are few-shot learners[C]//NIPS. 34th Conference on Neural Information Processing Systems. San Diego: NIPS, 2020: 1-25.
- [79] ARNAB A, DEHGHANI M, HEIGOLD G, et al. ViViT: a video vision transformer[C]//IEEE. 2021 IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2021: 6816-6826.
- [80] LI Y S, BACIU G. SAPCGA N: Self-Attention based generative adversarial network for point clouds[C]//IEEE. 19th IEEE International Conference on Cognitive Informatics and Cognitive Computing. New York: IEEE, 2020: 52-59.
- [81] ZHANG Kun-peng, JIA Ning, ZHENG Liang, et al. A novel generative adversarial network for estimation of trip travel time distribution with trajectory data[J]. Transportation Research Part C: Emerging Technologies, 2019, 108: 223-244.