

基于相互学习的短时交通流预测研究

刘忠伟¹, 李萍², 周盛¹, 闫豆豆², 李颖², 安毅生²

(1. 云基智慧工程有限公司, 广东 深圳 518000; 2. 长安大学 信息工程学院, 西安 710064)

摘要: 交通流预测是智能交通系统 (ITS) 的核心, 其中时空特性是最主要的特征; 由于不同道路之间存在复杂的空间相关性和时间依赖性, 因此交通流预测成为一项具有挑战性的任务; 目前, 基于图卷积神经网络的预测方法在网络局部以及整体的特征感知和提取方面, 仍存在优化空间; 为了解决以上问题, 提出了一种基于图神经网络的优化模型 (DMCRNN), 该模型以 DCRNN 为基准模型, 利用相互学习策略对其进行优化; 在训练过程中, 两个 DCRNN 网络之间相互学习、相互指导, 以此来增强每个网络的特征学习能力; 在 METR-LA 和 PEMS-BAY 两个真实数据集上验证优化策略的有效性; 结果表明, 经过优化后的模型预测误差显著降低, 在两个数据集上一小时的 MAE 与 DCRNN 相比分别降低了 0.15 和 0.12, 即相互学习优化策略具有较好的性能。

关键词: 交通流预测; 时空特性; 图神经网络; 知识蒸馏; 相互学习

Research on Short-term Traffic Flow Prediction Based on Mutual Learning

LIU Zhongwei¹, LI Ping², ZHOU Sheng¹, YAN Doudou², LI Ying², AN Yisheng²

(1. Yunji Smart Engineering Co., Ltd., Shenzhen 518000, China;

2. School of Information Engineering, Chang'an University, Xi'an 710064, China)

Abstract: Traffic flow prediction is the core of intelligent transportation systems (ITS), with spatiotemporal characteristics being the most important feature. Due to the complex spatial correlations and time dependencies between different roads, the traffic flow prediction has become a challenging task. Currently, prediction methods based on graph convolutional neural networks are further optimized on the feature perception and extraction of local and global networks. To address above issues, a diffusion mutual convolutional recurrent neural network (DMCRNN) optimized model based on graph neural networks is proposed. The model is based on the DCRNN as a benchmark model, and the mutual learning strategy is utilized to optimize it. During training, two DCRNN networks learn from and guide each other to enhance their respective feature learning capabilities. The effectiveness of the optimization strategy is verified on two real datasets of the METR-LA and PEMS-BAY. The results show that the optimized model significantly reduces the prediction errors, with a decrease in the MAE of 0.15 and 0.12 for one hour on the two datasets than that in the DCRNN, respectively, indicating that the mutual learning optimization strategy has a good performance.

Keywords: traffic flow prediction; spatiotemporal characteristics; graph convolutional neural networks; knowledge distillation; mutual learning

0 引言

随着城市机动车保有量的不断增加, 城市道路日益拥挤。为改善这一状况, 交通管理部门提出了利用智能交通系统 (ITS, intelligent transportation system) 缓解拥堵问题。其中, 交通流预测是智能交通系统的核心之一。通过收集历史数据, 可以预测未来某个时间段内的交通状况, 包括速度、流量和占有率等。根据预测时长^[1], 交通流预测可分为短期预测 (5~30 min) 和中长期预测 (超过 30 min)。交通流具有明显的时空特性, 不同道路之间复杂的空间相关性和时间依赖性给交通流预测带来了挑战。为了解决这些挑战, 过去几十年中涌现了许多交通流预测方法。最初, 线性回归等统计分析模型被广泛应用, 但是统计理论模型属于线性模

型, 在解决非线性问题时存在不足。受限于统计分析模型的不足, 机器学习的发展促使研究者探索支持向量机 (SVM, support vector machine)、K 最近邻 (K-Nearest Neighbors) 等基于数据驱动的方法进行交通预测。

1 交通流预测研究现状

近年来, 国内外对于交通流预测的方法大致可分为 3 大类: 统计分析模型、机器学习模型和深度学习模型。

1) 基于统计分析的预测模型: 统计分析模型利用概率论和数理统计等相关数学理论建模, 主要有历史平均模型 (HA, historical average)^[2]、差分整合移动平均自回归模型 (ARIMA, autoregressive integrated moving average model)^[3]、卡尔曼滤波器 (Kalman filtering)^[4] 等。历史平

收稿日期: 2023-04-26; 修回日期: 2023-06-05。

基金项目: 国家自然科学基金青年项目 (52002031); 国家自然科学基金面上项目 (52172325); 国家重点研发 (2021YFB1600104)。

作者简介: 刘忠伟 (1979-), 男, 大学本科, 高级工程师。

通讯作者: 李颖 (1986-), 女, 博士, 副教授。

引用格式: 刘忠伟, 李萍, 周盛, 等. 基于相互学习的短时交通流预测研究[J]. 计算机测量与控制, 2024, 32(4): 166-173.

均模型计算过程简单、速度快,但是模型是静态的,预测的准确性较低。ARIMA模型的主要目标是将时间序列数据转换为平稳序列,以便进行数学建模和预测,因此它对数据平稳性要求较高,主要适用于线性序列预测,但是交通流数据的不平稳性和多种非线性因素可能会降低ARIMA的预测性能。基于统计分析的模型在早期的交通流预测中表现较好,但它们假设交通流是静止的,属于线性预测,仅从概率的角度考虑空间依赖性,实际的交通流受多种因素影响,具有明显的非线性和复杂的时空依赖性。

2) 基于机器学习的预测模型:随着交通数据量的急剧增长以及神经网络的发展,研究者们开始使用机器学习进行交通流预测,相比于统计分析模型需要手工调参,机器学习模型比较灵活,可以在遍历数据集时可以自动调整参数。常见的机器学习预测模型有:K近邻(KNN, K-nearest neighbors)和支持向量回归(SVR, support vector regression)等。K最近邻算法是选取目标样本所在特征空间K个最近的样本,所选取K个样本数中所占数量最大的类别即为目标样本所属类别。Zheng等人^[5]提出了一种改进的KNN模型,在选择K近邻时引入了时间约束窗口并且对状态向量之间距离的局部极小值进行排序,为控制极值的影响,文章在主成分分析基础上,还提出了一种新的分析特征的算法,利用该模型来预测交通流。Hou等人^[6]引入了两层KNN算法模型对模式匹配过程做了改进,增加了状态模式匹配使得当前状态向量与历史样本趋势和最小。支持向量机是一种二分类算法,算法思想是求解凸二次规划的最优化。Wu等人^[7]应用SVR预测旅行时间,与其他预测模型相比,SVR可以提升预测性能。

3) 基于深度学习的预测模型:相比于机器学习,深度学习模型可以更好地挖掘数据隐藏的特性,处理复杂的非线性问题,目前,随着神经网络以及深度学习的快速发展,越来越多的研究者将其应用到交通邻域,以其强大的学习能力去提取交通大数据隐藏的时空特性,可以较为准确地预测未来时刻的交通状况。

在提取时间特征方面,循环神经网络(RNN, recurrent neural network)具有较好的性能。Ma等人^[8]在真实数据集上利用RNN进行拥堵预测,但是RNN在建模较长序列时会产生梯度爆炸等一系列问题。Qu等人^[9]将特征注入RNN,形成Feature Injected RNN(FI-RNNs),用堆叠的RNN去提取交通数据的序列特征,并将其与上下文因素结合,进一步挖掘交通数据潜在的时间特性。RNN的变体长短期记忆(LSTM, long-short term memory)和门控循环单元(GRU, recurrent neural network)可以有效解决模型训练过程中梯度爆炸等问题,阎嘉琳等人^[10]提出基于LSTM的城市交通道路预测模型,并在出租车GPS轨迹数据集上进行验证。Zhang等人^[11]考虑了天气因素的影响,使用GRU进行城市交通流预测。

在提取空间特征方面,卷积神经网络(CNN, convolution neural network)具有强大的特征提取能力。Ma等

人^[12]提出了一种基于图像的交通速度预测模型,将GPS轨迹数据作为输入,通过矩阵变换将交通路网转化为具有时空信息的图像,图像的两个维度分别表示路网的时间维度和空间维度,然后使用CNN提取图像特征进行交通流预测。Liu等人^[13]提出了Conv-LSTM,将卷积和LSTM组合形成一种端到端的结构提取交通流的时空特征,然后使用双向LSTM(Bi-LSTM)提取交通流的周期性特征。但是交通路网是一种典型的拓扑结构数据,将其处理为网格数据后使用CNN,忽略了路网本身的特征,因此研究者们尝试将CNN泛化到图域,形成图卷积神经网络GCN来处理交通路网、社交网络等拓扑结构数据。Li等人^[14]提出了一种扩散卷积循环神经网络(DCRNN, diffusion convolutional recurrent neural network),使用GCN和GRU共同提取交通流的时空相关性,首先在图上使用双向随机游走捕获空间依赖性,然后使用GRU提取时间特征。Yu等人^[15]提出了时空卷积网络(STGCN, spatio-temporal graph convolutional networks),该模型由时空卷积块堆叠而成,利用门控时间卷积和空间卷积提取出最基本的时间特征以及最相关的空间特征。由于GCN强大的空间特征提取能力,研究者们还尝试将注意力机制和Transformer等神经网络与GCN结合来预测交通流。Zheng等人^[16]提出了一种图上的多注意力网络(GMAN, graph multi-attention network),由多个注意力模块堆叠而成,包括对空间特性建模的空间注意力模块、对时间特性建模的时间注意力模块以及自适应融合时空特征的融合机制。空间注意力机制根据邻近节点对目标节点影响的大小分配不同的权值,时间注意力机制根据预测点历史时刻的影响不同分配不同的权值。Tian等人^[17]提出了一种包含时间注意力机制的深度学习网络框架解决城市交通事故预测问题,实现了较好的预测性能。此外,Cai等人^[18]将Transformer应用到图上来建模时间相关性,并且设计了4种新的位置编码策略编码时间序列的连续性和周期性。Li等人^[19]使用多目标粒子群优化算法优化深度信念网络的一些参数,增强了交通流的多步预测能力。

2 知识蒸馏相关理论

知识蒸馏是一种有效的模型压缩方法,其核心思想是将来自复杂教师模型的知识迁移到简单的学生模型上,从而提高学生模型的泛化能力。教师模型通常拥有较强的预测精度,但其训练过程比较复杂;相反,学生模型参数较少,训练过程较为简单,但其预测能力相对较弱。因此,知识蒸馏技术可以通过教师网络输出的软标签来监督学生网络的学习,从而提升学生模型的预测性能。根据蒸馏的方式和知识类型,知识蒸馏可分为离线蒸馏、在线蒸馏以及自蒸馏。在离线蒸馏中,教师模型经过预先训练,并生成软标签供学生模型使用。基于响应的知识蒸馏、基于特征的知识蒸馏以及基于关系的知识蒸馏则是离线蒸馏中按知识类型划分的3种主要方法。其中,基于响应的知识蒸

馏主要侧重于对教师网络最后一层的逻辑输出进行处理, 让学生网络去模仿教师网络的输出。2014 年 Hithon^[20]首次提出了基于响应的知识蒸馏, 同时也提出了蒸馏温度 T 的概念, 在高温 T 下, 学生网络模仿教师网络的输出, 蒸馏温度则是用来软化教师网络的软目标输出, 当温度系数值不断增大, 得到的输出标签更加平滑。基于特征的知识蒸馏, 其核心思想是学习教师网络层与层之间的关系, 从而提升学生网络的性能, Romero 等人^[21]提出的 FitNets 表示学生能够直接模仿教师网络的特征激活值。基于特征的知识蒸馏可以在学生网络学习时提供更多的信息, 但是由于教师网络与学生网络中间层存在维度差异, 以及在两个网络上如何选取对应中间层操作较为困难。Park 等人^[22]提出了基于关系的知识蒸馏, 重点关注模型内部特征之间的关系, 而不仅仅是模型输出。这种蒸馏方法方法认为捕捉和模仿教师模型内部特征之间的相互作用和依赖关系, 可以帮助学生模型更全面地理解任务, 从而学到更丰富的知识。

传统的离线蒸馏存在局限性, 需要提前训练一个复杂的教师网络, 会耗费大量的时间和计算资源, 且学生网络不能学习教师网络所蕴含的所有知识, 为了解决这些问题, 研究者在传统知识蒸馏的基础上提出了在线知识蒸馏, 即无需提取训练复杂的教师模型, 教师模型和学生模型在学习过程中互相更新。其中以 Zhang 等人^[23]提出的深度相互学习策略 (DML, deep mutual learning) 为代表, 两个或多个学生模型可以在训练过程中互相学习, 互相监督, 打破了传统的知识蒸馏预先设定好的强弱关系, 无需提前训练复杂的教师模型, 将传统的单向知识迁移更新为双向迁移, 即训练过程中两个或多个模型互为教师, 互为学生。Zhai 等人^[24]将相互学习策略泛化到图上, 提出了图相互学习模型 (MGL, mutual graph learning) 将相互学习由规则网格推广到图域。

自蒸馏是一种特殊的在线蒸馏, 指在没有新增复杂模型的情况下找到一个教师模型, 同样可以提供有效增益给学生模型, 意味着教师网络与学生网络应用同一个主体网络, 或者直接将自身的信息作为蒸馏损失的监督, 是模型内部的知识迁移。2019 年, Zhang 等人^[25]提出了一种自蒸馏的方式, 用模型的深层部分蒸馏和监督浅层部分。2021 年, Kim^[26]等人提出了渐进式自蒸馏 (PS-KD, progressive self-knowledge distillation), 用前边几轮的模型去蒸馏后边几轮模型训练。

综上所述, 知识蒸馏是一种通过知识迁移和蒸馏损失来提高模型预测精度的方法。这种方法可以满足交通流预测对于精准性的高要求。因此, 在本文中, 采用知识蒸馏技术对交通流预测模型进行了优化。

3 模型架构

3.1 预测问题定义

交通流预测是根据部署在道路上 N 个检测器预先收集到的数据来预测未来时刻的交通状况^[27]。交通路网可以定义为 $G = (V, E, W)$, 其中 V 为节点的集合, E 为两个节

点之间边的集合, $W \in R^{N \times N}$ 对应图 G 的邻接矩阵。实际应用中, 节点即为部署在道路上的检测器, 收集到的交通信息包括速度、流量、占有率等。对于一个交通网络, $X_t \in R^{N \times C}$ 表示在 t 时间步交通网络中所有节点的特征的集合, C 表示特征的数目。通过给定 M_0 个历史时刻的交通数据去测未来 M 个时刻内所有节点的交通特征, 即交通流预测问题可定义为:

$$[X^{(m-M_0+1)}, \dots, X^{(m)}; G] \xrightarrow{f(\cdot)} [X^{(m+1)}, \dots, X^{(m+M)}] \quad (1)$$

其中: $f(\cdot)$ 是一个经过训练的可学习的函数, 可以将历史时刻的交通数据映射到未来时刻。

3.2 DCRNN 模型

本文以 DCRNN 作为基准模型, 将交通流建模为有向图上的扩散过程, 通过图上的随机游走捕捉交通流的空间特征, 使用门控单元 GRU 提取时间特征, 将序列到序列框架和预定采样相结合进行交通流预测。图 1 即为 DCRNN 的网络架构, 是一种编码器-解码器结构, 其中编码器的功能是将一个长度不固定的输入序列转化成一个固定长度的上下文变量, 在该变量中编码输入序列信息; 解码器将该变量中的信息进行解码生成一个输出序列。

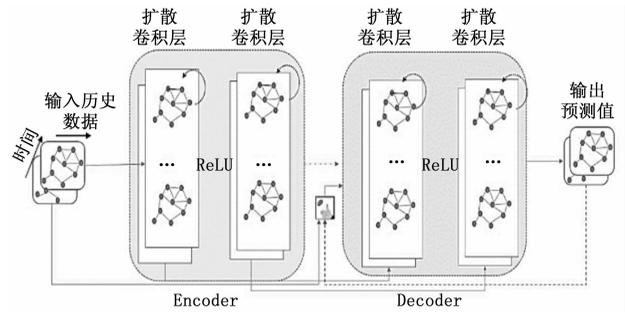


图 1 DCRNN 模型图

首先对历史交通流数据集按照时间步长进行划分, 将路网建模为有向图, 图中节点表示部署在道路上的检测器。然后将经过划分处理后的数据作为解码器的输入, 解码器由扩散卷积层堆叠而成, 以此来提取图中各个节点之间隐藏的空间相关性。使用门控循环单元 GRU 提取时间依赖性, 这是一种具有强大功能的循环神经网络 RNN 的变体, 用扩散卷积代替 GRU 中的矩阵相乘, 产生了扩散卷积门控循环单元 DCGRU。编码器的最终输出作为解码器的输入, 解码器的结构与编码器类似, 进行最终的预测输出。为了解决训练过程和测试过程中输入数据分布不一致的问题, DCRNN 在训练过程中采用计划采样, 网络将以概率 p 选择真实值作为下一步的输入, 以 $1-p$ 的概率选择模型自身的输出作为下一步的输入。

3.3 基于相互学习策略的模型优化

DCRNN 模型能够有效地提取交通数据中隐藏的时空相关性, 并预测未来交通状况。然而, 该模型在自身特征感知和提取方面存在一定的局限性。本文受到知识蒸馏中相互学习策略的启发, 对 DCRNN 进行了优化。利用两个或多个网络在训练过程中相互学习, 相互协同完成知识的传

递, 并在不增加模型复杂度的前提下优化模型的预测性能。为了降低训练过程中的参数量, 只使用两个未经训练的 DCRNN 模型构建基于相互学习的预测模型。首先将处理后的数据输入两个模型, 分别经过两个模型的 softmax 层输出预测值, 使用 KL (Kullback-Leibler Divergence, 用于计算不同概率分布之间的差异) 散度衡量两个网络预测值的匹配度, KL 值越小, 则表示两个网络之间的匹配度越高。除 KL 值之外, 使用交叉熵损失 (CE, cross entropy) 来衡量每个模型的预测值与真实值的差异。综上, 在搭建的预测模型中, 每个学生网络的损失由两部分组成: 传统的监督损失 CE 以及与同伴网络的匹配损失 KL。图 2 是优化模型 DMCRNN (Diffusion Mutual Convolutional Recurrent Neural Network) 的结构, 其中 P_1 和 P_2 分别表示两个网络在训练过程中输出的预测概率, 利用 $KL(P_2 || P_1)$ 计算两个网络预测概率之间的差异, KL 值越小, 表示两个预测值之间的差异越小。 P_1 与 P_2 之间 KL 散度计算公式如下:

$$D_{KL}(P_2 || P_1) = P_2 \log \frac{P_2}{P_1} \quad (2)$$

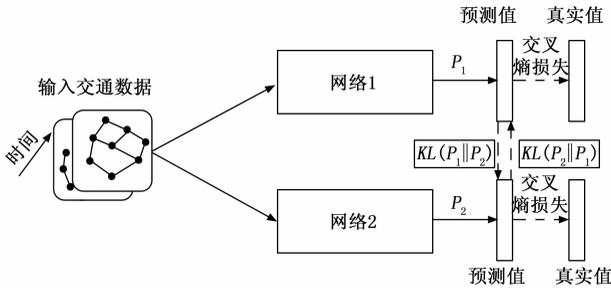


图 2 基于相互学习的预测模型

即网络 1 的损失为相互学习损失以及自身损失以权重系数进行叠加, 可表示为:

$$L_1 = L_{C1} + D_{KL}(P_2 || P_1) \quad (3)$$

同理可得网络 2 的损失:

$$L_2 = L_{C2} + D_{KL}(P_1 || P_2) \quad (4)$$

模型在每一次迭代过程中均采用相互学习算法, 对 DCRNN 模型进行优化, 两个网络分别利用对方的预测值去更新各自的参数, 直至收敛。不同于传统的知识蒸馏为单向蒸馏, 即知识仅从教师网络迁移到学术网络, 而相互学习是一种双向蒸馏, 两个网络在训练过程中不断更新优化, 直到参数收敛。基于相互学习的 DCRNN 模型优化算法框架如表 1 所示, 经过优化后可以提升 DCRNN 模型的预测性能。

4 实验

为了验证基于相互学习算法预测模型的有效性, 该研究在两个真实数据集 (METR-LA 和 PEMS-BAY) 上进行

实验并与其他几种基准模型进行对比。

算法: 基于相互学习的交通流预测算法

输入: 训练数据, 学习率 r

初始化: DCRNN 网络 Net_1 和 DCRNN 网络 Net_2

$t=0$

do: $t=t+1$ 随机抽取数据:

从 Net_1 中计算预测值 y_1 和真实值 y_{true} ;

从 Net_2 中计算预测值 y_2 和真实值 y_{true} ;

计算 Net_1 的随机梯度并更新网络参数:

a. 计算误差 $E_1 = (y_1 - y_{true})^2$;

b. 计算 Net_1 的梯度 dE_1/dw_1 ;

c. 更新 Net_1 的参数 $w_{1_new} = w_1 - r * dE_1/dw_1$;

更新 Net_1 的预测值 y_1 ;

计算 Net_2 的随机梯度并更新网络参数:

a. 计算误差 $E_2 = (y_2 - y_{true})^2$;

b. 计算 Net_2 的梯度 dE_2/dw_2 ;

c. 更新 Net_2 的参数 $w_{2_new} = w_2 - r * dE_2/dw_2$;

更新 Net_2 的预测值 y_2

While: 目标函数未收敛

4.1 数据预处理

METR-LA: 该数据集采样于美国洛杉矶高速公路上的 207 个传感器, 其中包含 2012 年 3 月 1 日到 2012 年 6 月 7 日所收集到的交通速度等信息。数据采集频率为 5 min 一次, 包含 34 272 条交通数据。

PEMS-BAY: 该数据集来源于加州高速公路, 包含湾区 325 个检测器, 数据采集频率为每五分钟一次, 包含从 2017 年 1 月 1 日到 2017 年 5 月 31 日所收集到的流量、速度等信息, 总共有 52 116 条交通数据。

如表 1 所示, 本文主要使用两个数据集中的速度信息, 按照 7 : 2 : 1 的比例划分数据集, 其中训练集占 70%, 测试集占 20%, 验证集占 10%。为了提高模型在梯度下降时收敛速度, 对数据集采用 Z-Score 来标准化:

$$z = \frac{x - \mu}{\sigma} \quad (5)$$

其中: z 表示归一化后的数据, x 表示原始输入数据, μ , σ 分别表示均值和方差。

构建图的邻接矩阵是交通流预测模型提取空间相关性的关键, DCRNN 模型采用基于距离的方法建立图的邻接矩阵, 利用交通路网上检测器之间的距离来计算节点之间的紧密程度, 加权邻接矩阵计算公式为:

$$W_{ij} = \exp \left[-\frac{dist(v_i, v_j)^2}{\sigma^2} \right] \quad (6)$$

其中: $dist(v_i, v_j) \leq \kappa$, W_{ij} 表示两个检测器 v_i 和 v_j 所连接的边的权值, $dist(v_i, v_j)$ 表示 v_i 和 v_j 在路网上的距离, σ 表示距离的标准差, k 表示阈值。

表 1 数据集基本信息

数据集	节点数	时间步长	时间跨度	采样间隔	训练集	验证集	测试集
METR-LA	307	34 373	2013-03-01~2013-06-30	5 min	33 974	3 435	6 850
PEMS-BAY	325	52 116	2017-01-01~2017-05-31	5 min	36 465	5 209	10 419

4.2 实验设置

基于 Pytorch 框架训练 DMCRNN 模型，训练时采用内含 4 张型号为 GeForce RTX 3090 显卡的服务器，具体的环境配置以及参数设置如表 2、表 3 所示。

表 2 硬件条件

名称	型号
操作系统	Ubuntu18.04
CPU	Intel(R) Xeon(R) Gold 5213W CPU @ 2.70 GHz
GPU	GeForce RTX 3090
CUDA	11.0
Python 版本	3.7.11
Pytorch 版本	1.7.1

表 3 实验参数设置

RNN 层数	2	优化器	Adam
RNN 神经元数目	64	初始学习率	0.001
batchsize	64	最大扩散步长	2
训练 epoch	100	蒸馏温度 T	8

实验过程中利用 1 小时的历史数据去预测未来 1 小时内的交通状况。并且训练过程中利用 dropout 和 early stopping 机制防止模型过拟合，随机扩散时最大扩散步长为 2。

4.3 评估指标

在评估优化后的模型预测性能时，采用如下 3 个指标评估真实值与预测值之间的差异：

1) 平均绝对误差 (MAE, mean absolute error)，可以直观显示出真实值与预测值之间的偏差，直观反映模型的预测结果，计算公式如下：

$$MAE = \frac{1}{n} \sum_{i=1}^n \sum_{i=1}^n |y_i - f(x_i)| \quad (7)$$

2) 平均绝对百分比误差 (MAPE, mean absolute percentage error)，可以评价真实值与预测值之间的符合程度，反映模型预测的准确程度，计算公式如下：

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - f(x_i)}{y_i} \right| \quad (8)$$

3) 均方根误差 (RMSE, root mean square deviation)，可以反映出真实值与预测值差值的变化，反映模型预测的精确程度。计算公式如下：

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [y_i - f(x_i)]^2} \quad (9)$$

其中： n 为时间步长， y_i 和 $f(x_i)$ 分别表示第 i 个时间步的真实值和预测值，3 个指标的数值越小，表示模型的预测精度越高。

4.4 基准模型介绍

为验证优化算法的有效性，我们在两个真实的数据集上进行了实验，并将实验结果与目前具有代表性的 7 种模型进行了比较。

HA^[2]：历史平均模型，即计算历史数据的平均值。

ARIMA^[3]：差分整合移动平均自回归模型，一种预测未来值的时间序列分析方法。

VAR^[28]：向量自回归模型，可以捕获交通流序列之间的成对关系，是一种更先进的时间序列模型。

SVR^[7]：支持向量机模型，利用支持向量机来解决回归问题，是一种线性模型。

FNN^[29]：前馈神经网络模型，其中包含多个隐藏层。

LSTM^[10]：长短期记忆网络，属于 RNN 的变体，在长时间序列问题上表现较好。

DCRNN^[14]：扩散卷积递归神经网络，利用扩散卷积网络和 GRU 分别提取空间特征和时间特征。

4.5 实验结果分析

我们在两个真实数据集上进行实验，根据预测结果，利用上述 3 个评估指标对预测性能进行分析。表 4 和表 5 分别展示了 DCMRNN 模型与其他基准模型在两个数据集上的性能比较，我们选取了 15 min，30 min 和 1 h 这 3 个时间步长进行比较。从表中可以看出本文提出的模型在两个数据集上的 3 种误差均是最低的，即预测结果更加接近真实值。尤其是随着时间步长的增加，其预测效果更好。从表 3 和表 4 中可以看出随着预测时间步长的增加，DMCRNN 模型与其他模型三项指标之间的差值越大，表明了该模型在较长时间步预测时仍具有较好的性能，模型具有鲁棒性。在 METR-LA 数据集上 MAE 值相较于 DCRNN 下降了 0.15，MAPE 下降了 0.4%，RMSE 下降了 0.55；在 PEMS-BAY 数据集上 MAE 值相较于 DCRNN 下降了 0.12，MAPE 下降了 0.2%，RMSE 下降了 0.31，表明增加相互学习算法后的模型优化效果显著。

为了进一步比较 DMCRNN 模型与基准模型 DCRNN，

表 4 DCMRNN 与基准模型在 METR-LA 数据集上的性能比较

方法	15 min			30 min			1 h		
	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE
HA	4.16	13.0	7.8	4.16	13.0	7.8	4.16	13.0	7.80
ARIMA	3.99	9.6	8.21	5.15	12.7	10.45	6.90	17.4	13.23
VAR	4.42	10.2	7.89	5.41	12.7	9.13	6.52	15.8	10.11
SVR	3.99	9.3	8.45	5.05	12.1	10.87	6.72	16.7	13.76
FNN	3.99	9.9	7.94	4.23	12.9	8.17	4.49	14.0	8.69
LSTM	3.44	9.6	6.3	3.77	10.9	7.23	4.37	13.2	8.69
DCRNN	2.77	7.3	5.38	3.15	8.8	6.45	3.60	10.5	7.59
DMCRNN	2.70	7.1	5.15	3.05	8.5	6.08	3.45	10.1	7.07

表 5 DCMRNN 与基准模型在 PEMS-BAY 数据集上的性能比较

方法	15 min			30 min			1 h		
	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE
HA	2.88	6.8	5.59	2.88	6.8	5.59	2.88	6.8	5.59
ARIMA	1.62	3.5	8.21	2.33	5.4	4.76	3.38	8.3	6.50
VAR	1.74	3.6	3.16	2.32	5.0	4.25	2.93	6.5	5.44
SVR	1.85	3.8	3.59	2.48	5.5	5.18	3.28	8.0	7.08
FNN	2.20	5.2	4.42	2.30	5.4	4.63	2.46	5.9	4.98
LSTM	2.05	4.8	4.19	2.20	5.2	4.55	2.37	5.7	4.96
DCRNN	1.38	2.9	2.95	1.74	3.9	3.97	2.07	4.9	4.74
DCMRNN	1.33	2.8	2.79	1.66	3.8	3.70	1.95	4.7	4.43

图 3、4 和 5 展示了两个模型在 METR-LA 数据集的 3 种评价指标以及损失可视化。图 3、4 和 5 中灰色线条表示基准模型 DCRNN 在验证过程中的指标变化, 黑色线条表示我们的模型 DCMRNN 的指标变化, 可以看出我们的模型相比于 DCRNN, 误差可以收敛到更低。图 6 为两个模型训练过程中损失曲线可视化图, 其中我们提出的模型 DCMRNN 收敛速度更快。图 7 是两个模型在测试集上 RMSE 的比较, 横坐标表示时间步长, 纵坐标表示 RMSE, 可以看出随着预测时间步长的增加, 两个模型之间的差异变大, 即我们提出的模型在较长时间步时表现较好。

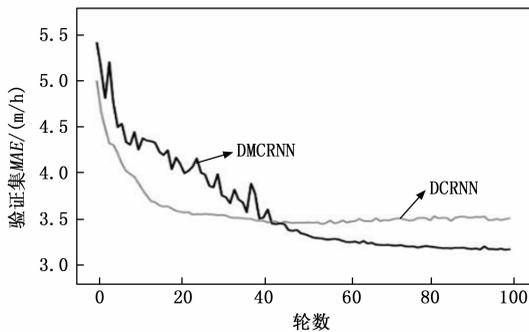


图 3 验证集 MAE 可视化曲线图

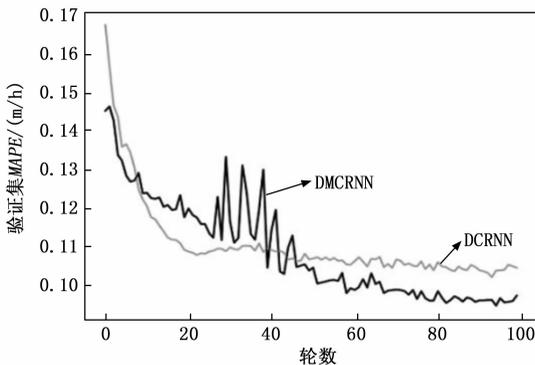


图 4 验证集 MAPE 可视化曲线

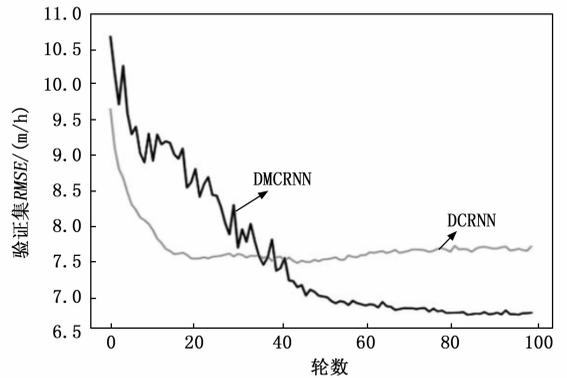


图 5 验证集 RMSE 可视化曲线图

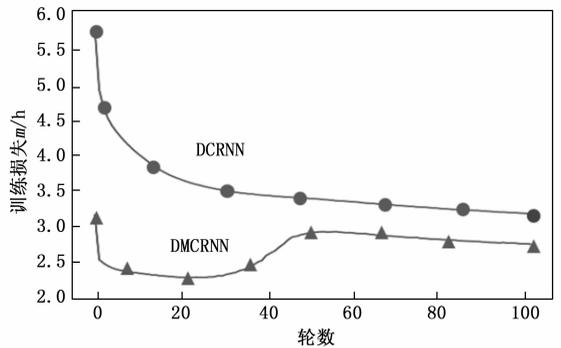


图 6 训练集损失可视化

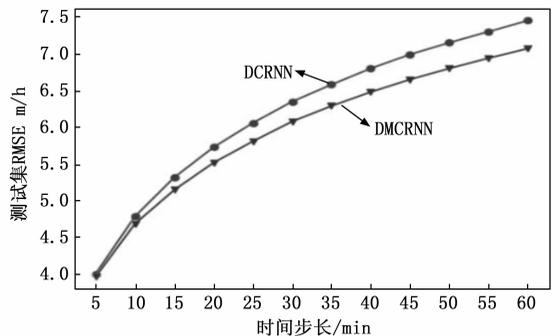


图 7 两个模型在不同时间步长上的 RMSE 比较

4.6 实验参数分析

基于相互学习优化过程中有两个重要的超参数, 第一个是不同温度系数对蒸馏结果的影响, 第二个是不同的权重系数对两部分损失函数的影响。这两个参数都来源于损

失函数, 损失函数是模型性能的关键之一, 温度 T 表示蒸馏温度, 来源于 KL 散度的计算过程中, 会对网络的性能

产生影响；而参数 α 表示两种损失在总损失中的比重。

实验过程中观察到 $T=8$ 时验证集上 MAE 以及损失曲线最终收敛的值比其他温度更低，因此选用 $T=8$ 作为基准蒸馏温度；将 α 系数设置为 $[0.1, 0.3, 0.5, 0.7, 0.9, 0.99]$ 进行实验，结果显示当 α 等于 0.9 或者 0.1 时，效果最差。 $\alpha=0.9$ 表示相互学习损失权重过大，网络自身的损失权重过低，极大影响了模型的预测性能。 $\alpha=0.1$ 表示相互学习损失权重过低，网络之间相互学习能力变弱，同样影响了预测性能。其他 3 个不同的权重系数对模型的影响整体相差不大，其中 $\alpha=0.5$ 在验证集上 MAE 以及损失曲线最终收敛的值较其他两个权重系数的曲线收敛结果更低一点。最终选取 $\alpha=0.5$ 作为最佳权重系数。

4.7 相互学习可扩展性实验

为了验证相互学习算法的普适性，实验中选取了 Shang 等人^[30]提出的 (GTS, graph for time series) 模型进行验证，GTS 是一种新的图优化方法，当图未知的时候，该方法可学习图结构以增强对多个多元时间序列的预测。GTS 也是以 DCRNN 为基准模型，通过优化图结构从而大幅度提升了预测性能。因此将 GTS 作为基准模型，增加相互学习算法，使两个 GTS 网络作为学生模型互相学习。同样在 METR-LA 和 PEMS-BAY 两个数据集上进行实验，表 7 选取了 1 小时的预测结果进行对比，实验结果可以看出，在增加相互学习算法的 GTS 模型预测误差相比于原来均有所降低，尤其是在 METR-LA 数据集上，再次证明了相互学习算法的有效性。

表 7 扩展实验结果

GTS	DML	METR-LA			PEMSBAY		
		MAE	MAPE/%	RMSF	MAE	MAPE/%	RMSF
✓		3.44	10.0	7.28	1.95	4.6	4.46
✓	✓	3.37	10.0	6.98	1.93	4.5	4.43

5 结束语

本文提出了一种基于相互学习策略的交通流预测优化模型 (DMCRNN)。该模型以 DCRNN 为基准模型，引入相互学习算法，并在两个真实数据集上进行了实验，取得了较好的结果。与其他基准模型进行对比证明了相互学习策略可以有效提高模型的预测性能。同时，设计了扩展性实验验证相互学习在其他模型上的效果，证明了相互学习的普适性。总之，优化后的模型表现良好，但训练过程中会生成两个模型，并联运算，导致训练时间相比基准模型较长，并且会生成两份优化参数。然而，在实际应用中只需要其中一个模型的优化参数。因此，未来工作可以考虑通过优化模型之间的并联方式来减少训练时间。

参考文献:

[1] TEDJOPURNOMO D A, BAO Z, ZHENG B, et al. A survey on modern deep neural network for traffic prediction; Trends, methods and challenges [J]. IEEE Transactions on Knowledge

and Data Engineering, 2020, 34 (4): 1544 - 1561.

[2] STEPHANEDES Y J, MICHALOPOULOS P G, PLUM R A. Improved estimation of traffic flow for real-time control (discussion and closure) [J]. Transportation Research Record Journal of the Transportation Research Board, 1980, 795: 28 - 39.

[3] WILLIAMS B M, HOEL L A. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results [J]. Journal of Transportation Engineering, 2003, 129 (6): 664 - 672.

[4] OKUTANI I, STEPHANEDES Y J. Dynamic prediction of traffic volume through Kalman filtering theory [J]. Transportation Research Part B Methodological, 1984, 18 (1): 1 - 11.

[5] ZHENG Z, SU D. Short-term traffic volume forecasting: A k-nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm [J]. Transportation Research Part C Emerging Technologies, 2014, 43: 143 - 157.

[6] HOU X, WANG Y, HU S. Short-term traffic flow forecasting based on two-tier k-nearest neighbor algorithm [J]. Procedia-Social and Behavioral Sciences, 2013, 96 (1): 2529 - 2536.

[7] WU C H, HO J M, LEE D T. Travel-time prediction with support vector regression [J]. IEEE Transactions on Intelligent Transportation Systems, 2004, 5 (4): 276 - 281.

[8] MA X, YU H, WANG Y, et al. Large-scale transportation network congestion evolution prediction using deep learning theory [J]. PloS One, 2015, 10 (3): e0119044.

[9] QU L, LYU J, LI W, et al. Features injected recurrent neural networks for short-term traffic speed prediction [J]. Neurocomputing, 2021, 451: 290 - 304.

[10] 阎嘉琳, 向隆刚, 吴华意, 等. 基于 LSTM 的城市道路交通速度预测 [J]. 地理信息世界, 2019, 26 (5): 79 - 85.

[11] ZHANG D, KABUKA M R. Combining weather condition data to predict traffic flow: a GRU-based deep learning approach [J]. IET Intelligent Transport Systems, 2018, 12 (7): 578 - 585.

[12] MA X, DAI Z, HE Z, et al. Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction [J]. Sensors, 2017, 17 (4): 818.

[13] LIU Y, ZHENG H, FENG X, et al. Short-term traffic flow prediction with Conv-LSTM [C] //2017 9th International Conference on Wireless Communications and Signal Processing (WCSP). IEEE, 2017: 1 - 6.

[14] LI Y, YU R, SHAHABI C, et al. Diffusion convolutional recurrent neural network: data-driven traffic forecasting [J]. ArXiv E-prints, 2017: ArXiv: 1707.01926.

[15] YU B, YIN H, ZHU Z. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting [C] //Proceedings of the 27th International Joint Conference on Artificial Intelligence, 2018: 3634 - 3640.

[16] ZHENG C, FAN X, WANG C, et al. Gman: A graph multi-attention network for traffic prediction [C] //Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34

- (1): 1234 - 1241.
- [17] TIAN Z, ZHANG S. Deep learning method for traffic accident prediction security [J]. *Soft Computing*, 2022, 26 (11): 5363 - 5375.
- [18] CAI L, JANOWICZ K, MAI G, et al. Traffic transformer: Capturing the continuity and periodicity of time series for traffic forecasting [J]. *Transactions in GIS*, 2020, 24 (3): 736 - 755.
- [19] LI L, QIN L, QU X, et al. Day-ahead traffic flow forecasting based on a deep belief network optimized by the multi-objective particle swarm algorithm [J]. *Knowledge-Based Systems*, 2019, 172: 1 - 14.
- [20] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network [J]. *Arxiv E-Prints*, 2015: arXiv: 1503.02531.
- [21] ROMERO A, BALLAS N, EBRAHIMI K S, et al. FitNets: Hints for Thin DeepNets [J]. *Arxiv E-Prints*, 2014: arXiv: 1412.6550.
- [22] PARK W, KIM D, LU Y, et al. Relational knowledge distillation [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 3967 - 3976.
- [23] ZHANG Y, XIANG T, HOSPEDALES T M, et al. Deep mutual learning [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 4320 - 4328.
- [24] ZHAI Q, LI X, YANG F, et al. Mutual graph learning for
- (上接第 165 页)
- [6] 李智超, 姚金杰, 孙兴丽, 等. 基于自适应滤波的北斗弱干扰源定位方法 [J]. *国外电子测量技术*, 2022, 41 (10): 33 - 38.
- [7] 张威, 马宏, 吴涛, 等. 一种基于泰勒级数展开的卫星 FDOA 地面干扰源定位算法 [J]. *无线电通信技术*, 2019, 45 (4): 385 - 390.
- [8] TIRER T, WEISS A J. High resolution direct position determination of radio frequency sources [J]. *IEEE Signal Processing Letters*, 2016 23 (2): 192 - 196.
- [9] 郝克钢. 基于参数网格优化的辐射源直接定位方法研究 [D]. 成都: 电子科技大学, 2021.
- [10] WEISS A J. Direct position determination of narrowband radio frequency transmitters [J]. *IEEE Signal Processing Letters*, 2004, 11 (5): 513516.
- [11] WEISS A J. Direct geolocation of wideband emitters based on delay and Doppler [J]. *IEEE Transactions on Signal Processing*, 2011, 59 (6): 25132521.
- [12] TIRER T, WEISS A J. High resolution localization of narrowband radio emitters based on Doppler frequency shifts [J]. *Signal Processing*, 2017, 141: 288298.
- [13] 王鼎, 张刚. 一种基于窄带信号多普勒频率测量的运动目标直接定位方法 [J]. *电子学报*, 2017, 45 (3): 591598.
- [14] 刘清, 谢坚, 王伶, 等. 卫星导航欺骗式干扰源高精度直接定位方法 [J]. *电子学报*, 2022, 50 (5): 1117 - 1122.
- camouflaged object detection [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 12997 - 13007.
- [25] ZHANG L, SONG J, GAO A, et al. Be your own teacher: Improve the performance of convolutional neural networks via self distillation [C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 3713 - 3722.
- [26] KIM K, JI B M, YOON D, et al. Self-knowledge distillation with progressiverefinement of targets [C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 6567 - 6576.
- [27] TEDJOPURNOMO D A, BAO Z, ZHENG B, et al. A survey on modern deep neural network for traffic prediction: Trends, methods and challenges [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 34 (4): 1544 - 1561.
- [28] HAMILTON J D. *Time series analysis* [M]. Princeton University Press, 2020.
- [29] MORE R, MUGAL A, RAJGURE S, et al. Road traffic prediction and congestion control using artificial neural networks [C] // *2016 International Conference on Computing, Analytics and Security Trends (CAST)*. IEEE, 2016: 52 - 57.
- [30] SHANG C, CHEN J. Discrete graph structure learning for forecasting multiple time series [C] // *Proceedings of International Conference on Learning Representations*, 2021: 66 - 71.
- [15] AMAR A, WEISS A J. Direct position determination of multiple radio signals [C] // *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montreal, QC, Canada, 2004: 81.
- [16] YI W, CHEN Z H, HOSEINNEZHAD R, et al. Joint estimation of location and signal parameters for an LFM emitter [J]. *Signal Processing*, 2017, 134: 100 - 112.
- [17] 陈芳香, 易伟, 周涛, 等. 应用膨胀算法提取目标的直接定位算法 [J]. *雷达科学与技术*, 2019, 17 (2): 161 - 167.
- [18] 唐元春, 陈端云, 夏炳森. 基于传播算子的卫星导航系统干扰源直接定位方法 [J]. *太赫兹科学与电子信息学报*, 2023, 21 (8): 985 - 991.
- [19] TZAFRI A, WEISS A J. High resolution direct position determination using MVDR [J]. *IEEE Transactions on Wireless Communications*, 2016, 15 (9): 6449646110.
- [20] 李星, 张利强, 盛兴, 等. 未知信源数量 DOA 估计方法 [J]. *通信技术*, 2023, 56 (2): 135 - 139.
- [21] GARCIA N, WYMEERSCH H, LARSSON E G, et al. Direct localization for massive mimo [J]. *IEEE Transactions on Signal Processing*, 2017, 65 (10): 24752487.
- [22] 赵晓丹, 张国梅, 尹佳文, 等. 利用神经网络逐级缩小定位区域的低复杂度多级干扰源直接定位方法 [J]. *西安交通大学学报*, 2022, 56 (7): 136 - 144.
- [23] 陈芳香. 分布式雷达多辐射源直接定位算法研究 [D]. 成都: 电子科技大学, 2019.