

目录

1	INTRODUCTION	2
2	Related Work	3
2.1	交通流量需求预测	3
2.2	多周期时间模式分析	4
3	PERLIMINARIES	5
4	Experiments	5
4.1	正文	5
4.2	参考	5
4.2.1	叶硕士论文	5
5	CONCLUSION	5
6	APPENDIX	5

Optimizing Traffic Flow Demand Prediction with Skip Connected Coupled Graph Convolutional Network using Self-Attention Fusion Mechanism

摘要 这项研究提出了一种新颖的方法来优化交通流量需求预测,采用了 Skip Connected Coupled Graph Convolutional Network (SCC-GCN) 结合了 Self-Attention Fusion Mechanism (SAFM)。先前的研究采用了周期性特征提取,但存在模型训练时间长、过拟合和结果不理想等局限性。为了克服这些问题,我们引入了一种新颖的自注意力融合机制,将三周期数据结合在一起,减少了训练时间,同时提高了模型的准确性。我们还提出了一个跳跃连接的耦合门控循环单元 (CGRU),用于保留历史节点信息,并在每个时间步骤使用局部和全局空间注意力来捕捉空间依赖关系,从而获得更大的隐藏空间依赖性。我们在公开可用的 NYCTaxi 和 NYCBike 数据集上进行了实验证明,我们提出的带有 SAFM 的 SCC-GCN 在交通流量需求预测方面优于几种最先进的方法,具有较高的预测准确性和计算效率,显示出在实际应用中的潜力。

1 INTRODUCTION

交通流量预测是城市计算中的一个关键任务,它可以优化交通资源的分配,提升智能交通系统的效率。然而,由于每天生成的大量交通数据,实时交通流量预测是一个复杂而具有挑战性的问题。先前的研究采用了卷积神经网络 (CNNs) 或循环神经网络 (RNNs) 来预测基于城市级网格地图的交通流量,这可以捕捉交通数据中的时空依赖关系。然而,这些模型在建模非欧几里德数据(如交通流量)的复杂空间关系方面存在局限性。

基于图构建的技术,如图卷积网络 (GCNs) 或图神经网络 (GNNs),近年来在研究城市交通流量的时空依赖性方面变得越来越受欢迎。这些方法可以将道路网络中的物理和语义信息结合起来,从而提供对交通动态的更全面理解。此外,有一种新兴趋势是使用多变量时间序列预测来分析交通数据中的多周期时间模式。

然而,这些方法仍然面临着在动态情况下从多个角度捕捉交通流量的重大挑战,这是由于交通数据的密集时空动态和城市中发现的可变多周期时间模式所导致的。其中一些技术挑战包括如下:

- 由于周期性数据噪声导致模型过拟合: 先前的模型依赖于周期性特征提取,这将数据分成不同的周期进行特征提取。然而,这种方法的缺点是对预测结果具有不同的权重,并且一些周期性数据包含噪声,这对于准确的预测结果并不有利。我们的方法通过自注意力融合三周期数据来解决这个挑战,从而缩短了模型训练时间并减少了过拟合。
- 无法在每个时间步骤提取空间依赖关系: 交通数据的密集时空动态为当前方法在每个时间步骤捕捉空间依赖关系提出了显著挑战。我们的方法采用了局部和全局空间注意力,在数据的每个时间步骤提取空间依赖关系以及全局空间依赖关系,从而获得更大的隐藏空间依赖性。

为了克服这些挑战,本研究提出了一种新颖的方法,利用了 Skip Connected Coupled Graph Convolutional

Network (SCC-GCN) 结合了 Self-Attention Fusion Mechanism (SAFM) 来改进交通流量需求预测。首先, 我们采用自注意力机制将多周期数据融合, 以缩短模型的训练时间并减少过拟合。随后, 我们提出了一个跳跃连接的耦合门控循环单元 (SCC-GRU), 在模型中保留了更多有关历史节点的信息。具体而言, 我们的模型可以通过同时采用局部和全局空间注意力机制来捕捉交通流量数据的时空依赖关系。我们的主要贡献如下:

- 我们开发了一个新颖的框架, SCC-GCN 与 SAFM 相结合, 可以有效地模拟交通流量数据的时空依赖关系, 提升了交通流量需求预测的准确性。
- 我们引入了一个跳跃连接的耦合图循环单元 (SCC-GRU), 可以保留历史节点信息, 并增强模型在每个时间步骤建模空间依赖关系的能力, 从而产生更多隐藏的空间依赖性。
- 我们设计了一个动态融合机制 SFAM, 可以高效地融合周期性数据并提取有用的特征, 从而改善了模型捕捉交通流量复杂时间模式的能力。
- 我们对来自多个城市的真实世界数据进行了全面的实验证明, 并展示了我们提出的 SCC-GCN 与 SAFM 在准确性、稳定性和鲁棒性方面优于最先进的方法。

本文的其余部分组织如下。我们在第 2 节回顾了相关工作。在第 3 节中描述了本研究中使用的定义。第 4 节和第 5 节详细阐述了 SCC-GCN 和 SAFM 的设计及评估。最后, 第 6 节总结了本文。

2 Related Work

本节介绍了在交通流量需求预测和多周期时间模式分析领域相关研究的文献综述。

2.1 交通流量需求预测

交通流量需求预测是智能交通系统中至关重要的任务, 它可以促进交通管理和规划。然而, 交通流量需求受到诸多因素的影响, 如天气、事件和道路状况, 使得准确预测变得困难。传统的方法, 如基于网格的方法, 使用网格地图来表示交通网络。例如, Li 等人提出了一个基于卷积神经网络 (CNNs) 和长短时记忆 (LSTM) 网络的深度学习框架来预测城市范围内的交通需求。Wang 等人设计了一个多视图图卷积网络来捕捉交通需求的时空相关性。然而, 这些方法忽视了交通网络的自然拓扑结构, 可能会丢失一些重要的信息。

最近, 图神经网络已经成为建模非欧几里德数据 (如交通网络) 的强大工具。因此, 许多研究采用了图神经网络来进行交通流量需求预测。例如, Jiang 等人提出了一种基于动态图和强化学习的长期交通流量预测方法。Chen 等人开发了一个动态时空图神经网络, 可以捕捉交通需求的不确定空间依赖关系。Yao 等人提出了一个用于交通预测的深度学习框架, 重新考虑了交通数据的时空相似性, 并学习了一个相似性感知的注意力机制来捕捉不同区域之间的动态相关性。Zhao 等人提出了一个用于交通预测的时空图卷积网络 (T-GCN), 结合了图卷积网络和门控循环单元, 以模拟交通数据的时空依赖关系。Guo 等人提出了一种基于注意力的时空图卷积网络 (ASTGCN) 用于交通流量预测, 利用空间注意机制来学习不同位置的重要性以及时间注意机制来学习不同时间间隔的影响。Bai 等人提出了一种用于交通预测的自适应图卷积循环网络 (AGCRN), 使用自适应邻接矩阵来捕捉动态的空间相关性, 并使用图卷积门控循环单元来捕捉交通数据的时间相关性。

尽管图神经网络具有许多优势，但这些方法仍然面临着从不同时间尺度分析交通需求的多周期模式的挑战

2.2 多周期时间模式分析

多周期时间模式分析是一种技术，用于发现和建模多元时间序列数据在不同时间尺度上的循环模式。例如，在交通流量数据中，可能存在影响交通量和速度的日常、周常和季节性模式。分析这些模式可以帮助理解数据的基本因素和动态特性，同时提高预测的准确性和可靠性。

已经提出了一些方法来对多元时间序列数据进行多周期时间模式分析。其中一些基于频域分析，如傅立叶变换或小波变换，将数据分解成不同的频率成分并识别出主导的周期性。然而，这些方法可能无法捕捉数据的非平稳和非线性特性，并且可能受到高计算复杂性和噪声敏感性的影响。

其他方法基于机器学习技术，如神经网络或深度学习模型，可以学习复杂灵活的数据表示，并捕捉多个序列之间的时空依赖关系。例如，Wang 等人提出了一种基于模型融合的时间序列预测方案，利用多元灰色模型和人工鱼群算法来优化设置。他们表明，他们的模型可以在来自不同场景（如交通流量和电力消耗）的真实数据系列上实现良好的预测准确性和效率。另一个例子是杨等人提出的一种称为空间-时空信息和交通模式相似性信息（STTPS）的方法，旨在通过考虑季节性和超近期模式的影响，将时间因素纳入交通分析。这种方法使得可以在交通流量数据的不同时间尺度上识别存在的规律性和变异性，从而提高了预测模型的准确性和可解释性。

然而，大多数现有方法倾向于关注每个周期内的规律性，并通过静态关系矩阵来捕捉空间模式，从而忽视了站点之间随时间变化的相互作用的动态性。这种限制限制了这些方法捕捉交通流量的更深层次的时空特征的能力。为了克服这一限制，我们提出了一种新颖的方法，利用了 Skip Connected Coupled Graph Convolutional Network (SCC-GCN) 结合了 Self-Attention Fusion Mechanism (SAFM) 来有效融合周期性数据并提取有用的特征。我们的方法旨在捕捉复杂的交通流量模式，从而提高预测的准确性和可靠性。

3 PERLIMINARIES

4 Experiments

4.1 正文

4.2 参考

4.2.1 叶硕士论文

表 3-1 NYCTaxi 数据集和 NYCBike 数据集详细信息

参数	NYCTaxi	NYCBike
开始时间	4/1/2016	4/1/2016
结束时间	6/31/2016	6/31/2016
训练集 (天)	63	63
验证集 (天)	14	14
测试集 (天)	14	14
时间片 (分钟)	30	30
研究范围 (千米)	8.42×14.45	8.42×14.45
站点个数	266	250

(1) NYCTaxi 数据集: 该数据集包含了 2016 年 4 月 1 日至 2016 年 6 月 31 日共 91 天纽约市大约 3500 万份出租车旅行记录数据, 包含以下信息: 上下车时间、上下车经纬度、行程距离等。实验中将前 63 天数据作为训练集, 最后 14 天数据作为测试集, 其余数据作为验证集数据。

(2) NYCBike 数据集: 这个数据集包含了纽约市人们每天使用的共享自行车订单的记录数据, 包含以下信息: 自行车上下车地点、上下车时间和旅行时间。训练集、验证集和测试集划分和 NYCTaxi 数据集一致。3.5.2

5 CONCLUSION

6 APPENDIX

A 表格: 符号说明

B 有效计算方程 2

方程 2 可以分解为两部分, 具有相同的时间复杂度, 即一部分是 $D_O^{-1}W$, 另一部分是 $D_I^{-1}W^T$ 。因此我们只展示第一部分的时间复杂度。

定义 $T_k(x) = (D_O^{-1}W)^k x$, 方程 2 的第一部分可以重写为

$$\sum_{k=0}^{K-1} \theta_k T_k(X_{:,p}) \quad (1)$$

由于 $T_{k+1}(x) = D_O^{-1}W T_k(x)$ 且 $D_O^{-1}W$ 是稀疏矩阵, 容易看出方程 4 可以使用 $O(K)$ 个递归稀疏-稠密矩

符号	说明
\mathcal{G}	一个图
\mathcal{V}, v_i	图的节点, $ \mathcal{V} = N$ 且第 i -个节点
\mathcal{E}	图的边
W, W_{ij}	图的权重矩阵及其元素
D, D_I, D_O	无向度矩阵, 入度/出度矩阵
L	归一化图拉普拉斯矩阵
Φ, Λ	L 的特征向量矩阵和特征值矩阵
$X, \hat{X} \in \mathbb{R}^{N \times P}$	一个图信号, 以及预测的图信号
$\mathbf{X}^{(t)} \in \mathbb{R}^{N \times P}$	时间 t 的图信号
$H \in \mathbb{R}^{N \times Q}$	扩散卷积层的输出
f_{θ}, θ	卷积滤波器及其参数
f_{Θ}, Θ	卷积层及其参数

Table 1 符号说明

阵乘法计算, 每个的时间复杂度为 $O(|\mathcal{E}|)$ 。因此, 方程 2 和方程 4 的时间复杂度均为 $O(K|\mathcal{E}|)$ 。对于稠密图, 可以使用谱稀疏化 (Cheng et al., 2015) 使其变得稀疏。

C 与谱图卷积的关系 2

证明。谱图卷积利用了归一化图拉普拉斯矩阵的概念 $L = D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}} = \Phi\Lambda\Phi^T$ 。ChebNet 将 f_{θ} 参数化为 K 阶 Λ 多项式, 并使用稳定的切比雪夫多项式基计算它。

$$\mathbf{X}_{:,p} \star_{\mathcal{G}} f_{\theta} = \Phi \left(\sum_{k=0}^{K-1} \theta_k \Lambda^k \right) \Phi^T \mathbf{X}_{:,p} = \sum_{k=0}^{K-1} \theta_k L^k \mathbf{X}_{:,p} = \sum_{k=0}^{K-1} \tilde{\theta}_k T_k(\tilde{L}) \mathbf{X}_{:,p} \quad (2)$$

其中 $T_0(x) = 1, T_1(x) = x, T_k(x) = xT_{k-1}(x) - T_{k-2}(x)$ 是切比雪夫多项式的基函数。设 λ_{\max} 表示 L 的最大特征值, $\tilde{L} = \frac{2}{\lambda_{\max}}L - I$ 表示图拉普拉斯矩阵的重新缩放, 将特征值从 $[0, \lambda_{\max}]$ 映射到 $[-1, 1]$, 因为切比雪夫多项式在 $[-1, 1]$ 中形成正交基。方程 5 可以被视为 \tilde{L} 的多项式, 我们将证明 ChebNet 卷积的输出与扩散卷积的输出类似, 只是存在常数比例的差异。假设 $\lambda_{\max} = 2$ 且 $D_I = D_O = D$ 用于无向图。

$$\tilde{L} = D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}} - I = -D^{-\frac{1}{2}}WD^{-\frac{1}{2}} \sim -D^{-1}W \quad (3)$$

因此, \tilde{L} 类似于负的随机游走转移矩阵, 因此方程 5 的输出也类似于方程 2 的输出, 只是存在常数比例的差异。

D MORE RELATED WORK AND DISCUSSION

Xie 等人 (2010) 引入了一种基于高斯过程 (GPs) 的方法。然而, GPs 很难扩展到大型数据集, 并且通常不适用于相对较长的交通预测, 例如 1 小时 (即提前 12 个步骤), 因为方差可能会累积并变得极大。

Cai 等人 (2016) 提出了使用时空最近邻 (ST-KNN) 进行交通预测。虽然 ST-KNN 考虑了空间和时间的依赖关系, 但它具有以下缺点。正如 Fusco 等人 (2016) 所示, ST-KNN 对每条单独的道路进行独立预测。一条道路的预测是其自身历史交通速度的加权组合。这使得 ST-KNN 难以充分利用邻近道路的信息。

此外, ST-KNN 是一种非参数方法, 每条道路都是单独建模和计算的 (Cai 等人, 2016), 这使得它难以推广到未见情况并且难以应用于大型数据集。最后, 在 ST-KNN 中, 所有的相似性都是使用手工设计的度量标准进行计算的, 其中包含了少量可学习参数, 这可能会限制其表征能力。

Cheng 等人 (2017) 提出了 DeepTransport, 通过明确地收集每条道路的一定数量的上游和下游道路, 然后对这些道路分别进行卷积来建模空间依赖关系。与 Cheng 等人 (2017) 相比, DCRNN 以一种更系统的方式建模了空间依赖关系, 即基于交通传播的性质将卷积推广到交通传感器图中。此外, 我们从随机游走的属性中推导出 DCRNN, 并展示了流行的谱卷积 ChebNet 是我们方法的一个特例。

所提出的方法也与图嵌入技术相关, 例如 Deepwalk (Perozzi 等人, 2014), node2vec (Grover 和 Leskovec, 2016), 它们为图中的每个节点学习了一个低维度的表示。DCRNN 也为每个节点学习了一个表示。学到的表示同时捕捉了空间和时间的依赖关系, 并且同时针对目标进行了优化, 例如未来的交通速度。

详细的实验设置

HA (Historical Average): 将交通流量建模为一个季节性过程, 使用前几个季节的加权平均作为预测值。所使用的周期为 1 周, 预测基于之前几周的聚合数据。例如, 对于本周三的预测, 将使用过去四个周三的平均交通速度。由于历史平均方法不依赖于短期数据, 其性能对于预测时域的小增量是不变的。

$ARIMA_{kal}$: 具有 Kalman 滤波器的自回归积分移动平均模型。阶数为 (3,0,1), 模型使用了 Python 包 statsmodel 进行实现。

VAR (Vector Auto-regressive) 模型 (Hamilton, 1994): 滞后数设定为 3, 模型使用了 Python 包 statsmodel 进行实现。

SVR (Support Vector Regression): 线性支持向量回归, 惩罚项设为 $C=0.1$, 历史观测值的数量为 5。以下是基于神经网络的方法:

FNN (Feed forward neural network): 具有两个隐藏层的前馈神经网络, 每个层包含 256 个单元。初始学习率为 $1e^{-3}$, 在第 50 轮开始, 每 20 轮减小到原来的 1/10。此外, 对于所有隐藏层, 使用了 0.5 的丢弃率和 L2 权重衰减 $1e^{-2}$ 。模型以批量大小 64 进行训练, 损失函数为 MAE。通过监控验证误差进行早期停止。

FC-LSTM (Encoder-decoder framework using LSTM): 使用具有窥视孔的 LSTM (Sutskever 等人, 2014) 的编码器-解码器框架。编码器和解码器都包含两个循环层。在每个循环层中, 有 256 个 LSTM 单元, L1 权重衰减为 $2e^{-5}$, L2 权重衰减为 $5e^{-4}$ 。模型以批量大小 64 进行训练, 损失函数为 MAE。初始学习率为 $1e^{-4}$, 在第 20 轮开始, 每 10 轮减小到原来的 1/10。通过监控验证误差进行早期停止。

DCRNN (Diffusion Convolutional Recurrent Neural Network): 包括两个循环层的编码器和解码器。在每个循环层中, 有 64 个单元, 初始学习率为 $1e^{-2}$, 并在第 20 轮开始, 每 10 轮减小到原来的 1/10。使用了在验证数据集上进行早期停止的策略。此外, 随机游走的最大步数, 即 K , 设为 3。对于计划的采样, 使用了阈值反 sigmoid 函数作为概率衰减:

$$\epsilon_i = \frac{\tau}{\tau + \exp(i/\tau)}$$

其中 i 是迭代次数, 而 τ 是控制收敛速度的参数。在实验中 τ 被设定为 3,000。该实现可以在 <https://github.com/liyaguang/DCRNN> 找到。

E.1 数据集

我们在两个实际的大规模数据集上进行实验：

- **METR-LA**: 该交通数据集包含了从洛杉矶县高速公路上的环路检测器中收集的交通信息（Jagadish 等人, 2014）。我们选择了 207 个传感器，并在实验中收集了从 2012 年 3 月 1 日到 2012 年 6 月 30 日的 4 个月的数据。观测到的交通数据点的总数为 6,519,002。
- **PEMS-BAY**: 该交通数据集由加利福尼亚交通局（CalTrans）的性能测量系统（PeMS）收集。我们选择了湾区的 325 个传感器，并在实验中收集了从 2017 年 1 月 1 日到 2017 年 5 月 31 日的 6 个月的数据。观测到的交通数据点的总数为 16,937,179。

这两个数据集的传感器分布在图 8 中可视化显示。

在这两个数据集中，我们将交通速度读数聚合成 5 分钟窗口，并进行 Z-Score 标准化。其中，70% 的数据用于训练，20% 用于测试，剩余的 10% 用于验证。为了构建传感器图，我们计算传感器之间的道路网络距离，并使用阈值化的高斯核（Shuman 等人, 2013）构建邻接矩阵。

$$W_{ij} = \begin{cases} \exp\left(-\frac{\text{dist}(v_i, v_j)^2}{\sigma^2}\right) & \text{若 } \text{dist}(v_i, v_j) \leq \kappa \\ 0 & \text{否则} \end{cases}$$

其中， W_{ij} 表示传感器 v_i 和传感器 v_j 之间的边权重， $\text{dist}(v_i, v_j)$ 表示从传感器 v_i 到传感器 v_j 的道路网络距离。 σ 是距离的标准差， κ 是阈值。

E.2 指标

假设 $x = x_1, \dots, x_n$ 表示实际值， $\hat{x} = \hat{x}_1, \dots, \hat{x}_n$ 表示预测值， Ω 表示观测样本的索引，定义如下指标：均方根误差（RMSE）

$$\text{RMSE}(x, \hat{x}) = \sqrt{\frac{1}{|\Omega|} \sum_{i \in \Omega} (x_i - \hat{x}_i)^2}$$

平均绝对百分比误差（MAPE）

$$\text{MAPE}(x, \hat{x}) = \frac{1}{|\Omega|} \sum_{i \in \Omega} \left| \frac{x_i - \hat{x}_i}{x_i} \right|$$

平均绝对误差（MAE）

$$\text{MAE}(x, \hat{x}) = \frac{1}{|\Omega|} \sum_{i \in \Omega} |x_i - \hat{x}_i|$$