

文章编号: 2095-2163(2021)08-0053-04

中图分类号: TP399

文献标志码: A

基于混合时空感知网络的城市区域人流量预测

张敏¹, 卿粼波¹, 王巧¹, 才虹丽², 陈杨²

(1 四川大学 电子信息学院, 成都 610065; 2 四川通信科研规划设计有限责任公司, 成都 610065)

摘要: 城市区域人流量预测对公共安全、交通管理有着重大的公共利益影响。然而,不同区域之间,以及受到时空相关性和许多外部因素的影响,给区域人流量的精准预测带来了极大的挑战。为进一步掌握城市人员流动聚集动向,本文提出一种基于混合时空感知网络模型,对重点密集区域的人流量进行建模,训练模型使之可以预测城市区域人流量,并将结果进一步与空间区域类型因子进行融合,最后可以分小时预测每个区域的人流量。选取北京市 2020-01-17~2020-02-15 重点区域数据信息,针对本文方法进行验证。实验结果表明,基于混合时空感知网络模型提升了城市区域人流量的预测效果。

关键词: 机器学习; 时空感知网络; 人流量预测

Prediction of urban population flow based on hybrid time-space perception network

ZHANG Min¹, QING Linbo¹, WANG Qiao¹, CAI Hongli², CHEN Yang²

(1 School of electronic information, Sichuan University, Chengdu 610065, China;

2 Sichuan communication scientific research planning and Design Co., Ltd., Chengdu 610065, China)

[Abstract] The prediction of urban regional passenger flow has a significant public interest impact on public safety and traffic management. However, the interaction between different regions, as well as the influence of spatiotemporal correlation and many external factors, has brought great challenges to the accurate prediction of regional people flow. In order to further grasp the trend of urban people flow and aggregation, this paper proposes a hybrid spatiotemporal perception network model to model the people flow in key intensive areas, train the model to predict the people flow in urban areas, and further fuse the results with the factors of spatial region type, and finally predict the people flow in each area hourly. In this paper, the key regional information of Beijing from January 17, 2020 to February 15, 2020 is selected to verify the proposed method. The experimental results show that the prediction effect based on the hybrid spatiotemporal perceptual network model is better than other existing methods.

[Key words] machine learning; space time perception network; population flow forecast

0 引言

城市区域人流量预测在智慧交通系统中具有重大现实意义。例如,当区域人流量过大、人群聚集程度较高时,如果不能及时察觉潜在的风险,进行科学的疏导和控制,很容易发生踩踏等事件^[1-2]。城市区域人流量预测在智慧交通系统中具有重大现实意义,人群流动建模为日益严峻的交通状况、公共卫生安全、资源配置、风险评估等问题提供有效的决策支持^[3]。

在城市区域人流量预测问题中,传统的建模思路难以同时有效的处理时序的两级周期性和空间依

赖性问题。2017年,Zhang等人^[4]建立了基于深度残差网络的城市人流量预测模型 ST-ResNet。其后,研究人员在此基础上陆续提出了 ConvGRU^[5]等一系列模型,将每个网格区域的人流量用规则矩阵表征卷积输入,表征人流量预测中的空间依赖关系。文献[6]中提出了 DeepSTN+模型,结合区域的功能特性,能够捕捉更大范围空间的依赖关系。文献[7]中使用空间卷积,来构建预测不规则区域的人流量模型 MVGCN 等等。然而,上述传统的区域人流量预测模型,没有充分挖掘人流量数据的时间特征,虽利用深度学习进行预测,但在数据量不足时,训练数据的性能会显著下降。

基金项目: 四川省科技计划项目(2018HH0143)。

作者简介: 张敏(1996-),女,硕士研究生,主要研究方向:多媒体通信与信息系统;卿粼波(1980-),男,博士,副教授,主要研究方向:图像处理、图像/视频编码通信、嵌入式系统;王巧(1996-),女,硕士研究生,主要研究方向:多媒体通信与信息系统;才虹丽(1973-),女,大专,高级工程师,主要研究方向:移动核心网、大数据应用;陈杨(1982-),女,硕士,中级工程师,主要研究方向:无线网、大数据应用。

通讯作者: 卿粼波 Email: qing_lb@scu.edu.cn

收稿日期: 2021-07-03

本文数据来源于 CCF 2020 重点区域人群密度预测^[8],在 Top1^[9]模型基础上,针对以上问题,考虑时序特征的两级周期性,一方面利用图神经网络(GNN)^[10]提取空间特征,另一方面将空间区域特性因素纳入其中,构建模型预测城市区域未来一周每小时的人流量分布。

1 基于混合时空感知的网络模型

1.1 问题分析

城市区域人流量预测相比一般的时序预测问题具有时序特征、区域特征,而时间特征具有两级周期性。图1为一个星期内,北京市内两种类型区域24小时人流量统计图。图中很明显地展示了城市区域人流量预测的时序特征与区域特性。

(1)天级别周期性:体现了每个区域在一天内小时级别的周期性。例如图两个区域人流量,由于人们每日作息影响呈现周期性变化,即早晨和晚上人流量骤减,中午达到高峰。

(2)周级别周期性:体现了每个区域周级别的周期性。图中两个区域人流量受工作日的人群出行规律的影响,工作日和周末的人流量差异较为明显。

(3)区域特性:由于人的移动是发生在物理世界中,直接受到区域属性的影响。从图中也可看出人流量的分布与功能区之间存在相关性。例如,北京站整体人流量较高,所以在构建模型时应当充分考虑区域类型与时间的相关性对人流量的影响。

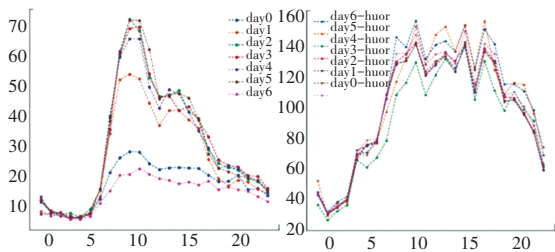


图1 两个区域人流量统计图

Fig. 1 Statistics of people flow in two regions

1.2 模型结构

本文提出的基于混合时空感知网络模型的完整网络结构如图2所示。其中包括分为数据层、特征层和融合层。数据层选择区域历史人流量和区域属性数据作为模型的输入。首先进行特征提取得到两个级别的时序特征,通过图神经网络(GNN)^[10]提取空间特征;将时序特征和空间特征输入到回归树模型和规则模型,对未来人流量进行连续预测;最后融合功能类型、时间影响因子、区域特性因子输出预测值。

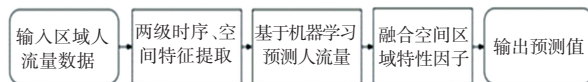


图2 模型框架图

Fig. 2 Model framework

1.3 特征提取

1.3.1 时序特征提取

本文对相关数据进行特征提取,得到区域人流量两级时序特征。人流量具有天级别和周级别的两级周期性,所以选择预测方式分为水平方向预测与垂直方向预测,如图3所示。

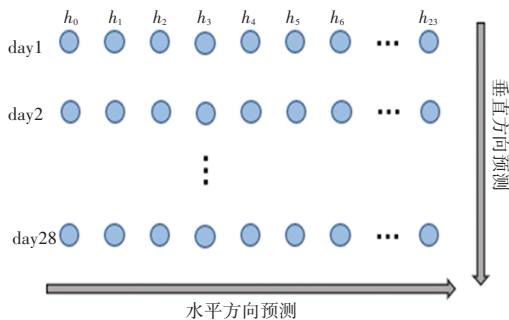


图3 两级时序特征提取

Fig. 3 Two level sequential feature extraction

其中,水平方向预测主要基于历史数据每天的相同小时进行预测,这样训练24个回归树模型进行预测,将预测的序列长度缩短到了7,垂直方向预测主要对历史训练数据进行统计建模。

1.3.2 空间特征提取

本文利用图神经网络,对不规则的空间特征进行提取。其基本原理如下:

将城市区域人流量的相关数据抽象成空间内有意义的模式与特征后,再将图转化成代数形式来对空间特征进行提取。

图4为以定点*i*计算出的核大小为1的图卷积。而定点*i*与4个一阶相邻定点的空间依赖关系则为图4中的 $f_{(1)}$ 、 $f_{(2)}$ 、 $f_{(3)}$ 、 $f_{(4)}$ 。

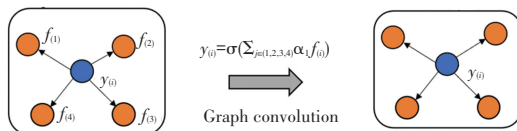


图4 区域人流量空间特征提取

Fig. 4 Extraction of spatial characteristics of regional passenger flow

1.4 基于机器学习算法的人流量预测

本文使用 LightGBM^[11]和 XGBoost^[12]两模型预

测人群密度预测值 $flow_{i,d,h}$ 。其中, $growth_{i,h}$ 表示区域 i 在第 h 小时的增长趋势因子, p_h 和 q_h 表示第 h 小时对应的两个模型, 则第 i 个区域在第 d 天第 h 小时的人流量预测值为:

$$flow_{i,d,h} = \frac{growt h_{i,h}}{2} \cdot (p_h^{(d)}(x_{i,h} \oplus f_{i,h}) + q_h^{(d)}(x_{i,h} \oplus f_{i,h})), \quad (1)$$

1.5 融合区域特性因子

区域特性因子 w^p 主要是计算同一时刻 t 内, 迁入某一区域的人流量 P_E 与迁出这一地区的人流量 P_R 的比值, 如公式(2)。

$$w^p = \frac{P_E}{P_R}, \quad (2)$$

为客观的选择人流量基准, 本文选择最近 3 天的平均人流量 $base_i^t$ 和回归树模型预测的第一天人流量 $base_i^1$ 进行加权, 作为基础人流量:

$$base_i = w^p \cdot base_i^t + w^1 \cdot base_i^1, \quad (3)$$

最后融合周级别周期因子 α 与天级别的影响因子 β , 可得到基于规则统计模型的人流量预测值为:

$$flow_{i,d,h}^{o2} = flow_{i,d,h} \times base_i \times \alpha_{i,d} \times \beta_{i,h}. \quad (4)$$

2 实验结果与分析

2.1 数据集

根据现实需求, 本文选取的数据集包括北京市 2020-01-17~2020-02-15 重点区域信息、100 个类别的重点区域小时级别的人群密度数据、城市间迁入迁出指数、网格联系强度。预测接下来 9 天北京市重点区域的人流量。所以要预测的序列长度为 216, 属于长期序列预测问题。数据集的统计信息见表 1。

表 1 数据集统计信息

Tab. 1 Statistical information of data set

城市	北京
时间跨度	20.01.17-20.02.15
重点区域	997 个
区域网格范围	200 m×200 m
网格联系强度	[0,1]
迁徙指数	[0,100]
城市	北京

2.2 基准方法与评价指标

基于本文所构建的模型, 用均方根误差 RMSE (Root Mean Square Error) 作为模型预测效果的评价指标。

$$RMSE = \sqrt{\frac{1}{\varepsilon} \sum_k (\varphi_k - \hat{\varphi}_k)^2}. \quad (5)$$

其中, ε 表示城市的区域总数; k 表示预测的时间区间总数; φ 是预测值; $\hat{\varphi}$ 是真实值。

为了验证本文模型在区域人流量预测问题上的可行性, 选择以下基准预测方法进行对比。预测方法包括: 时间序列预测模型和基于深度学习的预测模型。主要有长短期记忆网络 (LSTM)^[13]、自回归积分滑动平均模型 (ARIMA)^[14]、以及基础方法 (base)。

(1) ARIMA 模型: 自回归积分滑动平均模型, 是时间序列预测模型中非常经典的预测方法, ARIMA 的预测结果较好, 但模型对数据要求较高。

(2) LSTM 模型: 长短期记忆模型是一种循环神经网络, 擅长处理序列类型的数据。

2.3 实验对比

混合时空感知模型预测结果见表 2。

表 2 实验结果

Tab. 2 Experimental results

模型	RMSE
LSTM	16.48
ARIMA	8.87
Base[文献 9]	5.48
本文模型	5.39

3 结束语

本文提出一种基于混合时空感知网络模型, 用于解决城市区域人流量预测问题。该模型可以对影响区域人流量的两类因素 (即空间信息和时间信息) 进行建模, 根据天周期性和周周期性进行预测, 将时序特征和空间特征分别输入回归树模型和统计规则模型, 对未来人群密度进行连续预测。进一步对回归树模型和统计规则模型的预测结果进行加权求和, 融合后效果显著。通过对北京重点区域人流量数据验证, 证明本文提出的模型具有良好的效果。

当前大部分的城市区域人流量预测模型都只考虑了某种单一的数据集, 实际上, 本模型可以考虑其他因素用于改进城市区域人流量预测。为了验证模型的鲁棒性, 不同城市需要用到更多的数据来验证时空变化对预测精度的影响。同时为提高训练效率, 可减少迭代次数以提高训练结果的准确性。

(下转第 60 页)