“智联友道·滴滴”杯

第十五届全国大学生交通运输科技大赛

THE 15TH NATIONAL COMPETITION OF TRANSPORT SCIENCE AND TECHNOLOGY FOR UNDERGRADUATE STUDENTS

JS8-主题竞赛：交通大数据

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

基于大规模时空数据的短时交通流预测

指导导师：任桐鑫

队员：王鹏，单张拓，吕天傲，王家琪，黄沁

研究背景

对每条道路上时间段的车流量分布情况的研究，可以随时了解每条道路的交通情况达到快速掌握道路实时情况，及时改变出行路线措施或有效应对突发事件发生的情况的目的。通过合理的算法对交通流数据进行实时、高效的分析处理与预测，能更加准确得到未来一天内交通状况，避免拥堵。

解决方案

1 方案综述

本组采用对照方法，建立了具有创新性与先进性的两个实验方案，针对线圈数据建立了考虑时空的KNN模型和基于路段车辆流量拓扑图的GCN&LSTM深度学习方法，并设置了四个早起实验尝试的对照组以论证前两种方法的创新性和优越性。在浮动车部分，通过K-means++、整数规划等算法，并进行数学建模，实现了从浮动车数据到路段整体流量数据的跨越，并将此数据来优化线圈测得的路段数据。

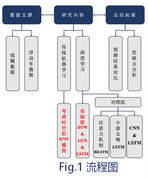


Fig.1 流程图

2 考虑时空的KNN模型

建立时空相似性的KNN预测模型，引入当前路段本身的历史数据和其上流路段的历史流量数据，建立时空矩阵，采用矩阵F-范数的倒数来衡量交通流之间的相似度，来寻找K个最相似的日期，并加权得到预测值。

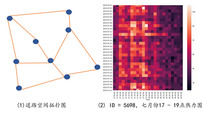


Fig.2 KNN

3 基于路段车辆流量拓扑图的GCN+ LSTM模型

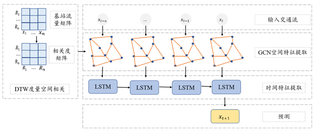


Fig.3 原理

本项目采用了动态时间规整（DTW）作为路段车辆流量相似性的定义方法。

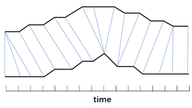


Fig.4 DTW

在神经网络预测方面采用图卷积神经网络（GCN）与长短期记忆人工神经网络（LSTM）结合的方式，通过DTW计算预测路段及其相邻路段的时空相关性，建立图邻接矩阵并对原始数据进行空间转换，并加入注意力机制构建GCN+ LSTM神经网络模型，通过参数设置与调整，得到训练完毕的预测模型。

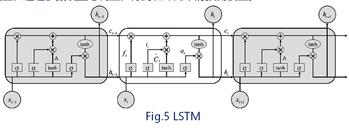


Fig.5 LSTM

4 浮动车

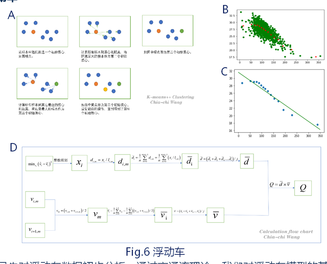


Fig.6 浮动车

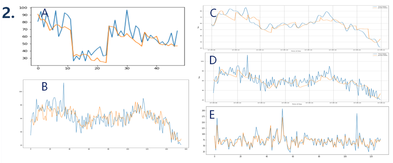
本项目先对浮动车数据初步分析，通过交通流理论，我们对浮动车模型的基本构想为通过浮动车建模得到平均车流密度和区间平均速度，从而得到交通流量，以此来优化线圈数据（线圈数据中有部分零错误点）。我们可视化出速度-密度散点图，并使用K-means++聚类，采用格林希尔兹（Greenshields）线性模型进行参数标定，得到该路段的速度-密度公式。然后，我们使用以整数规划为主要思想的数学建模，从而得出了该路段的平均流量。A图为K-means++算法，B图为聚类分析，C图为Greenshields线性模型拟合，D图为整个浮动车算法流程图。

模型训练结果

1.

建立时空相似性的KNN预测模型，引入当前路段本身的历史数据和其上流路段的历史流量数据，建立时空矩阵，采用矩阵F-范数的倒数来衡量交通流之间的相似度，来寻找K个最相似的日期，并加权得到预测值。图A、B、C为KNN在取K=4时ID=5698路段时7.28-7.30 的训练结果，平均MAPE约为11%。图D、E为考虑时空的KNN在取K=4时ID=5698路段时7.28-7.29 的训练结果，平均MAPE约为8%，可见相比直接使用KNN**提高了3个百分点。**

2.



图A为GCN+LSTM的部分预测结果，图B为注意力BiLSTM的预测结果，图C为小波去噪处理数据依频部分拟合后的部分预测结果，图D为小波去噪处理数据高频部分拟合后的部分预测结果，图E为CNN——注意力BiLSTM的部分预测结果。

结果分析

训练部分+提交部分

从训练结果中，我们不难看出，时空KNN和GCN+LSTM的效果最佳，误差处于一个较小的水平，MAPE达到9.23%和11.55%，MAE达到4.72和8.43，RMSE达到6.24和17.55。

我们发现，在各项评价指标中MAPE出现了与测试集较大的差别，而MAE和RMSE差异较小。最终决定提交GCN+LSTM的算法作为我们的最终算法。

我们分析MAPE有较大差别的原因，发现是对极端值没有进行处理。在未来的工作中，我们会对较大的数值进行进一步探讨，进行模型的研究和修改。同时，在周围路段数据得到后，使用时空矩阵等方法对数据进行处理，对结果进行优化。

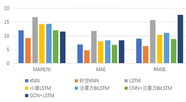


Fig.7 训练结果示意图

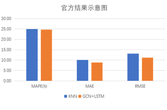


Fig.8 官方结果示意图

创新思想

1. 创新性地将空间因素嵌入到传统的KNN模型中，建立时空融合的KNN预测方法
2. 考虑路段间交通流量存在滞后效应，使用DTW度量路段间流量相似度
3. 使用图卷积神经网络将道路视为存在相互作用的拓扑结构进行建模，更贴合实际
4. 建立一个拉普拉斯矩阵便可以充分表示所有路段间的相似度，减小计算量
5. 对原始流量数据进行小波去噪的处理，降低噪声，便于更好地追踪交通流的特征
6. 将浮动车数据进行地图匹配，并提出可利用浮动车数据验证线圈预测数据准确性

应用前景与展望

本项目通过机器学习和神经网络智能算法实现了城区交通流路段级数据的预测，预测精度达到了较高的水平，经过更多的数据验证和进一步提升，可以广泛应用于城区道路交通乃至拥堵情况的预测，助力建设智慧城市，结合地图类APP和物联网，为居民城市出行提供有效参考，同时也为城区道路规划和建设提供参考性建议。

当然，本项目在预测精度方面也存在着提升的空间。取得完整的城区道路拓扑交通流数据，可期利用时空关联的GCN+LSTM取得精度更高的数据。另一方面，若能时间跨度更长的数据，提升训练数据的量级，可以使得训练的神经网络模型更加完备。

此外，探索更多更新的提升预测精度的人工智能和机器学习算法，也有提升结果的可能。

参考文献

[1]袁向红. 基于浮动车数据的高速公路交通量预测模型研究[D].北京交通大学,2019.

[2]薛亚鑫. 基于LSTM深度神经网络的短期交通流预测[D].北京交通大学,2019.

[3]庄瑞刚. 基于双向循环神经网络的交通流预测算法研究[D].西安电子科技大学,2019.

[4]薛泽龙. 基于多模型长短时记忆和时空关联的短时交通流预测[D].华南理工大学,2019.

[5] X.Y.Xu,J.Liu,H.Y.Li,and J.Q.Hu,‘ Analysis of subway station capacity with the use of queueing theory’ Transportation Research Part C Emerging Technologies, vol. 38, no. 1, pp. 28–43, Jan. 2014.

[6] P. Wei, Y. Cao, and D. Sun, “Total unimodularity and decomposition method for large-scale air traffic cell transmission model,” Transportation Research Part B, vol. 53, no. 3, pp. 1–16, Jul. 2013.