

基于网约车轨迹数据的路网拥堵识别及预测

第 40 期 PRP 结题答辩

徐麦 刘创

上海交通大学密西根学院

2022 年 3 月 9 日



背景



图 1: 堵车现象



图 2: 车辆轨迹数据可视化

随着城市和城际交通压力增大, 面对海量交通轨迹数据, 人们愈加需要**研究并分析数据特征**, 对于堵车现象建立相应**模型预测识别**。

关键词: **轨迹大数据, 深度学习 (神经网络), 预测模型**。

研究范围

研究区域——成都市市中心

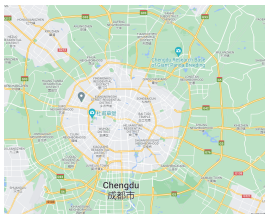


图 4: 成都市市中心概况

网约车司机 ID	opqlvh8hc1yjyh6iCBooqrkhdmi _s tBe
订单 ID	jyylseig5 _z oyi _r rrndqzyndmd1zpwI
Unix 时间戳	1477969147
经度	104.07513
纬度	30.72724

图 5: 原始数据样例

数据概况:

- 呈现很好的**结构性**
- 包含了详细经纬度信息和时间序列信息，能够从中提取**时空特征信息**
- 个别部分呈现**缺失值**仍需要预处理

① 研究背景与意义

② 研究内容与方法

③ 研究过程与结果分析

数据预处理与分析

算法设计

神经网络设计

实验设计

预实验

正式实验

实验结果

结论

④ 总结展望

⑤ 参考文献

预处理步骤

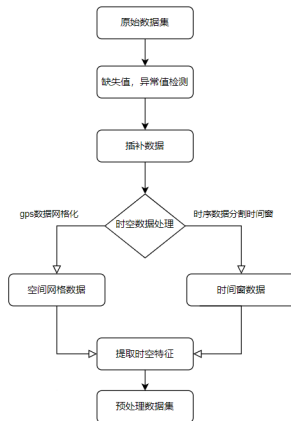






图 6: 流程图

数据特征分析：邻域连贯性

将预处理后数据后得到的**网格平均速度值**，呈现出**邻域连贯性**：

根据 Chen 的研究¹，一个交通网格在某一时刻的网格速度与其在空间和时间上邻近的网格的差距在一定的距离范围内可视为是连贯的。这是因为大部分时间段的交通状况很少出现突变的情况。

因此我们认为，同时考虑**一定时空邻域内的所有数据**对于预测某个时空网格的数据是有帮助的。

¹ Meng Chen, Xiaohui Yu, and Yang Liu. "PCNN: Deep Convolutional Networks for Short-Term Traffic Congestion Prediction". In: *IEEE Transactions on Intelligent Transportation Systems* 19.11 (2018), pp. 3550–3559. DOI: [10.1109/TITS.2018.2836523](https://doi.org/10.1109/TITS.2018.2836523).    

数据特征分析：周期性

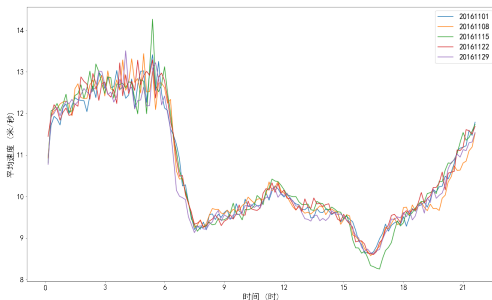


图 7: 2016 年 11 月每周周二网格平均速度趋势图

- 每周每天相同时间的平均速度趋势相似，呈现**周期性**。
- 相邻几天相同时间的平均速度趋势相似，呈现**邻域连贯性**。
- 平均车速折线两端**局部最低**分别对应**早晚高峰**（7-9；16-18）。

① 研究背景与意义

② 研究内容与方法

③ 研究过程与结果分析

数据预处理与分析

算法设计

神经网络设计

实验设计

预实验

正式实验

实验结果

结论

④ 总结展望

⑤ 参考文献

卷积神经网络

简介

卷积神经网络 (Convolutional Neural Network, CNN) 是深度学习领域最具代表性的算法之一。其核心思路是通过卷积运算和深层的网络结构，高效地捕捉高维数据中的**隐藏特征**，并完成分类和**回归**任务。

算法适用性

- 整个交通网络在同一时刻的数据可以视为一个由横纵坐标编码的**二维速度矩阵**。而不同时刻的数据间存在**邻域相关性**和**周期性**。
- 将间隔一定时间的数据矩阵叠加在一起，从多个通道一起输入神经网络，从而尝试让模型记忆并预测出交通拥堵的邻域相关性和周期性特征，最终提高预测性能。

① 研究背景与意义

② 研究内容与方法

③ 研究过程与结果分析

数据预处理与分析

算法设计

神经网络设计

实验设计

预实验

正式实验

实验结果

结论

④ 总结展望

⑤ 参考文献

输入层

设计

考虑到交通数据的邻域连贯性和周期性，我们采用**数据裁剪**来结合各个网格与它们空间上的邻域的数据；同时，每一个时刻的速度矩阵都会和与它在时间上相邻的速度矩阵从**多个通道**同时送入 CNN 进行处理，从而使 CNN 捕捉到数据在时空上的连贯性和周期性。

数据裁剪

- 本项目的研究区域是一个 15×15 的网格。我们采用长宽均为 10 的正方形滑动窗口对空间网格进行裁剪，是数据量增加到原来的 $6 \times 6 = 36$ 倍。
- 目的：使 CNN 捕捉到数据在**空间上的相关性**；扩大训练数据量

多通道输入

- 输入张量的结构可分为**标签层**和**数据层**。
- 标签层：待预测时刻 m 的速度网格。

神经网络搭建

隐藏层结构

隐藏层参考由 LeCun²设计的“**卷积-池化**”结构。特别地，考虑到卷积和池化操作可能会使特征图的尺寸缩小，我们选择在最后一个卷积层前增加一个**上采样层**来恢复特征图的尺寸

优化算法与目标函数

- 优化策略：Adam 优化器
- 目标函数：均方误差（MSE）

$$\frac{1}{n} \sum_{i=1}^n \left(Y_i - \hat{Y}_i \right)^2 \quad (1)$$

²Yann LeCun et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

① 研究背景与意义

② 研究内容与方法

③ 研究过程与结果分析

数据预处理与分析

算法设计

神经网络设计

实验设计

预实验

正式实验

实验结果

结论

④ 总结展望

⑤ 参考文献

性能评价标准

本研究中，我们综合四个性能指标来评价模型的回归性能：

- MSE: $\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$
- RMSE: $\sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}$
- MAE: $\frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$
- R^2 : $1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$

① 研究背景与意义

② 研究内容与方法

③ 研究过程与结果分析

数据预处理与分析

算法设计

神经网络设计

实验设计

预实验

正式实验

实验结果

结论

④ 总结展望

⑤ 参考文献

实验目的

为了尽量提高 CNN 性能，需要为神经网络找到合适的超参数，在本研究中包括**输入张量的形状**、**优化器学习率**以及训练的**批尺寸**。因此，在正式实验之前，我们选择用 11 月 1 日到 5 日的数据进行预实验，即用前四天的数据训练模型，以目标函数作为误差，预测第五天的数据。

输入张量的形状

将输入张量的形状 t 与 d 作为变量，实验结果如图8所示。
当 $t = 2, d = 2$ 时，误差达到最低。

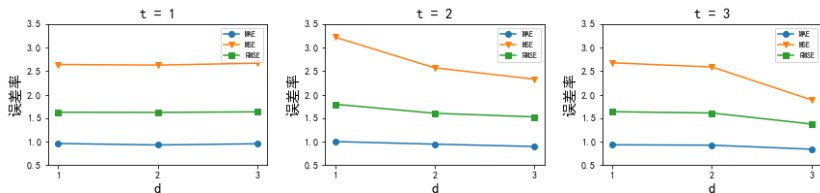


图 8: 对输入张量实验的预实验结果

学习率

将优化器的学习率作为变量，实验结果如图9所示。当学习率为0.0002时，误差达到最低。

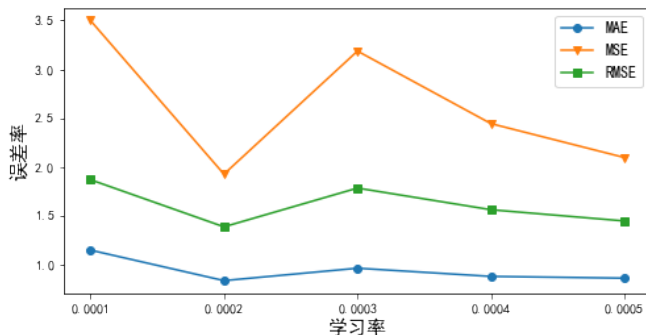


图 9: 对学习率的预实验结果

批尺寸

将批尺寸作为变量，实验结果如图10所示。在批尺寸为32时，误差达到最低。

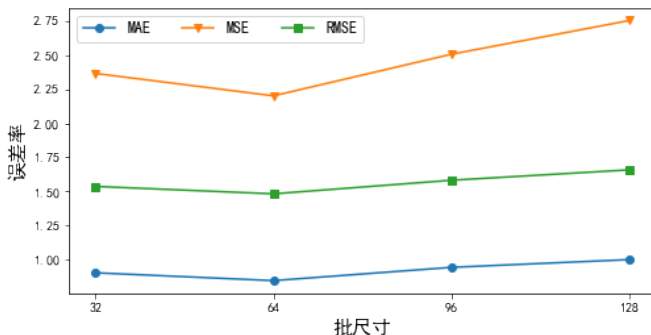


图 10: 对批尺寸的预实验结果

① 研究背景与意义

② 研究内容与方法

③ 研究过程与结果分析

数据预处理与分析

算法设计

神经网络设计

实验设计

预实验

正式实验

实验结果

结论

④ 总结展望

⑤ 参考文献

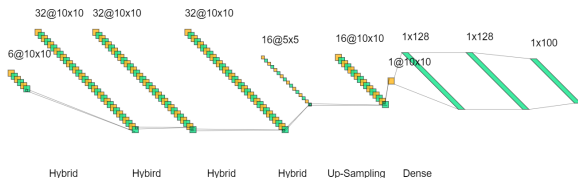
实验内容

实验目的

用 CNN 处理已经预处理后的数据集，用前 29 天的数据预测第 30 的全部网络速度数据。

实验设计

网络输入层和隐藏层结构如图11所示。Adam 优化器的学习率为 0.0002，训练批尺寸为 64. 前 29 天的速度数据在重整为特定形状的张量以后输入模型进行训练。最后，用训练所得的模型预测第 30 天的数据。



① 研究背景与意义

② 研究内容与方法

③ 研究过程与结果分析

数据预处理与分析

算法设计

神经网络设计

实验设计

预实验

正式实验

实验结果

结论

④ 总结展望

⑤ 参考文献

模型性能

对第 30 天数据预测的整体误差在表1中给出。其中**决定系数达到 0.89**，表明模型表现出较好的回归性能。

MAE	MSE	RMSE	R^2
0.617	1.068	1.034	0.893

表 1: 最终模型性能参数

结果可视化

图12为所研究路网中的其中 6 个网格的平均速度的预测值和真实值。

- 模型对于平均速度的**总体趋势**的预测较为准确。
- 不擅于预测**短时间范围内出现的速度突变**。

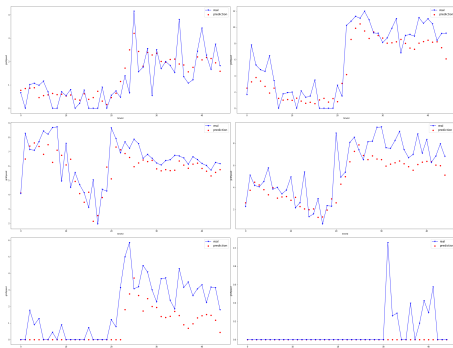


图 12: 预测结果折线图

结果可视化

图13展示了网格预测与真实情况的热力图

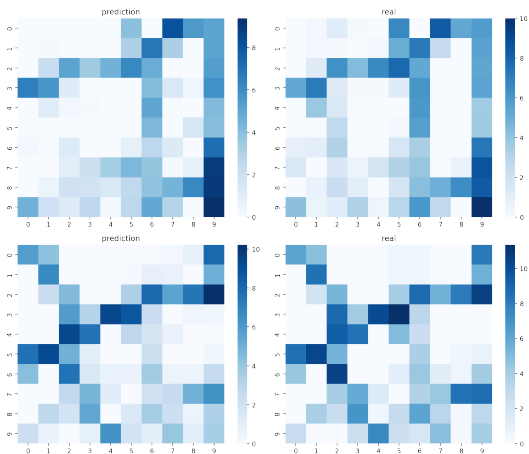


图 13: 预测结果热力图

① 研究背景与意义

② 研究内容与方法

③ 研究过程与结果分析

数据预处理与分析

算法设计

神经网络设计

实验设计

预实验

正式实验

实验结果

结论

④ 总结展望

⑤ 参考文献

研究结论

- 在本研究中，我们提出了一种用于预测短期交通拥堵状况的深度学习算法。整个算法基于卷积神经网络。考虑到交通数据往往具有连贯性和周期性，我们在输入层的选择和网络结构上做出了一定创新，将隔日与隔小时的数据相结合同时作为输入。
- 实验结果表明，模型对于一整日交通速度变化信息的预测表现出了**较强的回归性能**。这说明该算法设计可以用于交通拥堵状况的预测，从而辅助交通管理决策。

不足与应用前景

不足

- 基于网格化路段，忽略了路段的形状信息
- 训练数据完全基于网约车，无法提取更多路段车流信息
- 由于数据集的时间跨度较小以及网格化，模型对突变数据的预测能力较差
- 没有考虑更多影响交通的因素，如天气、节假日等

应用前景

本研究所提出的算法的预测值与真实值的误差相对较小，能够较好地预测网格车辆平均速度。而且模型能够实现同时对城市中心大片区域进行预测，对相对复杂、数量庞大的城市道路交通流量进行预测是一种行之有效的手段。

参考文献 I

- [1] Meng Chen, Xiaohui Yu, and Yang Liu. “PCNN: Deep Convolutional Networks for Short-Term Traffic Congestion Prediction”. In: *IEEE Transactions on Intelligent Transportation Systems* 19.11 (2018), pp. 3550–3559. DOI: 10.1109/TITS.2018.2835523.
- [2] Yann LeCun et al. “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.
- [3] Toon Bogaerts et al. “A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data”. In: *Transportation Research Part C: Emerging Technologies* 112 (2020), pp. 62–77.

Thanks!