PRP 论文草稿

绪论

交通拥堵是城市交通中非常常见的问题。城市化的推进和民众生活水平的提高同时也带来了城市车辆密度的持续上升，交通拥堵也因此愈发常见，影响民众的出行质量。应此需求，短期交通拥堵预测算法近年来在国内外蓬勃发展，目前可分为三类：概率推理模型、浅神经网络学习、深度学习。其中，概率推理模型和浅神经网络都需要对数据进行聚类分析，而深度学习不需要提前聚类即可直接处理数据，也因此在近年来获得越来越多的应用。

深度学习算法的核心是多个进行线性运算的隐藏层彼此相连，从而解决非线性问题。不同结构的隐藏层被用于提取数据蕴含的不同种类的信息。相较于浅神经网络，深度学习算法可以在对数据无任何事先了解的情况下提取出其中的特征。而对于隐变量错综复杂的交通预测问题，深度学习对数据特征的敏感性吸引越来越多的研究使用它作为预测算法的核心。

目前常用于交通信息预测领域的深度学习算法主要是卷积神经网络（CNN）和长短期记忆模型（LSTM）。考虑到交通信息在宏观上具有一定的周期性，收到[]的研究的启发，我们尝试借助CNN来实现交通拥堵的大范围预测。CNN在提取二维数据的特征上具有显著优势，而一个交通网络上的信息同样可由经度和维度两个维度编号，因此我们在预处理中选择将同一时间的交通信息转换成二维矩阵。同时，CNN的每个隐藏层都可以同时处理来自多个通道的二维矩阵，并提取它们所包含的共有特征。这一特性启发我们将间隔一定周期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周期性特征。

在本研究中，我们探索了通过合理的交通数据预处理和CNN的结构设计，测试CNN对大范围交通数据的预测性能，并进行可视化展示。所有的分析和计算都借助Python语言完成，其中，代码库Pandas和Keras分别用于进行数据的预处理和CNN的搭建、使用。

研究内容和方法

1. 问题定义以及重要参数

1.1 拥堵状态的区分

本研究中，我们以一个空间网格内的平均速度作为拥堵状态的表征。网格速度到拥堵等级的映射参考[]，并在表[]中给出

1.2 问题定义

在本研究中，我们对短期大范围交通拥堵预测问题的定义如下：给定一个地图网格和一个网格速度的序列{},i=0,1,2,…,m, j=0,1,2,…,n, t=0,1,…,T-1。其中，i和j分别表示数据所在网格的行索引与列索引，t是数据的时间索引。我们的目标是预测所有网格在时间点T的网格速度,i=0,1,2,…,m, j=0,1,2,…,n。

1. 数据预处理

2.1 数据来源与样式

本研究所用的数据来自于滴滴盖亚计划的开源数据，内容是2016年11月成都市中心的所有网约车订单数据。原始数据的样式如图[]所示，包括网约车的司机ID，订单ID，Unix时间戳，网约车在GCJ-02坐标系下的经度、维度，以及订单的日期。

2.2 数据清洗和转换

2.3 数据特征提取

1. 神经网络的构建