**A.交通数据集**

**1.数据集来源**

[1] Lan, T.; Cheng, H.; Wang, Y.; Wen, B. Site Selection via Learning Graph Convolutional Neural Networks: A Case Study of Singapore. Remote Sens. 2022, 14, 3579. https://doi.org/10.3390/rs14153579  
Web: <https://github.com/BlueSkyLT/siteselect_sg>

**2.数据集内容**

**bus\_line.csv**

包含5049个 Bus station及其路由信息

**bus\_vol.csv**

包含2021年7月、8月和9月的5018个公交车站及其乘客数量信息

**B.数据集处理与模型训练**

*数据内容较多，因此选取其中一个公交车线路进行分析尝试*  
**此处选择客流量较大、运营时间长、站点分布广的线路，如新加坡的178路公交线路，该线路连接多个居民区、商业区和学校，具有代表性。**  
此处仅对于178路的单个方向乘客数据进行分析，为stop\_id: 46009 -> 22009

**关于178路：** 178 公交 (Boon Lay Int) 有 70 站点 从 Woodlands Sq - Woodlands Int (46009) 出发并在 Jurong West Ctrl 3 - Boon Lay Int (22009) 结束

**1.处理步骤**

首先将两个文档中的数据进行关联，line文档中只有站点信息，需要与vol中的站点客流量数据进行筛选和关联。具体操作为使用Matlab程序preprocess.m选取line中的所有站点id，并在vol文件中匹配相应行单独保存到新的文件bus\_vol-178.csv中

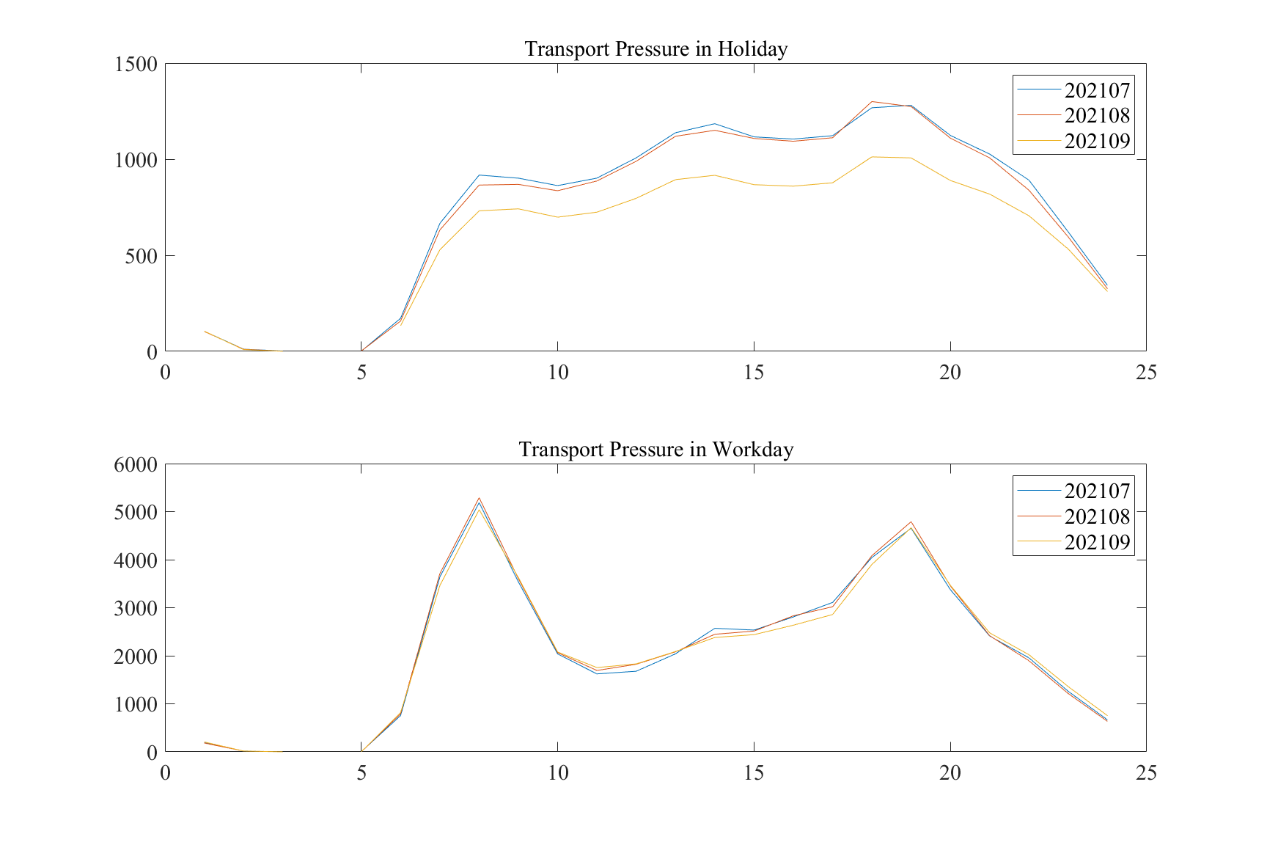
注意：为方便处理，将vol文件中day列的WD替换为0，H替换为1  
筛选完成后，将获得的矩阵保存在bus\_vol-178.csv文件中。

为了衡量某一时间段整条线路的交通压力，从而指导发车量计划，新建一个参数值为车站当前时间上车人数\*1-下车人数\*0.5

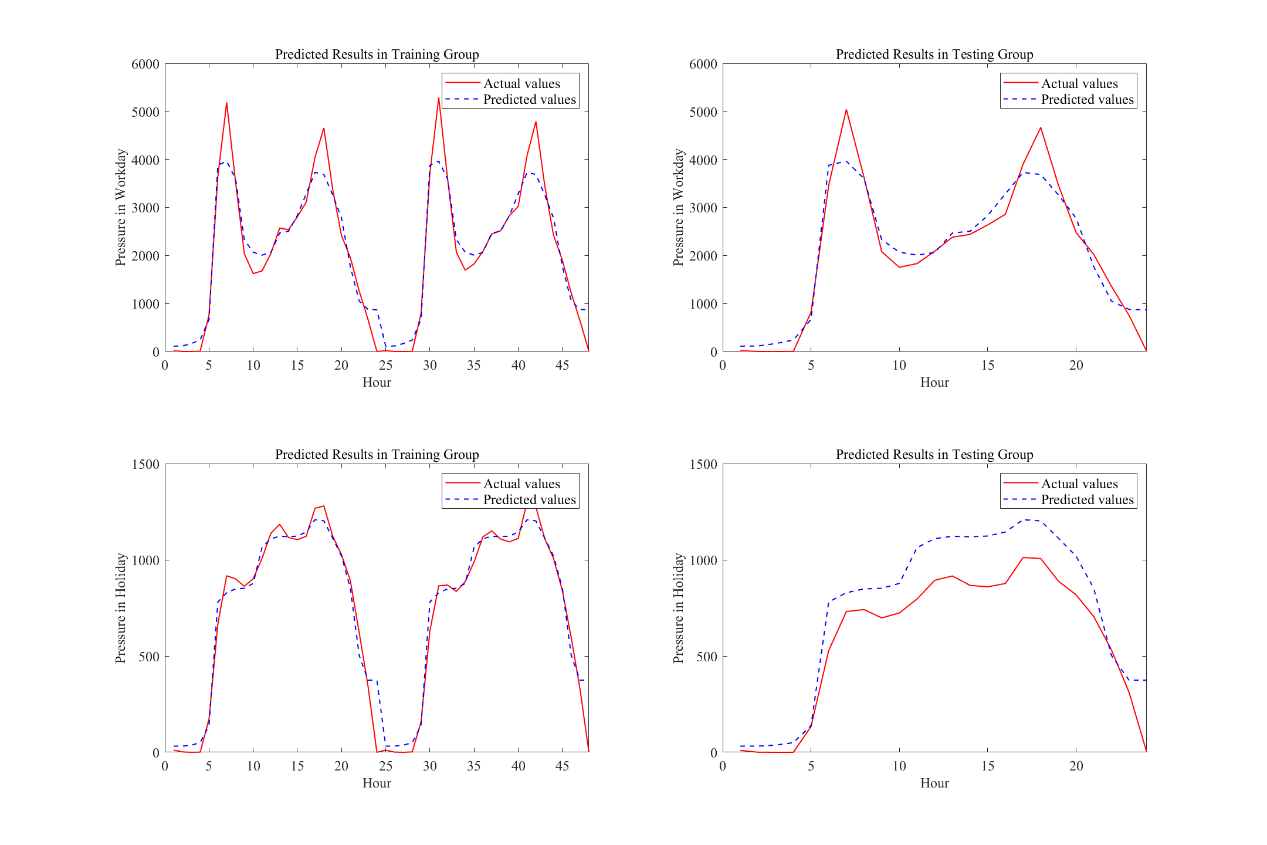
此处考虑到下车人数对于后续交通压力造成的影响没有上车那么大，但也不可以完全忽略

通过统计整条线路上所有站点的压力值之和，来反映整条线路的运输需求。当线路压力值较小时，则可以适当减少发车量，增大发车间隙时间。  
通过获得不同时间的压力值变化，结合RF随机森林预测模型，即可实现对于未来一定时间内公交线路压力值的可靠预测。  
**使用WD\_H\_Divide.m将bus\_vol-178.csv数据中WorkDays和Holidays分成两组，分别计算线路压力值，并对重复的时间段取均值  
压力值预测由Matlab文件pressurePred.m实现**

由于数据集限制，只能使用一个月的数据得到一个月中采样的每个小时的客流量，由此每个站点只能得到三组有效的24小时客流变化情况（WD和H分开计算和预测），因此用7&8月的数据作为训练集，9月作为测试集。

注意到文档中同一个月部分时间段进行了多次统计，这里对他们的客流量（进/出）取均值  


**2.RF训练与预测**

使用RF进行学习，并绘制预测与实际值对比曲线  
  
模型相关指标：均方根误差（RMSE）、决定系数（R²）

WD:

error1 = 460.3592

error2 = 428.9597

R1 = 0.9088

R2 = 0.9149

H:

error1 = 87.0491

error2 = 182.4739

R1 = 0.9632

R2 = 0.7400

**C.策略制定和仿真**

通过固定间隔发车和动态发车数量调整策略发车两者效果对比，在总发车量，人均等待时长，碳排放减少三个指标上进行模拟，总结策略优劣。

**此部分还未完成**

**不足之处**

1. 数据集比较片面，比较局限，只有7，8，9三个月的数据并且统计间隔较远，如果有更多数据补充也许能够扩大研究泛用性
2. 根据调查，新加坡的公共交通已经考虑到高峰时段（如早高峰，晚高峰）的用车压力问题，在对应时段增加公交车发车量，缩小等车间隔
3. 缩减发车量不可避免会导致受影响时间段居民等车时间延长，此外灵活的发车策略可能不利于部分民众的交通出行安排。因此需要找到方法将发车信息同步给想要搭乘的乘客，如与地图软件联动，或者在公交站台增设公交车到站预计时间指示