



2024 年（第 17 届） 中国大学生计算机设计大赛

大数据实践赛作品报告

作品编号：_____

作品名称：_____ 整车速达

——面向长三角地区的整车物流智能一体化决策方案

填写日期：_____ 2024 年 3 月 30 日

填写说明：

- 1、正文、标题格式已经在本文中设定，请勿修改；标题#的快捷键为“Ctrl+#”，正文快捷键为“Ctrl+0”；
- 2、本文档应结构清晰，突出重点，适当配合图表，描述准确，不易冗长拖沓；
- 3、提交文档时，以 PDF 格式提交；
- 4、本文档内容是正式参赛内容的组成部分，务必真实填写。如不属

目 录

中国大学生计算机设计大赛	0
大数据实践赛作品报告	0
第一章 作品概述	6
第二章 项目背景	7
2.1. 汽车物流行业发展态势	7
表格 1 近年来整车物流相关政策	7
2.2. 本作品要解决的痛点问题	8
2.2.1. 整车物流运输成本控制困难	8
2.2.2. 整车物流运输的联运能力差	8
第三章 集散中心选址方案	9
3.1. 现有整车物流运输模式	9
3.1.1. 公路运输	9
3.1.2. 公铁水联运	10
3.2. 现有整车物流运输节点	11
3.2.1. 港口节点	11
3.2.2. 铁路节点	13
3.3. 集散中心选址现存问题	15
3.3.1. 无法覆盖偏远地区	15
3.3.2. 需要增加城市中继点	15
3.4. 模型计算	15
3.4.1. 算法说明	15

3.4.2. 算法代码示例.....	16
3.4.3. 集散中心选址确定.....	19
第四章 运价成本决策方案	20
4.1. 数据来源说明	20
4.2. 数据预分析	21
4.2.1. 机构行情文本数据词频统计与关键词分析	21
4.2.2. 物流运价数据分析	23
4.3. 数据处理与清洗.....	25
4.3.1. AI 大模型处理文本数据.....	25
4.3.2. 标签编码	26
4.4. 模型构建	27
4.4.1. 算法选择与性能比较.....	27
4.4.2. 最优参数寻找	30
4.4.3. 随机森林的运价模型解析.....	32
4.4.4. 随机森林模型构建与预测函数.....	32
4.5. 运价辅助决策模型的总结和使用	32
第五章 利用 Dijkstra 算法优化运输线路.....	34
5.1. 基于启发式算法的合理取货半径.....	34
5.1.1. 算法思路.....	34
5.1.2. 算法实现.....	34
5.1.3. 结论.....	34
5.2. 基于合理取货半径的路径规划	35

5.2.1. 算法思路.....	35
5.2.2. 数据处理.....	35
5.2.3. 算法实现.....	39
5.2.4. 部分算法代码.....	40
5.3. 算法优化	42
5.3.1. 运输过程磨损费用.....	42
5.3.2. 中转费用	43
5.3.3. 基于动态规划的实际运输时效问题.....	43
第六章 整车物流大屏设计	43
6.1. 可视化大屏整体设计.....	43
6.2. 实时平均运价	44
6.3. 运价占比	44
6.4. 按季度平均价格变化.....	45
6.5. 整车平均运价预测	45
6.6. 集散中心选址	46
6.7. 不同路线运价对比图.....	46
第七章 项目总结	47
7.1. 分工与协作	47
7.2. 面对的困难和的解决方案.....	47
7.3. 未来展望	48

图表 1 思路流程图.....	7
图表 2 公路运输流程示意图.....	9
图表 3 宁波市主要港口.....	12
图表 4 上海市主要港口.....	12
图表 5 徐州市主要交通线路和枢纽.....	13
图表 6 芜湖市主要交通线路和枢纽.....	14
图表 7 杭州市主要交通线路和枢纽.....	14
图表 8 物流运价数据示例.....	20
图表 9 运输机构行情分析.....	21
图表 10 词频矩阵表.....	21
图表 11 词云图.....	22
图表 12 词频相关性图示.....	22
图表 13 偶发性因素示例.....	25
图表 14 标签编码.....	27
图表 15 得分为正数的模型.....	28
图表 16 学习曲线分析图.....	29
图表 17 随机森林模型.....	30
图表 18 最优参数.....	31
图表 19 默认参数.....	31
图表 20 示例测试集.....	33
图表 21 整车物流运输大屏展示.....	44

表格 1 近年来整车物流相关政策	7
------------------------	---

第一章 作品概述

本作品立足于长三角地区整车物流的实际情况，从整车物流运输选址、联运能力提升和运输成本控制等痛点出发，通过采用贪心算法、随机森林模型和 Dijkstra 等算法，对长三角地区的整车物流问题进行了深度研究和解决方案探索。具体包括以下四个方面的方案设计：

（1）集散中心选址方案

针对现有集散中心无法覆盖长三角地区所有城市且运输时效性无法保障的问题，首先使用启发式算法计算合理的集散中心覆盖范围，基于该合理覆盖半径使用贪心算法选择二级集散中心，以保证长三角地区所有城市均被集散中心覆盖。

（2）运价成本决策方案

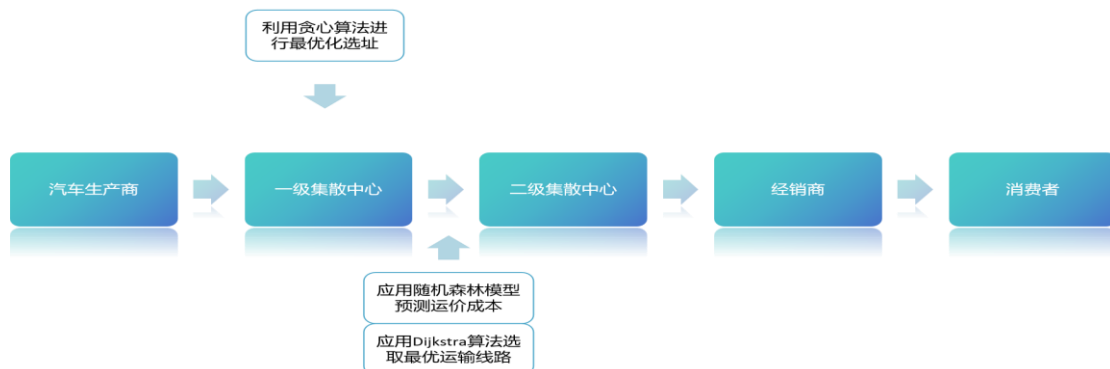
使用 Easyspider 爬取物流运价数据，并且整理成表格，通过预分析得到可能存在的特征值，使用的方法包括 AI 大模型分析，数值型数据清洗，词频统计，词语相关性分析。然后，使用这些特征值进行机器学习，比较多种机器学习模型，观察模型的综合得分，结合实际值与预测值的散点图以及学习曲线，确保不会发生欠拟合和过拟合的情况，从中选取性能最优的随机森林模型。然后，确定使用随机搜索算法，选择出最优参数下的随机森林模型，比对最优参数与默认参数的区别，确保使用的模型和参数均是现有情况的最优解。

（3）运输路线优化方案

通过 Dijkstra 算法可以更为简便的实现长三角地区整车物流运输的优化方案，从出发地点开始检索当前是否需要中转方案，并依靠 Dijkstra 算法多次更新当前运输方案，直至目标地点被收录，然后评估所有运输线路，择取最优路线。Dijkstra 算法优化了整车物流的联运能力，提高了整体的运输效率。

（4）可视化大屏设计

将各模块的数据将以直观、清晰的形式呈现在整车物流大屏上，为用户提供全面、准确的运输信息，帮助做出最佳的运输决策。



图表 1 思路流程图

第二章 项目背景

2.1. 汽车物流行业发展态势

政府对于新能源汽车产业的支持以及市场对新能源汽车的需求不断增长，也是推动长三角地区汽车制造企业发展整车物流运输的重要背景。

（1）高质量发展成为行业发展核心

中国的汽车行业正处于技术变革、产业创新和寻找新增长点的剧烈变化期。新能源和智能化汽车的发展异军突起。物流企业如想在未来仍能够维持足够的利润水平及市场竞争力，不仅要保证基本的物流运输能力，更要提高服务质量，未来物流企业竞争力的体现将从运输能力转向更高质的服务。

（2）第三方物流市场空间持续释放

在需求端，对于制造企业来讲，其核心业务为产品的生产加工及新产品研发等，物流业务属于非核心业务。为将更多资源集中于核心业务，避免因设备、仓库、人员等物流要素的投入而降低管理运作效率和资源利用效益，越来越多的制造企业将其物流业务外包给专业的物流公司，为第三方物流行业发展提供了充足动力。

（3）物流服务延伸扩展供应链两端

物流服务不断向供应链两端延伸，逐渐与制造业建立深度合作。物流企业最初只承担单一的第三方物流，到承担客户物流供应链的整体组织，进一步逐步拓展到全面介入企业的生产、销售阶段，并通过整合供应链上下游信息，优化企业各阶段的产销决策，物流企业专业化服务水平和效益显著提高。

（4）资源整合和战略合作需求凸显

目前，部分有实力的大型企业已经开始建设枢纽性基地，逐渐向“干线运输，支线分拨”的模式转型，即通过建设区域物流运营中心，扩展业务规模和服务范围，加强区域间的联动，构建与地区区域经济、产业体系、物流需求以及区位优势相适应的物流服务体系；在基地之间配合生产厂商的需求，通过干线运输集中转移，区域内再分别配送，显著提高物流效率。

表格 1 近年来整车物流相关政策

政策名称	实施年份	具体内容
《推动长三角一体化发展三年行动计划》	2020 年	提升区域物流一体化水平，优化整车物流网络布局。

政策名称	实施年份	具体内容
《国家物流枢纽布局和建设规划》	2018 年	在长三角地区布局一批国家级物流枢纽，提升整车物流服务能力。
《“十四五”信息化和工业化深度融合发展规划》与《“十四五”智能制造发展规划》	2021	推动智能制造、绿色制造示范工厂建设，构建面向工业生产全生命周期的数字孪生系统，推进相关标准的制修订工作
《“十四五”规划纲要》	2021	提出“探索建设数字孪生城市”
《原材料工业数字化转型工作方案（2024—2026 年）》	2024	显著的“工业互联网”融合应用效果，实现制造过程数字孪生的数字化转型标杆工厂

2.2. 本作品要解决的痛点问题

2.2.1. 整车物流运输成本控制困难

在传统模式下，信息不对称和效率低下会导致运输成本的增加。物流行业，特别是汽车物流行业，需要面对各种成本压力，包括：

- 运输成本：油料损耗、车辆损耗等费用。
- 人力成本：员工雇佣费用。
- 车辆折旧和维护的高成本：长期而言，运输车辆的折旧和维护成本也是企业不可忽视的开支。
- 运输路线和距离：运输路线的选择和运输距离的长短直接关系到物流运输的成本。最优的路径选择和规划问题也是企业在控制成本时需要考虑的关键因素。
- 延误和其他问题导致的额外成本。

2.2.2. 整车物流运输的联运能力差

在运输时间和路线上面花费的时间较长，不能根据有效的规划进行最快速度的送达，具体包括以下几个方面：

- 各种运输方式之间的协同问题：由于长三角地区内部包含了公路、铁路、水运等多种运输方式，但这些运输方式之间的衔接并不完善。如两种或两种以上的运输方式之间的基础设施配套不足、调度系统不统一

等，这些都导致联运效率低下，不能发挥出各种运输方式的优势。

- 信息化水平低：虽然长三角地区经济发达，但是物流行业的信息化建设还有待提高。信息传递不畅、数据共享难度大等问题，使得货物在各类运输方式之间的转换相对滞后，影响运输效率。
- 物流设施不足：长三角地区的物流设施与其经济规模和商品交易量相比还是略显不足，如货运站、集装箱码头、仓储设施等，这些不足会影响长三角地区整体的物流效率。

第三章 集散中心选址方案

3.1. 现有整车物流运输模式

3.1.1. 公路运输

目前，我国运输结构已进入新一轮优化调整。公路货运因其具有门到门直接运输、容易装卸车、适合近距离运输等特点，是目前中国物流产业的主要运输方式，有超过七成的货运量是通过公路运输来完成的。但是，以公路货运为代表的传统货运物流历年来一直存在着“小、散、乱、差”的情况，物流信息化程度不高、物流供需匹配效率较低、物流资源未能优化配置、物流服务质量不佳等一系列问题。

为此，公路运输重点转向中短途和两端短途接驳，逐渐形成节能高效的综合运输网络。公路运输的主要流程是：主机厂生产整车，生产完成后进行交接，无订单的整车，倒运至基地库，已有订单的整车直接公路运输至各 4S 店。



图表 2 公路运输流程示意图

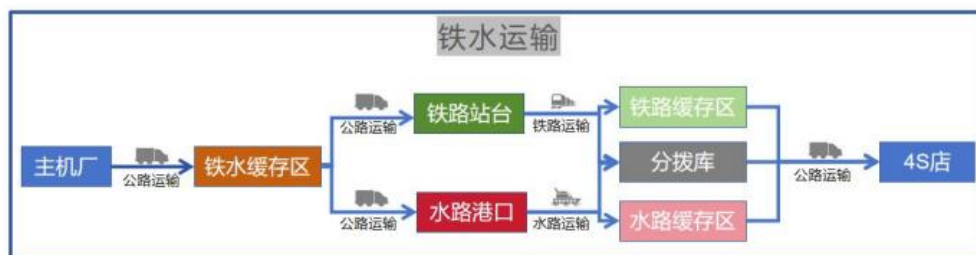


3.1.2. 公铁水联运

随着铁路运输业务量的不断增加，铁路、水路充分发挥其低成本、大批量的运输优势，承担更多中长距的批量干线运输业务，铁路运输模式逐渐呈现多样化发展，主要有站到站、站到店、站到库、厂到店等，尤其是铁路商品汽车“库前移”模式，在整体物流运作上可以有效发挥铁路批量运输和场地仓储优势。比如，中铁特货作为我国汽车铁路运输的主要承担者，在全国拥有 160 余个商品汽车装卸作业点，190 余个物流基地和铁路货场，可同时存储 26 万辆商品汽车，班、专列比例达 60%。



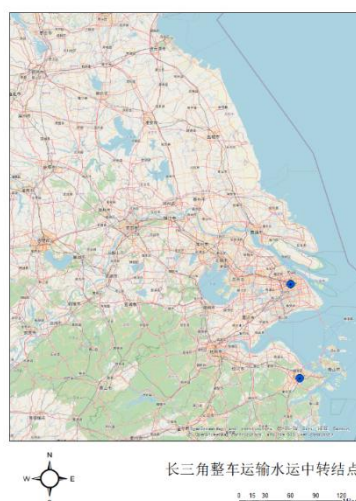
公铁水联运的主要流程是：主机厂生产整车，生产完成后进行交接并倒运至铁水缓存区，通过铁路/水路运输至指定站点后，通过公路运输至各 4S 店。



图表 2 公铁水联运流程示意图

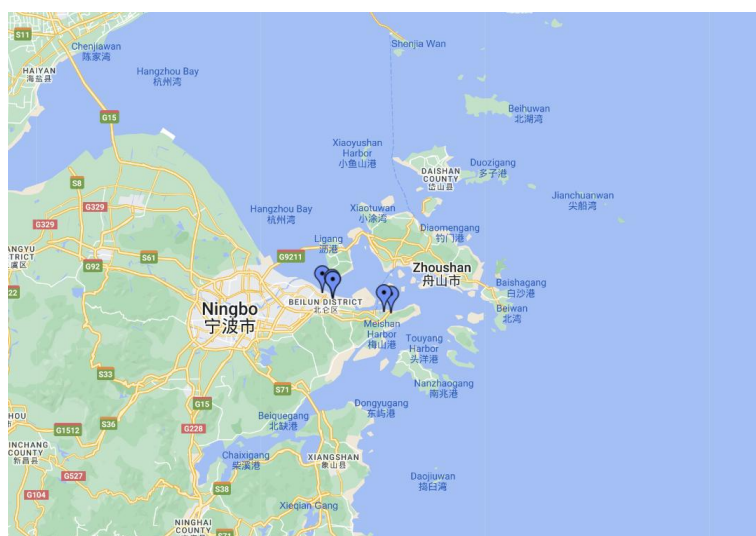
3.2. 现有整车物流运输节点

3.2.1. 港口节点



(1) 宁波港

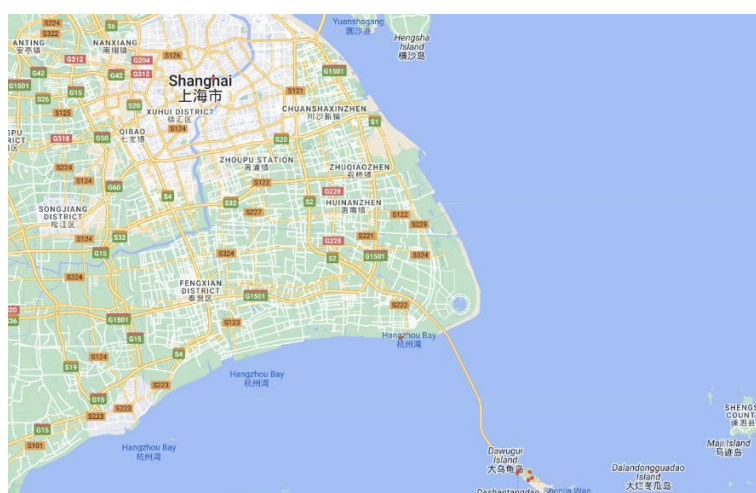
宁波港是中国东部沿海地区最重要的港口之一，位于浙江省宁波市，是中国四大主要港口之一，包括宁波港区和舟山港区。由北仑、梅山、象山、舟山、镇海等港区组成，总面积约为 338 平方公里，其中，北仑港区是中国第一大集装箱港口，舟山港区是中国主要的石油、液化气体进口港口之一。具有集装箱、散货、原油、液化气体、化工品、粮食等多种货物的装卸能力，是中国重要的外贸港口之一。经过多年的发展，宁波港得到了快速发展，是中国对外贸易的重要通道，也是“21 世纪海上丝绸之路”的重要节点之一。



图表 3 宁波市主要港口

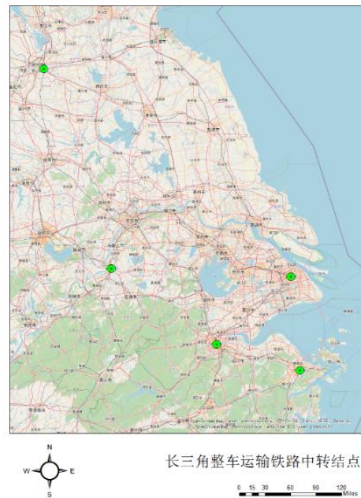
(2) 上海港

上海港是中国最大的港口，也是世界上最繁忙的港口之一，上海港位于长江口的北部，濒临长江入海口，毗邻东海，包括了若干个港区，主要有吴淞口港区、黄浦江港区、外高桥港区、杨浦港区等，其中，外高桥港区是中国最大的集装箱港区之一，拥有大型集装箱码头和现代化物流设施。总面积约为 3,619 平方公里。其拥有丰富的装卸能力和多样化的业务。它主要进行集装箱、散货、液体货物、原油、液化气体、铁矿石等货物的装卸和船舶服务。作为中国最大的港口之一，上海港在国际航运业中具有重要地位，与世界各地的港口建立了广泛的航运合作关系，成为全球贸易的重要枢纽之一。



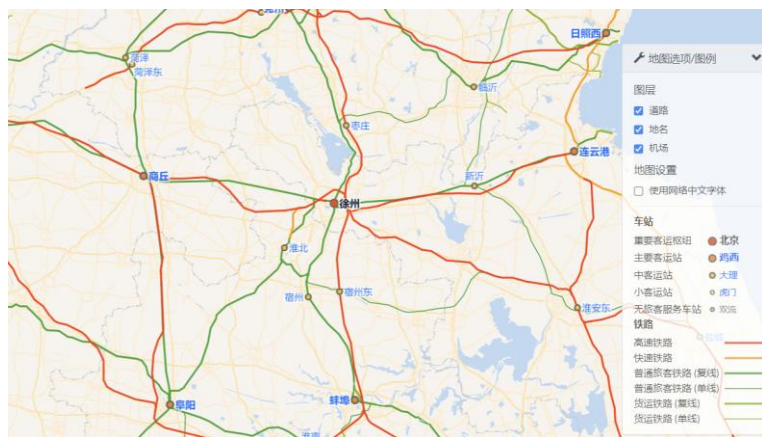
图表 4 上海市主要港口

3.2.2. 铁路节点



(1) 徐州

徐州位于中国东部，地处苏皖鲁豫四省交界处，是华东地区的交通要冲。其铁路枢纽连接了中国东部的多条主要铁路干线，包括京沪铁路、宁西铁路、连霍铁路、陇海铁路等，以及徐宿淮盐铁路等支线。这些铁路线路串联了中国东部地区以及其他地区的铁路交通，使得徐州成为铁路运输的重要中转站。徐州铁路枢纽不仅是各条铁路线路的交汇点和换乘站，还具有货运、客运、修理等多种功能。随着中国铁路网络的不断完善和发展，徐州铁路枢纽在近几十年来得到了进一步的发展和提升，成为中国东部地区不可或缺的交通枢纽之一。

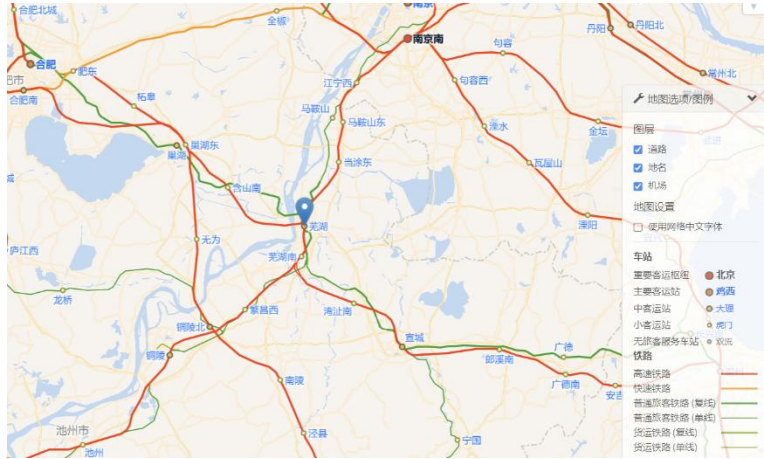


图表 5 徐州市主要交通线路和枢纽

(2) 芜湖

芜湖市位于安徽省东南部，长江南岸，地处长江中游，是连接江浙沪皖的交通枢纽。芜湖铁路枢纽地理位置优越，连接了长三角地区和华东地区的重要城市。其铁路枢纽连接了中国东部的多条主要铁路干线，包括京沪铁路、沪汉

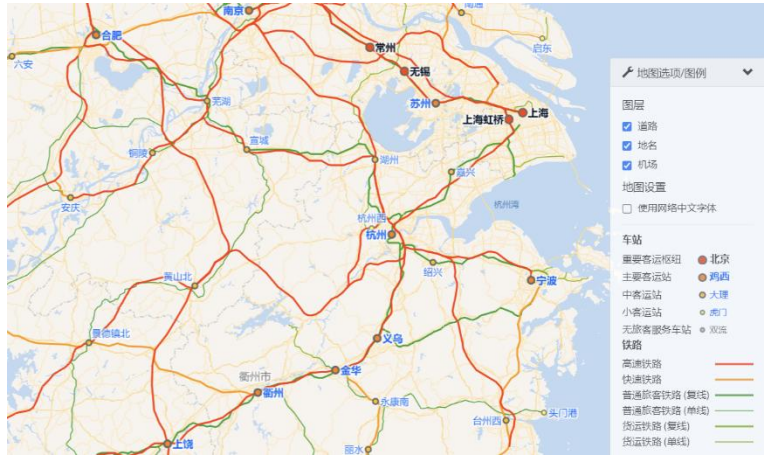
蓉铁路、皖赣铁路等，以及芜湖至宣城、芜湖至南京等支线。这些铁路线路串联了长三角地区、华东地区以及其他地区的铁路交通，使得芜湖成为铁路运输的重要中转站。2021 年发布的《安徽省综合立体交通网规划纲要》指出，建设芜湖、为全国性综合交通枢纽城市，作为国家综合立体交通网的重要节点、国家级客货运输组织和转换中心，在跨区域人员交流和国家战略物资的中转集散中发挥重要运输组织功能。



图表 6 芜湖市主要交通线路和枢纽

(3) 杭州

杭州市位于中国东南沿海地区，是浙江省的省会城市，地处长江三角洲经济区的核心位置。杭州铁路枢纽地理位置优越，连接了长三角地区和华东地区的重要城市。其铁路枢纽连接了中国东部的多条主要铁路干线，包括京沪铁路、杭福铁路、杭甬铁路等，以及杭州至宁波、杭州至湖州等支线。2018 年 10 月，杭州市围绕全力打造国际性区域交通枢纽要求，截至 2022 年底，“5433”工程重点项目已基本完工：一条条交通要线穿城而过，一座座交通枢纽拔地而起，杭州初步形成了集航空、铁路、公路、水路、城市轨道交通、城市快速路网等多种交通方式共同发展的内外交通格局，综合交通体系加快融合。



图表 7 杭州市主要交通线路和枢纽

3.3. 集散中心选址现存问题

3.3.1. 无法覆盖偏远地区

工厂将生产的整车通过水运或铁路等大宗物流方式运输至各个水运和铁路中转中心，这些中转中心通常位于各地的主要交通枢纽城市或物流集散地。然而，随着城市规模的扩大和交通网络的日益完善，一些偏远地区或次要城市可能无法直接被现有的中转中心覆盖到。

这就需要中转中心采取进一步的措施，将部分车辆运输至城市中继点，以弥补覆盖不到的地区。城市中继点通常位于各大城市或区域内的次要交通节点或物流中心，其作用是将来自中转中心的货物再次分拨至更广泛的目的地，比如各个 4S 店。

3.3.2. 需要增加城市中继点

为此，城市中继点充当了车辆运输的重要中转节点，通过城市中继点的再次分拨，可以实现车辆从集中到分散的过程。这种分拨模式可以更有效地满足消费者和 4S 店的需求，提高运输效率，同时也能够更好地适应各个城市的特定需求和交通状况。

因此，中转中心与城市中继点之间的协同运作，成为了整车运输中的重要环节，为整个供应链的顺畅运转提供了保障。同时，随着城市发展和交通基础设施的进一步完善，城市中继点的地位和作用也将逐渐得到提升，为车辆运输提供更加便捷、高效的服务。

3.4. 模型计算

3.4.1. 算法说明

贪心算法是一种在每一步选择中都采取当前状态下最好的选择，希望通过局部最优解来达到全局最优解的策略。

首先，需要计算两点之间的距离，使用的是大圆距离公式，它计算的是地球表面两点间的最短距离：设 R 是地球的半径，如果有两个点，它们的经纬度分别是 $(lon1, lat1)$ 和 $(lon2, lat2)$ ，这两个点之间的大圆距离 d 可以用下列公式计算：

将经纬度转换成弧度：

$$\text{rad}(\text{lon}) = \frac{\pi \cdot \text{lon}}{180}$$

$$\text{rad}(\text{lat}) = \frac{\pi \cdot \text{lat}}{180}$$

计算两点之间经度差和纬度差的弧度：

$$d\text{lon} = \text{rad}(\text{lon2} - \text{lon1})$$

$$d\text{lat} = \text{rad}(\text{lat2} - \text{lat1})$$

应用 Haversine 公式：

$$a = \sin^2\left(\frac{d\text{lat}}{2}\right) + \cos(\text{rad}(\text{lat1})) \cdot \cos(\text{rad}(\text{lat2})) \cdot \sin^2\left(\frac{d\text{lon}}{2}\right)$$

$$c = 2 \cdot \text{atan2}(\sqrt{a}, \sqrt{1-a})$$

$$d = R \cdot c$$

其次，在计算了距离之后，用它来确定点是否在集散中心的覆盖范围之内。若某点到中心点的距离不超过预设的覆盖半径，则可认为该点被覆盖。

对于未被一级集散中心覆盖的点，按照以下逻辑来选择二级集散中心：

- (1) 初始化未被覆盖的候选点集合为所有候选点。
- (2) 当未被覆盖的候选点集合不为空时：
 - 对于每一个候选集散中心，计算它能覆盖的未被覆盖候选点的数量。
 - 选择能覆盖最多未被覆盖候选点的集散中心成为二级集散中心。
 - 从候选点集合中移除被新选的二级集散中心覆盖的所有点。
- (3) 重复上述步骤，直到覆盖所有候选点或没有更多的候选集散中心可以选择。

通过这样的逻辑，可以得到一个贪心算法选取出的中心点集合，从而解决了集散中心选择的问题。

3.4.2. 算法代码示例

```
import math

import matplotlib.pyplot as plt
import folium

# 定义坐标点
class Point:
```

```

def __init__(self, name, lon, lat):
    self.name = name
    self.lon = lon
    self.lat = lat

# 计算两点的距离
def calculate_distance(p1, p2):
    R = 6371 # 地球半径 (千米)
    lon1, lat1, lon2, lat2 = map(math.radians, [p1.lon, p1.lat,
p2.lon, p2.lat])
    dlon = lon2 - lon1
    dlat = lat2 - lat1
    a = math.sin(dlat/2)**2 + math.cos(lat1) * math.cos(lat2) *
math.sin(dlon/2)**2
    c = 2 * math.atan2(math.sqrt(a), math.sqrt(1-a))
    distance = R * c
    return distance

# 判断点是否在集散中心的辐射范围内
def is_within_range(center, point, radius):
    return calculate_distance(center, point) <= radius

# 初步筛选未被一级集散中心覆盖的备选点
def filter_uncovered_points(primary_centers, candidate_centers,
primary_radius):
    uncovered_points = []
    for point in candidate_centers:
        if not any(is_within_range(center, point, primary_radius) for
center in primary_centers):
            uncovered_points.append(point)
    return uncovered_points

# 贪心算法选取二级集散中心
def select_distribution_centers(primary_centers, candidate_centers,
primary_radius, secondary_radius):
    # 筛选不被一级中心覆盖的点
    uncovered_points = filter_uncovered_points(primary_centers,

```

```

candidate_centers, primary_radius)

    secondary_centers = []

    while uncovered_points:
        best_center = None
        best_coverage = -1
        # 对每一个候选中心，找出能覆盖最多未覆盖坐标的中心
        for center in candidate_centers:
            coverage = sum(is_within_range(center, point,
secondary_radius) for point in uncovered_points)
            if coverage > best_coverage:
                best_coverage = coverage
                best_center = center
        # 更新未覆盖坐标列表
        uncovered_points = [point for point in uncovered_points if not
is_within_range(best_center, point, secondary_radius)]
        secondary_centers.append(best_center)
        candidate_centers.remove(best_center)

    return secondary_centers

# 定义一级和二级集散中心的辐射半径
PRIMARY_RADIUS = 0
SECONDARY_RADIUS = 200

# 定义一级集散中心
primary_centers = [
    Point('徐州', 117.283752, 34.204224),
    Point('宁波', 121.62454, 29.860258),
    Point('杭州', 120.210792, 30.246026),
    Point('芜湖', 118.433065, 31.352614),
    Point('上海', 121.473667, 31.230525),
]

# 定义二级集散中心的备选坐标
candidate_centers = [
    #此处省略

```

```

]

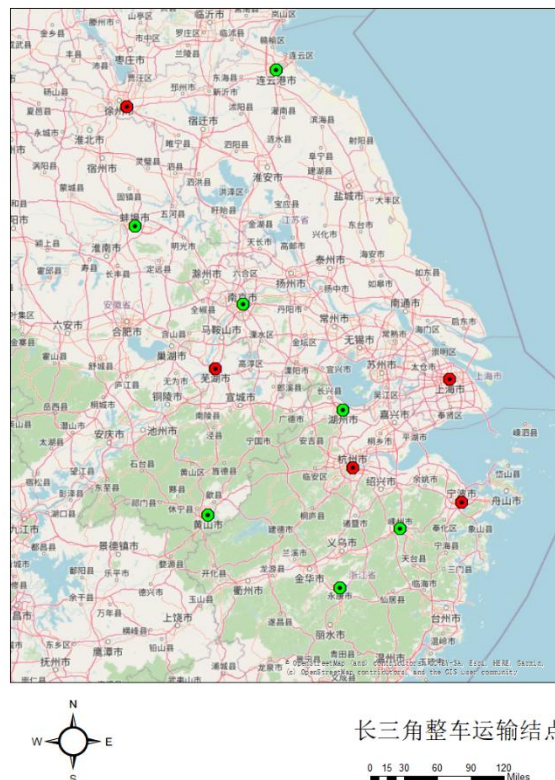
# 选取二级集散中心
secondary_centers = select_distribution_centers(primary_centers,
candidate_centers, PRIMARY_RADIUS, SECONDARY_RADIUS)

# 输出二级集散中心的名称
for center in secondary_centers:
    print(center.name)

```

3.4.3. 集散中心选址确定

将二级集散中心的辐射半径设置为 200KM，在这个问题中，贪心算法每次选择能够覆盖最多未被一级中心覆盖的点的城市作为二级集散中心。根据结果，被选取的城市是湖州市、永康市、蚌埠市、南京市、嵊州市、连云港市和黄山市。这些城市在备选坐标中，能够最大化覆盖未被一级中心覆盖的点，可以成为长三角地区的二级集散中心。



第四章 运价成本决策方案

4.1. 数据来源说明

本文的数据来自于[公路物流运价指数 \(0256.cn\)](http://0256.cn)

使用 easyspider 爬取了近年来的物流运价数据和机构行情分析报告

物流运价数据包括以下字段：

运输方 路线 价格 环比 日期
式

数据示例如下：

运输方式	路线	价格	环比	日期
整车运输-1	上海-南京	3464元/整	-0.72%	2024/3/6
整车运输-1	上海-南京	3464元/整	-0.72%	2024/2/28
整车运输-1	上海-南京	3464元/整	-0.72%	2024/2/21
整车运输-1	上海-南京	3464元/整	-0.72%	2024/2/14
整车运输-1	上海-南京	3464元/整	-0.72%	2024/2/7
整车运输-1	上海-南京	3464元/整	-0.72%	2024/1/31
整车运输-1	上海-南京	3464元/整	-0.72%	2024/1/24
整车运输-1	上海-南京	3464元/整	-0.72%	2024/1/17
整车运输-1	上海-南京	3464元/整	-0.72%	2024/1/10
整车运输-1	上海-南京	3464元/整	-0.72%	2024/1/3
整车运输-1	上海-南京	3464元/整	-0.72%	2023/12/27
整车运输-1	上海-南京	3464元/整	-0.72%	2023/12/20
整车运输-1	上海-南京	3464元/整	-0.72%	2023/12/13
整车运输-1	上海-南京	3464元/整	-0.72%	2023/12/6
整车运输-1	上海-南京	3464元/整	-0.72%	2023/11/29
整车运输-1	上海-南京	3464元/整	-0.72%	2023/11/15
整车运输-1	上海-南京	3464元/整	-0.72%	2023/11/8
整车运输-1	上海-南京	3464元/整	-0.72%	2023/11/1
整车运输-1	上海-南京	3464元/整	-0.72%	2023/10/25
整车运输-1	上海-南京	3464元/整	-0.72%	2023/10/18
整车运输-1	上海-南京	3464元/整	-0.72%	2023/10/11
整车运输-1	上海-南京	3464元/整	-0.72%	2023/10/4
整车运输-1	上海-南京	3464元/整	-0.72%	2023/9/27
整车运输-1	上海-南京	3464元/整	-0.72%	2023/9/20
整车运输-1	上海-南京	3464元/整	-0.72%	2023/9/13
整车运输-1	上海-南京	3464元/整	-0.72%	2023/9/6
整车运输-1	上海-南京	3464元/整	-0.72%	2023/8/30

图表 8 物流运价数据示例

机构行情分析包括以下字段：

标题 标题链接 行情

数据示例图：

标题	标题链接	行情
4月份中国	http://ir4月份中国公路物流运价指数为101.3点公路物流运价指数	2022-05-16 16:21:28 2022年4月份,由中国物流与采购联合会和林安物流集团联合调查的中国公路
3月份中国	http://ir3月份中国公路物流运价指数为100.3点公路物流运价指数	2022-04-19 10:02:10 2022年3月份,由中国物流与采购联合会和林安物流集团联合调查的中国公路
2月份中国	http://ir2月份中国公路物流运价指数为98.3点公路物流运价指数	2022-03-08 15:18:27 2022年2月份,由中国物流与采购联合会和林安物流集团联合调查的中国公路
1月份中国	http://ir1月份中国公路物流运价指数为100.5点公路物流运价指数	2022-02-11 16:32:43 2022年1月份,由中国物流与采购联合会和林安物流集团联合调查的中国公路
12月份中国	http://ir12月份中国公路物流运价指数为102.5点公路物流运价指数	2022-01-07 14:34:26 2021年12月份,由中国物流与采购联合会和林安物流集团联合调查的中国公
11月份中国	http://ir11月份中国公路物流运价指数为101.5点公路物流运价指数	2021-12-03 09:08:33 2021年11月份,由中国物流与采购联合会和林安物流集团联合调查的中国公
10月份中国	http://ir10月份中国公路物流运价指数为100.2点公路物流运价指数	2021-11-04 09:13:49 2021年10月份,由中国物流与采购联合会和林安物流集团联合调查的中国公
9月份中国	http://ir9月份中国公路物流运价指数为99.6点公路物流运价指数	2021-10-11 21:17:17 2021年9月份,由中国物流与采购联合会和林安物流集团联合调查的中国公路
8月份中国	http://ir8月份中国公路物流运价指数为100.1点公路物流运价指数	2021-09-06 09:10:15 2021年8月份,由中国物流与采购联合会和林安物流集团联合调查的中国公路
7月份中国	http://ir7月份中国公路物流运价指数为100点公路物流运价指数	2021-08-02 16:01:28 2021年7月份,由中国物流与采购联合会和林安物流集团联合调查的中国公路
6月份中国	http://ir6月份中国公路物流运价指数为100.1点公路物流运价指数	2021-07-06 08:44:43 2021年6月份,由中国物流与采购联合会和林安物流集团联合调查的中国公路
5月份中国	http://ir5月份中国公路物流运价指数为100.2点公路物流运价指数	2021-06-03 09:05:00 2021年5月份,由中国物流与采购联合会和林安物流集团联合调查的中国公路
4月份中国	http://ir4月份中国公路物流运价指数为100.2点公路物流运价指数	2021-05-06 14:31:58 2021年4月份,由中国物流与采购联合会和林安物流集团联合调查的中国公路
3月份中国	http://ir3月份中国公路物流运价指数为100.3点公路物流运价指数	2021-04-06 10:07:13 2021年3月份,由中国物流与采购联合会和林安物流集团联合调查的中国公路

图表 9 运输机构行情分析

4.2. 数据预分析

在机器学习的模型开始构建之前，首先对数据进行一定的预分析，确保使用的特征是有用的，有效的，同时排除不相关的因素，避免干扰机器学习的模型结果。

4.2.1. 机构行情文本数据词频统计与关键词分析

首先我们得出词频矩阵：

单词	词性	次数	条数	词频	TF-IDF
指数	名词	425	224	0.113333	0.084677
物流	名词	256	154	0.068267	0.062055
运价	名词	241	168	0.064267	0.056005
公路	名词	195	150	0.052	0.047859
中国	地名	154	126	0.041067	0.040883
回升	动词	116	46	0.030933	0.044149
回落	动词	76	36	0.020267	0.031031
零担	名词	70	42	0.018667	0.027363
区域	名词	63	41	0.0168	0.024798
月份	名词	61	43	0.016267	0.023682
增长	动词	60	30	0.016	0.025728
环比	简称略称	46	33	0.012267	0.019232
车型	名词	42	28	0.0112	0.018334
价格	名词	39	39	0.0104	0.015572
需求	动词	38	23	0.010133	0.01742
公路运输	习用语	38	34	0.010133	0.01576
市场需求	名词	34	24	0.009067	0.015426
运输	名动词	31	31	0.008267	0.013179
小幅	名词	30	25	0.008	0.013475
供给	名动词	30	16	0.008	0.014951
整车	名词	28	28	0.007467	0.012222
数据	名词	28	28	0.007467	0.012222
变动	名动词	28	14	0.007467	0.01436
平台	名词	28	14	0.007467	0.01436

图表 10 词频矩阵表

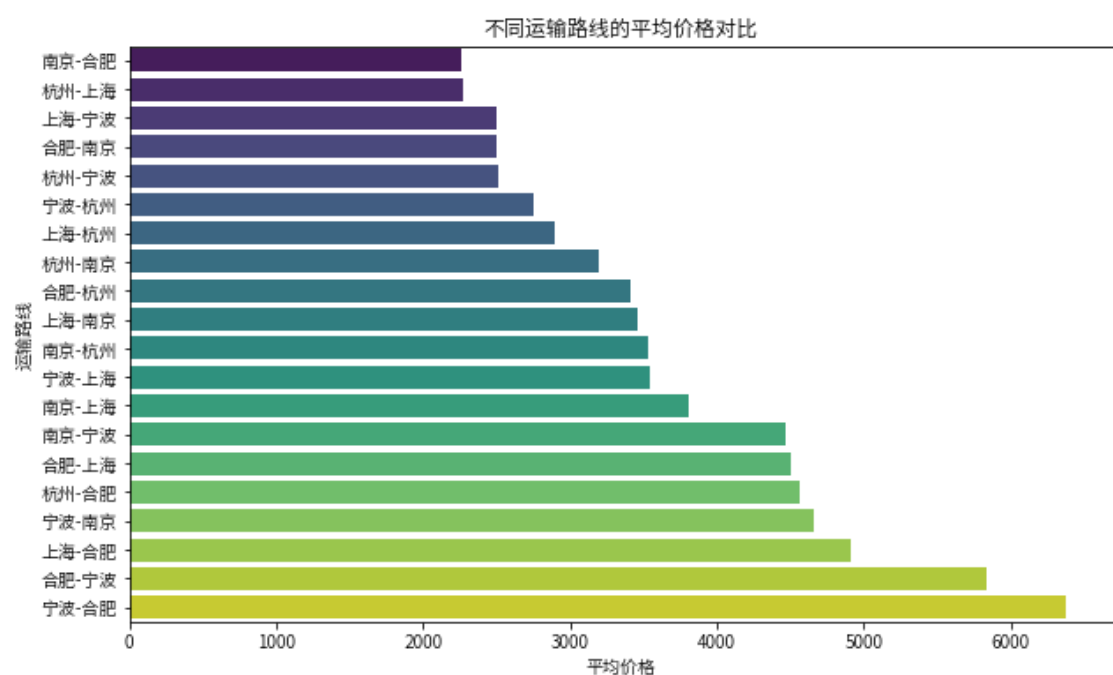
	指数	物流	运价	公路	零担	区域	月份	车型	价格	市场需求	小帽	整车	数据	平台	基点	经济	态势	市场	走势	去年同期	制造业	程度	趋势	疫情	平稳	线路	运力	总体	集团	商品	
商品	84				42	14	28				15	14								4	1										
集团	54	42	40	14		14													14					1				1			
总体	56	16	43	11		29	19		6	15	5	11				8	18	9	6	5	14	4	2	9	20		12		1		
运力	63	12	47	10		35	18		7	24	13	14				8	21	11	4	3	18	4	2	11	14	1		12			
线路	19	57	18	29		14	1		14	2	1		15	28		2	2	2	2		1			5	1		1				
平稳	61	20	45	16		31	18		5	18	8	12				12	18	9	7	6	16	1	3	8		1	14	20	1		
疫情	51	10	39	6		20	15		8	21	8	11				10	14	11	13	2	14	7	1		8	5	11	9			
趋势	24	28	23	28		3	1		16	6	1	1				17	1	6	7		2	14			1	3		2	2		
程度	39	32	27	30	3	13	7	2	18	8	6	1	5			17	7	5	1	1	6		14	7	1		4	4		1	
制造业	76	19	56	14		45	23		11	31	17	18				9	28	19	4	3		6	2	14	16	1	18	14			
去年同期	93	46	50	18	12	13	19	8		2	7	4	3			3	4		1		3	1		2	6		3	5	14	4	
走势	33	10	34	7		3	3		6	20	6	1				3	12	25		1	4	1	7	13	7	2	4	6			
市场	63	17	55	11		27	15		13	32	17	11				12	29		25		19	5	6	11	9	2	11	9			
态势	94	24	72	17		50	26		12	41	23	20				12		29	12	4	28	7	1	14	18	2	21	18			
经济	50	35	39	34		15	9		19	14	10	7						12	12	10	3	9	17	17	10	12	2	8	8		
基点	84	28	28	28																											
平台	28	112	28	56		28			28			28														28					
数据	73	69	56	38		49	18		21	22	12			28		7	20	11	1	3	18	5	1	11	12	15	14	11		1	
整车	98				42	14		28			15									4		1								14	
小帽	142	14	36	11	45	44	14	30	9	22		15	12			10	23	17	6	7	17	6	1	8	8	1	13	5		15	
市场需求	107	21	81	15		56	26		16		22	22				14	41	32	20	2	31	8	6	21	18	2	24	15			
价格	58	93	51	60		31	8			16	9		21	28		19	12	13	6		11										

通过自然语言处理，然后得出词语之间的相关性，可见车型，月份，区域，平台等因素与运价有着极高的相关性，在之后的机器学习中可考虑上述因素

4.2.2. 物流运价数据分析

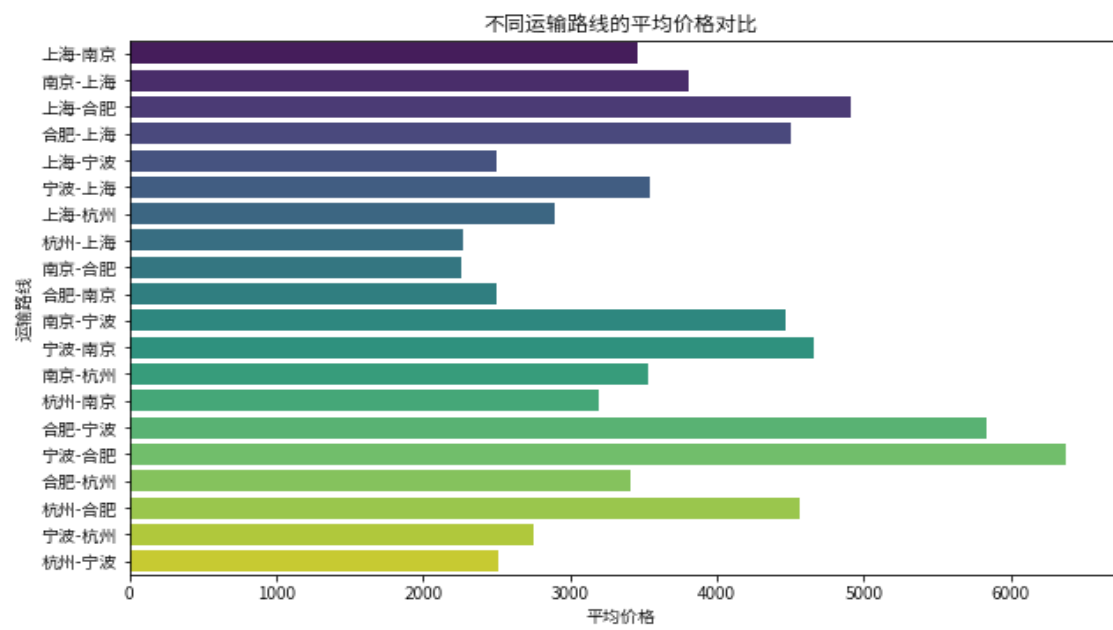
路线长度因素：

运输路线的长度和复杂性是决定运价的重要因素。一般而言，较长或更复杂的路线会导致更高的运输成本，从而影响价格。因此，可以说距离是影响因素中的一个主要因素，它基本上决定了运价的基础值。

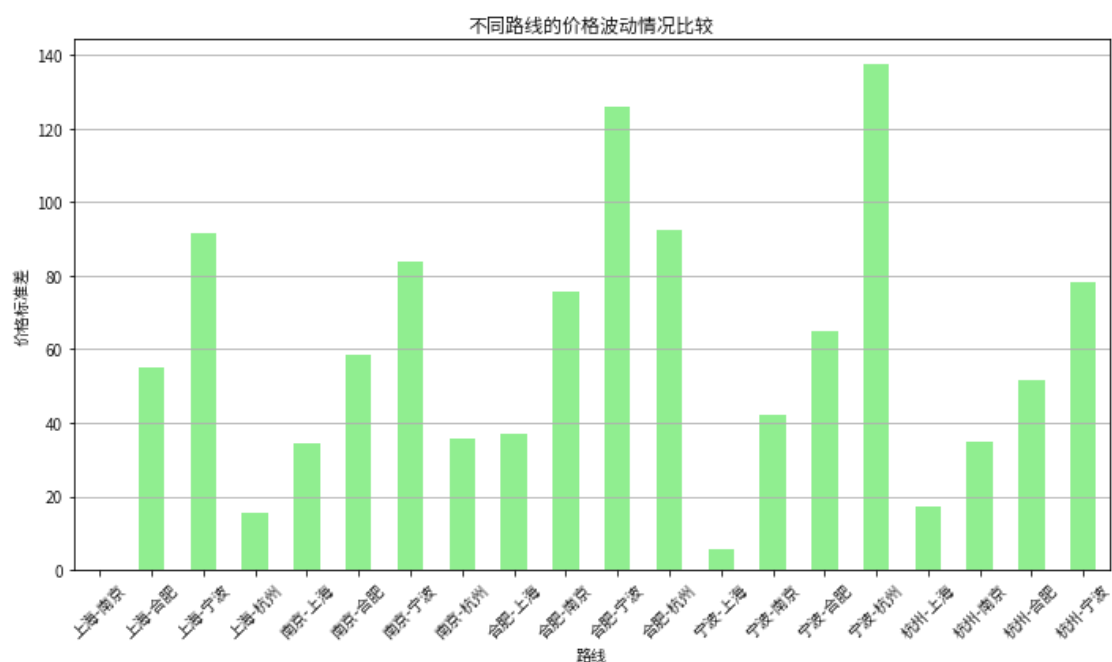


路线方向性因素

对运输路线进行平均价格比较后发现，即使是相同的路线，不同方向的运价也存在差异，比如上海到南京和南京到上海。



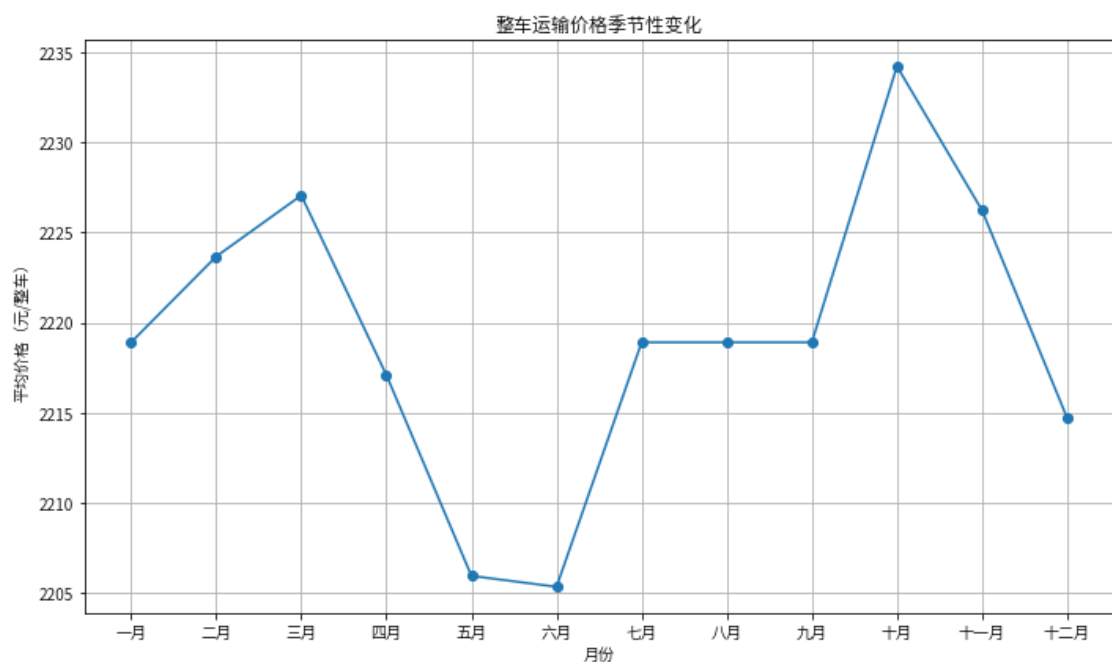
对于同一路径的不同方向，价格波动相差非常巨大。



月份季节性差异：

季节性因素也会对运价产生影响。不同季节，特定产品或服务的需求量可能会发生变化，从而导致运价的波动。例如，在某些季节需求量增加，运价可能会上涨；而在需求量较低的季节，运价可能会下降。

受限于数据，我们分析了运输价格的季节性变化的影响。



偶发性因素

除了常见的影响运价的因素之外，还存在一些偶发性因素，它们可能不是持续

性的，但在特定情况下会对运价产生显著影响。例如在 2022 年 1 月份调查报告[公路物流运价指数 2022 年 1 月份调查报告 \(0256.cn\)](https://0256.cn)，明确提及疫情这个突发因素的影响

从需求看，本月经济保持平稳运行态势，受传统节日因素，以及疫情多点散发影响，国内市场需求进一步放缓，公路物流需求稳中趋缓。PMI数据显示，1月份制造业PMI指数为50.1%，连续三个月运行在50%以上，新订单指数则持续位于50%以下。从供给看，春节返乡和部分地区疫情管控调节下，节前运力供给呈现小幅波动、震荡回落。综合来看，历史数据显示，春节假期公路运输迎来传统淡季，国内需求持续收窄，运价指数季节性回调，加之疫情因素影响，部分地区、线路运价短期波动幅度较大，本月运价指数呈现较快回落。

从后期走势看，年初经济实现平稳开局，纵观历史数据，节后公路运输市场处于调整恢复期，预计短期内运价指数可能震荡回升。但与往年相比市场需求不足、企业经营承压以及疫情防控问题仍较为突出，对公路运输供需和价格的短期影响较为明显，回升趋势或将有所趋缓。

图表 13 偶发性因素示例

总结

结合文本数据和运价数据，可以看出路线长度、路线方向性、月份季节性变化、偶发性因素、车型、月份、区域、平台等都是比较重要的特征因素。这些因素将被应用于后续的机器学习分析中，以更好地理解 and 预测运输价格的变化趋势。

4.3. 数据处理与清洗

4.3.1. AI 大模型处理文本数据

为了进一步分析机构行情文本数据，挖掘其中的偶然性突发因素对运价的影响，本文创新性地采用了语言大模型（chatglm）处理文本数据，以获取调研报告中有关突发因素的关键词。

在机器学习中，文本处理是一个重要的难点，传统的机器学习标签方法往往会丢失一些重要信息。因此，在进行编码之前，采用语言大模型提取关键词，以提高编码的准确性，减少信息噪音。同时，考虑到大模型的随机性和不稳定性，本文多次使用大模型生成数据，并从中提取出现次数最多的关键词，将其作为特征项加入到机器学习中。

使用提示词为：请对这段文本进行总结，包括日期：[日期] 关键词：[关键

词 1], [关键词 2], [关键词 3], ...摘要: [在这里提供一个简洁的摘要, 总结文本的

要点和主题。可以包括关键信息、重要观点和关联性内容。]

总结效果如下:

标题	标题链接	字段1	回答	概要	日期	关键词1	关键词2	关键词3	关键词4
4月份中国	http://ir4月份中国	日期	根据中国	2022/5/16	疫情	物流需求	价格变动	保物流畅通	运力供给
3月份中国	http://ir3月份中国	日期	根据中国	2022/4/19	环比上升	疫情影响	原材料价格	缓慢	运价指数
2月份中国	http://ir2月份中国	日期	2022年2月	2022/3/8	PMI数据	公路运输	市场需求	数回落	
1月份中国	http://ir1月份中国	日期	2022年1月	2022-02-1	2022年1月	下滑	整车指数	零担轻货	零担重货
12月份中国	http://ir12月份中国	日期	根据中国	2022/1/7	增长	整车指数	零担轻货	指数	
11月份中国	http://ir11月份中国	日期	根据2021	2021/12/3	环比增长	同比增长	中国物流	运价指数	
10月份中国	http://ir10月份中国	日期	在2021年	2021/11/4	原材料价格	需求增速	电商大促	疫情影响	经济下行
9月份中国	http://ir9月份中国	日期	2021年9月	2021	整车指数	零担轻货	零担重货	压力	运价指数
8月份中国	http://ir8月份中国	日期	2021年8月	2021/9/6	环比回升	增长	需求放缓	波动	市场需
7月份中国	http://ir7月份中国	日期	根据中国	2021/8/2	环比回落	PMI数据	短期季节	求增长	公路运
6月份中国	http://ir6月份中国	日期	根据中国	2021/7/6	物流市场需	运价指数	经济领土	输线路	
5月份中国	http://ir5月份中国	日期	2021年5月	2021/6/3	经济运行速	PMI数据	原材料市	运力供给	

将文本数据与价格变化数据通过日期进行拼接:

日期_x	年份月份	标题	标题链接	字段1	回答	概要	日期_y	关键词1	关键词2	关键词3	关键词4	关键词5	关键词6	关键词7	关键词8
2022-05	4月份中国	http://ir4月份中国	日期	根据中国	2022-05-16 00:00:00	疫情	物流需求	价格变动	保物流畅通						
2022-05	4月份中国	http://ir4月份中国	日期	根据中国	2022-05-16 00:00:00	疫情	物流需求	价格变动	保物流畅通						
2022-05	4月份中国	http://ir4月份中国	日期	根据中国	2022-05-16 00:00:00	疫情	物流需求	价格变动	保物流畅通						
2022-05	4月份中国	http://ir4月份中国	日期	根据中国	2022-05-16 00:00:00	疫情	物流需求	价格变动	保物流畅通						
2022-04	3月份中国	http://ir3月份中国	日期	根据中国	2022-04-19 00:00:00	环比上升	疫情影响	原材料价格	运力供给						
2022-04	3月份中国	http://ir3月份中国	日期	根据中国	2022-04-19 00:00:00	环比上升	疫情影响	原材料价格	运力供给						
2022-04	3月份中国	http://ir3月份中国	日期	根据中国	2022-04-19 00:00:00	环比上升	疫情影响	原材料价格	运力供给						
2022-02	1月份中国	http://ir1月份中国	日期	2022年1月	2022-02-11 00:00:00	环比上升	下滑	整车指数	零担轻货	零担重货	疫情影响	春节影响			
2022-02	1月份中国	http://ir1月份中国	日期	2022年1月	2022-02-11 00:00:00	环比上升	下滑	整车指数	零担轻货	零担重货	疫情影响	春节影响			
2022-02	1月份中国	http://ir1月份中国	日期	2022年1月	2022-02-11 00:00:00	环比上升	下滑	整车指数	零担轻货	零担重货	疫情影响	春节影响			
2022-01	12月份中国	http://ir12月份中国	日期	根据中国	2022-01-07 00:00:00	增长	整车指数	零担轻货	零担重货						
2022-01	12月份中国	http://ir12月份中国	日期	根据中国	2022-01-07 00:00:00	增长	整车指数	零担轻货	零担重货						
2021-09	8月份中国	http://ir8月份中国	日期	2021年8月	2021-09-06 00:00:00	环比回升	增长	需求放缓	运价指数						
2021-08	7月份中国	http://ir7月份中国	日期	根据中国	2021-08-02 00:00:00	环比回落	PMI数据	短期季节	市场需求						
2021-08	7月份中国	http://ir7月份中国	日期	根据中国	2021-08-02 00:00:00	环比回落	PMI数据	短期季节	市场需求						
2021-07	6月份中国	http://ir6月份中国	日期	根据中国	2021-07-06 00:00:00	物流市场	运价指数	经济领土	公路运输						
2021-07	6月份中国	http://ir6月份中国	日期	根据中国	2021-07-06 00:00:00	物流市场	运价指数	经济领土	公路运输						
2021-07	6月份中国	http://ir6月份中国	日期	根据中国	2021-07-06 00:00:00	物流市场	运价指数	经济领土	公路运输						
2021-06	5月份中国	http://ir5月份中国	日期	根据中国	2021-07-06 00:00:00	物流市场	运价指数	经济领土	公路运输						
2021-06	5月份中国	http://ir5月份中国	日期	2021年5月	2021-06-03 00:00:00	经济运行	PMI数据	原材料市	运力供给	降雨季节	宏观调控				
2021-06	5月份中国	http://ir5月份中国	日期	2021年5月	2021-06-03 00:00:00	经济运行	PMI数据	原材料市	运力供给	降雨季节	宏观调控				
2021-06	5月份中国	http://ir5月份中国	日期	2021年5月	2021-06-03 00:00:00	经济运行	PMI数据	原材料市	运力供给	降雨季节	宏观调控				
2021-05	4月份中国	http://ir4月份中国	日期	2021年4月	2021-05-06 00:00:00	林安物流	整车指数	零担轻货	零担重货	PMI指数					
2021-05	4月份中国	http://ir4月份中国	日期	2021年4月	2021-05-06 00:00:00	林安物流	整车指数	零担轻货	零担重货	PMI指数					
2021-05	4月份中国	http://ir4月份中国	日期	2021年4月	2021-05-06 00:00:00	林安物流	整车指数	零担轻货	零担重货	PMI指数					

4.3.2. 标签编码

常规机器学习操作包括去除缺失值和对字符串数据进行标签编码。编码后的结果如下图所示:

	运输方式	路线	环比	日期	年份月份	字段1	回答	概要	日期_y	\n关键词1	\n关键词2	\n关键词3	\n关键词4	\n关键词5	\n关键词6
0	0	0	3	32	9	5	0	4	9	3	3	5	5	3	2
1	0	0	3	31	9	5	0	4	9	3	3	5	5	3	2
2	0	0	3	30	9	5	0	4	9	3	3	5	5	3	2
3	0	0	3	29	9	5	0	4	9	3	3	5	5	3	2
4	0	0	3	28	8	3	0	7	8	8	7	6	7	3	2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
655	0	19	14	4	1	4	0	0	1	0	2	4	4	0	2
656	0	19	14	3	0	2	0	9	0	7	6	7	8	3	2
657	0	19	14	2	0	2	0	9	0	7	6	7	8	3	2
658	0	19	14	1	0	2	0	9	0	7	6	7	8	3	2
659	0	19	14	0	0	2	0	9	0	7	6	7	8	3	2

图表 14 标签编码

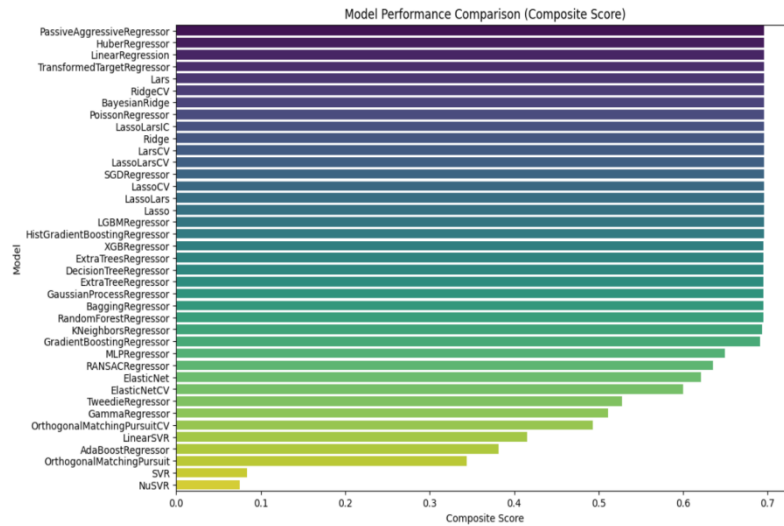
4.4. 模型构建

4.4.1. 算法选择与性能比较

为了寻找最优算法，首先使用 lazypredict 生成模型得分表。得分的计算方式是对参数进行加权，然后计算综合得分。
参数和权重如下：

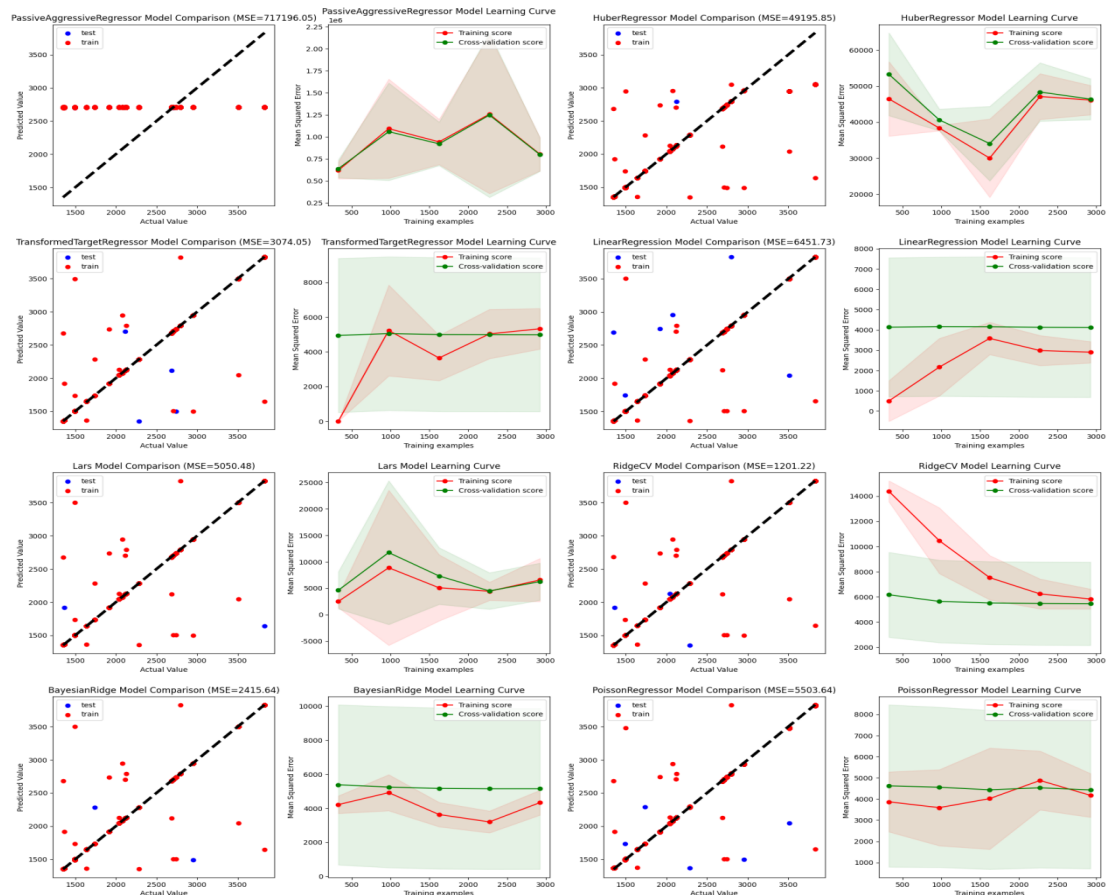
Adjusted R-Squared	0.4
R-Squared	0.3
RMSE	0.2
Time Taken	0.1

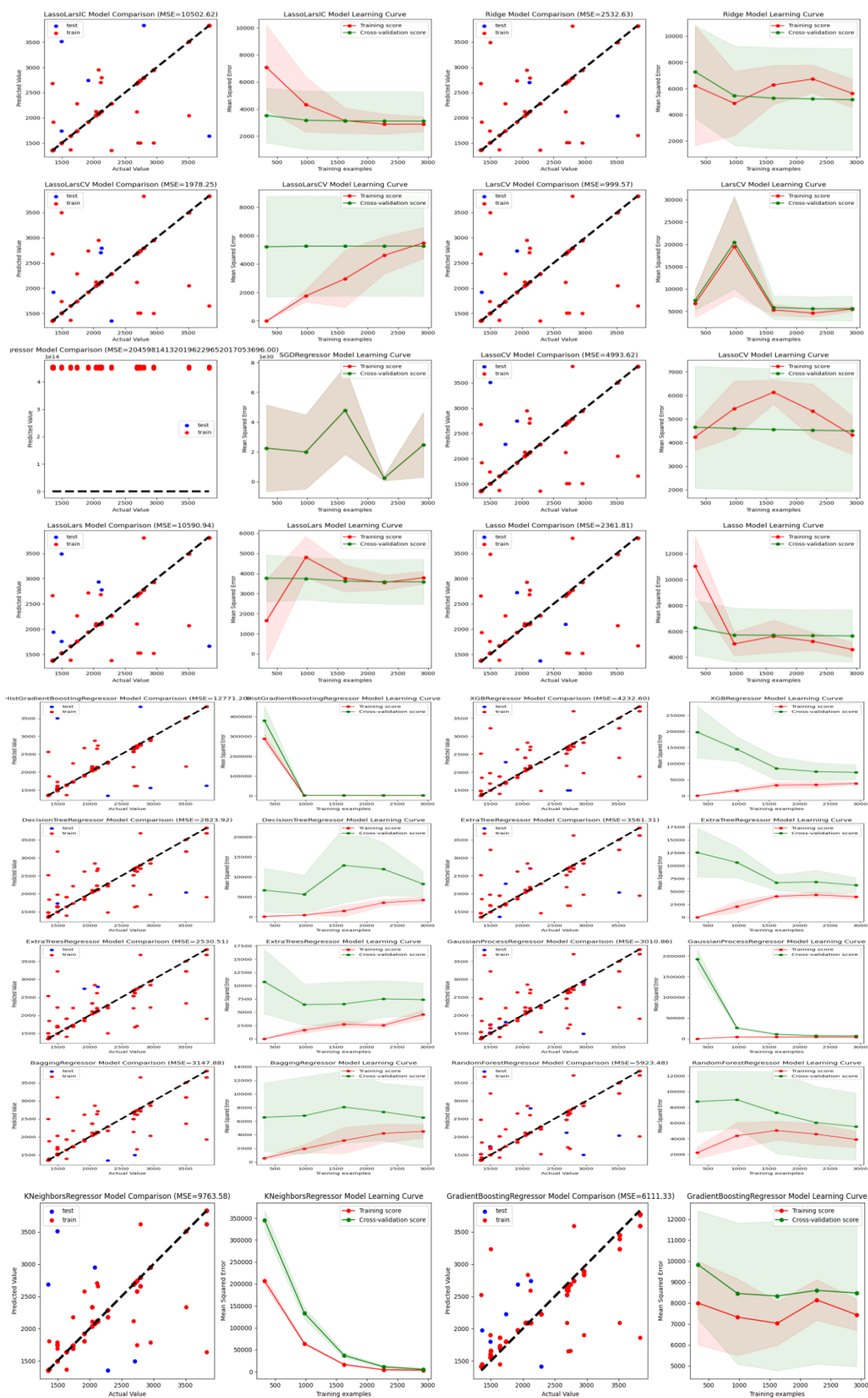
通过处理去除了得分为负数的模型，以下是经过筛选的得分为正数的模型：



图表 15 得分为正数的模型

观察得分前几的模型，例如 PassiveAggressiveRegressor、HuberRegressor 等，它们的得分大多在 0.7 左右，相差不大。因此，绘制学习曲线以便找到最优模型。通过分析学习曲线，可以确定是过拟合还是欠拟合，并找出训练效果最好的模型。

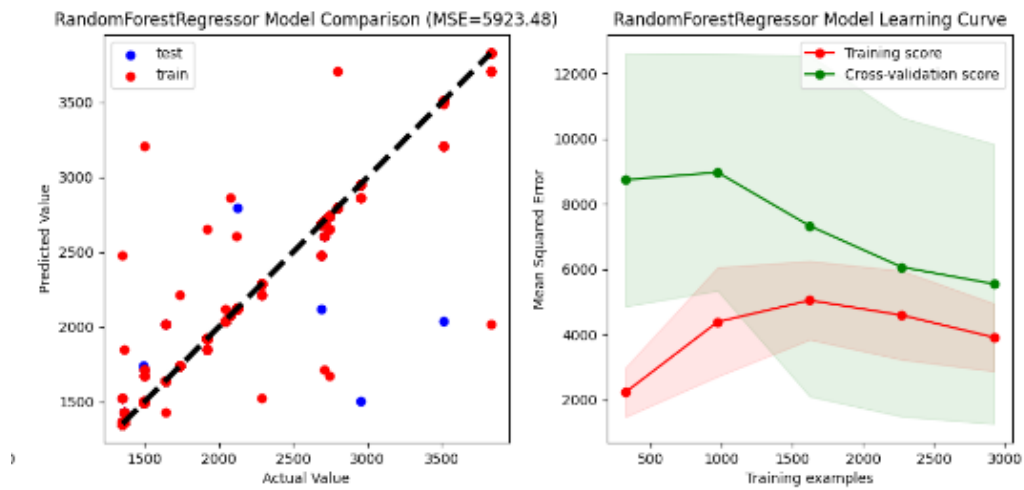




图表 16 学习曲线分析图

分析方法如下：

对于随机森林模型（RandomForestRegressor）：



图表 17 随机森林模型

实际值与预测值的散点图（左侧图）：观察散点图，实际值与预测值越接近标准线（图中虚线），说明模型的预测越准确。从图中可以看出，在训练集上的测试结果比较准确，而在测试集上略有差异，但大部分模型的实际值与预测值的差异并不大。

学习曲线图（右侧图）：图中采用了交叉验证，纵坐标为均方误差（Mean Squared Error）。随着训练样本的增加，无论是对于训练集还是交叉验证集，随机森林模型的均方误差逐渐下降。这表明模型既没有发生欠拟合也没有过拟合的情况。

综合分析：

通过比较多个模型的实际值与预测值的散点图和学习曲线图，可以得出随机森林（RandomForestRegressor）模型是比较适用的。

4.4.2. 最优参数寻找

使用随机搜索算法（random_search），自动搜索最优的深度、叶子节点数等参数，得到最优参数如下：

最大深度（max_depth）：6

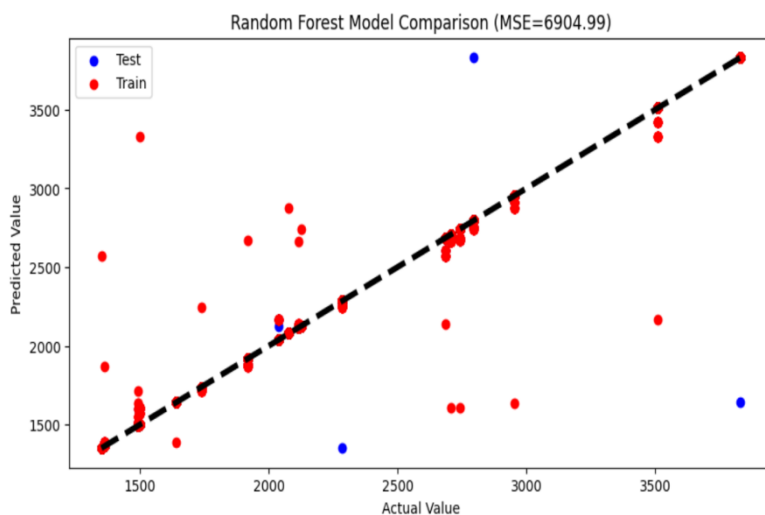
最小叶子节点数（min_samples_leaf）：11

最小分割样本数（min_samples_split）：4

树的数量（n_estimators）：449

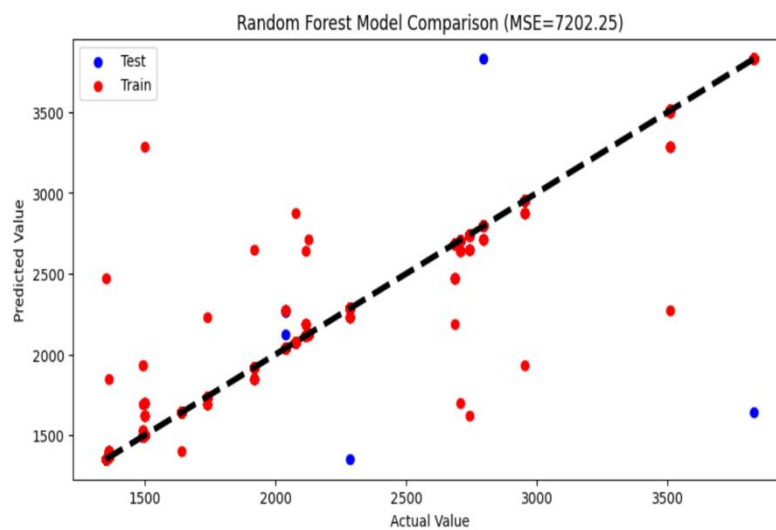
性能对比图如下：

最优参数：



图表 18 最优参数

默认参数：



图表 19 默认参数

可以看出，在最优参数情况下，均方误差（MSE）比默认参数小很多，这表明优化了随机森林模型的性能。同时，观察曲线的散点分布情况，可以发现最优参数下的点更靠近于标准线，这说明预测能力可以进一步提升。

4.4.3. 随机森林的运价模型解析

随机森林是一种集成学习方法，通过组合多个决策树来进行预测。其主要特点包括：

- **随机选择特征子集：**在构建每棵决策树的过程中，随机选择一部分特征进行划分，这有助于减少过拟合的风险。
- **随机选择样本：**对训练数据进行有放回的抽样，从中选择一部分样本来训练每棵决策树，这有助于增加模型的多样性。
- **基于投票或平均值进行预测：**对于分类问题，随机森林通过投票来确定最终的预测类别；对于回归问题，取所有决策树预测结果的平均值作为最终的预测值。

4.4.4. 随机森林模型构建与预测函数

随机森林的模型构建过程包括：

1. 从原始数据中随机选择一个样本集（有放回抽样），作为训练集。
2. 从所有特征中随机选择一个特征子集。
3. 根据选定的样本集和特征子集训练一个决策树。
4. 重复上述步骤多次，构建多棵决策树。
5. 对于分类问题，通过投票选择最终的预测类别；对于回归问题，取所有决策树预测结果的平均值作为最终的预测值。

决策树的预测函数表示为：

- **分类问题的预测函数：**

$$H(x) = \text{mode}h(x)$$

，其中

mode

表示取众数操作。

- **回归问题的预测函数：**

$$H(x) = \frac{1}{K} \sum_{k=1}^K h_k(x)$$

，其中 K 是决策树的数量。

4.5. 运价辅助决策模型的总结和使用

整车物流运价辅助决策模型的构建至此完成。尽管模型仍有进一步提升的空

间，但在比赛时间和资源限制下，只能达到有限情况下的最优解。
该模型极大地简化了运价预测过程，使用者只需加载已经训练好的模型，并提供符合特征值的数据，即可进行一定程度的运价预测。
具体使用方法如下：

```
python

import joblib

# 加载保存的模型

rf_model = joblib.load('random_forest_model.pkl')

# 预测值

y_pred = rf_model.predict(X_test)
```

其中`X\_test`为测试集的特征数据。模型的输入是特征向量，输出是目标变量的预测值。

示例测试集：

运输方式	路线	日期	关键词 1	关键词 2	关键词 3
整车运输-17.5	上海-南京	2024/3/6	疫情	物流需求	价格变动
整车运输-17.5	上海-南京	2024/2/28	疫情	物流需求	价格变动
整车运输-17.5	上海-南京	2024/2/21	疫情	物流需求	价格变动

图表 20 示例测试集

第五章 利用 Dijkstra 算法优化运输线路

5.1. 基于启发式算法的合理取货半径

5.1.1. 算法思路

- (1) 设定初始解空间，即可能的取货半径范围，初始化温度和冷却率。
- (2) 生成初始解，即随机选择一个初始的取货半径。在每一次迭代中，根据当前解生成一个邻域解，可以通过微调当前解来获得。
- (3) 计算当前解和邻域解的目标函数值。计算接受概率，根据 Metropolis 准则或 Boltzmann 分布来确定接受概率。根据接受概率决定是否接受邻域解作为新的当前解。
- (4) 更新当前解和温度。
- (5) 重复上述步骤直至满足终止条件。

5.1.2. 算法实现

定义目标函数 $f(r)$ 为取货半径为 r 时的总路程。

设 D_{ij} 表示从客户点 i 到客户点 j 的距离。

则目标函数 $f(r)$ 可以表示为：

$$f(r) = \sum_{i=1}^n D_{0i} + \sum_{i=1}^n D_{i0}$$

其中 n 为客户点数量， D_{0i} 表示从起始位置到客户点 i 的距离， D_{i0} 表示从客户点 i 到起始位置的距离。

在模拟退火算法的迭代过程中，需要根据当前解 r 和邻域解 r' 计算目标函数值，并根据 Metropolis 准则计算接受概率：

$$P(r' \rightarrow r) = e^{-\frac{f(r')-f(r)}{T}}$$

其中 T 为当前温度。

根据接受概率和随机数决定是否接受邻域解作为新的当前解。

5.1.3. 结论

模拟退火算法的目标是最小化总路程，即货车从起始位置出发，依次取货

再送达目的地所需的总路程。根据模拟退火算法得出的合理取货半径为 200 千米，在这个范围内，总路程相对较短，货车能够高效地完成货物的取送任务。

5.2. 基于合理取货半径的路径规划

5.2.1. 算法思路

在设定出发点位和目的地点位时，将一级集散中心作为可能的出发点位。这些集散中心通常具备较大的仓储空间和高效的物流处理能力，能够迅速将整车从生产线或仓库中调出，并安排合理的运输计划。

而经销商则被设定为可能的目的地点位。作为销售终端，经销商需要定期接收来自集散中心的整车，以满足市场需求。因此，需要根据经销商的分布情况和销售需求，合理规划运输路径，确保整车能够按时、按量地送达目的地。

整车物流从一级集散中心出发，当其 200 公里范围内存在目标点位时，整车物流可直接由一级集散中心出发前往目标经销商，当作为出发点的一级集散中心与目标经销商距离大于 200 公里时，必须经过其他二级集散中心进行整顿，再出发前往目标经销商。

5.2.2. 数据处理

关于路线距离

为了实现路线的可视化，调用了高德开放平台，利用平台所提供的路线规划 API 服务及其对外开放的道路数据，实现了在 python 中批量导入关于所选取城市点位的坐标，根据坐标构建请求参数，获得所有所选点位之间的距离参数，返回城市与城市的距离矩阵；

关于通过高德 API 获取道路路线代码部分

```
import requests
import json
import math

# 你的高德地图 API Key
amap_key = '41d5b8ee810bd7ad37b974f68d080313'

# 建立关于城市的编号
city_names=['湖州市','南京市','连云港市','嵊州市','蚌埠市','黄山市','永康市',
            '上海市','杭州市','宁波市','芜湖市','徐州市','苏州市','合肥市','玉
```

```

环县', '奉化市', '舟山市',

    '丽水市', '桐乡市', '乐清市', '苍南', '临海市', '温岭市', '上虞市', '富阳
市', '衢州市', '东阳市',

    '余姚市', '瑞安市', '诸暨市', '萧山市', '余杭', '义乌市', '慈溪市', '绍兴
市', '温州市', '台州市',

    '嘉兴市', '金华市', '金坛市', '溧阳市', '丹阳市', '泰兴市', '启东市', '如
皋市', '宜兴市', '海安',

    '江阴市', '太仓市', '张家港市', '淮安市', '宿迁市', '常熟市', '泰州市', '
镇江市', '盐城市', '吴江市',

    '扬州市', '昆山市', '常州市', '无锡市', '南通市', '铜陵市', '池州市', '淮
北市', '亳州市', '马鞍山市',

    '淮南市', '安庆市', '宿州市', '滁州市', '六安市', '阜阳市', ]

ids = list(range(len(city_names)))
list_cities = [ids, city_names]

city_coords = {
    '湖州市': (120.086881, 30.894178, ),
    '南京市': (118.796624, 32.059344, ),
    '连云港市': (119.221487, 34.596639, ),
    '嵊州市': (120.830505, 29.5615199, ),
    '蚌埠市': (117.388566, 32.916820, ),
    '黄山市': (118.337643, 29.714886, ),
    '永康市': (120.047356, 28.888990, ),
}

# 初始化一个距离矩阵，填充为无穷大
distance_matrix = [[float('inf')] * len(city_names) for _ in
range(len(city_names))]

# 设置对角线上的距离为 0
for i in range(len(city_names)):
    distance_matrix[i][i] = 0

# 路线规划 API 的基础地址
url = 'https://restapi.amap.com/v3/direction/driving'

for i in range(len(city_names)):

```

```

    for j in range(i+1, len(city_names)):
        origin_name = city_names[i]
        destination_name = city_names[j]
        else:
            print(f"未找到{origin_name}或{destination_name}的坐标信息")
    ")

# 打印距离矩阵
for row in distance_matrix:
    print(' '.join(map(lambda x: f"{x:.2f}" if isinstance(x, float)
and not math.isinf(x) else str(x), row)))

```

关于将 excel 数据导入 pathon 代码部分

```

import pandas as pd
import numpy as np

# 假设你的数字字符串如下
numbers_str = '''
0    202833 540523 215284 395333 268746 284995 153751 87151  241592
217209 539246 97251  353929 401415 255236 301618 345052 66477  399302
463494 306932 372321 164081 120064 299642 233683 192635 426573 170396
97456  87776  216895 207563 139274 388116 345738 89051  261047 157192
102010 199533 229059 269863 247226 69236  276112 189256 168587 184960
356046 449546 148100 216222 194105 354844 81476  220510 134906 126474
520862 332057 479152 726916 197301 505420 755107 628803 596325
751128 352135 287418 545221 220741 911057 768147 797347 814091 580294
909794 953686 817010 884858 675846 606427 620329 730183 701306 925991
667989 610306 577378 714549 703032 649760 888198 856342 602692 703148
423831 424991 416659 462158 661572 519058 462570 498333 501540 594915
523709 435911 377946 548110 452302 405889 543340 570590 394546 579656
460533 511017 578029 379957 354860 209329 118571 356170 136024 359229
220403 287212 167325 0
'''

# 遍历每一行，分割空格并转换为数字（或处理 inf）
for line in lines:
    row_numbers = []
    for num_str in line.split():

```

```

# 尝试将字符串转换为浮点数
try:
    num = float(num_str)
    # 检查是否为 inf, 并替换为 NaN 或其他值
    if num == float('inf'):
        num = np.nan # 或者你可以替换为其他值, 比如 0 或某个特定的占位符
    row_numbers.append(num)
except ValueError:
    # 如果转换失败 (比如非数字字符串), 则忽略或添加错误处理
    row_numbers.append(np.nan)
rows.append(row_numbers)

# 创建 DataFrame
df = pd.DataFrame(rows)
df.to_excel('distance.xlsx', index=False, header=False)

```

关于运价，通过 AI 大模型分析已有数据，并用随机森林算法对其进行学习预测，得到关于部分城市的整车物流的单位车次运价，以出发地点、目的地以及整车运输的单位车次运价为基础，可以构建一个运价与有向向量之间的关系表格，它反映了不同起点和终点之间整车运输的单位成本。为了深入分析这些数据并找出其中的规律，通过计算使得预测函数和实际值的残差最小的方式求得最佳线性函数。通过 SPSS，可以对关系表格进行一系列的数据分析，包括描述性统计、相关性分析等，以揭示不同城市对之间运价的可能关联和影响因素。

最小二乘法原理：

$$\text{Min } [S(b, a) = \sum_{i=1}^n (y_i - (bx_i + a))^2]$$

更为重要的是，还可以利用 SPSS 进行线性回归分析。线性回归是一种统计方法，用来预测因变量（在这里是单位车次运价）随城市之间的路线的变化情况。可以使用最小二乘法，通过线性回归分析，可以提取出已知运价的相关集散中心的特征值（x, y, distance），估计出模型中各个参数的数值，再根据其他集散中心的已知特征（x, y, distance），从而得到一个可以用来预测其他集散中心之间整车运输单位车次运价的公式或模型，从而构建出各集散中心的单位运价矩阵。

ANOVA^a

模型	平方和	自由度	均方	F	显著性
----	-----	-----	----	---	-----

1	回归	7991383.533	3	2663794.511	2.427	.103 ^b
	残差	17562081.670	16	1097630.104		
	总计	25553465.200	19			

a. 因变量：整车运输的单位车次运价

b. 预测变量：(常量), distance, x, y

		系数 ^a						
模型		未标准化系数		标准化系数	t	显著性	B 的 95.0% 置信区间	
		B	标准误差	Beta			下限	上限
1	(常量)	2203.098	602.587		3.656	.002	925.670	3480.526
	y	98.866	249.174	.119	.397	.697	-429.359	627.091
	x	-70.282	128.663	-.164	-.546	.592	-343.036	202.472
	distance	4.893	2.055	.493	2.381	.030	.536	9.250

a. 因变量：整车运输的单位车次运价

5.2.3. 算法实现

Dijkstra 算法是一种用于解决带权有向图中的单源最短路径问题的贪心算法。其核心思想是从图中选出一个节点作为起点，以该节点为中心，逐步向外扩展，直到找到通往图中所有其他节点的最短路径；

与 Floyd 算法相比，Dijkstra 算法更适用于已知两点间的最短路径计算，在计算速度上更优于 Floyd 算法；

Dijkstra 算法是从图中选出任一级集散中心作为起点，并标记为已访问。从图中选出一个未访问的节点，该节点到起点的距离是最小的（即当前的最短路径长度）。更新图中所有与选定节点相邻的节点的距离，如果通过该节点到达相邻节点的距离小于或等于该节点原来的距离，就更新这个节点的距离。将更新后的节点加入到优先队列中，以便进一步处理。重复步骤，直到优先队列为空或所有节点都被访问。

在整车物流环节中，通过结合距离矩阵和运价矩阵，并应用相应的系数，可以构建一个全新的嵌合矩阵。这一矩阵旨在更精准地反映整车物流的实际情况，为后续的路径优化提供数据基础。接着，利用 Dijkstra 算法，可以根据给定的出发点和目的地点，快速找出两者之间的预测最优路径。

为了将这一最优路径以直观的方式呈现出来，引用了 Python 中的多个强大库。其中，pandas 库用于数据处理和分析，帮助从复杂的数据中提取有用的信息；polyline 库则用于处理路径的几何信息，确保路径的连续性和准确性；cartopy 库则提供了丰富的地图绘制功能，使能够在地图上展示路径；而

matplotlib.pyplot 库则用于生成高质量的图表和图像。

此外，还借助了高德开放平台提供的 API 服务。通过发送请求并接收响应，能够获取到实时的路况信息、地理坐标等关键数据，进一步提升路径优化的准确性和实用性。

为了将这一最优路径以直观的方式呈现出来，引用了 Python 中的多个强大库。其中，pandas 库用于数据处理和分析，帮助从复杂的数据中提取有用的信息；polyline 库则用于处理路径的几何信息，确保路径的连续性和准确性；folium 库则提供了丰富的地图绘制功能，此外，还调用了高德卫星图像，并借助了高德开放平台提供的 API 服务，通过发送请求并接收响应，能够获取到实时的路况信息、地理坐标等关键数据，进一步提升路径优化的准确性和实用性。

5.2.4. 部分算法代码

```
import copy
import math
import pandas as pd
import json
from polyline import polyline
import requests
import numpy as np
import folium
from folium import plugins

    dist = distance[1][start]
    return path, dist

#建立关于城市的编号
city_names=['湖州市','南京市','连云港市','嵊州市','蚌埠市','黄山市','永康市',
            '上海市','杭州市','宁波市','芜湖市','徐州市','苏州市','合肥市','玉环县',
            '奉化市','舟山市','淮南市','安庆市','宿州市','滁州市','六安市','阜阳市',
            ],
ids = list(range(len(city_names)))
list_cities = [ids, city_names]
#建立关于城市的地理坐标
coordinates = {
    '湖州市': (120.086881,30.894178),
```

```

'南京市': (118.796624, 32.059344),
'连云港市': (119.221487, 34.596639),
'嵊州市': (120.830505, 29.5615199),
'蚌埠市': (117.388566, 32.916820),
'黄山市': (118.337643, 29.714886),
'永康市': (120.047356, 28.888990),

'上海市': (121.473667, 31.230525),
'杭州市': (120.210792, 30.246026),
'宁波市': (121.624540, 29.860258),

}

file_path = 'distance.xlsx'
# 读取 Excel 文件
try:
    df = pd.read_excel(file_path)
except Exception as e:
    df = None

# 检查 DataFrame 是否为空
if df is not None:
    # 将 DataFrame 转换为二维列表（即矩阵）
    distance_matrix = df.values.tolist()

    # 打印 DataFrame 以确认是否成功读取
    print("DataFrame 读取成功！")
else:
    print("DataFrame 读取失败！")

    # 去掉最后一个多余的分号（如果存在）
    merged_coordinates = merged_coordinates.rstrip(';')

# 给出的坐标系为 GCJ-02，如果需要测试 google 地图，需要进行坐标转换

```

```

tri = np.array(list(zip(Lat, Lon)))

start_coordinates = coordinates[city_names[jiedian[0]]]
end_coordinates = coordinates[city_names[jiedian[-1]]]
san_map = folium.Map(
    location=[latitude_1, longitude_1],
    zoom_start=10,
    # 高德街道图
    #
    tiles='http://webrd02.is.autonavi.com/appmaptile?lang=zh_cn&size=1&scale=1&style=7&x={x}&y={y}&z={z}',

    tiles='http://webst02.is.autonavi.com/appmaptile?style=6&x={x}&y={y}&z={z}', # 高德卫星图
    attr='default'
# 创建并添加终点标记
folium.Marker(
    location=[latitude_2, longitude_2],
    popup='终点',
    icon=folium.Icon(color='red', icon='fa-flag')
).add_to(san_map)

for lat, lon in zip(Lat, Lon):
    folium.Marker([lat, lon], color='red').add_to(marker_cluster)
san_map.save('showpoint.html')

```

5.3. 算法优化

5.3.1. 运输过程磨损费用

在整车物流的过程中，除了考虑到货物的安全和及时交付外，还必须充分考虑到整车运输过程中可能产生的车辆磨损费用。这一项费用可以被视为一个关键的参量，需要根据实际情况进行灵活调整和评估。

5.3.2. 中转费用

在整车物流的规划和执行中，还需要充分考虑到中转过程中可能产生的费用。这些中转费用包括但不限于装卸费、仓储费、中转站服务等。将中转费用作为一个可调节的参量，根据不同的地域，运输方式适当调整。通过优化物流方案，降低中转成本，提高运输效率，确保货物能够在最短的时间内以最经济的方式达到目的地。

5.3.3. 基于动态规划的实际运输时效问题

针对实际运输过程中受到天气、节假日、车流量等因素影响的问题，设计一个基于动态规划的算法，以实时数据为基础，对运输时效进行估算和优化。

状态定义：

将运输过程划分为不同的状态，考虑到天气、节假日、车流量等因素，定义每个状态下的运输时间和成本。

状态转移：

根据不同状态下的运输条件和成本，设计状态之间的转移规则。例如，若遇到恶劣天气，则运输时间可能会延长；若是节假日，则可能会有交通管制或交通疏导等情况。

动态规划求解：

利用动态规划算法，从起始状态出发，逐步计算出每个状态下的最优解，并记录路径。在计算过程中，考虑到不同因素对运输时效的影响，并进行合理的权衡和调整。

路径选择与优化：

根据求解出的最优路径，结合实际情况和运输需求，选择合适的路径进行运输。在路径选择过程中，需要综合考虑运输时效、成本、安全等因素，并进行优化调整。

第六章 整车物流大屏设计

6.1. 可视化大屏整体设计

将各模块的数据将以直观、清晰的形式呈现在整车物流大屏上，为用户提供全面、准确的运输信息，帮助做出最佳的运输决策。

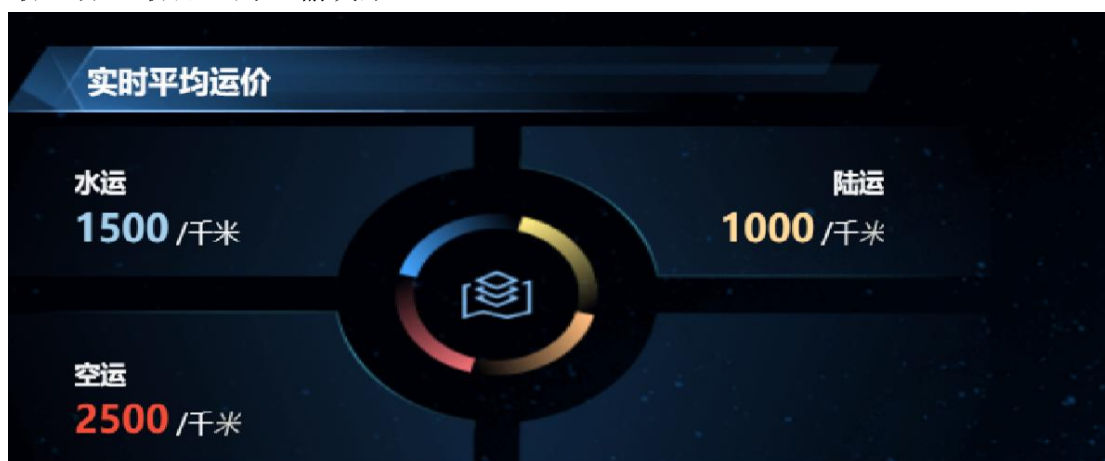


图表 21 整车物流运输大屏展示

6.2. 实时平均运价

仪表板 ID: 01; 组件标题: 实时平均运价; 组件类型: 展示表。

这一部分将针对不同的运输方式（水运、空运、陆运）提供当前的平均运价信息。运价数据将通过实时监测系统获取，并直接在大屏上展示，使用户可以及时了解各种运输方式的价格水平。这有助于用户根据实时的运价情况做出最经济、最合适的运输决策。

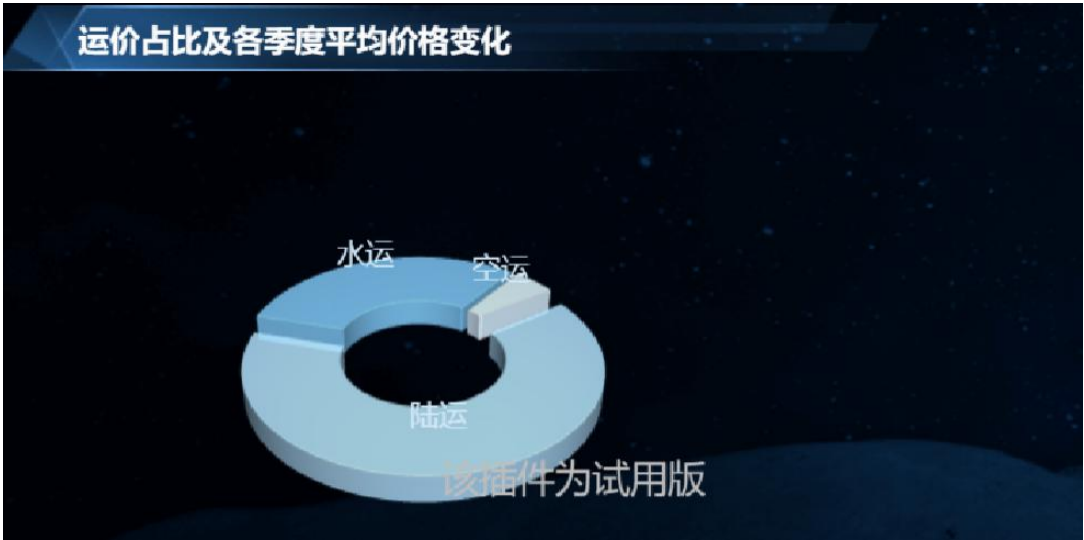


6.3. 运价占比

仪表板 ID: 02; 组件标题: 运价占比; 组件类型: 饼状图; 分类字段: 运输方式。

此模块将以图表形式展示长江三角区整车物流运输中各种运输方式的占比情况。用户可以清晰地看到不同运输方式在整车物流中的比重，以便更好地了

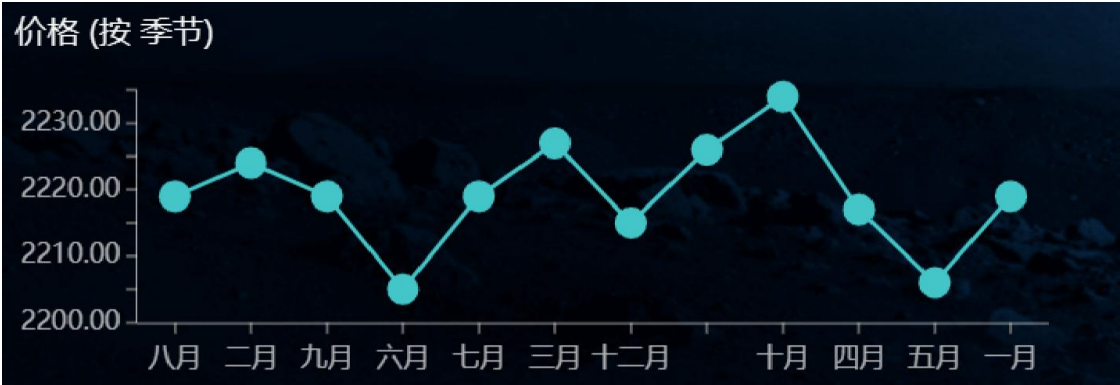
解行业的整体结构和趋势。用户在制定物流策略和规划运输路线时，可以考虑到主流运输方式的选择。



6.4. 按季度平均价格变化

仪表板 ID: 03; 组件标题: 按季度平均价格变化; 组件类型: 折线图; 分类字段: 月份; 数值字段: 平均运价。

此模块以对比图的形式展示了各个月份的平均价格的变化情况。由于不同季度的需求量不同，运价会在不同的季度出现波动，这有助于用户更好地了解运价的季节性变化规律。

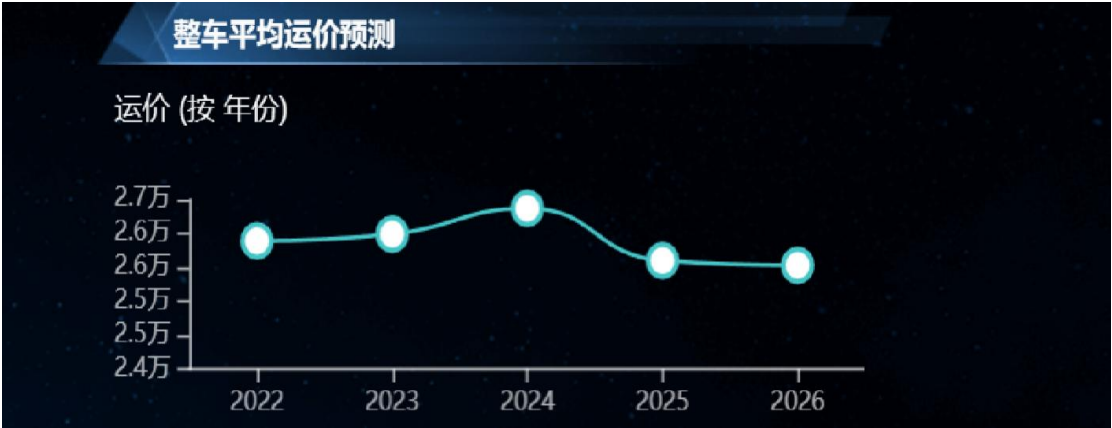


6.5. 整车平均运价预测

仪表板 ID: 04; 组件标题: 运价预测; 组件类型: 折线图; 分类字段: 年份; 数值字段: 运价。

这一模块将通过对比过去年份的运价进行分析，利用预测算法推算出未来可能的运价趋势。这有助于行业者提前了解未来运价的走势，从而做出更加准确

的运输成本预算和规划。例如，如果预测显示未来运价可能上涨，用户可以考虑提前采购或调整运输策略以应对可能的成本增加。



6.6. 集散中心选址

仪表板 ID: 05; 组件标题: 集散中心选址; 组件类型: 展示图; 分类字段: 选址地区。

我国针对整车物流结构进行了优化，此模块展示了水路、公路等各种运输方式下的优势集散中心选址情况。这些集散中心充分发挥着节点对于运输的关键作用，为整个物流系统的高效运作提供了支撑。

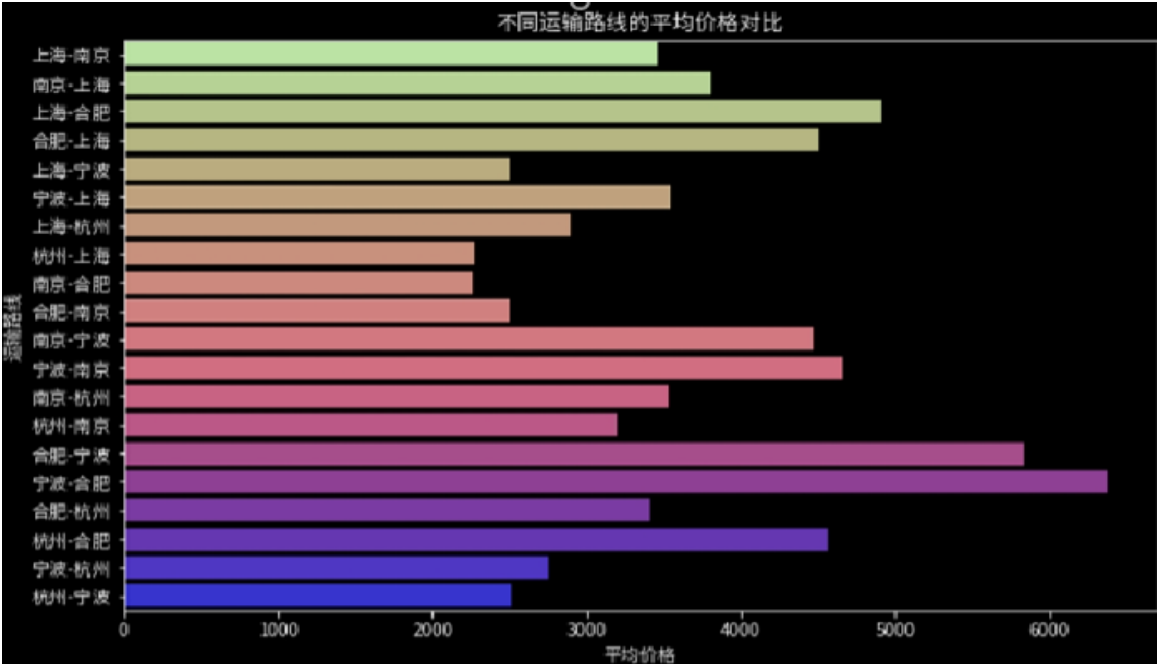


6.7. 不同路线运价对比图

仪表板 ID: 06; 组件标题: 不同路线运价对比; 组件类型: 柱状图; 分类字段: 路线; 数值字段: 运价。

这一模块将展示长江三角洲地区主要路线中运价的对比情况。不同路线的

运价可能受到地理、交通、市场需求等因素的影响，因此价格可能存在较大差异。通过此模块，用户可以清晰地比较不同路线的运价情况，从而选择最经济、最合适的运输路线。这有助于用户降低运输成本，提高运输效率。



第七章 项目总结

7.1. 分工与协作

作品的完成主要分为以下的三个阶段：

- 前期阶段：小组成员有着明确的分工包括作品的前期设计、文案的编辑和排版、算法、模型、以及数据大屏的前端设计。
- 在实施阶段：每周都会线下开会讨论，推进项目的实施，将本次作品拆分，由各自擅长的算法和模型技术进行分工，每个人负责本作品的一个部分。
- 最终阶段：经由讨论和作品汇总和指导老师对作品的指正，再进行修改和补充使作品更加的完善。

7.2. 面对的困难和的解决方案

团队是由不同专业组成的，在专业知识沟通方面以及技术交流方面存在部分障碍。本次作品也涉及到陌生新领域的开发，对于团队而言是个比较大的挑战。以下是面临的挑战和解决方案：

- **设计理念冲突：**早在项目的开始阶段，团队成员在设计理念上便存在诸多分歧。每个人都有自己独特的视角和创新思维，这使得确定统一的设计理念变得困难。
- **解决方案：**决定举行多次头脑风暴会议，使每个团队成员都有机会发表自己的观点。倾听并尊重每一个声音，并通过投票的方式确定了最终的设计理念。
- **技术瓶颈：**在实施设计过程中遇到了许多技术难题，比如功能实现的难度，以及某些设计的可操作性问题。
- **解决方案：**当遇到技术瓶颈时，齐心协力，积极寻求解决方案。进行大量的研究，查阅相关文献，甚至参加专业课程和研讨会，不断提升的技术水平。同时，也及时查找和使用现有的工具和框架来帮助解决问题。
- **时间管理：**虽然设定了明确的时间线和里程碑，但在项目进展过程中，发现个别任务的完成时间超出了预期，这给的时间管理带来了挑战。
- **解决方案：**为了更有效地管理时间，开始使用项目管理工具进行整体规划和个别任务追踪。通过这样的方式，可以更清晰地进行完各项任务。

7.3. 未来展望

希望通过的作品能够为长三角地区整车物流运输提供更好的借鉴，能够切实为企业提供可行的办法、及时响应国家号召推动我国物流的成本降低、推动整车物流运输效率的提高。甚至可以推动我国整车物流运输行业的进步！