

队伍编号	MC2302222
题号	C

基于 NSGA-II 遗传算法的物流网络调度运输和结构优化问题

摘 要

随着电商行业的快速发展，物流网络在保障电商行业正常运行中起到了举足轻重的作用。电商物流网络包括物流场地（接货仓、分中心、营业部等）和物流场地之间的运输线路。本文针对电商物流网络中的包裹应急调运与结构优化问题，展开深入研究。

针对问题一，为了建立线路货量的预测模型，预测 2023-01-01 至 2023-01-31 期间每条线路每天货量，本文构建了 SARIMAX 模型（具有外生变量的季节性和区域性整合移动平均模型）模型以提高模型的预测精度，先在时间上对货量数据进行差分，使其变得平稳后，再建立自回归模型和移动平均模型，将季节性因素和不由时间序列引起的外生变量纳入考虑，更准确预测未来货量值。最后，本文通过 SARIMAX 模型得出了线路 DC14→DC10、DC20→DC35、DC25→DC62 货运量的预测结果。

针对问题二，在关停 DC5 后，为了在将 DC5 货量分配到其他线路后满足使所有包裹正常流转，使货量线路的改变最小化，尽可能平衡各线路的工作负荷的三个要求，本文在预测出 DC5 的我们构建了以 NSGA-II 遗传算法为基础的模型，根据三个要求构建目标函数，制定以分配到每条线路的货运量组成的 n 维向量和分配的线路条数 n 作为决策变量，并根据货运量负载上限制定的约束条件，最后，本文将所有线路的线性规划问题进行联合优化，以得到最优的货量分配方案。

针对问题三，在关停 DC9 后，仍然需要满足使所有包裹正常流转，使货量线路的改变最小化，尽可能平衡各线路的工作负荷的三个要求，在此基础上，我们可以增删线路。在此，本文沿用了问题二中的 NSGA-II 遗传算法为基础的模型，在此，我们还引入了一个二元变量，以 0 和 1 分别代表线路的开放与否，使得优化器在寻找最优解时考虑线路是否开放的影响，并将其作为一个新的决策变量并引入该优化模型。最后，本文在次将所有线路的线性规划问题进行联合优化，以得到新的最优的货量分配方案。

针对问题四，研究对该物流网络的不同物流场地及线路的重要性进行评价的问题，本文选择了 TOPSIS 熵权法对其进行评估。本文首先构建了一个 AHP 层次分析法模型对 TOPSIS 熵权法中各指标的权重进行评估，其次，本文代入权重，使用 TOPSIS 熵权法，通过比较样本值与理想值的距离实现综合评价的方法，根据现有的数据，对各条线路进行评价排序。研究新增物流场地和线路的设置的问题，我们绘制了随机分布网络图，在对网络图节点进行聚类筛选后，我们得出对整个系统影响比较大的三个场地，据此，我们对新增的线路与对整个系统影响较大的三个场地进行了综合分析，得出了新增物流场地的处理能力及新增线路的运输能力应如何设置。最后，我们对我们的模型进行了鲁棒性分析。

关键词：电商物流网络；SARIMA 模型；NSGA-II 遗传算法；TOPSIS 熵权法；AHP 层次分析法；随机分布网络图

目录

基于 NSGA-II 遗传算法的物流网络调度运输和结构优化问题.....	1
摘 要.....	1
一、问题重述.....	1
1.1 问题背景.....	1
1.2 目标任务.....	1
二、模型假设及符号说明.....	2
2.1 模型假设.....	2
2.2 主要的符号说明.....	2
三、问题分析.....	3
3.1 针对问题一.....	3
3.2 针对问题二.....	3
3.3 针对问题三.....	3
3.4 针对问题四.....	3
四、问题一模型的建立与求解.....	4
4.1 历史数据预处理及可视化.....	4
4.2 历史货量数据分析.....	7
4.3 模型参数确定.....	9
4.3.1 模型阶数确定.....	11
4.3.2 模型参数估计.....	12
4.4 模型训练和诊断评估.....	12
4.5 模型预测.....	15
五、问题二模型的建立与求解.....	19
5.1 模型变量定义.....	19
5.1.1 参数.....	19
5.1.2 决策变量.....	19
5.2 定义目标函数.....	19
5.3 算法选择.....	19
5.4 定义约束条件.....	20
5.5 模型求解.....	20
六、问题三模型的建立与求解.....	22
6.1 模型变量定义.....	22
6.1.1 参数.....	23
6.1.2 决策变量.....	23
6.2 定义目标函数.....	23
6.3 算法选择与目标条件.....	24
6.4 模型求解.....	24
七、问题四模型的建立与求解.....	25
7.1 模型变量定义.....	25
7.1.1 模型的选取与假设.....	26
7.1.2 模型的建立与求解.....	26
7.2 模型选择与思路.....	27
7.3 利用 AHP 法对指标权重进行评估.....	27

7.4 熵权 TOPSIS 对场地重要性进行评估.....28

7.5 新增物流场地以及对其运输能力的分析.....29

八、模型评估.....30

九、参考文献..... 30

附录.....30

一、问题重述

1.1 问题背景

随着全球经济的快速发展和企业的扩张，物流网络运输的重要性不断提升。同时，各种自然灾害、社会突发事件和供应链中断等突发情况也层出不穷，诸如节假日促销活动，自然灾害、突发事件按等不可抗力因素，会导致电商用户的下单量会发生显著波动，出现物流场地被迫停用的情况，这将会对商家的货物流转产生一定程度的困难，增加运输成本，对物流网络运输产生较大的影响。为了解决这些问题，需要对物流网络运输进行应急调运和结构优化。应急调运是指在突发情况下，对物流网络运输进行及时的调配和调度，确保物流系统能够继续正常运转。而结构优化则是指通过优化物流系统的运输路线、运输方式、货物配送等方面，提高系统的效率和运输质量。措施具体可表现为被关停场地处理的包裹将被紧急分流到其余各线路，通过预测各物流场地及线路的货量，并据此合理安排运输、分拣等计划。物流系统的应急调运和结构优化，不仅可以帮助企业应对各种突发情况，提高业务的稳定性和安全性，最大程度地降低物流场地停用对物流网络的影响，保障物流网络的正常运行，从而降低运营成本，提高运营效率成为了目前电商的迫切需求。也可以为企业节约时间，提高运输效率和客户满意度。因此，物流网络运输应急调运及结构优化具有重要的现实意义和经济价值。

1.2 目标任务

问题一：

- 1、建立线路货量的预测模型，预测 2023-01-01 至 2023-01-31 期间每条线路每天货量。
- 2、给出其中的线路 DC14→DC10、DC20→DC35、DC25→DC62 的预测结果。

问题二：

- 1、假定场地 DC5 于 2023-01-01 开始关停，在此前预测基础上，将 DC5 货量分配到其他线路使所有包裹正常流转，并使货量线路的改变最小化，尽可能平衡各线路的工作负荷。
- 2、如有部分日期部分货量无法正常流转，也使其累计总量最小化。正常流转时，请给出因 DC5 关停导致货量发生变化的线路数及网络负荷情况；不能正常流转时，请给出因 DC5 关停导致货量发生变化的线路数、不能正常流转的货量及网络的负荷情况。

问题三：

- 1、在问题 2 中，如果被关停的物流场地为 DC9，允许对物流网络结构进行动态调整条件下，同样将 DC9 相关线路的货量分配到其他线路，使所有包裹尽可能正常流转，货量发生变化线路数最小化，且均衡各条线路的工作负荷。
- 2、如果存在部分日期没有满足要求的流转方案，应使期间未能正常流转的包裹日累计总量最小化。
- 3、正常流转时，给出因 DC9 关停导致货量发生变化的线路数及网络负荷情况；不能正常流转时，给出因 DC9 关停导致货量发生变化的线路数、不能正常流转的货量及网络的负荷情况，同时请给出每天的线路增减情况。

问题四：

- 1、据历史货量数据，对该网络的不同物流场地及线路的重要性进行评价。
- 2、如果打算新增物流场地及线路以改善网络性能，探讨分析新增场地应与哪几个已有场地互通线路，应如何设置新增物流场地的处理能力及新增线路的运输能力。
- 3、在考虑预测结果的随机性之上，探讨研究所建运输网络的鲁棒性。

二、模型假设及符号说明

2.1 模型假设

针对该题提出的问题，我们做出了如下的模型假设：

- 3、假设历史货量数据是可靠的，可以反映未来的货量变化趋势。
- 4、假设各条线路之间的货运量是相互独立的，运输时不会互相影响。
- 5、假设所有货物应该尽可能正常流转，不能出现滞留、丢失、损坏等情况。
- 6、假设物流场地的关停对网络的影响主要是货量分配不均和部分货物无法正常流转。
- 7、假设各个物流场地的处理能力和各条线路的运输能力限制是可靠的。
- 8、假设货量数据具有一定的周期性、趋势性和随机性。

2.2 主要的符号说明

符号	含义	单位	说明
X_i	线路 i 分配到的货量	件	
a_i	线路 i 的总流转货量	件	
k	DC5 线路货量能正常流 转的比率	%	
α	各线路货量标准差	件	代表工作负荷不均衡度
n	与 DC5 相关线路数		货量发生变化的线路数
X_n	DC5 分配给 n 条线路的货流量组成的 n 维向量		
ρ_k	ACF 图的一阶差分		
Φ_{kk}	PACF 图的一阶差分		

C_1	每条线路包裹运输总量	件
C_2	在历史数据时间内每条线路出现次数	
C_3	每条线路的最大运输货量	件

注：此为本文的主要符号说明，其他符号解释详见正文部分。

表 2.2 主要符号说明

三、问题分析

3.1 针对问题一

本题主要研究建立了一个线路货量的预测模型，对 2023 年 1 月期间每条线路每天的货量进行预测，并给出提交指定线路的预测结果。基于已有的截至预测期的前两年连续数据集，可得出其具有明显的时间依赖性、季节性、非特定性等特征，故选择采用时间序列相关的模型，此处我们采用常见的 ARIMA 模型先对货量序列的平稳性和自相关性进行检验和处理。然而，货量存在季节性变化规律，在促销活动期间，物流货量会迎来一年中的爆发式增长，因而在 ARIMA 模型的基础上引入季节性因素以及一些外生变量，例如促销活动和异常值等因素。因而我们最终选择 SARIMAX（具有外生变量的季节性和区域性整合移动平均模型）模型以提高模型的预测精度，先在时间上对货量数据进行差分，使其变得平稳后，再建立自回归模型和移动平均模型，将季节性因素和不由时间序列引起的外生变量纳入考虑，更准确预测未来货量值。

3.2 针对问题二

本题要求我们在问题 1 基础上，将 DC5 相关线路的货量分配到其他线路，使所有包裹可以正常流转，同时保持各线路符合均衡，最小化货量发生改变的线路数、未能流转的包裹累积量。我们采用基于 NSGA-II 遗传算法的多目标线性规划模型，具体地，我们先预测出 DC5 的出货量与入货量，将每条线路的货量分配问题看作一个线性规划问题，目标函数使尽可能使因 DC5 关停而变化的线路数，约束条件是使各线路共工作符合尽可能均衡，且所有包裹尽可能正常流转。最后我们将所有线路的线性规划问题进行联合优化，以得到最优的货量分配方案。

3.3 针对问题三

本题要求我们将 DC9 相关路线的货量分配到其他线路，并最小化 DC9 关停前后货量发

生变化的线路数、各条线路的工作负荷的不均衡度、无法正常流转的日累计货量。与问题 2 类似，我们将该问题的目标转化为一个线性规划问题，目标函数与约束条件与问题 2 相似。在此前模型基础上扩展到考虑允许每日关闭或新开线路，构建的规划方程需要将 DC9 相关线路的货量分配到其他线路。求解改进模型时，我们采用了一种滚动优化策略，即根据每天预测货量、已有线路和当前网络状态，重新调整线路分配方案。求解过程中，我们采用了启发式算法以提高求解效率，并在求解后对每天线路增减情况进行了统计。

3.4 针对问题四

本题要求对该物流网络的不同物流场地及线路的重要性进行评价，探讨新增物流场地和线路的设置，以及如何保证网络的鲁棒性。首先需要对不同场地和线路的重要性进行评估，在此，我们选择了熵权 TOPSIS 法，我们还构建了一个 AHP 层次分析法模型对 TOPSIS 中各指标的权重进行评估，以保证我们模型的准确性。其次，我们需要考虑是否增加场地及线路以及它们的指标选择的问题。在此，我们绘制了随机分布网络图，在对网络图节点进行聚类筛选后，我们得出对整个系统影响比较大的三个场地，据此，我们对新增的线路与场地进行了分析。最后，我们对我们的模型进行了鲁棒性分析。

四、问题一模型的建立与求解

4.1 历史数据预处理及可视化

我们对 2021 年到 2022 年每条线路的每天的货运量数据进行 SARIMAX 模型的数据预处理，包括数据清洗：缺失值和异常值处理；时间序列拆分：分为趋势性、季节性、随机性；特征提取：历史货量、节假日、天气等，为更好地理解货量数据的趋势、季节性和异常点等信息，我们采用绘制时间序列图、自相关图和偏自相关图等方式来观察数据的趋势和周期性，确保数据的平稳性和可处理性。

具体地,我们对收集到的线路的信息利用 python 中 pandas 库的 DataFrame 进行数据清洗,忽略过大过小值影响,接着将数据分类,历史数据多的应用到预测模型,历史数据少的进行求平均值。处理货量时序数据时,根据不同场地先划分路线,再按照日期进行排序,随后使用 ADF(Augmented Dicky-Fuller)检验进行了时间序列货量数据平稳性检测,考虑历史货量、季节性、趋势性等因素,使用折线图等方式呈现原始时间序列数据的趋势和季节性,了解数据在时间上的变化规律、季节性的影响以及可能存在的异常点。

首先对清洗过后的历史数据进行折线图和直方图的可视化操作(见图 4.1.1 和图 4.1.2),但由于该图重叠性和复杂性过高无法体现历史数据特征,故我们对 2021 年到 2022 年每个月所有线路货运量的数据以月为单位进行加和,重新绘制出以时间为横坐标,货运量为纵坐标的折线图(见图 4.1.1),我们可以看到在 3 月、6 月、9 月、11 月货流量在短时间存在激增,推测出这是促销活动导致的,因此引入外生变量来使我们的预测更加准确。

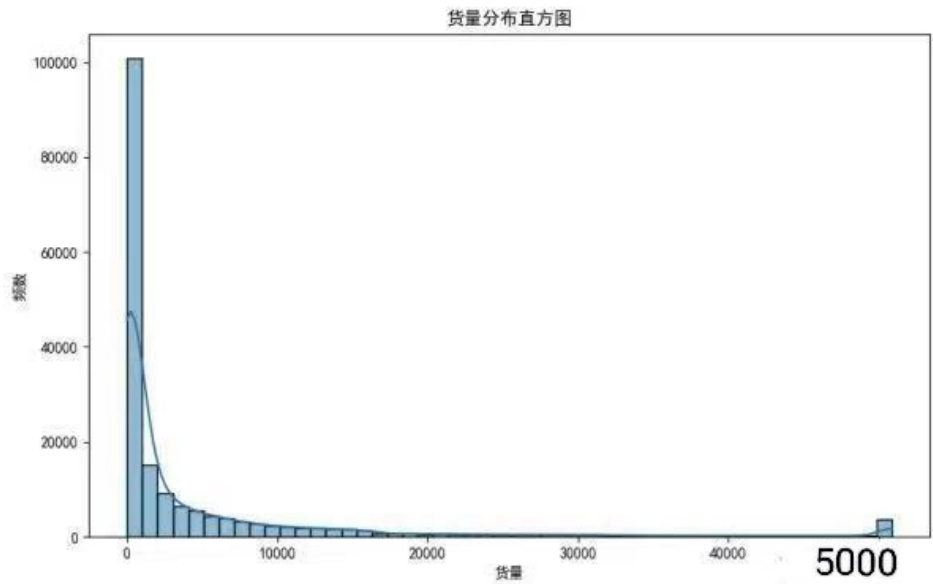


图 4.1.1

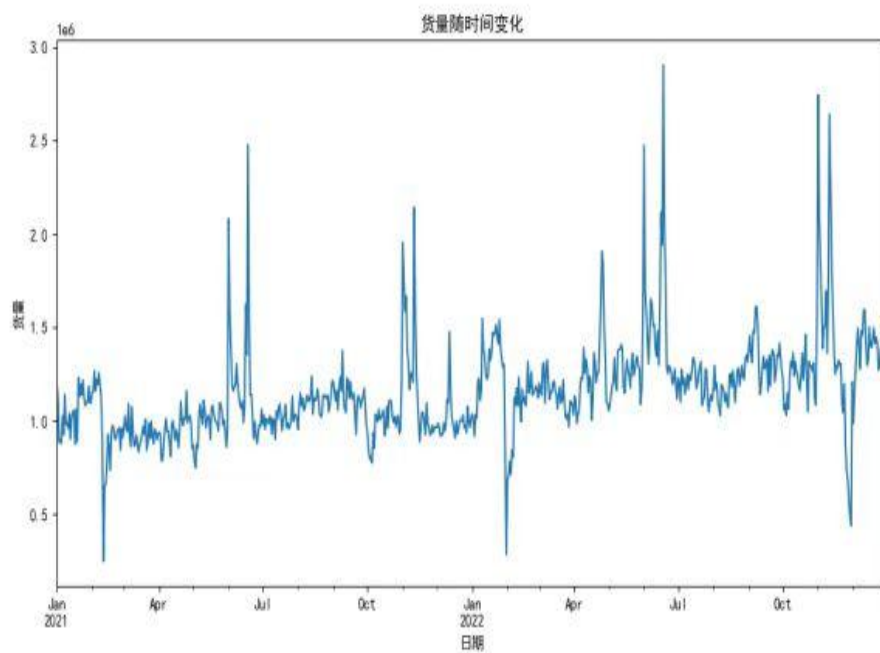


图 4.1.2



图 4.1.3

通过自相关图（ACF）和偏自相关图（PACF）（见图 4.1.4）来显示货量在不同时间点上的相关性程度，帮助理解时间序列数据中的季节性和趋势性变化以及数据结构，并且选择 AR 和 MA 模型的参数。

我们创建了一个 2 维的平面图，显示沿 x 轴的延迟值以及 y 轴上的相关性(-1 到 1 之间)。ACF 图中，纵轴表示自相关系数，即 y 轴上的数值表示货量数据间的相关性，横轴表示时间

间隔，即 x 轴上的数值表示不同时间点之间的跨度。当 x 轴跨度为 0 时，自相关系数等于 1，表示销量数据与其本身完全自相关。随着 x 轴跨度逐渐增大时，自相关系数逐渐趋于 0，表示数据之间的相关性逐渐减弱。我们将 x 轴上的延迟值限制为 30，让图表更易理解。

与自相关图类似，PACF 图也是一种描述时间序列数据在不同时间点上的相关程度的图形，但它显示的是在一定时间间隔内，两个序列的相关性除去其他自变量之后的残差相关。PACF 图中，纵轴表示偏自相关系数，即 y 轴上的数值表示数据间的相关性，横轴表示时间间隔，即 x 轴上的数值表示不同时间点之间的跨度。当 x 轴跨度为 0 时，偏自相关系数等于 1，表示数据与其本身完全自相关。随 x 轴跨度逐渐增大时，偏自相关系数趋于 0，表示所考察的时间间隔越来越远，两个序列之间的相关性越来越弱。

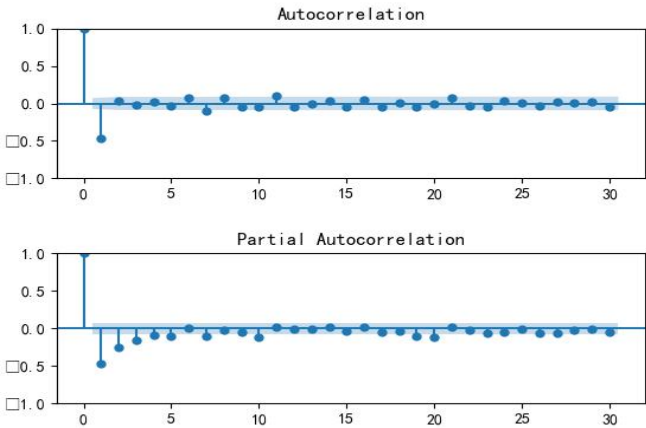


图 4.1.4-1(DC14→DC10)

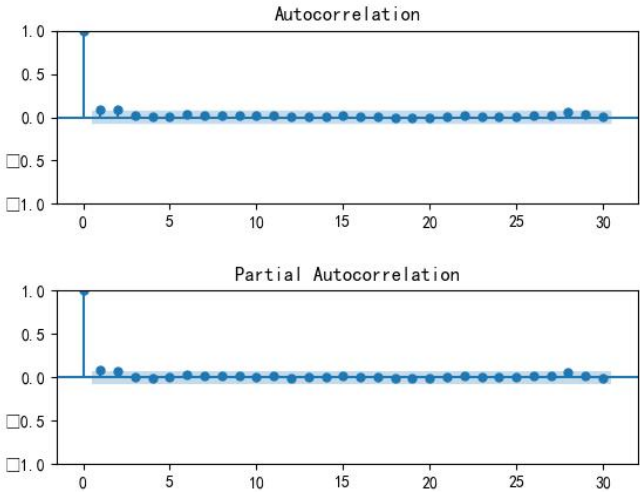


图 4.1.4-2(DC20→DC35)

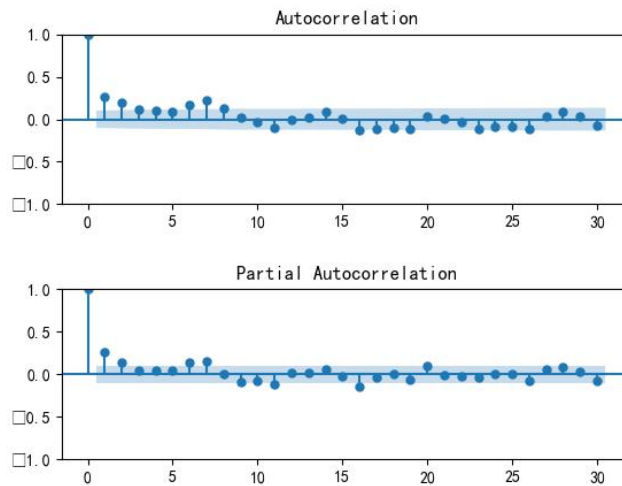


图 4.1.4-3(DC25→DC62)

4.2 历史货量数据分析

1、观察数据的趋势和周期性

● **ACF 图：**对于线路 DC14→DC10 及线路 DC20→DC35，ACF 值迅速衰减而随后在小范围内波动，表明该线路货量数据呈现出随机性。PACF 值没有在某些滞后阶数上出现显著的正负值，则表明这些滞后阶数上的货量数据不存在明显的趋势组成成分。

对于线路 DC25→DC62，由 AFC 图呈现出类正弦波型，说明货量数据与时间季节之间存在周期性的相关关系。这种周期性的相关性也称为“季节性自相关性”，说明数据存在季节性变化，此处季节波动的周期是每个季度单位内的货量数值。

● **PACF 图：**对于第一、第二条线路并没有出现显著的正负值，故不存在明显的趋势组成成分，但第三条线路 DC26→DC62 显示出的类正弦波形表示数据在图中所取的 0~30 阶数上存在拖尾或震荡的相关性，表明时间序列数据的趋势具有震荡或周期性的变化。

2、确定数据的平稳性(一阶差分)

由图可知，对于线路 DC20→DC35 以及线路 DC25→DC62，ACF 和 PACF 都在滞后阶数很小时便迅速衰减进而在零附近波动，并无出现在某些滞后阶数上呈现显著正值或负值的情况，故我们确保了这两条线路货量数据的平稳性。但对于线路 DC14→DC10，其 ACF 图和 PACF 图都在滞后阶数增长之后才衰减并且随后保持较大波动性，此处我们采用一阶差分方法将原始时序数据转化为平稳的时间序列数据，具体做法是对原始数据进行一次差分，将时间序列的差异化信息提取出来，剔除了趋势性和季节性成分，绘制对线路 DC14→DC10 的一阶差分图（图 4.2）从而使得时间序列能够更好的适应模型的假设前提，达到平稳序列的效果，提高模型的精确性和鲁棒性。

具体地,对每天货量数据点与其前一天点之差求出一个新的数据点,形成新的时间序列。数学上, **ACF** 图和 **PACF** 图的计算方式与 **ARIMA** 模型中的一样,只不过需要将滞后时间改为季节性滞后时间。具体来说,季节性 **ACF** 图的计算方式为:

$$\rho_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}$$

其中, Y_t 为时间序列的第 t 个观测值, k 为滞后阶数, n 为时间序列的总长度, $\bar{Y}$ 为时间序列的均值。

季节性 **PACF** 图的计算方式为:

$$\Phi_{kk} = \frac{\text{cov}(Y_t, Y_{t-k})}{\text{var}(Y_t)}$$

其中, $\text{cov}(Y_t, Y_{t-k})$ 表示时间序列 Y_t 和 $Y(t-k)$ 的协方差, $\text{var}(Y_t)$ 表示时间序列 Y_t 的方差, k 为滞后阶数。

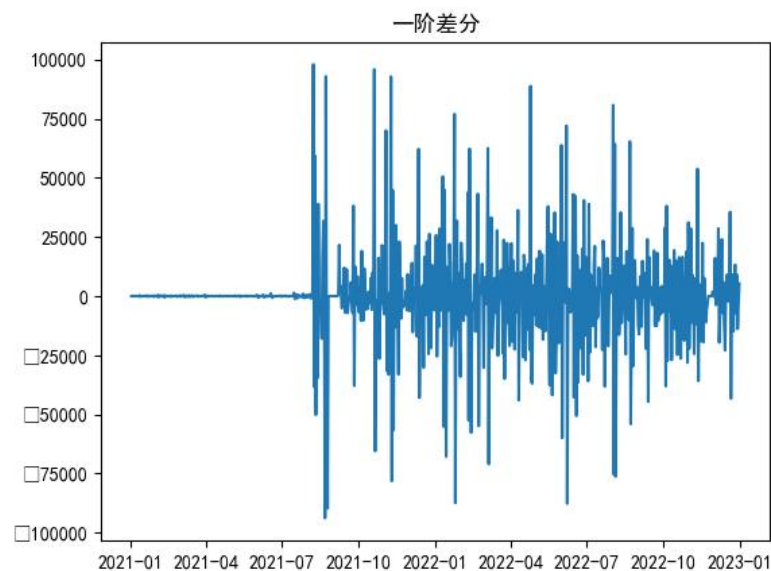


图 4.2

经过一阶差分处理后,再次绘制时间序列图,检测结果如下:

Test Statistic	-9.149761e+00
测试统计量	
p-value	2.706427e-15
基于“MacKinnon”的近似 p 值	
#Lags Used	1.900000e+01
使用的滞后数量	
Number of Observations Used	7.050000e+02
用于 ADF 回归的观察数	
Critical Value (1%)	-3.439660e+00
Critical Value (5%)	-2.865648e+00
Critical Value (10%)	-2.568958e+00
测试统计数据临界值为 1%，5%，10%	
D-type: float64	
Stationarity 是否平稳(1/0): 1	

综上所述，Test Statistic（测试统计值）小于所有临界值（Critical Value），且 P-value 是非常接近 0 的，小于 0.01，表明检验结果具有平稳性，故销量数据已处理为平稳时间序列。

4.3 确定模型参数

4.3.1 确定模型阶数

根据销量数据的趋势和周期性，选择适当的 SARIMAX 模型，包括季节性因素(s)、自回归(AR)、差分(I)、移动平均(MA)、外生变量(X)五部分。平稳序列在建模时涉及到包括确定 p、d、q 和 P、D、Q 等参数。我们参考 ACF 和 PACF 图，找到 ACF 和 PACF 函数早期的截尾点，从而推测出 p 和 q 的范围，并使用信息准则（如 AIC、BIC 等）来确认它们的最优值。

p: 与模型的自回归方面相关的参数，模型中包含的滞后观测数，即滞后阶数。

d: 与模型集成部分相关的参数，原始观测值差异的次数，即差异度。它影响到应用于时间序列的差分的数量。

q: 与模型的移动平均部分相关的参数。移动平均窗口的大小，即移动平均的阶数。我们选择通过 ACF 和 PACF 图来确定模型参数 p 和 q。

观察 ACF 图，找出自相关系数随着滞后期的增加而逐渐减小，超过置信区间后逐渐变为白噪声的点。根据 ARIMA 模型中 AR 部分的定义，需要找到最后一个超出置信区间的滞后期，记为 p 的值，此处观察得 p 值为 9。

观察 PACF 图，找出偏自相关系数随着滞后期的增加而逐渐减小，超过置信区间后逐渐变为白噪声的点。根据 ARIMA 模型中 MA 部分的定义，需要找到最后一个超出置信区间的滞后期，记为 q 的值，此处观察得 q 值为 8。

4.3.2 估计模型参数

根据确定的 p 和 q , 可以用最大似然方法来估计 ARMA 模型的参数。对于 SARIMA 模型, 我们考虑到季节因素的影响, 此时需要确定季节差分所需要的参数 P 、 D 和 Q , 并对模型进行对应的参数估计。

SARIMA 模型阶数的选择:在 ARIMA 模型中, 有时需要对数据的季节性进行建模, 因此需要增加一个季节性阶数, 记为 P 、 D 、 Q 。其中, P 和 Q 表示季节性 AR 和 MA 的阶数, D 表示季节性差分的阶数。可以通过类似 ACF 和 PACF 图的方法, 来确定季节性部分的阶数。SARIMAX 模型的阶数即为 p 、 d 、 g 和 P 、 D 、 Q 的组合。

接着, 我们运用网格搜索法 (Grid Search)帮助我们寻找 SARIMAX 模型中 seasonal_order 参数的最佳取值。

我们首先需要确定 seasonal_order 参数的取值范围, 例如季节性 AR 阶数 P 的范围可以设置为 0 至 3, 此处我们令其为 1, 季节性差分阶数 D 的范围可以设置为 0 或 1, 我们令其为 2, 季节性 MA 阶数 Q 的范围可以设置为 0 至 3, 我们令其为 3。

随后, 我们建立模型:根据上述步骤中确定的 seasonal_order 参数范围, 建立 SARIMAX 模型

接着确定模型评估指标: 我们选择 BIC (贝叶斯 3 信息准则)模型评估指标, 用于评估模型拟合效果。

再进行网格搜索:对于每一种 seasonal_order 的参数组合, 利用历史数据训练 SARIMAX 模型 4 并计算评估指标。选择评估指标最小的一组参数作为最佳取值。

最后验证模型:使用最佳的 seasonal_order 参数组合, 对一段时间内的数据进行预测, 并评估 5 预测效果。如果预测效果较好, 则可将该参数组合应用于实际预测。

4.4 模型训练和诊断评估:

我们使用历史数据训练模型, 并使用交叉验证等技术评估模型的准确性和稳定性。选择均方根误差 (RMSE)作为评估指标, 衡量模拟预测值与实际值的平均误差。

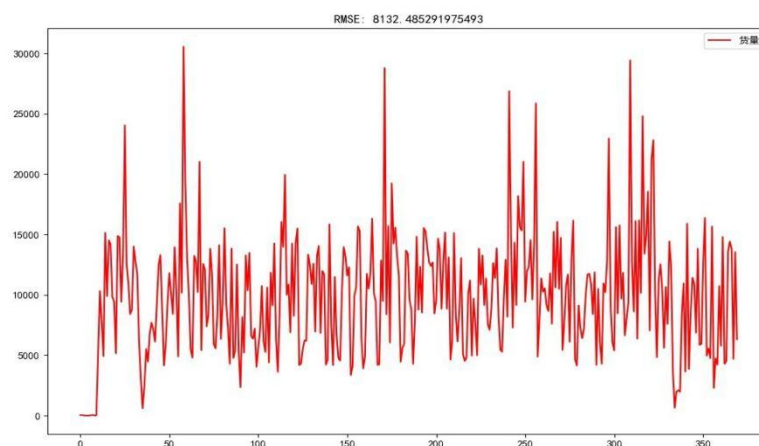


图 4.4.1(线路 DC14→DC10)

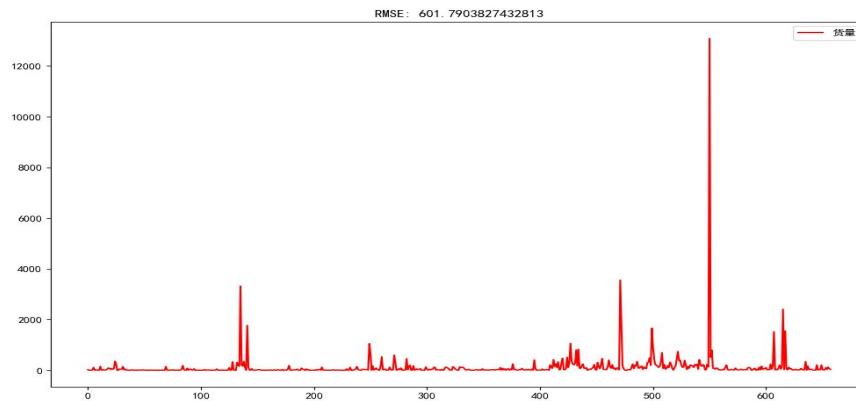


图 4.4.2(线路 DC20→DC35)

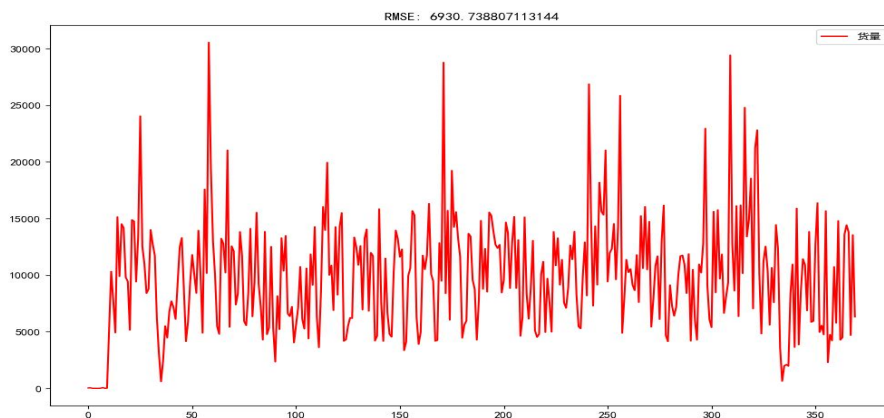


图 4.4.3(线路 DC25→DC62)

检验残差的分布及自相关性。在拟合好的 SARIMAX 模型中，需要对残差进行分析，检查残差是否符合扰动项的假设，并且需要检查残差中是否存在自相关性。可以通过直方图和自相关图等方式进行展示。

该模型中的标准化残差是指将 SARIMAX 模型拟合过程中得到的残差序列进行标准化处理，以消除不同高度的波动对建模结果的影响，更准确地评估模型的质量。标准化残差的计算方法如下：

首先，我们需要先获取模型的残差序列 `residuals`，然后计算它们的标准差，即 `np.std(residuals)`，最后将 `residuals` 中每一个元素减去平均值 `np.mean(residuals)`，再除以标准差 `np.std(residuals)`，即可得到标准化残差序列。

标准化残差通常应该服从均值为 0、标准差为 1 的标准正态分布，我们采用绘制标准化残差图、判断统计量是否服从正态分布等方式来检查模型的拟合质量。如图模型残差的分布特征所示，残差呈现单峰且近似于对称的正态分布，说明该模型对货量数据预测拟合效果较好。

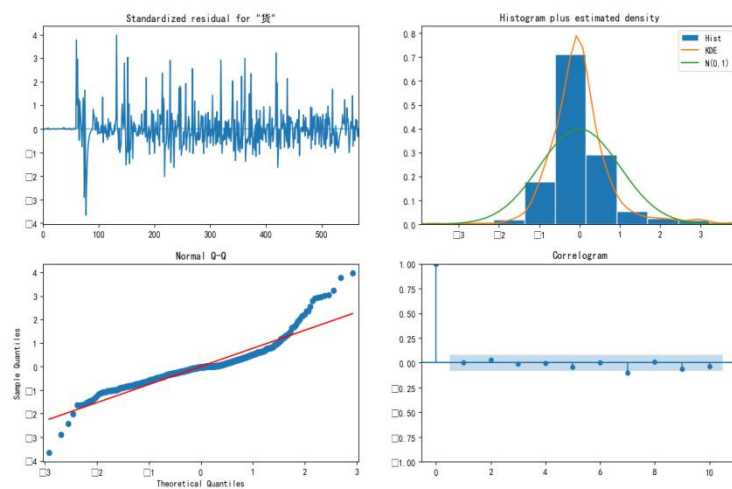


图 4.4.4(线路 DC14→DC10 残差检验)

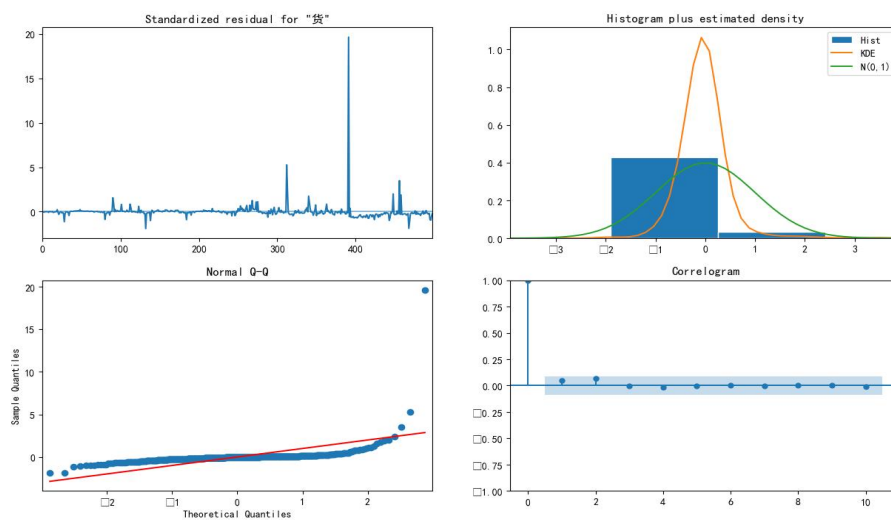


图 4.4.5(线路 DC20→DC35 残差检验)

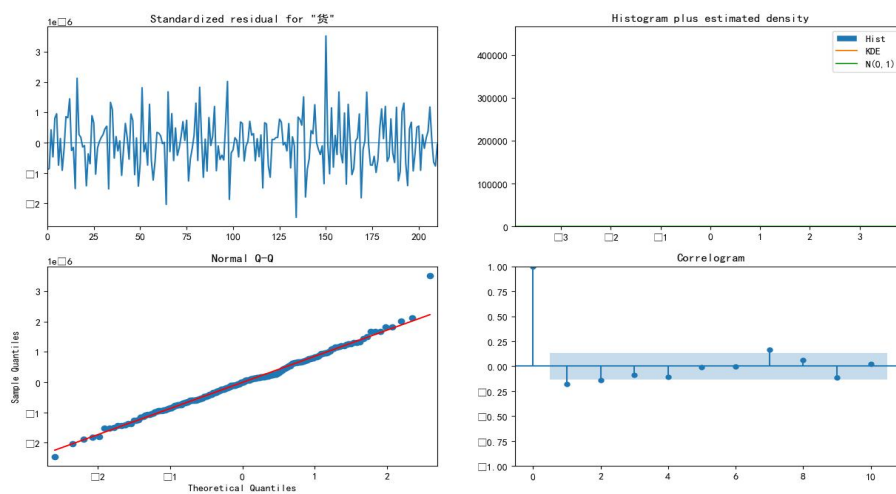


图 4.4.6(线路 DC25→DC62 残差检验)

另外，直方图加密度估算图（“histograms plus estimated density”）显示数据的分布特征。该图形包含两部分内容：直方图和密度估计曲线。

直方图用于表示数据的分布情况，它将数据分成若干个等距区间，并统计每个区间中数据的频数或频率。经过归一化处理，直方图可以表示数据的相对密度，从而反映数据的分布形状，如正态分布、偏态分布或者多峰分布等。

密度估计曲线表示的是数据的概率密度函数，即连续变量在给定值处的密度值。通过对数据进行核密度估计，可以得到一个平滑的曲线，更准确地展现了数据的概率密度函数。

4.5 模型预测

SARIMAX 模型可以概括为以下步骤和公式：

1.差分(Differencing):为了消除时序数据中的季节性和趋势性，在进行模型训练之前需要对数据进行差分处理。差分公式为：

$$\Delta y_t = y_t - y_{t-1}$$

其中， y_t 为原始数据中的某时刻的观测值， y_{t-1} 为差分后的值。

2.自回归模型 (AR): 自回归模型是指用前一时刻的观测值来预测当前时刻的观测值，通常用 p 来表示用几个前一时刻的观测值来预测当前时刻的观测值。AR(p)模型公式为：

$$y_t = \sum_{i=1}^p \Phi_i y_{t-i} + \epsilon_t$$

其中， Φ_i 为 AR 系数， ϵ_t 为误差项。

3. 移动平均模型(MA):移动平均模型是指用前一时刻的误差来预测当前时刻的观测值通常用 q 来表示用几个前一时刻的误差来预测当前时刻的观测值。MA(q)模型公式为：

$$y_t = \sum_{i=1}^q \Phi_i y_{t-i} + \epsilon_t$$

其中， Φ_i 为 AR 系数， ϵ_t 为误差项。

4. 季节性模型(Seasonal): 考虑到时序数据中存在的季节性变化，需要引入季节性模型.季节性模型通常用 P 和 Q 表示，分别表示前 P 个季节性差分和前 Q 个季节性误差的线性组合。

SARIMA(p,d,q)(P,D,Q)s 模型公式为：

$$y_t = \Phi \cdot y_{t-1} + \sum_{i=1}^p \Phi_i y_{t-i} + \theta \cdot \epsilon_{t-1} + \sum_{i=1}^q \theta_i \cdot \epsilon_{t-i} + \epsilon_t + \sum_{i=1}^P \Phi_i \cdot \Delta y_{t-i} + \sum_{i=1}^Q \theta_i \cdot \Delta \epsilon_{t-i}$$

其中， ϕ 和 Δ 分别为 AR 和 MA 的一阶系数， ϕ_i 和 Δ_i 分别为季节性 AR 和季节性 MA 的一阶系数， Δy_{t-i} 和 $\Delta \epsilon_{t-i}$ 分别为前 i 个季节性差分和前 i 个季节性误差的线性组合。

经过诊断确认建好的模型可以用于时间序列的预测，将预处理后的数据输入模型进行训练。我们使用了早停策略以防止过拟合并加速训练过程，使用验证集评估模型性能。在每个训练周期后，根据评估指标检查模型在验证集上的性能。如果性能不再提高，则停止训练。我们使用 SARIMAX 模型对未来 2023 年 1 个月内每条线路的每日货运量数据进行单步的预测。特别地，对题目中要求的 DC14→DC10、DC20→DC35、DC25→DC62 3 条线路预测结果进行展示（见图 4.5）。

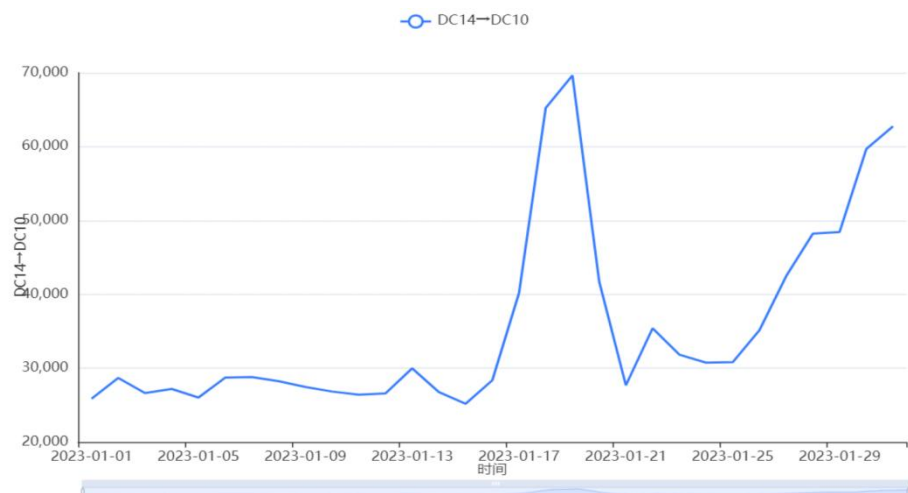


图 4.5.1(DC14→DC10 预测结果)

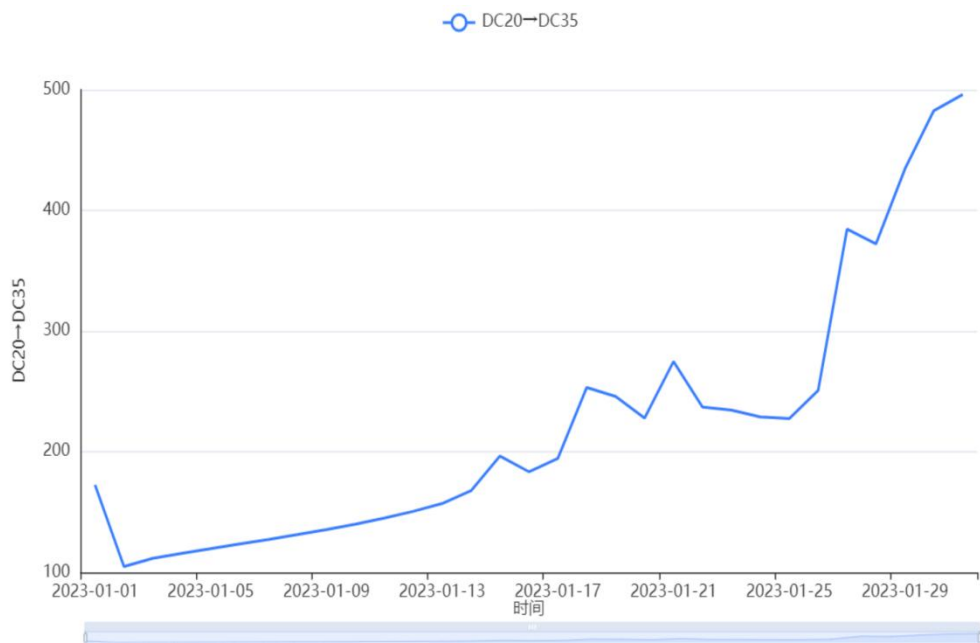


图 4.5.2(DC20→DC35 预测结果)

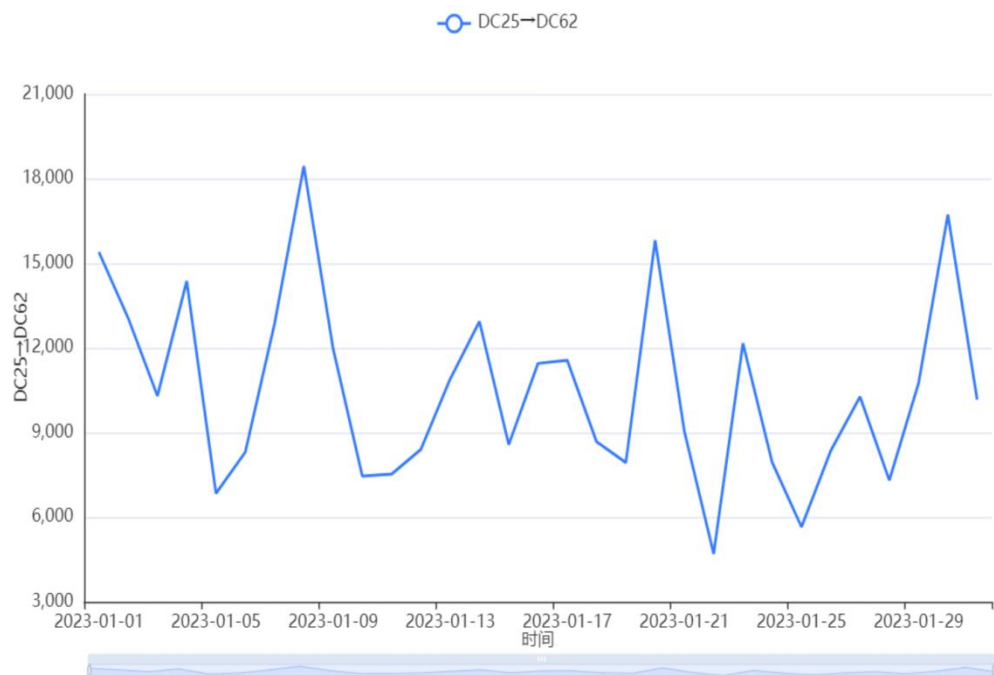


图 4.5.3(DC25→DC62 预测结果)

我们给出三条路线的具体预测结果数据，

如下表所示：

时间	DC14→DC10	DC20→DC35	DC25→DC62
2023-1-1	25859.63	172.49	15399.86
2023-1-2	28683.34	104.84	13060.06
2023-1-3	26624.80	111.64	10298.50
2023-1-4	27186.35	115.75	14368.13
2023-1-5	26018.86	119.71	6841.27
2023-1-6	28715.31	123.51	8304.37
2023-1-7	28801.59	127.36	12856.09
2023-1-8	28229.25	131.32	18449.43
2023-1-9	27452.49	135.49	11976.04
2023-1-10	26834.76	139.98	7453.94
2023-1-11	26406.29	144.96	7527.39
2023-1-12	26574.16	150.60	8403.75
2023-1-13	29978.08	157.10	10896.64
2023-1-14	26757.39	167.66	12937.43
2023-1-15	25177.18	196.41	8570.76
2023-1-16	28357.35	183.41	11451.84
2023-1-17	40174.83	194.51	11557.97
2023-1-18	65243.93	253.25	8672.20
2023-1-19	69647.47	245.78	7934.51
2023-1-20	41739.79	227.82	15813.83
2023-1-21	27673.21	274.75	9036.09
2023-1-22	35402.72	237.11	4698.90
2023-1-23	31830.46	234.48	12164.78
2023-1-24	30748.93	228.89	7937.34
2023-1-25	30809.01	227.48	5652.11
2023-1-26	35135.69	250.70	8355.79
2023-1-27	42462.04	384.47	10268.74
2023-1-28	48224.74	372.25	7307.94
2023-1-29	48445.07	434.25	10751.81
2023-1-30	59678.18	482.61	16727.27
2023-1-31	62744.69859	496.03	10163.56

表 4.5(预测结果详细数据)

五、问题二模型的建立与求解

5.1 模型变量定义：

首先，我们为多目标线性规划模型(MLP)定义一些参数、决策变量和约束条件来表示货量分配方案和网络负荷情况：

5.1.1 参数：

- X_i : 线路 i 分配到的货量 (单位: 件)
- a_i : 线路 i 的总流转货量 (单位: 件)
- k : DC5 线路货量能正常流转的比率 (单位: %)
- α : 各线路货量标准差, 代表工作负荷不均衡度 (单位: 件)
- n : 与 DC5 相关线路数, 即货量发生变化的线路数

5.1.2 决策变量：

- X_n : DC5 分配给 n 条线路的货流量组成的 n 维向量
- 新增货流量的线路数 n

其中, i 表示线路编号

5.2 定义目标函数：

基于①最小化因 DC5 关停而发生货量变化的线路数②最小化各线路工作符合不均衡度③最小化未能正常流转的包裹日累计量的目标, 我们定义 3 个最小化目标函数：

$$\text{minimize } g(n) = n$$

$$\text{minimize } k = \frac{a - \sum_{i=1}^n X_i}{a}$$

$$\text{minimize } \alpha = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

其中, a 值具体为线路 DC5 1 月份的预测总运货量 3781385。

5.3 算法选择

NSGA-II (Non-dominated Sorting Genetic Algorithm II) 是一种经典的多目标优化遗传算法。它通过简化非支配排序过程和增加拥挤度距离的计算来提高优化效率和搜索质量，故 NSGA-II 算法可以广泛应用于各种优化问题。以下展示该遗传算法的原理和流程图：

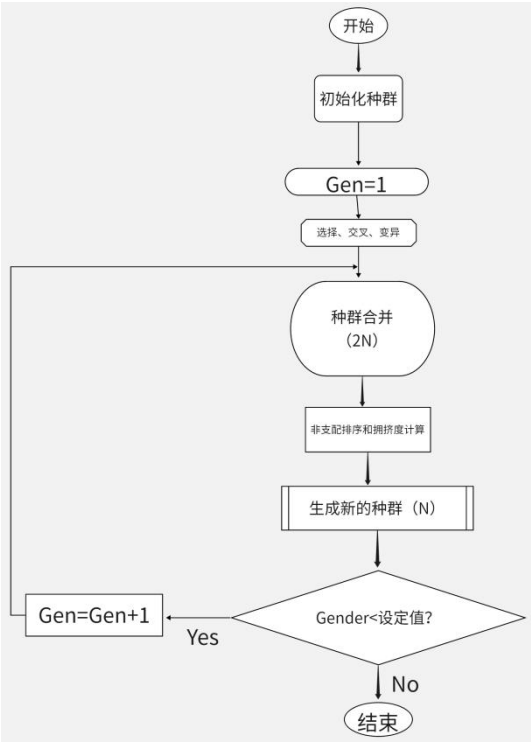


图 5.3 NSGA-II 遗传算法

5.4 定义约束条件：

- 每条线路新增的负载运货量/运输能力约束： $X_i \leq m$

考虑到该模型为全局模型，对于所有线路来说， m 为货量转移负载能力的上线，由历史数据可得场地 DC5 的出入路线中最大运出货量 15505，故 m 值为：15505。

- 每条线路新增的负载接收货量/处理能力约束： $\sum_{i=1}^n X_i \leq j$

由历史数据可得场地 DC5 的出入路线中最大接收货量为 63906，故 j 值为：63906。

5.5 模型求解：

首先，基于问题 1 中的预测模型，获取 2023-01-01 至 2023-01-31 期间每条线路每天的预测货量，接着就是识别受 DC5 关停影响的线路，包括所有以 DC5 为始发地或目的地的线路为每条受影响的线路。在为每条受影响的线路找到候选替代线路时，应避免包含 DC5，并确保始发地和目的地之间的距离尽可能短。为了找到候选替代线路，我们需要考虑避免包含 DC5 的线路，同时确保始发地和目的地之间的距离尽可能短。

求解过程我们使用多目标线性规划模型为每条受影响线路分配货量。考虑到该问题本质为多目标优化问题，我们利用线性规划方法配合启发式算法（NSGA-II 遗传算法）进行迭代得出最优解，即优化的分流方案，使各个线路的负载尽可能平衡，并尽量减少运输量变化的线路数。

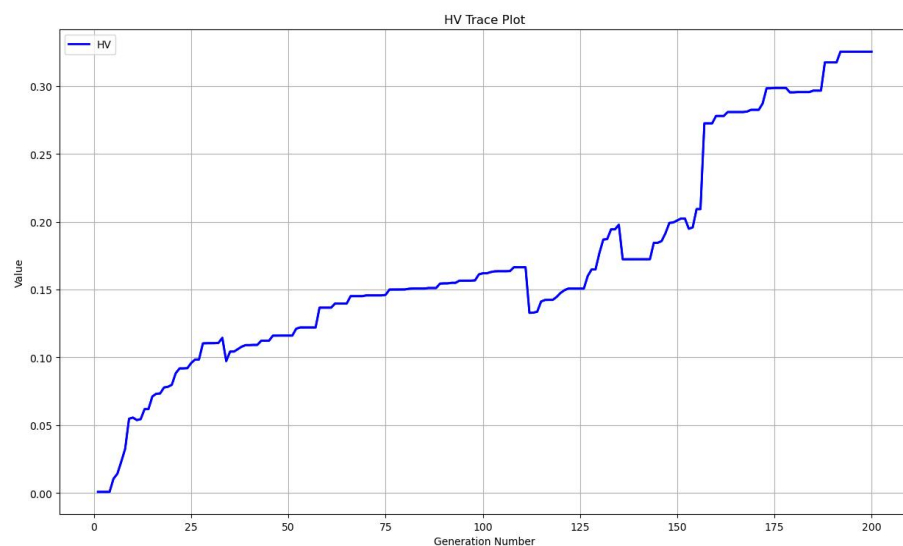


图 5.5.1(种群数量图)

我们根据问题的规模和复杂度以确定种群数量的选择。如果种群数量过小，则可能会导致搜索空间覆盖不全，从而无法找到全局最优解。如果种群数量过大，则算法的运行时间和计算资源需求将会增加，同时也可能导致过度拟合和搜索效率降低。我们通过绘制收敛曲线（图 5.5.1）和评估算法的性能来选择最优的种群数量。

在绘制收敛曲线后，我们根据图像的趋势来选择最合适的种群数量。通常情况下，随着种群数量的增加，算法的性能会提高，直到达到一个最优点后开始下降。如图，我们观察到一个种群数量点 52，在其之前收敛曲线呈现出随着种群数量的增加而不断改善的趋势，在其之后拟合程度值迎来爆发式增长，代表算法出现拟和过度，故我们选择 52 作为最有种群数量，使得算法达到最优点的同时又不会过度拟合或过度计算。

绘制出最优解后，根据要求，我们希望上述三个目标函数值(惩罚项)都是最小的，所以我们选择三个坐标分别都较小的点作为最优解，即点（0.0464,584.5378,13.5646），用橙色箭头表示，具体情况如下图（图 5.5.2）所示：

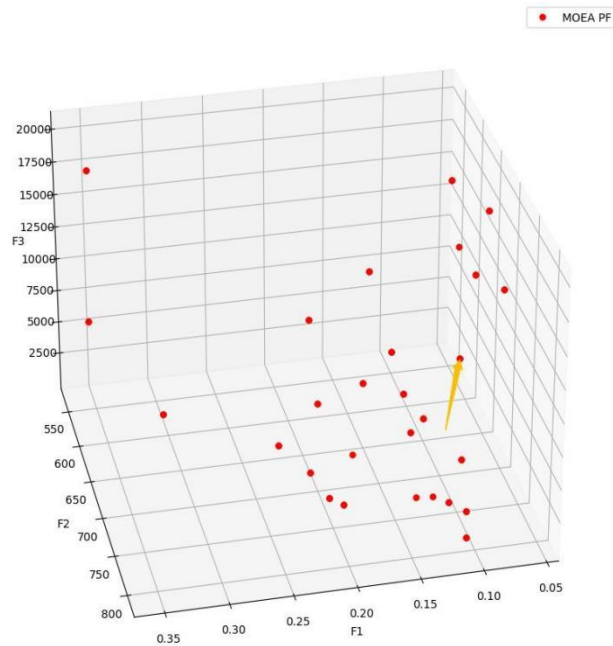


图 5.5.2(最优解选择)

由图可得, x 值(f_1)为 10^{-2} 数量级, k 值数量小说明发生异常未能流转的货量很少, z 值(f_3)小说明各个线路分配后工作负荷情况较均衡, 很好地符合题设要求。

求解上述 MLP 模型后, 我们可以获得 DC5 关停时货量分配方案。通过分析求解结果, 我们可以得到因 DC5 关停导致货量发生变化的线路数及网络负荷情况。

具体地, 我们得出的最优解为 (主要部分举例):

线路 DC5→DC3 的货量全部分配到线路 DC5→DC44
 线路 DC5→DC10 的货量全部分配到线路 DC5→DC80
 线路 DC5→DC4 的货量全部分配到线路 DC5→DC50
 此时目标函数值为: 2.0969e+0.8

网络负荷情况为:

线路 DC5→DC44 货量为 39144
 线路 DC5→DC80 货量为 135739
 线路 DC5→DC50 货量为 41747。

六、问题三模型的建立与求解

6.1 模型变量定义:

该问题在问题 2 的线路货量预测基础上, 考虑重新分配并每日动态调整。我们对模型的改进

具体为增设一个二元决策变量 δ_i 、成本参数 C_i 和目标成本函数 $\sum_{\delta_i=1} C_i$ ，削减了参数 n ，因为改进后的模型实现了每天线路开闭的动态调整，无需考虑统一分配到 n 条线路的问题。

6.1.1 参数：

- X_i : 线路 i 分配到的货量 (单位: 件)
- a_i : 线路 i 的总流转货量 (单位: 件)
- k : DC5 线路货量能正常流转的比率 (单位: %)
- α : 各线路货量标准差, 代表工作符合不均衡度 (单位: 件)
- C_i 为场馆成本, 将开放场地的成本设为 0, 关闭场地的成本设为一个极大的正数 (例如 1000000), 该做法目的为凸显场地开放成本的代价。

6.1.2 决策变量：

- δ_i 为线路 i 货量变化指示变量, 当线路 i 的货量发生变化 (场地新增) 时, δ_i 取值为 1, 否则 (场地关闭) 取值为 0。

其中, i 表示线路编号。

6.2 定义目标函数：

基于①最小化因 DC5 关停而发生货量变化的成本值②最小化各线路工作符合不均衡度③最小化未能正常流转的包裹日累计量的目标, 我们在此前基础上重新定义 3 个最小化目标函数：

$$\text{minimize } \sum_{\delta_i=1} C_i$$

$$\text{minimize } k = \frac{a - \sum_{i=1}^n X_i}{a}$$

$$\text{minimize } \alpha = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

其中, a 值具体为线路 DC9 1 月份的预测总运货量 471875。

- 目标函数将决策变量值为 1 的成本参数 C_i 求和，要求是使其和尽可能小。

6.3 算法选择与约束条件：

在问题 3 情境下，考虑上述的模型改进步骤后，问题 2 的启发式算法与运输能力约束、处理能力约束皆适用，故我们沿用此前的算法解决该问题。

6.4 模型求解：

首先，我们找到与 DC9 相关的线路（始发地或者目的地为 DC9 的线路）。接着遍历 2023 年 1 月份的日期范围，对期间的每一天，进行如下的具体步骤执行：

- ①：寻找该日期的受影响路线及其预测货量数据。
- ②：识别候选替代线路，我们避免了该线路包含 DC9，并确保始发地和目的地之间的距离尽可能短。同时，允许新开线路，但运输能力限制为已有线路历史运输能力的最大值。
- ③：在满足线路运输能力和场地处理能力限制的条件下，使用多目标线性规划方法为每条受影响的线路分配货量。目标为最小化流转异常的包裹日累积量、货运量发生改变的线路数量、各线路工作负载的不均衡度。
- ④：根据多目标线性规划问题的解决方案，确定当前日期的货量分配方案。同时，记录因 DC9 关停导致货量发生变化的线路数、不能正常流转的货量及网络负荷情况。此外，记录每天的线路增减情况。

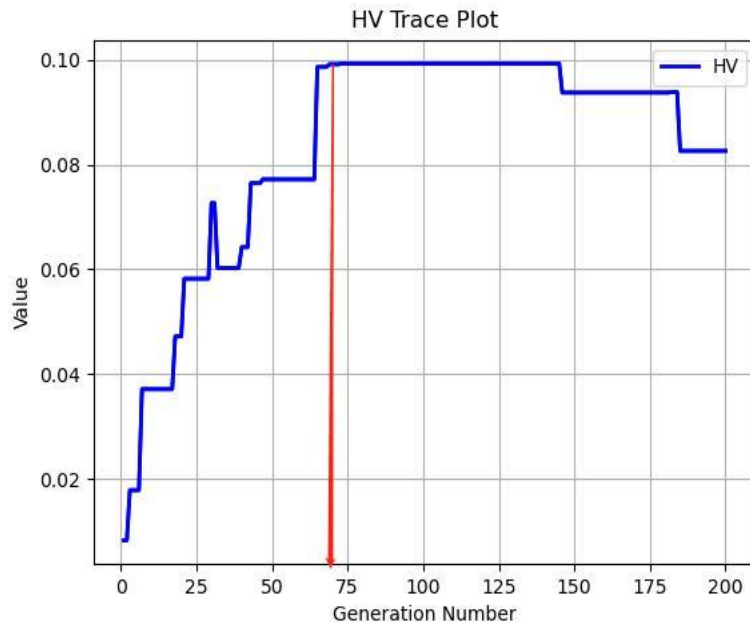


图 6.4.1(种群数量图)

与问题 2 类似，我们通过绘制收敛曲线（图 6.4.1）和评估算法的性能来选择最优的种群数量。由观察得，在 60 处取得最优种群数量，遗传算法的拟合效果最佳。

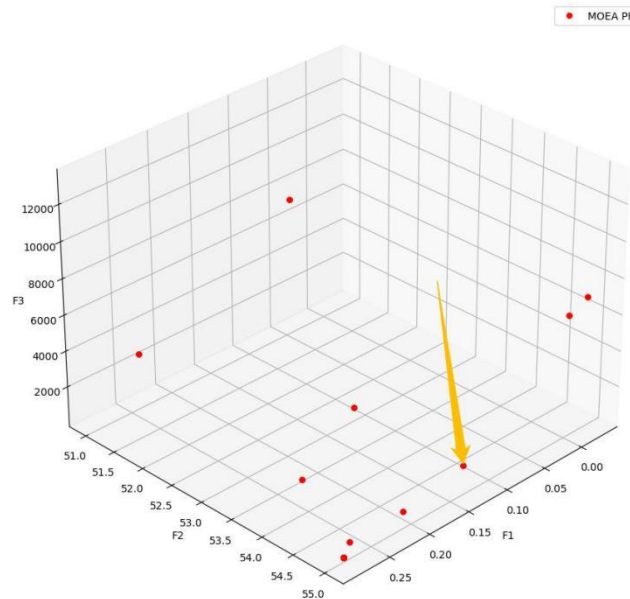


图 6.4.2（最优解选择）

同理，我们在得到模型最优解后选取出 3 个惩罚项值皆较小的值，即点（0.1295，54.9554，34.2344），用橙色箭头表示，具体情况如上图（图 6.4.2）所示：

由图可得，x 值数量小说明发生异常未能流转的货量很少，y 值为优化后的线路数 55，说明我们删减了 26 条线路，z 值(f3)小说明各个线路分配后工作负荷情况较均衡，较好地符合题设要求。

求解上述 MLP 模型后，我们可以获得 DC5 关停时货量分配方案。通过分析求解结果，我们可以得到因 DC5 关停导致货量发生变化的线路数及网络负荷情况。

具体地，我们得出的最优解为：

线路 DC9→DC3 的货量全部分配到线路 DC5→DC45

线路 DC9→DC14 的货量全部分配到线路 DC5→DC82

线路 DC9→DC5 的货量全部分配到线路 DC5→DC53

此时目标函数值为：2.0969e+0.8

网络负荷情况为：

线路 DC5→DC45 货量为 1888830

线路 DC5→DC82 货量为 2245422

线路 DC5→DC53 货量为 1684056。

七、问题四模型与分析

7.1 模型数据预处理以及可视化展示

我们使用 K-means 聚类法查找最接近某个聚类中心的数据点来查找历史数据中运输最频繁

的线路和场地，并将其分配到该聚类中心所表示的簇中，从而实现聚类的目的。具体包括以下步骤：

- 1、随机设置 k 个场地作为聚类中心；
- 2、计算每个场地点到这 k 个聚类中心的距离，并将场地点分配给与其最相似的聚类中心所代表的簇；
- 3、根据每个簇中所有场地点的均值来更新 k 个聚类中心；
- 4、重复步骤 2 和 3，直到聚类中心的位置不再改变或达到预设的迭代次数。

接着，我们使用 `python` 对这些运输最为频繁的场地和线路进行可视化实现，绘制出随机分布网络图，如下图所示（见图 7.1）

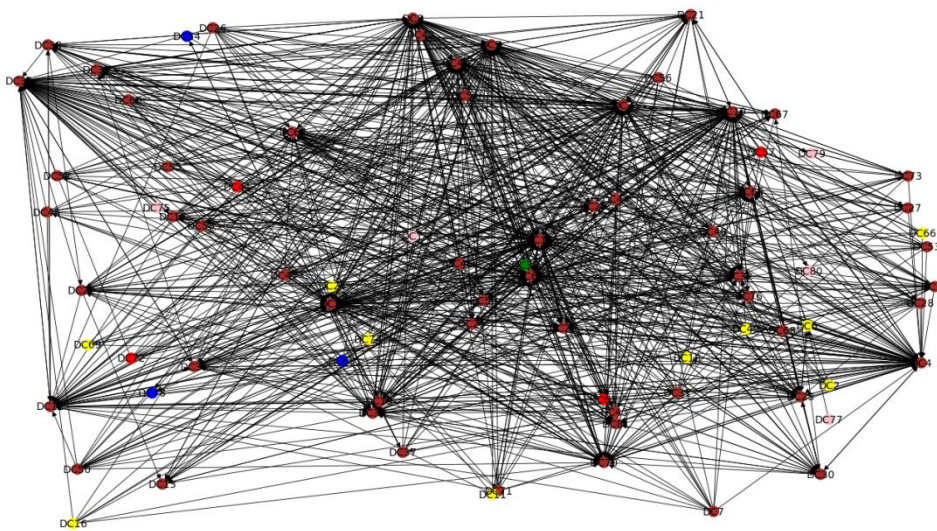


图 7.1(随机分布网络图)

7.2 模型选择与思路

在问题四中，我们首先需要对不同场地和线路的重要性进行评估，在此，我们选择了 TOPSIS 熵权法，我们还构建了一个 AHP 层次分析法模型对 TOPSIS 熵权法中各指标的权重进行评估，以保证我们模型的准确性。其次，我们需要考虑是否增加场地及线路以及它们的指标选择的问题。在此，我们绘制了随机分布网络图，在对网络图节点进行聚类筛选后，我们得出对整个系统影响比较大的三个场地，据此，我们对新增的线路与场地进行了分析。

最后，考虑到检验建立的模型在面对数据异常、模型假设不满足等情况时的表现，我们进行鲁棒性分析。

首先，我们通过异常值检验其鲁棒性。由于 SARIMAX 模型对异常值比较敏感，因此需要通过异常值检测等方法识别和调整异常值，避免它们对模型的影响，此处我们没有检测出显著的异常值。

我们再次对预测数据进行平稳性检验，如果数据不平稳，需要对其进行差分和满足白噪声检验等步骤，使其变得平稳。

满足上述条件后，我们使用历史数据对模型进行拟合，得出预测数据与历史数据不存在很大差异，确保模型能够较好地拟合数据。

7.3 利用 AHP 法对指标权重进行评估

我们使用 AHP 方法进行指标权重评估时，首先确定目标和评价指标：明确决策问题的具体目标，并确定所有相关的评价指标。

构建层次结构模型：将目标与指标之间的层次结构模型化为一个树形结构，其中目标层次为最高层，指标层次为下一层，以此类推。在 AHP 中，通常只会有 3~5 个层次。

两两比较指标权重：使用 1~9 的比较刻度（1 表示两个因素同等重要，9 表示其中一个因素比另一个因素非常重要），对每一层的指标进行两两比较，得到一个权重矩阵。

计算层次总体一致性：检验每一层次的一致性，即评估处理每一层次的评估者意见是否一致。一般来说，判断标准是判断矩阵的一致性指标 CI 是否小于或接近于 0.1。

计算指标权重：在一致性指数 CI 小于或接近于 0.1 的条件下，计算权重向量，即每个指标的重要性权重值。通常使用各种软件工具进行计算。

检查权重的合理性：得出权重向量后，需要对其进行检查，确保权重在合理范围内
我们在准则层的基础上建立一个 2x2 的比较矩阵，如下所示：

	C1	C2	C3
C1	1	2	3
C2	$\frac{1}{2}$	1	$\frac{3}{2}$
C3	$\frac{1}{3}$	$\frac{2}{3}$	1

附注：C1:每条线路包裹运输总量

C2:在历史数据时间内每条线路出现次数

C3:每条线路最大包裹运输量

用算术平均值法求权向量的公式：

$$\varpi_i = \frac{1}{n} \sum_{j=1}^n \frac{a_{ij}}{\sum_{k=1}^n a_{kj}}, (i=1,2,\dots,n)$$

注：n 表示矩阵阶数， a_{ij} 表示矩阵中第 i 行第 j 列的元素。

经计算，得出算术平均法权重为：

[0.54545455 0.27272727 0.18181818]

7.4 熵权 TOPSIS 对场地重要性进行评估

熵权 TOPSIS 是基于 TOPSIS（技术优选排序法）和熵权法的决策分析方法，常常用于多指标评价问题中。场地重要性评估是其中的一种应用，用来评估不同场地的重要性，为后续的规划和决策提供参考。

具体来说，熵权 TOPSIS 评估场地重要性可以分为以下步骤：

- 1、确定评价指标：根据具体问题，确定评价场地重要性的指标体系。选择合适的指标，既要具有反映场地特点的代表性，也要具有实际可操作性。
 - 2、数据标准化：将每一个指标的值标准化，将指标值转换为无量纲的得分，消除了数据值取值范围和单位的影响。
 - 3、计算权重：使用熵权法，根据指标数据的信息量计算出各个指标的权重。
 - 4、构建决策矩阵：将所有场地的评价指标值，带入决策矩阵中，形成一个场地数量行，指标个数列的矩阵。
 - 5、计算距离：通过计算场地矩阵到正理想解的距离和负理想解的距离，得出每个场地的相对距离。
 - 6、计算排序得分：按照距离远近给场地进行排名，距离越近，排序得分越高，反之亦然。
 - 7、得出评估结果：根据排序得分，得出不同场地的重要性评估结果。
- 我们使用熵权 TOPSIS 评估场地重要性，了解不同场地的特点和优劣，制定适当的规划和决策。同时，熵权 TOPSIS 还能够考虑到不同指标之间的相互作用和权重，提高评价结果的准确性和可信度。

线路	线路的得分/%	线路	线路的得分/%
DC1	0.000229326	DC11	1.790514933
DC2	0.482050043	DC12	5.797055124
DC3	9.519933209	DC13	3.291931784
DC4	39.05354722	DC14	76.21619047
DC5	17.16675415	DC15	47.38469158
DC6	0.132975344	DC16	1.744330551
DC7	33.85211982	DC17	54.70326924
DC8	48.47001922	DC18	4.931549884
DC9	18.55730394	DC19	55.01056759
DC10	66.10841893	DC20	49.18730299
DC21	42.10174196	DC31	41.73423974
DC22	54.52181988	DC32	1.205228074
DC23	54.36560201	DC33	2.761218451
DC24	1.650514401	DC34	39.42270159
DC25	50.34487313	DC35	12.56119751
DC26	39.78402529	DC36	56.028289
DC27	15.87189484	DC37	15.54701999
DC28	47.62926184	DC38	48.8278847
DC29	0.011231921	DC39	0.44812592
DC30	46.07494593	DC40	39.28824657
DC41	9.379737855	DC51	52.91637407
DC42	44.51243973	DC52	27.97522739
DC43	42.53126473	DC53	5.742256841
DC44	51.28836031	DC54	0.012707118
DC45	14.14501211	DC55	30.69501273
DC46	47.64068185	DC56	17.49101619

DC47	45.26050469	DC57	10.08880124
DC48	42.93221179	DC58	21.55823109
DC49	4.790689679	DC59	7.095422779
DC50	41.09265513	DC60	6.876883551
DC61	0.171708839	DC71	34.74642056
DC62	27.94651987	DC72	6.025661623
DC63	0.537575407	DC73	38.59731357
DC64	0.990478535	DC74	5.686250235
DC65	0.815756086	DC75	0.000114663
DC66	9.613299161	DC76	0.213050468
DC67	1.859968905	DC77	0.000114663
DC68	13.38840218	DC78	0.15701523
DC69	3.196433918	DC79	0.000114663
DC70	11.05521594	DC80	0
		DC81	100

表 7.4 场地重要性评估

由上表可得出，DC81、DC14、DC10 三条线路在“每条线路包裹运输总量、每条线路出现次数、每条线路最大包裹运输量”这三项指标中贡献最大，重要性评估得分最高。

7.5 新增物流场地以及对其运输能力的分析

根据问题一的预测结果，我们可以发现 DC10、DC14、DC4 场地对整个物流网络的影响比较大，因此，如果打算新增物流场地及线路，我们可以考虑在与 DC10、DC14、DC4 场地距离较近的地方新增物流场地，然后新增线路连接新场地和原有场地。这样可以缓解因 DC10、DC14、DC4 关停而导致的网络负荷问题，同时可以更加灵活地处理包裹流转。

新增物流场地的处理能力应根据预测结果和实际情况进行调整。如果新增场地与已有场地的距离较近。则处理能力可以设置为与现有场地相近或略高即可，如果新增场地与已有场地距离较远，则需要考虑到场地间的物流运输成本等因素，来确定处理能力的大小。

新增线路的运输能力应该根据场地的转运能力进行调整。新线路的运输能力应该与已有线路的最大运输能力相近或略高，以确保流量能够合理分配，

考虑以上的分析，我们在原有 81 个场地的基础上构建 82 号场地，将 82 号场地作为中转站，并设置与之相连的路线：

新增站点：82

新增路线：4→82(300000)，36→82(300000)，23→82(200000)，82→8(300000)，
82→9(200000)，82→10(200000)，82→4(200000)

每条线的容量取原线路容量的平均值，即将 82 号所包含的路线，分担原有路线运输量的 1/2。

结合我们在 7.1 中绘制的随机分布网络图，我们对加入 82 号场地后的线路进行了负荷量可视化处理，我们可以从随机分布网络图中看到仅 36→82，23→82 处于满负荷状态。

八、模型评估

优点：

- 1、在问题一中，我们使用了 SARIMAX 模型，此模型中，我们不仅考虑到了货流量可能存在季节性的变化这一因素，更考虑到了存在六一八、双十一等大促可能造成货流量数值激增的问题，因而我们引进了外生变量以对我们的 ARIMA 模型进行优化。
- 2、在问题二中，我们建立了 NSGA-II 模型，将原本很复杂的分配给每条线路的货流量这 n 个决策变量进行简化，将其放在一个 n 维向量中作为单一决策变量输入，将许多的复杂决策变量简化为单个决策变量，使模型变得简单直观。
- 3、在问题三中，我们引入一个二元变量，以 0 和 1 分别代表线路的开放与否，使得优化器在寻找最优解时考虑场地是否开放的影响。
- 4、在问题四中，我们使用熵权 TOPSIS 法对我们的线路进行重要性评估，我们合理地提取了四个评估指标，使得我们的评估更加准确可信。

缺点：

- 1、鉴于我们需要应用的数据集非常的大，我们 NSGA-II 模型的求解过程可能非常耗时，因此我认为我们的模型存在在运行时间和计算资源的需求上较大的缺点。
- 2、在我们对模型的评估过程中，我们对指标间关系的假设较为简单。由于指标权重是由层次分析法得出的，因此我们熵权 TOPSIS 法的结果可能会受到指标权重的选择过于主观的影响。

九、参考文献

[1]申永生,何世伟,穆美如等. 综合运输体系下货运流量分配优化模型与算法[C]//Intelligent Information Technology Application Association.Proceedings of the 2011 International Conference on Software Engineering and Multimedia Communication (SEMC 2011 V4) .INFORMATION ENGINEERING RESEARCH INSTITUTE,2011:4.

[2]李睿. 复杂网络上资源分配及流量控制问题研究[D].浙江大学,2018.

[3]石褚巍,马昌喜,麻存瑞.基于两阶段鲁棒优化的可靠性物流网络设计[J/OL].交通运输系统工程与信
息:1-20[2023-04-17].<http://kns.cnki.net/kcms/detail/11.4520.U.20230330.1452.014.html>.

[4]张婉琳.基于 GM(1,1)-ARIMA 组合模型的宁波港口物流需求预测分析[J].物流工程与管理,2023,45(01):32-35.

<https://kns.cnki.net/KXReader/Detail?invoice=OeuX3%2FMMKaREdCRU6rKhkgHMfNePiFKK3P5qQbR1HJmv9m%2FWoWLGmReKKnorrV60rHXRCkphxdnAxwAbmYL%2BaU5TDG5VtfrsSmx5K0qAnL1sDC%2FcmASYu5JoeOZh509PQxMGtj%2FESLt9TCXFFY4vElsInnVv86d0n1sChaHhMk%3D&DBCODE=CJFD&FileName=SPCY202301009&TABLEName=cjfdlast2023&nonce=DE3B32DBF611477A8E45340A94F34D5B&uid=&TIMESTAMP=1681379294149>

[5]卢尔赛,张改平,赵良等.基于灰色 ARIMA 组合模型的砂石骨料物流需求预测——以浙江省为例[J].交通运输研究,2022,8(05):84-90.DOI:10.16503/j.cnki.2095-9931.2022.05.010.

<https://kns.cnki.net/KXReader/Detail?invoice=xDBQV53GvNvK7wfkIjJ9QTfInUHWIRfGuDrz%2FXqxXNbkKDSCTo%2Bb7LI4xivmg2d49BTUvCDHJfHOBiSdG%2FbGgfU%2FJ%2Fr25i9CRMmF09ChBvWwxrDJNBbpccPvZC04TXuYdLZUlmw4nE3VgKEgezpDqIXHyNXuYQHlLaIMtyUCA%3D&DBCODE=CJFD&FileName=JTBH202205010&TABLEName=cjfdlast2022&nonce=70B50DE087B54CCFA6226D929A606A4D&uid=&TIMESTAMP=1681379607234>

[6]杨喜文. 基于 NSGA- II 算法的正逆向物流网络中回收处理中心选址[D].大连海事大学,2020.DOI:10.26989/d.cnki.gdlhu.2020.001270.

[https://kreader-cnki-net-s.libvpn.scnu.edu.cn:20080/Kreader/CatalogViewPage.aspx?dbCode=cmd&filename=1020087827.nh&tablename=CMFD202101&compose=&first=1&uid=WEEvREcwSIJHSldSdmVpaVVVQVRBWXFHMVRQYmcxVnlqbXVtUFFUTlg5MD0=\\$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4IQMovwHtwkF4VYPoHbKxJw!!](https://kreader-cnki-net-s.libvpn.scnu.edu.cn:20080/Kreader/CatalogViewPage.aspx?dbCode=cmd&filename=1020087827.nh&tablename=CMFD202101&compose=&first=1&uid=WEEvREcwSIJHSldSdmVpaVVVQVRBWXFHMVRQYmcxVnlqbXVtUFFUTlg5MD0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4IQMovwHtwkF4VYPoHbKxJw!!)

[7]石褚巍,马昌喜,麻存瑞.基于两阶段鲁棒优化的可靠性物流网络设计[J/OL].交通运输系统工程与信息:1-20[2023-04-17]

[8]杨进伟. 基于 GM/BP 神经网络组合预测模型的区域物流量预测研究[D].中南大学,2011.