

---

# 数学建模选读及文献检索 课程总结报告

姓 名: Donny SHEN  
专 业: 数理基础科学  
学 号:   
完成日期: 2024. 11. 08

大连理工大学

Dalian University of Technology

## 一、任务 1

1、请利用课堂上所讲的 Web of Science 数据库，完成以下任务，要求写出检索过程及分析报告。

(1) 以“囚徒困境”问题为主题进行文献检索，选取 5 年内该专题发表文章较多的三个期刊，将期刊名、期刊所属学科领域、最新影响因子等结果制成表格。

(2) 任选相关领域发表文章较多的一位作者，检索其最新发表的文章或被引次数最多的文章，翻译其摘要。

选择数据库: 所有数据库 合集: All

文献 被引参考文献

主题 Prisoner's Dilemma

AND 主题 prisoner dilemma

出版日期 最近 5 年

+ 添加行 高级检索

清除 检索

二、图 1-1：检索关键词

首先通过关键词搜索与出版日期筛选得出五年内关于“囚徒困境”主题的文献。

显示 25 共计 689 条目  
71 条记录 (5.014%) 不包含所分析字段的数据

全选	字段: 出版物/来源出版物名称	记录数	1,416 的百分位
<input type="checkbox"/>	CHAOS SOLITONS FRACTALS	76	5.367%
<input type="checkbox"/>	APPLIED MATHEMATICS AND COMPUTATION	70	4.944%
<input type="checkbox"/>	ARXIV	48	3.390%
<input type="checkbox"/>	PHYSICA A STATISTICAL MECHANICS AND ITS APPLICATIONS	41	2.895%
<input type="checkbox"/>	SCIENTIFIC REPORTS	38	2.684%
<input type="checkbox"/>	CHAOS WOODBURY N Y	21	1.483%
<input type="checkbox"/>	CHAOS	20	1.412%
<input type="checkbox"/>	EPL	20	1.412%
<input type="checkbox"/>	CHAOS AN INTERDISCIPLINARY JOURNAL OF NONLINEAR SCIENCE	19	1.342%
<input type="checkbox"/>	EPL EUROPEAN LETTERS	10	1.247%

三、图 1-2：对期刊文章数量降序筛选

第二步，通过对期刊所发表的文章进行降序排列，从而实现文章数量比较，并

制作表格。由于 ARXIV 的特殊性，将其单独列出，不计入排名。

期刊名	学科领域	影响因子 2023	影响因子五年
APPLIED MATHEMATICS AND COMPUTATION	MATHEMATICS APPLIED	3.5	3.1
PHYSICA A-STATISTICAL MECHANICS AND ITS APPLICATIONS	PHYSICS, MULTIDISC IPLINARY	2.8	2.6
CHAOS SOLITONS & FRACTALS	MATHEMATICSINTERD ISCIPLINARY APPLICATIONS	5.3	5.4

四、表 1-1：期刊表格

全选 <input type="checkbox"/>	字段: 作者	记录数	41的百分位
<input type="checkbox"/>	Wang J	4	9.756%
<input type="checkbox"/>	Xia C	3	7.317%
<input type="checkbox"/>	Xia Cy	3	7.317%
<input type="checkbox"/>	Chen G	2	4.878%
<input type="checkbox"/>	Chen L	2	4.878%
<input type="checkbox"/>	Cheng L	2	4.878%
<input type="checkbox"/>	Fan S	2	4.878%
<input type="checkbox"/>	Fan Sh	2	4.878%
<input type="checkbox"/>	Li X	2	4.878%
<input type="checkbox"/>	Li X P	2	4.878%
<input type="checkbox"/>	Li Xp	2	4.878%
<input type="checkbox"/>	Li Y	2	4.878%

图 1-3：筛选作者

通过对作者进行降序筛选，可以看出 Wang J 为发表文章最多的人。其最新发表的文章为：The link weight adjustment considering historical strategy promotes the cooperation in the spatial prisoner's dilemma game.

查询文章，其摘要为：Nothing remains the same in social networks, especially the closeness of relationships between selfish individuals. With the passage of time, this relationship may change dynamically and have an important impact on cooperation. In this paper, we propose a new evolutionary game model to investigate the evolution of cooperation, in

---

which the link weight is adaptively adjusted by comparing the individual payoff with her/his surrounding environment, and the learning ability of individual is affected by his historical strategies at the same time. To be specific, if the focal individual's payoff is greater than the average one of his nearest neighbors, the link weight between the focal individual and nearest neighbors will be increased by one unit; however, if the focal individual's payoff is smaller than the average one of his nearest neighbors, the link weight between them will be reduced by one unit; otherwise, the link weight between them will be unchanged. In addition, we use a specific parameter  $\epsilon$  to determine the link weight adjustment range. Meanwhile, the focal individual will decide how to learn from her/his neighbor's strategy according to his strategy of the last  $M$  game rounds when she/he updates the current strategy. Through extensive Monte Carlo simulations, we find that dynamic adjustment of link weights can significantly promote the evolution of cooperation. Particularly, the parameter  $\delta$  determining the intensity of weight adjustment has an optimal value regarding the level of cooperation, and then the cooperation has significantly been improved with the growth of  $\epsilon$ . Also, there is an optimal memory length  $M$  as far as the emergence and persistence of cooperation is concerned. The current results are extremely conducive to understanding how selfish individuals in social dilemma dynamically adjust their relationships to promote the collective cooperation. (C)

2020 Elsevier B.V. All rights reserved.

译文如下：在社会网络中，没有什么是一成不变的，尤其是自私的个体之间的亲密关系。随着时间的推移，这种关系可能会发生动态变化，并对合作产生重要影响。本文提出了一个新演化的博弈模型来研究合作演化，在这个模型中，权重是通过比较个体的报酬和周围环境来进行自适应调整的，同时个体的学习能力也会受到其历史策略的影响。具体来说，如果焦点个体的报酬大于其近邻的平均报酬，则焦点个体与近邻之间的联系权重将增加一个单位；但如果焦点个体的报酬小于其近邻的平均报酬，则它们之间的联系权重将减少一个单位；否则，它们

---

之间的联系权重将保持不变。此外，我们还使用特定参数  $\epsilon$  来确定链接权重的调整范围。同时，焦点个体在更新当前策略时，会根据邻居在过去  $M$  轮博弈中的策略来决定如何学习邻居的策略。通过大量的蒙特卡罗模拟，我们发现动态调整链接权重能显著促进合作的发展。特别是，决定权重调整强度的参数  $\delta$  对于合作水平有一个最佳值，然后随着  $\epsilon$  的增长，合作得到了显著改善。此外，就合作的出现和持续性而言，存在一个最佳记忆长度  $M$ 。目前的研究结果非常有助于理解社会困境中的自私个体如何动态调整关系以促进集体合作。

## 二、任务 3

在现代社会中，传染病、信息、谣言等传播问题备受关注。这些现象的传播过程具有复杂的非线性特性，难以通过传统的解析方法进行精确描述。元胞自动机和蒙特卡洛方法提供了有效的手段来模拟和分析这种传播过程。元胞自动机是一种离散的数学模型，通过简单的局部规则来描述系统状态的演化，特别适用于描述空间传播动态行为。而蒙特卡洛方法则是一种基于概率统计的数值模拟方法，常用于研究随机过程和复杂系统中的不确定性。本模型基于元胞自动机给出一个感染模型。

### 模型描述：

#### 1. 元胞自动机模型

元胞自动机模型将传播过程简化为网格中的元胞状态变化，每个元胞可以代表个体或者空间单位。每个元胞的状态在每个时间步长中根据其邻域中的元胞状态和特定的传播规则进行更新。

- **空间结构：** 使用二维网格表示传播区域，每个网格点代表一个元胞，初始时部分元胞处于“感染”状态，其余元胞处于“易感”状态。
- **状态设定：** 每个元胞有三种可能的状态：易感 (S)、感染 (I)、免疫或恢复 (R)。
- **传播规则：** 在每个时间步长中，每个感染元胞有一定概率  $ppp$  感染其邻近的易感元胞。邻近的定义是摩尔邻域 (Moore neighborhood，包含上下左右以及对角线共 8 个方向的元胞)。感染概率  $ppp$  可以根据传播力、

---

接触次数等参数进行设定。

- **恢复与免疫：** 每个感染元胞在经过一定时间 TTT 后会转为免疫或恢复状态，不再具有传染性。
- **演化过程：** 模拟演化过程中，观察系统中感染元胞数量的变化，研究不同初始条件和传播参数对整个系统的影响。

具体的模型步骤如下：

1. 初始化二维网格，设定初始感染元胞位置，其余元胞设为易感状态。
  2. 设定感染概率  $ppp$  和恢复时间 TTT。
  3. 在每个时间步长中，遍历所有元胞，根据其邻域内感染元胞的数量和感染概率  $ppp$  决定易感元胞是否被感染。
  4. 对于已感染的元胞，记录其感染时间，当感染时间达到 TTT 时，该元胞转为免疫状态。
  5. 重复以上步骤，直到系统中不再有感染元胞，或达到预设的最大时间步数。
- **个体建模：** 将整个系统中的个体视为一个样本集合，利用随机数模拟个体间的接触和传染过程。
  - **状态设定：** 每个个体有三种状态：易感 (S)、感染 (I)、免疫或恢复 (R)。
  - **传播概率：** 在每个时间步长中，根据一定的传播概率  $ppp$  和接触次数，确定哪些个体由易感状态转为感染状态。传播概率可以由以下公式给出：

$$P_{\text{感染}} = 1 - (1 - p)^n$$

- **多次模拟：** 通过多次独立模拟，得到每次传播过程的演化路径，并通过统计分析计算感染人数的期望值、方差等，揭示系统的整体传播特性。
- **演化过程：**
  - 初始化个体集合，设定初始感染个体，其余个体为易感状态。
  - 在每个时间步长中，随机选择个体进行接触，根据传播概率确定易感个体是否被感染。
  - 对于已感染的个体，经过一定时间后转为免疫状态。
  - 重复上述过程，进行多次独立模拟，统计每次模拟中感染个体数量的变化情况。

## 2024 年妈妈杯 C 题论文 摘 要

电商物流网络中分拣中心是一个关键的环节，负责将订单按照不同的流向进行分拣以便将包裹送达到消费者手中。因此提升分拣中心的管理效率对整体网络的履约效率和运营成本至关重要。本文主要根据分拣中心的历史货量与物流网络配置等信息来预测分拣中心每天或者每小时的货量。同时需要根据货物的预测结果对人员进行排班。

针对第一个问题，题目要求利用分拣中心的历史货量数据构建货量预测模型，对 57 个分拣中心未来 30 天每天及每小时的货量进行预测。我们采用 LSTM 模型根据每个分拣中心历史的历史货量数据来对此时序预测任务进行建模，首先将历史数据整理为时序数据，然后初始化 LSTM 模型进行训练并进行评估，最后对 57 个分拣中心未来 30 天每天及每小时的货量进行预测。

针对第二个问题，题目在第一问的基础上增加了运输线路这一维度数据，需要在运输线路发生变化时对分拣中心的货量进行预测。我们利用与第一问类似的做法，首先将附件三中的涉及的数据进行处理，然后利用始发分拣中心、到达分拣中心的路线、货量作为特征构建并训练 LSTM 模型。对于新增路线，我们对附件三、四进行合并确定新增运输路线的情况，以附件 3 中的平均货量作为其初始货量。最后，用训练得到的模型对附件 4 中每个分拣中心的货量进行预测。

针对第三个问题，我们需要根据分拣中心的货量情况，安排每天出勤的正式员工和临时工数。在这个问题中，我们需要最小化总人天数作为目标函数，同时要尽量让每个班次的小时人效接近一个平均效率。我们通过定义变量、参数、约束条件和优化目标将此问题转化为一个混合整数规划问题，然后采用求解器进行求解。最后得到了每个分拣中心每天每个班次中正式员工和临时员工的人数。

针对第四个问题，我们需要对 SC60 这个分拣中心给出未来 30 天每天六个班次中，每名正式员工将在哪些班次出勤，每个班次需要雇佣多少临时工。我们首先根据约束条件明确定义需要解决的问题以及设计目标函数，然后采用遗传算法进行建模求解。

关键词：物流网络、LSTM、混合整数规划、遗传算法

## 目录

摘 要 .....	1
1 引言 .....	3
1.1 问题背景 .....	3
1.2 问题重述 .....	3
2 模型假设 .....	4
3 符号说明 .....	4
4 问题一 .....	5
4.1 数据描述和问题一分析 .....	5
4.2 模型的建立 .....	6
4.3 问题一求解 .....	8
5 问题二 .....	10
5.1 数据描述与问题二分析 .....	10
4.2 问题二求解 .....	11
6 问题三 .....	11
6.1 问题三分析 .....	11
6.2 模型的建立 .....	12
6.3 问题三的求解 .....	12
7 问题四 .....	13
7.1 问题四分析 .....	13
7.2 建模的建立 .....	14
7.3 问题四的求解 .....	15
8 模型评估与进一步讨论 .....	17
8.1 模型的优点 .....	17
8.2 模型的缺点 .....	17
8.3 总结 .....	17
9 参考文献 .....	18

具体代码见：

<https://github.com/DonnySHEN/mathematical-modelling-code>



# 1 引言

## 1.1 问题背景

电商物流网络中，分拣中心是至关重要的环节，它负责将订单按照不同的流向进行分拣，以确保包裹能够快速准确地送达消费者手中。对分拣中心的货量预测是电商物流网络中一个关键的研究问题，准确的货量预测是后续管理和决策的基础。如果管理者能够提前了解各个分拣中心未来一段时间需要处理的货量，就可以提前做好资源安排，从而提高效率并降低成本。因此，分拣中心管理效率的提升对整体网络的履约效率和运营成本有着重要的影响。

本文旨在利用模型算法通过分析物流中心的数据，包括货量、货物流向、包裹分拣速度等信息，结合模型算法，可以实现更精准的货量预测与人员排班计划。这样的优化能够有效地调配人力资源，使得物流中心在不同时间段和不同工作需求下都能够保持高效运转，提高订单处理速度，降低等待时间，进而提升客户满意度并降低运营成本。

## 1.2 问题重述

问题一：建立货量预测模型，对 57 个分拣中心未来 30 天每天及每小时的货量进行预测，将预测结果写入结果表 1 和表 2 中。

问题二：过去 90 天各分拣中心之间的各运输线路平均货量如附件 3 所示。若未来 30 天分拣中心之间的运输线路发生了变化，具体如附件 4 所示根据附件 1-4，请对 57 个分拣中心未来 30 天每天及每小时的货量进行预测。并将预测结果写入结果表 3 和表 4 中

问题三：假设每个分拣中心有 60 名正式工，在人员安排时将优先使用正式工，若需额外人员将使用临时工。请基于问题 2 的预测结果建立模型，给出未来 30 天每个分拣中心每个班次的出勤人数，并写入结果表 5 中。要求在每天的货量处理完成的基础上，安排的人天数（例如 30 天每天出勤 200 名员工，则总人天数为 6000）尽可能少，且每天的实际小时人效尽量均衡。

问题四：研究特定分拣中心的排班问题，这里不妨以 SC60 为例，假设分拣中心 SC60 当前有 200 名正式工，请基于问题 2 的预测结果建立模型确定未来 30 天每名正式工及临时工的班次出勤计划，即给出未来 30 天每天六个班次中，每名正式工将在哪些班次出勤，每个班次需要雇佣多少临时工，并写入结果表 6 中。每名正式工的出勤率（出勤的天数除以总天数 30）不能高于 85%，且连续出勤天数不能超过 7 天。要求在每天货量处理完成的基础上，安排的人天数尽可能少，每天的实际小时人效尽量均衡且正式工出勤率尽量均衡。

## 2 模型假设

1、数据可信度与完整性：假设提供的数据是真实有效的，并且可能包含一些异常情况，但这些异常情况也是合理存在的。

2、缺失值处理：假设我们对缺失值的处理方式不会对预测结果产生显著影响。

3、新旧路线变化影响：假设新旧路线的变化直接影响到货量的增减，并且这种影响可以通过历史数据来量化和预测。

4、人员排班和劳动法规：假设每个分拣中心都遵守当地劳动法规，包括工作时间、连续工作日数限制和休息时间等，并且可以根据需求调整正式工和临时工的人数，且临时工的调整可以在短时间内进行。

5、工作平衡：假设特定分拣中心的排班需要平衡正式工的出勤率，使其不超过 85%，并且尽量均匀分配工作日，以避免劳动力集中和疲劳。

6、稳定的市场需求：假设市场需求是相对稳定的，即在预测的时间段内，市场需求不会受到剧烈的波动或突发事件的影响。

7、稳定的供应链环境：假设供应链的环境相对稳定，包括供应商、运输网络和分销渠道等方面的稳定性，以便进行准确的预测和规划。

8、工作效率假设：假设在排班安排下，工作人员的工作效率是相对稳定和可预测的，不受到外部因素的影响。

这些假设为我们建立模型提供了一些先验知识和限制条件，帮助我们更好地理解问题的背景和范围，以及在模型建立和预测过程中需要考虑的因素。

## 3 符号说明

本文使用的关键数学符号位于表 3-1 中。

表 3-1：本文使用的符号

变量名称	定义
$x_i$	第 $i$ 个班次的正式工人数
$y_i$	第 $i$ 个班次的临时工人数
$E_i$	第 $i$ 个班次的实际小时人效
$P$	预测的每班次需求的包裹量
$E_{max}$	正式工的最大小时人效（25 包裹/小时）
$E_{temp}$	临时工的最大小时人效（20 包裹/小时）
$x_{i,j,t}$	第 $t$ 天，第 $i$ 名员工在第 $j$ 个班次工作，则该变量为 1，否则为 0
$N_{formal}$	正式工的最大人数限制（200 人）

$h$	每天的实际小时人效
$r$	正式工的出勤率
$\lambda$	权重参数
$\mu$	权重参数
$\text{demand}_t$	第 $t$ 天的货量需求
$y_{i,j,t}$	第 $t$ 天, 第 $i$ 名临时工在第 $j$ 个班次工作, 则该变量为 1, 否则为 0
$E_i$	第 $i$ 个班次的实际小时人效
$p$	预测的每班次需求的包裹量

## 4 问题一

### 4.1 数据描述和问题一分析

该物流网络包括 57 个分拣中心, 每个分拣中心过去 4 个月的每天货量数据以及过去 30 天的每小时货量数据。我们选取了 8 个分拣中心每个月货量的平均值进行统计, 绘制的折线统计图如下图 4-1 所示; 对 SC48 分拣中心某两天每个小时的货量进行统计, 绘制的折线统计图如下图 4-2 所示。

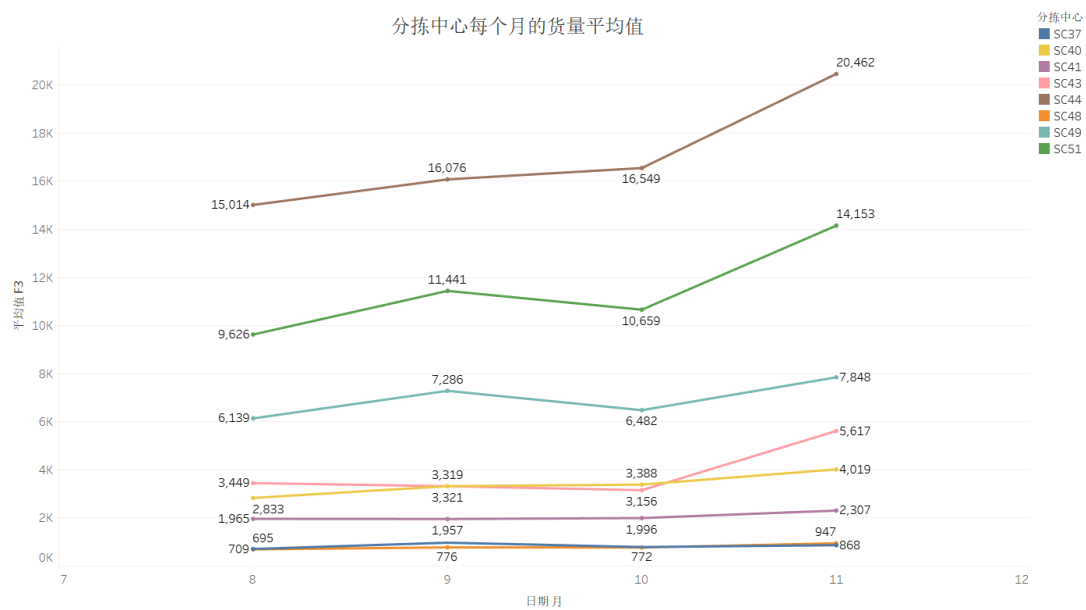


图 4-1 分拣中心每个月货量平均值折线图

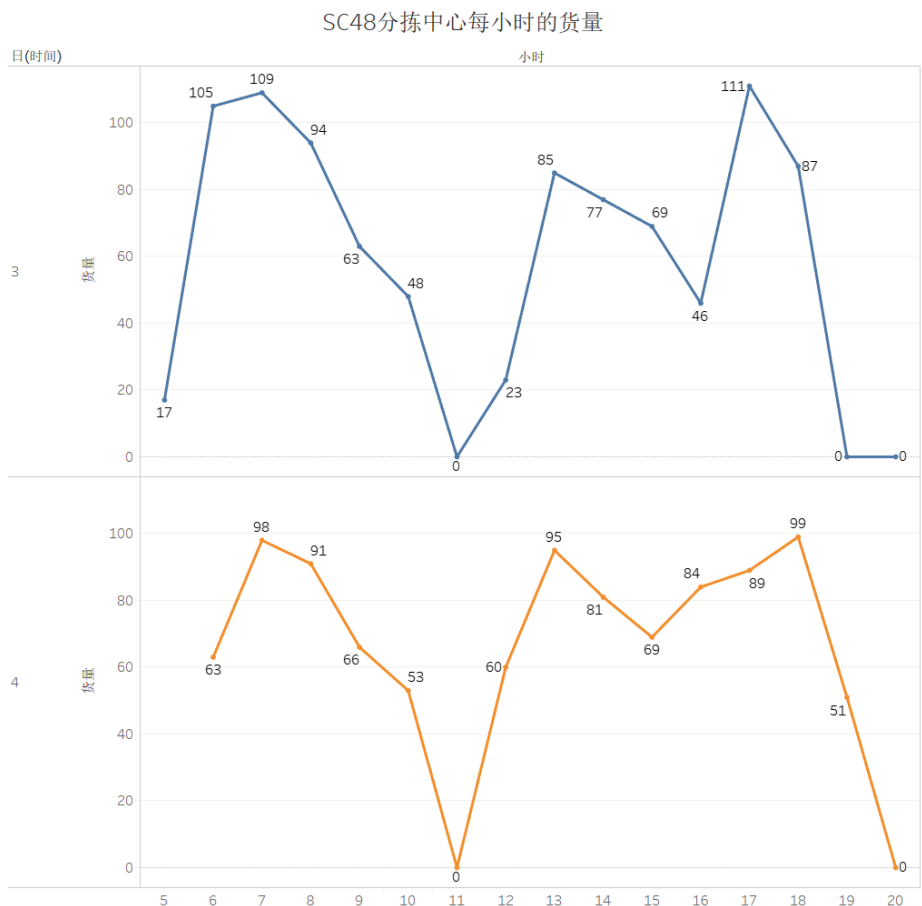


图 4-2 SC48 分拣中心每小时的货量

从图 1 和图 2 的统计结果可知，分拣中心的货量在不同月份以及每天的不同时段都有其周期性变化的趋势。因此我们可以采用时间序列预测算法来解决货量的预测问题。

## 4.2 模型的建立

时间序列任务时间序列数据预测任务是指利用过去的时间序列数据来预测未来的数值或趋势。这种任务通常涉及到对时间序列中的趋势、季节性、周期性以及随机噪声等方面进行建模，以便进行准确的预测。

LSTM 模型，LSTM 全称为长短期记忆（Long Short-Term Memory）网络，是一种特殊的递归神经网络（RNN）。与传统的前馈神经网络不同，LSTM 能够利用时间序列对输入数据进行分析，这使得它在处理诸如自然语言处理（NLP）或者连续数据点（例如股票价格）等任务时表现出色。LSTM 的核心在于它的内部门结构，这些门能够控制信息的存储、更新和遗忘。这样的设计使得 LSTM 能够解决传统 RNN 面临的梯度消失和梯度爆炸问题，从而有效地学习和记忆长期依赖关系。LSTM 的结构如下图所示：

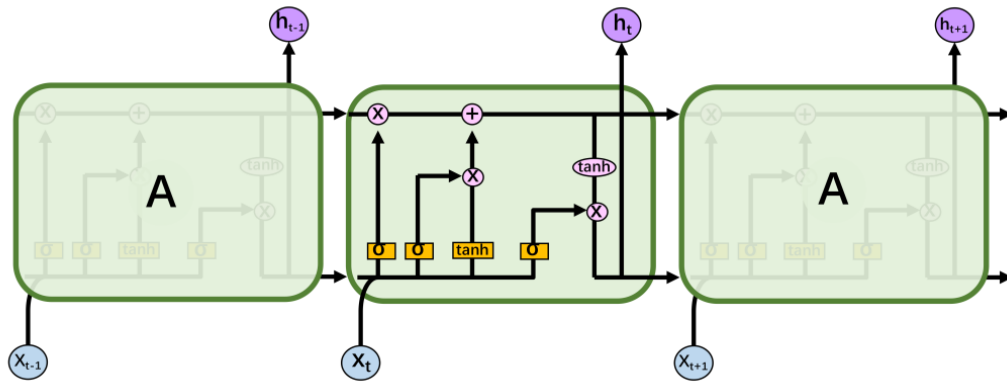


图 4-3 LSTM 模式的四个交互层

4-

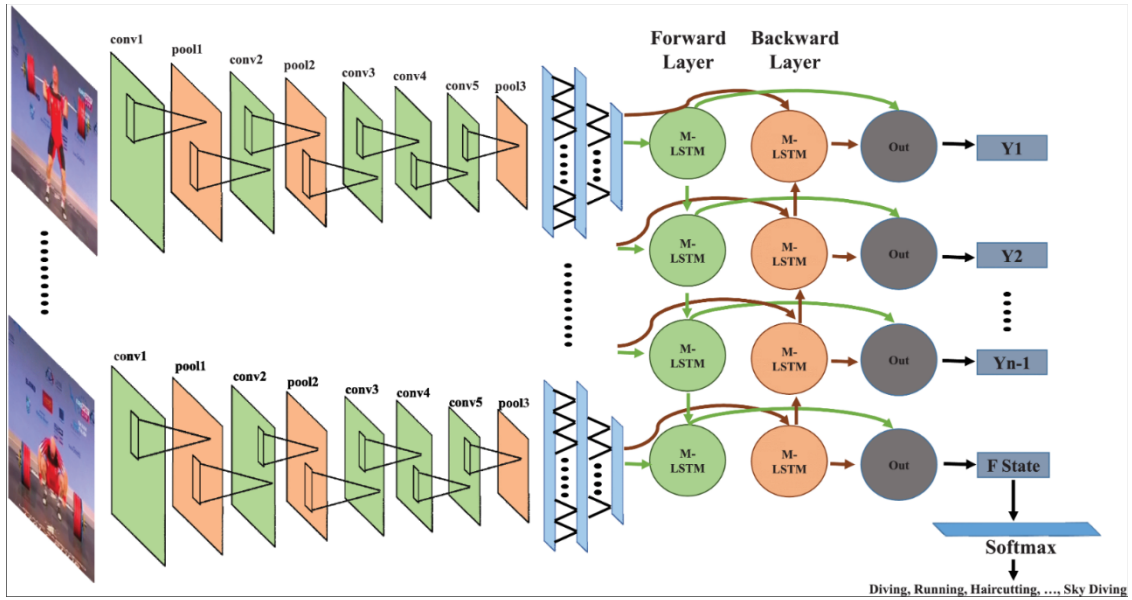


图 4-4 LSTM 网络结构

神经网络节点计算公式如下：

$$\begin{aligned}
 i_t &= \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{(t-1)} + b_{hi}) \\
 f_t &= \sigma(W_{if}x_t + b_{if} + W_{hf}h_{(t-1)} + b_{hf}) \\
 g_t &= \tanh(W_{ig}x_t + b_{ig} + W_{hg}h_{(t-1)} + b_{hg}) \\
 o_t &= \sigma(W_{io}x_t + b_{io} + W_{ho}h_{(t-1)} + b_{ho}) \\
 c_t &= \text{forget\_gate\_output} * c_{(t-1)} + i_t * g_t \\
 h_t &= o_t * \tanh(c_t)
 \end{aligned} \tag{1}$$

其中  $i_t$  表示输入门，决定了多少新的输入信息应该被加入到细胞状态中。这里的  $\sigma$  表示 sigmoid 激活函数， $W_{ii}$  和  $b_{ii}$  分别表示权重和偏置项， $x_t$  是当前时间步的输入， $h_{(t-1)}$  是前一个时间步的隐藏状态。

$f_t$ 表示遗忘门，控制着从细胞状态中丢弃什么信息。这同样使用 sigmoid 函数来决定哪些信息保留，哪些信息遗忘。

$g_t$ 表示候选值，是一个可能会被加入到细胞状态的新信息。这里使用的是 tanh 激活函数，它可以输出介于-1 和 1 之间的值。

$o_t$ 表示输出门，决定了下一个隐藏状态应该包含多少当前细胞状态的信息。这也是通过 sigmoid 函数来计算的。

$c_t$ 表示细胞状态，是 LSTM 单元的“记忆”部分，它结合了过去的信息（通过遗忘门）和新的信息（通过输入门和候选值）。

$h_t$ 表示隐藏状态，包含了当前时间步的输出信息，它是细胞状态的一个过滤版本，由输出门控制。

### 4.3 问题一求解

本文需要利用历史货量数据对货量进行预测即为一个时间序列预测任务，我们构建 LSTM 模型来对货量进行预测。以 SC48 分拣中心的数据作为示例，首先我们对数据进行标准化操作，再将时间序列数据转换为监督学习问题所需的输入特征和目标值。然后，我们将构建好的数据输入 LSTM 模型中进行训练，训练了 2 个 LSTM 模型，分别对每天及每小时的货量进行预测。训练好后模型的测试结果入下图 5 所示：

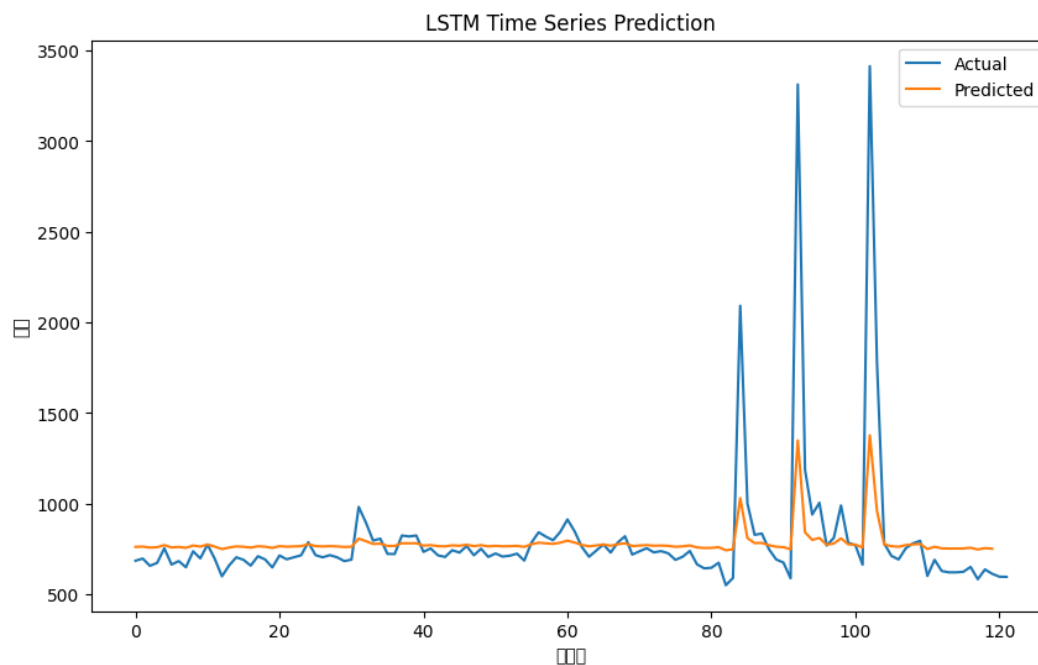


图 4-5 LSTM 每日货量预测结果

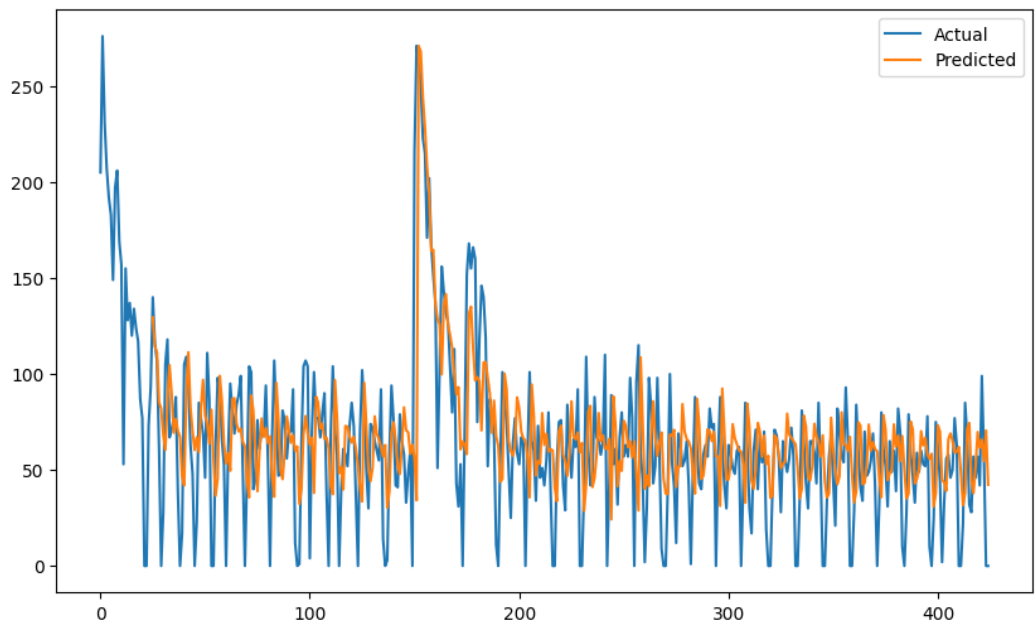


图 4-6 LSTM 每小时货量预测结果

通过计算得到了模型的 RMSE 为: 32.6859，可以看出 LSTM 模型具有较好的预测效果。因此，我们分别对 57 个分拣中心未来 30 天每天及每小时的货量进行预测，并把结果写入结果表 4-1 和表 4-2 中。部分预测的结果如表 4-1、表 4-2 所示

表 4-1 未来 30 天每天货量预测结果

分拣中心	日期	预测货量	预测货量(取整)
SC48	2024/4/12	770.654792	770
SC48	2024/4/13	798.2795019	798
SC48	2024/4/14	802.8281794	802
SC48	2024/4/15	803.5818539	803
SC48	2024/4/16	803.7068546	803
SC48	2024/4/17	803.7275742	803
SC48	2024/4/18	803.7310097	803
SC48	2024/4/19	803.7315859	803
SC48	2024/4/20	803.7316926	803
SC48	2024/4/21	803.7316926	803

表 4-2 未来 30 天每小时货量预测结果

分拣中心	日期	小时	预测货量
SC54	2023/12/1	1	106.8180391
SC54	2023/12/1	2	106.8180391
SC54	2023/12/1	3	106.8180391
SC54	2023/12/1	4	106.8180391
SC54	2023/12/1	5	106.8180391
SC54	2023/12/1	6	106.8180391
SC54	2023/12/1	7	106.8180391
SC54	2023/12/1	8	106.8180391
SC54	2023/12/1	9	106.8180391
SC54	2023/12/1	10	106.8180391

## 5 问题二

### 5.1 数据描述与问题二分析

此文新增了分拣中心各个运输路线过去 90 天的平均货量数据，我们对数据进行可视化，做出的分拣中心不同路线货量统计气泡图如下图 6 所示，气泡的颜色越深气泡越大说明这条线路的货量越大，从图中我们可以看出从 SC6 到 SC12 这条路线的货量是最大的，同时也可以发现有很多路线的货量比较小，因此不同路线的货量的差异是比较大的，这可以作为我们预测分拣中心货量的一个重要条件。



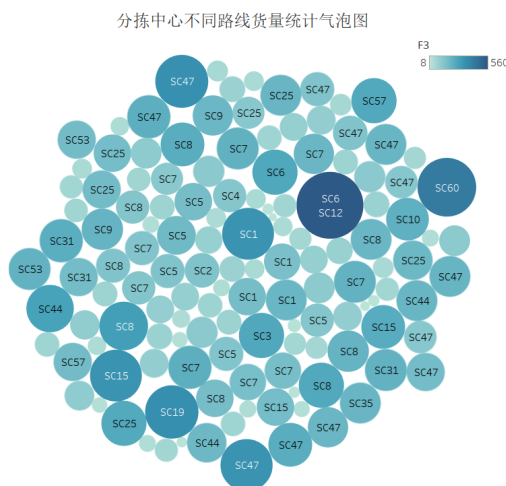


图 5-1 同路线货量统计气泡图

## 4.2 问题二求解

问题 2 需要对 57 个分拣中心未来 30 天每天及每小时的货量进行预测。我们首先将附件三中的涉及的数据进行处理，然后利用始发分拣中心、到达分拣中心的路线、货量作为特征构建 LSTM 模型。对于新增路线，我们对附件三、四进行合并确定新增运输路线的情况，以附件 3 中的平均货量作为其初始货量。模型的训练与第一问的过程类似，先把输入数据处理为 LSTM 模型需要的输入数据格式，然后构建 LSTM 模型训练 100 个 epoch 得到我们的预测模型。最后，用训练得到的模型对附件 4 中每个分拣中心的货量进行预测并将预测结果保存到结果 3、结果 4 文件中。

## 6 问题三

### 6.1 问题三分析

混合整数规划（Mixed Integer Programming, MIP）问题是一类数学优化问题，其目标是在满足一系列约束条件的情况下，最大化或最小化一个目标函数。与线性规划（LP）问题类似，MIP 问题的目标函数和约束条件都是线性的，但是在 MIP 问题中，决策变量中既包含连续型变量（即实数变量），又包含离散型变量（即整数变量）。这就是“混合”（Mixed）的含义。MIP 问题在实际中有广泛的应用，例如生产调度、物流优化、网络设计、投资组合优化等领域。

## 6.2 模型的建立

这个问题可以被视为一个混合整数规划问题。在这个问题中，我们需要最小化总人天数作为目标函数，下面是问题建模和解决的步骤，变量和参数的定义：

目标函数

$$\text{Minimize } \sum_{i=1}^6 (E_i - E_{avg})^2 \quad (2)$$

尽量让每个班次的小时人效 $E_i$ 接近一个平均效率 $E_{avg}$ ，可以通过最小化方差来实现，其中 $E_i$ 表示第 $i$ 个班次的实际小时人效。

约束条件：

要处理完成每天的货量。

$$x_i * E_{max} * 8 + y_i * E_{temp} * 8 \geq P_i, \text{ for all } i \quad (3)$$

其中 $x_i$ 表示第 $i$ 个班次的正式工人数， $y_i$ 为第 $i$ 个班次的临时工人数， $E_{max}$ 为正式工的最大小时人效， $E_{temp}$ 为临时工的最大小时人效， $P_i$ 为预测的每班次需求的包裹量。我们需要保证安排的员工能完成每天的包裹货量。

正式工使用不超过 60 人。

$$\sum_{i=1}^6 x_i \leq N_{formal} \quad (4)$$

其中 $x_i$ 表示第 $i$ 个班次的正式工人数， $N_{formal}$ 为正式工的最大人数限制（60 人）。每天的 6 个班次的正式员工数不超过 60。

人员非负性

$$x_i \geq 0, y_i \geq 0, \text{ for all } i \quad (5)$$

其中 $x_i$ 表示第 $i$ 个班次的正式工人数， $y_i$ 为第 $i$ 个班次的临时工人数，人员不能是负数。

每个正式工每天最多一个班次

此约束需要保证如果一个正式工安排在一个班次，那么其他班次中不能再出现此人。可以通过其他逻辑或二进制变量进行建模，但此处我们假设一人一班次即可。

## 6.3 问题三的求解

接下来，我们将使用 Python 中的 PuLP 包来构建和求解这个模型。并将结果写入结果表 5 中。预测结果如表 6-1 所示：

表 6-1 分拣中心排班表

分拣中心	日期	班次	正式工人数	临时工人数	总人天数
------	----	----	-------	-------	------

SC1	2023/12/1	00:00-08:00	60	61	121
SC1	2023/12/1	05:00-13:00	60	8	68
SC1	2023/12/1	08:00-16:00	40	0	40
SC1	2023/12/1	12:00-20:00	53	0	53
SC1	2023/12/1	14:00-22:00	27	0	27
SC1	2023/12/1	16:00-24:00	27	0	27
SC1	2023/12/2	00:00-08:00	60	56	116
SC1	2023/12/2	05:00-13:00	60	7	67
SC1	2023/12/2	08:00-16:00	40	0	40
SC1	2023/12/2	12:00-20:00	53	0	53

## 7 问题四

### 7.1 问题四分析

遗传算法（Genetic Algorithm, GA）是一种启发式优化算法，受到达尔文进化论的启发，模拟了生物进化过程中的遗传机制和自然选择原理。遗传算法通过模拟生物种群的遗传、交叉和变异等过程，逐步优化解空间中的解，以求解优化问题。

遗传算法的基本思想是将问题的解表示为种群中的个体，每个个体都有一定的适应度值，反映了其在解空间中的优劣程度。通过选择、交叉和变异等操作，不断迭代优化种群中的个体，直到找到满足停止条件的最优解或近似最优解。

遗传算法的主要特点包括：

**全局搜索能力：**遗传算法具有较强的全局搜索能力，能够在复杂的解空间中搜索到较优解或近似最优解；**并行性：**遗传算法的并行性较强，可以同时种群中的多个个体进行操作，加速优化过程；**适应性：**遗传算法具有自适应性，能够适应不同问题的特点和求解需求，并通过调整参数和操作来优化搜索过程；**灵活性：**遗传算法具有较强的灵活性，可以应用于各种不同类型的优化问题，并且可以结合其他优化方法和启发式算法进行改进和扩展。

遗传算法在实际应用中被广泛应用于复杂的优化问题，如组合优化、函数优化、调度问题、机器学习、神经网络训练等领域。它不仅适用于单目标优化问题，还可以用于多目标优化和约束优化等问题。由于其简单、有效和通用的特点，遗传算法成为了求解复杂优化问题的重要工具之一。

遗传算法基本流程图：

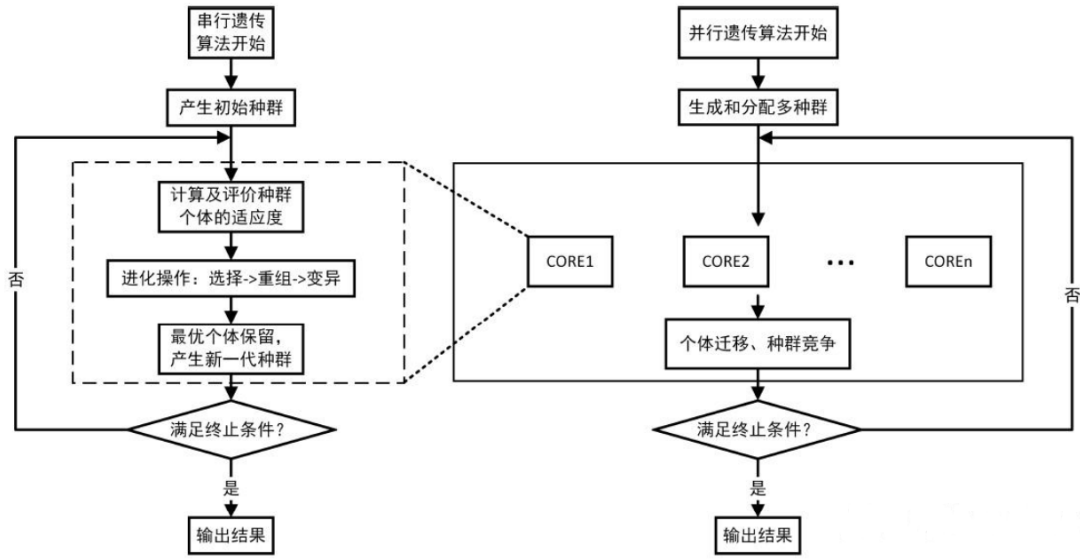


图 7-1 遗传算法流程图

## 7.2 建模的建立

在这个问题中，我们的目标是最小化总的人天数和实际小时人效的不均衡度，同时使正式工出勤率尽量均衡，下面是采用遗传算法对问题进行建模和解决的步骤，变量和参数的定义如下：

定义问题：明确定义需要解决的问题，包括目标函数、约束条件。

目标函数：

$$\text{Minimize } \sum_{i,j,t} x_{i,j,t} + \sum_{i,j,t} y_{i,j,t} + \lambda h + \mu \lambda \quad (6)$$

目标是最小化总的人天数和实际小时人效的不均衡度，同时使正式工出勤率尽量均衡。这可以通过一个加权的形式来表示。其中  $x_{i,j,t}$ ，表示第  $t$  天，第  $i$  名正式员工在第  $j$  个班次工作， $y_{i,j,t}$  表示第  $t$  天，第  $i$  名零时员工在第  $j$  个班次工作。 $h$  为每天实际小时人数， $\lambda$  和  $\mu$  为权重参数。

约束条件：

要处理完成每天的货量。

$$x_{i,j,t} * 8 * E_{max} + y_{i,j,t} * 8 * E_{temp} \geq P_i, \text{ for all } i \quad (7)$$

其中  $x_{i,j,t}$ ，表示第  $t$  天，第  $i$  名正式员工在第  $j$  个班次工作， $y_{i,j,t}$  表示第  $t$  天，第  $i$  名零时员工在第  $j$  个班次工作， $E_{max}$  为正式工的最大小时人效， $E_{temp}$  为临时工的最大小时人效， $P_i$  为预测的每班次需求的包裹量。我们需要保证安排的员工能完成每天的包裹货量。

正式工使用不超过 200 人。

$$\sum_{i=1}^6 x_{i,j,t} \leq N_{\text{formal}} \quad (8)$$

其中 $x_{i,j,t}$ ，表示第 $t$ 天，第 $i$ 名正式员工在第 $j$ 个班次工作， $N_{formal}$ 为正式工的最大人数限制（200 人），正式员工不能超过 200 人。

人员非负性

$$x_{i,j,t} = 0 \text{ 或 } 1, y_{i,j,t} = 0 \text{ 或 } 1, \text{ for all } i, j, t \quad (9)$$

其中 $x_{i,j,t}$ ，表示第 $t$ 天，第 $i$ 名正式员工在第 $j$ 个班次工作， $y_{i,j,t}$ 表示第 $t$ 天，第 $i$ 名零时员工在第 $j$ 个班次工作。

每名正式工的出勤率不能高于 85%。

$$\frac{1}{30} \sum_{j,t} x_{i,j,t} \leq 0.85, \forall i \quad (10)$$

其中 $x_{i,j,t}$ ，表示第 $t$ 天，第 $i$ 名正式员工在第 $j$ 个班次工作。我们需要计算其一周的出勤率，使其不能高于 85%。

连续出勤天数不能超过 7 天。

$$\sum_{j,t=s}^{s+6} x_{i,j,t} \leq 7, \forall i, s \quad (11)$$

其中 $x_{i,j,t}$ ，表示第 $t$ 天，第 $i$ 名正式员工在第 $j$ 个班次工。需要保证正式工不能连续工作 7 天。

每个正式工和临时工每天最多一个班次

$$\sum_{j=1}^6 x_{i,j,t} \leq 1, \sum_{j=1}^6 y_{i,j,t} \leq 1 \forall i, t \quad (12)$$

其中 $x_{i,j,t}$ ，表示第 $t$ 天，第 $i$ 名正式员工在第 $j$ 个班次工作。 $y_{i,j,t}$ 表示第 $t$ 天，第 $i$ 名零时员工在第 $j$ 个班次工作。需要保证每天每个员工只安排一个班次。

### 7.3 问题四的求解

初始化种群：根据需求随机生成一定数量的个体组成初始种群。每个个体都对应着问题的一个解。

计算适应度：对每个个体利用适应度函数进行评估，计算其适应度值。适应度值反映了个体在解空间中的优劣程度。

选择操作：根据个体的适应度值，选择一定数量的个体作为父代，用于产生下一代个体。常见的选择操作包括轮盘赌选择、锦标赛选择等。

交叉操作：对选出的父代个体进行交叉操作，产生新的个体。交叉操作模拟了自然界中的基因交换过程，以产生新的个体。

变异操作：对交叉后的个体进行变异操作，引入随机扰动以增加搜索空间的多样性。变异操作有助于避免陷入局部最优解。

更新种群：将父代个体和新生成的个体合并，组成下一代种群。

重复迭代：重复进行选择、交叉、变异和更新种群的操作，直到达到终止条

件为止。通常，终止条件可以是达到最大迭代次数、目标函数值收敛到某个阈值、达到时间限制等。

输出结果：在算法停止迭代后，输出种群中最优个体对应的解作为最终的优化结果，以及其对应的目标函数值。

训练好模型后我们对未来 30 天每名正式工及临时工的班次出勤计划对员工进行安排，并把结果写入结果表 6 中，其中的部分结果如下表 7-1 所示：

表 7-1 分拣中心出勤计划表

分拣中心	日期	班次	出勤员工
SC1	2023/12/1	3	SC1-EMP3
SC2	2023/12/1	3	SC2-EMP1
SC2	2023/12/1	5	SC2-EMP3
SC3	2023/12/1	1	SC3-EMP1
SC3	2023/12/1	2	SC3-EMP2
SC3	2023/12/1	3	SC3-EMP3
SC4	2023/12/1	2	SC4-EMP1
SC4	2023/12/1	5	SC4-EMP2
SC4	2023/12/1	1	SC4-EMP3
SC5	2023/12/1	6	SC5-EMP1

选取其中三名员工的排班情况进行可视化，可视化结果如下图 7-1 所示。图中横坐标表示一个月的天数，纵坐标表示一天的 6 个班次，色块所在的坐标表示该名员工在这天的第几个班次值班。

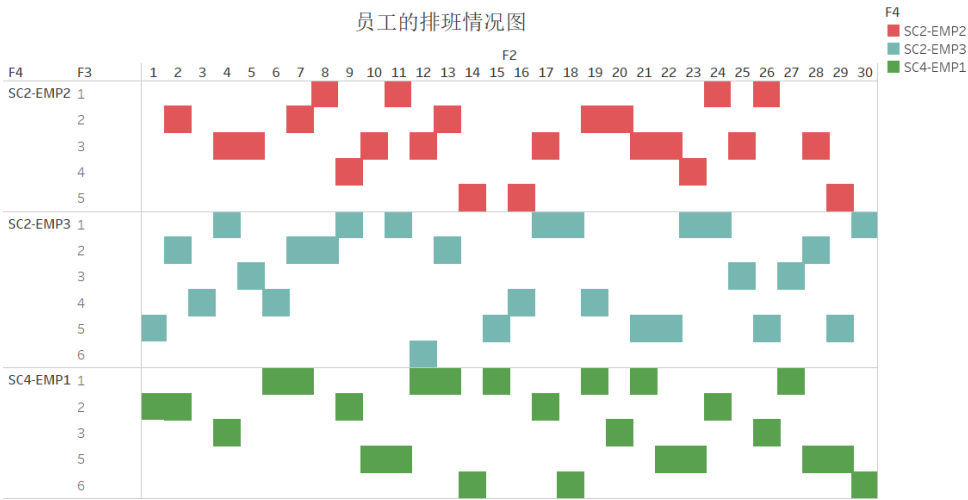


图 1-1 员工排班情况图

## 8 模型评估与进一步讨论

### 8.1 模型的优点

利用 LSTM 来处理和预测时间序列数据，可以有效地处理长序列数据并且不会受到梯度消失或梯度爆炸的影响。

MIP 能够灵活处理各种约束条件和目标函数，包括线性约束、非线性约束和离散变量，适用于复杂的问题。采用混合整数规划（MIP）求解人员排班问题够找到问题的全局最优解，确保了人员排班方案的最优性。

遗传算法是一种启发式搜索算法，能够灵活处理复杂的约束条件和目标函数，适用于各种类型的优化问题，包括排班问题。同时其能够进行全局搜索，不容易陷入局部最优解，有助于找到较优的排班方案。

### 8.2 模型的缺点

LSTM 模型通常需要大量的数据进行训练，以学习数据之间的复杂关系，如果数据量不足，容易导致过拟合。

在某些情况下，MIP 求解器可能会陷入局部最优解中，无法找到全局最优解，导致排班方案不够优化。

遗传算法有许多参数需要调整，如种群大小、交叉概率、变异概率等，对参数的选择和调整需要一定的经验和试验。

### 8.3 总结

电商物流网络中，分拣中心是一个关键的环节，负责将订单按照不同的流向进行分拣以便将包裹送达到消费者手中。本文主要根据分拣中心的历史货量与物流网络配置等信息来预测分拣中心每天或者每小时的货量。同时需要根据货物的预测结果对人员进行排班旨在提高分拣中心的管理效率，为电商物流网络的顺畅运作提供支持。

## 9 参考文献

- [1] 李宁. 物流配送网络规划问题研究 [D]. 北京交通大学 [2024-04-13]. DOI:10.7666/d.y1963088.
- [2] 袁际军. 现代物流配送路线优化研究 [D]. 武汉理工大学 [2024-04-13]. DOI:10.7666/d.y674307.
- [3] 张惠玲, 鲁鹏, 牟智伟, 等. 护士排班问题的研究 [J]. 管理学家: 学术版, 2014(1):2. DOI:10.3969/j.issn.1674-1722.2014.01.577.
- [4] 李加加, 沈吟东. 基于演化算法的多级别多岗位护士排班问题研究 [C]//第 36 届中国控制会议论文集 (B). 2017.
- [5] Nair, Vinod et al. "Solving Mixed Integer Programs Using Neural Networks." ArXiv abs/2012.13349 (2020): n. pag.
- [6] 李文静, 王潇潇. 基于简化型 LSTM 神经网络的时间序列预测方法 [J]. 北京工业大学学报, 2021, 47(5): 480-488. DOI: 10.11936/bjutxb2020120032
- [7] 王英伟, 马树才. 基于 ARIMA 和 LSTM 混合模型的时间序列预测 [J]. 计算机应用与软件, 2021. DOI:10.3969/j.issn.1000-386x.2021.02.047.