

XGBoost在故障预测及遗传算法在排班优化中的应用

何 娅*, 宋雅欣, 曾银香

重庆理工大学理学院, 重庆

收稿日期: 2024年11月24日; 录用日期: 2024年12月18日; 发布日期: 2024年12月26日

摘要

为了解决工业生产线中的故障智能识别与预测问题, 本研究利用XGBoost算法及相关性等方法对生产线故障数据进行了深入研究, 并基于遗传算法对排班问题进行了优化分析。首先整理了A工厂中生产线的详细记录, 同时构建XGBoost模型对生产线M201进行训练和预测, 提取各类故障的发生次数和持续时间, 揭示了故障与产量之间的关系。此外, 还探索了遗传算法在排班问题中的适用性, 并设计出相应的排班优化模型。研究结果显示, XGBoost模型在故障预测中具有较高的拟合度, 能够准确预测故障的发生及其持续时间, 并且预测得到了各装置每月的故障总次数和最长与最短的持续时长, 并且发现产品合格率与故障之间的关系。在排班优化方面, 遗传算法表现出了良好的适用性, 能够有效解决排班问题中的复杂性和不确定性。本研究将XGBoost算法应用于工业生产线故障预测中, 提高了故障识别的准确性和及时性。同时, 遗传算法在排班优化中的应用, 也为工业生产的调度和管理提供了新的思路和方法。这些研究成果对于推动工业智能化进程和提高生产效率具有重要意义。

关键词

XGBoost算法, 故障持续时间预测, 遗传算法, 排班优化

Application of XGBoost in Fault Prediction and Genetic Algorithm in Scheduling Optimisation

Ya He*, Yixin Song, Yinxian Zeng

College of Science, Chongqing University of Technology, Chongqing

Received: Nov. 24th, 2024; accepted: Dec. 18th, 2024; published: Dec. 26th, 2024

*通讯作者。

Abstract

In order to solve the problem of intelligent identification and prediction of faults in industrial production lines, this study conducted an in-depth study of production line fault data using XGBoost algorithm and correlation and other methods, and optimised the scheduling problem based on genetic algorithm. Firstly, the detailed records of the production line in Factory A were collated, while the XGBoost model was constructed to train and predict the production line M201, extracting the number of occurrences and duration of various types of faults, and revealing the relationship between the faults and the output. In addition, the applicability of genetic algorithms in scheduling problems was explored and corresponding scheduling optimisation models were designed. The results show that the XGBoost model has a high degree of fit in fault prediction, can accurately predict the occurrence of faults and their durations, and predicts the total number of faults and the longest and shortest durations of each unit per month, and finds the relationship between product qualification rate and faults. In scheduling optimisation, genetic algorithms show good applicability and can effectively solve the complexity and uncertainty in scheduling problems. In this study, XGBoost algorithm is applied to industrial production line fault prediction to improve the accuracy and timeliness of fault identification. Meanwhile, the application of genetic algorithm in scheduling optimisation also provides new ideas and methods for industrial production scheduling and management. These research results are of great significance for promoting the process of industrial intelligence and improving production efficiency.

Keywords

XGBoost Algorithm, Fault Intelligent Recognition, Genetic Algorithm, Scheduling Optimisation

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

科技浪潮推动了工业领域的智能化进程,自动化设备的广泛应用显著提升了生产效率与产能。然而,要实现全面自动化,必须解决生产线中的故障识别与预测问题。故障智能报警技术能够有效监控生产线状态,及时预警,防止生产中断和经济损失。在汽车制造业等领域的研究中,已探索出多种基于传统机器学习及深度学习算法的故障识别与预测方法[1]。XGBoost 算法作为一种高效的集成学习方法,实际运用中,在速度控制和生产优化方面表现出色[2][3],但仍面临模型拟合度不高、预测准确性不足等问题[4]。不过 XGBoost 算法在处理非线性关系和复杂数据时更具有高效性,可以实现对生产线故障的及时预警和有效预防,从而减少生产中断,提高整体生产效率。因此,本研究旨在通过 XGBoost 算法提高生产线故障预测的准确性和可靠性,为工业智能化转型提供创新性解决方案。此外,遗传算法以其强大的全局搜索能力被广泛应用于排班流程等领域[5]。在工业生产线中,故障的智能识别与预测对于提高生产效率、降低维护成本具有重要意义。并且随着企业规模的扩大和生产效率的提升,合理的人员配置也是一个重要问题,不合理的排班不仅会影响员工的工作效率,也会对企业的生产效率和产品质量造成不良影响。因此,研究并解决排班问题对于 A 企业的可持续发展具有重要意义,通过引入遗传算法,优化 A 企业的排班问题,以提高生产效率、降低成本并提升员工满意度。

2. 模型与方法介绍

2.1. XGBoost 算法在故障预测中的应用

XGBoost 是一种基于 GBDT 的高效、灵活、可扩展的梯度提升算法，是 GBDT 算法的一种变种，也归属于一种监督算法。此外，它还是一种 boost 算法，具有伸缩性强、便捷的可并行构建模型的特点。XGBoost 与 GBDT 的基本思想一致，都是利用加法模型与前向分步算法实现学习的优化过程，XGBoost 是利用加法模型与前向分步算法实现学习的优化过程，其预测过程如下：

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \quad (1)$$

其中， f_k 表示回归 X 树， K 为回归树的数量[6][7]。

XGBoost 在 GBDT 的基础上进行了很多改进，通过优化目标函数，提高了模型的精确度和训练速度，并且与其他机器学习算法相比，XGBoost 在处理非线性关系和复杂数据模式方面更具优势。此外，其内置的交叉验证功能有助于评估模型性能，确保预测的可靠性。

2.2. 遗传算法在排班优化中的应用

2.2.1. 遗传算法基本原理及优势分析

遗传算法(Genetic Algorithm, GA)是一种模拟自然选择和遗传机制的优化算法，通过模拟生物进化过程中的选择、交叉和变异等操作，不断迭代优化求解问题的解空间。与其他优化算法相比，遗传算法具有更强的鲁棒性和全局搜索能力[8]。遗传算法的基本步骤包括选择优良个体、交叉组合生成新个体、变异以避免局部最优，直至满足终止条件，标准遗传算法的流程如图 1 所示[9]：

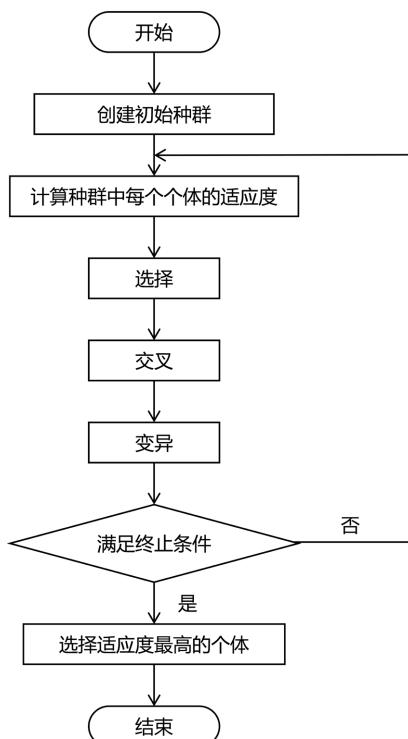


Figure 1. Genetic algorithm flowchart

图 1. 遗传算法流程图

遗传算法通过模拟生物进化过程，不断迭代优化，然后找到合理的排班方案，具体步骤如下：

- 1) 编码：将排班方案表示为染色体，即一个由基因组成的序列。每个基因可以代表某个员工在某个时间段内的排班情况。
- 2) 初始化种群：随机生成一定数量的染色体作为初始种群，每个染色体代表一个可能的排班方案。
- 3) 适应度评估：根据适应度函数计算每个染色体的适应度值，适应度值越高表示排班方案越优。
- 4) 选择：根据适应度值选择一定数量的优秀染色体作为父代，用于生成下一代。
- 5) 交叉：对父代染色体进行交叉操作，生成新的子代染色体。交叉操作可以保留父代染色体的优秀基因，同时引入新的基因组合。
- 6) 变异：对子代染色体进行变异操作，以一定概率改变某些基因的值。变异操作可以增加染色体的多样性，避免陷入局部最优解。
- 7) 迭代：重复上述步骤，直到达到预设的迭代次数或适应度值不再显著提高为止。最终得到的染色体即为最优排班方案。

2.2.2. 遗传算法在排班问题中的适用性

排班问题涉及多个约束条件，如班次需求、工人技能等，而遗传算法能够灵活处理这些约束，通过适应度函数评估排班方案的质量，实现优化目标。与其他优化算法相比，遗传算法具有更强的鲁棒性和全局搜索能力，可以设置更加合理的排班方案。

3. 数据预处理及可视化分析

3.1. 数据来源及预处理

本文数据来源是第 12 届泰迪杯 A 题数据(<https://www.tipdm.org>)，选取了 A 工厂中编号为 M101 至 M110 的生产线和操作人员作为研究对象，针对这些生产线上各类执行设备和检测设备的工作情况进行了详细记录。全面收集整理了每一道工序的执行信息，包括工序的状态、工序间的衔接情况以及产量合格数等关键数据。其初始数据的具体说明如图 2 所示：

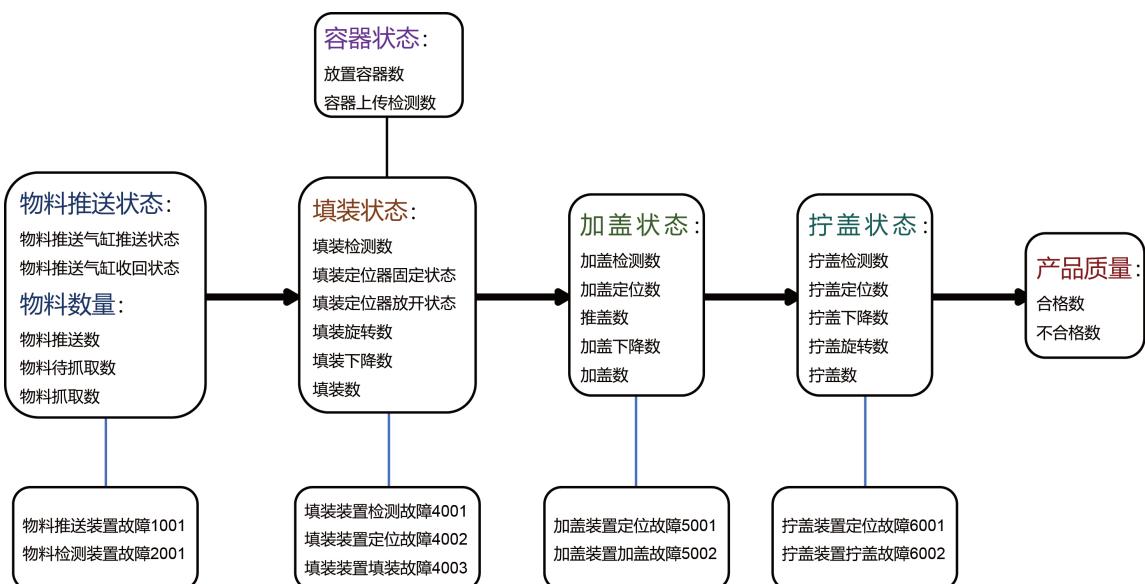


Figure 2. Initial feature state
图 2. 初始特征状态

由于处始数据量庞大,十条生产线的数据读取后共有 10,485,750 条记录。经检查,数据中未发现异常值,因此无需进行剔除处理。为了更有效地进行描述性统计分析,我们根据每天的不同时间段以及各产品线的生产时间进行了分层随机抽样。经过处理后,最终样本量减少至 787,230 条,这大幅降低了数据规模,同时保持了数据的代表性。

针对企业发展,我们对 A 工厂的工人进行了工龄的统计,得到工龄比例如表 1 所示:

Table 1. Ratio of length of service
表 1. 工龄比例

工龄	比率
1	0.2
2	0.2
3	0.1
4	0.2
5	0.2
6≤	0.1

可以看出该工厂大部分人的工龄是均匀分布在 1 年到 5 年之间的,特别是集中在前两年的工龄段和第四、五年的工龄段。

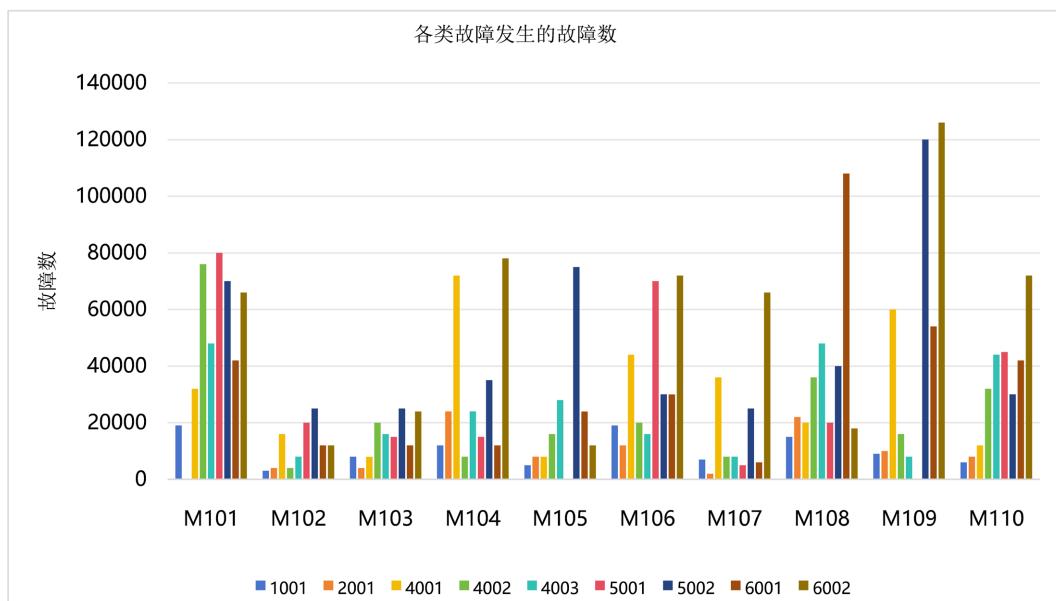
3.2. 数据可视化与相关性分析

然后我们将十条生产线的 35 个指标进行 Spearman 相关性分析,得到大多数指标与故障情况之间呈现正相关,结合故障运用和实际情况,为了更好地提高模型性能和贴合测试集,我们将全部数据进行拟合,将预处理之后的数据进行收集,对类故障数进行整理,详细如表 2 所示:

Table 2. Number of faults occurring for each type of fault
表 2. 各类故障发生的故障数

故障编号	M101	M102	M103	M104	M105	M106	M107	M108	M109	M110	总计
1001	19,019	3003	8008	12,012	5005	19,019	7007	15,015	9009	6006	103,103
2001	0	4002	4002	24,012	8004	12,006	2001	22,011	10,005	8004	94,047
4001	32,008	16,004	8002	72,018	8002	44,011	36,009	20,005	60,015	12,003	308,077
4002	76,038	4002	20,010	8004	16,008	20,010	8004	36,018	16,008	32,016	236,118
4003	48,036	8006	16,012	24,018	28,021	16,012	8006	48,036	8006	44,033	248,186
5001	80,016	20,004	15,003	15,003	0	70,014	5001	20,004	0	45,009	270,054
5002	70,028	25,010	25,010	35,014	75,030	30,012	25,010	40,016	120,048	30,012	475,190
6001	42,007	12,002	12,002	12,002	24,004	30,005	6001	108,018	54,009	42,007	342,057
6002	66,022	12,004	24,008	78,026	12,004	72,024	66,022	18,006	126,042	72,024	546,182

从表 2 可以看出,总故障数为 2,374,464,类故障 5002 和类故障 6002 的故障类型更为普遍,M102 和 M108 在多数故障类型中故障数较低,其他的故障数普遍呈现正态性,类故障图如图 3 所示:

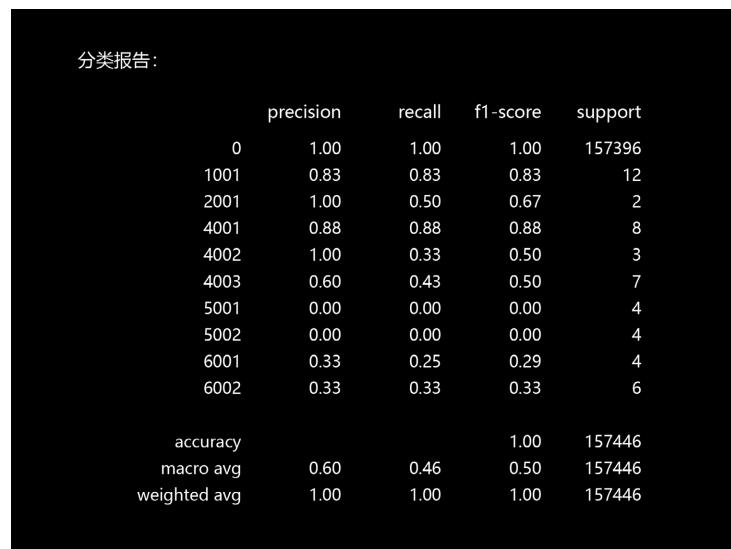
**Figure 3.** Failure count chart for each production line**图 3. 各生产线故障数图**

4. XGBoost 算法在故障智能识别中的应用

4.1. 模型建立及预测

4.1.1. 模型建立及检验

由于初始数据的庞大，而 XGBoost 算法在处理大规模数据集时表现优异，对于初始数据，首先将九个故障指标设为被检测变量 y ，并将物料推送气缸推送状态、物料推送气缸收回状态等 25 个变量作为输入特征 X ，使用 XGBoost 算法进行模型拟合。对于总共 787,230 条数据，以 8:2 的比例划分为训练集和测试集。该模型的分类结果如图 4 所示：

**Figure 4.** Model analysis diagram**图 4. 模型分析图**

将该模型应用于预测集后,可以看出预测值的准确率接近于 1,这表明算法具有很高的拟合度。通过将模型的预测值与测试集的实际值进行对比,我们构建了混淆矩阵如图 5 所示:

混淆矩阵:										
[157384	2	0	1	0	2	1	0	2	4]
[2	10	0	0	0	0	0	0	0	0]
[1	0	1	0	0	0	0	0	0	0]
[1	0	0	7	0	0	0	0	0	0]
[2	0	0	0	1	0	0	0	0	0]
[4	0	0	0	0	3	0	0	0	0]
[4	0	0	0	0	0	0	0	0	0]
[4	0	0	0	0	0	0	0	0	0]
[3	0	0	0	0	0	0	0	1	0]
[4	0	0	0	0	0	0	0	0	2]]

Figure 5. confusion matrix

图 5. 混淆矩阵

4.1.2. 故障持续时间预测

故障预测能够提前发现生产线的潜在问题,为维修和预防提供宝贵时间,减少生产中断,提高整体生产效率。针对已有的生产线,我们取时间状态一致的特征值众数,模拟了长度为 1,048,575 的生产线 M201 的所有 X 值,并进行建模和预测,首先将已建立的 XGBoost 模型用于预测故障类型,确定每次故障开始的时间点,假设初始时间为 0,计算每次故障的持续时间。预测的故障概率分布如图 6 所示:

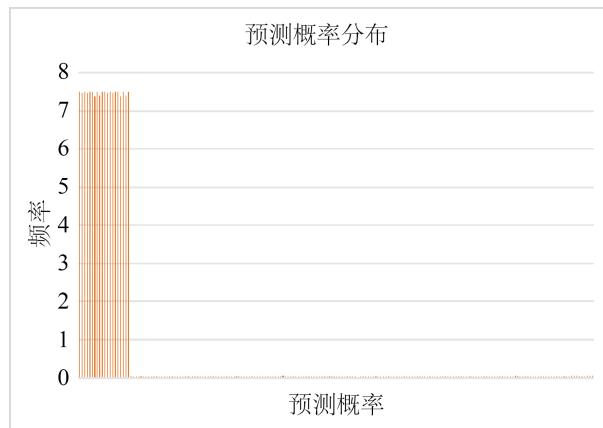


Figure 6. Predicted probability distribution plot

图 6. 预测概率分布图

对持续预测的结果进行整理,每类故障持续时间详细如表 3 所示。

从表 3 可以看出,9 类故障的总故障数共为 17,328 次,累计故障时长为 133,595 秒。其中,类故障 2001 的累计时长最少 3874 秒,持续时间也是最短的,即积累日期为 158,开始时间为 3323 秒,持续时间是 16 秒。而故障持续最长的是类故障 4001 开始时间是 1973 秒的 4001,持续时间高达 219 秒。并且故障累计时间最长和总故障数最多的是 4003,总故障数最少的是 4002,只有 92 次。

根据故障预测的结果,可以建立故障预警系统,并根据 XGBoost 算法提供的特征重要性,对生产线进行故障排查和维护保养,并优化生产流程。

Table 3. Class failure duration prediction**表 3. 类故障持续时间预测**

故障编号	日期	开始时间	持续时长/秒	累计时长/秒	总故障数
1001	320	15,309	173	7613	401
2001	158	3323	16	3874	595
4001	85	1973	219	6454	128
4002	256	26,353	174	4875	92
4003	353	25,624	91	36,584	6080
5001	249	9290	77	14,652	2184
5002	192	23,023	196	28,507	3591
6001	316	6248	151	20,231	2655
6002	159	18,920	21	10,805	1602

4.2. 故障与产量关系分析

整理合格数、不合格数以及九类故障数据时,对于每个生产的产品,结果只有两种:合格或不合格。通过将当日累积的合格数与不合格数相加,可以得到当日单个生产线的总产量。同样地,将当日合格数除以当日总产量,即可计算出当日的合格率。预测整理的故障数据对各变量求平均后,得到表 4:

Table 4. Annual average value of variables**表 4. 变量年均值**

	变量	数量
	产量	1437.023077
	合格率	0.999444223
持续时间	物料推送装置故障 1001	6.165384615
	物料检测装置故障 2001	8.576923077
	填装装置检测故障 4001	5.292307692
	填装装置定位故障 4002	9.919230769
	填装装置填装故障 4003	6.457692308
	加盖装置定位故障 5001	5.613461538
	加盖装置加盖故障 5002	6.403846154
	拧盖装置定位故障 6001	10.38461538
	拧盖装置拧盖故障 6002	4.751923077
故障次数	物料推送装置故障 1001	0.111538462
	物料检测装置故障 2001	0.157692308
	填装装置检测故障 4001	0.088461538
	填装装置定位故障 4002	0.184615385
	填装装置填装故障 4003	0.103846154
	加盖装置定位故障 5001	0.1
	加盖装置加盖故障 5002	0.103846154
	拧盖装置定位故障 6001	0.180769231
	拧盖装置拧盖故障 6002	0.088461538

观察表4可知,生产线的合格率高达99.94%,各类故障平均每天发生的次数较低,且不同类型的故障发生频率相差不大。然而,在各类故障的平均持续时间上,物料检测装置故障2001、填装装置定位故障4002和持续时间_拧盖装置定位故障6001,故障持续时间较长,因此,需要对这三类故障进行重点监测与维护。

将工龄与表4的数据进行结合,绘制出相关性的热力图为图7:

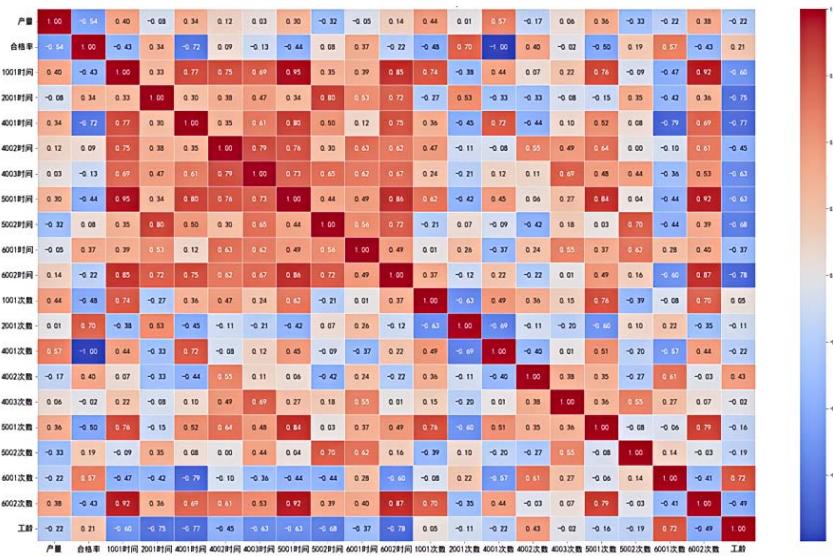


Figure 7. Heat map of correlation of variables

图7. 各变量相关性热力图

观察热力图结果,可得出以下结论:

- 1) 产量与其他变量的相关性较弱,可能是因为生产线中的故障次数和持续时间对其影响较小。
- 2) 类故障4001的故障次数与合格率呈极强的负相关性。该故障导致传送带无法停止,填装装置定位器无法固定容器,影响后续流程,降低合格率。企业应重点加强对此故障的监督和维修。
- 3) 类故障1001的持续时间与类故障5001、类故障6002的持续时间及故障次数呈正相关。该故障导致物料无法推出或气缸无法收回,影响后续加盖和拧盖流程。
- 4) 类故障2001的持续时间与类故障5002的持续时间呈正相关。物料检测失败导致物料推送和传送带停止,影响后续流程。
- 5) 类故障4001的持续时间与类故障5001的持续时间呈正相关。容器上传但未检测到,导致产品无法正确定位和加盖。
- 6) 类故障5001的持续时间与类故障6002的持续时间及故障次数呈正相关。若定位器无法固定产品,则产品无法加盖,影响后续拧盖流程。故障次数越多,持续时间越长,加剧了拧盖装置的故障。

然后根据图7我们可以发现产量与合格率存在较弱的负相关性,表明产量越多,故障也增多,合格率因此下降。对于合格率,大部分指标都与其有着较弱的负相关性,故障多了势必造成合格率下降。对于工龄,9类故障的持续时间与工龄均呈现较弱的负相关性,说明随着工龄的增加,对故障的应对处理能力会增强。总的来说,企业应加重对物料推送装置故障1001、物料检测装置故障2001和填装装置检测故障4001三处的监测和维修,因为它们处于生产线的前端也是重要的位置节点,这三处的故障发生,会进一步引致后续加盖装置和拧盖装置环节的故障发生,从而严重拉低产品质量。

5. 遗传算法在人员排班优化中的应用

5.1. 问题描述与模型建立

优化排班方案对于提高生产效率、降低人工成本具有重要意义，不合理的排班可能导致人员配置偏差，影响生产效率。为了解决这一问题，我们采用遗传算法来实现工业生产线中的排班优化，相较于传统的人工调度或简单的规则调度，遗传算法能够搜索全局最优解，更适应于动态变化的生产环境。

我们以如下情况为例，在A企业进入旺季时，生产线需要加大单量以满足市场需求，假设排班时间为1周，一年的排班方案可以以周排班方案循环，生产线共10条，需要每天运行，班次分为早班8~16点，中班16~24点以及晚班0~8点，考虑到员工上5休2，因此在班次设计中，增加了休息班次，需要42名工人，考虑到A工厂的操作人员工龄比例，每个工龄需要人数如表5所示。

Table 5. Number of workers

表5. 工人数

工龄	比率	数量
1	0.2	8
2	0.2	8
3	0.1	4
4	0.2	9
5	0.2	9
6≤	0.1	4

以这42名工人为研究对象，通过算法设计，得到最佳的操作人员排班方案。

5.2. 遗传算法优化排班方案的设计与实现

5.2.1. 遗传算法函数设计

每条生产线之间没有差异，因此生产线与操作人员之间可以任意搭配。以最小化排班次数为目标，目标函数为：

$$\min \sum_{d \in D, s \in S, e \in E} x_{d,s,e} \quad (2)$$

$$x_{d,s,e} = \begin{cases} 1, & \text{周 } d \text{ 班次 } s \text{ 员工 } e \text{ 上班} \\ 0, & \text{周 } d \text{ 班次 } s \text{ 员工 } e \text{ 不上班} \end{cases} \quad (3)$$

其中，周 $D = 1, 2, \dots, 7$ ，班次 $S = 1, 2, 3$ ，即早班，中班和晚班，员工编号 $E = 1, 2, \dots, 42$ 。

约束条件为：

$$\sum_{e \in E} x_{d,s,e} \geq N_{d,s}, \forall d \in D, s \in S \quad (4)$$

$$\sum_{s \in S} x_{d,s,e} \leq 1, \forall d \in D, e \in E \quad (5)$$

$$\begin{cases} x_{d,3,e} + x_{d+1,1,e} \leq 1 \\ x_{7,3,e} + x_{1,1,e} \leq 1 \end{cases} \quad (6)$$

$$\sum_{d \in D, s \in S} x_{d,s,e} = 5, e \in E \quad (7)$$

其中, $N_{d,s}$ 为每个班次每天需要在岗的人数, 即 10 人。

5.2.2. 适应度函数设计

适应度函数的设计旨在对排班方案的质量进行量化评估, 从而为遗传算法提供评价标准, 评价不同方案的质量, 算法优先考虑那些能平衡工作负荷且满足所有约束条件的方案。适应度函数首先遍历每一天的班次, 然后对每个班次中的工人数与期望值之间的偏差进行累加。如果某个班次的工人数与期望值相差较大, 则适应度值会受到惩罚。适应度函数的目标是尽量减少每天每个班次工人数与期望值之间的偏差, 从而使得排班方案更加合理。

5.2.3. 遗传算子选择

① 选择算子: 选择操作采用的是锦标赛选择方法。通过对种群中个体的适应度得分进行排序, 选择适应度最好的两个个体作为父母。

② 交叉算子: 交叉操作采用的是单点交叉方法。通过随机选择一个交叉点, 然后将两个父个体在交叉点之后的基因串进行交换, 生成两个新的子个体。

③ 变异算子: 变异操作采用的是随机变异方法。通过对每个个体的每个基因进行随机变异, 如果随机数小于变异率, 则随机选择一个非休息的班次进行替换。特别地, 还考虑了前一天是晚班的情况, 保证了第二天不会安排早班。

5.3. 结果与优化分析

在排班配置问题中, 使用遗传算法并设置合理的参数是一种有效的优化方法, 针对只有 42 位工人的情况, 算法需要在有限的资源下进行优化, 并综合考虑多种排班需求。为了确保算法的性能和效果, 根据遗传算法的设计, 利用 Python3.9, 模型参数设置如表 6 所示。

Table 6. Parameterisation
表 6. 参数设置

参数	参数值
种群大小	100
变异率	0.1
迭代次数	100

该方案能够满足所有工作需求, 同时考虑到工人的工龄、工作量、休息时间等因素。通过迭代优化, 遗传算法给出一周的最佳人员排班方案如表 7 所示。

Table 7. Work schedule
表 7. 人员排班表

日期	班次	M101	M102	M103	M104	M105	M106	M107	M108	M109	M110
1	早	B001	B006	B007	B009	B011	B020	B028	B031	B039	B040
1	中	B004	B012	B013	B014	B023	B027	B032	B033	B036	B041
1	晚	B002	B010	B017	B019	B024	B026	B029	B030	B034	B038
2	早	B003	B008	B015	B016	B018	B022	B025	B035	B037	B042
2	中	B001	B005	B012	B014	B021	B024	B028	B031	B040	B041

续表

2	晚	B004	B007	B019	B020	B023	B029	B030	B033	B036	B039
3	早	B002	B003	B008	B013	B017	B021	B022	B032	B037	B042
3	中	B005	B009	B010	B011	B018	B025	B026	B027	B035	B038
3	晚	B004	B006	B007	B015	B016	B023	B030	B031	B033	B034
4	早	B001	B014	B015	B016	B018	B022	B027	B035	B036	B040
4	中	B002	B003	B006	B009	B010	B017	B024	B037	B038	B042
4	晚	B008	B011	B012	B013	B019	B020	B021	B028	B029	B039
5	早	B005	B009	B023	B025	B026	B029	B032	B034	B039	B041
5	中	B002	B007	B008	B016	B024	B027	B033	B035	B037	B040
5	晚	B001	B004	B012	B015	B018	B022	B030	B031	B038	B042
6	早	B003	B005	B006	B013	B019	B025	B028	B032	B036	B041
6	中	B008	B009	B010	B011	B014	B017	B020	B021	B026	B034
6	晚	B002	B004	B007	B016	B018	B022	B027	B030	B031	B035
7	早	B006	B011	B012	B015	B023	B024	B029	B034	B036	B037
7	中	B001	B013	B014	B019	B020	B021	B025	B038	B041	B042
7	晚	B003	B005	B010	B017	B026	B028	B032	B033	B039	B040

合理的排班能确保关键岗位始终有合适的员工在岗，降低因人员不足或配置不当导致的生产停滞风险。遗传算法不仅能有效平衡员工工作量，避免某些员工过度劳累，而其他员工则相对轻松的情况，从而提高了员工满意度和工作效率，还有助于降低企业成本，因为合理的排班能减少不必要的加班和人员浪费，从而实现资源的最优配置。

6. 结论与建议

通过对生产线故障数据的深入研究，本文利用 XGBoost 算法构建了一个高效的故障识别与预测模型。该模型不仅具有较高的拟合度和预测准确率，还对生产线 M201 上的故障次数进行了详细分析，并预测了故障持续时间。研究结果表明，生产线每月总故障数为 17,328 次，累计故障时长为 133,595 秒。其中，填装装置故障 4001 最长持续时间为 219 秒，物料检测装置故障 2001 最短持续时间为 16 秒。通过相关性分析，我们发现产品合格率与填装装置故障 4001 的故障次数之间存在显著负相关关系，这表明控制此类故障对于提升产品合格率至关重要。此外，其他故障之间的复杂相关性也被揭示，这些故障多发生在生产线的关键前端位置，企业应加强对这些环节的监测和维护工作，以防止后续工序受到影响，从而严重影响整体的产品质量。在人力资源优化方面，本研究应用遗传算法制定了详细的排班计划。通过迭代优化，得到了一周内 42 名员工的最佳操作人员排班方案。该方案考虑了员工的工作时间和休息安排，确保了排班的均衡性和生产效率的最大化。

建议企业推广智能化排班系统，根据生产需求和员工情况自动调整排班方案，提高工作效率和员工满意度。建立故障监测与预警系统，利用传感器和物联网技术，实时监测生产线上的设备状态和运行数据。建立完善的故障数据库，收集并整理各类故障的发生时间、持续时间、影响范围等信息。同时加强对员工的技能培训，提高他们的故障识别和处理能力。

参考文献

- [1] 梁闯. 汽车生产线转毂测试机构滚动轴承故障诊断与预测方法研究[D]: [博士学位论文]. 沈阳: 沈阳工业大学, 2022.
- [2] 谢威炜, 曹曦, 蒋勉, 陈勇, 黄玮. 基于 BPNN-XGBoost 组合模型的瓦楞纸板线湿部生产速度预测方法[J]. 包装工程, 2024, 45(9): 210-217.
- [3] 刘孝保, 严清秀, 易斌, 姚廷强, 顾文娟. 基于集成学习和改进粒子群优化算法的流程制造工艺参数优化[J]. 中国机械工程, 2023, 34(23): 2842-2853.
- [4] 贺鑫来, 孙庚, 汪敏捷, 等. 基于时序数据的列车牵引系统故障预测方法[J/OL]. 现代电子技术: 1-9.
<https://link.cnki.net/urlid/61.1224.TN.20240620.1437.002>, 2024-10-07.
- [5] 刘亚红. 公交车辆智能排班与时刻表优化方法研究[D]: [博士学位论文]. 北京: 北京邮电大学, 2023.
- [6] 丁海博, 张睿, 崔丽玲. 基于 XGBoost-LightGBM 的保险理赔预测研究[J]. 计算机时代, 2023(5): 61-65.
- [7] 李天举, 谢志峰, 张侃弘, 等. 基于集成学习的烟草异常数据挖掘研究与应用[J]. 计算机技术与发展, 2020, 30(11): 128-135.
- [8] 张明哲. 两种生物进化算法的膜计算模型研究[D]: [硕士学位论文]. 重庆: 重庆大学, 2018.
- [9] 李少波, 宋启松, 李志昂, 张星星, 枘龙炫. 遗传算法在机器人路径规划中的研究综述[J]. 科学技术与工程, 2020, 20(2): 423-431.