

基于多阶段动态规划的电子产品生产过程决策研究

刘盈颖, 叶敏怡*

广州大学数学与信息科学学院, 广东 广州 510006

DOI:10.61369/ASDS.2025040008

摘要 : 为推进电子产品生产过程的绿色低碳发展, 本文提出了基于二项分布正态近似的抽样检测方案, 此方法便于实际操作, 但考虑其缺乏灵活性且对样本的依赖性较强, 本文提出了基于序贯概率比检验的抽样检测, 并给出两种检测方法的优缺点比较。分析四个生产阶段在不同决策下企业可能面临的成本费用, 考虑到不同生产阶段之间相互影响, 选择构建基于动态规划方法的多阶段决策过程最优化模型。利用该模型的马尔可夫性质, 从阶段4到阶段1逐步递推, 分别定义各个阶段的状态转移方程、成本函数以及动态规划方程, 从而得到零配件检测决策、成品检测决策、不合格品的处理决策以及退回的不合格品的处理决策。将 m 道工序、 n 个零配件的生产过程视为一个多阶段的马尔可夫流程, 再采用强化学习算法中的 Q-learning 方法进行建立模型, 得到最优决策。

关键词 : 二项分布正态近似; 序贯检测; 多阶段动态规划; 马尔可夫; Q-learning

Research on Decision-making in the Production Process of Electronic Products Based on Multi-stage Dynamic Programming

Liu Yingying, Ye Minyi*

School of Mathematics and Information Science, Guangzhou University, Guangzhou, Guangdong 510006

Abstract : In order to promote the green and low-carbon development of the production process of electronic products, this paper proposes a sampling detection scheme based on the normal approximation of binomial distribution, which is convenient for practical operation, but considering its lack of flexibility and strong dependence on samples, this paper proposes a sampling detection based on the sequential probability ratio test, and gives a comparison of the advantages and disadvantages of the two detection methods. Considering the mutual influence between different production stages, a multi-stage decision-making process optimization model based on dynamic programming method is selected to be constructed. Based on the Markov properties of the model, the state transition equation, cost function and dynamic programming equation of each stage are defined step by step from stage 4 to stage 1, so as to obtain the spare parts inspection decision, the finished product inspection decision, the processing decision of nonconforming products and the processing decision of returned nonconforming products. The production process of m process and n spare parts is regarded as a multi-stage Markov process, and then the Q-learning method in the reinforcement learning algorithm is used to establish the model to obtain the optimal decision.

Keywords : normal approximation of binomial distribution; sequential detection; multi-stage dynamic programming; Markov; Q-learning

引言

根据国务院于2025年5月23日召开的会议, 中国正在推进制造业的绿色低碳发展, 并审议通过了《制造业绿色低碳发展行动方案(2025 - 2027年)》。会议强调, 推动制造业的绿色转型是大势所趋, 必须加快绿色科技创新和先进绿色技术的推广应用, 以实现传统产业的深度绿色转型。

在制造业中, 电子信息制造业占据收入占比大, 因此电子信息制造业发展中可能造成的资源浪费、环境污染等问题不可忽视, 本文从电子产品生产过程入手, 建立对应数学模型, 既能保证企业利润, 又能在一定程度上减少资源的浪费, 实现绿色制造。

为简化研究, 本文将研究过程主要分为零件检测阶段和后续生产阶段, 其中零件检测部分只假设了零件1和零件2, 后续生产过程具

基金项目: 中华人民共和国教育部2025年广州大学大学生创新训练项目(202511078351)。

作者简介: 刘盈颖(2004-), 女, 广东潮州人, 广州大学数学与信息科学学院本科生在读, 研究方向: 数学与应用数学。

通讯作者: 叶敏怡(2004-), 女, 广东中山人, 广州大学数学与信息科学学院本科生在读, 研究方向: 数学与应用数学。

体划分为4个阶段：

1. 对零配件是否进行检测：

在这一阶段，企业需要决定是否对采购的零配件进行质量检测。如果对零配件进行检测，可以筛选出不合格的零配件并将其丢弃，降低产品不合格的风险，但这会增加生产成本。如果不进行检测，次品零配件将直接进入成品的装配环节，这可能会影响后续成品的质量并增加不合格品退回调换产生损失的风险。

2. 对装配好的每一件成品是否进行检测：

在成品装配完成后，企业可以选择对成品进行质量检测。检测后的合格成品可以直接进入市场销售，而不合格成品则需要进一步处理（如拆解或丢弃）。如果不进行检测，所有成品均直接进入市场，即不合格成品也流入市场，导致售后退回调换以及企业信誉的损失。

3. 对检测出的不合格成品是否进行拆解：

检测出的不合格成品可以选择进行拆解，重复使用其中的零配件。这一过程会产生拆解成本，但部分零配件可以重新利用，降低对新零配件的需求。若选择不拆解，则直接将不合格成品丢弃，已投入的装配和材料成本全部损失。

4. 对用户购买的不合格品的处理：

对于用户退回的不合格的成品，企业将无条件予以调换，但这会产生一定的损失，包括物流费用以及企业信誉等；对于退回的成品，企业可以选择将其拆解并重复前述步骤，或者直接丢弃。

四个阶段的具体流程如图 1 所示：

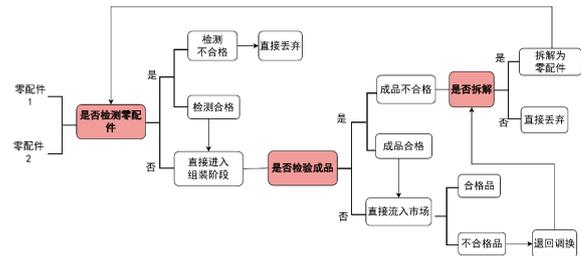


图 1 生产过程的具体流程图

一、模型假设与符号说明

（一）模型假设

1. 在装配过程中零配件无被损耗现象；
2. 企业准备检测的这批零配件样本量足够大；
3. 零件1和零件2相互独立，二者的合格与否互相不产生影响；
4. 企业将无条件予以调换的产品只限于不合格品；
5. 抽样检测过程中每个部件被抽到的概率相等，抽测具有随机性；
6. 合格的零配件在拆解后可以循环足够多次数使用；
7. 对于不合格产品，企业除了无条件予以退还外无其他经济补偿。

（二）符号说明

符号	含义
P	次品率
P_0	标称次品率
P_f	成品次品率
N	生产成品数
P_m	成品市场售价
A	成品的装配成本
T_f	成品检测成本
R	不合格品的拆解成本
D	调换损失（物流成本和企业信誉损失）
C_i	第 i 个零配件的购买单价
T_i	第 i 个零配件的检测费用

P_i	第 i 个零配件的次品率
x_i	第 i 个零配件是否检测的决策变量
x_i	第 i 个半成品是否检测的决策变量
y	成品是否检测的检测变量
zx	不合格成品是否拆解的决策变量
w	退回调换的产品是否拆解的决策变量

二、研究模型的建立

（一）基于二项分布^[1]正态近似的抽样检测方案

设零配件的次品率为 p ，标称次品率为 p_0 ，抽样的样本量为 n ，次品检测量为 x ，决策标准量为 c ，那么依据题目条件有假设检验：

$$\begin{aligned} H_0 : p &\leq p_0 \\ H_1 : p &> p_0 \end{aligned} \quad (1)$$

次品检测量服从二项分布，即：

$$x \sim \text{Binomial}(n, p) \quad (2)$$

在本题中，当样本量 n 较大时，二项分布可以用正态分布来近似，即：

$$x \sim N(np_0, np_0(1-p_0)) \quad (3)$$

标准化，得到标准检验统计量：

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \quad (4)$$

由此可以通过比较 Z 与标准正态分布表中的临界值来决定是

否拒绝 H_0 。

若直接基于置信水平确定样本量，那么对于样本量的计算有：

$$n = \left(\frac{Z_{\alpha/2}}{E} \right)^2 p_0 (1 - p_0) n \quad (5)$$

其中， $Z_{\alpha/2}$ 是显著性水平 α 对应的正态分布临界值， E 是允许的最大误差。

本文假设允许最大误差为 0.05，次品率 $p_0 = 0.1$ 。

1. 在 95% 置信水平下的检测样本量

此时 $\alpha = 0.05$ ，因此 $Z_{0.025} \approx 1.96$ ，代入公式取整可得 $n = 139$ 。为确定决策次品数阈值 c ，可以使用精确二项分布计算得到，即有不等式：

$$P(X \geq c | p = 0.1) \leq 0.05 \quad (6)$$

2. 在 90% 置信水平下的检测样本量

此时 $\alpha = 0.1$ ，因此 $Z_{0.05} \approx 1.64$ ，代入公式整可得 $n = 98$ 。同样地，对于决策次品数阈值，有不等式：

$$P(X \geq c | p = 0.1) \leq 0.1 \quad (7)$$

(二) 基于序贯概率比检验的抽样检测^[2]

1. 拒收检测方案

针对情形一，标称次品率为 $p_0 = 0.1$ ，需要在 95% 的置信度下判断是否拒收。序贯检验步骤如下：

设定检验参数：

显著性水平 $\alpha = 0.05$ ；检验功效 $\beta = 0.10$ 。

计算序贯检验^[3]的两个阈值：

上限阈值：

$$A_1 = \frac{1 - \beta}{\alpha} = \frac{1 - 0.10}{0.05} = 18 \quad (8)$$

下限阈值：

$$B_1 = \frac{\beta}{1 - \alpha} = \frac{0.10}{1 - 0.05} \approx 0.105 \quad (9)$$

构造序贯检验的似然比：

设 $p_1 = 15\%$ ， p_1 是设定的较高的次品率，用于跟 p_0 进行比较。

由此可以计算似然比函数为：

$$\Lambda_1 = \left(\frac{p_1}{p_0} \right)^k \times \left(\frac{1 - p_1}{1 - p_0} \right)^{n - k} \quad (10)$$

2. 接收检测方案

针对情形二，标称次品率为 $p_0 = 0.1$ ，需要在 90% 的置信度下判断是否接收。序贯检验步骤如下：

设定检验参数：

显著性水平 $\alpha = 0.10$ ；检验功效 $\beta = 0.05$ 。

计算序贯检验的两个阈值：

上限阈值：

$$A_2 = \frac{1 - \beta}{\alpha} = \frac{1 - 0.05}{0.10} = 9.5 \quad (11)$$

下限阈值：

$$B_2 = \frac{\beta}{1 - \alpha} = \frac{0.05}{1 - 0.10} \approx 0.0556 \quad (12)$$

构造序贯检验的似然比：

设 $p_2 = 5\%$ ， p_2 是设定的较低的次品率，用于跟 p_0 进行比较。

因此可以计算似然比函数为：

$$\Lambda_2 = \left(\frac{p_2}{p_0} \right)^k \times \left(\frac{1 - p_2}{1 - p_0} \right)^{n - k} \quad (13)$$

其中 k 是检测中发现的不合格零配件数； n 是已检测的总样本数。

依据建立的抽样检测方案，在实际生产过程中模型流程的实施步骤如下：

① 从零配件中逐个抽样，每次检验一个零配件，记录数据；

② 在每次抽样后，更新似然比的值；

③ 根据决策规则做出决策。

具体流程图如图 2 所示：

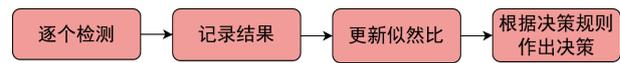


图 2 实际操作流程

(三) 基于动态规划^[4]方法的多阶段决策过程最优化模型

为求生产过程最佳决策路径，本文使用动态规划模型解决该多阶段决策问题。多阶段决策过程最优化的目标是达到整个生产过程的总体效果最优。由于各阶段决策间存在有机的联系，每个阶段决策的执行都将影响到下一阶段的决策，从而对总体效果产生影响。因此在做出短阶段决策时不应仅考虑本阶段最优，还应考虑其对最终目标的影响，选择对整个过程而言最优的决策。

在解决多阶段决策^[5]动态规划模型时，通常由后往前逐步计算，这是利用了模型的马尔可夫性质：不管前面阶段的状态和决策如何，对前面阶段的决策所形成的状态来说后部分过程的策略必须构成最优策略，因此，可以将多阶段决策问题的求解过程看成是一个连续的递推过程。

在每个阶段中，通过选择最优的检测、装配、拆解或丢弃策略，使得动态规划递推的解能够最小化企业的总成本或最大化企业的收益。最终，依据各阶段的最优决策路径，企业可以制定零配件检测、成品检测及不合格品处理的综合策略，从而实现成本优化。

对于每一个生产阶段，可以描述其动态规划基本方程：

阶段 4 退回不合格品的处理决策

在阶段 4 中，企业无条件予以调换不合格产品，该过程中会产生退回调换损失（包括物流成本以及企业信誉损失等），针对退回的不合格产品，企业可选择将其直接丢弃，也可以选择拆解成品，使零部件重新回到阶段 1。

该阶段的状态转移方程为：

$$f_4(w) = \begin{cases} -D \times p_f \times N & w=0 \\ -(D+R) \times p_f \times N & w=1 \end{cases} \quad (14)$$

成本函数为：

$$C_4(w) = \begin{cases} D \times p_f \times N & w=0 \\ (D+R) \times p_f \times N & w=1 \end{cases} \quad (15)$$

动态规划方程为：

$$V_4(w) = -C_4(w) \quad (16)$$

阶段3 不合格成品的处理决策

在阶段3中, 成品装配后, 若成品被检测为不合格, 企业需决定是否进行拆解。如果拆解, 会产生一定的拆解成本, 拆解后的零配件将返回阶段1; 如果不拆解, 企业将产生丢弃成本, 直接造成损失。

该阶段的状态转移方程为:

$$f_3(z) = \begin{cases} -p_f \times (A+T_f) \times N + (1-p_f) \times P_m \times N + f_3(0) & z=0 \\ -R \times N + f_3(1) & z=1 \end{cases} \quad (17)$$

成本函数为:

$$C_3(z) = \begin{cases} p_f \times (A+T_f) \times N & z=0 \\ R \times N & z=1 \end{cases} \quad (18)$$

动态规划方程为:

$$V_3(z) = -C_3(z) + \max_w V_4(w) \quad (19)$$

阶段2 成品的装配与检测决策

在阶段2中, 企业需决定是否对装配完成的成品进行检测。如果检测, 需要支付一定的检测费用, 如果不检测, 不合格成品直接流入市场销售会导致退回调换损失, 退回的不合格成品将进入阶段3进行处理。

该阶段的状态转移方程为:

$$f_2(y) = \begin{cases} -p_f \times D \times N + (1-p_f) \times P_m \times N + f_2(0) & y=0 \\ -T_f \times N + (1-p_f) \times P_m \times N + f_2(1) & y=1 \end{cases} \quad (20)$$

成本函数为:

$$C_2(y) = \begin{cases} p_f \times D \times N & y=0 \\ T_f \times N & y=1 \end{cases} \quad (21)$$

动态规划方程为:

$$V_2(y) = -C_2(y) + \max_z V_3(z) \quad (22)$$

阶段1 零配件的采购与检测决策

该阶段作为整个生产过程的起点, 如果未检测零配件, 使得不合格的零配件进入组装阶段, 后续的成品质量会受影响。由于检测需要一定的成本, 因此企业需要决定是否对采购的零配件进行检测。

该阶段的状态转移方程为:

$$f_1(x_1, x_2) = \begin{cases} -\sum_{i=1}^2 T_i \times N + \sum_{i=1}^2 c_i \times N + f_1(1) & (x_1=1, x_2=1) \\ -(T_1 \times N + \sum_{i=1}^2 c_i \times N + P_2 \times D \times N) + f_1(1) & (x_1=1, x_2=0) \\ -\sum_{i=1}^2 c_i \times N + T_2 \times N + P_1 \times D \times N + f_1(1) & (x_1=0, x_2=1) \\ -\sum_{i=1}^2 c_i \times N + \sum_{i=1}^2 p_i \times D \times N + f_1(1) & (x_1=0, x_2=0) \end{cases} \quad (23)$$

成本函数为:

$$C_1(x_1, x_2) = \begin{cases} \sum_{i=1}^2 T_i \times N + \sum_{i=1}^2 c_i \times N & (x_1=1, x_2=1) \\ T_1 \times N + \sum_{i=1}^2 c_i \times N + P_2 \times D \times N & (x_1=1, x_2=0) \\ \sum_{i=1}^2 c_i \times N + T_2 \times N + P_1 \times D \times N & (x_1=0, x_2=1) \\ \sum_{i=1}^2 c_i \times N + \sum_{i=1}^2 p_i \times D \times N & (x_1=0, x_2=0) \end{cases} \quad (24)$$

动态规划方程为:

$$V_1(x_1, x_2) = -C_1(x_1, x_2) + \max_y V_2(y) \quad (25)$$

(四) 基于 Q-learning 算法^[6]的生产过程决策方案

对于 m 道工序、n 个零配件的生产过程, 本文假设中间每一道工序生产出若干件半成品, 最后生产出一种成品。

这一生产决策问题可以视为一个多阶段、多决策的马尔可夫决策过程, 针对此类问题, 可以使用强化学习算法中的 Q-learning 方法解决。

Q-learning 是一种强化学习方法, 它允许智能体学习在一个环境中采取何种行动可以获得最大的奖励。该算法试图学习一个 Q 表, 该表记录了每个状态-动作对的 Q 值。算法通过迭代更新 Q 表, 使得对于每一个状态, 智能体能够选择具有最高预期奖励的动作。

Q-learning 算法模型^[7]建立如下:

1. 状态空间

状态 S 可以表示为当前生产阶段、零配件的状态、半成品的状态和成品的状态。状态可以表示为一个向量:

$$S = (s_1, s_2, s_3, \dots, s_n, s'_1, s'_2, \dots, s'_m, s_{fin}) \quad (26)$$

其中, s_{fin} 表示成品的检测决策。 s'_j 表示第 j 个半成品的检测决策; s_{fin} 表示成品的检测决策。以上决策变量取 0 时表示未检测, 取 1 时表示检测合格, 取 2 时表示检测不合格。

2. 动作空间

动作 A 可以表示对当前生产阶段的决策。可以表示为一个向量:

$$A = (a_1, a_2, \dots, a_n, a'_1, a'_2, \dots, a'_m, a_{fin}) \quad (27)$$

其中, a_i 表示第 i 个零配件的决策; a'_j 表示第 j 个半成品的决策; a_{fin} 表示成品的决策。 a_i 取 0 时表示不检测, 1 表示检测, 2 表示检测后不合格丢弃; a'_m 和 a'_j 取 0 时表示不检测, 取 1 时表示检测, 取 2 时表示检测后不合格拆解, 3 表示检测后不合格丢弃。

3. 奖励

奖励 R 依据决策的成本和收益来定义, 表示为:

$$R = P_m - \sum_{i=1}^n [c_i(1+p_i) + a_i T_i] - \sum_{j=1}^m [A_j + a'_j T_j + p_j(A_j + D_j)] - [A_f + a_f T_f + p_f(A_f + D_f)] - L_r \quad (28)$$

其中, P_m 表示成品的市场价格, C_i 表示第 i 个零配件的成本, T_i 表示第 i 个零配件的检测成本, A_j 表示第 j 个半成品的组装成本, T_j 表示第 j 个半成品的检测成本, D_j 表示第 j 个半成品的拆解成本, A_{fin} 表示成品的组装成本, T_{fin} 表示成品的检测成本, D_{fin} 表示成品的拆解成本, L_r 表示成品的替换损失。

奖励是成品的市场价格减去所有相关的成本, 包括零配件的成本、检测成本、组装成本、拆解成本和替换损失。

4. Q 值更新

Q-Learning 的核心是 Q 值的更新^[8]。Q 值表示在状态 S 下采取动作 A 的预期累积奖励。其更新规则为:

$$Q(S, A) \leftarrow Q(S, A) + \alpha [R + \gamma \max_{A'} Q(S', A') - Q(S, A)] \quad (29)$$

其中 α 为学习率, γ 是折扣因子, S' 是下一个状态, A' 是下一个动作。

通过对 Q-learning 算法模型^[9]的建立, 可以依据实际生产流

程情况对生产过程给出最优决策方案。

将上述模型应用于两道工序、八个零配件的情境中，可得具体的生产流程如图 3 所示：

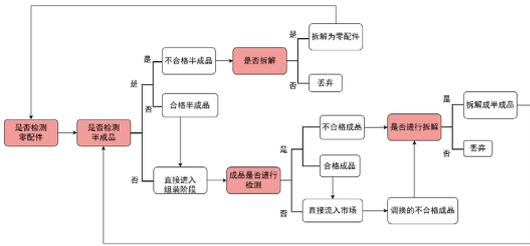


图 3 生产流程图

三、实验结果与分析

(一) 零件检测阶段的决策求解结果

1. 基于正态分布的抽样检测方案

①在 95% 的置信水平下，检测 139 个零配件，若检测到 20 个次品，则拒收这批零配件；

②在 90% 的置信水平下，检测 98 个零配件，若最多检测到 14 个次品，则接收这批零配件。

2. 基于序贯概率比检验的抽样检测规则

(1) 拒收方案

①如果 $\Lambda_1 > 18$ ，则认为次品率超过标称值，拒收这批零配件；

②如果 $\Lambda_1 < 0.105$ ，则认为次品率不超过标称值，接收这批零配件；

③如果 $0.105 \leq \Lambda_1 \leq 18$ ，则继续检测下一个零配件。

(2) 接收方案

①如果 $\Lambda_2 > 9.5$ ，则认为次品率超过标称值，不接受这批零配件；

②如果 $\Lambda_2 < 0.0556$ ，则认为次品率不超过标称值，接收这批零配件；

③如果 $0.0556 \leq \Lambda_2 \leq 9.5$ ，则继续检测下一个零配件。

为更好说明抽样检测方案，本文随机生成次品样本进行模拟。在每一步计算似然比。若似然比超过上限阈值，则拒收该批次产品；若低于下限阈值，则接收该批次产品；若似然比位于上下限之间，则继续检测。

某次随机模拟的结果如图 4、图 5 所示：

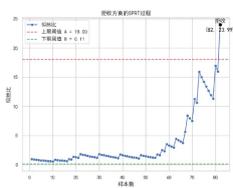


图 4 拒收方案的 SPRT 过程

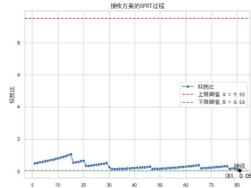


图 5 接收方案的 SPRT 过程

图 4 和图 5 分别展示了拒收和接收检测方案中的序贯概率比检验 (SPRT) 过程。图中的蓝色线条代表似然比随检测样本数量的变化过程，红色虚线为拒收阈值，绿色虚线为接收阈值。在每一步检测中，似然比逐渐逼近上限或下限，从而得出最终的决策。

3. 两种检测方法的比较

基于二项分布正态近似的抽样检测给出了确切的检测样本量以及最大可接受的次品量，在实际生产过程中更易理解且实施难度较小，但是此类方法通常需要在检测前确定一个固定的样本量，缺乏灵活性，一旦数据分布情况与预期存在较大差异，可能会导致资源浪费或结论不准确。

基于序贯概率比检验^[10]的抽样检测并未给出确切的检测样本量，而是给出了不同样本数量下的似然比，在实际操作中对检验人员的专业素养要求较高，操作过程较为繁琐。但是由于未固定确切检测值，而是在检测过程中动态调整检验策略，具有更强的灵活性，也使得检测的可靠性更强，同时可以避免不必要的检测，节省成本。

(二) 基于动态规划的生产过程决策

针对零部件 1、2，成品和不合格成品，本文设定了以下六种情况，具体数据如表 1 所示：

表 1 六类假设情况

情况	零配件 1			零配件 2			成品			不合格成品		
	次品率	购买单价	检测成本	次品率	购买单价	检测成本	次品率	装配成本	检测成本	市场售价	调换损失	拆解费用
1	10%	4	2	10%	18	3	10%	6	3	56	6	5
2	20%	4	2	20%	18	3	20%	6	3	56	6	5
3	10%	4	2	10%	18	3	10%	6	3	56	30	5
4	20%	4	1	20%	18	1	20%	6	2	56	30	5
5	10%	4	8	20%	18	1	10%	6	2	56	10	5
6	5%	4	2	5%	18	3	5%	6	3	56	10	40

假定最后成品共有 100 件，按照上述设定的六类情况，可以计算出各自对应的最优效益，具体决策结果如表 2 所示：

表 2 各情况的决策情况

情况	最大平均盈利 (元)	零件 1 是否检测	零件 2 是否检测	成品 是否检测	不合格成品 是否拆解
1	25.34	是	否	否	否
2	24.5	是	否	否	是
3	25.34	是	否	否	否
4	25.76	是	否	否	否
5	25.23	否	否	否	否
6	26.6	否	否	否	是

由上述结果可发现，仅检验零配件 1 且不拆解这一决策为选择最多的最优决策。由于零配件 1 的检测成本较低，可以提前识别出一些次品，从而避免将次品直接装配成成品，降低次品率。选择该决策大部分原因是因为拆解费用较高且成品合格率较高，检测零配件 1 后，装配出的成品大概率是合格的，则拆解并没有带来额外的收益反而增加了成本。因此，检测零配件 1 可以有效控制成本，而不进行拆解进一步减少了额外的处理费用，从而在此策略下实现单件利润最大化。

不同的决策策略都是基于具体情况的次品率、检测成本、装

配成本和拆解费用进行的优化选择。关键在于在检测和拆解之间找到一个平衡,使得合格率提升的同时,整体成本得到控制,从而实现单件产品的利润最大化。

(三) 基于 Q-learning 的生产决策

假定工序 $m=2$ 、零配件个数 $n=8$, 可以从中抽取数据并投入到 Q-learning 算法的训练之中。得到最优生产决策方案如表 2 所示:

表 2 最优生产决策

8个零配件	3个半成品	成品	退回的不合格品
不检测	不检测	检测并拆解次品	不拆解

四、模型评价

(一) 模型的优点

1. 模型的高效性: 基于 SPRT 的抽样检测方案确保在较少的样本量下能够尽快做出可靠的拒收或接收决策, 不仅降低了检测成本, 还保证了较高的检验效率, 具有高效性。

2. 模型的灵活性: SPRT 允许在数据的采集过程中动态地进行假设检验, 而不需要预先确定固定样本量, 根据样本结果计算似然比动态地调整抽样次数, 具有较高的灵活性。

3. 模型的可行性: 通过动态规划算法采取清晰的递归结构并写出状态转移方程, 将复杂问题分解为一系列相对简单的子问题, 并通过解决这些子问题来找到原问题的解, 简化了复杂循环过程, 使得整个求解过程更具可行性。

4. 模型的普适性: 在制定 m 道工序, n 个零件决策方案模型时不确定具体生产流程如零件和半成品具体组装情况等, 但 Q-learning 算法能够通过与环境交互来学习适应这种不确定性, 并找到最优的策略。

(二) 模型的缺点

问题二采用动态规划进行求解需要遍历 16 种可能状态找到最优解, 倘若生产流程更加复杂、零配件数量增加, 都有可能导致计算量急剧增加, 算法运行时间非常长且状态空间爆炸。

五、结论

本文针对电子产品生产过程的决策问题, 从零件的采购检测、成品的检测、不合格品的拆解以及退回不合格品的处理四个阶段入手, 建立了基于动态规划的电子产品生产过程的决策模型, 在降低企业成本的同时也降低电子产品生产过程中造成的资源浪费, 符合绿色发展理念。主要成果如下:

(一) 序贯检验在电子产品零件检测中的应用

(二) 基于动态规划的多阶段决策过程最优化模型

(三) 基于 Q-learning 算法的生产过程决策模型

本研究为步骤相对简单的电子产品生产决策提供了理论支持, 未来可进一步结合更加复杂的生产过程, 完善模型, 使得该模型适用于多样化的生产领域。

参考文献

- [1] 张雪媛. 二项分布及其应用的历史研究 [D]. 天津财经大学, 2015.
- [2] 崔岩, 崔迪. 计数序贯抽样检验方法在元器件验收中的应用 [J]. 长春光学精密机械学院学报, 1996, (04):68-72.
- [3] 于超. 基于统计过程控制和序贯概率比检验的高速铁路轨道结构监测数据异常诊断 [D]. 北京交通大学, 2021.
- [4] 宋辉, 崔梦天. 动态规划在企业生产与储存管理中的应用 [J]. 科技广场, 2006, (6):114-116.
- [5] 沈建新, 周儒荣, 廖文和. 动态规划模型在生产库存优化中的应用研究 [J]. 机械设计与制造工程, 2002, 31(6):112-113.
- [6] 张峰, 刘凌云, 郭欣欣. 基于改进 Q-学习算法的多阶段群体决策模型 [J]. 控制与决策, 2019, 34(09):1917-1922.
- [7] Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., & Mordatch, I. Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. Advances in Neural Information Processing Systems, 2017, 30: 6379-6390.
- [8] 刘波. 一种基于强化学习的机械车间任务调度节能优化支持系统研究 [D]. 重庆大学, 2013.
- [9] 星海. 面向复杂批次过程的强化学习控制方案 [D]. 厦门大学, 2021.
- [10] Wald, A. (1945) Sequential Tests of Statistical Hypotheses. The Annals of Mathematical Statistics, 16, 117-186.