



中国研究生创新实践系列大赛
“华为杯”第二十届中国研究生
数学建模竞赛

学 校

上海海事大学

参赛队号

23102540029

队员姓名

1. 刘 艺
2. 黎禧妮
3. 陈振华

中国研究生创新实践系列大赛
“华为杯”第二十届中国研究生
数学建模竞赛

题 目： 大规模创新类竞赛评审方案研究

摘 要：

当前大规模创新类竞赛日益增多，通常采用两到三阶段评审，但由于没有标准答案，不同评委对同一作品的评分极有可能存在显著差异。对于这一情况，尽管已有多种评审方案，但它们对大规模创新类竞赛评审的适用性依然有限。因此，本研究重点在于针对大规模创新性竞赛，为其建立更为合理和公平的评审方案。

针对问题一，本问题旨在确保评审方案的交叉性，我们建立了一个数学规划模型。在模型中，将专家是否评审该作品作为 0-1 变量，在满足评审规定条件的约束下，我们旨在最大化评审专家对作品的交集数量，同时保证交集数量尽可能地均衡。为了衡量这种均衡性，我们引入了均衡性指标和交叉值指标。采用 NSGA-II 算法进行求解。最终得到的交叉量为 30000，标准差为 1.6258。每位专家所评审的作品数量和作品集合，详见表 4-1。同时，我们还统计了 125 位评审专家的交叉评审作品数量，详见图 4-8。

针对问题二，我们首先深入分析了数据的原始成绩分布特征。通过仔细考虑每位专家的评分习惯和分布特点以及作品的得分特点，设计了两种评审方案——专家角度评审和综合评审，主要评价指标包括分析专家评分标准差 SD_j 、评审作品数量、极评指标 ESI 、作品标准差、极差值、复议情况等。进一步地，我们确定了两个关键评价指标，即专家专业性评分(ER)、作品得分合理性评分(SRA)。然后，我们通过线性回归方法，建立了一个初步的标准分计算模型。再根据第二次评审阶段修正得出最终的标准分计算模型（见下式），根据最终模型，我们对调整后的成绩进行了进一步的特征分析，并对比了获奖等级。

$$STD_{ij} = 0.765625 * ER_{ij} - 0.291511 * SRA_i + 3.76465 * std_{ij}$$

针对问题三，通过分析两阶段成绩和极差，以及比较两阶段与不分阶段的评审方案，我们在问题二确定的专家专业性指标以及作品得分均值和标准差的基础上，增加了**创新性评价指标**，并将所有指标作为线性回归、决策树、随机森林和 XGBoost 的输入指标，经过回归指标的评估，最终选择了**随机森林**作为极差的拟合模型。接着，我们根据第二阶段专家评审调整“极差”模型，并探讨了**极差与创新性指标的相关性**（见下表）。

	阈值	关系式
数据 2.1	> 270.24	成绩=1.38*创新性评价指标+0.35*极差+266
	≤ 270.24	成绩=-0.45*创新性评价指标+1.50*极差+219.69
数据 2.2	> 249.32	成绩=-5.13*创新性评价指标+1.70*极差+250.56
	≤ 249.32	成绩=-3.37*创新性评价指标-0.90*极差+218.75

针对问题四，本问题致力于创新类竞赛构建一个全面且公正的评审模型，确保所有参赛作品都能获得公正且合理的评估。评审模型包括**两个关键阶段**：在第一阶段，我们结合问题一的模型和问题二的相关指标筛选出有资格进入第二评审阶段的参赛作品。在第二阶段中，我们以问题二得出的极评指标（偏差）最大化作为目标函数，以专家专业性评分作为**约束条件**，建立第二阶段的优化模型。为了在有限的时间内找到接近最优的评审分配方案，我们采用了**遗传算法**，得到**极评指标和为 2654.73**。最终，我们使用问题二中得出的标准分计算模型来确定参赛作品的最终排名。

关键词：NSGA-II；线性回归；随机森林；两阶段评审模型；评价指标

目 录

一. 问题重述	6
1.1 问题背景	6
1.2 问题提出	6
二. 总技术路线图	7
三. 符号说明	8
四. 问题一：基于 NSGA-II 的“交叉分发”方案研究	10
4.1 问题描述与分析	10
4.2 模型建立	10
4.3 模型评价指标	11
4.4 模型求解	12
4.4.1 算法设计	12
4.4.2 算法步骤和流程	14
4.5 求解结果与分析	16
五. 问题二：评审方案设计与标准分改进	20
5.1 问题描述与分析	20
5.2 数据预处理	20
5.2.1 数据转换和整合	20
5.2.2 原始成绩分布特征分析	20
5.2.3 评价指标确立	23
5.2.4 评分公式计算	26
5.3 基于线性回归的标准分计算模型	26
5.3.1 回归模型介绍	26
5.3.2 回归模型构建	27
5.3.3 模型结果评估	29
5.3.4 标准分计算模型改进	30

5.3.5 调整后成绩分布特征	32
5.3.6 获奖等级对比分析	33
六. 问题三：基于随机森林的“极差”模型.....	35
6.1 问题描述与分析	35
6.2 数据预处理	35
6.3 数据探索与分析	36
6.3.1 两阶段评审方案对比	36
6.3.2 评价指标确立	38
6.4 建立极差模型	39
6.4.1 模型介绍	39
6.4.2 模型训练与评估	40
6.5 极差调整与分析	42
6.6 极差与创新性相关性分析	43
七. 问题四：评审模型建立与求解.....	45
7.1 问题描述与分析	45
7.2 两阶段模型建立	45
7.3 模型求解	46
7.3.1 算法设计	46
7.3.2 算法步骤	48
7.4 求解结果与分析	49
7.4.1 实验设计	49
7.4.2 求解结果	50
八. 模型评价与改进.....	51
8.1 模型和算法的优点	51
8.2 模型改进和展望	51
参考文献	52

一. 问题重述

1.1 问题背景

创新类竞赛的数量正日益增多，但由于创新类竞赛没有标准答案，需要评审专家根据评审框架进行独立评审。因此，同一份作品可能会得到不同的评分。尤其在大规模竞赛中，由于评审专家数量众多，评分的极差问题变得更为突出。单纯依赖评分总和进行排序显然不是一个理想的评审方法。因此，探讨创新类竞赛评审方案的公正性科学性变得尤为重要。

为了进一步探索合理的创新类竞赛评审方法，本研究需通过给定的模拟的数据集，利用数据集建立数学模型，以探讨更为公平和合理的评审方案。问题涉及以下几个关键方面：

1. “交叉分发”方案：需要建立有效的数学规划模型，尽可能保证分发方案的均衡性、交叉性、公正性、覆盖率和分散性，可能涉及到数学建模、运筹学方法和智能算法的应用。

2. 评审方案设计与标准分改进：解决方法要考虑设计多个评审方案，重新设计标准分计算，整合评审数据，验证模型，提高透明度和提供反馈，确保公正和准确评审。

3. 处理和调整极差：考虑“创新类”大赛的极差问题。由于作品创新性强，评审意见可能差异大，导致同一作品的评分存在极差。问题要求根据提供的数据，分析两阶段评审方案的优劣，并建立一个处理和调整极差的模型，以实现更公平、准确的评审。

4. 设计完整评审模型：建立优化模型，并针对所给的数据研究进行求解，并对现行的评审方案给出改进的具体建议。

1.2 问题提出

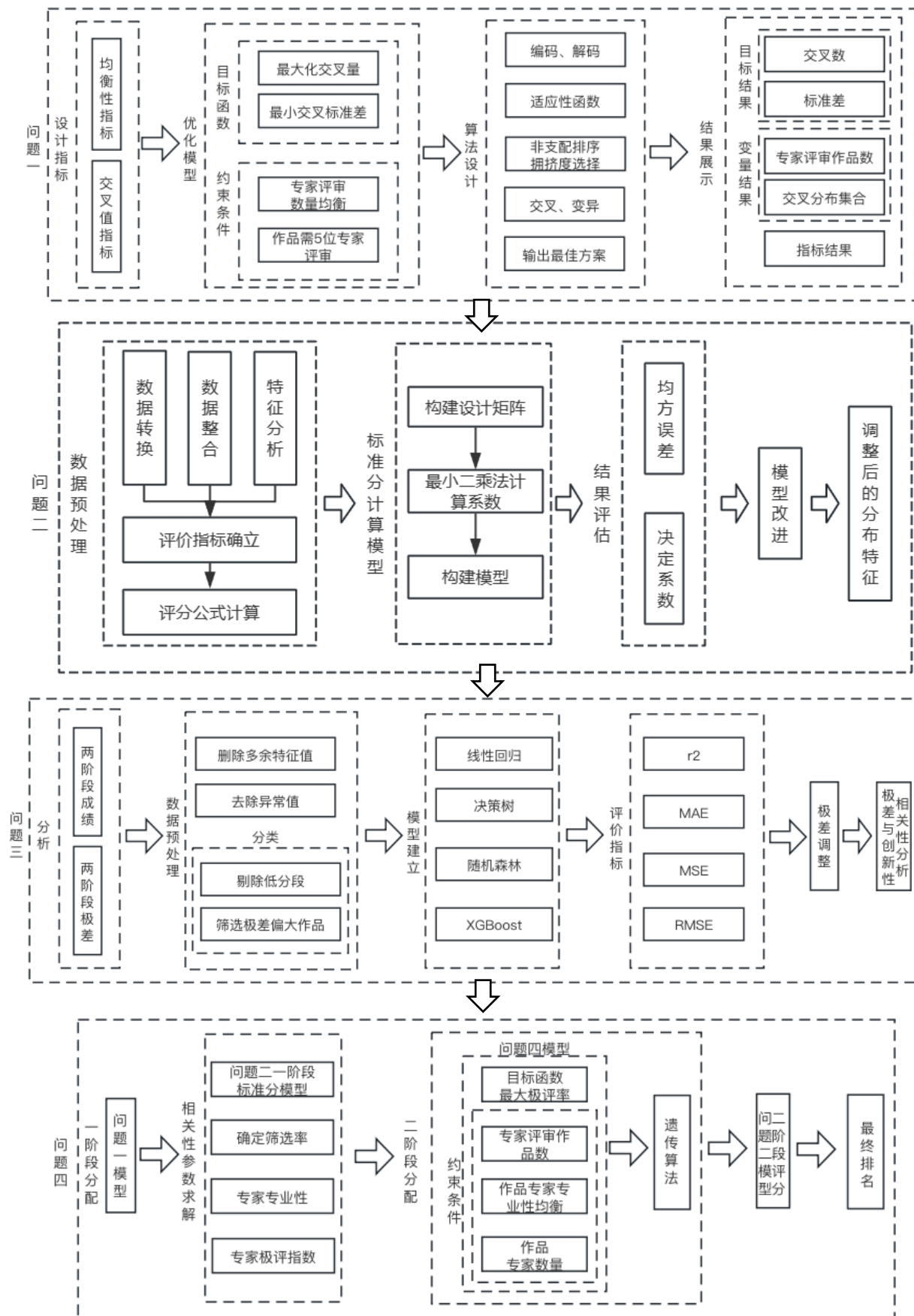
问题 1：优化“交叉分发”方案。建立数学规划模型，根据评审规则，制定分发方案，在满足约束条件下，最大化评审专家之间作品的交集数量以及确保交叉数量尽可能均衡，讨论相关指标，确保评审方案的均衡性、交叉性、公正性、覆盖率和分散性。

问题 2：设计评审方案并改进标准分计算模型。根据数据 1 的数据分析原始成绩特征，确定评价指标，设计新的评审方案，建立线性回归模型，改进标准分计算。

问题 3：通过分析模拟数据建立极差模型，可以采用随机森林等机器学习方法对非高分和非低分段作品的“大极差”进行精确自动化调整。有效地发掘评估高创新性的作品。

问题 4：利用问题一模型获取初始方案与评分，使用问题二模型确定标准分，计算最终成绩及排名。超过阈值的作品进入第二阶段，应用问题二模型优化评审分配并得出最终得分。本问题聚焦第二阶段模型建立。

二. 总技术路线图



三. 符号说明

索引

i 作品的索引

j 评审专家的索引

集合

I 作品集合, $i \in I = \{1, 2, 3, \dots, I_{max}\}$

J 评审专家集合, $j \in J = \{1, 2, 3, \dots, J_{max}\}$

E_j 评审作品 j 的专家集合

参数

L 每个专家评审的最小的作品数

U 每个专家评审的最大的作品数

m 作品数量

n 评审专家的数量

N_i 第 i 个作品的评审专家数量

N_j 专家 j 评审的作品总数

变量

x_{ij} 第 i 个作品被评审专家 j 评审为 1, 否则为 0。

OR_{j_1, j_2} 对于任意两位专家 j_1 和 j_2 , 它们的交集值。

评价指标

RWB 评审工作量均衡性 (Reviewer Workload Balance)

MEB 作品曝光均衡性 (Manuscript Exposure Balance)

Max_OR 最大交叉值 (Maximum Overlap Rate)

Min_OR 最小交叉值 (Minimum Overlap Rate)

AOR 平均交叉率 (Average Overlap Rate)

RDM 评审多样性 (Reviewer Diversity Metric)

SD_j 专家 j 评分标准差

ESI 极评指标

ER 专家专业性评分

AER 平均专家专业性评分

SRA	作品得分合理性评分
IG	综合评审得分
std_{ij}	专家 j 评价作品 i 的评分专业性得分
STD_{ij}	改进后的专家 j 评价作品 i 的评分专业性得分
FSE	各个参赛作品的最终得分
I	创新性评价指标

四. 问题一：基于 NSGA-II 的“交叉分发”方案研究

4.1 问题描述与分析

问题一中提出在评审过程中作品分配给不同评审专家时应会存在某种重叠部分，这种重叠的程度对作品的可比性起到关键作用。如果部分评审专家之间的重叠部分过于显著，那么与其他评审专家之间的重叠部分则可能偏少，从而使得作品的可比性受损。因此，问题针对 m 组作品以及 n 名评审专家（每个专家评审作品的份数为 $5m/n$ ），我们制定一个合理的“交叉分发”策略，建立规划模型，在满足约束 1-2 的条件下，确定实施细节，最大化评审专家对作品的交集数量以及确保交叉数量尽可能均衡，来保证评审的整体合理性。

约束 1：每份作品恰好被 5 位评审专家评审（题目要求）；

约束 2：每位评审专家评审的作品数量应当是均匀的。

对于该“交叉分发”方案，我们规定它需要满足的性质：

①**均衡性**：评审工作量均衡和作品曝光均衡，确保每位评审员都有相似的工作量以及确保每个作品都被相同数量的评审员评审；

②**交叉性**：评审员之间评审的作品的交集确保评审可比性；

③**公正性**：确保每位评审员应该看到尽可能多的不同作品，而不是总是评审相同的作品集。

④**覆盖率**：确保所有作品都至少被一个评审员评审；

⑤**分散性**：确保评审员之间不总是评审相同的作品。

4.2 模型建立

（1）决策变量

第 i 个作品被评审专家 j 评审，则 $x_{ij} = 1$ ，否则为 0。

$$x_{ij} \in \{0,1\}, \quad \forall i,j$$

（2）目标函数

我们的目标是最大化整体的交叉率，目标函数的意图是最大化所有可能的评审专家对的交集以及交叉数量尽可能均衡，具体目标函数如下式：

$$\text{Max } Z = \sum_{j_1, j_2}^n OR_{j_1, j_2}, j_1 \neq j_2 \quad (4-1)$$

$$\text{Min} \sum_{j_1, j_2}^n (OR_{j_1, j_2} - OR_{j_3, j_4})^2, j_1 \neq j_2 \quad (4-2)$$

(3) 约束条件

约束(4-2)表示每份作品恰好被 5 位评审专家评审:

$$\sum_{i=1}^m x_{ij} = 5, \forall i \in \{1, 2, \dots, m\} \quad (4-3)$$

约束(4-3)每位评审专家评审的作品数量应当是均匀的。设定每个评审专家评审的最小和最大作品数为 L 和 U ，其中， L 和 U 根据实际需求来确定， $U = L + a$ ，其中 a 是一个较小的整数，例如 5，以考虑到某些专家可能评审稍微多一些的作品，则:

$$L \leq \sum_{i=1}^n x_{ij} \leq U, \forall j \in \{1, 2, \dots, n\} \quad (4-4)$$

约束(4-5)表示对于任意两位专家 j_1 和 j_2 ，它们的交集值为它们评审的作品的交集数量。

$$OR_{j_1, j_2} = \sum_{i=1}^m x_{ij_1} * x_{ij_2}, j_1 \neq j_2 \quad (4-5)$$

综上所述，问题一建立“交叉分发”策略的规划模型如下:

$$\begin{aligned} \text{Max} \quad & Z = \sum_{j_1, j_2}^n OR_{j_1, j_2} \\ \text{Min} \quad & \sum_{j_1, j_2}^n (OR_{j_1, j_2} - OR_{j_3, j_4})^2 \\ \text{s.t.} \quad & \sum_{i=1}^m x_{ij} = 5, \forall i \in \{1, 2, \dots, m\} \\ & L \leq \sum_{i=1}^n x_{ij} \leq U, \forall j \in \{1, 2, \dots, n\} \end{aligned}$$

4.3 模型评价指标

以确保“交叉分发”方案的公平性和均衡性。以下是考虑这些因素后的评价指标:

1. 均衡性指标

①评审工作量均衡性*Reviewer Workload Balance (RWB)*: 所有评审员评审的作品数量的标准差。数值越小，均衡性越好。确保每位评审员都有相似的工作量。

$$RWB = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^m x_{ij} - \frac{5m}{n} \right)^2} \quad (4-6)$$

②作品曝光均衡性*Manuscript Exposure Balance (MEB)*: 所有作品被评审的次数的标准差。数值越小，均衡性越好。确保每个作品都被相同数量的评审员评审。

$$MEB = \sqrt{\frac{1}{m} \sum_{j=1}^m \left(\sum_{i=1}^n x_{ij} - 5 \right)^2} \quad (4-7)$$

2.交叉值指标

对于任意两位专家 j_1 和 j_2 ，定义它们的交集值为它们评审的作品的交集数量。

$$OR_{j_1, j_2} = \sum_{i=1}^m x_{ij_1} * x_{ij_2}, j_1 \neq j_2 \quad (4-8)$$

最大交叉值 $Maximun\ Overlap\ Rate\ (Max_OR)$:

$$\begin{aligned} Max_OR \\ = \max (OR_{1,2}, OR_{2,3}, \dots, OR_{j_1, j_2}, \dots) \end{aligned} \quad (4-9)$$

最小交叉值 $Minimum\ Overlap\ Rate\ (Min_OR)$:

$$Min_OR = \min (OR_{1,2}, OR_{2,3}, \dots, OR_{j_1, j_2}, \dots) \quad (4-10)$$

平均交叉值 $Average\ Overlap\ Rate\ (AOR)$: 所有评审员对作品的交集的平均大小。数值越大，交叉值越高。

$$AOR = \frac{1}{C_n^2} (OR_{1,2} + OR_{2,3} + \dots + OR_{j_1, j_2} + \dots) \quad (4-11)$$

4.4 模型求解

4.4.1 算法设计

NSGA-II (非支配排序遗传算法 II, Non-dominated Sorting Genetic Algorithm II) 是一种在多目标优化问题中广泛使用的算法。它的设计初衷是为了解决两个常见的多目标遗传算法问题: 1) 选择压力不足, 导致收敛速度缓慢; 2) 无法保持种群的多样性, 导致 Pareto 前沿上的解分布不均。

(1) NSGA-II 算法编码解码

针对问题一确立最优“交叉分发”方案, 首先对当前种群进行非支配排序。在这个排序过程中, 将会根据个体之间的支配关系为每个个体分配一个等级。本文基因编码由 3000×125 的矩阵构成, 每一行的矩阵由随生成的 0~1 的实数填充, 如下图 4-1 所示。

根据上述编码结果, 每行生成的随机数中, 我们选取最大的五个数, 并在其对应位置标为 1, 其余位置标为 0。本文基因解码由 3000×125 矩阵构成, 每一行代表一个作品, 每一列代表一个评审专家, 如图 4-2 所示, 矩阵中的 1 表示该作品被该专家进行了评审, 如

下图所示，表示作品 1 和作品 3 被专家 1 进行评审，作品 2 和作品 3 被专家 3 进行评审。

		评审专家						
		1	2	3	124	125	
作品	1	1	0.9	0	0	0.95	
	2	0.1	0.3	1	0.4	0.2	
	3	0.98	0.32	0.95	0.31	0.4	
							
	2998	0.61	0.23	0.99	0.15	1	
	2999	0.11	0.17	0.98	0.62	0.13	
	3000	1	0.25	0.97	0.29	0.78	

图 4-1 染色体编码

		评审专家						
		1	2	3	124	125	
作品	1	1	1	0	0	1	
	2	0	0	1	0	0	
	3	1	0	1	0	0	
							
	2998	0	0	1	0	1	
	2999	0	0	1	0	0	
	3000	1	0	1	0	0	

图 4-2 染色体解码

(2) 交叉

我们采取了经典的交叉方式。首先在父代 1 中随机挑选若干位点并将其转移到子代染色体。接着，从父代 2 中获取子代未包含的部分，并按照原有顺序填补到子代中。

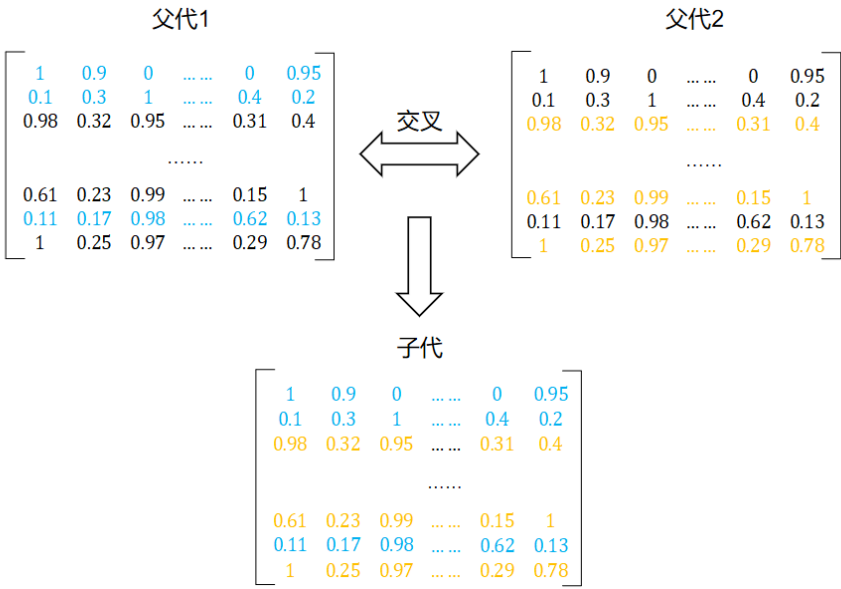


图 4-3 染色体交叉

(3) 变异

变异随机选择染色体位置进行变异。

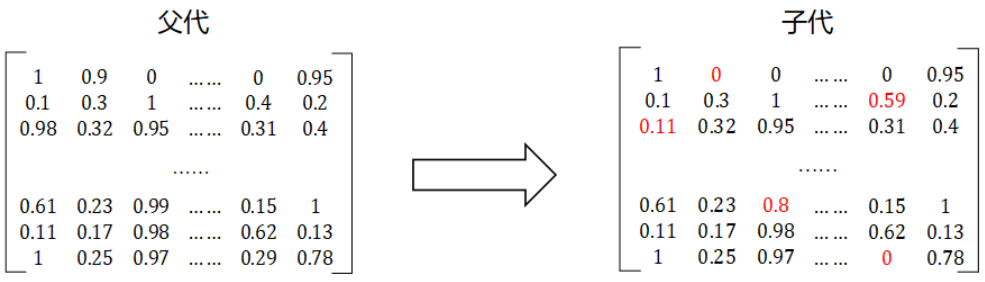


图 4-4 染色体变异

(4) 适应性函数 f

适应性函数由交叉值和标准差组成，其中：

作品交叉值定义：对于每对专家组 j_1 和 j_2 （其中 $j_1 \neq j_2$ ），其交叉值 OR_{j_1,j_2} 是这两位专家同时评审的作品数量。具体来说，如果 E_{j_1} 是评审专家 j_1 的作品集合，而 E_{j_2} 是评审专家 j_2 的作品集合，那么交叉数 OR_{j_1,j_2} 就是这两个集合的交集的大小。例如，假如我们有 5 位专家，分别标记为 j_1, j_2, j_3, j_4, j_5 。他们评审一下作品集： $E_{j_1} = \{A, B, C, D\}$; $E_{j_2} = \{B, C, D, F\}$; $E_{j_3} = \{D, E, F, G\}$...那么专家 j_1 和 j_2 的交叉值 $OR_{j_1,j_2} = |E_{j_1} \cap E_{j_2}| = |\{A, B, C, D\} \cap \{B, C, E, F\}| = |\{B, C\}| = 2$ 。适应性函数具体公式如下所示：

$$f_1 = \sum_{j_1, j_2}^n OR_{j_1, j_2}, j_1 \neq j_2 \quad (4-12)$$

作品交叉值的标准差定义：为了计算标准差，我们首先需要计算交叉数的平均值和方差。对于特定的作品 i ，其交叉值的标准差描述了作品 i 与其他所有作品的交叉值得变化程度。这表示了作品 i 在专家评审中与其他作品得相似性的稳定性，具体公式如下所示：

$$f_2 = \sqrt{\frac{\sum_{j_1, j_2}^n (OR_{j_1, j_2} - OR_{j_3, j_4})^2}{C_n^2}} \quad (4-13)$$

将这两个指标合并，可以为每个作品得到一个双目标适应性函数值，其中一个目标是最大化交叉值，另一个目标是最大化标准差。

4.4.2 算法步骤和流程

基于 NSGA-II 算法求解评审分配的问题：

初始化：

步骤 1.1: 初始化种群，生成随机染色体。

步骤 1.2: 解码染色体并计算种群中每个染色体的适应度值。

排序：

步骤 2.1: 对初始种群进行非支配排序。

步骤 2.2: 计算每个染色体的拥挤距离。

选择：

步骤 3.1: 选择种群中的父代。这通常是基于它们的非支配排序和拥挤距离。

交叉和变异:

步骤 4.1: 使用选择的父代通过交叉和变异技术生成子代。

集成与更新:

步骤 5.1: 将当前种群和子代合并。

步骤 5.2: 再次进行非支配排序和计算拥挤距离。

步骤 5.3: 根据非支配排序和拥挤距离选择下一代种群。

终止条件:

步骤 6.1: 检查是否达到最大迭代次数。若已达到，则转至步骤 7；否则，返回步骤 3 继续迭代。

输出:

步骤 7.1: 输出 Pareto 前沿中的染色体及其适应度值，显示评审分配结果。

本问题采用的遗传算法的具体流程图如下所示:

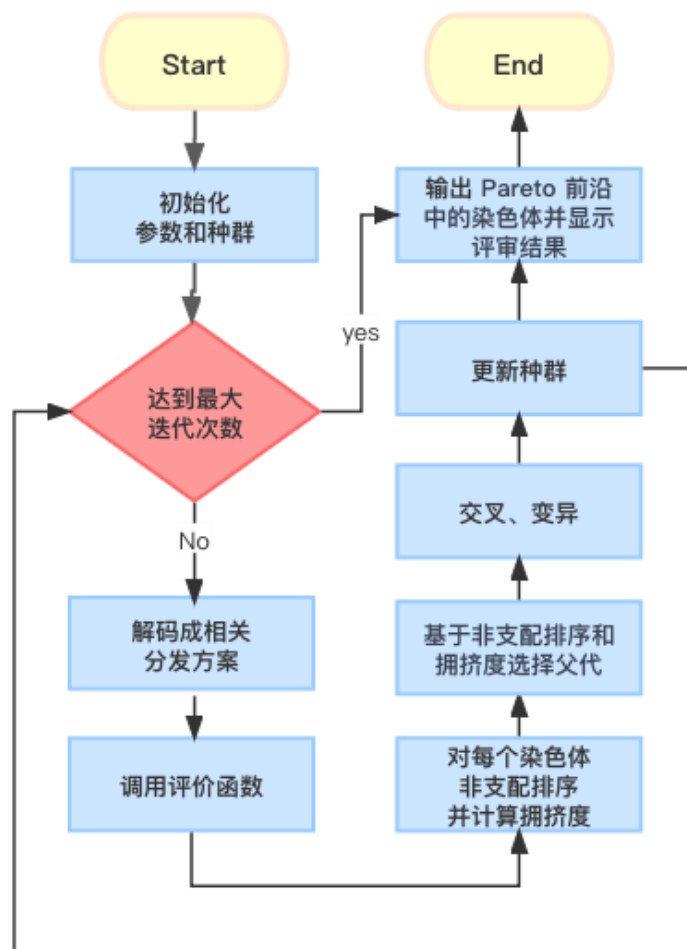


图 4-5 遗传算法求解流程图

4.5 求解结果与分析

(1) 求解结果

在本问题中，我们的目标是在维护评审结果的可比性的同时，最大程度地保证方案的均衡性、交叉性、公正性、覆盖率和分散性，最终“交叉分发”方案的结果如下。

问题一模型最终只生成 1 个解集，最终结果为 (30000, 1.6258)，如图 4-6。

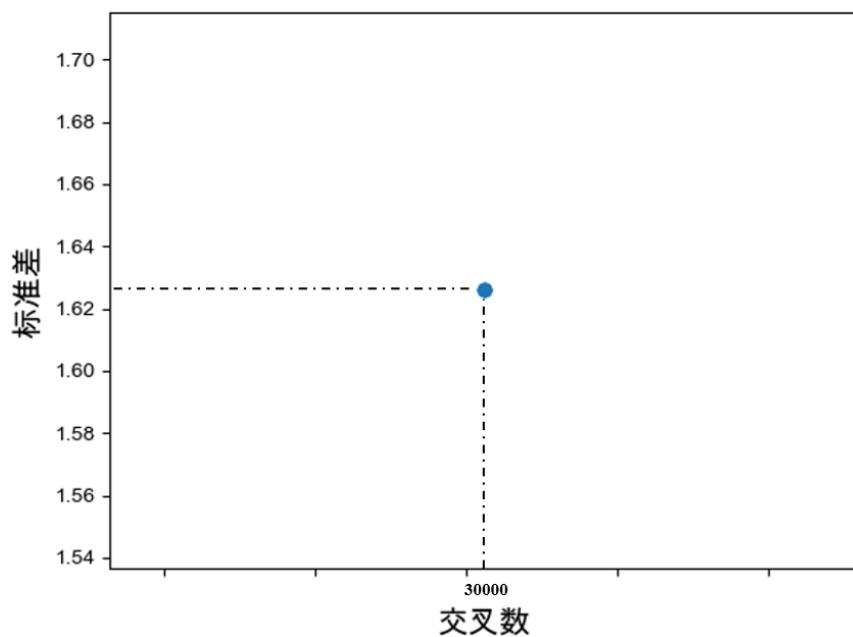


图 4-6 帕累托最优解

方案最终为 125 为评审专家分配的作品数量，保证一定的均衡性，如图 4-7。

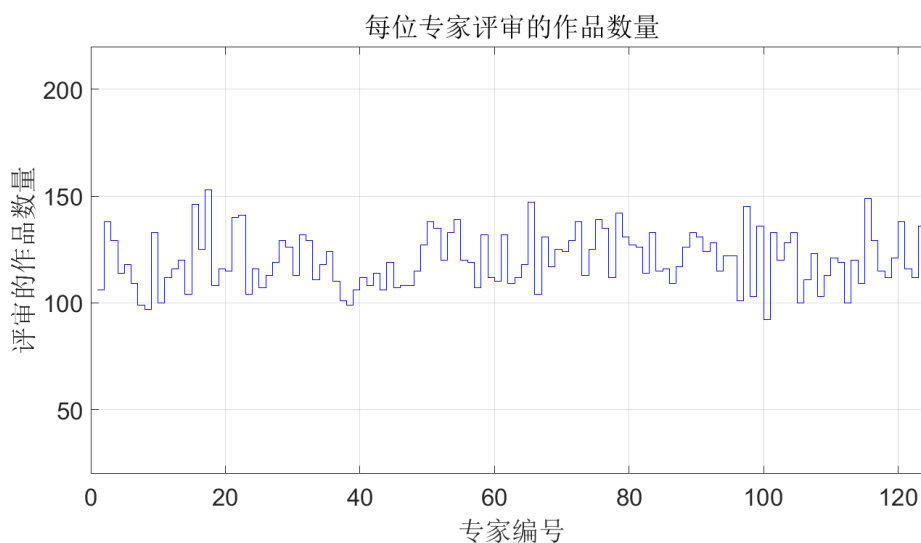


图 4-7 每个专家评审作品数的阶梯图

本问题使用的表格记录包括了专家评审作品集合的详细信息。该表格提供了对竞赛评审过程中每位专家所评审的作品数量和作品集合。如下表 4-1 所示。

表 4-1 专家评审作品集合信息表

专家编号	集合总数	评审作品集合
Reviewer_1	106	[18, 25, 42, 49, 59, 93, 106, 112,, 2880, 2934, 2935]
Reviewer_2	138	[60, 71, 87, 145, 162,, 2943, 2968, 2987, 2992]
Reviewer_3	114	[34, 53, 99, 105,, 2957, 2960, 2966, 2981]
.....		
Reviewer_60	112	[13, 43, 77, 125, 129,, 2868, 2962, 2986, 2989]
Reviewer_61	110	[18, 64, 78, 80,, 2929, 2951, 2992]
Reviewer_62	132	[10, 24, 131, 147,, 2791, 2831, 2854, 2934]
.....		
Reviewer_123	136	[34, 78, 88, 197,, 2904, 2930, 2937, 2939, 2974]
Reviewer_124	117	[6, 22, 28, 32, 34, 80, 117,, 2828, 2855, 2887, 2991, 2996]
Reviewer_125	119	[32, 66, 69, 107, 149, 222,, 2933, 2948, 2956, 2979]

我们绘制了 125 个评审专家交叉评审作品数量的矩阵热力图，以了解在评审过程中不同专家之间作品的交叉情况。每个热力格代表对应两个专家间评审作品的交叉数，颜色越深交叉数越多，从图 4-8 中可看出，颜色分布比较均衡，满足“交叉分发”方案的均衡性。

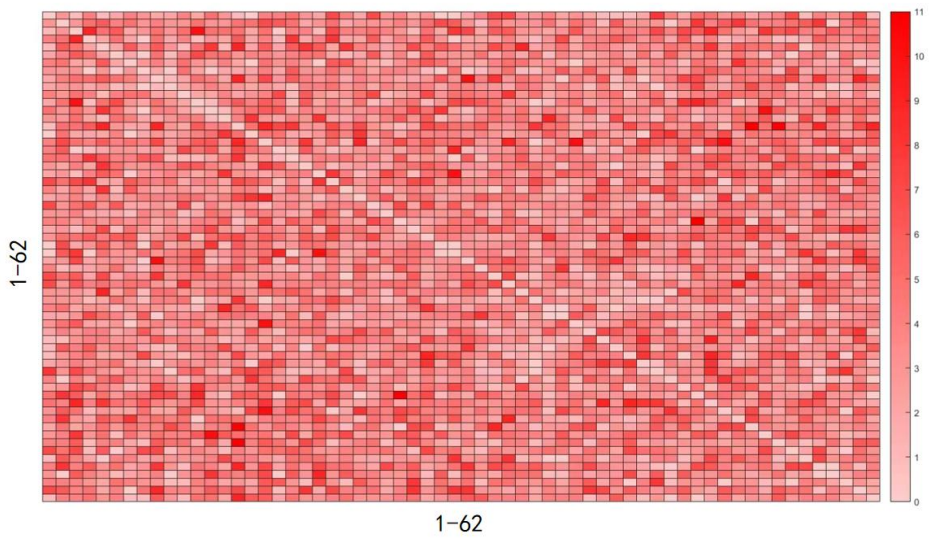


图 4-8(a) 评审专家作品交叉情况热力图

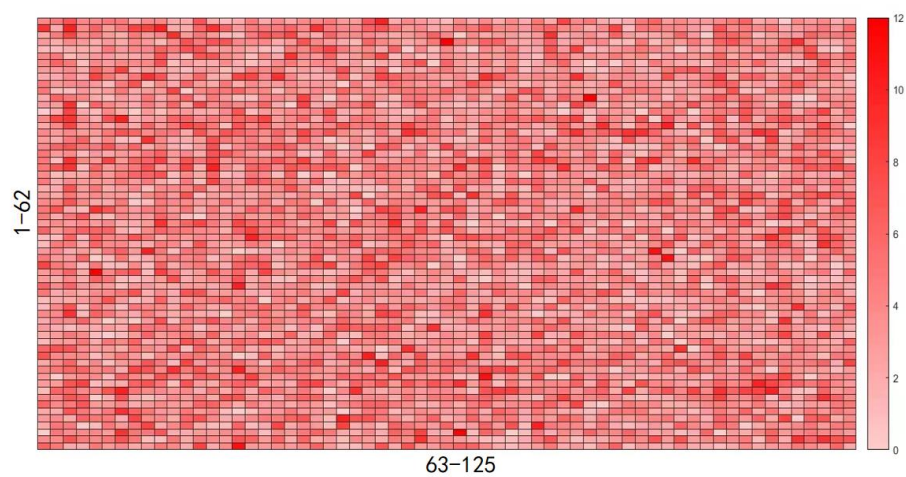


图 4-8(b) 评审专家作品交叉情况热力图

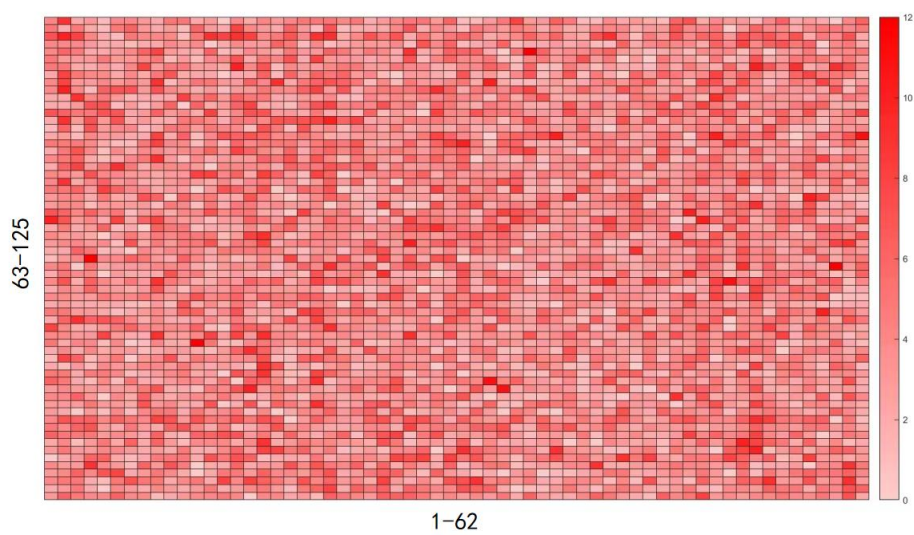


图 4-8(c) 评审专家作品交叉情况热力图

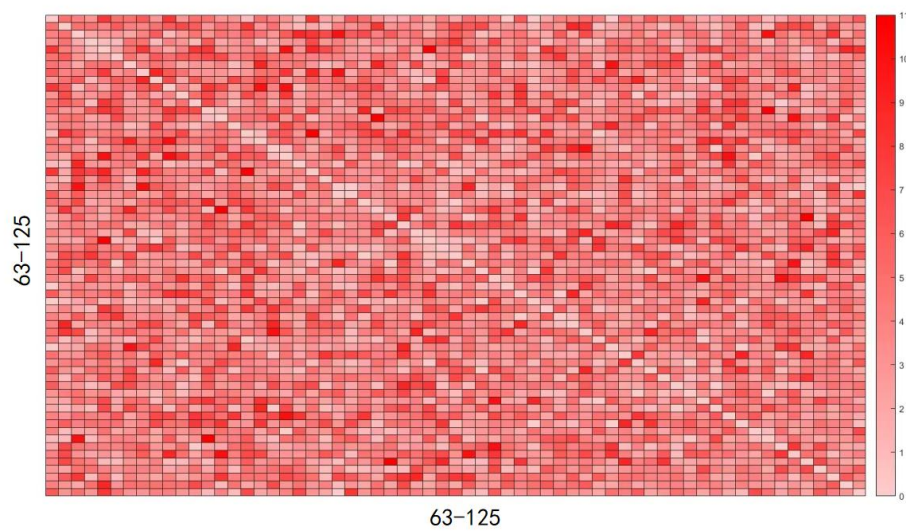


图 4-8(d) 评审专家作品交叉情况热力图

（2）评价指标分析

我们定义“交叉分发”方案中保证方案的均衡性、交叉性、公正性、覆盖率和分散性的评价指标计算结果如下表 4-2 所示：

表 4-2 评价指标结果

评价指标	指标值
<i>RWB</i>	12.63
<i>MEB</i>	0.0
<i>Max_OR</i>	12
<i>Min_OR</i>	0
<i>AOR</i>	3.87

1. 评审工作量均衡性*Reviewer Workload Balance (RWB)*: 这个指标测量的是评审员之间工作量的差异性，若值为 0，这意味着所有的评审员都有完全相同的工作。在本方案中，*RWB* 为 12.63，这意味着评审员之间的工作量存在一定的差异。

2. 作品曝光均衡性*Manuscript Exposure Balance (MEB)*: 这个指标测量的是作品被评审的均衡性。理想的值同样为 0，表明每部作品都被相同数量的评审员评审。*MEB* 为 0.0，这是一个完美的得分，表明每个作品都得到了相同的评审次数。

3. 最大交叉值*Maximun Overlap Rate (Max_OR)*: 这代表了两位评审专家之间评审作品的最大交集数量。*Max_OR*为 12，这意味着存在某两位评审专家，他们至少共同评审了 12 个相同的作品。

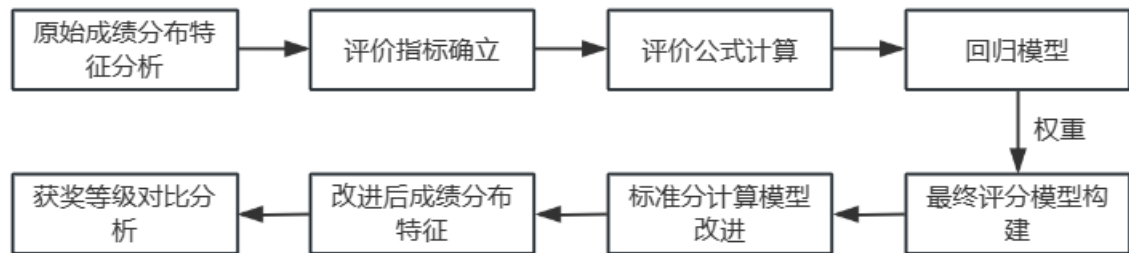
4. 最小交叉值*Minimum Overlap Rate (Min_OR)*: 与*Max_OR*相反，这是两位评审专家之间评审作品的最小交集数量。*Min_OR*为 0，表明存在某两位评审专家他们没有共同评审的作品。

5. 平均交叉值*Average Overlap Rate (AOR)*: 这表示了所有可能的评审员组合评审的平均交集作品数量。*AOR*为 3.87，这意味着平均而言，任意两位评审员之间共同评审了大约 4 个作品。

五. 问题二：评审方案设计与标准分改进

5.1 问题描述与分析

在大规模的创新竞赛评审过程中，各位专家评审的作品主要各不相同，且每位专家仅能审阅到全部作品集合的一小部分。这一情况使得依赖标准分的评审方案的基本假设可能变得不太适用。因此，选择或设计两种或更多的评审方案，设计新的标准分（公式）计算模型，以确保评审的公正性和准确性。以下为本问题的思路流程：



5.2 数据预处理

5.2.1 数据转换和整合

为了使数据更符合分析的需求，我们将执行以下步骤：

等级编码：对“奖项”进行等级编码，将作品的获奖情况：“一等奖”、“二等奖”、“三等奖”和“未获奖”分别编码为 1、2、3 和 4。

添加复议标记：根据是否存在复议分，为数据添加一个新的特征变量“复议标记”。具体来说，如果存在复议分，该标记为 1，否则为 0。

数据排序：根据“专家编号”和“作品编号”分别对数据进行排序整合。

5.2.2 原始成绩分布特征分析

为每位专家分析他们评分的分布特点，并为每位专家绘制分布图。这包括计算每位专家的分数平均值、中位数、众数、标准差、偏度和峰度，并绘制直方图来表示分数的分布。以下为部分专家的评分分布直方图。考虑到图的数量和可视化效果，只选择了 5 位专家进行展示。每个直方图中，还包括了一个核密度估计曲线（KDE），它提供了一个平滑的分布形状。

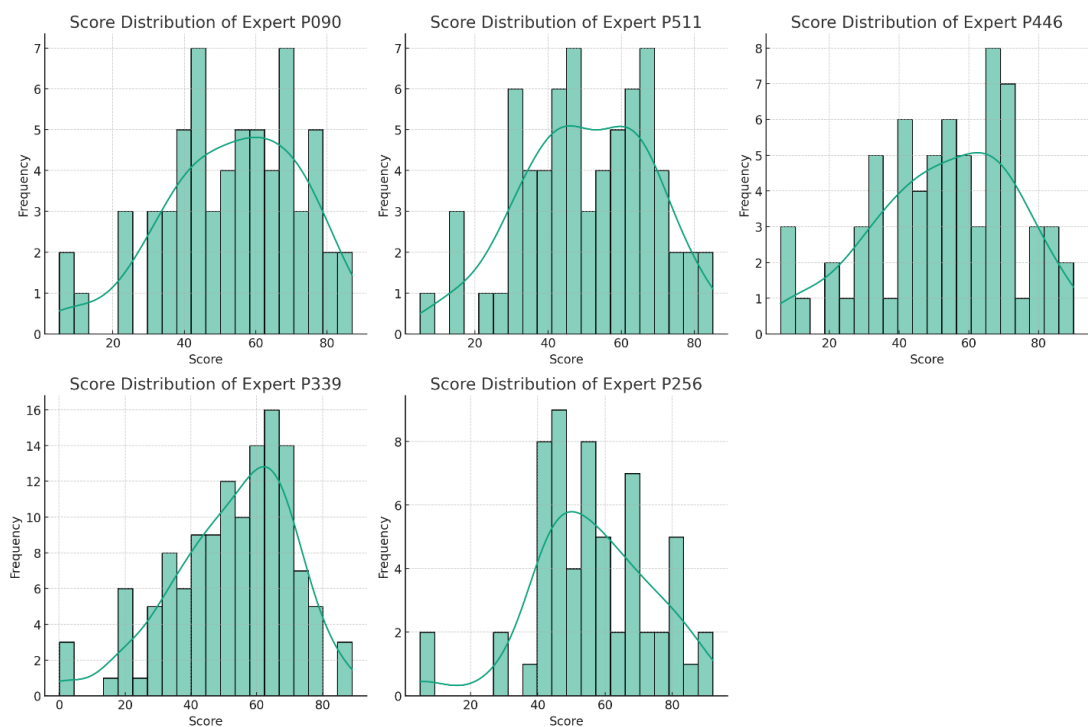


图 5-1 部分专家评分分布直方图

通过观察这些图，我们发现多数专家评分分布呈双峰形态，说明这些专家在评分时倾向于给出两个主要的分数区间，而不是一个连续的分数范围；不同专家的评分习惯不同，某些专家的评分更为集中，而其他专家更为分散，每位专家的评分习惯和评价标准存在差异；分数范围的差异，大多数专家打分范围大致相同，但也有一些专家的范围较窄或较宽。

为了更好地理解和展示每位专家的评分习惯和分布特点，我们可以使用偏度和峰度的值来绘制一个散点图，其中偏度作为 x 轴，峰度作为 y 轴。通过这样的图，我们可以直观地看到每位专家的打分数据的形状特点。

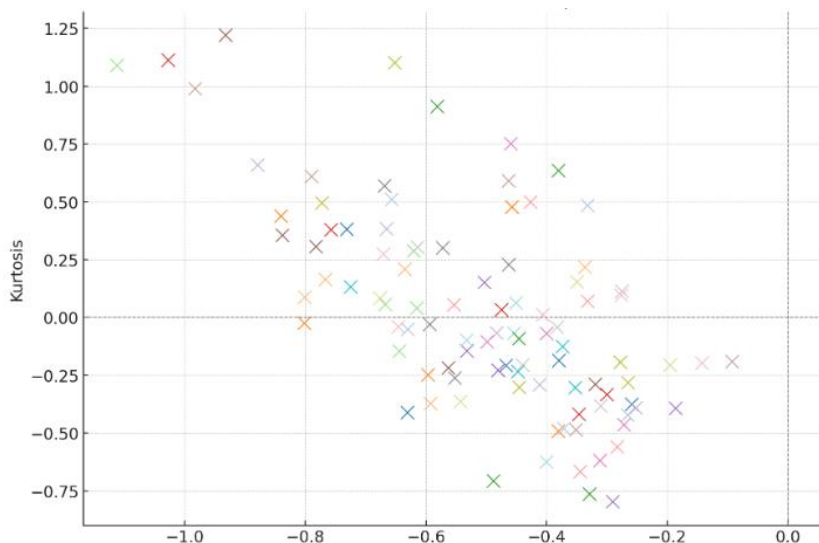


图 5-2 专家评审分数偏斜与尖锐度分析图

Skewness 偏度描述了分布的偏斜程度。正值表示分布向左偏斜，负值表示分布向右偏斜。**Kurtosis** 峰度描述了分布的尖锐程度。正值表示分布比正态分布更尖锐，负值表示分布比正态分布更平坦。从图中可以观察到：

偏斜程度：大多数专家的打分数据都呈现出轻微的左偏（偏度 <0 ）。这意味着他们的分数分布的右尾部分比左尾部分更长，且多数得分位于平均分的左侧。

尖锐程度：大多数专家的打分数据的峰度接近于 0，这意味着他们的分数分布的形状与正态分布相似。没有专家的打分数据呈现出明显的高尖或扁平的特点。

分析作品得分的分布特点，并绘制分布图。这包括计算分数的平均值、中位数、众数、标准差、偏度和峰度，并绘制直方图来表示分数的分布。

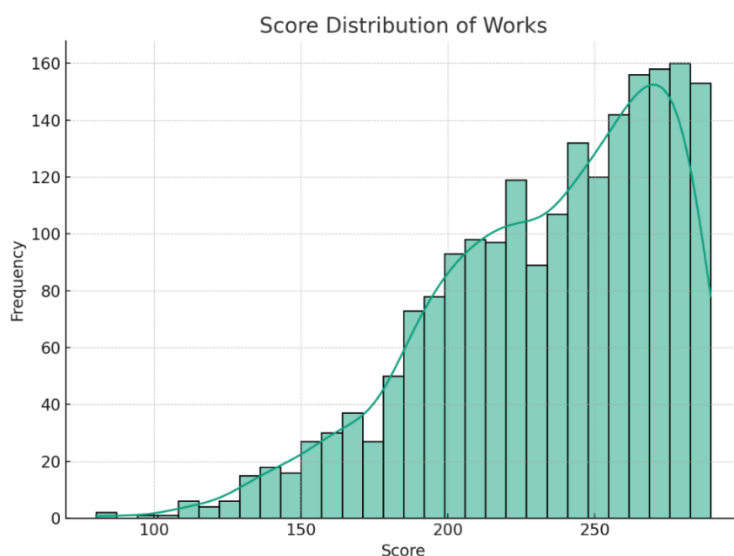


图 5-3 作品得分的分布直方图

我们通过计算得出：**Skewness**：偏度为 -0.73 。该值小于 0，表明分布是左偏的，即分布的左尾部分比右尾部分更长，多数得分位于平均分的右侧。**Kurtosis**：峰度为 -0.02 。该值接近 0，表明该分布与正态分布的峰态相似。

从直方图中，我们可以观察到作品得分主要集中在 200-260 的范围内，且分布略微左偏。这意味着较高的分数比较常见，而较低的分数则较为稀少。

为了分析专家评审打分与作品最终得分的关系特征，我们使用散点图来直观地展示两者之间的关系。此外，我们在散点图上画出线性回归的拟合线，以更好地展示两者之间的线性关系。为了简洁，我们选择为几位专家绘制此类图。

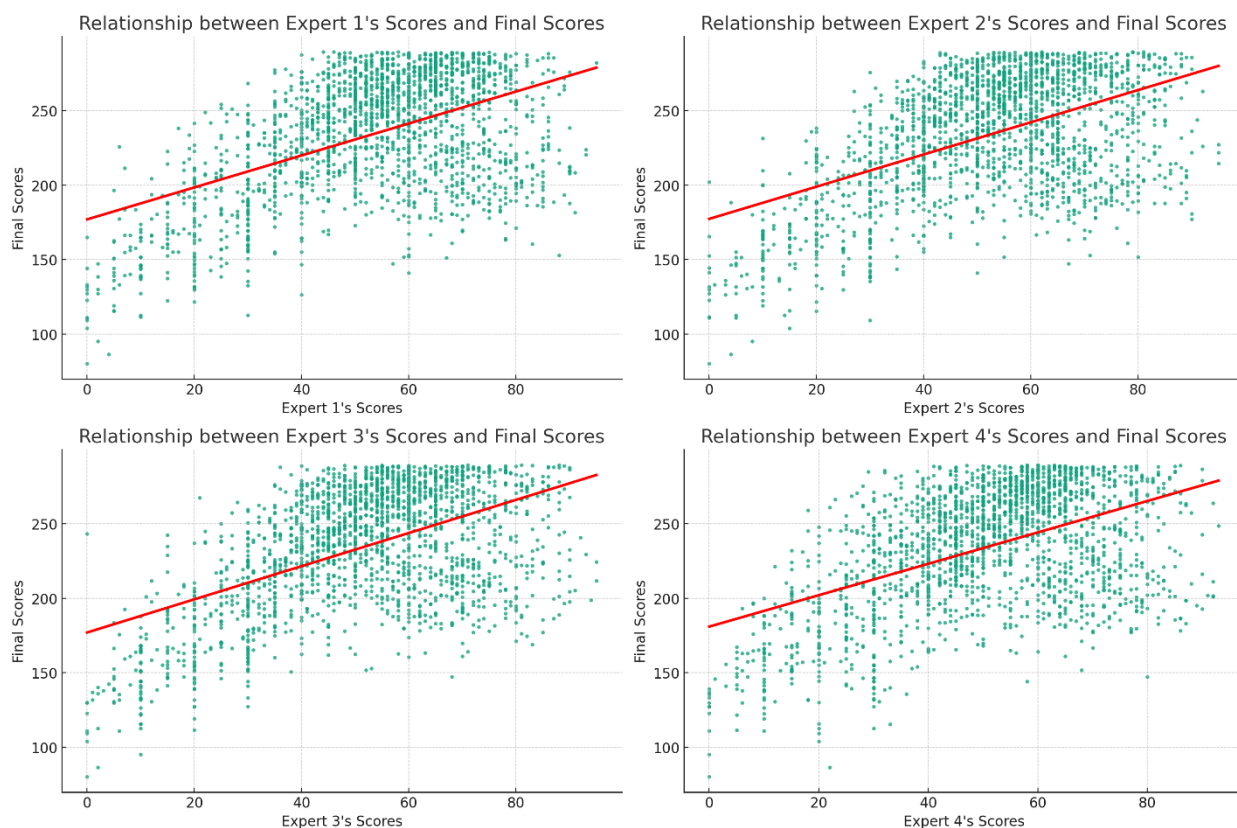


图 5-4 部分专家评分与作品最终得分的关系特征散点图

这些图展示了部分专家评分与作品最终得分之间的关系。每个散点图中的红线代表线性回归的拟合线，我们可以看出：

正相关：每位专家的打分与作品的最终得分都呈现出正相关关系。这意味着当一位专家给出较高的分数时，作品的最终得分也很可能较高，反之亦然。

分散度：各专家的打分与作品的最终得分的散点图分散度不同，这可能与他们的打分习惯和评价标准有关。

拟合线：红色的线性拟合线为我们提供了一个直观的方式来理解专家打分与作品最终得分之间的关系。对于大多数专家，他们的打分与作品的最终得分都较为接近这条线，这意味着他们的打分与作品的最终得分有很强的线性关系。

5.2.3 评价指标确立

（1）评审方案一：专家角度

从专家角度建立评价指标，目的是从不同的角度全面地了解专家在评审作品中的评分习惯和一致性，以及与其他评审专家的评分差异。我们可以考虑以下几个方面：专家评分标准差、专家评审作品数量、以及评审专家的评分偏好。

① 专家评分标准差

专家标准差描述了一个专家对每个作品评价的偏差程度。具体来说，它计算了每个专家对每个作品的原始评分与该作品的标准分之间的差异的方差。一个较小的专家标准差表示该专家的评分与大多数其他专家的评分更为接近，而一个较大的专家标准差表示该专家的评分与其他专家存在较大差异，计算公式如下所示：

$$SD_j = \sqrt{\frac{\sum (\text{作品标准分} - \text{作品原始分})^2}{\text{某个专家评审作品总数}}} \quad (5-1)$$

② 评审作品数量

评审作品数量评审作品的数量作为评价指标可以帮助我们了解评审专家的评分经验和工作量。通常，评审的作品数量越多，意味着该专家在评审中的经验更为丰富，对竞赛作品有更全面的了解。但这也可能带来一个问题，即当评审的作品数量过多时，可能会导致评审疲劳，进而影响到评分的准确性和一致性。

② 评审专家的评分偏好

每位专家的评分可能是一个作品得到的所有评分中的最高分、最低分，或者中间的分数。为了评估一个专家的评分与整体评分的差异，我们可以计算极评指标 *Extremity Scoring Index*:

$$ESI = |a - b| \quad (5-2)$$

$$p_j + b_j + q_j = N_j \quad (5-3)$$

$$a = \frac{p_j}{N_j}, b = \frac{b_j}{N_j}, b = \frac{b_j}{N_j} \quad (5-4)$$

其中， N_j 为专家 j 评审的作品总数； p_j 表示专家 j 给出的评分是某个作品得到的所有评分中的最高分的数量； b_j 表示专家 j 给出的评分是某个作品得到的所有评分中的最高分的数量； q_j 表示专家 j 给出的评分既不是最高分也不是最低分的作品数量。当极评指标较小时，表示专家的评分相对稳定和一致，不至于偏向某个极端。

(2) 评审方案二：综合评审

为了更准确地评估每个作品的真实水平，我们建立一个综合评审方案，该方案主要考虑了两个关键因素：专家的专业性评分和作品的得分合理性。

专家专业性评分(Expertise Rating): 反映了专家在评审中的专业性和一致性。这部

分主要基于“专家角度”的评审方案，考虑了专家的评分习惯、一致性和经验。

作品得分合理性评分(Score Rationality Assessment): 反映了作品在所有评审专家中的整体表现和得分的一致性。这部分主要基于“作品角度”的评审方案，考虑了作品的评分分布、一致性和复议情况。其中相关影响因素为:

①作品的评分分布

查看每个作品的平均评分、最高评分、最低评分以及计算最终的标准差，平均评分反映了作品的整体质量，而最高评分、最低评分和标准差则可以帮助我们了解评审专家之间的评分差异。标准差可以帮助我们判断作品的整体受欢迎程度以及评审专家间的评分一致性。

相关特征变量: 我们看了每个作品的平均评分、最高评分、最低评分以及评分的标准差。

$$\text{平均分} = \frac{\sum \text{每个评审专家的评分}}{\text{评审专家的数量}} \quad (5-5)$$

$$\text{标准差} = \sqrt{\frac{\sum (\text{评分} - \text{平均分})^2}{\text{评审专家的数量}}} \quad (5-6)$$

②作品的极差值

极差是一种统计量，用于描述数据集中的最大值和最小值之间的差异，是最大值与最小值之差。对于每个作品，计算其评分的极差（最高分与最低分之差）以了解评审专家间的评分差异。一个大的极差可能意味着评审专家之间对作品的看法存在较大的分歧。

当某个作品的评分极差较大时，这意味着评审专家之间的评价存在很大的分歧。一些专家可能认为该作品很好，而另一些专家可能认为它较差；相反，当极差较小时，这意味着所有评审专家对该作品的评价都相对一致。

③复议分

复议分是当作品或项目的初始评分被认为可能不公正或有误时，可以请求对其进行复审模型建立，主要目的是确保评审过程的公正性和准确性，为评审对象提供一个纠正可能的评分误差或偏见的机会。复议不仅仅是对初始评分的简单修正，而是要对整个作品进行全面、深入的再次评估，确保得到的复议分是准确和公正的，因此，相对于原始分数，复议分会更加准确。在问题二中，我们将复议分替代原始评分，作为该作品的最终评分。

5.2.4评分公式计算

(1) 专家专业性评分(Expertise Rating)

我们最终收集了以下三个关键指标：专家评分标准差（越小越好）、评审作品数量、极评指标（越小越好），为了确保每个指标在分析中具有平等的重要性，我们对原始数据进行了标准化处理。这使得每个指标的平均值为 0，标准差为 1。我们使用主成分分析 (PCA) 来确定每个指标的权重。这种方法基于数据的方差和相关性来确定权重。

PCA 的结果为：专家评分标准差权重：0.2573；评审作品数量权重：0.6797；极评指数权重：0.6869。

使用上述权重，我们为每个专家计算了最终评分，公式如下：

$$\text{专家专业性评分(ER)} = 0.2573 * \text{专家评分标准差} + 0.6797 * \text{评审作品数量权重} + 0.6869 * \text{极评指数}$$

(2) 作品得分合理性评分(Score Rationality Assessment)

我们通过以下指标：原始分（越小越好）、标准分（越小越好）、第二次评审标准分极差（越小越好）、极差（越小越好）以及复议标记，并通过熵权法确定指标权重。熵权法的主要步骤如下：标准化数据、计算每个指标的信息熵、计算每个指标的差异系数、计算每个指标的权重。

熵权法的结果为：原始分权重：0.0146；标准分权重：0.0258；第二次评审标准分极差权重：0.3611；极差权重：0.0221；复议标记权重：0.5764。

使用上述权重，我们为每个作品计算了最终评分，公式如下：

$$\text{作品得分合理性评分(SRA)} = 0.0146 * \text{原始分} + 0.0258 * \text{标准分} + 0.3611 * \text{第二次评审标准分极差权重} + 0.0221 * \text{极差} + 0.5764 * \text{复议标记}$$

我们将专家专业性评分(ER)、作品得分合理性评分(SRA)做为评审方案二的关键指标，根据主成分分析法计算它们之间的权重，结果为专家专业性评分：0.6053、作品得分合理性评分 0.3947。使用上述权重，综合评审得分(IntegratedGrade)的公式如下：

$$\text{综合评审得分(IG)} = 0.6053 * \text{专家专业性评分(ER)} + 0.3947 * \text{作品得分合理性评分(SRA)}$$

5.3基于线性回归的标准分计算模型

5.3.1回归模型介绍

我们选择线性回归方法来确定权重。对于线性回归，我们将一元变量推广到多原变量，

设多元函数式为 $f(x_1, x_2 \dots x_n) = w_1x_1 + w_2x_2 + \dots + w_nx_n + b$ 。使用线性代数的向量概念对该式进行整理，为方便，记 $w_0 = b$ （可把上函数式的 b 视为 $b*1$ ）。

那么此时，我们构造一个权重向量 w 和特征向量 x ：

$$w = (w_0, w_1, w_2, \dots, w_n) \quad (5-7)$$

$$x = (1, x_1, x_2 \dots, x_n) \quad (5-8)$$

上述的多元函数式则可以写成 $f(x) = wx$ 或 $f(x) = xw$ ，是完全等价的。此时，我们可以以损失函数写成如下形式：

$$L(w) = \sum_{i=1}^m (y_i - w^T x^{(i)})^2 \quad (5-9)$$

其中， y_i 为第 i 个真实值， $x(i)$ 为第 i 个样本的特征向量。

注：真实值，标签值，样本值这三个词意思是相同的。

5.3.2回归模型构建

数据准备：

对于评审方案一，我们根据平均专家专业性评分(AER)对数据 1 中的一等奖、二等奖、三等奖和未获奖的四个等级的作品分别进行排序，根据排序结果选择一等奖的前 100%以及二等奖、三等奖、未获奖的前 50%做为用来进行回归的数据集。

$$AER = \frac{1}{N_i} \sum_j ER_j$$

其中， N_i 为第 i 个作品的评审专家数量。

对于评审方案二，我们根据综合评审得分(IG)对数据 1 中的一等奖、二等奖、三等奖和未获奖的四个等级的作品分别进行排序，根据排序结果选择一等奖的前 100%以及二等奖、三等奖、未获奖的前 50%做为用来进行回归的数据集。

将数据分为训练集和测试集。训练集用于训练模型，而测试集用于验证模型的性能。训练集占总数据集 70%，测试集占总数据集 30%。

模型假设：

① 评审方案一：专家角度

我们旨在构建一个从专家评审角度考虑专家的评分习惯、一致性和经验，计算各个参赛作品的最终得分(Final Score Evaluation)。为了实现这一目标，该方法考虑了三个主要的影

响因素：标准分(std_{ij})、专家评分专业性(ER_{ij})和第*i*个作品的平均专家评分专业性(AER_i)。

具体地，我们的模型假设最终得分是这三个变量的线性组合：

$$FSE_1 = \alpha * \sum_j std_{ij} + \beta * \sum_j ER_{ij} + \gamma * AER_i$$

该方案的标准分计算模型为：

$$STD_{1ij} = \beta * ER_{ij} + \alpha * std_{ij}$$

② 评审方案二：综合评审

我们旨在构建一个综合评价模型，从专家专业水平评估和作品得分合理性角度，计算各个参赛作品的最终得分。综合评审方案考虑了评审方案一的标准分(std_{ij})、专家评分专业性(ER_{ij})、第*i*个作品的平均专家评分专业性(AER_i)以及作品得分合理性评分(SRA_i)，我们的模型假设最终得分是这四个变量的线性组合：

$$FSE_2 = \alpha * \sum_j std_{ij} + \beta * \sum_j ER_{ij} + \gamma * AER_i + \delta * SRA_i$$

该方案的标准分计算模型为：

$$STD_{2ij} = \beta * ER_{ij} + \delta * SRA_i + \alpha * std_{ij}$$

设定成本函数：

成本函数（或损失函数）衡量模型的预测值与实际值之间的差距。我们的目标是通过调整 $\alpha, \beta, \gamma, \delta$ 的值来最小化成本函数。在线性回归中，我们通常使用均方误差作为成本函数。通过最小化成本函数，我们可以找到最佳的模型参数。 \hat{y}_i 是模型的预测值， y_i 是实际观察值， d 是观察值的数量。

评审方案一：

$$\hat{y}_i = \alpha * \sum_j std_{ij} + \beta * \sum_j ER_{ij} + \gamma * AER_i$$

$$J_1(\alpha_i, \beta_i, \gamma_i) = \frac{1}{2d} \sum_{i=1}^m \left(\left(\alpha * \sum_j std_{ij} + \beta * \sum_j ER_{ij} + \gamma * AER_i \right) - y_i \right)^2$$

评审方案二：

$$\hat{y}_i = \alpha * \sum_j std_{ij} + \beta * \sum_j ER_{ij} + \gamma * AER_i + \delta * SRA_i$$

$$J_2(\alpha_i, \beta_i, \gamma_i) = \frac{1}{2d} \sum_{i=1}^m \left(\left(\alpha * \sum_j std_{ij} + \beta * \sum_j ER_{ij} + \gamma * AER_i + \delta * SRA_i \right) - y_i \right)^2$$

选择优化技术，最小二乘法是一种常用的线性回归优化技术。它通过数学公式直接计算出使成本函数最小的参数值。这种方法非常高效，但要求特征矩阵必须是可逆的。通过最小二乘法，我们能够从训练数据中估计出这些权重，从而优化我们的模型。

5.3.3模型结果评估

我们需要评估两个评价方案所建立模型的性能。常用的评估指标包括均方误差 (MSE) 和决定系数 (R^2)。我们通过这些指标来确定模型的好坏，判断是否能够评估和排名参赛作品提供一个可靠和有效的工具，从而帮助确保比赛的公平性和客观性。

其中：

$$MSE = \left(\frac{1}{n} \right) * \sum (y_i - \hat{y}_i)^2$$

MSE 计算了观测值与模型预测值之间的差异的平方，然后取平均值。 MSE 值越小，表示模型的拟合效果越好。

$$R^2 = 1 - (SSR - SST)$$

SSR 是回归平方和（回归模型解释的变异性的总和）， SST 是总平方和（因变量的总变异性）。 R^2 的值介于 0 和 1 之间，越接近 1 表示模型能够更好地解释因变量的变异性，拟合效果越好。

两个评审方案的评估结果见下表：

表 5-2 回归模型评估结果		
	评审方案一	评审方案二
MSE	279.15	241.63
R^2	0.79	0.82

根据线性回归方法，我们最终求出了两种评审方案的权重值：

表 5-3 评审方案权重值

	评审方案一	评审方案二
α	1.1161	0.910745
β	0.95079	0.961569
γ	-6.25542	-5.07002
δ	-	-3.84561

我们根据求得得权重值得出两个评审方案的最终得分模型：

$$FSE_1 = 1.1161 * \sum_j std_{ij} + 0.95079 * \sum_j ER_{ij} - 6.25542 * AER_i$$

$$STD_{1ij} = 0.95079 * ER_{ij} + 1.1161 * std_{ij}$$

$$FSE_2 = 0.910745 * \sum_j std_{ij} + 0.961569 * \sum_j ER_{ij} - 5.07002 * AER_i - 3.84561 * SRA_i$$

$$STD_{2ij} = 0.961569 * ER_{ij} - 3.84561 * SRA_i + 0.910745 * std_{ij}$$

已知， MSE 评估指标越小越好， R^2 评估指标越大越好，所以我们最终选择评审方案二——考虑专家的专业性评分和作品的得分合理性的综合评审做为我们最终得分模型，以及标准分计算模型。

$$STD_{ij} = 0.961569 * ER_{ij} - 3.84561 * SRA_i + 0.910745 * std_{ij}$$

5.3.4标准分计算模型改进

在 5.3.3 节中，建立的回归模型仅考虑了第一阶段的 5 位专家分数。第二阶段经专家协商一致后，选出一等奖作品的排序。第二阶段有 3 位专家评审，因此总共有 8 位专家评审进入第二阶段的作品，而未进入第二阶段的作品由 5 位专家评审。不同的评审专家数量将导致不同的结果。通过比较两个评审方案的评价指标结果（见表 5-3），我们将分析是否需要改进我们的标准分计算模型。

表 5-3 考虑不同阶段的评价方案的评估指标

	评审方案一_5	评审方案一_8	评审方案二_5	评审方案二_8
MSE	279.15	41.17	241.63	41.01
R^2	0.79	0.91	0.82	0.91

从 MSE 角度看，增加评审专家数目显著降低了 MSE 。 R^2 在两个方案中都有所提升，特

别是在 8 位评审专家的情况下，两个方案的 R^2 都是 0.91，表明模型拟合度很高。可以看出无论是评审方案一还是评审方案二，考虑第二阶段的专家评审之后，模型的性能都会提升。

我们考虑第二阶段的数据，改进两个评审方案，得出的新权重为：

表 5-4 评审方案新权重值

	评审方案一	评审方案二
α	3.65416	3.76465
β	0.761719	0.765625
γ	-2.92333	-3.01172
δ	-	-0.291511

我们根据求得新权重值得出两个评审方案新的最终得分模型以及标准分计算模型：

$$FSE_1 = 3.65416 * \sum_j std_{ij} + 0.761719 * \sum_j ER_{ij} - 2.92333 * AER_i$$
$$STD_{1ij} = 0.761719 * ER_{ij} + 3.65416 * std_{ij}$$
$$FSE_2 = 3.76465 * \sum_j std_{ij} + 0.765625 * \sum_j ER_{ij} - 3.01172 * AER_i - 0.291511 * SRA_i$$
$$STD_{2ij} = 0.765625 * ER_{ij} - 0.291511 * SRA_i + 3.76465 * std_{ij}$$

通过应用更新后的得分模型，我们能够计算出两种不同评审方案下所有作品的最终得分。如图 5-5 所示，两种评审方案的第二阶段得分均优于第一阶段，这一结果验证了我们模型的改进效果。

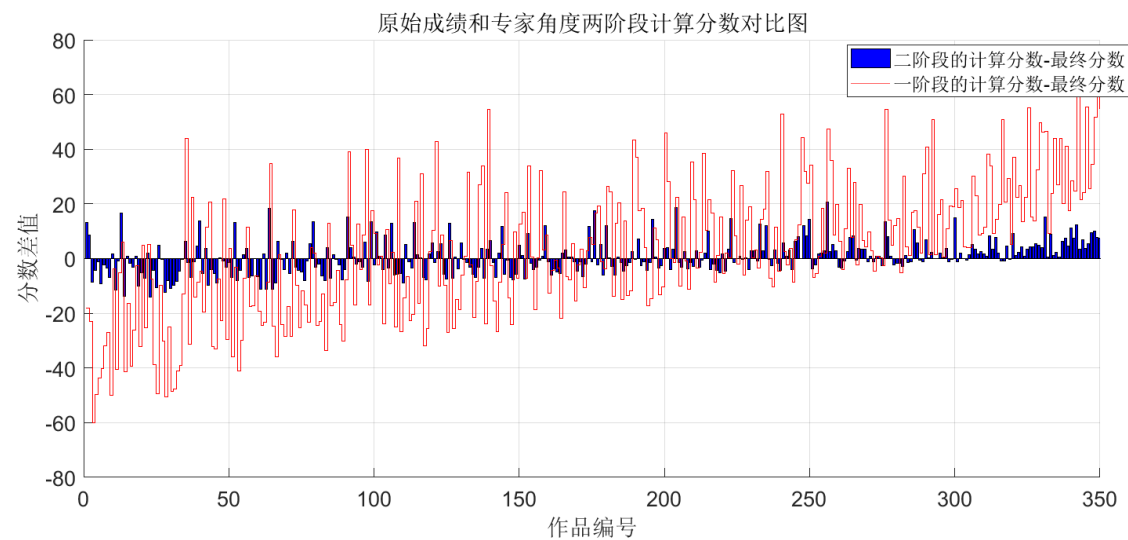


图 5-5(a)评审方案一两阶段对比

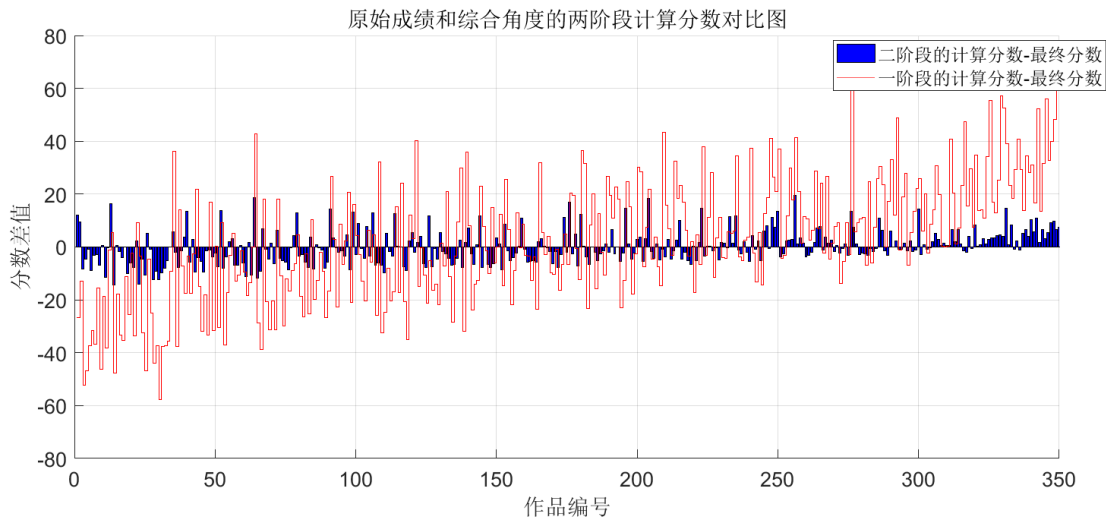


图 5-5(b)评审方案二两阶段对比

所以我们最终选择的评审方案二——考虑专家的专业性评分和作品的得分合理性的综合评审做为我们最终得分模型，以及改进之后标准分计算模型：

$$STD_{ij} = 0.765625 * ER_{ij} - 0.291511 * SRA_i + 3.76465 * std_{ij}$$

5.3.5调整后成绩分布特征

为每位专家分析他们打分的分布特点，均值：不同专家的平均打分范围从 246 到 279 不等。标准差：所有专家的标准差大约在 32 到 40 之间，显示出一定的分数分布差异。分布：从箱型图中可以看出，每个专家的打分分布都有一些不同，但大体上都是集中在某个范围内。

作品最终得分分布特征，均值与标准差：作品的最终得分的均值为 210.67，标准差为 20.41，显示了作品得分的分散程度。分布：从作品最终得分的直方图可以看出，大多数作品的得分集中在 200 到 220 之间。这表明大多数作品的质量相对集中，没有出现极端高分或低分的情况。四分位数：作品最终得分的第一四分位数（Q1）为 196.52，第三四分位数（Q3）为 222.90，显示了大部分数据（50%）的分布范围。

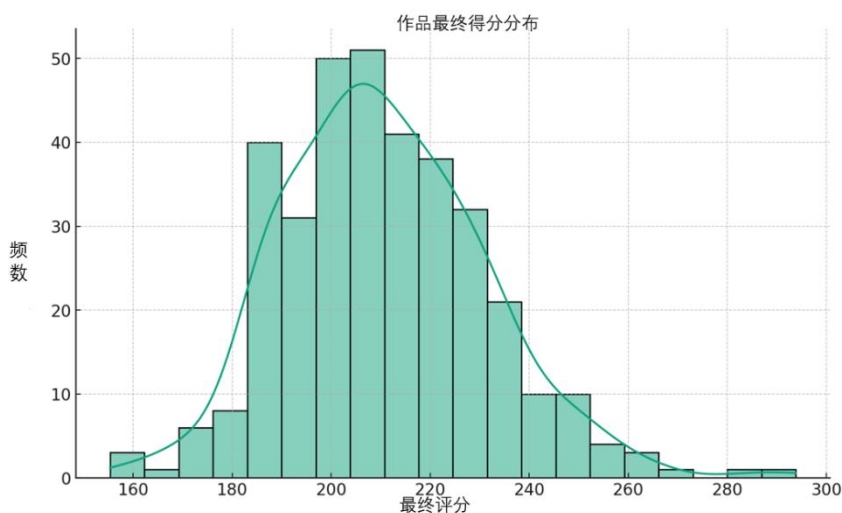


图 5-6 作品最终得分的直方图

通过计算每个作品的专家打分的平均值，并与最终得分进行比较，发现平均专家打分与作品最终得分之间存在很强的正相关关系（相关系数=0.906）。这意味着，一般情况下，专家给出的平均分越高，作品的最终得分也越高。这种强相关关系表明，专家打分在决定作品最终得分方面起着关键作用。评审员的评分标准和一致性对作品的最终评价有着直接的影响。

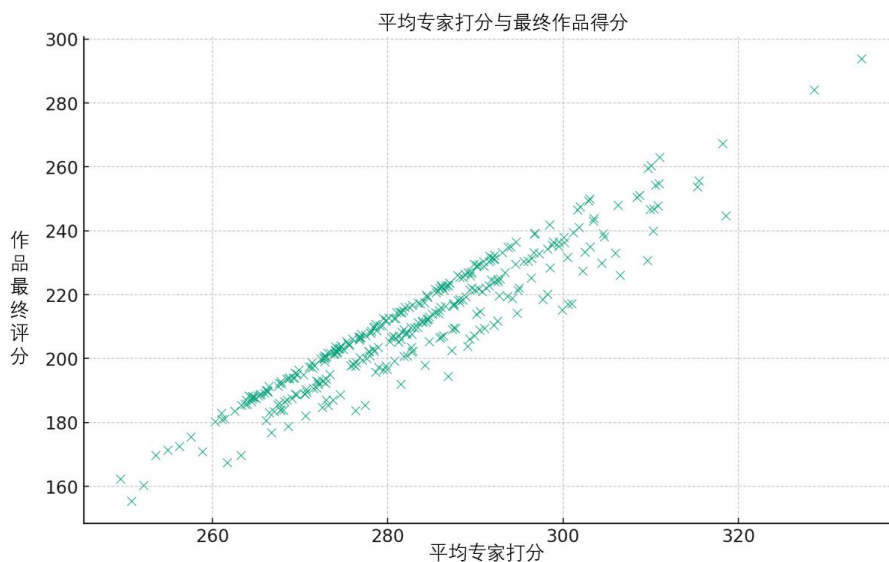


图 5-7 平均专家打分与作品最终得分关系

5.3.6 获奖等级对比分析

在第一阶段结束后，我们根据调整后的作品最终成绩进行了深入的分析。根据一阶段的数据确定作品是否获奖，根据二阶段的数据，确定我们三个获奖等级的比例，最终算出了各个奖项的比例率：一等奖为 1.34%，二等奖为 13.30%，三等奖为 18.51%，未获奖为

66.85%。我们将调整后的作品成绩进行了排序，并确定了新的获奖等级排名。

为了对比每个作品的新旧评分等级，我们使用了混淆矩阵。在这个矩阵中，1、2、3、4 分别代表一等奖到三等奖以及未获奖。Y 轴显示了实际的获奖等级，X 轴显示了调整后计算出的每个作品的获奖等级。混淆矩阵的每个单元格表示两个评分等级的数量。对角线的单元格表示原始评级和调整后的评级相同的作品数量。

通过混淆矩阵的热图表示，我们可以清晰地看出各个等级的分布情况，其中颜色的深浅代表了数量的多少。从分析中我们发现，在 2015 个作品中，经过专家评估后，有 67 个作品的评级发生了变化；而从作品本身的角度评估，有 97 个作品的评级发生了变化。

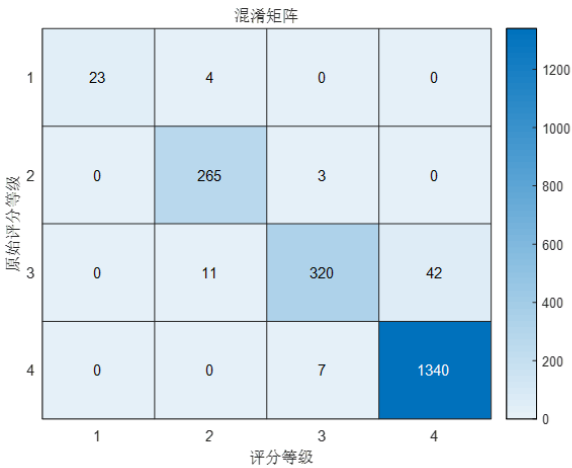


图 5-8 专家角度和原始评分等级混淆矩阵图

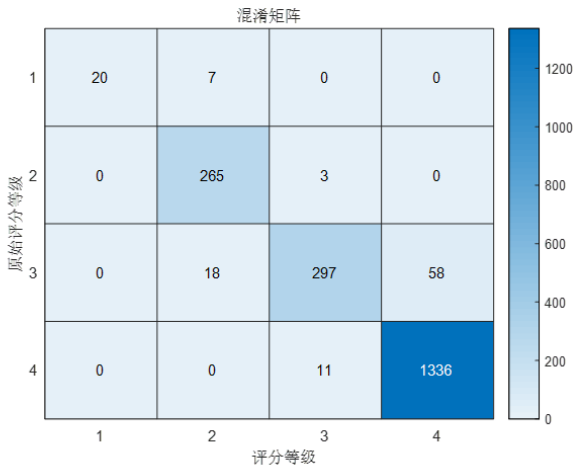
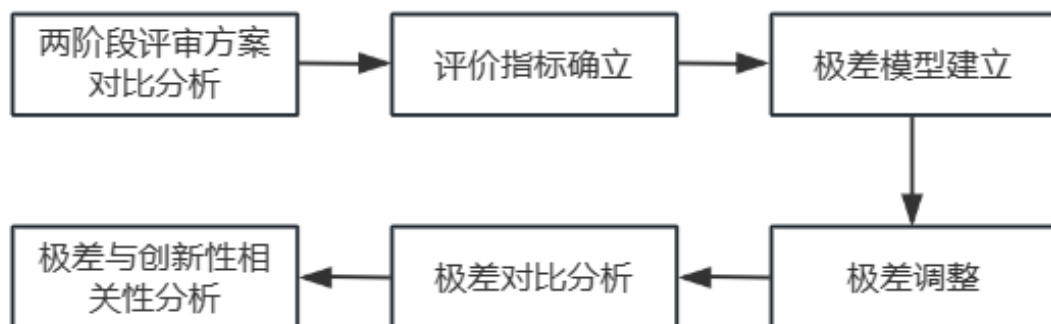


图 5-9 综合角度和原始评分等级混淆矩阵图

六. 问题三：基于随机森林的“极差”模型

6.1 问题描述与分析

在“创新类”大赛中，面临的主要挑战是极差大和评分不一致问题。通过对模拟数据 2.1 和 2.2 的分析，我们观察到在第一阶段评审中，由于专家对作品的不同看法和评价标准，导致了较大的极差。这种极差在一定程度上反映了作品的创新性，但同时也带来了评分的不稳定性。进入第二阶段评审，作品在更严格和权威的评审下，极差有所缩小，但仍然存在部分极差大的作品。这些作品的极差需要通过复议和调整来进行缩小，以确保评分的公正性和准确性。通过建立极差模型，并结合随机森林等机器学习方法，我们可以对非高分且非低分段的作品的“大极差”进行更为精确和自动化的调整，从而更好地发掘和评价创新性较高的作品。问题思路流程如下：



6.2 数据预处理

删除多余特征值：“学校编码”并不直接反映作品的质量或价值，我们同样也不会考虑学校作为影响作品评分的一个因素，为了保证每个作品都得到公正和一视同仁的对待，将“学校编码”这一特征值的样本数据删除。

去除异常值：按照作品的最终得分对所有作品进行排序，得到一个按得分排序的新数据排名。比较每个作品的得分排序与真实排序之间的差异，通过 3-sigma 计算每个作品的排名差异。数据 2.1 阈值：233.0-247.56，数据 2.2 阈值：611.65-2385.38，超过阈值的数据点被标记为异常值剔除。数据 2.1 没有发现异常值，数据 2.2 剔除的异常值见下表，可以看出被剔除的名次都为 8498，我们猜测名次为 8498 的作品可能被判定为“违规论文”。具体异常值见下表：

表 6-1 数据 2.2 异常值

最终成绩	名次	奖项	学校编码
331.48	8498	未获奖	S513
290.3	8498	未获奖	S374
286.91	8498	未获奖	S202
278.34	8498	未获奖	S142
277.75	8498	未获奖	S069
277.64	8498	未获奖	S499
274.38	8498	未获奖	S269
273.32	8498	未获奖	S166
271.63	8498	未获奖	S231
270.11	8498	未获奖	S149
268.22	8498	未获奖	S269
266.65	8498	未获奖	S142
266.08	8498	未获奖	S269

剔除低分段：低分段属于淘汰范围，我们使用 1-sigma 原则来确定低分段的阈值。在一个正态分布的数据集中，大约 68%的数据点会落在平均值的一个标准差之内。我们可以认为低于平均值减去一个标准差的数据点属于低分段，应该被剔除。

$$z_{score} = \frac{x - \mu}{\sigma}$$

其中， x 是数据点的值； μ 是均值； σ 是标准差。

通过计算我们选择负的 z-score 来确定低分段的样本。则 $z < -2$ 作为低分段的标准，最终数据 2.1 剔除了 37 个低分段作品，数据 2.1 剔除了 289 个低分段作品。

筛选极差偏大作品：我们可以使用正的 z-score 来确定极差偏大的样本。我们可以选择 $z > 2$ 作为极差偏大的标准，筛选出新的数据集。

6.3 数据探索与分析

6.3.1 两阶段评审方案对比

在基于问题三给定的数据 2.1 和数据 2.2 进行分析后，我们对第一阶段和第二阶段的

极差和最终成绩进行了深入探讨。我们根据数据 2.1 和数据 2.2 中第一阶段每个专家的标准分，计算得出第一阶段的综合得分，并进行了详细对比。

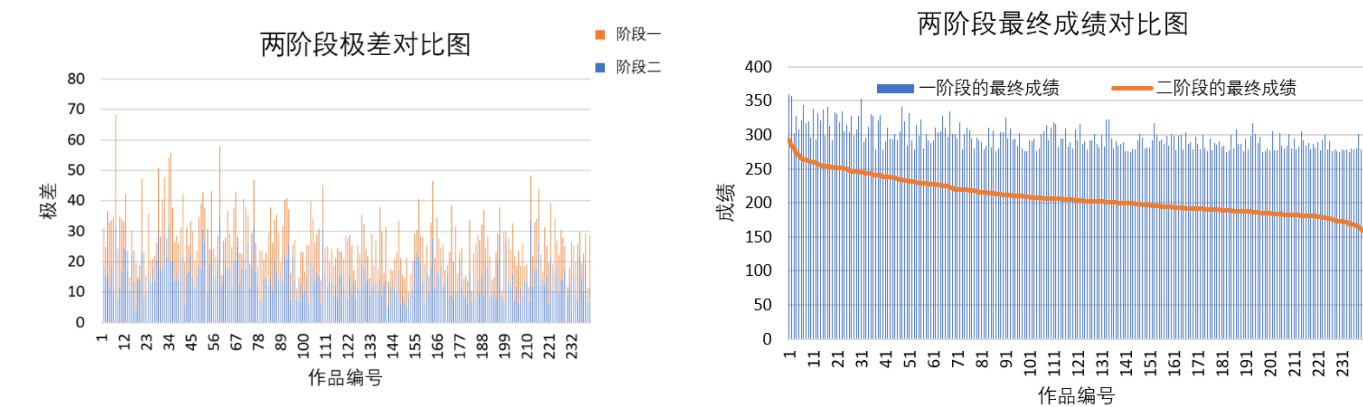


图 6-1 数据 2.1 两阶段极差和最终成绩的对比图

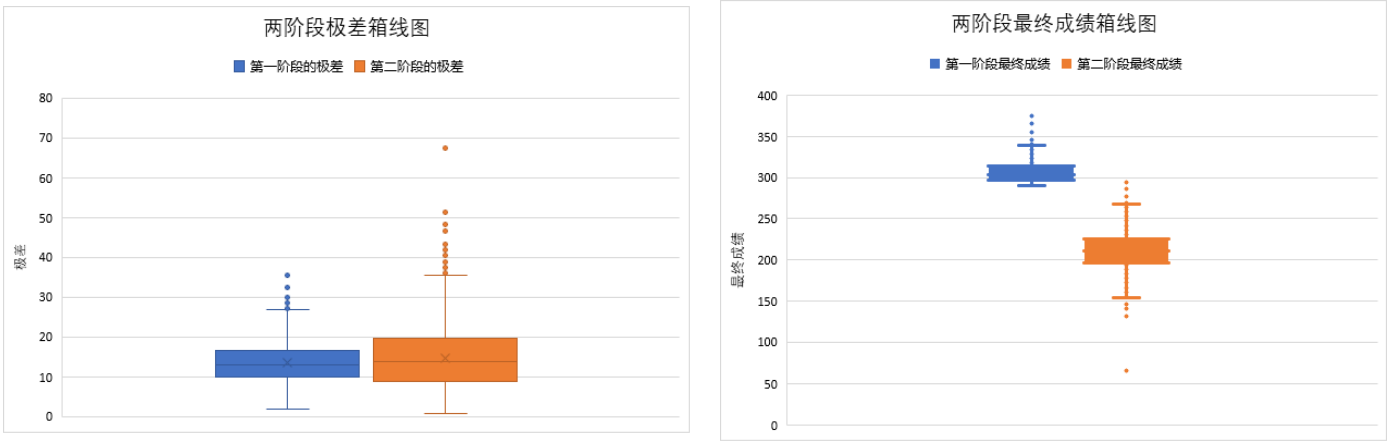


图 6-2 数据 2.2 两阶段极差和最终成绩的箱线图

图 6-1 是数据 2.1 两个阶段的极值和最终成绩的对比图，第二阶段的极差整体较第一阶段小。这显示第二阶段的评分标准更为一致，评审间的分歧较小。同时，第二阶段的最终成绩整体也低于第一阶段，且变化幅度更大。这可能暗示第二阶段的评审标准更为严格，或者作品的质量在第二阶段被更明确地区分出来；

图 6-2 是数据 2.2 两个阶段的极值和最终成绩箱线图，第二阶段的箱体较宽，展示了一个更广泛的数据分布。这表明在第二阶段，作品的评分或者评分标准的差异更大。尽管第二阶段的极差中位数仅略低于第一阶段，但最终成绩的中位数却大幅下降。这进一步强调了第二阶段评审的严格性和对作品质量的细致判断。须线的展示也表明，不论是极差还是最终成绩，第二阶段的数据分布都比第一阶段广，再次证实第二阶段评分的多元性和差异性。此外，第二阶段的异常值更多，可能表示第二阶段的作品具有更高的创新性。

6.3.2评价指标确立

我们继续沿用问题二设计评审方案采用的评价指标：专家评分标准差 SD_j 、评审作品数量、极评指标 ESI ，以及作品的标准差和均值。

在问题三我们需要考虑作品的创新性。为了实现这一点，我们需要分析每个作品的评分分布，若一个作品的评分呈现出广泛的分布（即部分专家给出了高分，而其他专家给出了低分），这可能意味着该作品具有争议性。这种争议性可能源于作品包含的一些创新元素，这些元素并不被所有人接受。

在分析作品评分分布时，我们通常利用一系列统计指标来量化评分的分布特性。在这里，我们决定计算以下基本统计指标：平均分、标准差、中位数、四分位距和偏度，以帮助我们更好地理解评分分布。

平均分：提供了所有专家对作品的平均评分信息。

标准差：揭示了评分数据的分散程度。一个较大的标准差表明评分之间的差异较大，这可能意味着对作品的看法存在分歧。

中位数：描述了评分数据的中心趋势。

四分位距：展示了评分数据的分散程度。

偏度：揭示了评分数据的不对称性。正偏度表明大部分评分较低，而负偏度表明大部分评分较高。

$$Skewness = \frac{n\sqrt{n-1}}{n-2} * \left(\frac{\sum(\text{评分} - \text{平均分})^3}{(n * s^3)} \right)$$

其中， n 是专家评分数量， s 是标准差。

为了量化作品的创新性，我们进一步构建了一个创新性评价指标的公式，综合考虑了上述统计指标，从而更全面地评估每个作品的创新性。

$$\text{创新性评价指标}(I) = \omega_1 * \left(\frac{\mu}{\sigma} \right) + \omega_2 * |Skewness| + \omega_3 * \left(\frac{Max}{Q_3} \right)$$

其中， ω_1, ω_2 和 ω_3 是权重系数， μ 是平均分， σ 是标准差， $|Skewness|$ 表示偏度的绝对值， Max 是最高分， Q_3 是四分位距。 μ/σ 是标准差与平均分的比值，这个比值可以反映评分的变异程度。 $|Skewness|$ 是评分偏度的绝对值，它可以反映评分分布的不对称性。 Max/Q_3 是四分位距与最大分的比值，这个比值可以反映评分的分散程度。

我们通过权重的灵敏度分析，尝试不同的权重组合，并计算每个组合下的创新性指标。

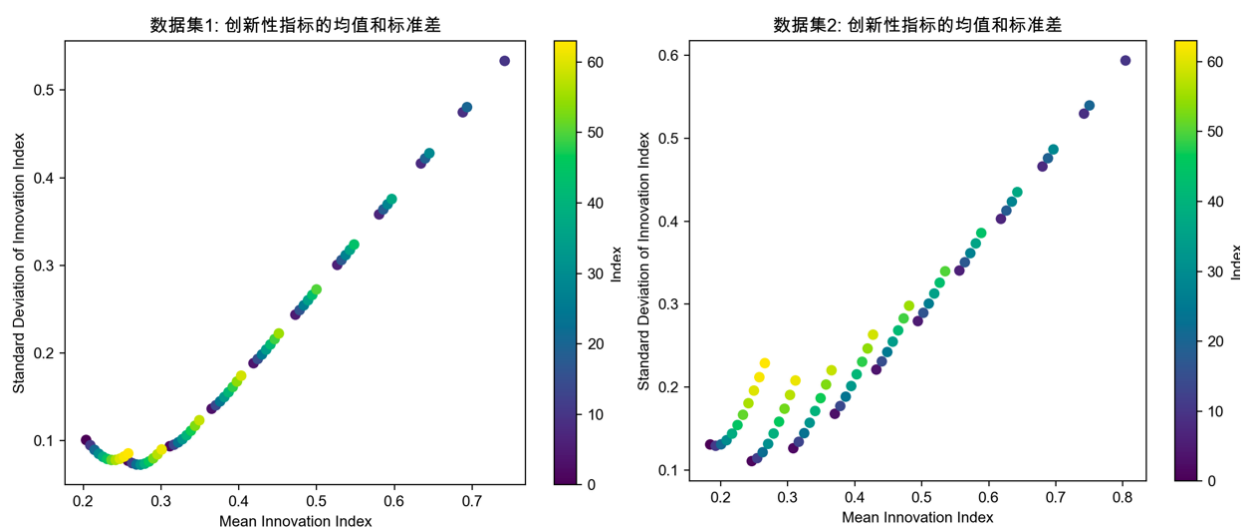


图 6-3 创新性指标的均值和标准差

上图分别展示了两个数据集下不同权重组合对创新性指标的影响。颜色代表了不同的权重组合索引，横坐标代表创新性指标的均值，纵坐标代表创新性指标的标准差。通过这个分析，我们可以选择一种权重组合，使得创新性指标的均值和标准差都较高，这意味着我们的创新性指标能够捕捉到更多的变异和差异。

考虑到有两个不同的数据集，我们选择在两个数据集上都表现较好的权重组合，即创新性指标的均值和标准差都排在前 25% 的权重组合，并查看它们的分布。最终我们选择具体的权重组合来计算最终的创新性指标： $\omega_1 = 0.1, \omega_2 = 0.7, \omega_3 = 0.2$ 。

$$\text{创新性指标}(I) = 0.1 * \left(\frac{\mu}{\sigma}\right) + 0.7 * |\text{Skewness}| + 0.2 * \left(\frac{Max}{Q_3}\right)$$

6.4 建立极差模型

6.4.1 模型介绍

(1) 线性回归

线性回归是一种监督学习算法，用于预测一个连续变量（因变量或响应变量）基于一个或多个预测变量（特征）的值。它假设因变量和预测变量之间存在线性关系。在简单线性回归中，只涉及一个自变量和一个因变量，它们之间的关系可以用一条直线表示。多元线性回归则涉及两个或更多的预测变量。线性回归的主要目标是找到最佳拟合直线，使得实际观测点和预测点之间的差的平方和最小。

(2) 决策树

决策树是一种采用树形结构对样本进行分类的算法。决策树由节点和有方向边组成，其中节点分为非叶子节点和叶节点两种类型。决策树的每个非叶子节点代表在一个特征上

的检测，决策树的每个分支代表了一个输出结果，而决策树的每个叶节点则代表了一个类别。使用决策树对测试样本进行分类时，从根节点开始，根据非叶子节点所代表的特征，对测试样本中相应的特征进行检测，并根据输出结果选择相应的分支，将样本分配到其子节点。之后，继续向下移动，直至达到叶节点，则叶节点所代表的类别就是测试样本的分类结果。

（3）随机森林

随机森林算法实质上是对决策树算法的一种改进，是通过多个决策树进行分类的算法。单个决策树的分类能力有限，达不到理想的分类效果，因此随机森林通过随机的方式生成大量的决策树，并通过投票法将这些决策树结合起来，从而对这些决策树的分类结果进行统计，获得最可能的分类结果。

使用随机森林实现分类，首先要使用训练集构建随机森林分类器，然后使用构建的随机森林分类器对测试数据进行分类。随机森林的分类过程大致如下：

- 1.采用自助抽样法，从 n 个训练样本中有放回地随机抽取 n 个样本作为一个决策树的训练集，并将未被抽到的样本作为该决策树的袋外数据。重复执行 N_{tree} 次，得到 N 个训练集和袋外数据；

- 2.使用得到的 N_{tree} 个训练集构建 N_{tree} 个决策树；

- 3.对于决策树的每个节点，从 m 个特征中随机抽取 N_f 个特征作为候选特征，然后从 N_f 个候选特征中选出一个最佳的分裂特征；

- 4.每个决策树最大限度地进行分裂，直到节点中的训练样本都属于同一类别；

- 5.将生成 N_{tree} 个决策树组成随机森林，对于输入的测试样本，随机森林中的每个决策树会进行投票，得票数最多的类别则为该测试样本的分类结果。

（4）XGBoost(Extreme Gradient Boosting)

XGBoost 是一个优化的梯度增强机器学习算法，它结合了弱学习器（通常是决策树）来形成一个强学习器。与其他梯度增强方法相比，XGBoost 在其损失函数中包括正则化项，有并行处理能力，内置交叉验证，可以处理缺失值，并提供了对树的复杂性的正则化。因其高效、灵活和出色的预测能力，XGBoost 在多种机器学习任务中都广受欢迎。

6.4.2模型训练与评估

输入评价指标：专家评分标准差 SD_j 、评审作品数量、极评指标 ESI 、作品的标准差和

均值，以及创新性指标。

数据集划分：从进入第二阶段的作品数据集中，我们选取 70%的数据作为训练集，用于模型的训练；剩余的 30%作为测试集，用于评估模型的性能。我们选择了四个不同的模型进行训练。

训练策略：对于每一个筛选和划分好的数据子集，我们将其分别输入到四个模型中进行训练。这意味着，每个模型会针对每个评分等级的数据进行训练，从而得到针对不同评分等级的四个独立的模型。

模型评估：在进行模型选择时，评估指标的选择至关重要。本次分析将考虑以下几个主要评估指标：分别是 R^2 (决定系数)、MAE（平均绝对误差）、MSE（均方误差）、RMSE（均方根误差），每个指标都通过具体的数学公式进行计算，如下所示：

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$RMSE = \sqrt{MSE}$$

针对数据 2.1 和数据 2.2, 我们应用了四种不同的机器学习模型：随机森林、线性回归、决策树和 XGBoost。每个模型都被评估和比较基于上述指标，结果如下表所示。

表 6-2 数据 2.1 四个模型评价指标

<div>模型 指标</div>	随机森林	线性回归	决策树	XGBoost
R^2	0.79	0.74	0.72	0.76
MAE	3.29	3.55	3.75	3.40
MSE	18.28	21.00	23.50	19.75
RMSE	4.28	4.58	4.84	4.45

表 6-3 数据 2.2 四个模型评价指标

<div>模型 指标</div>	随机森林	线性回归	决策树	XGBoost
R^2	0.76	0.71	0.69	0.73
MAE	3.85	4.10	4.30	3.95
MSE	23.47	26.20	28.70	24.95
RMSE	4.85	5.12	5.36	4.99

在详细分析上述结果后，我们可以看到，随机森林在所有模型中展现出最好的性能，具有最高的 R^2 值和最低的 MAE、MSE 和 RMSE，线性回归和 XGBoost 表现相似，但均不如随机森林，决策树在所有模型中表现最差。基于上述分析和评估结果，我们建议选择随机森林作为最优模型，这是因为它在所有评估指标上都表现出了最佳性能。

6.5 极差调整与分析

依据建立的极差模型，将非高分且非低分段作品的数据集输入到上述模型中，对第一阶段中非高分且非低分段的作品的“大极差”进行了调整。针对数据 2.1 和数据 2.2，我们获取了调整后的极差数据，并与原始极差数据进行了深入比较。如下图所示，调整后的极差明显小于原始极差，表明通过模型调整，我们有效降低了各作品间的评分差异，使评分更趋于一致。

图中的误差线基于标准偏差进行绘制，旨在展示每个数据点相对于各自数据集平均值的偏离程度及其分布范围。通过观察可见，调整后的数据的误差线显著缩短。这一现象表现了调整后的极差数据点间的差异减小，评分的一致性得到了显著提升。

此外，更短的误差线也意味着每个作品的评分更加稳定，不容易受到极端值的影响，从而使得整个评审过程更加公平和合理。通过这一系列调整，大大提升了评审结果的准确性和可靠性，为后续的评审决策提供了坚实的数据支持。

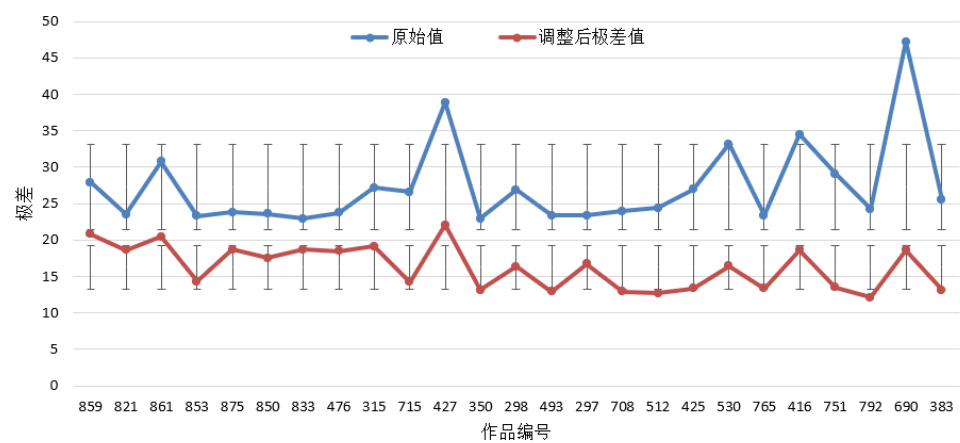


图 6-4 数据 2.1 调整后极差对比图

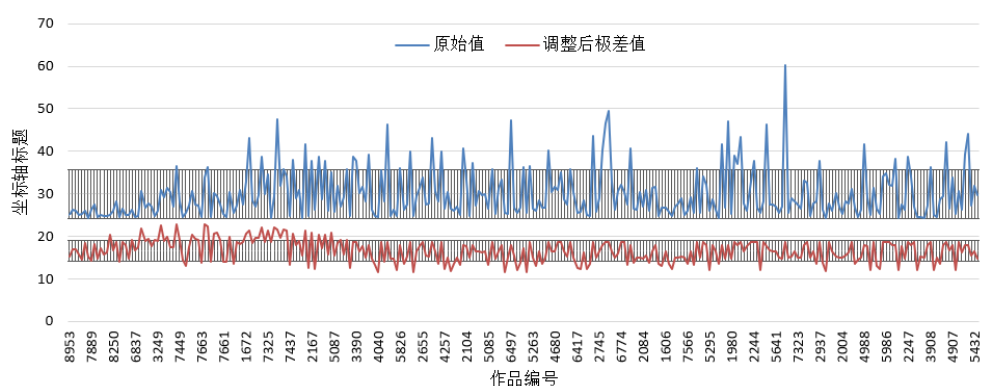


图 6-5 数据 2.2 调整后极差对比图

6.6 极差与创新性相关性分析

对于非高且非低分段的作品来讲，极差与创新性的关系可能不是简单的线性关系，初步推测，作品的成绩在大于一定值的时候，极差和创新性模型成正相关，成绩小于一定值的时候，极差和创新性模型成负相关。这种情况通常可以使用分段回归（或阶梯回归）来分析。在这种方法中，我们将数据分成两部分（基于成绩的阈值），然后分别对这两部分数据进行回归分析。我们采用数据可视化来确定成绩的阈值。

我们根据筛选后数据 2.1 和数据 2.2 的数据进行可视化，画出核密度估计图，见下图。

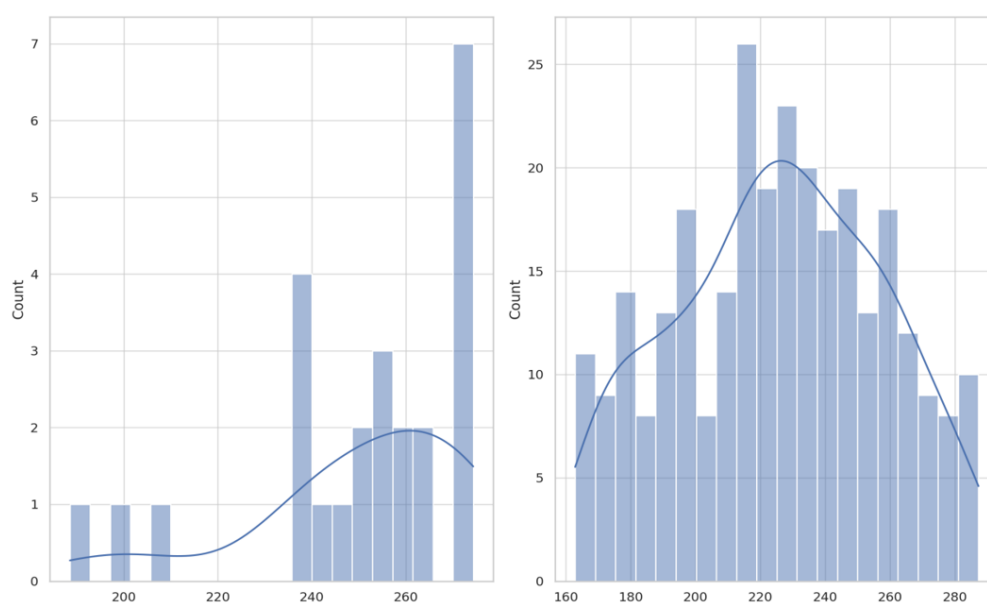


图 6-6 作品成绩 KDE 图

确定的两个数据集的阈值分别是：数据集 2.1 的阈值: 270.24; 数据集 2.1 的阈值: 249.32。使用这些阈值进行分段回归分析，结果如下表：

表 6-4 回归分析相关指数

		创新性的系数	极差的系数	截距	MSE	R ²
数据 2.1	> 270.24	1.38	0.35	266.00	0.94	0.73
	≤ 270.24	-0.45	1.50	219.69	461.70	0.63
数据 2.2	> 249.32	5.13	1.70	250.56	91.39	0.705
	≤ 249.32	3.37	-0.90	218.75	568.25	0.61

表 6-5 创新性和极差的相关性关系式

	阈值	关系式
数据 2.1	> 270.24	成绩=1.38*创新性指标+0.35*极差+266
	≤ 270.24	成绩=-0.45*创新性指标+1.50*极差+219.69
数据 2.2	> 249.32	成绩=-5.13*创新性指标+1.70*极差+250.56
	≤ 249.32	成绩=-3.37*创新性指标-0.90*极差+218.75

数据集 2.1: 对于成绩大于 270.24 的数据: 创新性与成绩呈正相关: 创新性越高, 成绩越高。极差与成绩也呈正相关: 极差越大, 成绩越高。这意味着在高分段, 创新性和极差都是提高成绩的有利因素。

对于成绩小于或等于 270.24 的数据: 创新性与成绩呈负相关: 创新性越高, 成绩越低。极差与成绩也呈正相关: 极差越大, 成绩越高。这意味着在低分段, 创新性不利于提高成绩, 而极差仍然是一个有利因素。

数据集 2.2: 对于成绩大于 249.32 的数据: 创新性与成绩呈正相关: 创新性越高, 成绩越高。极差与成绩也呈正相关: 极差越大, 成绩越高。这意味着在高分段, 创新性和极差都是提高成绩的有利因素。

对于成绩小于或等于 249.32 的数据: 创新性与成绩呈负相关: 创新性越高, 成绩越低。极差与成绩呈负相关: 极差越大, 成绩越低。这意味着在低分段, 创新性和极差都不利于提高成绩。

在两个数据集中, 我们看到了创新性和极差与成绩之间复杂的关系。这种关系依赖于成绩的阈值, 也就是说它依赖于作品的分数。一般来说, 在非高非低的成绩段, 偏高分数的创新性可能有利于提高成绩, 极差也可能是一个有利因素。对于偏低的分数, 这种趋势可能会颠倒。

七. 问题四：评审模型建立与求解

7.1 问题描述与分析

对创新类竞赛，我们需建立一个完整的评审模型以确保每个作品得到公平和全面的评价。在第一阶段，我们利用问题一构建的优化模型来得到初始分配方案和初始评分。接下来，我们使用问题二第一阶段标准分模型来确定标准分，从而计算最终成绩并进行排名，只有排名超过某个阈值的作品才会进入第二阶段。第二阶段中，我们使用问题二的模型来进一步优化评审分配，再利用问题二第二阶段标准分模型得到第二阶段的最终得分。因为第二阶段的分配模型的缺少，本问题重点在于第二阶段模型的建立。目标函数的目的是最大化每个作品的创新性指标和第一阶段中每位专家的极评指数。我们还加入了多个约束条件以确保每份作品只被三位评审专家评审，每位评审专家评审的作品数量是均匀的，以及每份作品所分配的专家专业性是平衡的。我们使用遗传算法来求解这个优化问题。



7.2 两阶段模型建立

（1）第一阶段模型

第一阶段模型与问题一模型相同，此处省略模型公式。从第一阶段模型得到以下指标，分别是： ESI_j 第一阶段专家 j 的极评指数、 ER_i 专家专业性指标以及初始分配结果，利用初始分配结果的专家初始评分和问题二得到的标准分得到最终成绩，再对其进行排名，排名大于 $RATE$ 的作品选择进入第二阶段。

（2）第二阶段模型

① 目标函数

我们的目标是创新性比赛， ESI_j 表示第一阶段专家 j 的极评指数，越大表示专家的偏好越明显，具体目标函数如下式：

$$Max \sum_i \sum_j ESI_j * x_{ij} \quad (7-1)$$

② 约束条件

约束(7-2)表示第二阶段的每份作品恰好被 3 位评审专家评审:

$$\sum_{i=1}^m x_{ij} = 3, \forall i \in \{1, 2, \dots, m\} \quad (7-2)$$

约束(7-3)每位评审专家评审的作品数量应当是均匀的。设定每个评审专家评审的最小和最大作品数为 L 和 U ，其中， L 和 U 根据实际需求来确定， $U = L + a$ ，其中 a 是一个较小的整数，例如 5，以考虑到某些专家可能评审稍微多一些的作品，则:

$$L \leq \sum_{i=1}^n x_{ij} \leq U, \forall j \in \{1, 2, \dots, n\} \quad (7-3)$$

约束(7-4)表示每份作品所分配的专家专业性尽可能平衡:

$$L \leq ER_i \sum_{j=1}^m x_{ij} \leq U, \forall i \in \{1, 2, \dots, m\} \quad (7-4)$$

综上所述，问题四对“创新类”竞赛构建的第二阶段评审模型如下:

$$\begin{aligned} &Max \sum_i \sum_j ESI_j * x_{ij} \\ &s.t \quad \sum_{i=1}^m x_{ij} = 3, \forall i \in \{1, 2, \dots, m\} \\ &\quad L \leq \sum_{i=1}^n x_{ij} \leq U, \forall j \in \{1, 2, \dots, n\} \\ &\quad L \leq ER_i \sum_{j=1}^m x_{ij} \leq U, \forall i \in \{1, 2, \dots, m\} \end{aligned}$$

7.3 模型求解

7.3.1 算法设计

遗传算法是一种优化和搜索算法，它的工作原理是模仿自然界中的进化过程。在遗传算法中，一个解决方案可以被看作是一个“个体”，而一组可能的解决方案构成了一个“种群”。每个个体都有一个与之相关的“适应度”，用于评估该个体解决给定问题的能力。遗传算法通过反复进行以下几个步骤来改进种群中个体的质量：选择（基于适应度选择个体），交叉（或重组，将两个个体组合以产生一个或多个后代），和变异（随机修改个体的部分）。通过多代的进化，种群中的个体将逐渐改进，最终，算法将找到问题的一个近似最优解。遗传算法被广泛应用于函数优化、机器学习、调度问题等多个领域，由于它的全局搜索能力，能够在复杂的搜索空间中找到满意的解。

(1) 解码和编码

针对问题四建立评审模型，首先对当前种群进行非支配排序。本文基因编码由 3000×125 的矩阵构成，每一行的矩阵由随生成的 0~1 的实数填充，如下图所示。

		评审专家						
		1	2	3	124	125	
作品	1	1	0.9	0	0	0.95	
	2	0.1	0.3	1	0.4	0.2	
	3	0.98	0.32	0.95	0.31	0.4	
			
	2998	0.61	0.23	0.99	0.15	1	
	2999	0.11	0.17	0.98	0.62	0.13	
	3000	1	0.25	0.97	0.29	0.78	

图 7-1 染色体编码

根据上述编码结果，每行生成的随机数中，我们选取最大的五个数，并在其对应位置标为 1，其余位置标为 0。本文基因解码由 3000×125 矩阵构成，每一行代表一个作品，每一列代表一个评审专家，如下图所示，矩阵中的 1 表示了该作品被该专家进行了评审，如下图所示，表示作品 1 和作品 3 被专家 1 进行评审，作品 2 和作品 3 被专家 3 进行评审。

		评审专家						
		1	2	3	124	125	
作品	1	1	1	0	0	1	
	2	0	0	1	0	0	
	3	1	0	1	0	0	
			
	2998	0	0	1	0	1	
	2999	0	0	1	0	0	
	3000	1	0	1	0	0	

图 7-2 染色体解码

(2) 交叉

我们采取了经典的交叉方式。首先在父代 1 中随机挑选若干位点并将其转移到子代染色体。接着，从父代 2 中获取子代未包含的部分，并按照原有顺序填补到子代中。

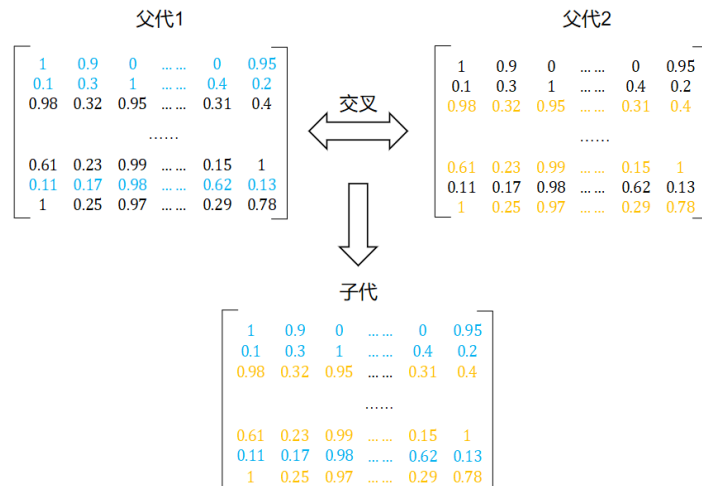


图 7-3 染色体交叉

(3) 变异

变异随机选择染色体位置进行变异。

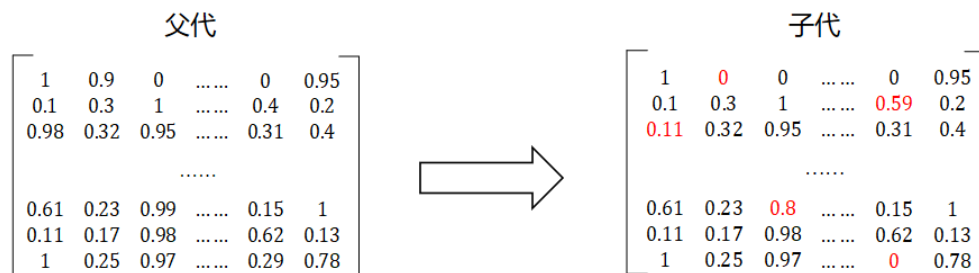


图 7-4 染色体变异

(4) 适应性函数 f

适应性函数是评价染色体（或解决方案）优劣的标准，在这个评审模型中主要用于评估某个分配方案的效果。在这个模型中，适应性函数与目标函数类似，主要是为了最大化创新性指标 θ 和第一阶段中每位专家的极评指数 ESI_j 的乘积之和。

通过这个适应性函数，我们可以评估每个染色体的好坏，并在遗传算法的每一代中选择适应度较高的染色体进行交叉和变异，以生成新的解决方案，并逐渐找到最优的作品分配方案。

$$f = \sum_i \sum_j ESI_j * x_{ij} \quad (7-5)$$

7.3.2 算法步骤

基于遗传算法求解评审分配问题：

步骤 1：初始化种群，生成随机染色体。

步骤 2：解码染色体并计算每个染色体的适应度值。

步骤 3：根据适应度值，选择种群中适应度最高的一半作为父代。

步骤 4：使用父代通过交叉和变异技术生成子代，并更新种群。

步骤 5：检查是否达到最大迭代次数。若已达到，则转至步骤 6；否则，返回步骤 2 继续迭代。

步骤 6：解码并输出最佳染色体及其适应度值，显示评审分配结果。

遗传算法具体流程图如下所示：

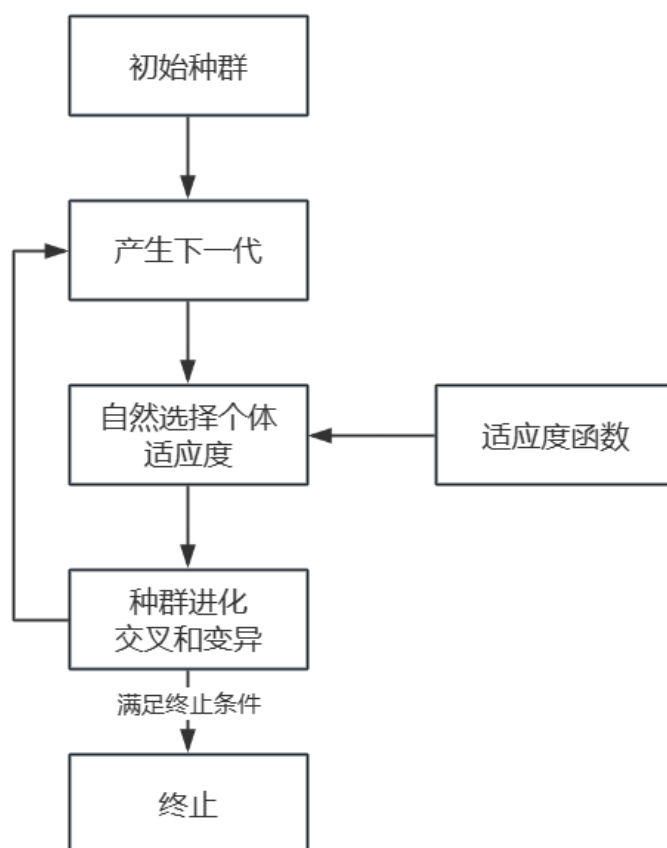


图 7-5 遗传算法流程图

7.4求解结果与分析

7.4.1实验设计

我们以问题一的背景数进行计算，前 30%的数据进入第二阶段，即前 900 份作品进入第二阶段；选择问题二背景，利用问题二中的专家专业性指标，选择专业性排名前 30 位专家的编号数据和专业性水平数据，如下表所示：

表 7-1 专家编号和专业性水平表

专家编号	专业性水平	专家编号	专业性水平
P487	117.6048649	P493	116.1066641
P285	117.4714296	P270	116.0864242
P263	117.2564183	P437	115.9967272
P308	117.1519343	P370	115.8976919
P497	117.1022252	P324	115.8161278
P371	117.0212161	P063	115.7258017
P364	116.8987788	P504	115.6235972
P255	116.7099178	P666	115.5803019
P115	116.6866203	P443	115.5536071
P476	116.6834812	P399	111.0703692
P283	116.6425496	P075	110.3700983
P422	116.6367701	P357	110.043683
P079	116.2821023	P143	110.0241887
P344	116.2173739	P299	109.9815622
P311	116.2057739	P405	109.9724862

7.4.2求解结果

得到具体分配方案结果和创新性比赛的迭代图，其中，创新性迭代 50 次左右达到最优解，为 2654.73。

表 7-2 创新性比赛第二阶段分配方案结果表

作品编号	专家 1	专家 2	专家 3
1	P487	P497	P115
2	P311	P079	P283
.....			
899	P364	P308	P283
900	P364	P285	P422

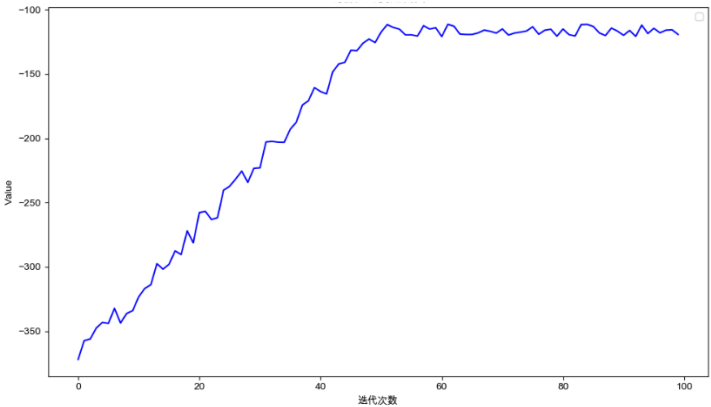


图 7-6 遗传算法迭代图

八. 模型评价与改进

8.1 模型和算法的优点

(1) NSGA-II

NSGA-II（非支配排序遗传算法 II）是一种用于解决多目标优化问题的遗传算法。NSGA-II 使用快速非支配排序方法，可以有效地找到 Pareto 前沿，从而得到多个优秀的候选解、保持种群多样性、具有较强的全局搜索能力、广泛应用、可以同时考虑多个目标函数进行优化；NSGA-II 涉及多个目标和多个约束条件时，计算复杂性可能会增加、缺乏确保收敛性的显式机制。

(2) 线性回归

线性回归是一种统计学方法，用于建模和分析两个变量之间的关系。其目的是找到最佳拟合直线，描述这种关系。实现简单，易于理解和解释，计算效率高，可以用于实值和连续的输出预测；但对异常值敏感，不能捕捉复杂的数据模式。

(3) 随机森林

随机森林是一种集成学习方法，通过整合多个决策树的预测来改善模型的准确性和稳定性。可以处理大量的输入变量，能够处理缺失值，不容易过拟合，能够很好地处理非线性关系；在大数据集上可能会很慢，需要足够的树来确保模型性能，可能导致大量的计算。模型和算法的缺点。

8.2 模型改进和展望

- 1.使用更高效的算法:考虑使用更高效和先进的算法或模型，例如深度学习或增强学习算法，以提高模型性能。
- 2.特征工程:进一步进行特征选择和特征工程以提高模型的精度和效率。可能包括更深入的特征选择、创建新的特征或使用特征降维技术。
- 3.调优和正则化:通过更精细的超参数调整和使用正则化技术来避免过拟合，提高模型的泛化能力。
- 4.集成学习或堆叠模型:利用集成学习方法或模型堆叠来提高模型的稳定性和精度。

参考文献

- [1] Holland, J. H. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press. Cui Y D, Huang B X. Reducing the number of cuts in generating three-staged cutting patterns [J]. *European Journal of Operational Research*, 2012, 218: 358-365.
- [2] Puchinger J, Raidl G R. Models and algorithms for three-stage two-dimensional bin packing [J]. *European Journal of Operational Research*, 2007, 183(3): 1304-1327.
- [3] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- [4] 杨松, 邓斌, 杨绍棠. (2012). 随机森林及其在数据挖掘中的应用. *统计与决策*, 3, 81-84.
- [5] 张铃铛, 周志强. (2006). *遗传算法:理论、应用与软件实现*. 清华大学出版社.