



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

南京林业大学

23102980125

参赛队号

1.徐溢琪

队员姓名

2.张志伟

3.王昕

---

# 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的建模与研究

## 摘 要：

随着科技的发展和教育制度的改革，近年来涌现出一批以“创新”为主题的竞赛项目。这类竞赛的运行模式为，参赛队伍提交文档、视频或幻灯片等文本形式的作品，专家对参赛队伍提交的作品评阅判分，一份作品将由多位专家独立进行评阅打分，按照具体的赛事细则，选出其中质量好的作品，最终制定获奖名单。这类竞赛规模大、创新性强，容易出现因作品分配不合理、专家个人因素引起评分的系统性和随机误差等问题。若不能减小或解决上述问题，将直接影响到最终获奖名单，从而影响竞赛的公平性。本文为，而探究一套科学、合理、有效的竞赛评分机制。该评分机制考虑到了作品科学分配、专家个体打分与整体的一致性关系及评审等因素。

针对问题一，为制定一份专家评分可比性强且不同专家与专家之间评分可比性尽可能一致的作品分配方案，本文以多目标优化的思想为基础，综合考虑专家评阅论文的工作量的均衡量和专家评阅论文的交叉量这两个因素，将论文的分配量和论文的交叉量作为方案指标，建立均匀-交叉多目标作品分配模型。为求解该模型，本文提出一种基于 NSGA-II 算法的多目标遗传进化算法（Multi-Objective Genetic algorithm, MO-GA 算法）。该算法不受模型参数规模的限制，有较强的自适应性，能自主协调各目标函数之间的关系，得到 Pareto 最优前沿上的多样性解集。针对问题一中的参数限制（3000 支队伍和 125 位评委），给出基于 MO-GA 算法的均匀-交叉多目标作品分配模型的具体实施步骤。结果表明，基于该模型的最优作品分配量为每位专家 120 份，最优交叉量为 1.92。

针对问题二，首先对每份作品的原始分数和标准分进行描述性统计分析，原始数据的变异系数为 0.292，判断存在异常值，经过清洗得到筛选后的数据。然后，利用 Shapiro-Wilk 检验、P-P 图和 Q-Q 图对全体作品的原始分数和标准分进行正态性检验，结果总体上符合正态分布。进一步绘制专家打分箱型图，并通过 K-means 聚类分析了专家打分习惯，其中打分偏高、平均和偏低的专家比例分别为 15.46%、79.38%和 5.16%。考虑到专家作品集合的学术水平分布不同，本文选择现有模型(2)并提出了基于一次评审极差的标准分评阅模型。以获奖总人数为基准，采用名次的相对变化量和相对变化率作为评价指标，结果显示基于一次评审的模型较现有模型在提升相对变化量和变化率方面分别高出 73.9%和 4.6%。此外，本文还通过 Pearson 相关性分析证实了名次与标准分均值之间存在显著负相关关系，结果为-0.947。考虑到专家打分习惯和标准分模型对排名的影响，本文从调整标准化分数计

---

算系数和基于离差改进专家权重两个角度出发,建立了基于离差的权重标准分全局优化的评审模型。并采用重合度和乱序度作为模型评价指标,以第二阶段评审中专家协商一致认定的一等奖排序作为基准。该模型的重合度和乱序度分别为 9 和 122。在此基础上,本文进一步提出了加权系数改进模型,该模型的重合度和乱序度分别为 14 和 83,改进模型的重合度提高了 55.56%,重合度降低了 32.97%。结果表明,改进后的模型能有效减小乱序度和提高匹配度,更好地反映专家共识,验证了模型的效果。

针对问题三,分别分析了比较模拟数据 2.1 和 2.2 进入和未进入第二阶段评审的两组数据,共得到 4 组对比数据。在总体打分不变的情况下,两阶段评审方案的数据分布更加集中,评分结果更优。对两阶段的数据中的极差、名次、最终成绩、一阶段平均标准分等指标通过 Kendall 系数进行一致性检验,得协调系数为  $W = 0.802$ ,相关度完全一致。为区别高低分段的极差与创新性差异,选出创新性作品,本文引入作品争议度,以作品创新度为衡量作品是否具备创新性的指标,建立基于极差-争议度的作品创新度评价模型。利用 K-means 并结合附件 2.1 和 2.2 数据,对作品创新度进行聚类,提出基于创新度的二轮分数修正模型,用于对极差中等的作品进行极差修正。

针对问题四,本文构建了一个完整的评审模型,包含以下四个步骤:第一步,建立一个多目标优化的评审专家分组模型,目标是 최소화总体出错率和最大化评审速度。第二步,对作品进行密码加密,采用编程随机生成作品序列号,以保障作品评审过程中的公平公正。第三步,在问题一的作品分配模型基础上,增加专家回避本校作品的约束,得到改进的作品分配模型。第四步,采用问题二中改进后的方案三作为作品评审模型。通过这四步得到完整的评审模型和流程。最后,本文从评分机制、作品分配、评委选择等角度分析现有方案的不足,并给出目标化的改进建议,提出未来需要收集具有论文编码的分发方案、论文分发过程中的时间损耗、已有创新类竞赛的创新性打分依据等数据以不断优化评审方案。

关键词: 创新类竞赛, 交叉分发, 标准分计算模型, 权重, 多目标优化, 群组决策评分

---

## 目录

一、问题重述 .....	5
1.1 问题背景 .....	5
1.1.1 创新类竞赛发展现状 .....	5
1.1.2 创新类竞赛的共性问题 .....	5
1.2 问题提出 .....	6
二、技术路线图 .....	8
三、模型假设与符号说明 .....	10
3.1 模型假设 .....	10
3.2 符号说明 .....	10
四、问题一模型建立与求解 .....	11
4.1 问题一分析及建模思路 .....	11
4.2 基于均匀-交叉的多目标作品分配模型 .....	11
4.2.1 数据观察与模型的选用 .....	11
4.2.2 模型的建立与求解 .....	11
4.3 模型的求解 .....	12
五、问题二模型建立与求解 .....	16
5.1 问题分析及建模思路 .....	16
5.2 数据的预处理与相关分析 .....	17
5.2.1 竞赛成绩分布情况分析 .....	17
5.2.2 专家打分特点分析 .....	19
5.2.3 作品名次与第一阶段分数的关系分析 .....	21
5.3 基于传统方案的评审模型 .....	22
5.4 基于离差的权重标准分全局优化的评审模型 .....	27
5.4.1 模型的建立 .....	27
5.4.2 模型的求解与方案三分析 .....	29
5.4.3 模型以及方案的改进 .....	31
5.4.4 改进模型的求解 .....	31
六、问题三模型建立与求解 .....	34
6.1 问题分析及建模思路 .....	34
6.2 评审方案优劣比较分析 .....	34
6.2.1 基于模拟数据 2.1 数据的评审方案优劣分析 .....	34
6.2.2 基于模拟数据 2.1 数据的评审方案优劣分析 .....	36
6.2.3 两阶段评审方案相比不分阶段评审方案的优劣分析 .....	37
6.3 基于极差-争议度的论文创新度评价模型的建立与分析 .....	38
6.3.1 基于极差-争议度的论文创新度评价模型 .....	38
6.3.2 创新度的分类 .....	38
6.3.3 基于创新度的修正分数模型 .....	39
七、问题四模型建立与求解 .....	40
7.1 问题分析及建模思路 .....	40
7.2 完整评审模型的建立 .....	40
7.2.1 评审专家分组模型的建立 .....	40
7.2.2 作品加密 .....	41

---

7.2.3 作品分配模型的建立 .....	41
7.2.4 作品评审模型的建立 .....	42
7.3 对现行的评审方案的建议 .....	42
八、模型评估与改进 .....	44
九、参考文献 .....	45
十、附录 .....	46

## 一、问题重述

### 1.1 问题背景

#### 1.1.1 创新类竞赛发展现状

自党的十八大以来，我国一直坚定不移地实施创新驱动发展战略，把科技自立自强作为国家发展的战略支撑。在现阶段建设国家创新体系和实现社会经济转型的背景下，创新人才的培养成为关乎我国是否能实现民族振兴的有力保障<sup>[1]</sup>。为了贯彻落实国家的方针和政策，全国各大高校担任了培养创新人才的重要使命<sup>[2]</sup>，主要体现在引入创新类比赛、开设创新类课程和开展创新类实践项目这三个方面。其中，参加创新类竞赛是快速提升学生能的最有效途径<sup>[2]</sup>。

本文依据《全国普通高校大学生竞赛目录榜单》和中国研究生创新实践系列大赛官网（数据部分来源 [www.csadge.edu.cn](http://www.csadge.edu.cn) 和 <http://www.cahe.edu.cn>），分析 2021 年和 2022 年全国各省参加创新类竞赛的参赛情况，下图 1.1 分别为 2021 年和 2022 年全国各省高校参加创新类竞赛热力气泡图，基于图 1.1 全国各省市参加创新类竞赛的高校百分比数量逐年递增，其中江苏省、浙江省、广东省，为全国各省参赛情况排名前三。由图可知，创新类竞赛备受关注，且关注度逐年递增。

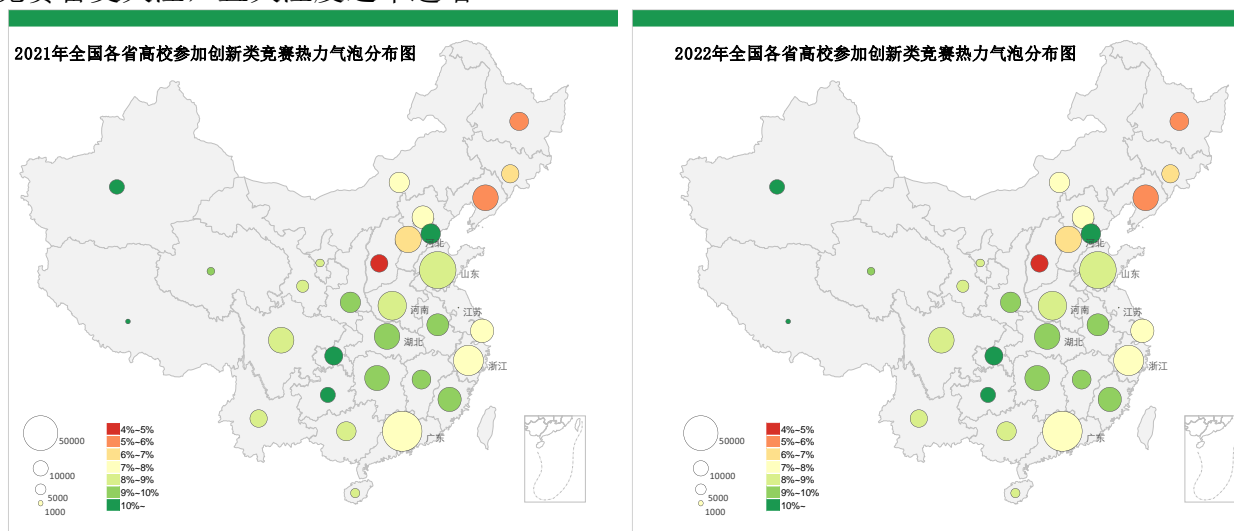


图 1.1(a) 2021 年全国各省高校参加创新类竞赛热力气泡图

图 1.1 (b) 2022 年全国各省高校参加创新类竞赛热力气泡图

图 1.1 近年来全国各省高校参加创新类竞赛热力气泡图

#### 1.1.2 创新类竞赛的共性问题

创新类竞赛没有标准答案，需要评审专家根据命题人（组）提出的评审框架（建议）独立评审。所以，对同一份作品，不同评委的评分可能存在较大差异。实际上，当竞赛规模庞大，评委人数众多时，评分存在极大波动的问题更为突出。显然，简单依据多个评委评分的总和进行排名不是一个好的竞赛评审方案。

创新类竞赛评审主要面临以下九个问题：

1、评审方案设计与优化问题。评审方案的设计直接影响竞赛的公平性和科学性。需要考虑作品的合理分配机制、评委组成、评分标准与机制设计等。优化评审流程，提高评审效率也至关重要。

2、评分差异与评审结果可比性问题。不同评委对同一作品可能存在较大的打分偏差。

---

这将影响评审结果的可比性。需要研究方法来检验评分的一致性,增强不同评委结果的可比性。

3、评分标准化方法问题。传统标准分计算方法在创新类竞赛中存在缺陷。需研究适合创新类竞赛特点的评分标准化方法,降低评分波动。

4、两阶段评审机制有效性问题。许多竞赛采用两阶段评审,但其效果和必要性有待验证。需对比研究两阶段与单阶段评审的优劣。

5、极差问题的合理处理。创新类竞赛中评分极差问题突出。需要研究在不同评审阶段对极差问题的合理处理,提高评审公信力。

6、创新性评估客观性问题。创新性评估主观性大,需要增强评估的客观性。可构建包含多个指标的评估体系,引入更多定量分析。

7、评审数据建模与分析。应用数据挖掘与建模方法,能提高评审的科学性与准确性。需要研究适合的建模方法。

8、专家协商机制的设计。专家协商能有效解决评分分歧。需要设计高效的专家协商流程与机制。

9、评审方案合理性评估。需要构建评估模型,对不同评审方案的优劣进行比较分析,选择最优方案。

因此,探讨大规模创新类竞赛评审方案的公正性、公平性和科学性意义重大。

## 1.2 问题提出

本文主要研究大规模创新类竞赛评审方案研究。为探索大规模创新类竞赛评审的好方法,本文主要从作品“交叉分发”模型、标准分计算模型、基于极差的论文创新性研究和合理的评审模型进行分析研究。

问题一:“均匀-交叉分发”模型的研究

通常,作品在不同的评审阶段都是随机分发给多位评委进行独立评审。不同专家的评审作品集合的出现作品交集可以增加不同专家所给成绩的可比性。但是需要对作品交集进行合理化控制。依据专家工作量和专家评阅论文交集量,建立作品“均匀-交叉分发”模型。在给定 3000 支参赛队、125 位评委、每份作品由 5 位评委评审的前提要求下,求解最优的“交叉分发”方案,使不同评委之间评分具有最大可比性。

问题二:标准分评审方案及标准分计算模型研究

大规模创新类竞赛中,不同评委评审的作品样本有限,传统标准分评审方案的假设可能不成立。为此,需选择新的评审方案,分析专家原始评分和标准分的分布特征。然后,设计适用于大规模创新类竞赛的新的标准分计算模型。本文从调整标准化分数计算系数和基于离差改进专家权重两个角度出发,建立了整体优化的基于离差权重标准分评审模型。利用第二阶段评审中专家协商一致认定的一等奖样本数据对模型进行验证和改进,采用乱序度和匹配度作为评价指标,使模型结果更具公信力。

问题三:基于“极差”模型的论文创新型挖掘研究

“创新类”竞赛的特点是作品没有标准答案,需要通过创新来解决难题。由于评审专家视角不同,同一作品的评分通常存在较大差异。问题三要求根据所给数据中的所有作品在两个不同阶段的成绩和极差的整体变化,对比分析按照不分阶段的评审方案和分两个阶段的评审方案的参数,来比较两种方案的优劣判断分两个阶段是否有必要。分析极差大和创新性强之间的关系,并建立极差模型,并根据所给数据研究第一阶段非高非低分数段极差较大的作品的程序化处理方法。

问题四:“创新类”竞赛完整评审模型的研究与求解

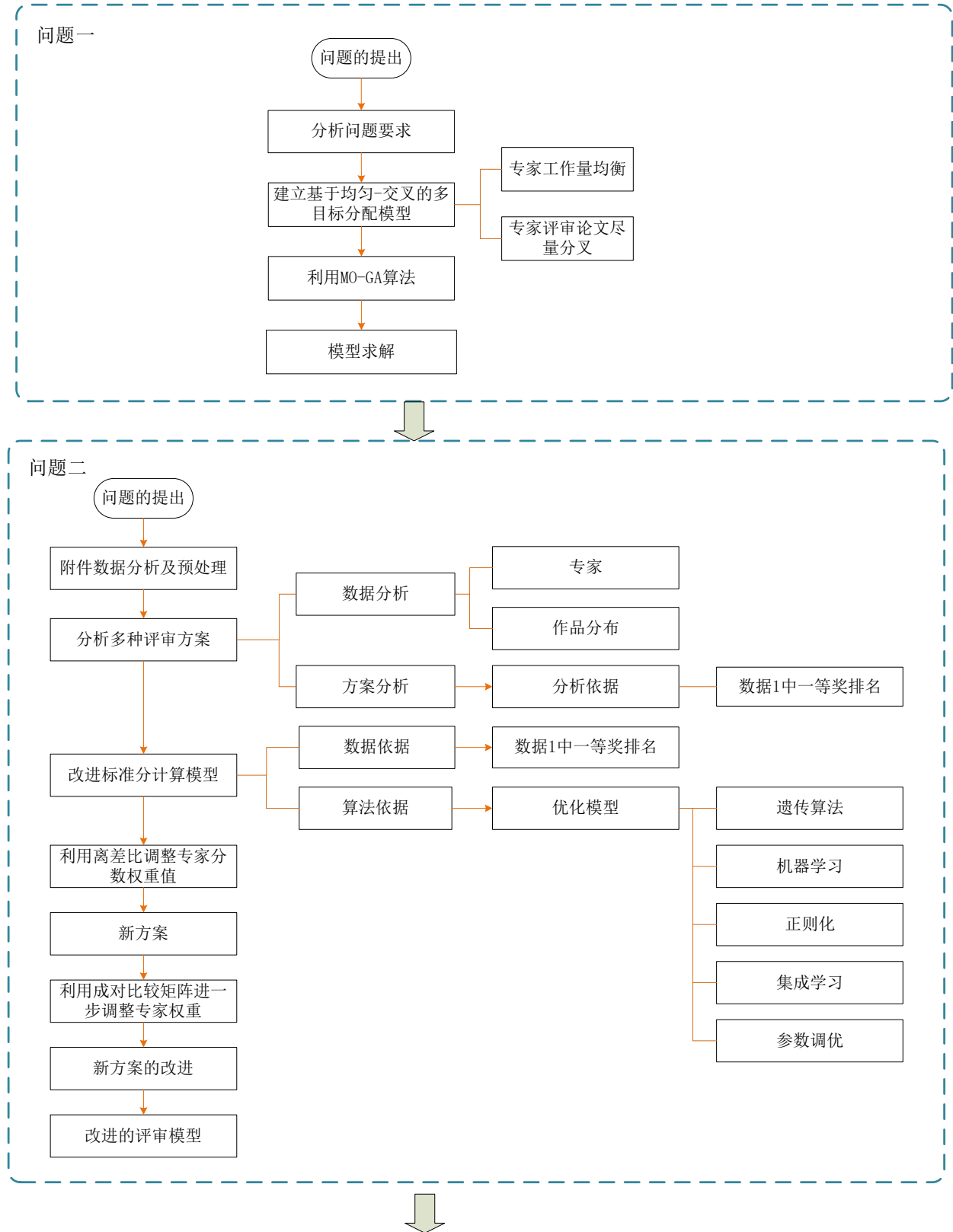
针对创新类竞赛给出一个完整的评审优化模型,并研究如何利用给定数据求解该模型。也可以基于现有评审方案,给出改进建议,包括未来需要收集的数据。

---

由于现在竞赛种类丰富，参赛者的多样性和题目的灵活性使答案多种多样，还存在各位专家打分习惯的差异等各种因素，因此问题四要求针对“创新类”竞赛，提出一个完整的评审模型。针对所给数据对其进行求解，并且对现行的评审方案给出改进的具体建议以及未来需要收集哪些相关数据作为评审的参考



## 二、技术路线图



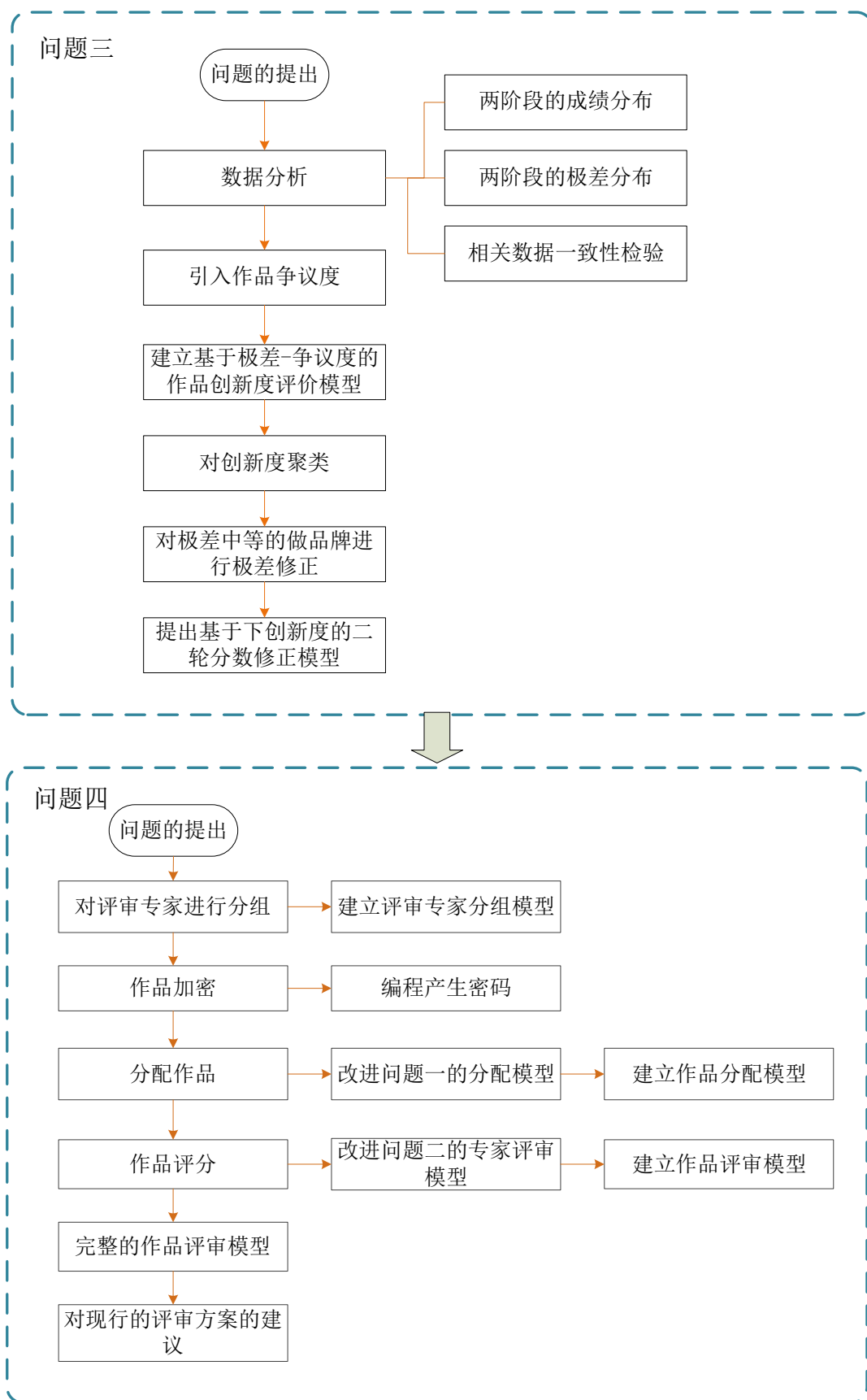


图 2.1 技术路线图

### 三、模型假设与符号说明

#### 3.1 模型假设

##### 全局假设

忽略专家在不同领域的专业性差异，默认每位专家在不同领域的专业度一致。

##### 问题一假设

忽略专家来源地区、所在单位、专业领域以及年龄等客观性的个人因素对分发、评审作品的影响。

##### 问题二假设

各位评审专家都是公正、无私的，阅卷时无感情因素，即对于两篇论文 A 和 B，如果该评委认为 A 较 B 水平高，则有 A 得分高于 B 的得分。

##### 问题三假设

假设作品最终名次越靠前、总分越高，即表示作品的创新性越好。

各位评审专家都是公正、无私的，阅卷时无感情因素，即对于两篇论文 A 和 B，如果该评委认为 A 较 B 水平高，则有 A 的创新性高于 B。

##### 问题四假设

每个专家评审的论文尽可能广泛；每个学校的答卷尽可能平均分给每个专家。

#### 3.2 符号说明

符号	内容说明
$x_{ik}(i \in (1, m))$	第 $k$ 篇作品由第 $i$ 位专家评审
$Q_{ij}$	分给 $i, j$ 两位专家论文的重合数量
$a_i$	每位专家评审作品数量最小值
$b_i$	每位专家评审作品数量最大值
$x_{ik}$	若评委 $i$ 评审了 $k$ 号作品，则 $x_{ik} = 1$ ，否则 $x_{ik} = 0$ 。
$a_{ij}$	第 $j$ 位专家对第 $i$ 份作品的原始评分
$a_{ij}'$	第 $j$ 位专家对第 $i$ 份作品的改进标准分
$\Delta a_0$	改进后的标准分模型调整后的分数的平均分
$w_s$	第 $j_s(1 \leq s \leq k)$ 个专家的权重
$c_i$	第一阶段的综合总分
$d_i$	第二阶段得分
$z_i$	总成绩
$q_{ij}$	评委 $i, j$ 交叉论文的均分

## 四、问题一模型建立与求解

### 4.1 问题一分析及建模思路

一般情况下，大规模的创新类竞赛中，参与评审工作的专家数目受到多方条件约束，远远少于参赛作品的总数目，但每份作品需要经过多位专家评审打分，因此存在专家与专家之间存在作品重叠的情况。现实情况中的分发作品策略通常是采取随机策略，随机分配策略导致任意两位专家评审的作品交集存在大小差异，会影响不同评审专家所给成绩的可比性。问题一在参赛队伍数量、专家数量和每位作品探寻一种“交叉分发”方案，

据问题一的要求，需针对 3000 支参赛队和 125 位评审专家通过建模、分析确定最优的“交叉分发”方案并讨论该方案的有关指标和实施细节。

竞赛的目的不仅是测试参赛者对知识的掌握和应用能力，而且需要进行优胜劣汰即要进行选拔择优，因此在决策时必须做到科学、公正。但是由于选题的多样化和评委进行评分的标准也存在差异，难免产生评分系统误差和随机误差。此外，由于工作量巨大、时间紧迫等原因，每位评审专家不可能审阅每篇论文。为了使决策更加科学化，将每篇论文随机分发由几位评委逐一打分。

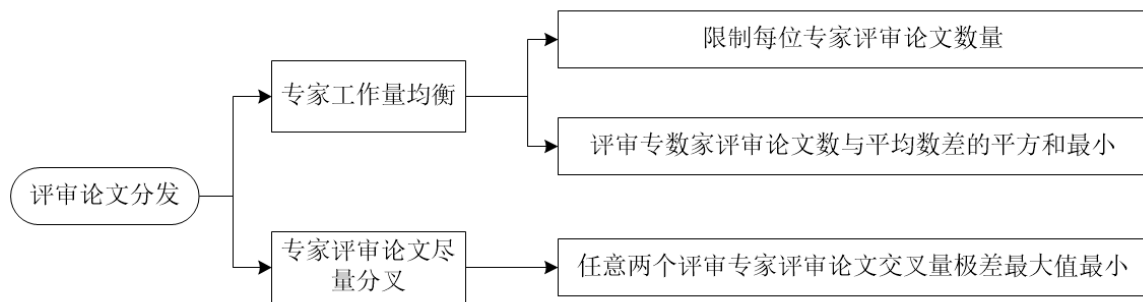


图 4.2 基于均匀-交叉的多目标作品分配模型研究思路图

### 4.2 基于均匀-交叉的多目标作品分配模型

#### 4.2.1 数据观察与模型的选用

在问题一中未提供具体的分发作品的数据附件，因此不考虑对原始数据的分析，重点讨论模型的选用。

在实际情况中，竞赛组委会制定分配方案时，除了需要考虑专家分配到的作品之间的交叉量处于合适的阈值，还需要考虑到任意两位专家之间的评审工作量是均匀的，需要限制分配给每位专家评审的作品数量。将问题一中的“交叉分发”方案转换为一类多目标优化问题，通过设定决策变量、目标函数和约束条件建立基于均匀—交叉的分配作品多目标优化模型，进行求解。

#### 4.2.2 模型的建立与求解

为保证竞赛的公平，在分配作品前对作品进行密封、编号，并遵循以下的分配规则：

- 1、每篇论文分配给 5 个评委进行独立打分；
- 2、论文的分配尽可能均匀；
- 3、专家评委之间尽可能交叉评阅论文；

假设本次竞赛共收到  $n$  份参赛作品，将参赛作品编号设为  $C_1, C_2, \dots, C_n$ ；设参加此次阅卷的评委有  $m$  位，评委编号分别为  $S_1, S_2, \dots, S_m$ 。限定每位专家评审作品数量的阈值区间为  $[a_i, b_i]$ 。

(1) 决策变量的设定

$x_{ik} = 1 \text{ or } 0$ , 若评委 $i$ 评审了 $k$ 号作品, 则 $x_{ik} = 1$ , 否则 $x_{ik} = 0$ ;

$a_i, b_i$ 分别表示评委 $i$ 能够评阅论文数量的上、下界;

$Q_{ij}$ 表示分给 $i, j$ 两位专家论文的重合数量,  $Q_{ij} = \sum_{k=1}^m (x_{ik} \wedge x_{jk})$ 。

假设本次竞赛共收到 $n$ 份参赛作品, 将参赛作品编号设为 $1 - n$ ; 设参加此次阅卷的评委有 $m$ 位, 评委编号分别为 $1 - m$ 。每份论文由随机的 $t$ 位评委评阅打分, 其中 $t < n$ 。限定每位专家评审作品数量的阈值区间为 $[a_i, b_i]$ 。

(2) 目标函数的建立

基于上述分析, 本文所提出的均匀-交叉多目标优化模型中, 存在以下两类目标: 第一类目标要求分发给每位评审专家的论文数量尽可能接近, 因此每位评审专家评阅的论文数量与平均分发作品数量之间的平方和尽可能地小; 第二类目标要求专家评选作品尽可能交叉, 因此任意两位评审专家评阅作品的交叉量的最大值与最小值之差(极差)的最大值尽可能地小。

第一类目标函数的数学模型为式(1):

$$\min \sum_{i=1}^m \left( \sum_{k=1}^n x_{ik} - \frac{t \times n}{m} \right)^2 \quad (\text{其中, } t = 5) \quad (1)$$

第二类目标函数的数学模型为式(2):

$$\begin{cases} z = \max(\max Q_{ij} - \min Q_{ij}) \\ \min(z) \end{cases} \quad (2)$$

其中(1)  $x_{ik}$  ( $i \in (1, m)$ 的自然数,  $k \in (1, n)$ 的自然数)表示为第 $k$ 篇作品由第 $i$ 位专家评审。(2)  $Q_{ij}$ 表示分给 $i, j$ 两位专家论文的重合数量,  $Q_{ij} = \sum_{k=1}^m (x_{ik} \wedge x_{jk})$ 。 $x_{ik} \wedge x_{jk} = 1 \text{ or } 0$ , 当 $x_{ik} \wedge x_{jk} = 1$ , 表示第 $k$ 篇论文由 $i$ 和 $j$ 两个评审专家共同审阅, 否则 $x_{ik} \wedge x_{jk} = 0$ 。

(3) 约束条件的制定

基于上述分析, 本文所提出的均匀-交叉分配模型中, 存在以下的约束性条件: (1) 需要限制每位专家评审作品的数量; (2) 每份作品需要由 $t$ 位评委评阅。此外需要注意 $x_{ik} \wedge x_{jk} = 1 \text{ or } 0$ , 制定式(3)中的约束条件:

$$s. t. \begin{cases} a_i \leq \sum_{k=1}^n x_{ik} \leq b_i \\ \sum_{i=1}^m x_{ik} = t, (t = 5) \\ Q_{ij} = \sum_{k=1}^n (x_{ik} \wedge x_{jk}) \\ x_{ik} = 0 \text{ or } 1 \end{cases} \quad (3)$$

其中(1)  $a_i \leq \sum_{k=1}^n x_{ik} \leq b_i$ 是对每位评审专家阅读论文的数量进行限制, 通过设置 $a_i$ 和 $b_i$ 的大小来保证每位评审专家的阅卷数量相对均衡; (2)  $\sum_{i=1}^m x_{ik} = t$ 是每份论文由 $t$ 位评审专家评阅; (3)  $Q_{ij}$ 表示分给 $i, j$ 两位专家论文的重合数量; (4)  $x_{ik} = 1 \text{ or } 0$ , 若评委 $i$ 评审了 $k$ 号作品, 则 $x_{ik} = 1$ , 否则 $x_{ik} = 0$ 。

### 4.3 模型的求解

相较于单目标优化问题, 多目标优化问题下的子目标之间可能存在互斥关系, 导致可能不存在一个绝对最优解可以使得全部的目标都达到最优。因此, 只能在所有子目标之间不断进行调试, 使目标函数无限逼近理想化的最优解, 最终得出一个折中妥协后的综合效

果最优方案<sup>[3]</sup>。多目标优化算法（Non-dominated Sorting Genetic Algorithm II, NSGA-II 算法）是一种求解多目标优化问题的有效算法，在求解过程和迭代过程中不受求解规模限制，且能自主地协调各目标函数之间的关系，得到 Pareto 最优前沿上多样性解集<sup>[4]</sup>。

本文提出一种基于 NSGA-II 算法的多目标遗传进化算法（Multi-Objective Genetic algorithm, MO-GA）用于求解上述多目标优化模型。该算法融合标准 GA 算法和 NSGA-III 算法，具备自适应参数优化和决定种群初始化、进化阶段的约束条件。下表 4.1 简述改进优化算法与传统算法的区别。

表 4.1

不同点	算法	
	标准 NSGA-II 算法	多目标遗传进化算法（MO-GA）
1	采用 0-1 二进制形式编码	多维矩阵编码
2	不存在约束条件算子	支持在种群生成和进化阶段设定约束条件
3	种群交叉和变异概率为常数定值	种群交叉和变异概率迭代自适应调整

利用 MO-GA 算法求解多目标优化模型，流程主要包括以下几步：

步骤 1：设置和输入初始化参数，参数分为模型参数和算法参数。模型参数在本模型中均取自自然数 1。算法参数：重复迭代次数 $g_{ag}$ 、种群规模 $p_{po}$ 、种群交叉和变异概率自适应参数： $P_{m1} = 0.1, P_{m2} = 0.01$ ；

步骤 2：生成初始化种群，采用二维矩阵 0-1 编码，随机生成比 $p_{po}$ 规模大的初始种群 $A_{ini}$ ；

步骤 3：计算初始种群 $A_{ini}$ 中所有个体的第一类目标函数值 $O_1$ ，第二类目标函数值 $O_2$ 和约束度 $H_i$ 。写入多维矩阵 $D_{ini}$ 中；

步骤 4：以 $D_{ini}$ 中三个参数 $O_1$ 、 $O_2$ 、 $H_i$ 为依据，对初始种群 $A_{ini}$ 进行层级划分和 Euclidean 距离计算，并根据每个个体的层级和 Euclidean 距离，在对应的多维数组中选出最优的满足种群规模的个体组成初始的父代种群 $A_{far}$ ；

步骤 5：通过帝国主义竞争法在种群 $A_{far}$ 中选择一部分的个体通过遗传，交叉等操作形成种群 $A_c$ ，变异操作形成种群 $A_m$ ，合并 $A_c$ 和 $A_m$ 组成子代种群 $A_{ger}$ ，合并种群 $A_{far}$ 和种群 $A_{ger}$ 为种群 $A_{mid}$ ；

步骤 6：计算种群 $A_{mid}$ 中所有个体的 $O_1$ 、 $O_2$ 和 $H_i$ ，得到 $D_{mid}$ ，对种群 $A_{mid}$ 进行层级划分和 Euclidean 距离计算，并根据每个个体的层级和 Euclidean 距离，依次选出 $p_{po}$ 个满足种群规模的个体形成新的种群 $A_{far}$ 。

步骤 7：进行迭代运算，并判断是否达到最大迭代次数，若达到，则终止运算，输出结果，否则一直重复步骤 5。

运用 MO-GA 算法求解多目标优化模型流程图如图 4.2 所示。

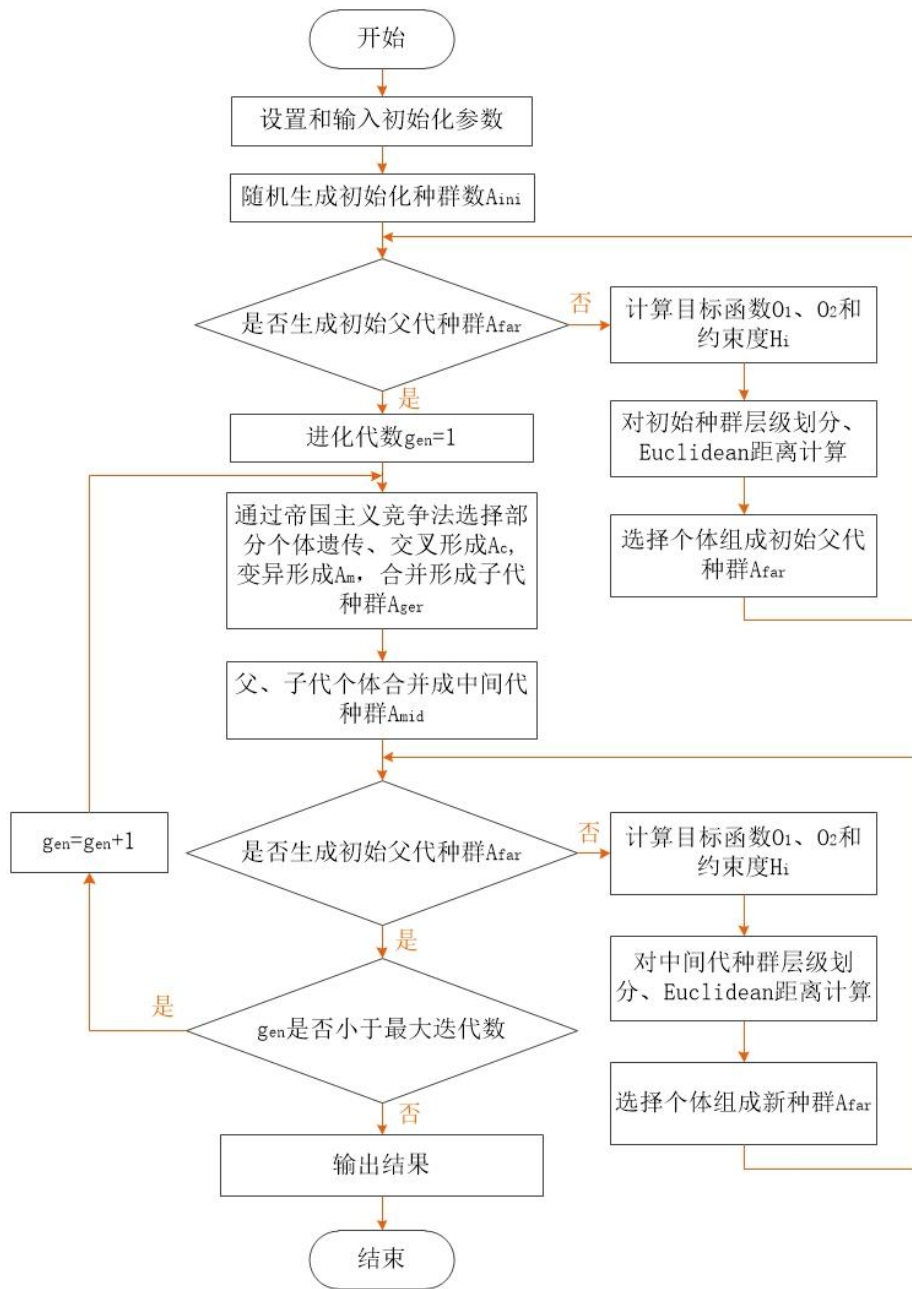


图 4.2 运用 MO-GA 算法求解多目标优化模型流程图

运用 MO-GA 算法求解本文所提出的基于均匀一交叉多目标分配模型时，采取以下步骤：

步骤 1：输入模型的相关参数（其中最大迭代次数为 125 次）；

步骤 2：选取未分配的作品形成种群；

步骤 3 对作品  $k$  是否分配给评委  $i$  进行检测；

1) 作品  $k$  是否已经分配给了评委  $i$ ；

2) 判断作品  $k$  的分配是否超过了评委  $i$  能评阅量的上限；

如果作品  $k$  分配给评委  $i$  是可行的，计算相关目标函数  $O_1$ 、 $O_2$ ，约束条件  $H_i$ ，通过 Euclidean 距离计算选出新的目标种群  $A_{far}$ ；

步骤 4 经过对作品  $k$  分配给所有评委进行检测，如果作品  $k$  可以分配给多个评委，则比较这些可行分配方案使目标函数值的增量，形成新的种群集合  $A_{mid}$ ，随机选取使目标

---

函数值增量最小的分配集作为作品  $k$  的最终分配；

步骤 5 重复步骤 3 和步骤 4，直到作品成功分配给 5 个评委；

步骤 6 判断是否超过最大迭代次数 125，若没有，重复步骤 2 至步骤 5，直到所有论文分配结束。



## 五、问题二模型建立与求解

### 5.1 问题分析及建模思路

问题二要求选择几种评审方案，基于所给数据 1 分析每位专家、每份作品原始成绩、调整之后成绩分布的特点，对这几种方案实行的结果进行排序并进行数据分析比较优劣。设计新的标准分模型再根据一等奖作品的成绩进行改进标准分计算模型。

针对问题二的要求，首先对附件数据 1 进行数据分析及预处理；其次对现有的两种方案进行分析，其中方案一就是数据 1 所实行的方案，对数据 1 中每位专家的打分情况以及每份作品的原始成绩以及标准分进行分析，利用数据 1 对传统的方案二进行分析，通过与方案一中具有高可信度的一等奖排名进行对比从而评价方案的优劣；接着针对传统方案二存在的问题提出方案三，其一是改进标准分计算模型，以此平衡评审专家打分的误差，其二是利用离差比调整各部分专家分数权重值，得到评审模型，并对利用数据 1 中一等奖排名对方案三进行验证方案优劣性；最后再利用成对比较矩阵进一步调整专家权重，得到改进后的方案以及改进后的评审模型，最终再次利用数据 1 中一等奖排名对方案三进行验证方案优劣性。如图所示 5.1 所示为评审模型改进的研究思路。

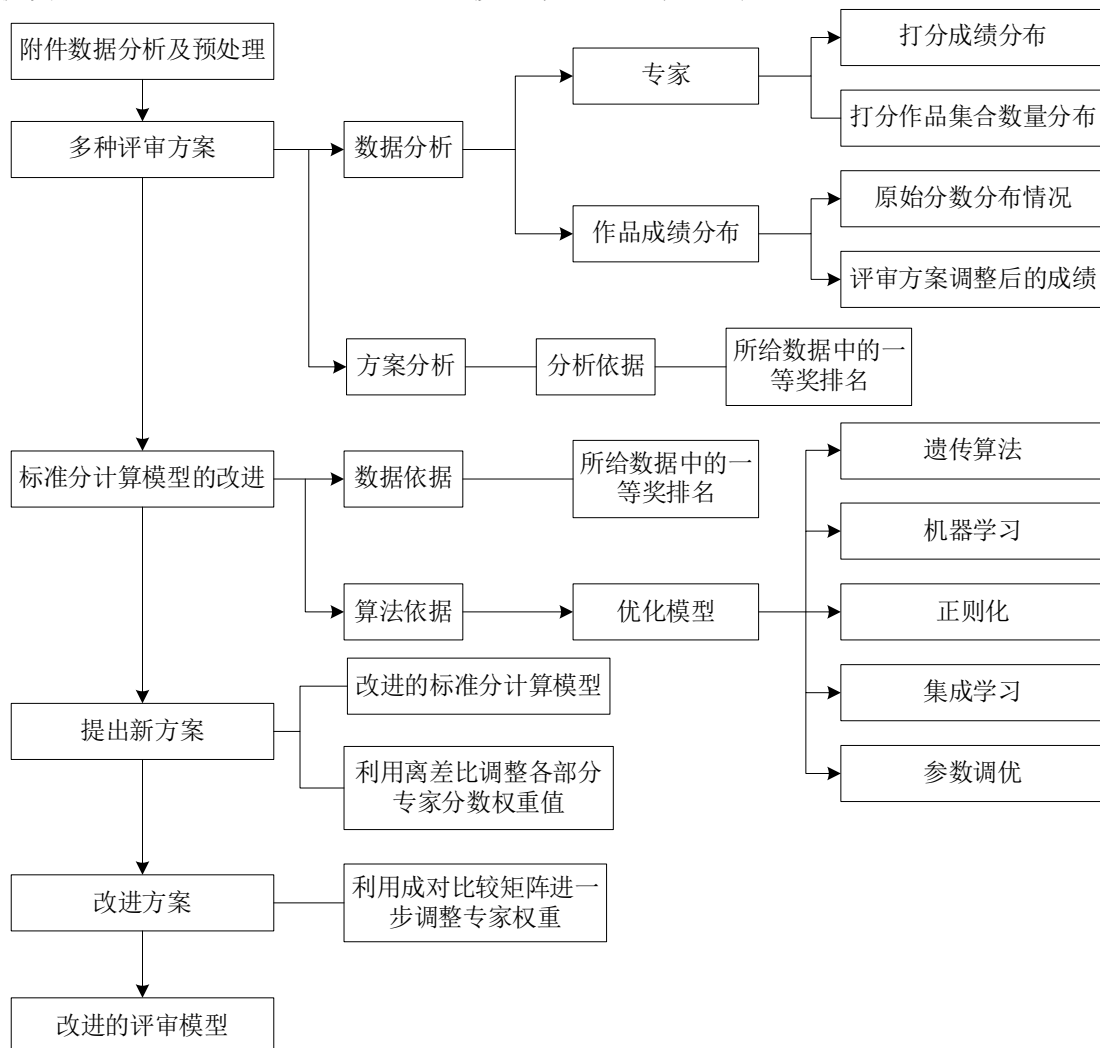


图 5.1 评审模型改进的研究思路

5.2 数据的预处理与相关分析

5.2.1 竞赛成绩分布情况分析

基于数据 1 可以得出竞赛获奖分布情况如表 5.1 和图 5.2 所示：共有 2018 支队伍参赛，一等奖有 27 支队伍占比 1.3%，二等奖有 268 支队伍占比 13.3%，三等奖有 373 支队伍占比 18.5%，其中进入第二阶段评审的有 352 支队伍占比 17.4%。

表 5.1 原始数据获奖及分布情况

获奖类别	数量	原始占比
一等奖	27	1.34%
二等奖	268	13.3%
三等奖	373	18.51%
未获奖	1347	66.85%

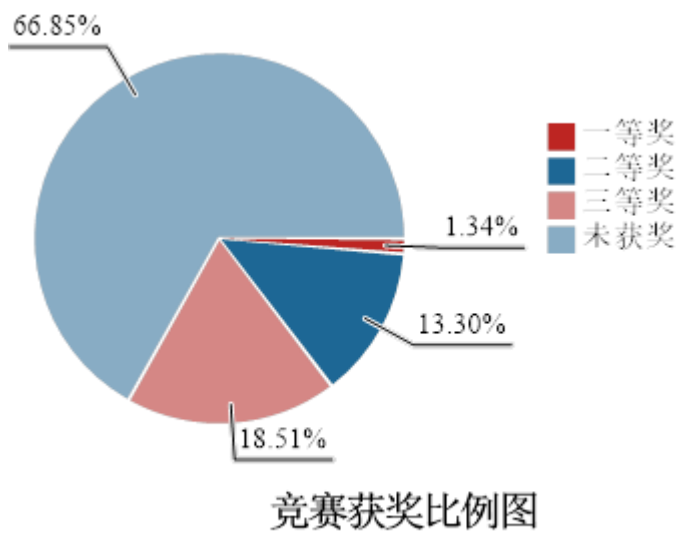


图 5.2 竞赛获奖情况

对数据 1 中的所有作品的分数和排名进行 Shapiro-Wilk 检验，查看其显著性，观察是否符合正态分布。得到结果如表 5.2 所示。

表 5.2 最终成绩描述性统计结果

变量名	样本量	中位数	平均值	标准差	偏度	峰度	S-W 检验
最终成绩	2018	241.61	234.048	39.837	-0.728	-0.024	0.945 (0.000***)

上述结果展示了最终成绩描述性统计的结果，包括中位数、平均值等，用于检验数据的正态性。显著性 P 值为 0.000\*\*\*，水平呈现显著性，因此数据不满足正态分布。其峰度（-0.024）绝对值小于 10 并且偏度（-0.728）绝对值小于 3，因此结合正态分布直方图和正态性检验 P-P 图和 Q-Q 图进行进一步分析，正态分布直方图、正态性检验 P-P 图、正态性检验 Q-Q 图分别如图 5.3、5.4、5.5 所示。

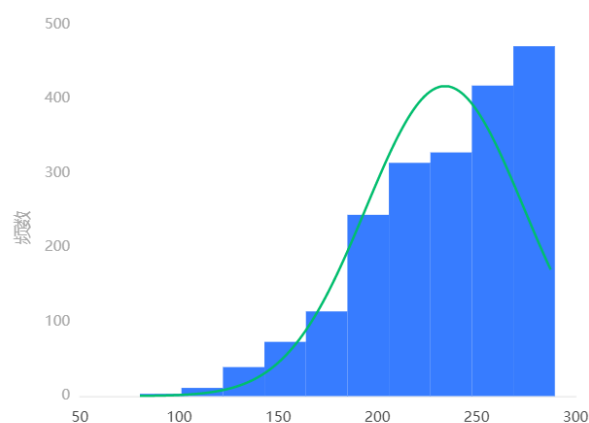


图 5.3 最终成绩数据的正态性检验直方图

根据正态性检验直方图，可以发现若正态图基本上呈现出钟形（中间高，两端低），则说明数据虽然不是绝对正态，但基本可接受为正态分布。

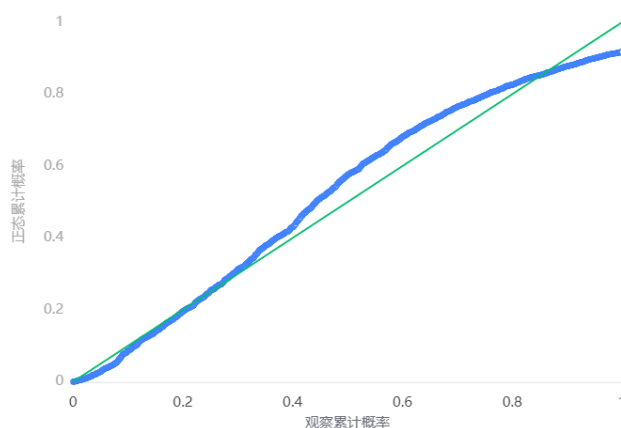


图 5.4 正态性检验 P-P 图

根据正态性检验 P-P 图，观察最终成绩计算观测的累计概率（P）与正态累计概率（P）的拟合情况，拟合程度越高越服从正态分布。

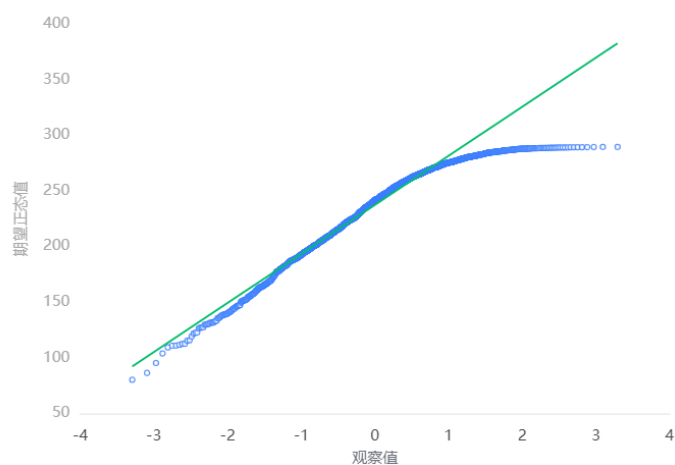


图 5.5 正态性检验 Q-Q 图

根据正态性检验 Q-Q 图，比较观测值与预测值（假定正态下的分布）不同分位数的概率分布，从而检验是否吻合正态分布规律。并且将实际数据作为 X 轴，将假定正态时的数据分位数作为 Y 轴，作散点图，散点与直线重合度越高越服从正态分布，散点差异愈大越

不服从正态分布。

根据上述的检验结果，最终成绩的描述性统计结果中峰度绝对值小于 10 并且偏度绝对值小于 3，结合正态分布直方图、PP 图或者 QQ 图可以描述为基本符合正态分布。

根据所给数据进行数据分析，发现其中存在部分参赛队有专家打零分的情况，考虑这部分参赛队伍存在作品质量极差的情况，对后续的数据分析没有参考意义，这里选择对其进行剔除，通过数据筛选得到所需要剔除的结果如图 5.6 所示。

```
In [72]: df_zeros
```

```
Out[72]:
```

	名次	奖项	11原始分	11标准分	12原始分	12标准分	13原始分	13标准分	14原始分	14标准分	15原始分	15标准分
452	1998	未获奖	0	25.52	5	17.84	30	36.11	0	17.57	0	30.10
562	2005	未获奖	5	12.30	20	29.38	10	23.10	33	31.39	0	19.20
620	1939	未获奖	10	26.87	30	36.11	30	29.52	15	29.30	0	30.10
636	2013	未获奖	2	25.05	8	26.90	10	3.59	0	25.21	5	14.56
663	1993	未获奖	2	28.91	15	20.88	0	16.64	20	33.19	10	30.53
741	2010	未获奖	0	25.52	5	17.84	6	21.24	10	13.20	20	33.05
763	1960	未获奖	0	29.98	10	25.40	10	29.89	30	29.04	20	29.91
831	2008	未获奖	10	29.08	0	29.98	20	10.53	5	24.57	10	17.46
837	2009	未获奖	0	19.93	0	27.76	0	15.74	0	18.23	0	29.34
896	1958	未获奖	10	22.72	0	29.66	10	32.76	20	32.84	0	26.46
1027	1888	未获奖	20	31.59	0	23.86	30	35.76	45	45.64	0	28.74
1101	2015	未获奖	0	20.19	0	4.05	0	23.86	0	13.31	0	18.84
1189	1891	未获奖	0	29.98	45	27.90	10	27.60	55	43.52	33	35.96
1219	1988	未获奖	0	19.93	0	27.76	1	24.45	30	29.52	0	30.10
1230	1983	未获奖	30	34.96	30	36.11	6	21.24	0	21.50	1	21.66
1286	2000	未获奖	0	16.91	0	29.66	0	27.10	0	22.59	0	26.46
1309	1995	未获奖	0	20.19	0	29.66	0	27.10	0	22.59	3	30.14
1391	2014	未获奖	4	13.36	4	2.36	2	15.37	22	31.53	0	23.78
1404	1967	未获奖	5	24.87	0	29.22	15	30.35	10	26.98	40	29.93
1410	1997	未获奖	2	15.37	0	27.76	20	30.10	0	23.78	2	30.21
1415	1985	未获奖	0	24.50	1	27.66	30	29.52	0	27.89	0	23.61
1420	1980	未获奖	10	22.72	10	34.69	1	27.66	0	26.46	10	25.25
1470	2011	未获奖	0	15.93	30	34.96	0	16.79	20	23.62	0	18.08
1520	1300	未获奖	59	58.13	55	58.25	0	22.59	78	50.12	65	54.08
1593	2012	未获奖	0	15.93	15	27.15	0	19.18	20	23.62	0	18.08
1608	1974	未获奖	5	22.56	30	29.52	20	30.10	0	23.78	20	33.05
2012	1937	未获奖	15	25.63	0	29.98	5	32.17	26	39.52	10	25.25
2014	1706	未获奖	59	51.55	0	29.22	45	48.92	60	52.34	30	20.01

图 5.6 存在专家打零分情况的队伍

### 5.2.2 专家打分特点分析

对数据 1 进行数据处理和筛选，得到一共有 97 位专家对此次竞赛进行评分，每位专家评审论文的数量在 102-105 份范围内。由于每位专家的打分习惯各不相同，为了了解每位专家的打分特点，这里计算出所有专家打分的平均值、标准差、中位数等，其中部分如图 5.7 所示。

	P005	P022	P056	P069	P070	P082	P090	P118	P120
25%	40	45.75	48	45	32.75	50	53	42	40.5
max	90	86	87	80	80	87	88	81	86
min	2	0	6	0	2	0	30	20	2
50%	50	54	58	58	50	58	60	62.5	62
75%	65	60	67.5	65	60	65.25	66	70	71.5

图 5.7 部分专家打分情况

对专家打分情况进行分析，如图 5.8 所示，可以看出整体的打分中心和分散情况。

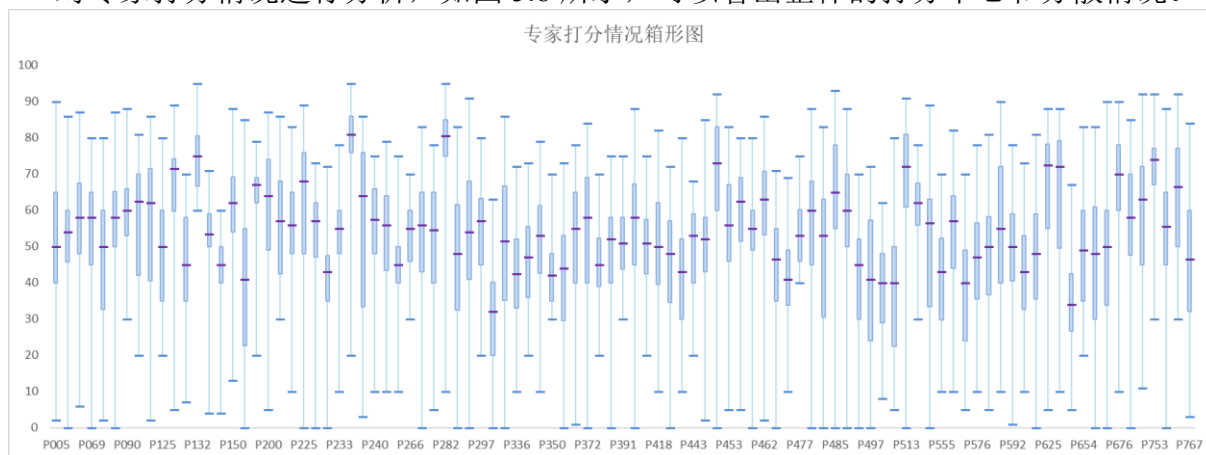


图 5.8 所有专家打分情况箱形图

对 97 位专家打分的均值进行排序分析，在 97 位专家中存在平均打分偏高和偏低的情况，其中 P753、P132、P282 和 P235 这四位专家打分的平均值均在 70 分以上，打分偏高；其中 P322、P635、P573、P511 以及 P506 这五位专家打分的平均值均在 40 分以下，打分偏低。

对 97 位专家打分的标准差进行排序分析，可以发现 P170、P350、P477 以及 P132 这四位专家打分的标准差均在 10 分以内，表明这些专家打分情况较为稳定；而 P294、P070、P163、P287、P658、P632、P334、P120、P484、P225、P549 以及 P236 等十二位专家打分的标准差均在 20 分以上，表明这些专家打分存在不稳定的情况。

对 97 位专家打分的四分位数范围进行排序分析，可以发现 P170、P138、P753、P143、P256、P235 以及 P282 这七位专家所打分数四分位极差较小均在 10 分以内，表明这几位专家所打的分数比较集中，其中 P170 和 P138 两位专家所打分数最为集中，分别主要集中在 62-69 分以及 50-59 分；而 P658、P120、P334、P587、P163、P484、P497 以及 P236 这八位专家所打分数四分位极差较大均在 30 分以上，表明这几位专家所打分数较为分散。下图 5.9 为专家打分情况分布图。

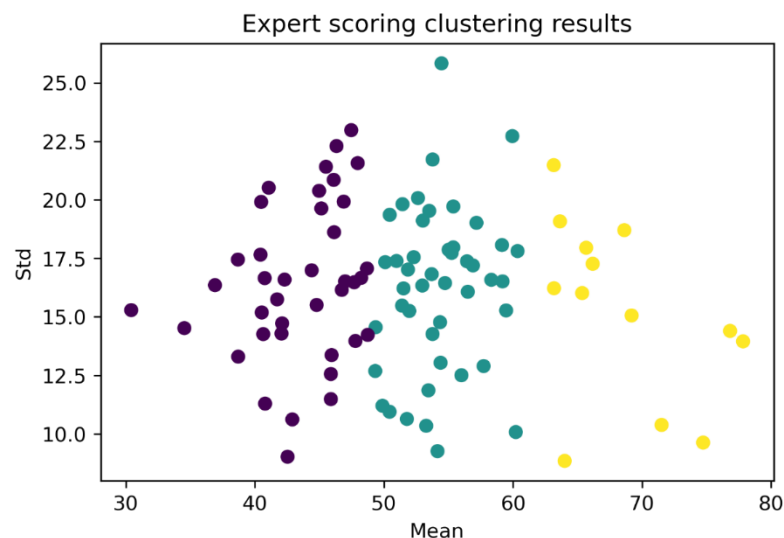


图 5.9 专家打分情况分布图

### 5.2.3 作品名次与第一阶段分数的关系分析

对 2018 个作品的原始分数进行数据处理，计算出第一阶段原始分数的平均分、标准差、中位数以及极差，并对他们进行相关性分析，得到最终作品名次与第一阶段原始分数的平均分、标准差、中位数以及极差的相关系数如表 5.3 所示。

表 5.3 最终作品名次与第一阶段原始分有关参数的相关系数

	平均分	标准差	中位数	极差
作品名次	-0.910	0.122	-0.879	0.113

相关系数热力图如图 5.10 所示。

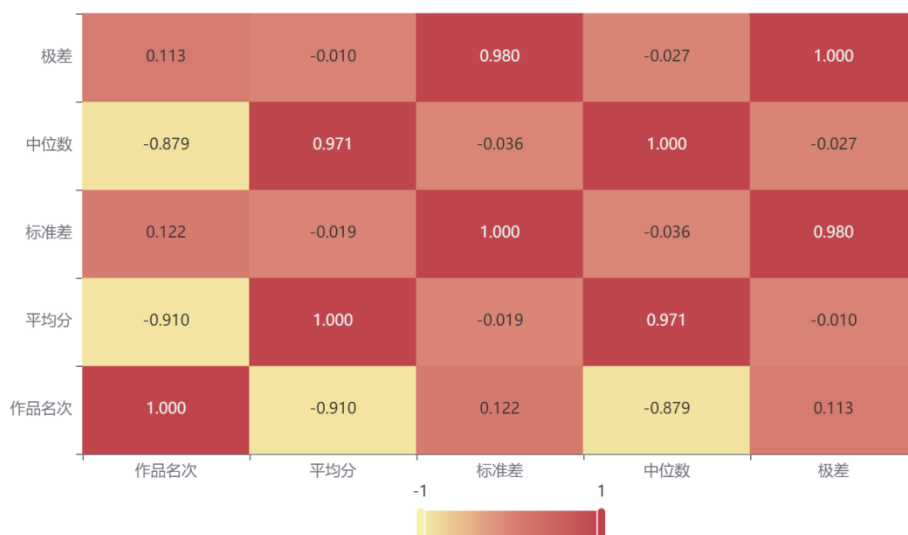


图 5.10 最终作品名次与第一阶段原始分有关参数的相关系数热力图

通过相关性分析可以得到：最终作品名次与第一阶段原始分数的平均分和中位数呈很大程度的负相关，同时与标准差和极差呈较小程度的正相关。

对 2018 个作品的第一阶段标准分数进行数据处理，计算出第一阶段标准分数的平均分、标准差、中位数以及极差，并对他们进行相关性分析，得到最终作品名次与第一阶段标准分的平均分、标准差、中位数以及极差的相关系数如表 5.3 所示。

表 5.4 最终作品名次与第一阶段标准分有关参数的相关系数

	平均分	标准差	中位数	极差
作品名次	-0.947	0.098	-0.936	0.097

相关系数热力图如图 5.11 所示。

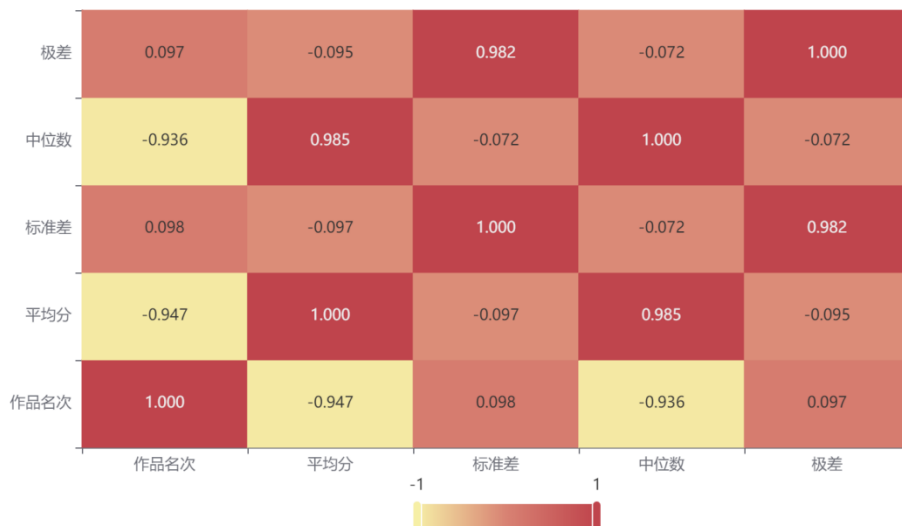


图 5.11 最终作品名次与第一阶段标准分有关参数的相关系数热力图

通过相关性分析可以得到：作品名次与标准分的平均分和中位数呈很大程度的负相关，同时与标准差和极差呈较小程度的正相关。

对比最终作品名次与原始分和标准分的有关参数的相关系数，如图 5.12 所示。



图 5.12 最终名次与第一阶段原始分和标准分有关参数的相关系数对比

### 5.3 基于传统方案的评审模型

这里对不同方案的成绩分布进行分析，分析的方案一为数据 1 所实行的方案：第一次评审后进行第二次评审，同一份作品第一阶段评审的五位专家评审的原始分数根据题中所给标准分计算公式计算得到标准分，对这五个标准分进行求均值，再与第二阶段三位专家评审的标准分共 4 份成绩相加得到总成绩进行排序。

分析的方案二为传统方案：只实行一次评审，同一份作品由五位专家进行评审，所得五个原始分数中去掉最高分和最低分，将剩余三位专家评审的原始分数相加得到总成绩进行排序。方案二实行简单但是存在不够科学的问题，因此这里对其进行改进，改进后的方



案：只实行一次评审，同一份作品由五位专家进行评审，所得五个原始分数根据公式计算得标准分，并去掉最高分和最低分，将剩余三位专家评审的标准分相加得到总成绩进行排序。

对上述的方案二及其改进后的方案二与方案一进行对比。其中在方案一中原先获得一等奖、二等奖、三等奖以及未获奖的队伍在实行改进前后的方案二后获奖的情况如图 5.13 和表 5.4 所示。

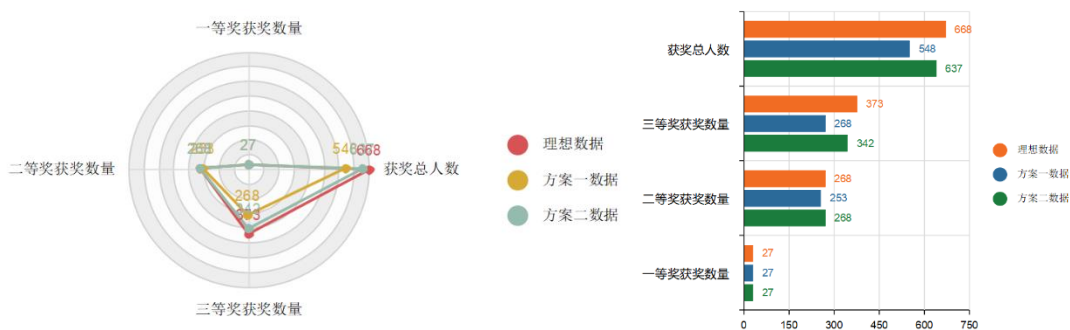


图 5.13 不同方案下获奖数量情况

在方案一中获得一等奖的队伍在方案二和改进后的方案二下仍然全部获奖；在方案一中获得二等奖的队伍在方案二下存在部分未进入获奖行列，268 支队伍中只有 253 支队伍仍然获奖，在改进后的方案二下仍然全部获奖；在方案一中获得三等奖的队伍在方案二和改进后的方案二下存在部分未进入获奖行列，373 支队伍中分别只有 268 和 342 支队伍仍然获奖；在方案一中所有 668 支获奖队伍在方案二和改进后的方案二下只有 548 和 637 支队伍仍然获奖，即分别包括 119 和 31 支原先没有获奖的队伍进入了获奖行列。

表 5.4 方案二和改进后获奖队伍中各原始获奖等级分布情况

类别	方案二和改进后获奖队伍中各原始获奖等级分布数量及其变化						相对变化量	相对变化率
	理想数量	方案二		方案二改进后				
		数量	占比	数量	占比			
一等奖	27	27	100%	27	100%	0	0	
二等奖	268	253	94.40%	268	100%	+15	5.93%	
三等奖	373	269	72.12%	342	91.69%	+73	27.14%	
未获奖	0	119	-	31	-	-88	-73.95% ↓	

在方案一下获得一等奖、二等奖、三等奖和未获奖的队伍在方案二和方案二改进后的相对变化率分别为 0、5.93%、27.14%和-73.95%，说明用标准分代替原始分进行评奖对原先一等奖的队伍获奖没有影响，对原先二等奖、三等奖和未获奖的队伍获奖的影响递增。

其中在方案一中原先获得一等奖、二等奖、三等奖以及未获奖的队伍进入第二阶段评审的情况如图 5.14 和表 5.5 所示。



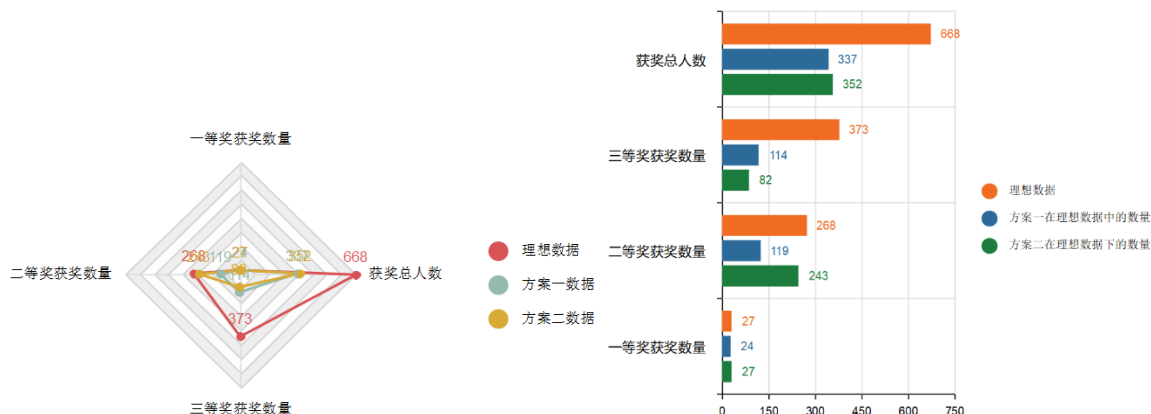


图 5.14 不同方案下进第二阶段评审数量情况

在方案一中获得一等奖的队伍在方案二下存在部分未进入第二阶段评审行列，27 支队伍中只有 24 支队伍进入第二阶段评审，在改进后的方案二下仍然全部进入第二阶段评审；在方案一中获得二等奖的队伍在方案二和改进后的方案二下均存在部分未进入第二阶段评审行列，268 支队伍中分别只有 199 和 243 支队伍仍然进入第二阶段评审行列；在方案一中获得三等奖的队伍在方案二和改进后的方案二下存在更多进入第二阶段评审行列，由原来的 57 支队伍分别增加到 114 和 82 支队伍；在方案一中所有获奖队伍中在方案二下存在部分未进入第二阶段评审，在改进后的方案二下仍然全部进入第二阶段评审，即在方案二下有 15 支原先没有获奖的队伍进入了第二阶段评审。

表 5.5 方案二进入第二阶段评审的队伍中各原始获奖等级分布情况

类别	方案二进入第二阶段的队伍中各原始获奖等级分布数量及其变化						相对变化量	相对变化率
	理想数量	方案二		方案二改进后				
		数量	占比	数量	占比			
一等奖	27	24	88.89%	27	100%	+3	12.50% ↑	
二等奖	268	199	74.25%	243	90.67%	+44	22.11% ↑	
三等奖	57	114	200%	82	143.86%	-32	-28.07% ↓	
未获奖	0	15	-	0	-	-15	-100.00% ↓	

在方案一下获得一等奖、二等奖、三等奖和未获奖的队伍在方案二和方案二改进后进入第二阶段评审中的相对变化率分别为 12.50%、22.11%、-28.07%和-100%，说明用标准分代替原始分进行判定原先一等奖、二等奖、三等奖和未获奖的队伍是否进入第二阶段评审均有影响，且影响递增。

其中在方案一中原先获得一等奖、二等奖、三等奖以及未获奖的队伍获得一等奖或二等奖的情况如图 5.15 和表 5.6 所示。

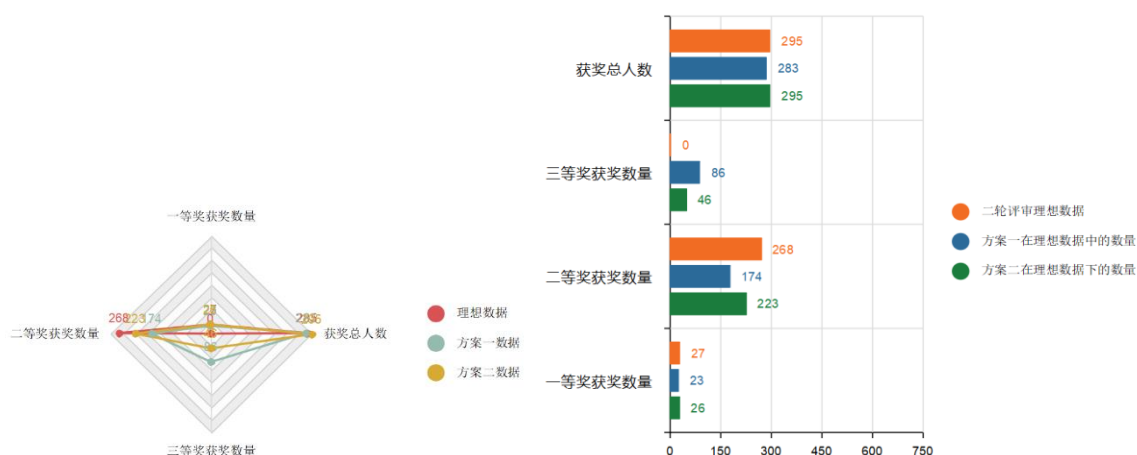


图 5.15 不同方案下获一等奖二等奖数量情况

在方案一中获得一等奖的队伍在方案二和改进后的方案二下存在部分未获得一等奖或二等奖，27 支队伍中分别只有 23 和 26 支队伍仍然获得一等奖或二等奖；在方案一中获得二等奖的队伍在方案二和改进后的方案二下均存在部分未获得一等奖或二等奖，268 支队伍中分别只有 174 和 223 支队伍仍然获得一等奖或二等奖；在方案一中获得三等奖的队伍在方案二和改进后的方案二下存在多支队伍获得一等奖或二等奖，分别有 86 和 46 支队伍进入一等奖和二等奖行列；在方案一中未获奖的队伍在方案二下存在 12 支队伍进入一等奖和二等奖行列。

表 5.6 方案二获得一等奖或二等奖的队伍中各原始获奖等级分布情况

类别	方案二获得一等奖或二等奖各原始获奖等级分布数量及变化						相对变化量	相对变化率
	理想数量	方案二		方案二改进后				
		数量	占比	数量	占比			
一等奖	27	23	85.19%	26	96.30%	+3	13.04%	
二等奖	268	174	64.93%	223	83.21%	+49	28.16%	
三等奖	0	86	-	46	-	-40	-46.51%	
未获奖	0	12	-	0	-	-12	-100.00%	

在方案一下获得一等奖、二等奖、三等奖和未获奖的队伍在方案二和方案二改进后的的相对变化率分别为 13.04%、28.16%、-46.51%和-100%，说明用标准分代替原始分进行判定原先一等奖、二等奖、三等奖和未获奖的队伍是否获得一等奖和二等奖均有影响，且影响递增。

其中在方案一中原先获得一等奖、二等奖、三等奖以及未获奖的队伍获得一等奖的情况如图 5.16 和表 5.7 所示。

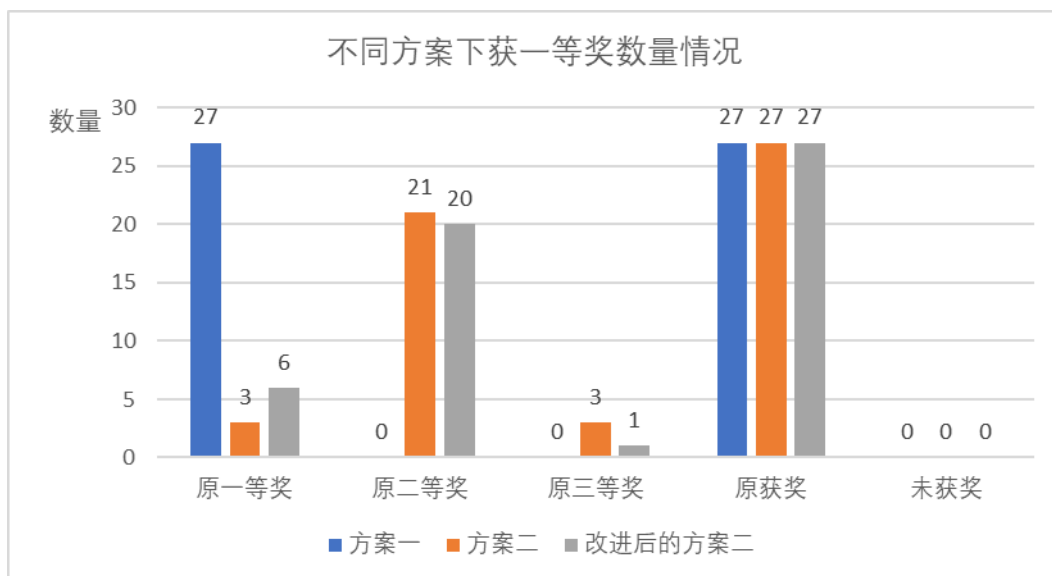


图 5.16 不同方案下获一等奖数量情况

在方案一中获得一等奖的队伍在方案二和改进后的方案二下存在大部分未获得一等奖，27 支队伍中分别只有 3 和 6 支队伍仍然获得一等奖；在方案一中获得二等奖的队伍在方案二和改进后的方案二下分别存在 21 和 20 支队伍获得一等奖；在方案一中获得三等奖的队伍在方案二和改进后的方案二下分别存在 3 和 1 支队伍获得一等奖；在方案二和改进的方案二下获得一等奖的队伍均在原先获奖的行列中。

表 5.7 方案二一等奖队伍中各原始获奖等级分布数量

类别	方案二一等奖队伍中各原始获奖等级分布数量						相对变化量	相对变化率
	理想数量	方案二		方案二改进后				
		数量	占比	数量	占比			
一等奖	27	3	11.11%	5	18.52%	+2	+66.67%	
二等奖	0	21	-	20	-	-1	-4.76%	
三等奖	0	3	-	1	-	-2	-66.67%	
未获奖	0	0	-	0	-	-	-	

在方案一下获得一等奖、二等奖、三等奖和未获奖的队伍在方案二和方案二改进后的的相对变化率分别为 66.67%、-4.76%和-66.67%，说明用标准分代替原始分进行判定原先一等奖、二等奖、三等奖队伍是否获一等奖均有影响。

由于方案一的第二评审阶段评选出的一等奖作品排序经过专家协商取得一致的，因此一等奖作品排序具有最大的可信度，因此这里可以用来判断方案二和改进后的方案二的优劣性。用方案二和改进后的方案二对一等奖作品进行重新评分排序，得到结果如图 5.17 所示。

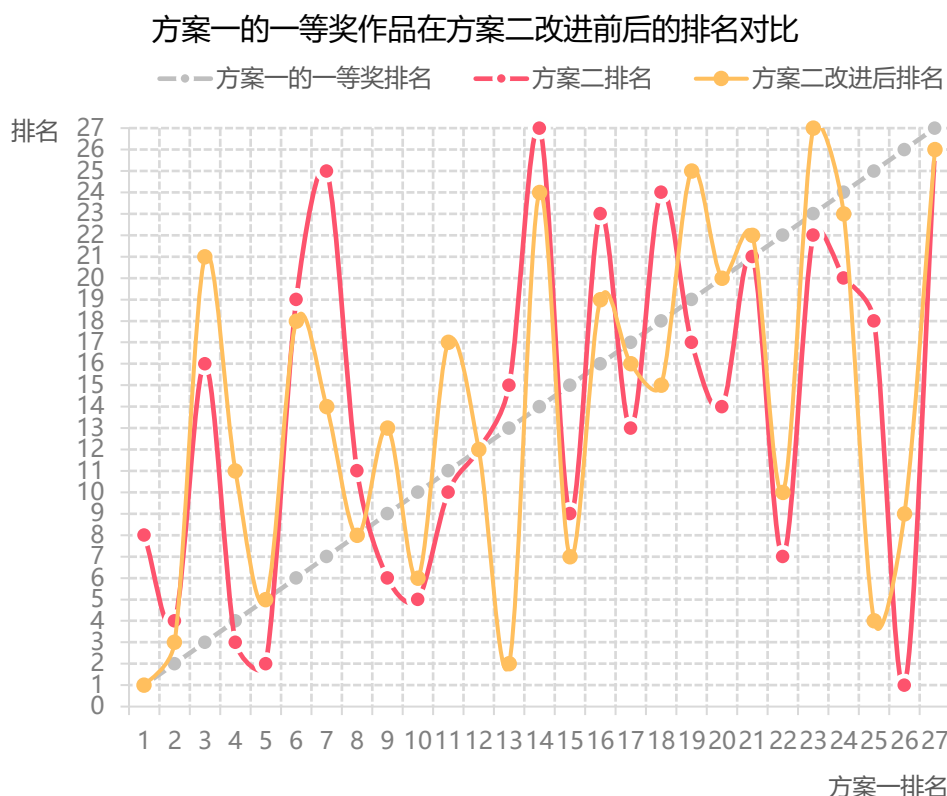


图 5.17 方案一的一等奖作品在方案二改进前后的排名对比

基于方案一中的一等奖作品按方案二和改进后的方案二实行得到的排名可以发现：用标准分代替原始分进行评分排序得到的结果的可信度更高，因此后续设计方案中可以考虑对原始分数进行标准化来评分排序。

#### 5.4 基于离差的权重标准分全局优化的评审模型

在理想情况下，我们期望专家被分配到的作品集合与整个作品集合的学术水平质量分布是一致的，即每位专家所评阅的作品中有一定量的好作品，也有一定量的差作品。但在现实情况中，会存在某位专家被分配到的作品中都是较好的作品，某位专家分配到的作品都是较差作品的情况，此时专家对作品的评分与实际情况将存在一定量的误差。体现在把较优作品中的差作品认为是全局差作品，从而打低分，把较差作品中的较优作品认为是全局优作品，从而打高分。问题二给出标准分模型在判分过程中和客观情况存在误差。此外不同专家的打分习惯不同，有些专家偏向于打高分，有些专家偏向于打低分，我们希望尽可能减小专家个人因素对一份作品最终评分的误差影响，建立一种能客观合理反映一份作品在整个作品中排名的评分模型。为解决两方面问题，本节提出一种基于离差的权重标准分全局优化评审模型。

##### 5.4.1 模型的建立

###### 1、改进标准化分数模型平衡评卷专家打分的误差

近年来有不少学者提出相关的改进方案，集中在 T 分数系数改进法，T 分数单目标优化加权法[5]和 T 分数极差平方和法[4]。不难看出，T 分数的结果对作品评审的重要性。

变量的假设： $a_{ij}$ 表示第*j*位专家对第*i*份作品的原始评分，（ $i = 1, 2, \dots, n$ ）； $x_j$ 表示第*j*位专家打分的标准分。

传统的 T 分数模型如下式（4）所示：

$$\begin{cases} \bar{a}_j = \frac{1}{n} \sum_{i=1}^n a_{ij} \\ s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (a_{ij} - \bar{a}_j)^2} \\ x_j = \delta_0 + \delta \times \frac{a_{ij} - \bar{a}}{s} \end{cases} \quad (4)$$

式(4)中的 $x_j = \delta_0 + \delta \times \frac{a_{ij} - \bar{a}}{s}$ 在问题二所给附件中,  $\delta_0$ 取 50,  $\delta$ 取 10。文献[6]中 $\delta_0$ 取 60,  $\delta$ 取 10, 而在文献[7]中 $\delta_0$ 取 70,  $\delta$ 取 10。本文认为这样的取值方法带有很强的主观性质, 容易导致较大的误差, 为了减小专家个人因素对作品最终成绩的干扰, 本文提出一种全局优化标准分方法。首先将每位专家的打分均值平移到同一基准点。这一基准点选取所有专家打分的均值(即 $\bar{a}_j$ )的均值, 即 $\delta_0 = \frac{1}{m} \sum_{j=1}^m \bar{a}_j$ 。利用方差压缩的思想把所有专家打分的离散度(即方差 $\delta_i$ ,  $i = 1, 2, \dots, m$ )调整到同一标准, 这个标准我们选取所有专家打分的方差的均值, 即 $\delta = \frac{1}{m} \sum_{i=1}^m \delta_i$ 。然后对所有专家所给的分数进行调整。若 $a_{ij}$ 表示第 $j$ 位专家对第 $i$ 份作品的原始评分, ( $i = 1, 2, \dots, n$ ), 那么经过本文提出一种全局优化标准分方法调整后的分数 $x_j'$ 为式(5)所示:

$$x_j' = \delta_0 + \delta \times \frac{a_{ij} - \bar{a}}{s} \quad (5)$$

因此第 $j$ 位专家对第 $i$ 份作品的改进标准分 $a_{ij}'$ 为式(6)所示:

$$a_{ij}' = \delta_0 + \delta \times \frac{a_{ij} - \bar{a}}{s} \quad (6)$$

通过调整 $\delta_0$ 和 $\delta$ , 满足所有专家的均值都相同, 方差也都相同, 从而去掉了不同的专家个人因素对评卷的干扰。若此时某份试卷任意换一位评审专家, 分数也不会相差太远。

## 2、利用离差比调整各部分专家分数权重值

根据之前的分析, 根据原始分求得标准分后, 对第一阶段五位专家评审的标准分直接去取平均或者去掉最高分和最低分效果都没有很优。因此这里提出利用离差比来作为各位评审专家的权重来调节, 取加权平均值, 离差比较大的专家所打的分数相应的权重就较小。

第 $i$ 份作品分别由 $j_1, j_2, \dots, j_k$ 这 $k$ 位专家单独评审, 根据上述改进后的标准分模型调整后的分数分别为 $a_{ij_1}', a_{ij_2}', \dots, a_{ij_k}'$ , 直接取平均得到 $\Delta a_0'$ 。每个分数相对平均值的偏差 $\Delta a_{ij_1}', \Delta a_{ij_2}', \dots, \Delta a_{ij_k}'$ , 那么第 $j_s$  ( $1 \leq s \leq k$ )个专家的权重为(6)式所示:

$$w_s = \frac{1}{k-1} \cdot \frac{\sum_{t=1}^k |\Delta a_{ij_t}'| - |\Delta a_{ij_s}'|}{\sum_{t=1}^k |\Delta a_{ij_t}'|} \quad (7)$$

对这个作品的所有分数进行加权平均, 得到最后的综合总分:

$$c_i = \sum_{s=1}^k w_s \cdot a_{ij_s}' \quad (8)$$

本文设计的方案三: 进行两个阶段的评审, 首先第一个阶段每一份作品都分别有 5 位专家进行评审打分, 并且所打分数均按上述(5)的改进标准化分数模型来计算得到每个原始分数的标准分, 其次在对第一阶段评审的 5 个标准分按上述(6)、(7)式进行计算权重以及第一阶段的综合得分, 再进行第二阶段每一份作品由 3 位专家进行评审打分, 所打分数均按上述(5)的改进标准化分数模型来计算得到每个原始分数的标准分, 最终的成绩

定为第一阶段的综合得分与第二阶段 3 位专家所打原始的标准分一共四份分数之和。

因此方案三采用上述的基于离差的权重标准分全局优化评审模型，即不仅对方案一中计算标准分的模型进行改进，还利用离差比调整第一阶段评审的各部分专家分数权重值计算得分来代替方案一中用五位专家标准分均值的结果。

假设采用方案三后第一阶段第  $i$  份作品分别由  $j_1, j_2, \dots, j_5$  这 5 位专家单独评审，根据上述改进后的标准分模型调整以及利用比调整 5 位专家分数的权重得到第一阶段综合得分  $c_i$  如 (9) 式所示：

$$c_i = \sum_{s=1}^5 w_s \cdot a_{ij_s}' \quad (9)$$

假设采用方案三后第二阶段第  $i$  份作品分别由  $q_1, q_2, q_3$  这 3 位专家单独评审，因此这 3 位专家评审的原始分数  $a_{iq_s}$  的改进标准分  $a_{iq_s}'$ ：

$$a_{iq_s}' = \delta_0 + \delta \times \frac{a_{iq_s} - \bar{a}}{s} \quad (10)$$

第二阶段得分  $d_i$ ：

$$d_i = \sum_{s=1}^3 a_{iq_s}' \quad (11)$$

因此采用方案三作品所得总成绩  $z_i$ ：

$$z_i = c_i + d_i \quad (12)$$

#### 5.4.2 模型的求解与方案三分析

据上述的方案三，基于数据 1 中的得分情况，通过将数据 1 中的作品进行按方案三模型进行数据计算成绩，由于方案一中原先一等奖的 27 个作品的最终排序是由各位专家协商一致通过的，具有最高的可信度，因此这里通过对原一等奖的 27 个作品按照方案三得到的最终成绩，按照求得成绩进行重新排名，将得到的新排名与原排名进行对比分析来评价方案三的优劣性。如图 5.18 和图 5.19 所示是方案一中 27 个一等奖作品在方案三下的得分情况以及原一等奖实行方案二、方案二改进后以及方案三后的排名情况。

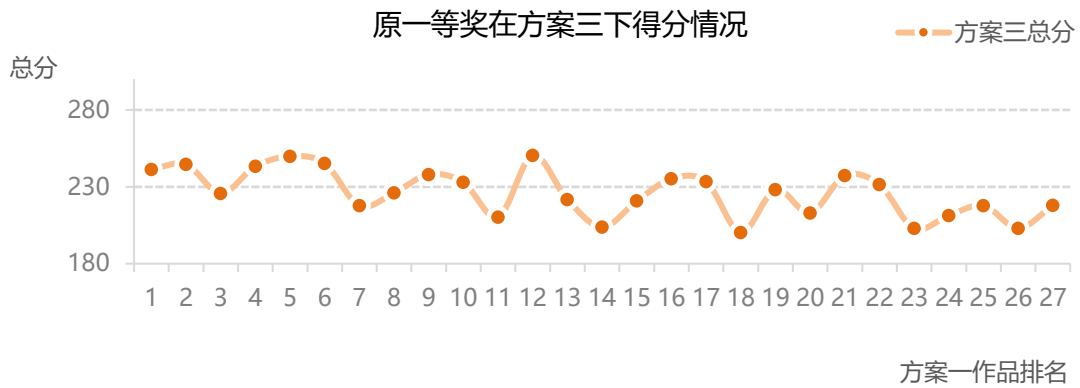


图 5.18 原一等奖在方案三下得分情况  
将几种方案实施之后的原一等奖作品的排名进行对比。

原一等奖在不同方案下的排名情况

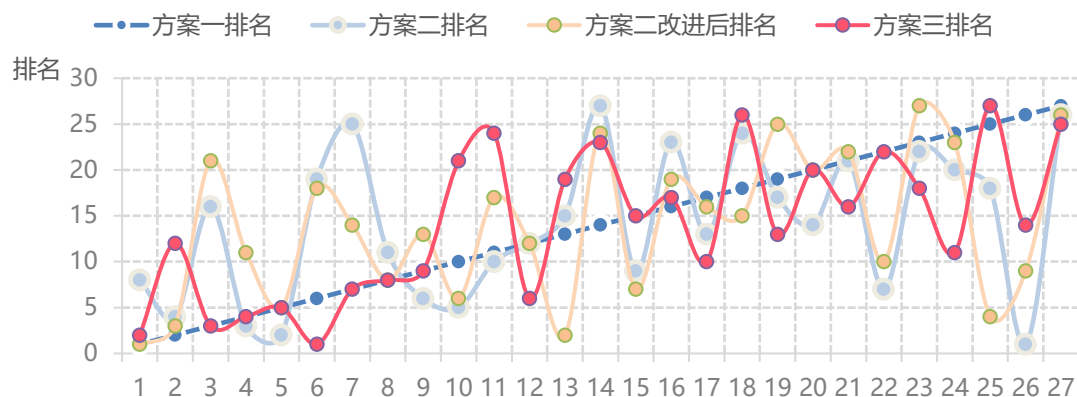


图 5.19 原一等奖在不同方案下的排名情况

从上图可以看出采用方案三对原一等奖的 27 个作品进行排序，得到的排名结果与原排名的吻合度较高。

对采用方案二改进后以及方案三的排名与原排名进行对比分析，得到重合度和乱序度的结果如表 5.8 所示。

表 5.8 方案二改进后、方案三的排名和原排名对比

原排名	方法二改进后排名	重合度 2	乱序度 2	方案三排名	重合度 3	乱序度 3
1	1	1	0	2	0	1
2	3	0	1	12	0	10
3	18	0	15	3	1	0
4	10	0	6	4	1	0
5	6	0	1	5	1	0
6	19	0	13	1	0	5
7	12	0	5	7	1	0
8	13	0	5	8	1	0
9	15	0	6	9	1	0
10	5	0	5	21	0	11
11	22	0	11	24	0	13
12	8	0	4	6	0	6
13	2	0	11	19	0	6
14	25	0	11	23	0	9
15	7	0	8	15	1	0
16	16	1	0	17	0	1
17	17	1	0	10	0	7
18	11	0	7	26	0	8
19	24	0	5	13	0	6
20	20	1	0	20	1	0
21	21	1	0	16	0	5
22	9	0	13	22	1	0
23	27	0	4	18	0	5
24	23	0	1	11	0	13
25	4	0	21	27	0	2
26	14	0	12	14	0	12
27	26	0	1	25	0	2



通过对上表 5.8 分析, 可得方案二改进后的排名与原排名的重合度为 5, 重合率约为 0.19, 乱序度为 166; 方案三的排名与原排名的重合度为 9, 重合率约为 0.33, 乱序度为 122。因此采用方案三得到的排名相较于采用改进后的方案二得到的排名与原排名的重合度更高以及乱序度更低, 因此方案三在一定程度上相较于改进后的方案二更优越、可信度更高, 但是仍然存在乱序的问题, 所有 27 个作品的方案三排名与原排名的整体趋势存在一定程度的差异, 因此这里对方案三所提出模型进行改进。

#### 5.4.3 模型以及方案的改进

这里考虑对方案三所提出模型进行改进: 上述第一阶段的评审中对 5 位专家所打分数进行计算得到 5 个改进后的标准分, 再进行利用离差比调整各部分专家分数权重值计算第一阶段的综合得分, 这里可以先对原始评分进行 T 分数处理, 再用成对比较矩阵进一步调整各个专家的权重再计算第一阶段的综合得分。

因此改进的方案三是指按照上述改进的模型进行算得第一阶段的综合得分, 其第二阶段得分与改进前的方案三相同, 即第一阶段改进后的综合得分与第二阶段原方案三的得分相加得到的就是各作品采用方案三的最终成绩。

基于上述问题二的假设, 各位评审专家在主观评分中的权重并不代表专家的水平高低, 而是为了减少专家之间评分的系统误差。特别是在残缺评分矩阵中, 如果采用 T 分数法处理, 可以消除专家之间平均分及方差不等的误差, 但考虑到每位专家评审的论文不同, 评审的论文的整体水平也存在差异, 因此将每位评委的平均分和方差换算成一个值并不合理。因此合理的做法是可以先对原始评分进行 T 分数处理, 再用成对比较矩阵进一步调整各个专家的权重<sup>[5-6]</sup>。

设  $q_{ij} = \bar{y}_i^j / \bar{y}_j^i$  分别表示评委  $i, j$  交叉论文的均分, 比如评委  $i, j$  共同评审了编号为  $k_1, k_2, \dots, k_r$  的  $r$  份论文, 则  $\bar{y}_i^j = \frac{1}{r} \sum_{t=1}^r y_{ik_t}$ ,  $\bar{y}_j^i = \frac{1}{r} \sum_{t=1}^r y_{jk_t}$ 。因此矩阵  $A = (q_{ij})_{n \times n}$  就表示专家均分成对比较矩阵<sup>[7-8]</sup>。对于专家  $i, j$  共同评审的编号为  $k_1, k_2, \dots, k_r$  的  $r$  份论文, 从理想情况上来说, 如果两个专家打分标准完全相同, 则应该有  $y_{ik_t} = y_{jk_t}$  ( $t=1, 2, \dots, r$ ) 成立<sup>[9]</sup>, 但事实上存在专家打分系统误差和随机误差, 上式往往不能完全成立<sup>[7,9]</sup>。因此设评委  $i$  的评分权重为  $w_i$ , 则应该有  $\bar{y}_i^j w_i \approx \bar{y}_j^i w_j$ , 即  $q_{ij} = \bar{y}_i^j / \bar{y}_j^i \approx \bar{y}_j^i w_j / \bar{y}_i^j w_i$ 。所以

$$A = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{n1} & q_{n2} & \cdots & q_{nn} \end{bmatrix} \approx \begin{bmatrix} w_1/w_1 & w_2/w_1 & \cdots & w_n/w_1 \\ w_1/w_2 & w_2/w_2 & \cdots & w_n/w_2 \\ \vdots & \vdots & \ddots & \vdots \\ w_1/w_n & w_2/w_n & \cdots & w_n/w_n \end{bmatrix} \quad (13)$$

一般来说  $q_{ij} w_i - w_j$  的值并不为零, 但可以选择一组  $\{w_1, w_2, \dots, w_n\}$  使其误差平方和最小, 即:  $\min\{z = \sum_{i=1}^n \sum_{j=1}^n (q_{ij} w_i - w_j)^2\}$

为了保持评分的百分制, 价值系数  $\{w_1, w_2, \dots, w_n\}$  受约束于:  $\sum_{i=1}^n w_i = n, w_i > 0 (i = 1, 2, \dots, n)$

专家  $i$  对论文  $k$  的最终评分为, 取论文  $k$  的平均分  $\bar{y}_i^k = \frac{1}{t} \sum_{p=1}^t y_{ipk}$  作为第一阶段的综合得分, 第二阶段与原方案三的计算方法一致即第二阶段的三位专家进行评分后仍然按照上述式 (6) 改进化的标准分模型进行计算标准分, 最终将第一阶段的得分与第二阶段三位专家的标准分相加即为最终成绩, 对其进行排名得到该竞赛在实行改进后方案三的排名。

#### 5.4.4 改进模型的求解

据上述的改进后的方案三, 基于数据 1 中的得分情况, 通过将数据 1 中的作品进行按方案三改进后的模型进行数据计算成绩, 因此这里仍然通过对具有高可信度的原一等奖的



27 个作品按照改进后的方案三得到的最终成绩分析，按照求得成绩进行重新排名，将得到的新排名与原排名进行对比分析来评价改进后的方案三的优劣性。如图 5.20 和图 5.21 所示是方案一中 27 个一等奖作品在改进后方案三下的得分情况以及原一等奖实行方案三以及改进后的方案三的排名情况。

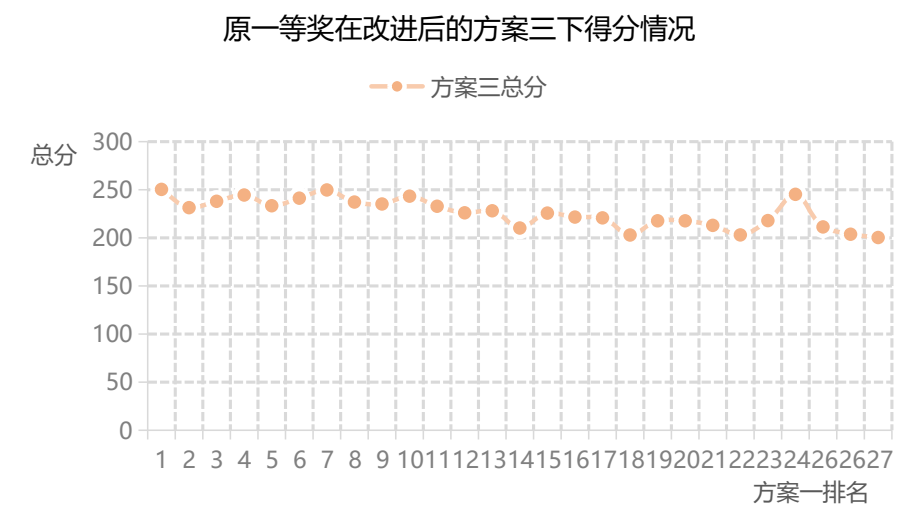


图 5.20 原一等奖在改进后的方案三下得分情况

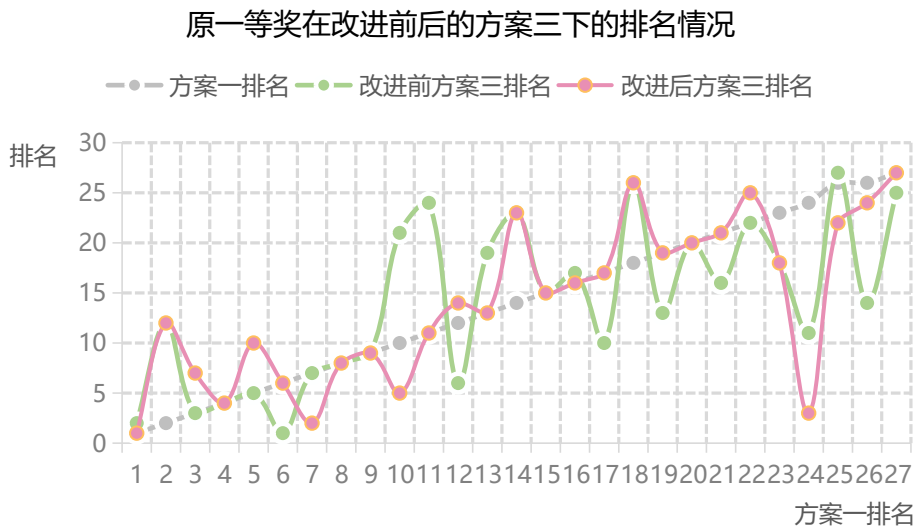


图 5.21 原一等奖在改进前后的方案三下的排名情况

从上图可以看出采用改进后的方案三对原一等奖的 27 个作品进行排序，得到的排名结果与原排名的吻合度较高。对采用方案三以及改进后方案三的排名与原排名进行对比分析，得到重合度和乱序度的结果如表 5.9 所示。

表 5.9 方案三改进前后的排名和原排名对比

原排名	方案三排名	重合度 3	乱序度 3	方案三改进后排名	重合度 4	乱序度 4
1	2	0	1	1	1	0
2	12	0	10	12	0	5
3	3	1	0	7	0	21
4	4	1	0	4	1	0

5	5	1	0	10	0	5
6	1	0	5	6	1	0
7	7	1	0	2	0	4
8	8	1	0	8	1	0
9	9	1	0	9	1	0
10	21	0	11	5	0	5
11	24	0	13	11	1	0
12	6	0	6	14	0	10
13	19	0	6	13	1	0
14	23	0	9	23	0	2
15	15	1	0	15	1	0
16	17	0	1	16	1	0
17	10	0	7	17	1	0
18	26	0	8	26	0	5
19	13	0	6	19	1	0
20	20	1	0	20	1	0
21	16	0	5	21	1	0
22	22	1	0	25	0	4
23	18	0	5	18	0	9
24	11	0	13	3	0	2
25	27	0	2	22	0	3
26	14	0	12	24	0	8
27	25	0	2	27	1	0
总和		9	122		14	83

通过对上表 5.9 分析，可得方案三改进后的排名与原排名的重合度为 14，重合率约为 0.63，乱序度为 83。因此采用改进后的方案三得到的排名相较于改进前的方案三得到的排名与原排名的重合度更高以及乱序度更低，其中重合度提高了约 56%，乱序度降低了约 32%，由此可以看出改进后的方案三更优越、可信度更高。

## 六、问题三模型建立与求解

### 6.1 问题分析及建模思路

针对问题三，本文首先分析了两阶段成绩和极差分布的特点，对极差、名次、最终成绩、一阶段平均标准分利用 Kendall 系数进行一致性检验，得出协调系数，相关度完全一致。为区别高低分段的极差与创新性差异，选出创新性作品，本文引入作品争议度，以作品创新度为衡量作品是否具备创新性的指标，建立基于极差—争议度的作品创新度评价模型。利用 K-means 并结合附件数据，对作品创新度进行聚类，并提出基于创新度的二轮分数修正模型，用于对极差中等的作品进行极差修正。

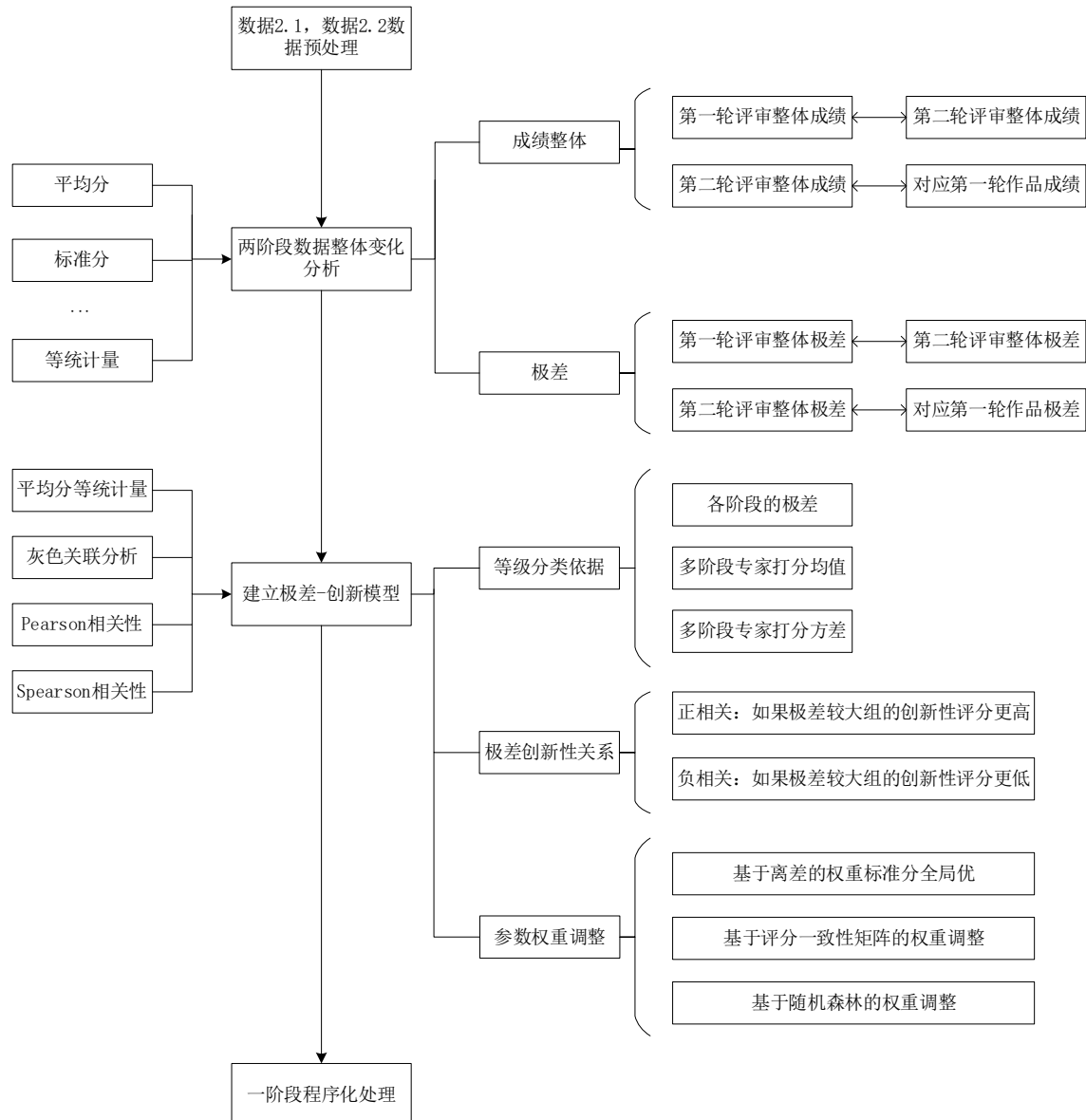


图 6.1 “极差”模型的建立研究思路

### 6.2 评审方案优劣比较分析

#### 6.2.1 基于模拟数据 2.1 数据的评审方案优劣分析

##### 1 两阶段的成绩整体的变化分析

两阶段的成绩整体的变化数据如表 6.1 所示。

表 6.1 模拟数据 2.1 两阶段的成绩整体的变化数据

类别	mean	std	min	25%	50%	75%	max
第一阶段所有打分描述	50.05109	9.92029	7.71	44.43	50.73	56.34	85.99
第二阶段所有打分描述	49.98592	9.880633	25.38	42.73	48.72	56.475	85.33
第 2 阶段作品 1 阶段打分述	59.30645	6.776417	36.33	54.48	59.09	63.51	85.99

对数据 2.1 中的第一阶段的所有 4425 个标准分和第二阶段所有 720 个标准分以及进入第二阶段的所有 240 个作品在第一阶段的标准分情况进行分析，得到结果如图 6.2 所示。两个阶段的成绩整体变化趋势：第二阶段

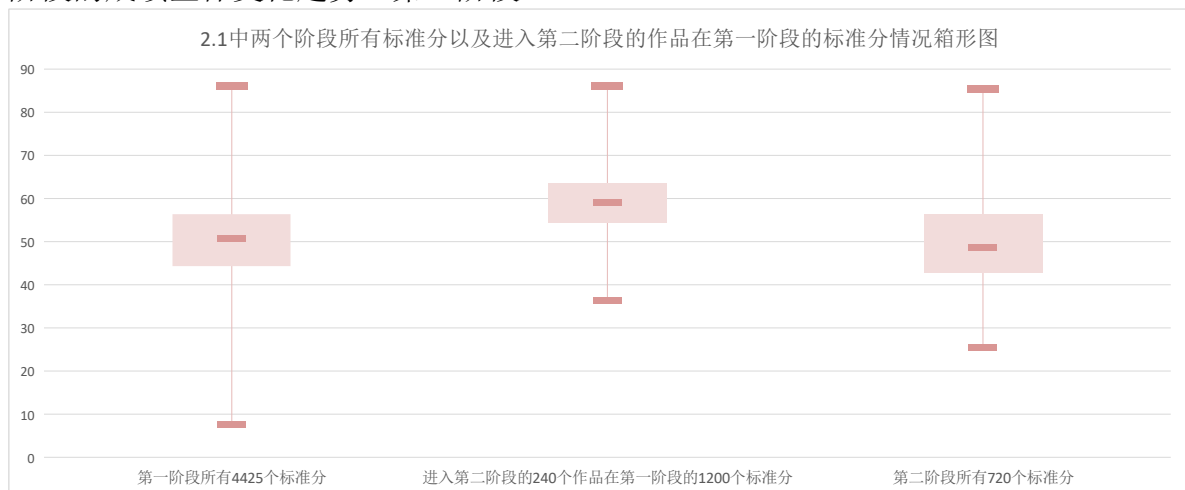


图 6.2 2.1 中两个阶段所有打分以及进入第二阶段的作品在第一阶段的打分情况箱形图

通过上图可以看出，第二阶段的中位数更接近箱体的下限值即是一个右偏的分布，且上边的箱须更长，长须向较高数值延伸，使其正偏态；进入第二阶段的所有作品的第一阶段标准分近似对对称分布。

第一阶段所有标准分与第二阶段所有标准分对比：

(1) 第二阶段标准分最高分和均分相较于第一阶段标准分微小程度降低，最低分相较于第一阶段更高；

(2) 第二阶段标准分整体相较于第一阶段标准分有所提高；

(3) 第二阶段标准分主要集中分布的范围比第一阶段稍广，第一阶段标准分相对集中并且主要集中在 45-55 分，第二阶段标准分主要集中在 42-55 分；

(4) 第二阶段标准分极差相比于第一阶段有大幅度减小。

进入第二阶段的所有作品在第一阶段和第二阶段标准分对比：

(1) 第二阶段标准分的整体、最高分和均值相对于第一阶段有所降低，其中均值降低了近 10 分；

(2) 第二阶段标准分主要集中分布的范围比第一阶段稍广，第一阶段标准分相对集中并且主要集中在 55-65 分，第二阶段标准分主要集中在 42-55 分；

(3) 第二阶段比第一阶段的箱和箱须更长，因此具有更大方差即分数波动更大。

基于上述的两种对比分析，可以发现第二阶段的成绩比第一阶段的成绩偏低，并且方差也更大；

(4) 第二阶段标准分极差相比于第一阶段有近 10 分的增大。

对数据 2.2 中的第一阶段的所有 46645 个标准分和第二阶段所有 4500 个标准分以及进入第二阶段的所有 1500 个作品在第一阶段的 7500 个标准分情况进行分析，得到结果如图

6.2 所示。

2 两阶段极差整体的变化分析

两阶段的极差整体的变化数据如表 6.2 所示。

表 6.2 模拟数据 2.1 两阶段的极差整体的变化数据

类别	mean	std	min	25%	50%	75%	max
第二阶段整体极差	12.47883	6.206357	1.02	7.8075	12.215	16.68	36.76
第一阶段整体极差	12.93866	5.596951	2.05	8.84	12.16	15.59	47.19
第 2 阶段作品的 1 阶段极差	14.61025	6.061826	3.85	10.035	13.855	18.025	34.64

对数据 2.1 中的进入第二阶段的所有 240 个作品在第一、二阶段的标准分极差进行分析，得到结果如图 6.3 所示。

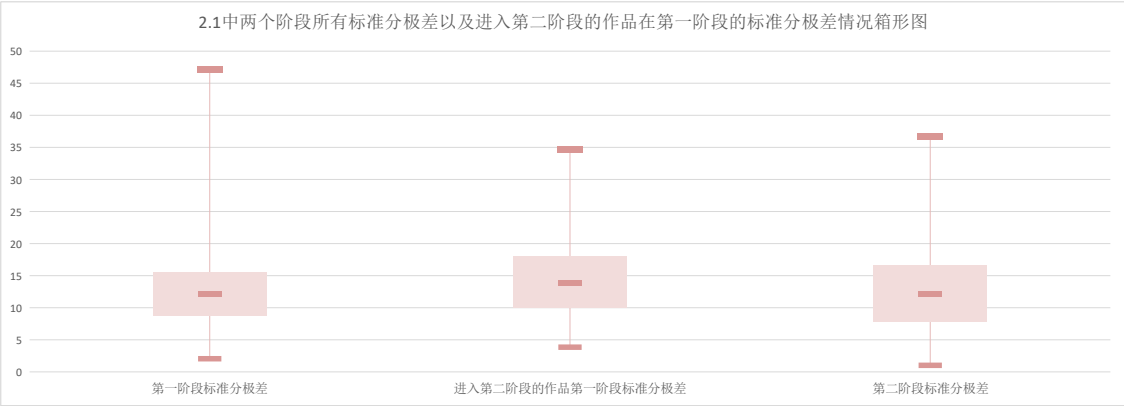


图 6.3 2.1 中两个阶段极差及进入第二阶段的作品在第一阶段的极差情况箱形图

6.2.2 基于模拟数据 2.1 数据的评审方案优劣分析

1 两阶段的成绩整体的变化分析

两阶段的成绩整体的变化数据如表 6.3 所示。

表 6.3 模拟数据 2.2 两阶段的成绩整体的变化数据

类别	mean	std	min	25%	50%	75%	max
第一阶段所有打分描述	50.001806	9.94886148	4.46	44.38	50.71	56.76	93.76
第二阶段所有打分描述	50.012311	9.8268765	-3.03	43.84	49.68	56.37	85.82
第 2 阶段作品 1 阶段打分	61.33503	5.8032254	42.01	57.37	61.145	65.22	93.76

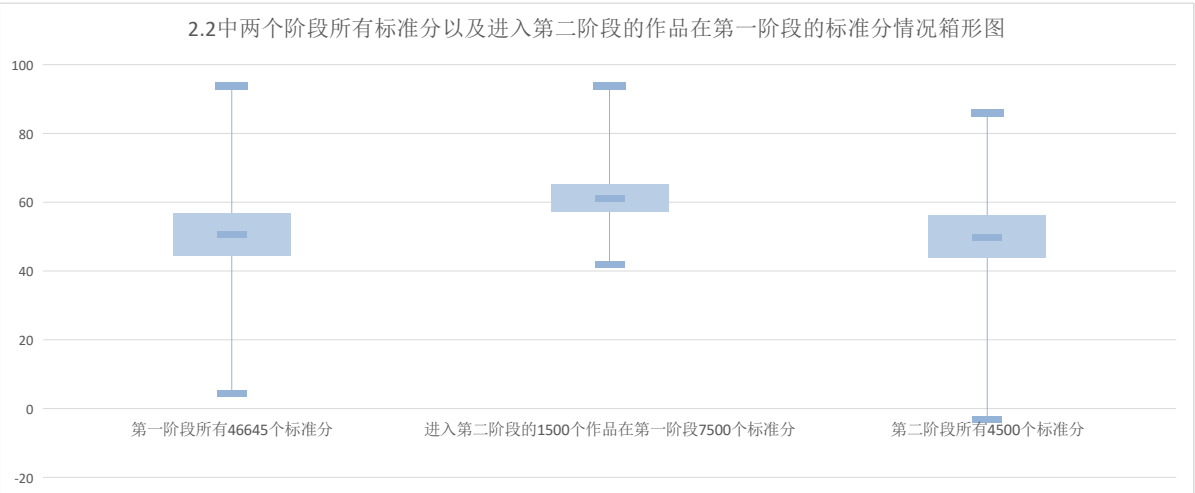


图 6.4 2.2 中两个阶段所有打分以及进入第二阶段的作品在第一阶段的打分情况箱形图

通过上图可以看出，第一阶段所有作品标准分近似对对称分布；二阶段的中位数更接近箱体的下限值即是一个右偏的分布，且上边的箱须更长，长须向较高数值延伸，使其正偏态。

第一阶段所有标准分与第二阶段所有标准分对比：

（1）第二阶段标准分最高分和均分相较于第一阶段标准分微小程度降低，最低分相较于第一阶段更低；

（2）第二阶段标准分主要集中分布的范围相近；

（3）两个阶段标准分的极差相近。

进入第二阶段的所有作品在第一阶段和第二阶段标准分对比：

（1）第二阶段标准分的整体、最高分和均值相对于第一阶段有所降低，其中均值降低了多于 10 分；

（2）第二阶段标准分主要集中分布的范围比第一阶段更广，第一阶段标准分相对集中并且主要集中在 58-62 分，第二阶段标准分主要集中在 42-58 分。

（3）第二阶段比第一阶段的箱和箱须更长，因此具有更大方差即分数波动更大。

（4）第二阶段标准分极差相比于第一阶段有大幅度增大，考虑其中原因是作品本身存在严重问题在第二阶段评审时被发现。

基于上述两个数据的对比分析，可以发现第二阶段的成绩比第一阶段的成绩更低，并且方差也更大。

2 两阶段极差整体的变化分析

两阶段的极差整体的变化数据如表 6.4 所示。

表 6.4 模拟数据 2.2 两阶段的极差整体的变化数据

类别	mean	std	min	25%	50%	75%	max
第二阶段整体极差	14.59708	8.0002454	0.56	8.6425	13.81	19.515	67.32
第一阶段整体极差	12.20387	6.0392756	0.94	8.03	11.15	15.09	60.23
第 2 阶段作品的 1 阶段极差	13.44331	5.1476998	1.83	9.825	12.94	16.57	35.54

对数据 2.2 中的进入第二阶段的所有 1500 个作品在第一、二阶段的标准分极差进行分析，得到结果如图 6.4 所示。

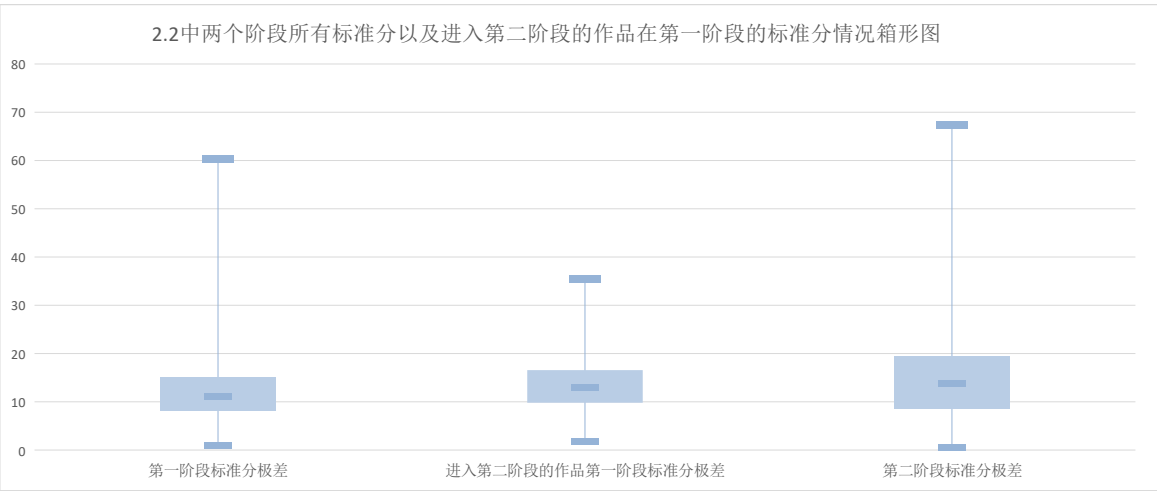


图 6.4 2.2 中两个阶段极差及进入第二阶段的作品在第一阶段的极差情况箱形图

对比数据 2.1 和 2.2 中进入第二阶段的作品在第一、二阶段的标准分极差，可以发现第二阶段相较于第一阶段的极差整体得到减小，但是极差的最大值更大、最小值更小。

6.2.3 两阶段评审方案相比不分阶段评审方案的优劣分析

### 6.3 基于极差-争议度的论文创新度评价模型的建立与分析

#### 6.3.1 基于极差-争议度的论文创新度评价模型

基于“创新类”竞赛的评分高低部分取决于作品是否具备创新性，专家与专家之间很难有一致的看法，尤其是对高分段和低分段的作品，低分段的大极差往往可能是单个专家的意见与其他多数专家意见相悖导致出现极差大的现象，可以看成该作品争议度大。对附件 2.1 中的最终成绩、名次、第一阶段极差和第一阶段标准分均值利用 Kendall 一致性检验进行整体性关联分析，结果如下表所示。

表 6.5 附件 2.1 中的最终成绩、名次、第一阶段极差和第一阶段标准分均值一致性检验

Kendall's W 分析结果					
名称	秩平均值	中位数	Kendall's W 系数	X <sup>2</sup>	P
最终成绩	3.739	210.88	0.802	710.134	0.000***
名次	3.017	148			
1 阶段极差	1.044	13.44			
1 阶段标准成绩的均值	2.2	57.948			

注：\*\*\*、\*\*、\*分别代表 1%、5%、10%的显著性水平

Kendall 系数一致性检验的结果显示，整体数据呈现一致性，模型的 Kendall 协调系数  $W = 0.802$ ，因此相关性的程度为几乎完全一致，认为极差与文章创新性之间具有一定关系。为客观地评选作品的创新度。本文引入作品水平争议度，提出一种基于极差-争议度的论文创新度评价模型。

设第  $i$  份作品的创新度为  $Ino_i$ ，设第  $i$  份作品的争议度为  $H_k$ 。第  $i$  位专家对第  $k$  份作品的标准分为  $f_{ik}$ ， $\bar{f}_k$  为第  $k$  份作品得分的平均值，假设一份作品有  $n$  位评委打分，第  $k$  份作品的争议度为：

$$H_k = \sqrt{\sum_{i=1}^n (f_{ik} - \bar{f}_k)^2} \times x_{ik}$$

$$\begin{cases} Ino_i = \frac{\max\{f_{ik}\} - \min\{f_{ik}\}}{H_k} \\ H_k = \sqrt{\sum_{i=1}^n (f_{ik} - \bar{f}_k)^2} \times x_{ik} \end{cases} \quad (14)$$

在该模型中，若第  $i$  篇作品的争议度越小，且极差大越则说明该作品的创新度越好。反之，创新度越小。

#### 6.3.2 创新度的分类

利用 K-means 对作品创新度一名次进行聚类分析，将作品创新度分为 5 类，得出不同名次下，作品创新度的分布特点。下图 6.5 为作品创新度分布特点。

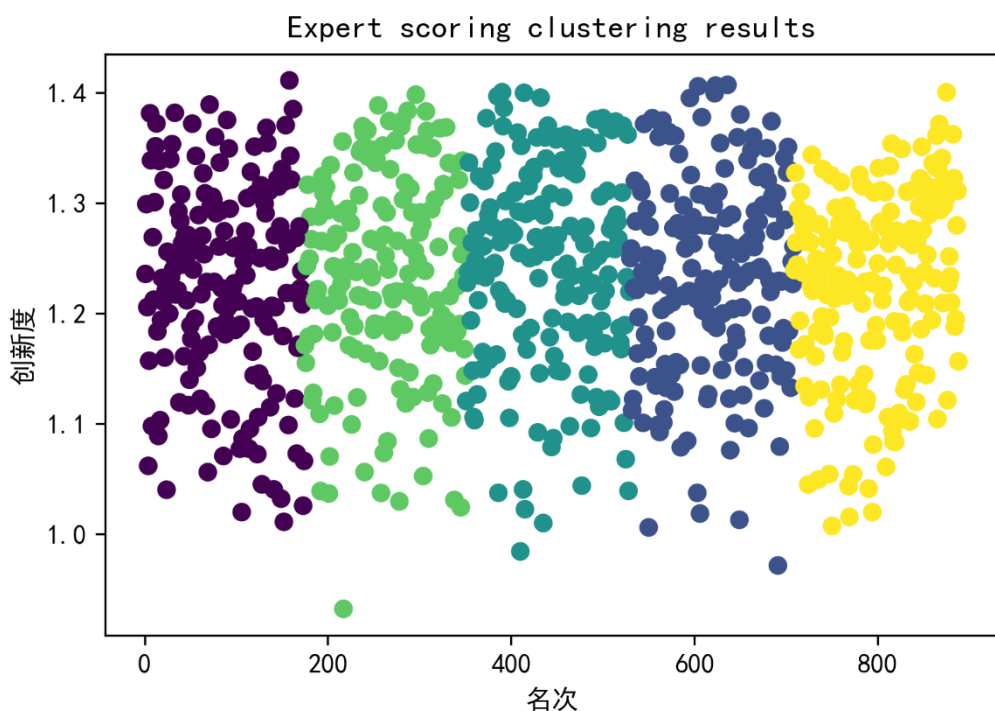


图 6.5 作品创新度分布特点

### 6.3.3 基于创新度的修正分数模型

本节基于 6.3.1 提出的基于极差-争议度的论文创新度评价模型。对二阶段的分数进行调整，将作品创新度作为衡量的指标之一，建立基于创新度的修正分数模型。

设第 $i$ 份作品的调整分数为 $Ex_i$ ， $P_i$ 为组成调整分数各部分的权重系数， $J_i$ 为组成调整分数各部分分数。

$$Ex_i = \sum_{i=1}^n P_i \times J_i \quad (15)$$

使用灰色关联法对各部分的权重进行计算，得出第 $i$ 份作品的调整分数为 $Ex_i$ 。



## 七、问题四模型建立与求解

### 7.1 问题分析及建模思路

由于现在竞赛种类丰富，参赛者的多样性和题目的灵活性使答案多种多样，还存在各位专家打分习惯的差异等各种因素，因此问题四要求要求针对“创新类”竞赛，提出一个完整的评审模型，针对所给数据对其进行求解，并且对现行的评审方案给出改进的具体建议。

建立完整的评审模型需要根据评审的步骤来分步建立模型，按照对评审专家进行分组、作品加密、分配作品以及作品评分这四个步骤，因此首先建立评审专家分组模型，其次进行作品加密，再建立作品分配模型及作品评审模型，最终得到完整的评审模型。对现行的评审方案给出改进的具体建议主要考虑从目前作品评审过程中存在的一些会造成评审结果误差的主观性和客观性的因素出发，以减少这些因素对评审结果的影响而提出改进的建议。如图所示 7.1 所示为建立完整的评审模型的研究思路。

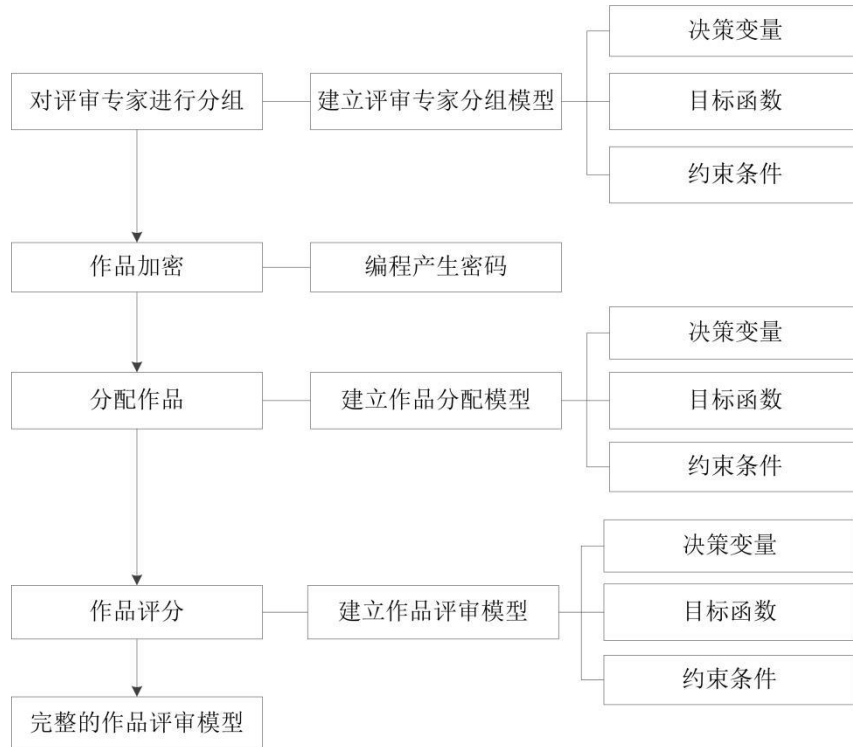


图 7.1 建立完整的评审模型的研究思路

### 7.2 完整评审模型的建立

#### 7.2.1 评审专家分组模型的建立

对于评审作品的专家分组优化问题可描述如下：假设有  $mn$  个评审专家，需要将专家分为  $n$  个专家组，每组  $m$  人。对于专家  $i$ ，其评审速度为  $v_i$ ，评审一致率为  $O_i (0 < O_i < 1, i=1, 2, \dots, mn)$ 。其中  $O_i$  表示专家对每个作品的评价和公众一致的概率。假设某个评价与公众评价不一致即为出现差错，那么  $(1-O_i)$  为专家  $i$  出现差错的概率。在一个专家组中，只有所有专家都出现差错才会造成错判，因此一个有  $m$  个专家的评审组的出错概率为  $\prod_{i=1}^m (1-O_i)$ 。假设在评审作品的过程中中不允许跨组交叉评审，其中专家组  $j$  的评审速度取决于最慢评审专家的速度，问题的目标是（1）评审的总出错率达到最小（2）评审的速度达到最快。

(1) 决策变量的设定

$f_{ij} = 1 \text{ or } 0$ , 若专家 $i$ 被分到 $j$ 组则为1, 否则为0。

(2) 目标函数的建立

第一类目标函数的数学模型为式(16):

$$\min F_1 = \min \sum_{j=1}^m \prod_{i=1}^{mn} (1 - O_i) f_{ij} \quad (16)$$

第二类目标函数的数学模型为式(17):

$$\max F_2 = \max \sum_{j=1}^m \min \{v_i | f_{ij} = 1, \forall i\} \quad (17)$$

(3) 约束条件的制定

基于上述分析, 这里提出的评审专家分组模型中, 存在以下的约束性条件:

$$s. t. \begin{cases} \sum_{i=1}^{mn} f_{ij} = m, j = 1, 2, \dots, n \\ \sum_{j=1}^n f_{ij} = 1, i = 1, 2, \dots, mn \\ f_{ij} = 0 \text{ or } 1, i = 1, 2, \dots, mn, j = 1, 2, \dots, n \end{cases} \quad (18)$$

其中: (1)  $\sum_{i=1}^{mn} f_{ij} = m, j = 1, 2, \dots, n$ 是对各组的专家数约束; (2)  $\sum_{j=1}^n f_{ij} = 1, i = 1, 2, \dots, mn$ 限定一个专家只分配到一个组; (3)  $f_{ij} = 0 \text{ or } 1$ 为变量类型约束。

### 7.2.2 作品加密

为了保证作品

这里对作品加密采用密码技术, 目前人们将密码理论与技术分成两大类: 一类是基于数学的密码理论与技术, 包括公钥密码、分组密码、序列密码、认证码、数字签名、Hash函数、身份识别等; 另一类是非数学的密码理论与技术, 包括信息隐藏、量子密码、基于生物特征的识别理论与技术等<sup>[10]</sup>。

为了保证作品的安全以及评审结果的公正, 对作品编号进行加密是非常有必要的, 常见的加密方法过程繁琐包括古典密码、移位密码、公钥密码等, 但为了让加密更加方便快捷, 这里可以考虑采用编程的方法产生一组随机序列来作为试卷序号的密码, 详细过程这里不再赘述。

### 7.2.3 作品分配模型的建立

这里在问题一的作品分配模型的基础上进行改进, 增加规定即评审专家必须回避本校作品, 在收集作品完成后, 先将各个作品按照学校进行编号, 每个评审组由 $G$ 个来自不同学校, 每组评审 $N$ 个作品。设评委数为 $m$ , 用0-1变量 $F_{ijk}$ 表示第 $i$ 所学校的专家的第 $k$ 个作品, 第 $j$ 所学校作品总数为 $n_j$ ,  $\sum_{j=1}^G n_j = N$ 。问题的目标: (1) 专家所在学校的作品分配给各专家尽可能接近加权平均数 (2) 所有专家评审的作品总数尽可能接近。

(1) 决策变量的设定:

$F_{ijk} = 1 \text{ or } 0$ , 若第 $i$ 所学校的评审第 $j$ 所学校的第 $k$ 份答卷则为1, 否则为0。

(2) 目标函数的建立

$$\min \left[ \sum_{j=1}^M (z_j - y_j), (v - u) \right] \quad (19)$$

(3) 约束条件的制定

基于上述分析，这里提出的作品分配模型中，存在以下的约束性条件：

$$s. t. \left\{ \begin{array}{l} \sum_{k=1}^{n_i} F_{iik} = 0, i = 1, 2, \dots, m; \\ \sum_{i=1}^m F_{ijk} = 5, j = 1, \dots, G, k = 1, \dots, n_j \\ e_{ij}y_j \leq \sum_{k=1}^{n_j} F_{ijk} \leq e_{ij}z_j, i = 1, \dots, m \\ u \leq \sum_{j=1}^G \sum_{k=1}^{n_j} F_{ijk} \leq v, i = 1, \dots, m \\ F_{ijk} = 1 \text{ or } 0, i = 1, \dots, m; j = 1, \dots, G; k = 1, \dots, n_j \end{array} \right. \quad (20)$$

其中  $e_{ij} = \frac{3N}{m(N-n_i)} n_j$  表示  $i$  专家评审  $j$  学校作品的加权平均数。

约束条件中：（1）专家必须回避其所在学校的作品；（2）每个作品经 5 个不同的专家评审；（3） $y_j$ 、 $z_j$  为所有专家评审第  $j$  所学校答卷的下、上界；（4） $u$ 、 $v$  所有评委评阅答卷总数的下、上界；（5） $F_{ijk} = 1 \text{ or } 0$  为变量类型约束。

#### 7.2.4 作品评审模型的建立

这里评分仍然用上述问题二中改进后的方案三的评审模型，同样采取方案三的具体评审过程即进行两个阶段的评审：第一阶段完成上述的专家分组、作品加密后，第一阶段的作品分配中每个作品分配给 5 位专家进行评审，利用上述问题二中改进的标准分模型即式（6）将第二阶段 5 位专家打的 5 个原始分数进行调整为标准分，利用改进后的方案三中提出的用成对比较矩阵进一步调整 5 位专家的权重，计算得到第一阶段的综合得分；对第一阶段中所有作品进行排序按照竞赛的具体要求筛选部分作品进入第二阶段评审，第二阶段的作品分配中每个作品分配给 3 位专家进行评审，同样利用改进的标准分模型将 3 个原始分调整为标准分，最终的得分为第一阶段综合得分与第二阶段 3 位专家的标准分之和。

具体模型参照问题二中改进后的方案三中的评审模型。

#### 7.3 对现行的评审方案的建议

现如今竞赛种类丰富、参赛者的多样性、题目的灵活性以及评分的主观性致使评审过程中存在着各种各样的问题，这些问题在一定程度上对评分结果具有一定影响，针对这些问题提出了一些建议：

（1）评分差异与评审结果可比性问题：不同评委对同一作品可能存在较大的打分偏差，这将直接影响评审结果的可比性。因此针对这一问题需要研究方法来检验评分的一致性以此提高不同评委评分的可比性。

（2）评分标准化方法问题：通常大型竞赛都会对评分进行标准化，评分标准化的方法直接影响评审结果。因此针对这一问题需根据不同特点的竞赛研究合适的评分标准化方法，减小评分方差，降低评分波动。

（3）评审方案设计与优化问题：评审方案的设计直接影响竞赛的公平性和科学性。针对这一问题需要考虑作品的合理分配机制、评委组成、评分标准与机制设计等，除此之外优化评审流程，提高评审效率也至关重要。

（4）专家协商机制的设计：在进行筛选竞赛中最优的少数作品时会采用专家协商机制，能有效解决评分分歧，因此可以考虑研究设计高效的专家协商流程与机制。

（5）创新性评估客观性问题：创新性评估主观性大，需要增强评估的客观性。因此针

---

对这一问题可构建包含多个指标的评估体系，引入更多定量分析。

(6) 极差问题：创新类竞赛中评分极差问题突出，对评分结果也有较大影响。对这一问题进行合理处理利用考虑研究在不同评审阶段对极差问题的合理处理，提高评审公信力。

(7) 评审方案合理性评估问题：可以构建评估模型从而对不同评审方案的优劣进行比较分析，针对不同竞赛情况选择最优方案。

(8) 未来可以考虑从收集更多会对评分结果产生影响的因素：例如收集评审历史记录包括评审专家的评分、复议过程和调整；收集更多有关作品特征包括关键词、引用数量、论文结构等以帮助评估学术质量；收集评审专家和参赛队伍的反馈信息，以不断改进评审流程和模型。

综上所述，评审模型考虑建立多目标优化模型，目标函数可以考虑作品质量打分、评委评分一致性等，设置模型约束条件，如评委评审作品数量限制，确定模型变量，如作品分配策略、评委权重等；模型的求解部分可以考虑根据给定数据，确定模型参数选用求解算法,如线性规划、启发式算法等对模型进行求解；分析现有评审方案存在的问题，如评分标准不一致、评委作品样本数量有限等，增加评委交叉评审、调整评分机制、优化作品分配等，提出新增的数据收集计划，如评委打分习惯、评分偏差等数据。

---

## 八、模型评估与改进

### 8.1 模型优点

1、问题一提出的均匀—交叉多目标作品分配模型，综合考虑专家分配的工作量和分配论文的交叉量，是一类适用于智能算法的多目标优化模型。

2、问题二提出的基于离差的权重标准分全局优化的评审模型，考虑专家个人打分习惯和对打分的影响，从两个方面提出该优化模型。

3、问题三提出的基于极差—争议度的作品创新度评价模型，引入争议度作为衡量作品创新度的因素，采用创新度为评价指标，一定程度上减小低分段极差大作品的影响，并基于创新度提出二轮修正模型，用于修正极差位于中间部分的作品得分。

4、问题四提出的基于多目标优化的评审专家分组模型，考虑到出错率和总评审速度这两个方面，因此考虑对专家进行合理分组使总的出错率达到极小，并且使总的评审速度达到极大；其次对作品进行编程加密，简单易行不易被破解，而且可以在一定程度保证作品的安全以及竞赛的公正；提出的作品分配模型也在之前的基础上进行改进，即保证评审专家必须回避本校作品，也可以提高竞赛的公平性；作品评审模型采用问题二中改进后的方案三即引入改进的标准分模型和专家评审权重，可以在很大程度上减少专家之间评分的系统误差，因此完整的评审模型对降低出错率、提高评审速度、增强作品的安全性和竞赛的公正性这几个方面有一定程度上的作用。

### 8.2 模型改进点

1、问题一中提出均匀—交叉多目标作品分配模型，未考虑到论文分配时间损耗的影响，可继续增加时间最小化目标函数，进行模型改进。

2、问题二中提出的基于离差的权重标准分全局优化的评审模型，虽在一定程度上优化了求解的准确性，但依旧不准确，后续还要进行不断优化。

3、问题三提出的极差—争议度的作品创新度评价模型，由于竞赛时间有限，只考虑了两个因素对创新度的影响，模型的准确性较低，后期将平均值、标准分纳入衡量因素中，进行模型的优化。

4、在理论和实践上仍处于积极探索和有待完善的阶段，尚有一些问题如应用于缺损评分系统仍存在较大误差等问题有待解决。

### 8.3 模型的推广

1、问题一中提出的模型是典型的多目标优化问题，模型不仅适用于本题中创新性竞赛的论文分配中，还可以适用于最优物资分配、生产线任务分配问题中。所提的创新算法，自适应强，可以求解各种优化问题。

2、问题二中提出的模型是多个自变量因素对因变量的影响，适用于回归性分析类、多因素优化类问题。

3、问题三中提出的模型可推广适用于大规模创新型竞赛中极差大，争议度高的文章进行创新性评估问题。

4、问题四中提出的模型根据竞赛的具体要求和实际情况，对一些如标准分模型或专家权重计算等方面进行改进，具有较强的实用性，对其它主观性较强的评分竞赛亦有很强的适用性，但是仍然需要不断对其方法、手段和步骤在理论上进行不断摸索。

---

## 九、参考文献

- [1]曾水兵.从“工程育人”到“文化育人”——大学创新人才培养的再思考[J].江苏高教,2012(05):38-41.DOI:10.13236/j.cnki.jshe.2012.05.014.
- [2]顾雷.创新创业类竞赛对大学生能力培养的研究[J].湖北开放职业学院学报,2019,32(24):13-14.
- [3]许玉德,刘一鸣,沈坚锋等.轨道养护计划多目标模型及求解[J].同济大学学报(自然科学版),2018,46(07):926-933.
- [4]郭东威,丁根宏.群组决策主观评分型竞赛名次的优化模型[J].工程数学学报,2022,39(03):379-388.
- [5]Kim S H, Ahn B S. Interactive Group Decision Making Procedure Under Incomplete Information[J]. European Journal of Operational Research, 1999, 116(3).
- [6]Ramanathan R, Ganesh L S. Group Preference Aggregation Methods Employed in AHP: An Evaluation and an Intrinsic Process for Deriving Members' Weightages[J]. European Journal of Operational Research, 1994, 79(2).
- [7]郭东威,丁根宏,毛俊诚等.群决策论文名次的优化模型[J].统计与决策,2016(18):80-83.DOI:10.13546/j.cnki.tjyjc.2016.18.021.
- [8]史晓峰.基于多目标优化的竞赛评卷系统设计研究[D].哈尔滨工
- [9]郭东威,丁根宏,毛俊诚等.群决策论文型竞赛名次的加权 T 分数法[J].中国科技论文,2015,10(17):2059-2063.
- [10]冯登国.国内外密码学研究现状及发展趋势.2002,23(5):18~26.

---

## 十、附录

### 问题 1 模型求解部分代码

```
import pulp
# 定义问题
prob = pulp.LpProblem("ExpertAllocation", pulp.LpMinimize)

# 定义变量, X[i][j]表示专家 i 是否被分配给作品 j
N = 125 # 评审专家数量
M = 3000 # 作品数量
K = 5 # 每份作品需要评审的专家数量
X = pulp.LpVariable.dicts("ExpertAssignment", ((i, j) for i in range(N) for j in range(M)), 0, 1,
pulp.LpBinary)

# 定义目标函数, 最小化交叉评审次数
# 假设可比性度量 C[i][j]为随机值, 这里简化为 0 和 1 之间的随机数
import random
C = {(i, j): random.random() for i in range(N) for j in range(M)}
prob += pulp.lpSum(X[i, j] * C[i, j] for i in range(N) for j in range(M))

# 添加约束条件
# 每份作品必须由 K 位专家评审
for j in range(M):
    prob += pulp.lpSum(X[i, j] for i in range(N)) == K

# 每位专家可以同时评审多个作品
for i in range(N):
    prob += pulp.lpSum(X[i, j] for j in range(M)) <= 1

prob.solve()
print("Optimal objective value:", pulp.value(prob.objective))

for v in prob.variables():
    if v.varValue == 1:
        print(v.name, "=", v.varValue)

import pandas as pd

allocation = pd.DataFrame.from_dict(X, orient="index", columns = ["allocation"])
allocation = allocation[allocation["allocation"]==1]
allocation.to_excel('问题 1 专家-论文分配情况.xlsx', index=False)
```

### 问题 2 模型求解部分代码

```
import pandas as pd
```

---

```

import matplotlib.pyplot as plt
df = pd.read_excel(r'C:\Users\OS\Jupyter_file\shumo\问题 2 按列显示专家打分.xlsx', ) #
header=None
df

# 按专家编号分组,并获取原始分这一列
expert_scores = df.groupby(['专家编码'])['原始分'].apply(list)

# 初始化结果数据框
result_df = pd.DataFrame()
# 遍历每个专家
for expert, scores in expert_scores.iteritems():

    # 获取打分描述统计
    score_series = pd.Series(scores)
    stats = score_series.describe()

    # 将专家编号及统计结果加入结果数据框
    result_df[expert] = stats

# 将结果写入 Excel 文件
result_df.to_excel('experts_stats.xlsx')

print('结果已保存到 experts_stats.xlsx')

# 对结果数据框进行转置
result_df = result_df.T

# 将结果写入 Excel 文件
result_df.to_excel('experts_stats.xlsx')

# 观察专家打分情况聚类
# 选择特征
X = df[['mean', 'std']]
# KMeans 聚类,设定聚类数为 5
kmeans = KMeans(n_clusters=3)
kmeans.fit(X)
# 获取样本聚类标签
labels = kmeans.labels_

# 添加到 DataFrame
df['cluster'] = labels
# 将结果写入 Excel 文件
df.to_excel('experts_stats.xlsx')
# 绘制聚类结果散点图
plt.scatter(X['mean'], X['std'], c=labels)

```



---

```
plt.xlabel('Mean', )
plt.ylabel('Std', )
plt.title('Expert scoring clustering results',)

plt.savefig(fname='clustering.png', format='png', bbox_inches = 'tight', dpi=300, )
plt.show()
# 输出按聚类结果排序
print(df.sort_values('cluster'))
```

```
# 获取作品标准分的数据分布
```

```
import pandas as pd
import matplotlib.pyplot as plt
```

```
# 读取数据
```

```
df = pd.read_excel('问题 2 作品对应的标准分.xlsx')
```

```
df_2 = df[['标准分 1', '标准分 2', '标准分 3', '标准分 4', '标准分 5']]
```

```
# 计算每行的平均分和标准差
```

```
AVE= df_2.mean(axis=1)
```

```
std = df_2.std(axis=1)
```

```
media = df_2.median(axis=1)
```

```
jc = df_2.max(axis=1) - df_2.min(axis=1)
```

```
df_2['平均分'] = AVE
```

```
df_2['标准差'] = std
```

```
df_2['中位数'] = media
```

```
df_2['极差'] = jc
```

```
df_2['作品名次'] = df['名次']
```

```
df_2.to_excel('问题 2 作品标准分分析 analysis.xlsx', index=False)
```

```
df_2
```

```
问题 3 数据分析代码
```

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
df = pd.read_excel('数据 2.1 - 副本.xlsx')
```

---

```

cols = ['11 专家编码', '12 专家编码', '13 专家编码', '14 专家编码', '15 专家编码']
data = df[cols].apply(lambda x: ','.join(x.astype(str)), axis=1)

unique_codes = set(data.str.split(',').sum())

# 将唯一编码保存到 Excel 文件
unique_df = pd.DataFrame(list(unique_codes), columns=['唯一编码'])
unique_df.to_excel('数据 2.1 专家编码.xlsx', index=False)

# 对数据 2.1 进行继续处理
import pandas as pd

df = pd.read_excel('数据 2.1 - 副本.xlsx')
df.head(5)

columns = ['名次', '第 2 次评审标准分极差', '11 标准分', '12 标准分', '13 标准分', '14 标准分', '15 标准分', '1 极差', '21 标准分', '22 标准分', '23 标准分']
df_1 = df[columns]
df_1.head(3)

ext1 = df_1['1 极差']

# plt.boxplot([ext1, ext2], labels=['第一阶段', '第二阶段'])

plt.rcParams['font.sans-serif'] = ['SimHei'] # 用来正常显示中文标签
plt.rcParams['axes.unicode_minus'] = False # 用来正常显示负号

fig = plt.figure(1, (7,3), dpi=250)
ax = plt.subplot(111)
plt.sca(ax)

plt.boxplot(ext1)# /60 是为了得到分钟
plt.ylabel('极差分值')
plt.xlabel('第一次评审阶段极差')
plt.title('第一次评审阶段极差')
plt.savefig('第一次评审阶段极差.svg')

plt.show()

ext1_des = df.describe()
ext1_des
ext1_des.to_excel('数据 2.1 所有数据按列的 describe.xlsx', index=False)

```

---

```
# 选取所有第一阶段打分数据，并对其进行 describe
```

```
a = pd.read_excel('数据 2.1 - 副本.xlsx', usecols=['11 标准分','12 标准分','13 标准分','14 标准分','15 标准分'])
```

```
a = a.values.ravel() # 将 DataFrameFlatten 拼接为一维数组
```

```
x1 = pd.Series(a).describe() # 对 Series 进行 describe 描述统计
```

```
ext1_des['第一阶段所有打分描述'] = x1
```

```
ext1_des
```

```
ext1_des.to_excel('数据 2.1 所有数据按列的 describe.xlsx')
```

```
# 筛选出进入第二轮的所有数据，再对其打分情况进行分析
```

```
x2 = df[df['21 标准分'].notna()]
```

```
x2.head(3)
```

```
# 第二阶段所有打分 describe
```

```
a = x2[['21 标准分','22 标准分','23 标准分']]
```

```
a = a.values.ravel() # 将 DataFrameFlatten 拼接为一维数组
```

```
x1 = pd.Series(a).describe() # 对 Series 进行 describe 描述统计
```

```
ext1_des['第二阶段所有打分描述'] = x1
```

```
x1
```

```
# 第 1 阶段 进入第 2 阶段的作品 所有打分描述
```

```
a = x2[['11 标准分','12 标准分','13 标准分','14 标准分','15 标准分'],]
```

```
a = a.values.ravel() # 将 DataFrameFlatten 拼接为一维数组
```

```
x1 = pd.Series(a).describe() # 对 Series 进行 describe 描述统计
```

```
ext1_des['第 1 阶段 进入第 2 阶段的作品 所有打分描述'] = x1
```

```
x1
```

```
a = x2[['1 极差'],]
```

```
a = a.values.ravel() # 将 DataFrameFlatten 拼接为一维数组
```

```
x1 = pd.Series(a).describe() # 对 Series 进行 describe 描述统计
```

```
ext1_des['第 1 阶段 进入第 2 阶段的作品 极差描述'] = x1
```

```
ext1_des = ext1_des.T
```

```
ext1_des.to_excel('数据 2.1 所有数据按列的 describe.xlsx')
```

```
ext1_des
```

```
## 展示数据中获奖以及未获奖描述 未获奖、三等奖、二等奖、一等奖
```

```
df = pd.read_excel('数据 2.1 - 副本.xlsx')
```

```
x2 = df[df['奖项']!='未获奖']
```

```
a = x2[['11 标准分','12 标准分','13 标准分','14 标准分','15 标准分'],]
```

```
a = a.values.ravel() # 将 DataFrameFlatten 拼接为一维数组
```

---

```

x1 = pd.Series(a).describe() # 对 Series 进行 describe 描述统计

ext1_des=ext1_des.T
ext1_des['未获奖'] = x1

x2 = df[df['奖项']=='三等奖']
a = x2[['11 标准分','12 标准分','13 标准分','14 标准分','15 标准分']]
a = a.values.ravel() # 将 DataFrameFlatten 拼接为一维数组
x1 = pd.Series(a).describe() # 对 Series 进行 describe 描述统计
ext1_des['三等奖'] = x1

x2 = df[df['奖项']=='二等奖']
a = x2[['11 标准分','12 标准分','13 标准分','14 标准分','15 标准分']]
a = a.values.ravel() # 将 DataFrameFlatten 拼接为一维数组
x1 = pd.Series(a).describe() # 对 Series 进行 describe 描述统计
ext1_des['二等奖'] = x1

x2 = df[df['奖项']=='一等奖']
a = x2[['11 标准分','12 标准分','13 标准分','14 标准分','15 标准分']]
a = a.values.ravel() # 将 DataFrameFlatten 拼接为一维数组
x1 = pd.Series(a).describe() # 对 Series 进行 describe 描述统计
ext1_des['一等奖'] = x1
ext1_des

ext1_des=ext1_des.T
ext1_des.to_excel('数据 2.1 所有数据按列的 describe.xlsx')

ext1_des[['未获奖','三等奖','二等奖','一等奖']]

ext1_des=ext1_des.T
ext1_des.to_excel('数据 2.1 所有数据按列的 describe.xlsx')

```