



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 国防科技大学

参赛队号 23910020010

1.黎 森

队员姓名 2.姚雪湄

3.谭云雷

关注公众号，获取更多资料

中国光谷·“华为杯”第二十届中国研究生 数学建模竞赛

题 目 创新型竞赛评审全流程优化设计方法探索

摘 要：

随着社会创新活力日益增强，企业、学会等举办的创新型竞赛正向着参赛规模扩大化，参赛地域全国化以及参赛水平多样化的方向发展，在有限的评审资源的约束下，如何设计一整套完整的评审流程，将成为支撑比赛公平性的关键一环。科学的评审流程应该具有评阅双向均匀化、一致性高、容错性强、发掘创新型准等特点。本文以“创新型”竞赛为任务背景，问题一至问题三分别致力于解决评审全流程中的分发评审作品、分数标准化计算、极差调整、多阶段赋分排名等关键环节的科学性与问题优化。

问题一，首先分析了在 3000 份作品，125 位评委，5 次评阅的条件下，能够产生 7750 个评委对，每对评委之间的平均交集数为 3.87。其次，构建了论文分配模型，根据作品的评阅次数要求、评委的工作量应保持在同一水平的要求，建立了模型的约束条件。将评委作品集合的交集尽可能均衡的要求转化为两个目标函数：(1) 彼此之间的存在交集的评委对尽可能多；(2) 所有评委的最大交集数与最小交集数之差尽可能少。即先保证所有评委之间有交集，再考虑作品交集均衡这一目标。基于该模型，设计了基于模拟退火的两阶段交叉算法求解，通过改变交叉幅度与交叉对象，分两阶段进行方案优化。经过 10 次实验，求得的分发方案的评委最大最小交集差的平均值为 94，没有交集的评委对数的平均值为 104.9，交集度为 98.64%，明显优于另外两种随机分配方案。

问题二部分，核心问题在于设计标准化计算公式以及改善成绩计算方案。首先，我们从每个专家、每件作品的角度进行成绩分析，专家总体层面，存在评分均值差异大的情况，作品层面也存在极差大的问题。个体层面，不同专家的评分偏好差异大。我们分别采用现有的两个方案进行排名，并提出了优劣对比指标：排名重合度、平均乱序度，最大乱序度以及分数标准差。通过计算，去除极差的方案二效果优于仅采用标准化的方案一。我们针对针对这两方面，我们提出了 WLT 修正公式——基于评分尺度系数折算的综合评分加权 T 分数法。该公式由专家综合评分、评分尺度系数以及 T 分数法标准化三部分组成：(1) 专家综合评分考虑了评委维度的偏差度与区分度指标，其中偏差度 w 指标反映了组内误差，而区分度 s 反映了组间误差，二者通过计算比值即可得到每位专家的综合评分，综合评分反映了专家的可靠程度。(2) 评分尺度从作品、专家两个维度综合运用数据，作为标准分之间进行加和的赋权值。(3) T 分数法，即原题中的标准化公式。我们利用修正公式对第一阶段数据进行标准化，结合第二阶段评分重新排名，将得到的排名结果与一等奖排名进行重合度计算，得到的重合度为 11.11%，平均乱序度为 6，分数标准差为 0.4974，小于现有方案，具有较好的一致性。

问题三部分，(1) 我们首先基于附件数据分析了两阶段前后的整体成绩与极差变化。从排名变化的统计指标来看，第二阶段中获得一等奖作品的逆序度为 0.2680，乱序度为 328，而入选第二阶段作品的整体逆序度为 0.3，乱序度为 12250，均高于一等奖作品。对比两阶段与不分阶段的优劣方面，通过绘制三个阶段极差的频率直方图和箱线图发现，第二阶段的分数均匀性优于第一阶段。(2) 其次，我们建立了程序化极差调整模型，主要通过确定极差阈值、识别调整模式、分类器模型自主学习预测以及计算调整幅度四个方面完成自动化调整极差的功能。其中极差调整阈值为 20，超过 20 的即为大极差，而极差调整模式包括六类，即调整最小值、中间值以及最大值中的一个或两个。分类预测模型的选择方面，我们通过训练，得到 Adaboost 集成优化、KNN 以及 Bagging 三类模型的准确率分别为 85%、69.2%、82.1%，因此最终选择 Adaboost 模型。最后我们利用训练好的模型对数据

2.1 中的第一阶段非高非低类作品进行极差调整，筛选出了大极差作品为**编号 199、248、250、260、266、269、273** 七件代表性作品，认为其具有创新型的潜能，因此调整极差后能够进入第二轮评审。

问题四在现有流程基础上，将评审流程分为：(1) 论文分配 (2) 基于序关系的淘汰机制 (3) 极差分数修正，WTF 修正分数排序 (4) 最终复审四个阶段。在论文分配阶段，基于问题一中的论文分配模型，进一步考虑了学校与评委、作品之间的关系，对论文分配模型进行了完善。在基于序关系的淘汰机制阶段，结合等级制作品评阅水平准确度较高的优点，在这一阶段进行 3 次论文评阅，将所给分数转化为正态上侧百分比 (NUD) 排名并，根据排名划分了高序列 (S)、中间序列 (M)、低序列 (L)3 个等级。然后，将 3 次评阅结果都是低序列 L,L,L 或者 1 次中序列 2 次低序列 L,L,M 的作品淘汰。剩余作品继续完成 2 次评阅，极差修正后，将 5 次评阅的结果进行 WTF 标准化，进入复审阶段，复审阶段与原有流程保持一直。对基于序关系的筛选机制进行合理性分析后，在数据集 2.2 上完成了该流程的实验。最终结果显示，在 (2) 阶段被淘汰的作品共有 895 个，进入复审阶段的作品与原有方案的重合度为 100%。不改变现有结果的基础上，评委的评阅工作量相较于原有方案减少了 17.77%。

关键词：多目标规划；相关性分析；集成优化；排名方案评估

关注公众号：建模忠哥，获取更多资料

目录

1. 问题重述	5
1.1 问题背景	5
1.2 问题提出	5
2. 模型假设	6
3. 符号说明	6
4. 问题一的模型建立与求解	7
4.1 问题的描述分析	7
4.2 多目标的论文交叉分配数学模型	7
4.2.1 论文最大交叉量分析	7
4.2.2 目标函数与约束	7
4.3 模型求解和分析	8
4.3.1 基于模拟退火的两阶段任务交叉算法	8
4.3.2 两种随机策略的分配方案	10
4.3.3 实验结果与分析	11
5. 问题二的模型建立与求解	11
5.1 问题的描述分析	11
5.2 模型的建立与求解	12
5.2.1 成绩分布特点的分析与可视化	12
5.2.2 不同方案下的作品排名	15
5.2.3 方案优劣的比较与分析	17
5.2.4 标准化修正模型：基于评分尺度系数折算的综合评分加权 T 分数法	18
6. 问题三的模型建立与求解	22
6.1 问题的描述分析	22
6.2 数学模型的建立	22
6.2.1 两阶段方案整体性分析	22
6.2.2 极差模型的建立与求解	25
6.3 模型求解和分析	35
7. 问题四的模型建立与求解	36
7.1 问题描述与分析	36
7.2 数学模型的建立	37
7.2.1 评审流程说明	37
7.2.2 考虑学校的论文分配模型	37
7.2.3 基于序关系的筛选流程	39
7.3 模型求解和分析	40
7.3.1 求解结果与分析	40
7.3.2 改进建议	40
8. 模型评价	42
8.1 模型的优点	42
8.2 模型的缺点	42
8.3 模型的改进和展望	43

参考文献.....	43
附录 A 复议作品模式识别训练集	44
附录 B 决策树模型预测散点图与混淆矩阵结果图.....	45
附录 C KNN（三次）模型预测散点图与混淆矩阵结果图	46
附录 D Bagging 集成树模型的最小分类误差图.....	47
附录 E 两种模式下的样本训练属性集.....	48
附录 F 问题三核心代码：程序化调整极差.....	49
附录 G 问题三核心代码：基于集成优化的分类器模型	56

公众号关注：建模忠哥，获取更多资料

1. 问题重述

1.1 问题背景

现阶段创新类竞赛的发展与组织日趋成熟，参赛人数呈逐年递增的趋势。其中规模较大的竞赛，一般采用两阶段（网评、现场评审）或三阶段（网评、现场评审和答辩）评审。但是，由于该类竞赛开放程度高，对于特定问题往往没有一个具体的标准答案，需要评审专家根据命题人提出的评审框架或者建议独立评审。所以，对同一份作品，不同评委的评分可能存在较大差异。在实际情况中，当竞赛规模大，评委的人数众多时，关于作品创新的争议性问题会更加凸显。

目前，各项创新类竞赛不断地调整评审方案，主要采用数理统计手段或者科学规范的评审流程来更加公正地评估作品的真实水平。针对该问题目前主要的处理方案如下：

1. 对每位评审专家的评分进行标准化 0，按作品将标准分相加得每件作品的总分，然后依总分排序；
2. 去掉同一份作品得分中的最高分、最低分，再将剩余评分相加，最后依总分排序；
3. 同一份作品如果专家的评分差异（极差）较大，组织相关专家协商调整，将调整后得分相加，再依总分排序；
4. 当竞赛规模很大时，首先利用上述方案对作品进行初选，再对初选入围的作品组织专家评审（第二阶段评审）或经过答辩等环节确定获奖名单。

以上方案都有自己的优势，但也存在一定的局限性。现有方案中针对大规模创新类竞赛评审研究较少，流程模式简单，简单地依据多位评委评分的总和进行排序并不能够反映出作品的真实水平，即与客观值之间存在偏差。因此，探讨大规模创新类竞赛评审方案的公正性、公平性和科学性具有深远意义。

客观角度而言，如果能够增加每件作品的评阅次数，对于作品的评审工作将会更加有利，得到的结果也会更加公平。但是，由于客观条件的限制如专家数量、资金、时间等，该条件难以满足。评审专家减少后，评审工作的误差也会相应地变大。不过，大部分大规模创新类竞赛获奖比例小于 50%，有些误差并不影响是否能够获奖。因此，在不影响评审结果条件下，为了适应评审专家人数较少的现状，许多竞赛评审方案采用两阶段评审办法。

1.2 问题提出

针对创新类竞赛评审的新特点，为探讨研究新的评审机制，提出以下问题：

问题一：针对 3000 只参赛队伍，125 位评委，每份作品评阅 5 次，建立数学模型确定最优的交叉分发方案，使评委的作品交集尽可能均衡并讨论方案的相关指标与实施细节。

问题二：基于 2 种以上已有或者自行设计的评审方案，分析专家、作品原始成绩、调整后成绩的分布特点，采用不同方案进行排序并比较方案优劣。针对评审问题，设计新的标准分计算方法，并利用可信度较高的数据改进标准分计算模型

问题三：根据已有数据，讨论两阶段的成绩整体变化和两阶段极差整体变化情况分析两阶段法与不分阶段评审方案的优劣。基于极差调整数据建立极差模型，并给出自动化调整第一阶段非高非低分作品的大极差的办法。

问题四：针对“创新类”竞赛，给出完整的评审模型并研究其求解过程，或者对现行的评审方案给出改进的具体建议。

2. 模型假设

为简化模型，便于计算与分析，本文提出以下几点假设：

- 创新类竞赛中作品总体呈现成绩与创新性正相关的趋势，但仍存在少部分创新性较强、提出了首创性理念或成果的作品被误判的情况。
- 调整极差的模式存在主观性，调整后主要影响平均值与极差，因此调整极大值与同时调整极大值与中间值的效果一致即可判定为同一调整模式。
- 评委的打分体现了他对作品优劣的判断，当他认为作品 A 优于作品 B 时，那么分值应该与其判定具有一致性，即 A 的分数要大于 B 的分数。
- 假设当前的数据中，所有作品都是可比的，不存在两篇论文围绕两个互斥的主题进行设计的情况。
- 所有的评委在进行作品评阅前都经过了相关的培训，熟悉评分的标准与细则，能够做到公平、客观的打分。

3. 符号说明

本文使用的符号和含义说表如表1所示：

表 1: 符号与含义说明

符号列表	含义说明
k	第 k 篇作品, $k \in [1, n]$
n	作品数
t	每篇作品的评委评阅数
m	评委数
m_{ij}	第 i 位评委是否属于学校 j
s_{kj}	第 k 份作品是否属于学校 j
x_{kj}	论文 k 与评委 i 之间的决策关系
d_{iz}	评委 z 与评委 i 之间是否存在作品的交集
s_{iz}	评委 i 与评委 z 之间的论文的数量
$cross$	方案的中评委的交叉程度
i	第 i 位评委, $i \in [1, m]$
y_{ik}	第 i 位评委对第 k 篇作品的修正标准化评分
μ_i	第 i 位评委的评分均值
σ_i	第 i 位评委的评分方差
a_{ik}	第 i 位评委对第 k 篇作品的原始评分
w_i	第 i 位评委的偏差度
s_i^2	第 i 位评委的区分度
λ_i	第 i 位评委的综合评分
S'_k	第 k 篇作品的理想分数
W_{ik}	第 i 位评委对第 k 篇作品的近似评分尺度
W_i	第 i 位评委的最优评分尺度
y_k	第 k 篇论文的标准分集合
y'_k	第 k 篇论文的修正标准分集合

符号列表	含义说明
R	排名重合度
$\text{degree}_{\text{disorder}}$	平均乱序度
$\text{degree}_{\text{max}}$	最大乱序度
σ_{plan}	分数标准差
η	常数项
D_{up}	上偏距离
D_{down}	下偏距离
Dev_d	偏差距离差
Dev_{degree}	偏差度

4. 问题一的模型建立与求解

4.1 问题的描述分析

在评审过程中，常规思路采取的论文分配方法为随机分配，每份作品要求多位评委独立评审。为了增加不同评审专家所给成绩之间的可比性，不同的专家评审的作品集合之间应该要有一些交集，但是部分专家之间的交集过大时，意味着另一部分专家的交集必然减少，会导致成绩可比性的减弱。针对该问题，前人已经做过许多基础性的研究 [文献][文献]，但存在以下问题：

- 目标函数的设置大多数为评审专家的工作量大致相当，即每位评委所分配的作品数量之间的差值应该尽可能少，没有考虑到评委评阅作品的也需要存在交集
- 该问题本质上为组合优化问题，是 NP-hard 问题，当问题规模上升时如何快速得出有效合理的分配方案是一个难题

4.2 多目标的论文交叉分配数学模型

4.2.1 论文最大交叉量分析

针对参赛队伍总数 n 、评委评阅次数 t 、评委数量 m 确定时，可以计算出不同评委之间的交叉组合一共有 $m * (m - 1) / 2$ 种，而通过每篇论文的最大评阅次数 t 则可以计算出，一只参赛队伍能够让 t 个不同评委产生 $t * (t - 1) / 2$ 个的交叉对，一共有 n 只参赛队伍，那么能够确定所有参赛队伍能够带来的评委总交叉次数为 $n * t * (t - 1) / 2$ 。因此可以分析出每个评委之间的交叉组合的平均值应该为 $n * t * (t - 1) / (m * (m - 1))$

由此，考虑 3000 只参赛队伍，125 位评委，每个队伍的参赛作品需要被评阅 5 次，可以计算出不重复的评委交叉对总数为 7750 个，作品能够带来的总的交叉次数为 30000 个。因此，评委交叉对的平均交叉次数为 3.8710 次，即当前的评阅条件下，最理想的分配情况是所有评委之间都能够保证评阅的作品之间相互有交集，次数为 3.8710。

4.2.2 目标函数与约束

由题目要求可知，建立的论文交叉分配方案重点关注不同评委之间的作品要求存在适当数量的交集。决策变量 x_{ji} 表示作品 j 与评委 i 之间的决策关系， $x_{ji} = 1$ 意味着作品 j 被评委 i 评阅；等于 0 则反之； $c_{iz} = 1$ 则表示评委 i 与评委 z 之间的作品交集个数； $d_{iz} = 1$ 表示评委 i 与评委 z 之间是否存在交集， $d_{iz} = 1$ 意味着两位评委之间存在交集，为 0 时则反之。因此，我们不妨假设，对于该分配问题，我们需要满足两个方面要求：

$$\begin{aligned} \max \left\{ f_1 = \sum_{i=1}^{m-1} \sum_{z=i+1}^m d_{iz} \right\} \\ \min \{ f_2 = \max c_{iz} - \min c_{iz} \} \end{aligned} \quad (4.1)$$

目标函数：(1) 目标函数 1：评委评阅的作品之间存在交集，即无交集时 0 的个数最多，该目标效益越低；评委之间全部存在交集时，该目标效益最高为 7750。
(2) 目标函数 2：交集的数量应该要尽可能均衡，即评委之间的最大交集数与最小交集数的差尽可能小，反映了分配方案总体交集均匀的程度。

$$\begin{cases} \sum_{i=1}^m x_{ji} \leq \lceil n/m \rceil + 1 \\ \sum_{i=1}^m x_{ji} \geq \lfloor n/m \rfloor - 1 \\ \sum_{j=1}^n x_{ji} = 5 \\ c_{iz} = \sum_{i=1}^m (x_{ji} \wedge x_{zi}) \\ d_{iz} = \begin{cases} 1 & \text{otherwise} \\ 0 & \text{if } c_{iz} == 0 \end{cases} \\ x_{ji} = 0 \text{ or } 1, i = 1, 2, \dots, m; j = 1, 2, \dots, n \end{cases} \quad (4.2)$$

约束说明：为平衡所有评委的工作强度，将评委评阅的论文数量限制在一个合理区间。在约束 4.2 中：约束 1 表示每位评委的工作量不能过大，限制为每位评委需要评阅的数量的平均数向上取整加 1；约束 2 表示每位评阅的数量为平均数向下取整减 1；约束 1 和约束 2 保证了评委的评阅数量的差值最多 3，使得各评委间的工作量相对均衡；约束 3 表示每个参赛作品必须被评阅 5 次；约束 4 表示两位评委评阅作品交集数量的总和；约束 5 表示，当两位评委评阅的作品集有交集时，变量 c_{jk} 的取值；决策变量为 0-1 变量。

4.3 模型求解和分析

4.3.1 基于模拟退火的两阶段任务交叉算法

实际求解过程中，该问题的决策变量维度为 3000×125 ，经过分析后，由于中间变量 c_{jk} 和 d_{jk} 的存在，计算目标函数值时存在大量非线性约束。在合理的时间与有限的计算条件下很难使用 Lingo、Cplex 等求解器进行计算，存在非线性约束难以转化、数据溢出、求解时间长等难点。目前，针对大规模问题，比较有效的处理方式是采用多阶段的优化流程机制，设计符合背景逻辑的操作算子，基于简单高效的元启发式算法进行求解。基于上述分析，针对问题一，设计了基于模拟退火的两阶段任务交叉算法，通过分阶段规划优先满足各个评委评阅的作品集合中必须要有交集存在这一目标，然后通过一定的启发式规则，指导计算机选择符合特殊条件的评委进行小段的任务交换，以满足交集的数量要尽可能的均衡。算法步骤如下：

- 步骤 1：**将所有任务分配到评委
 - 步骤 2：**随机选择两位评委，将除相同作品之外的其他作品进行随机点位作品交换；
 - 步骤 3：**满足小段优化过程，转入步骤 4；否则，重复步骤 2；
 - 步骤 4：**选择交集数最大的评委，选择一位交集数较小的评委，将两者进行小段点位的作品交换；
 - 步骤 5：**满足终止条件则停，否则，重复步骤 4.
- 算法流程机制，如图 1 所示：

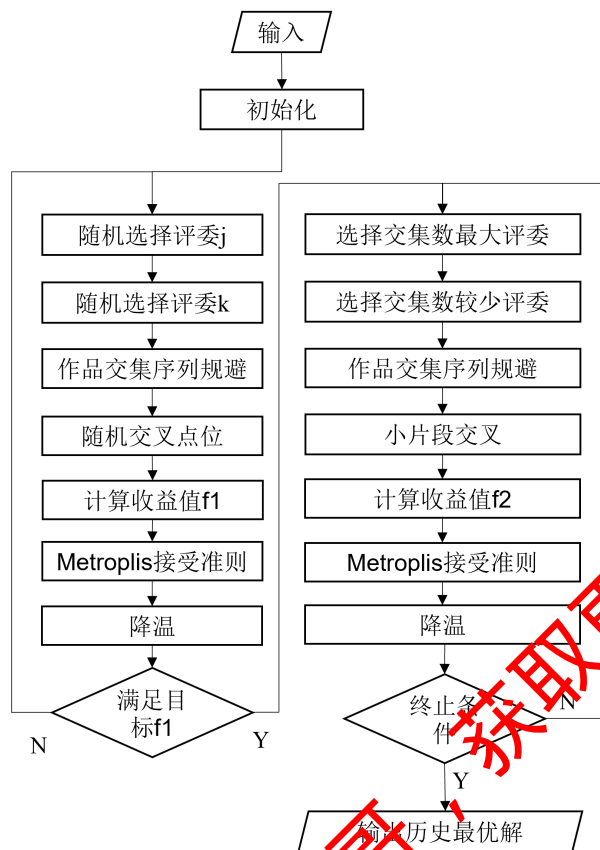


图 1: 算法流程机制

流程机制说明: 第一阶段为以 f_1 为主要目标进行优化。先要满足所有评委之间存在交集。这个阶段中，评委的选择主要以随机、首个最大交集、与其他评委交集最少等算子确定，同时，交叉点位也是基于随机数进行选择，目的在于通过随机兼一定程度的优化方向指引使得各个评委评阅的作品之间建立起交集。

进入第二阶段后，更加注重平衡过程，选择的评委为交集最大与交集最少的评委，当存在多个时采用最后一个数据。选择最大与最小的目的在于使得二者差值能够更加均衡，小片段交叉则保证了进行片段交换后，两位评委与其他评委作品之间的交集被破坏的概率减小，增加了可行解出现的可能性。

编码方式与交叉操作: 由于该问题决策变量过大，因此结合问题背景设计了合适的编码方式。考虑到每个评委所评阅的作品数目大致相同，因此，能够把作品的序号作为基因，和其他作品序号共同组成了评委 i 的编码片段。交叉方式如图2。



图 2: 评委作品集合交叉示意图

在进行交叉操作时，需要注意，由于两个评委之间的作品可能存在重复的情况，而一个作品只能由五位不同的专家独立评阅。因此，交叉之前，需要确定两位评委评阅的作品中重叠的部分在哪，将这部分编号放于片头，该片头之后的作品能够等数目的随意交叉。

该编码与交叉方式的好处在于，只要不交换作品集合集合中的已有项，其他作品无如何交叉变换也不会影响每个作品被若干个评委独立评价，也不会改变作品的评阅次数。

4.3.2 两种随机策略的分配方案

1、作品优先的随机分配方案

步骤 1：遍历作品集合；

步骤 2：选择 s 个不同评委为该作品评分；

步骤 3：剔除评阅数量已到 n/m 个作品的评委；

步骤 4：重复步骤 2-3 至所有作品完成遍历。

优先满足作品的评阅次数要求，随机选择不同的评委为作品打分，能够满足不同评委工作强度一致的要求，随机性的存在也能够使得不同作品将评委串联起来。

2、评委优先的随机分配方案

步骤 1：生成 1 组 $1-m$ 的随机数，取前 n/m 个作品分给评委，直到剩余随机数无法再分配给下一位评委，剩余随机数保存；

步骤 2：把上一次剩余的随机数与下一组随机数拼接，继续分给评委，直到剩余随机数无法再分配给下一位评委，剩余随机数保存；

步骤 3：重复步骤 2，直到生成 s 组随机数。如果最后一位评委分到的论文不足 m/n 篇，随机选取不重复的随机数补足。

评委优先的随机分配策略则是优先满足评委的评阅次数要求，将一组任务序列打乱以后，每个评委按照顺序领取相应的评阅作品。但该方法的缺点很明显，该方案分配的随机性仅仅来自于每组作品的序列，则属于同一组的评委的评阅作品的交集则必然为 0。每次评阅作品会导致 n/m 个评委之间相互没有交集。

3、指标说明

为了对比两个分配方案的实际效果，在计算机中进行模拟实验，每种策略生产 10 种不同的方案，而后统计两者在 $f1$ 与 $f2$ 上的目标值大小。一共有 125 位评委，考虑不同评委之间不重复的两两配对，最多有 7750 对不同评委评阅了相同作品，将 $(7750 - \text{无交集的评委对个数}) / 7750$ 作为一个指标，反映分配方案内各个专家评阅作品的交叉程度，计算方法如式 4.3。

$$cross = 2 \frac{\sum_{i=1}^{n-1} \sum_{z=i+1}^n d_{iz}}{n(n-1)} \quad (4.3)$$

其他指标如最大交集、最小交集、最大与最小的差也可能衡量任务分配方案的结果是否均衡。

4.3.3 实验结果与分析

表 2: 作品分配方案相关指标计算

实验 编号	作品优先的分配方案				评委优先的分配方案				两阶段任务交叉分配方案			
	最大 交集	最小 交集	差 值	无 交集	最大 交集	最小 交集	差 值	无 交集	最大 交集	最小 交集	差 值	无 交集
1	15	1	15	1	15	0	15	1533	9	0	9	98
2	15	2	15	2	14	0	14	1548	9	0	9	101
3	14	3	14	3	15	0	15	1570	10	0	10	108
4	16	4	16	4	13	0	13	1547	9	0	9	105
5	17	5	17	5	16	0	16	1546	9	0	9	107
6	15	6	15	6	14	0	14	1554	10	0	10	109
7	17	7	17	7	15	0	15	1543	9	0	9	108
8	15	8	15	8	16	0	16	1537	10	0	10	102
9	16	9	16	9	17	0	17	1547	10	0	9	104
10	15	10	15	10	16	0	16	1548	10	0	10	107
平均	14.3	0	14.3	145.5	15.1	0	16	1547.9	9.4	0	9.4	104.9
交叉度		98.12%				80.03%					98.64%	

由表2可见，基于评委优先的分配方案在 3000 只队伍，125 位评委的分配中取得的效果最差。原因如下：由于 3000 只队伍分配每组随机数刚好能够使得 25 位评委满足评阅要求，所以每组随机数之间没有交叉。这意味着，在 5 轮作品分配中，每一轮的 25 位评委之间是没有交集的（因为 3000 份作品，刚好能够被 25 位评委刚好分到 120 份作品），在这个过程中，会导致 $(1+2+\dots+24) = 300$ 对评委之间没有交集，5 组则为 1500，如果没有经过方案优化，则评委有限的分配方案评委与评委之间将存在较大的间隔，作品的交集数量偏少。该方法在分配时虽然实现简单，但容易带来过多评委之间没有交集的情况出现。

两阶段任务交叉分配方案充分利用了评委作品集合的特点，设计了类似基因片段的交叉操作，通过交叉操作，使得所有评委之间能够交换作品集，寻找最大与最小的评委进行片段交叉对于缩小二者之间的差值也有一定的帮助。因此，交集的最大最小值较随机分配方案较低，平均差值为 9.4，评委之间的交叉度最大为 98.64%。

5. 问题二的模型建立与求解

5.1 问题的描述分析

问题二研究的核心在于第一阶段的评审方案制定、评审数据的分布特点分析以及标准化公式的改进。由于每位作家只能评阅一部分的作品，而对于不同专家拿到的作品集合水平可能存在差异性，导致不同专家的评阅分数分布存在差异。标准化公式的目的在于消除组间差异性，使得不同专家的成绩分布尽量趋于一致。因此问题的研究逻辑如图3所示：

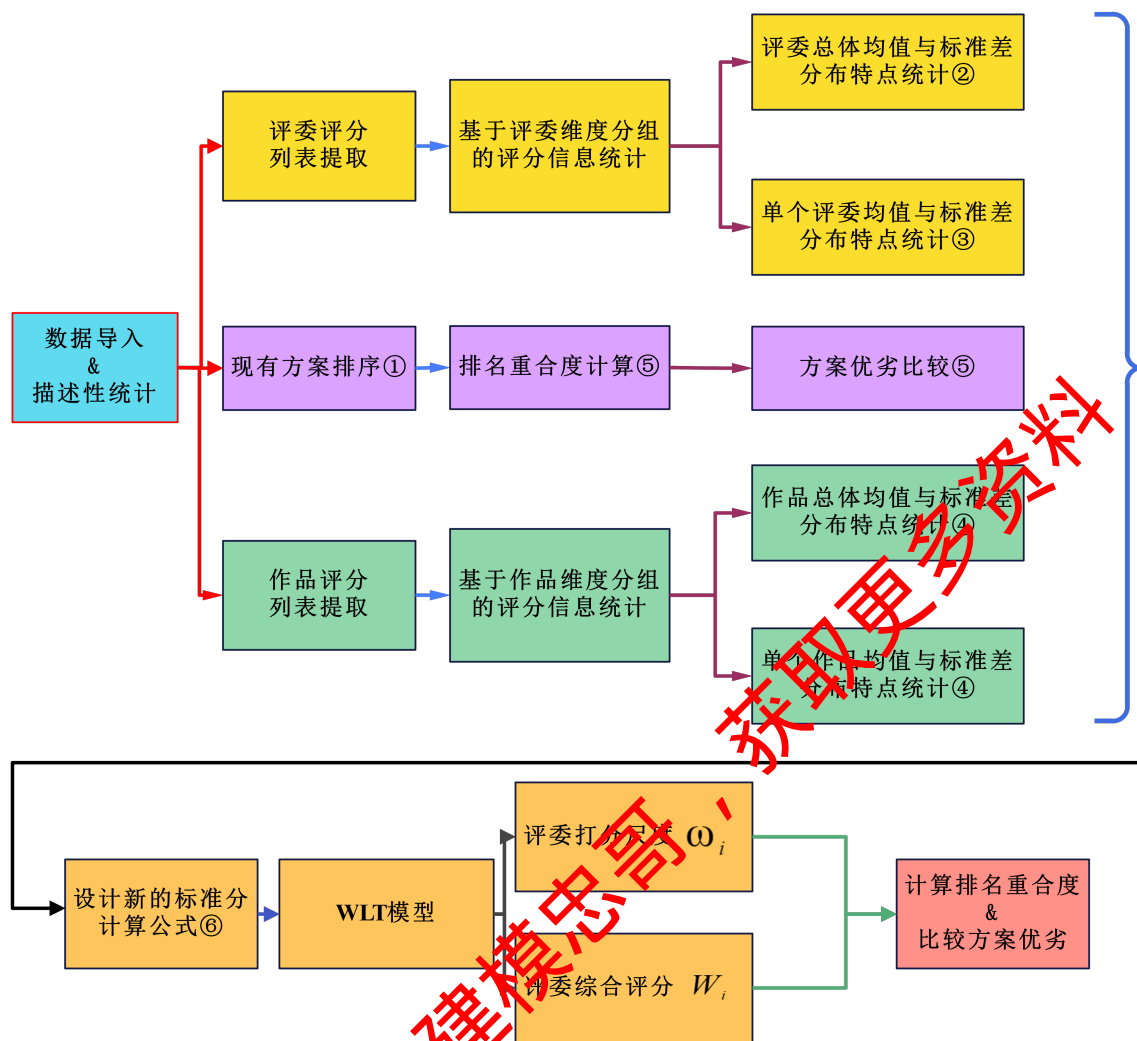


图 3: 第二问流程图

首先通过描述性统计、相关性分析等方法对原始分数进行分析处理，研究其分布特点，在此基础上，选择两类方案进行作品排序，评价方案的优劣，从中吸取经验，改进现有的标准化公式。

5.2 模型的建立与求解

5.2.1 成绩分布特点的分析与可视化

1. 单个评委评分的分布特点与分析

通过数据 1 提供的每位专家的打分情况，可以计算得出每位专家评审作品中的最大值，上四分位数，中位数，下四分位数和最小值，从而得到各位专家评审分数的箱线图如图4所示。其中，横坐标代表两轮参与评审的 97 位专家的序号，与数据 1 中专家编号不同。图4可以反映出每位评委的打分分布情况和总体倾向。

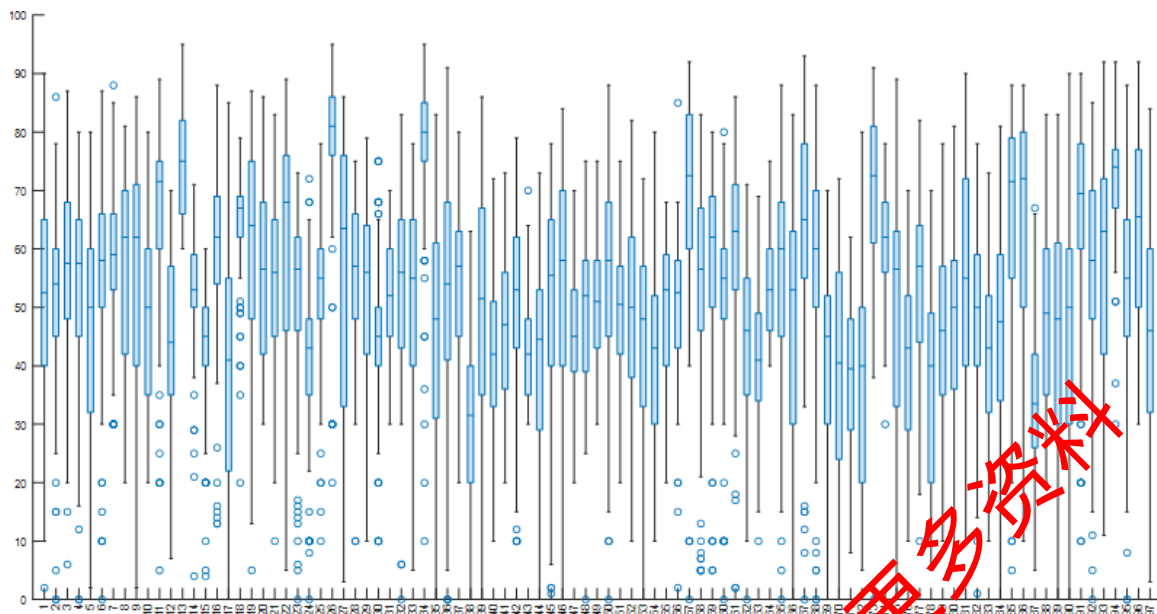
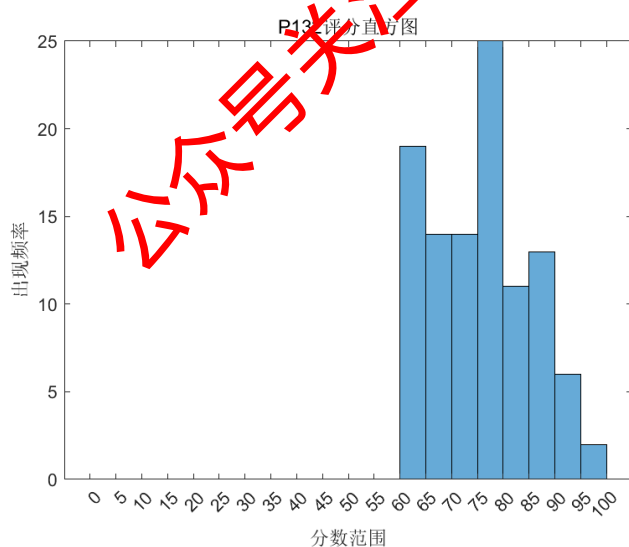
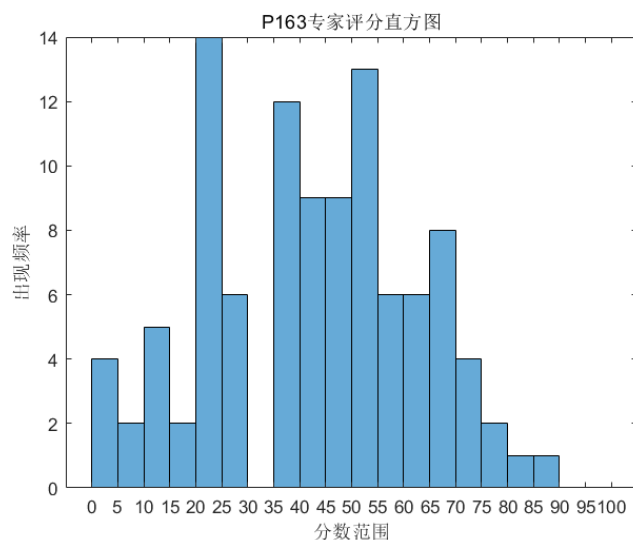


图 4: 各位专家评审分数的箱线图

因此，能够大致认为该评委能够做到公平公正，打分具有很强的区分度。其中，序号为 17(P163)、38(P322)、72(P511)、78(P573) 等评委对应的箱线图明显低于其他评委，且无异常值出现，说明上述 4 名评委给各个作品的评分普遍偏低。其中，序号为 38(P322) 的评委对应的评分分布直方图如图5(a) 所示，可以看出 P322 评委审评的 103 篇作品中，评分在 $[20, 45]$ 区间内的作品数大约为 63 篇，即位于 $[20, 45]$ 区间内的作品占大多数，且该评委评审的作品最高分不超过 65 分，说明 38 号 (P322) 评委评审的作品质量普遍偏低，或者反映出该评委的评审标准相较于其他评委更加严格。图5 (b) 对应的 17 号 (P163) 评委打分的 103 篇作品中同样有 63 篇分布在 $[20, 55]$ 区间内，分数总体较低，但位于 $[65, 70]$ 区间内的作品数为 16 篇，高于 P322 评委的 0 篇，说明作品质量普遍较低的情况下，P163 评委的打分更加客观公正，且区分度更高。



(a) P132 评委评分直方图



(b) P163 评委评分直方图

图 5: 箱线图整体偏低的评分直方图

同时,也存在打分相对较高的评委,例如序号为 13(P132)、26(P235)、34(P282)、57(P449)等的评委,所给分数主要集中在高分段。以图6(a)(b) 对应的 13 号 (P132) 评委和 34 号 (P282) 评委为例。

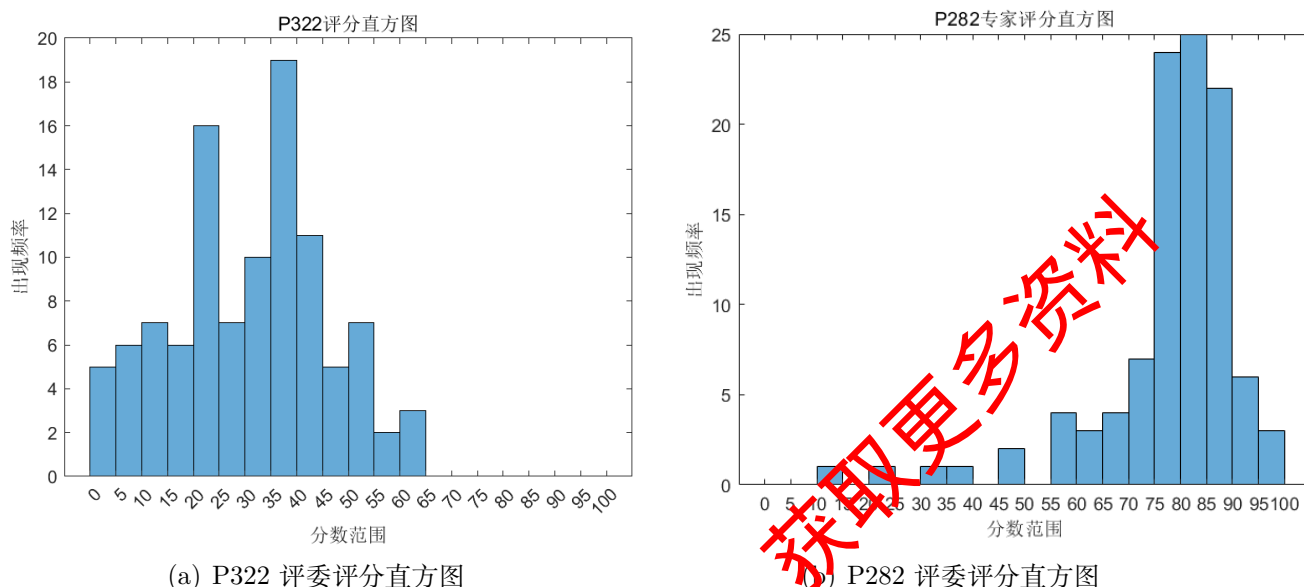


图 6: 箱线图整体偏高的评分直方图

P132 评委评审的 104 篇作品的分数均高于 60 分,且分数主要分布在区间 [60,80] 内,一共有 72 篇,说明该评委评审的作品质量普遍偏高,或者反映出该评委的评审标准相较于其他评委更加宽松。相比之下,34 号 (P282) 评委评分的作品中有 71 篇位于区间 [75,90] 内,作品分数总体高于 P132 评委,但从图4和图6 (b) 中可以看到该评委的评分存在较低的异常值点,说明在作品整体质量较高的情况下,P282 评委的评分标准比 P132 更加严格,且区分度更强。

由于评委分到的作品质量随机性较强,因此认为评委的评审标准和水平等主观因素会评分分布造成较大的影响,后续将基于评委所给分数与其他评委所给分数的偏离程度与区分度,对评委的评分标准和水平等进行分析。

2. 原始分数与标准分数的分布特点与分析

将作品第一阶段的评分作为一个总体,以单个作品的 5 个评分为一组,统计分析每个作品原始分均值、标准分均值、极大值、极小值以及极差 5 个指标的分布情况,得到不同参数组合下的参数分布直方图与拟合曲线如图7所示。

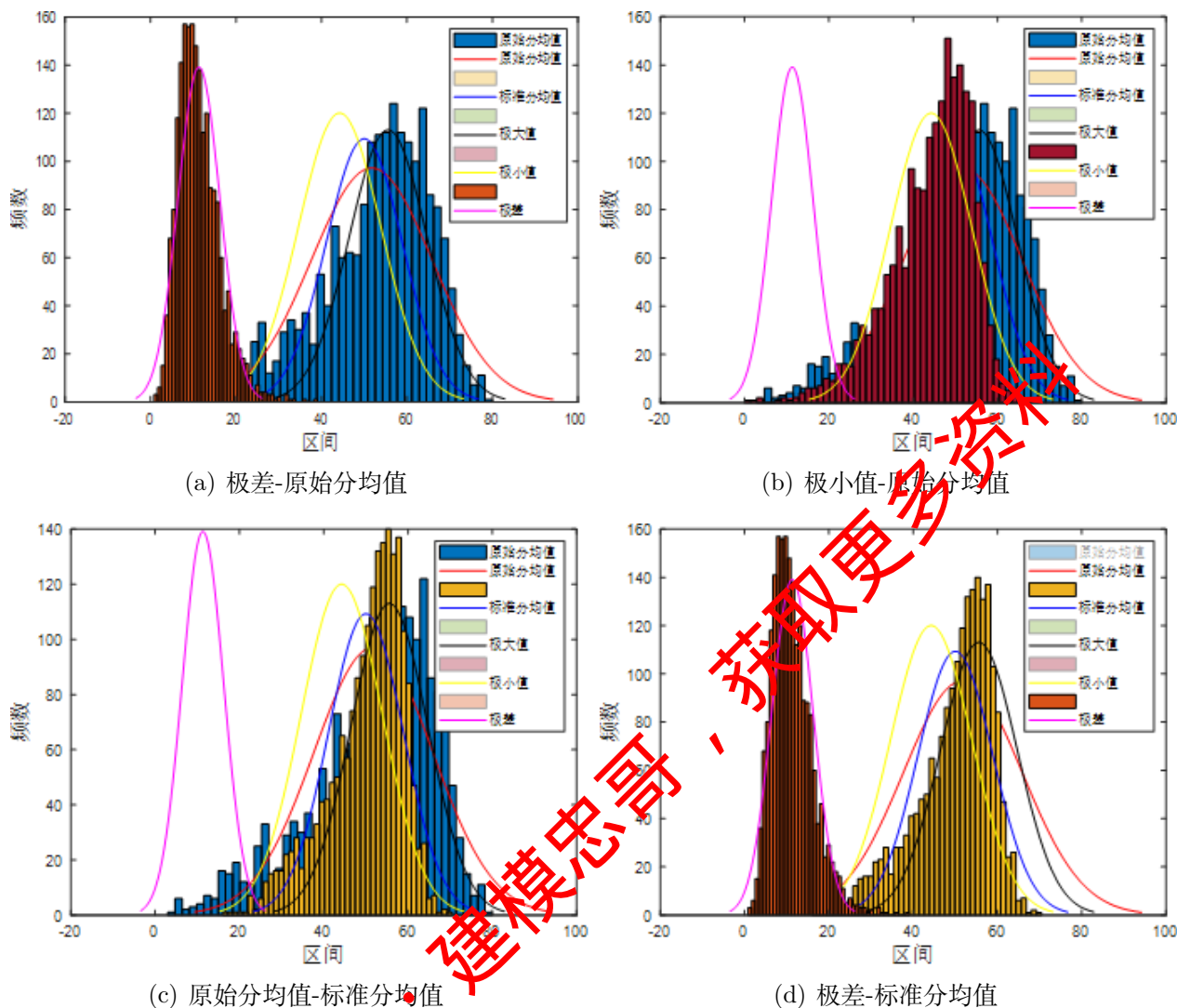


图7: 不同指标组合下的作品总体分布直方图

由图7(a)(b)可知, 除极差外, 其余4个指标分布均呈现左偏态特征, 说明不同评委审阅的作品总体水平存在差异性。根据图7(c)可知, 标准分分布相对原始分分布峰度更高, 且相对更加集中, 即转换为标准分后, 得分均值在高分段的部分作品转移到了中间相对密集的分段, 一定程度上说明了标准分减弱了不同作品评分的差异性, 使得高分段的作品整体分数下降。此外, 标准分分布(蓝线)更接近于极大值(黑线)与极小值(黄线)分布, 反映出标准化并不能克服极大差带来的影响。

5.2.2 不同方案下的作品排名

本问采用2种方案进行对比和分析。方案(1)按照附件1的公式对5位评审专家的评分进行标准化, 然后将每个作品的5个标准分取平均值作为第一轮的评审分数, 表达式为:

$$y_{ik} = 50 + 10 \times \frac{a_{ik} - \mu_i}{\sigma_i} \quad (5.1)$$

$$y_k = \frac{1}{t} \sum_{i=1}^t y_{ik} \quad (5.2)$$

其中, t 代表每篇作品的评审数, k 代表作品的序号, 即第 k 篇作品, i 代表第 i 位评委, y_{ik} 代表第 i 位评委对第 k 篇作品的标准化评分。最后将第二轮中 3 位评审专家的原始分数与第一轮的评审分数再取均值, 即可得到该作品的最终得分。方案 (2) 是去掉同一份作品得分中的最高分、最低分, 然后将剩余评分取平均分作为该作品第一轮的评审分数, 第二轮的处理方式与方案 (1) 相同, 将第二轮中 3 位评审专家的原始分数与第一轮的评审分数再取均值, 即可得到该作品的最终得分。方案 (1) 与方案 (2) 计算得出的数据 1 前 30 篇作品的排名如表3所示。

表 3: 两种方案的前 30 篇作品排名与最终分数

数据 1 排名与分数		方案 (1)		方案 (2)	
排名	分数	排名	分数	排名	分数
1	70.44	1	69.44	1	71.60
2	68.67	2	68.67	2	70.65
3	67.20	4	66.04	5	70.22
4	66.04	5	65.97	4	69.82
5	65.97	13	65.56	3	69.34
6	64.89	6	64.89	9	67.53
7	64.47	8	64.27	11	67.19
8	64.27	9	64.07	6	66.98
9	64.07	12	63.55	8	66.78
10	63.74	3	63.20	10	66.67
11	63.66	15	62.86	26	65.67
12	63.55	18	62.48	12	65.65
13	62.73	16	62.11	15	65.34
14	62.73	17	61.97	7	64.69
15	62.66	19	61.60	17	64.50
16	62.13	21	61.47	20	64.14
17	61.97	47	61.14	19	64.09
18	61.66	22	61.12	22	64.01
19	61.60	25	60.68	32	63.98
20	61.51	26	60.67	14	63.41
21	61.47	27	60.65	21	63.41
22	61.12	20	60.47	23	63.40
23	61.01	7	60.22	29	63.36
32	60.96	30	60.19	24	63.30
24	60.76	33	60.03	13	63.27
25	60.68	29	59.94	16	63.25
26	60.67	34	59.88	41	62.95
27	60.65	35	59.80	18	62.44
28	60.52	38	59.48	31	62.44
29	60.43	39	59.36	55	62.37
30	60.19	14	59.19	34	62.27

5.2.3 方案优劣的比较与分析

由于题目假设在第二轮中获得一等奖的 27 件作品进评委协商复议后，其排名与分数具有较高的可信性，因此方案 (1)(2) 得到的分数与排名与上述一等奖作品的重合度和偏离程度可以作为方案 (1)(2) 优劣的评价标准，同时也可以作为标准分计算方法的改进与修正依据。

将“方案 (1) 标准化分数”与“方案 (2) 去除极值分数”取平均值后，作为第 4 位评审分数，加入到前 352 件作品（附件中只有前 352 件作品进入第二轮评审）的第二轮最终分数计算中，最终得到这 352 件作品的三个最终排名结果，如图8所示，可以看到方案 (2)（蓝色）的波动小于方案 (1)（红色）。

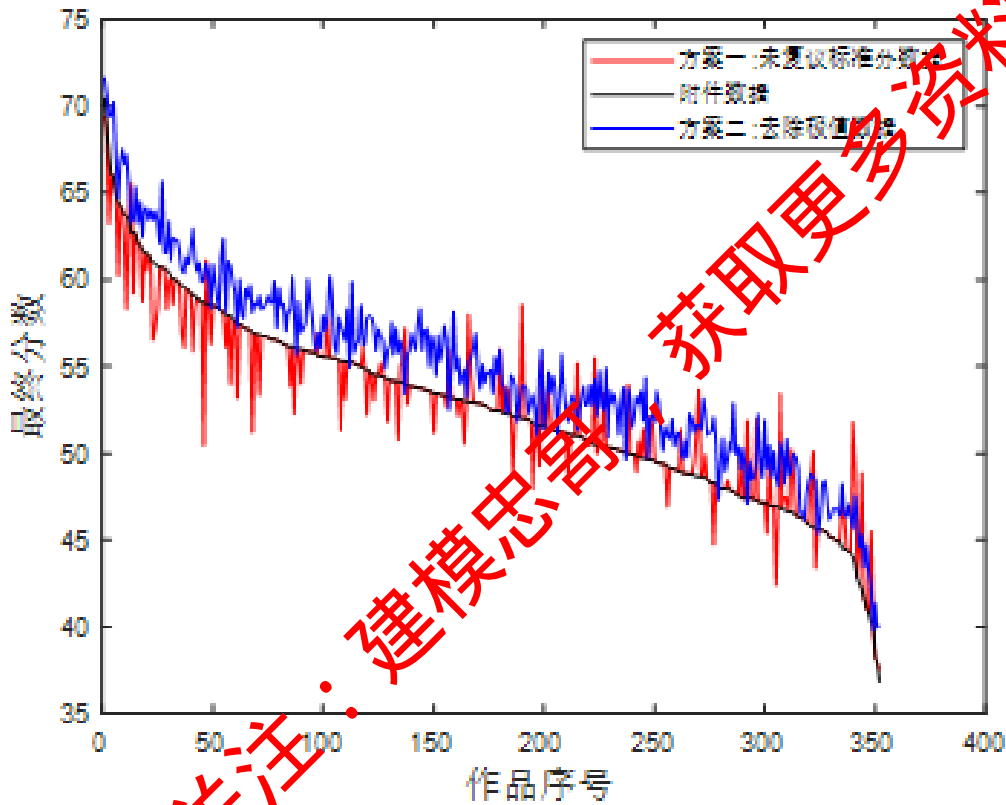


图 8: 方案 (1) 与方案 (1) 对比图

为进一步量化方案 (1) 与方案 (2) 的差异，提高比较结果的说服力，基于 27 个一等奖作品的分数和排名，本节提出 4 个评估指标：排名重合率 R ，平均乱序度 $degree_{disorder}$ ，最大乱序度 $degree_{max}$ ，分数标准差 σ_{plan} [1]。

1. 排名重合率 R

设有两个排名向量 var_1 、 var_2 ，计算两个向量对应位置中元素相同的个数，除以单个向量中元素的总数，百分化后即两个排名结果的重合率：

$$R = \frac{\sum var_1 \oplus var_2}{n_{var}} \times 100\% \quad (5.3)$$

2. 第 k 个作品在两个排名向量中的名次分别为 $Rank_{var_1}$ 、 $Rank_{var_2}$ ，则 $|Rank_{var_1} - Rank_{var_2}|$ 即为该作品的乱序度，将向量中所有作品的乱序度相加即为该排名向量的总乱序度，除以向量中元素个数即为平均乱序度：

$$degree_{disorder} = \frac{\sum |Rank_{var_1} - Rank_{var_2}|}{n_{var}} \quad (5.4)$$

3. 对于 $\forall \text{var}_1, \forall \text{var}_2$, 有:

$$\text{degree}_{\max} = \max(|\text{Rank}_{\text{var}_1} - \text{Rank}_{\text{var}_2}|) \quad (5.5)$$

4. 存在分数向量 $\text{grade}_1, \text{grade}_2$, 则:

$$\sigma_{\text{plan}} = \frac{\sum (\text{grade}_{1i} - \text{grade}_{2i})^2}{n_{\text{grade}}} \quad (5.6)$$

通过计算, 得到方案 (1) 与方案 (2) 的 4 类指标值如表4所示, 方案 (2) 在各项指标上均优于方案 (1), 因此同样可以反映出方案 (2) 通过去除极大极小值得到的排序结果优于方案 (1) 通过标准化分数得到的排序结果。

表 4: 两种方案优劣对比的指标列表

	排名重合度	平均乱序度	最大乱序度	分数标准差
方案 (1)	11.11%	5.1538	30	3.9746
方案 (2)	22.22%	4.2692	15	2.8212

5.2.4 标准化修正模型: 基于评分尺度系数折算的综合评分加权 T 分数法

由5.2.1节评委评分的分布特点可以发现评委评分标准和水平等因素可能会对作品分数分布产生较大影响, 因此为使得新的标准分计算方法更加合理客观, 首先应当考虑评委之间的差异。

假设有 m 个评委, 对 n 篇作品进行评分。本节引入偏差度和区分度两个指标 [2], 从

表 5: 各位评委作品打分情况

作品	评委 1	评委 2	评委 3	...	评委 m	总评
1	x_{11}	x_{12}	x_{13}	...	x_{1m}	\bar{x}_1
...
n	x_{n1}	x_{n2}	x_{n3}	...	x_{nm}	\bar{x}_n

不同维度对评委之间的差异进行量化评估。

1. 偏差度指标

偏差度指标体现了每个评委打分与选手真实水平的吻合程度, 偏差越小说明裁判打出的分数越符合选手的真实成绩。第 j 位评委的偏差度计算公式为:

$$w_j = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_i)^2 \quad (5.7)$$

2. 区分度指标

区分度指标表现了评委打分对不同作品的区别程度, 区分度越大说明该评委的打分比较客观, 不存在“老好人”的现象。第 j 位评委的区分度指标计算公式为:

$$s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \quad (5.8)$$

3. 综合评估模型

通过结合偏美度和区分度进行交叉分析, 给出第 j 位评委的统一评价指标: $\lambda_j = \frac{s_j^2}{w_j}$, λ_j 表示第 j 名评委最终获得的评估分值。评估分值 λ_j 越高, 说明该评委的评分水平越高。结合上述分析, 建立综合评估模型:

$$\begin{cases} w_j = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_i)^2, j = 1, 2, \dots, m \\ s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2, j = 1, 2, \dots, m \\ \lambda_j = \frac{s_j^2}{w_j}, j = 1, 2, \dots, m \end{cases} \quad (5.9)$$

基于附件给出的数据, 结合综合评估模型, 可以计算得出各位专家的综合评分。每名专家的综合评分与评分均值和标准差构成的散点图如图9(a)(b)所示。

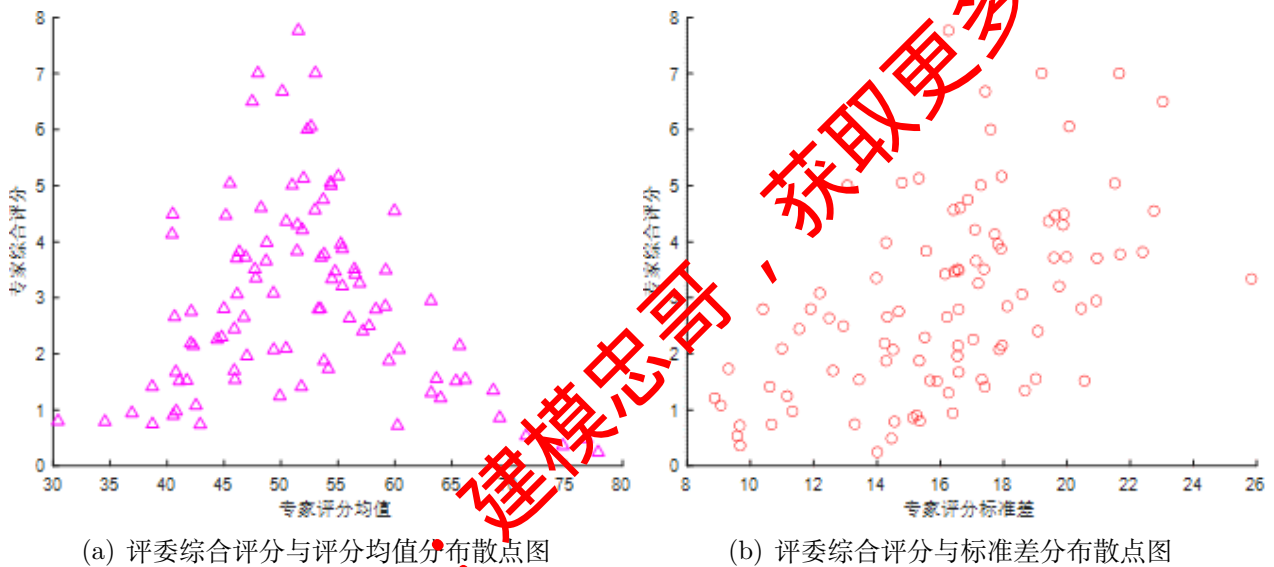


图 9: 评委综合评分与均值以及标准差散点图

由图9 (a) 可知, 评分均值位于 $[45, 60]$ 分数区间的专家综合评分最高, 说明该位置的专家打分水平相对较高, 相比之下, 评分均值偏大或者偏小的专家评分水平相对较弱。同理, 由图9 (b) 可知评分标准差位于较高分数的段的专家评分水平相对较高, 标准差偏低则说明存在“老好人”现象, 或者专家评分水平相对较弱。

因此可以进一步反映出, 专家的评分水平存在较大差异, 因此在设计新的标准公式计算方法时, 核心在于如何消除评委评分的组间分布差异。

由于不同评委的权威性与水平不同, 同一评委对于作品的评阅分数整体存在一个评分偏差度, 定义为评分尺度 w_i 。假设共有 m 个评委, n 篇作品, 第 i 位评委对第 k 篇作品的评分标准分为 y_{ik} , 一篇作品共有 t 位评阅评委进行评阅, 一位评委评阅 r 篇作品。

该标准分计算方法的核心步骤如下:

1. 利用第 k 篇作品的五个评分确定理想分数 S_k' ;
2. 利用第 i 个评委的所有评阅作品的理想分数与实际 T 分数得到与理想分数数量相同的评价尺度 W_{ij}' , 其关系式为 $S_{ij}' = W_{ij}' \times y_{ij} (j = 1, 2, \dots, r)$;
3. 利用这 r 个权重求平均得到该评委的最优评价尺度 w_i' ;

4. 计算每一篇论文的最终分数 $y'_k = \frac{1}{t} \sum_{i=1}^t W'_i y_{ik}$, 最后根据 y'_k 进行排名。

综合上述, 得到多个评委标准分综合计算公式如下:

$$y'_k = \frac{1}{t} \sum_{i=1}^t \left(\frac{1}{r} \sum_{j=1}^r \frac{S'_{ij}}{y_{ij}} \right) \times y_{ik} \quad (5.10)$$

上述给出了如何基于同一个作品的标准分得到第一轮排名分数的计算公式, 而题目要求给出新的标准分计算公式, 考虑上文中的两个结论: 第一, 不同评委评分结果的可靠性(评分水平)不同, 可以用评委综合评分衡量评分结果在所以评分结果中的重要程度; 第二, 去掉极值相比于标准化分数更能缩小第一轮评审分数的波动。因此, 基于以上两条结论, 我们提出新的标准化分数计算公式如下:

$$y'_{ik} = w_i y_{ik} = \frac{n-1}{n} \sum_{j=1}^r \frac{(a_{ij} - \bar{a}_i)^2}{(a_{ij} - \bar{a}_j)^2} y_{ik}, i = 1, 2, \dots, t \quad (5.11)$$

综合以上两阶段模型, 得到初审评分优化模型:

$$\begin{cases} y'_k = \frac{1}{t} \sum_{i=1}^t \left(\frac{1}{r} \sum_{j=1}^r \frac{s'_{ij}}{y_{ij}} \right) \times y_{ik} \\ y'_{ik} = w_i y_{ik} = \frac{n-1}{n} \sum_{j=1}^r \frac{(a_{ij} - \bar{a}_i)^2}{(a_{ij} - \bar{a}_j)^2} y_{ik}, i = 1, 2, \dots, t \end{cases} \quad (5.12)$$

基于上述公式, 设计求解流程如下所示:

- 步骤 1: 构建每一位评委的评审作品集, 每一件作品的得分集;
- 步骤 2: 计算每一位评委的评审分数平均值 \bar{a}_i , 每件作品得分平均值 \bar{a}_j ;
- 步骤 3: 计算作品集与得分集的组内方差;
- 步骤 4: 计算得分集中五位评委各自的综合评分权重 w_i , 构建权重向量:

$$w_k = \{w_{i_1k}, w_{i_2k}, w_{i_3k}, w_{i_4k}, w_{i_5k}\}$$

步骤 5: 令 $w_k \times y_{ik}$, 得到作品的修正标准分 y'_{ik} ;

步骤 6: 计算评委 $i_l (l = 1, 2, 3, 4, 5)$ 的评分尺度 W_{i_l} ;

步骤 7: 计算作品上的一轮评审最终得分 y'_k 。

将数据 1 的一轮评审数据代入上式计算, 新的作品排名如表??所示, 我们发现修正方案在排名准确性效果上相对于原有方案 (2) 效果有所减弱。

表 6: 三类方案前 30 名作品序号与最终分数

数据 1 分数与排名		方案一		方案二		修正方案	
排名	分数	排名	分数	排名	分数	排名	排名
1	70.44	1	69.44	1	71.60	1	69.79
2	68.67	2	68.67	2	70.65	5	67.80
3	67.20	4	66.04	5	70.22	2	67.73
4	66.04	5	65.97	4	69.52	7	66.28
5	65.97	13	65.56	3	69.34	3	65.74
6	64.89	6	64.89	9	67.53	10	65.69
7	64.47	8	64.27	11	67.19	4	65.64
8	64.27	9	64.07	6	66.98	9	64.14

数据 1 分数与排名		方案一		方案二		修正方案	
排名	分数	排名	分数	排名	分数	排名	排名
9	64.07	12	63.55	8	66.78	6	64.10
10	63.74	3	63.20	10	66.67	11	63.88
11	63.66	15	62.66	26	65.67	20	63.87
12	63.55	10	62.48	12	65.65	8	63.29
13	62.73	16	62.11	15	65.34	14	63.10
14	62.73	17	61.97	7	64.69	15	63.09
15	62.66	19	61.60	17	64.50	12	62.95
16	62.11	21	61.47	20	64.14	32	62.71
17	61.97	47	61.14	19	64.09	18	62.67
18	61.66	22	61.12	22	64.01	48	62.08
19	61.60	25	60.68	32	63.98	17	62.03
20	61.51	26	60.67	14	63.41	35	61.82
21	61.47	27	60.65	21	63.41	35	61.73
22	61.12	20	60.47	23	63.40	23	61.68
23	61.01	7	60.22	29	63.36	19	61.63
32	60.96	30	60.19	24	63.30	21	61.54
24	60.76	33	60.03	13	63.27	13	61.54
25	60.68	29	59.94	16	63.25	47	61.45
26	60.67	34	59.88	41	62.95	16	61.21
27	60.65	35	59.80	18	62.44	31	60.96
28	60.52	38	59.48	31	62.44	46	60.95
29	60.43	39	59.36	55	62.37	43	60.73
30	60.19	14	59.19	34	62.27	51	60.71

在表7反映数据集中程度的各指标上来看，修正方案的作品分数标准差小于方案（1）与方案（2），说明修正方案的作品评分离散度相对较小且较集中；最大乱序度与平均乱序度相对较高，因此下一步可以考虑去除极值，对计算方法进行优化改进。

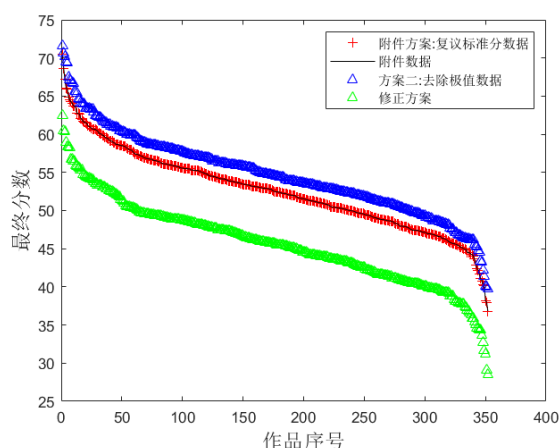
表 7: 三种方案优劣对比的指标列表

	排名重合度	平均乱序度	最大乱序度	分数标准差
方案 (1)	11.11%	5.1538	30	3.9746
方案 (2)	22.22%	4.2692	15	2.8212
修正方案	11.11%	6.2692	30	0.4974

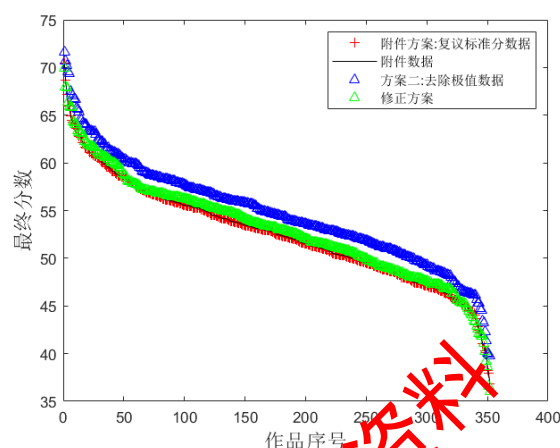
在实际求解拟合的过程中，27 份一等奖作品的修正分数与附件提供的参照分数的差值近似一个常数，如图10所示，绿色散点值即为经过修正公式计算得到的作品分数，曲线形状与真实数据形状相近，但是整体偏下，因此修改模型，在模型中增加一项常数项 η ，通过仿真模拟，以步长为 0.5，取值范围为 [20, 40]，遍历 η 得到当 $\eta = 32.5$ 时拟合效果较好。

综上所述，最终确定第一轮初审评分优化模型为：

$$\begin{cases} y'_k = \frac{1}{t} \sum_{i=1}^t \left(\frac{1}{r} \sum_{j=1}^r \frac{s'_{ij}}{y_{ij}} \right) \times y_{ik}' + \eta \\ y'_{ik} = w_i y_{ik} = \frac{r-1}{r} \sum_{j=1}^r \frac{(a_{ij} - \bar{a}_i)^2}{(a_{ij} - \bar{a}_j)^2} y_{ik}, i = 1, 2, \dots, t \end{cases} \quad (5.13)$$



(a) 修正方案存在常数偏差



(b) 增加常数项的修正方案曲线

图 10: 修正方案存在常数偏差

6. 问题三的模型建立与求解

6.1 问题的描述分析

极差大是创新类竞赛的一个重要特点。对于低分段的作品来说，极差大主要是由于某一位或几位专家对作品及其不认可，此部分作品极差大并不影响整体的获奖情况；而对于能够获得入围第二阶段评审的作品来说，第一阶段的分数也会参与最终成绩的计算，对作品的最终成绩会产生较大影响；对于能够获得三等奖而在三等奖后半部分的作品来说，较大的极差可能会使得边缘类的作品与奖项之交臂，因此在第一阶段也要尽量考虑极差的影响，同时，通过调整极差，会产生一部分名词变化特别大的作品，这部分作品往往也会与创新性较强而不被理解的原因有关，建立极差模型可以帮助筛选出这一部分作品。

题目要求如下：

- 从成绩整体、极差整体两个方面分析两阶段前后的变化情况；
- 分析两阶段评审方案的优劣；
- 建立极差模型，使其能够根据第一阶段评审分数自动调整非高非低分数段作品的“大极差”问题，同时能够在一定程度上发掘这一部分中具有创新性潜力的作品。

6.2 数学模型的建立

6.2.1 两阶段方案整体性分析

6.2.1.1 数据预处理

我们将“数据 2.1.xlsx”与“数据 2.2.xlsx”中的数据表导入 matlab 中，生成了基于专家维度的评阅作品信息，评阅分数整体信息等数据表。其中，在处理数据 2.1 时，我们发现“专家评分分布”数据表中存在异常数据，如表所示，有五位专家只评阅了一份作品，通过查阅“专家评阅作品信息”数据表，发现五位专家评阅的作品均为同一份，因此判定该作品数据属于异常数据，因此在原数据表中删除，最终数据 2.1 中得到作品数为 884 件。

表 8: 异常专家信息列表

专家编号	评阅数量	原始平均分	标准化平均分	原始标准差	标准化标准差
3	119	48.70338983	49.87313559	12.80378505	9.947245781
37	120	63.63865546	50.07638655	14.87731244	10.00719256
...
223	1	NaN	NaN	NaN	NaN
271	1	NaN	NaN	NaN	NaN
275	1	NaN	NaN	NaN	NaN
381	1	NaN	NaN	NaN	NaN
611	1	NaN	NaN	NaN	NaN
...
625	119	75.90677966	50.0759322	14.33004571	10.06862007
704	120	60.22689076	49.97462185	15.98380376	10.91604575

6.2.1.2 两阶段成绩整体的变化分析

分析两阶段成绩整体变化时，我们将同一作品的五个分数按照标准分相加取平均值，得到每个作品第一阶段的评审成绩，并根据该成绩进行排名，得到作品在第一阶段的排名，以该排名作为横坐标轴变量，附件数据提供的第二阶段排名为纵坐标轴数据，得到两个阶段排名的线性相关图如图12所示。通过观察，我们发现前 27% 的作品两个阶段的排名不存在明显的线性相关关系，甚至有部分作品存在较大的名次变化，说明对于入围第二轮评审的前 27% 参赛队伍而言，第一阶段的评选与最终成绩之间没有必然联系。

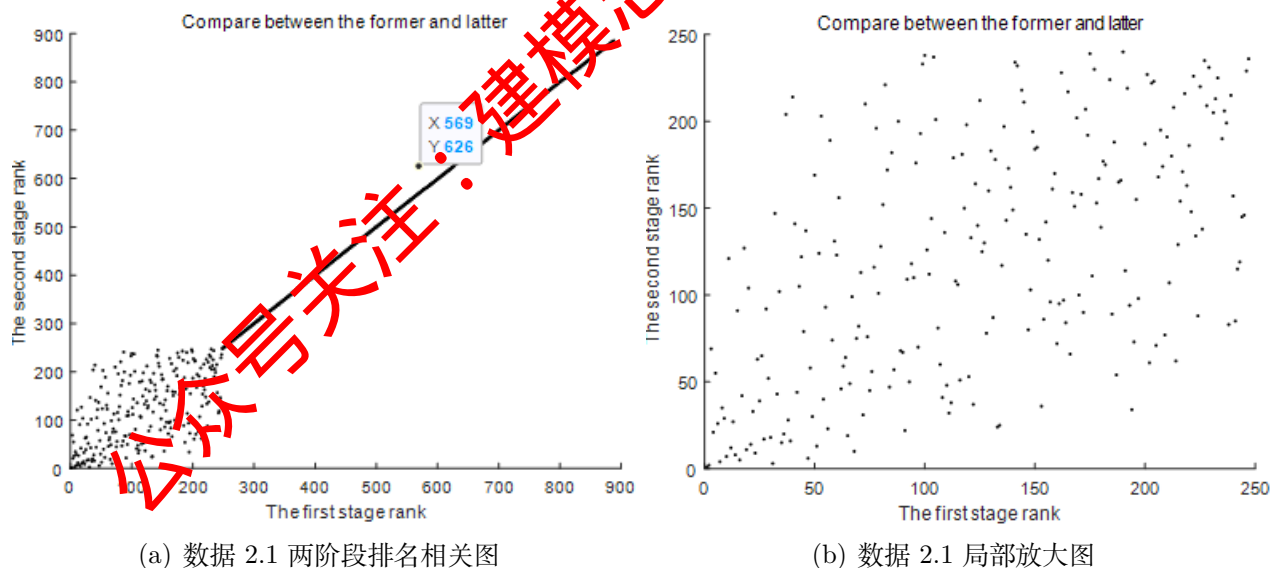


图 11: 数据 2.1 中两阶段作品成绩排名相关性分析

从统计指标上分析，排名靠前的获奖队伍的序相关性指标如表??所示，高水平作品在两阶段的排名较稳定，受到评审因素的影响较小，而入围作品总体呈现较大的名次变化。分析原因，由题设第二阶段参与评审的的评委打分更加稳定，值得信任。从中可以看出，第一阶段由于作品数量过多，评委参评数量有限，导致优秀论文的评选出现了巨大的偏差。因此，在第二阶段评委再次评分与修正时，前后结果会有明显的差别。

6.2.1.3 两阶段极差整体的变化分析

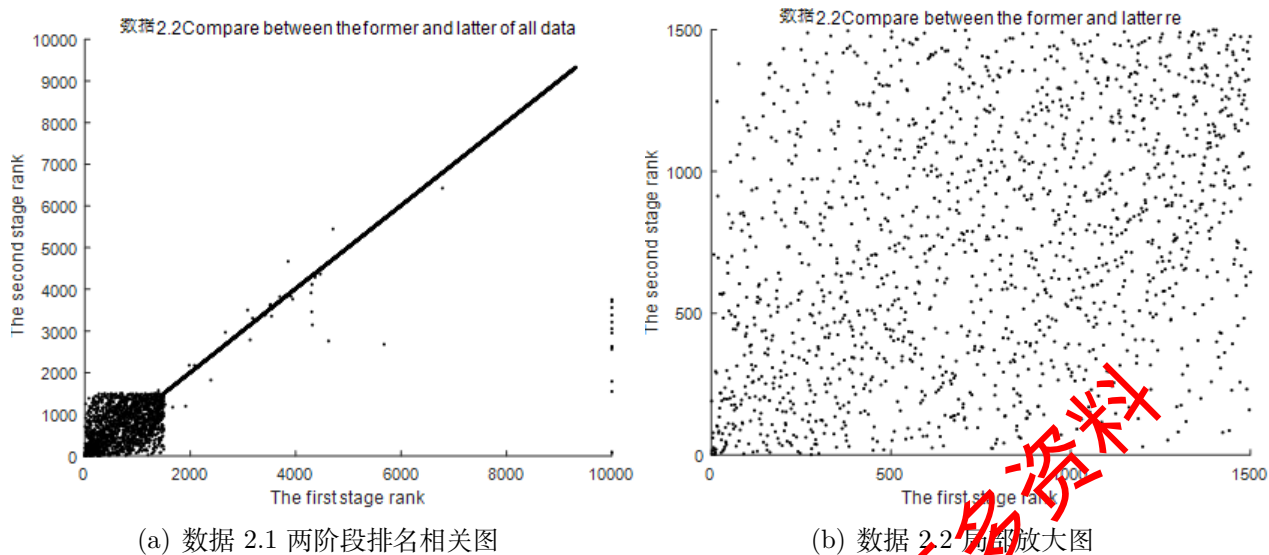


图 12: 数据 2.2 中两阶段作品成绩排名相关性分析

极差总体一定程度上反映了专家总体的评分一致性，极差总体较大说明专家总体对于单个作品的评分差异性较大，会降低作品评审的可信性。目前，根据题目提供的现有方案，第二阶段评审具有更加权威的特点，评审分数的差异性也相对较小，这一结论可以通过图 13(a)(b) 散点图可以直观地反映，第二阶段的极差分布峰值、均值以及离散程度均小于第一阶段，大极差的数量同样也少于第一阶段。图 13(d) 反映了第一阶段异常的“大极差”作品数量较多。

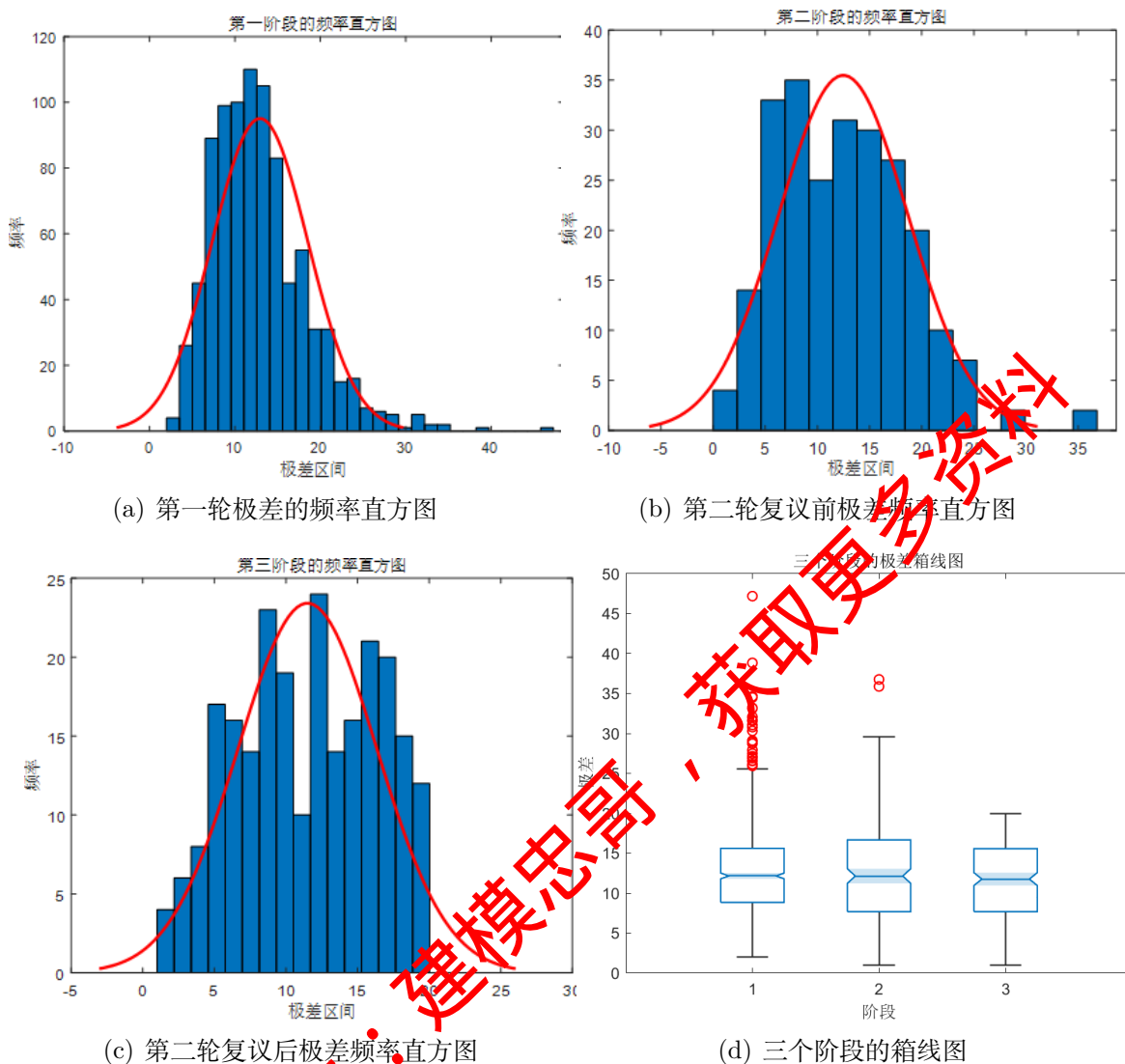


图 13: 不同阶段极差总体分布直方图与箱线图

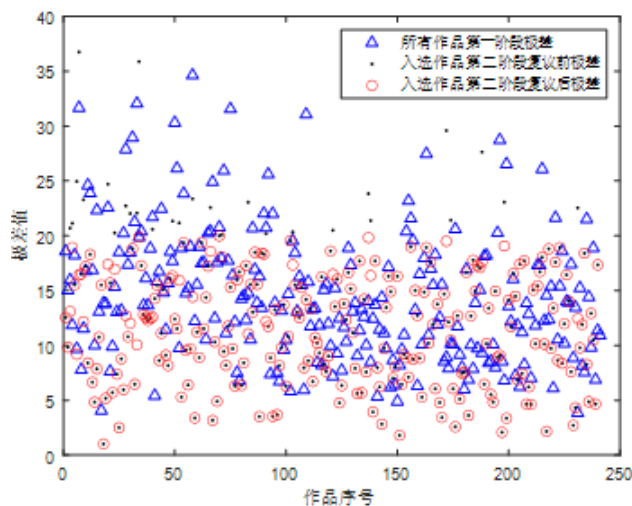
6.2.1.4 两阶段评审方案与不分阶段评审方案的优劣分析

根据图13, 从三个阶段极差的频率直方图和箱线图能够直观地看出, 第一阶段所有作品的极差分布总体而言较为均匀, 在 10 前后较为集中, 但也存在大量的离散点, 说明争议性作品较多。而第二轮评审复议前, 能够发现作品的极差相对而言比较大的占多数, 说明极差大与作品的创新性有一定关联。经过专家调整后, 在第三轮复议过程中, 所有作品的极差都小于 20。通过图14的散点图也可以直观地看到第二阶段的大极差数量少于第一阶段的大极差数量。

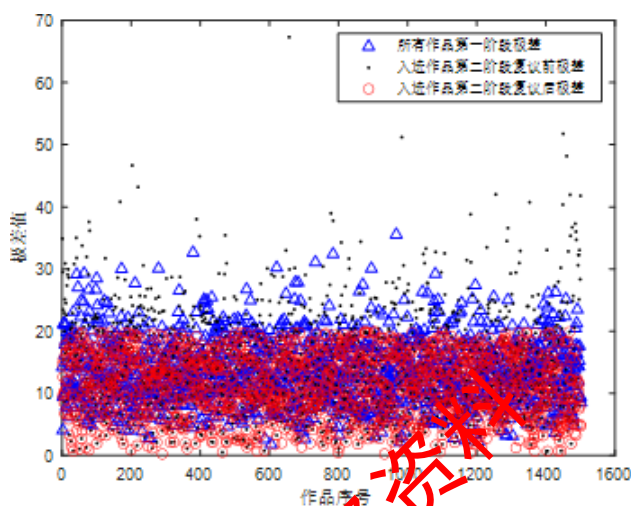
通过表中数据也可以直接看到 3 个阶段的极差的均值、方差、变异系数都在不断减小, 说明专家打分的一致性变强, 在一定程度上能够反映作品的真实水平。

6.2.2 极差模型的建立与求解

在建立极差模型前, 我们通过图表分析, 描述性统计以及相关性分析对各阶段极差数据进行了研究, 在这里做出说明, 极差数据有三个样本总体, 即第一阶段极差, 第二阶段复议前极差以及第二阶段复议后极差。



(a) 数据 2.1 不同阶段极差散点图



(b) 数据 2.2 不同阶段极差散点图

图 14: 不同阶段极差散点图

要建立极差模型，并实现对第一阶段非高非低作品“极差”的自动化调整，首先需要明确极差处理的流程与判定条件，这里给出极差调整程序流程图如15所示：

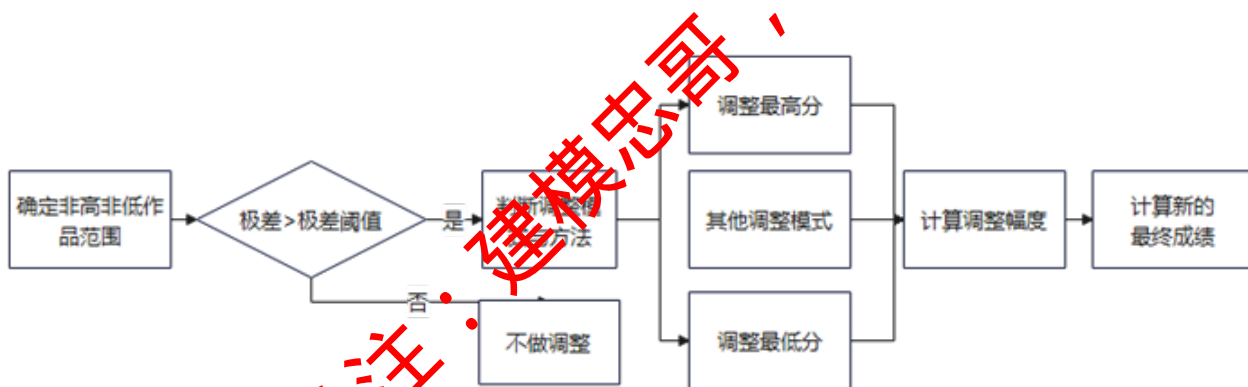


图 15: “极差”模型流程图

要建立极差模型，我们需要解决四个方面的问题：

- 确定非高非低分数段的定义；
- 确定极差调整阈值；
- 极差调整的模式识别；
- 极差调整幅度计算；

6.2.2.1 定义非高非低分数作品范围

对于第一阶段非高非低分数段的上下界定义为：

“非高”分数线 θ_1 ：最低分数低于于入选第二轮评审的最低平均分数线，

“非低”分数线 θ_0 ：最高分数高于获奖最低平均分数。

极差的调整可能导致获奖作品范围以及入选第二阶段的作品范围发生变化，而 θ_1 确保了在 θ_1 以上的作品一定能够进入第二轮评审，而 θ_0 以下的作品也一定不能获奖，我们需要讨论的是介于这两条分数线之间的作品集合。

6.2.2.2 确定极差阈值

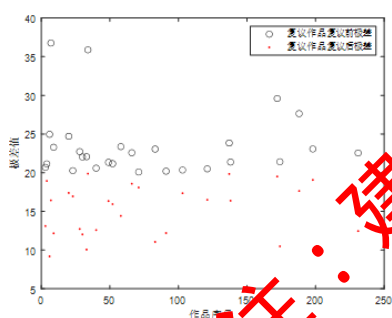
专家打分难免存在一定的分歧，当极差大于某一个阈值时，我们才需要进行极差调整。因此，根据比赛数据形成经验性规律，从而确定极差阈值，可以帮助我们识别哪些作品需要调整极差。以“数据 2.1.xlsx”的数据为样本，如表所示，表中所列的 28 个作品是经过专家协商，进行了极差调整的，而第二轮其他的 222 的作品并未进行极差调整，因此，确定极差调整的阈值，需要研究这 28 个样本数据。

因为极差阈值的几何意义为二维平面上的一条直线，将大极差样本与其他样本分割开来，因此属于分类识别中的线性分类问题。这里我们采用支持向量机进行分类识别。支持向量机 (Support Vector Machine, 简称 SVM)，是一种有监督学习方法，可被广泛应用于统计分类以及线性回归。使用支持向量机方法的核心在于确定超平面方程，这里分界线方程比较简单，为

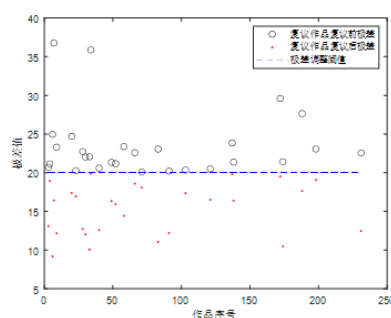
$$y = b \pm \delta$$

其中， $\pm\delta$ 为分界线的上下界。通过求解我们得到分界直线为 $y = 20 \pm \delta = 0$ ，如图 16(b) 所示，两类样本的极值分布在分界线上，因此上下界为 0。

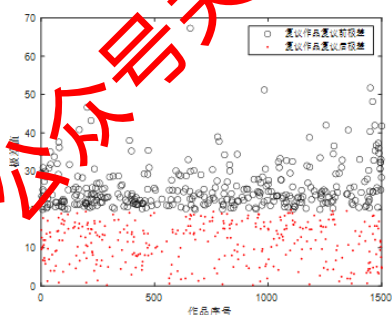
利用“数据 2.2.xlsx”的复议作品数据进行对比，数据 2.2 中共有 342 件复议作品，经过求解得到极差阈值同样为 20，如图 16(d) 因此我们将此作为极差调整的经验阈值，认为极差超过 20 的作品需要启动复议程序。



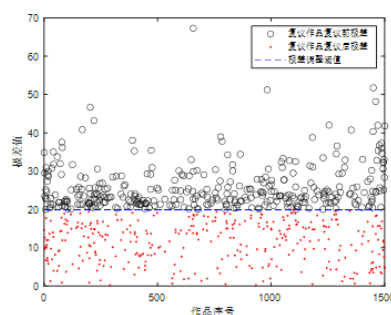
(a) 数据 2.1 待分类数据



(b) 数据 2.1 确定 SVM 分类界限



(c) 数据 2.2 待分类数据



(d) 数据 2.2 确定 SVM 分类界限

图 16: 确定极差阈值

6.2.2.3 基于集成优化算法的极差调整模式识别

确定了极差阈值后即可从作品中筛选出待调整的“大极差作品”。下一步要确定调整作品的哪一位或几位评审的分数，不同于第二轮的评委经过协商一致商定分数来调整策略，

第一轮要求在没有人干预的情况下完成极差的调整，这就要求我们根据极差调整的历史数据总结学习规律，提取分数调整对象的选择模式。

在具体的模型建立与求解环节，我们利用现有数据构建属性集，基于现有数据，筛选出经过复议的作品集合（其中数据 2.1 共有 28 个复议作品，数据 2.2 共有），通过数据矩阵分解与分类存储实现不同类别调整策略的三维空间展示，从而识别出作品的调整策略。在此基础上，得到对样本的分类标签，与我们构建的属性集一同作为训练样本，分别采用三次 KNN 算法、决策树、随机森林以及 Adaboost 集成优化的提升树等模型进行模式识别训练，选取准确率最高的模型封装保存，实现对后续比赛极差调整方法的自动识别。

表 9: 复议作品数据集实例

作品序号	标准分一	复议分一	标准分二	复议分二	标准分三	复议分三
...
3	78.11	78.11	76.69	76.69	57.41	65
...
20	71.29	68.29	50.9	50.9	52.51	66
...
23	71.17	65.17	50.9	50.9	67.85	67.85
...

以数据 2.1 为例，首先进行复议作品数据集的构建。如表??所示，通过观察数据，我们初步得出以下结论：第一，最高分、中间分与最低分均有可能作为调整对象；第二，最高分与最低分调整后可能不再是最高分或最低分；第三，可能同时调整多个分数。

将数据存储到 matlab 矩阵变量中，利用作品标号标签与序号标签共同对作品数据进行定位，构建不同维度的子数据集，以此进行样本分类可视化展示，如图 2 7 所示，展示了 28 份复议作品的分数变化二维可视化图。我们可以直观地观察到，一部分标记为红色的复议后分数与复议前分数重合，一部分作品序号对应 5 个标记分数，说明有多个分数发生变化。

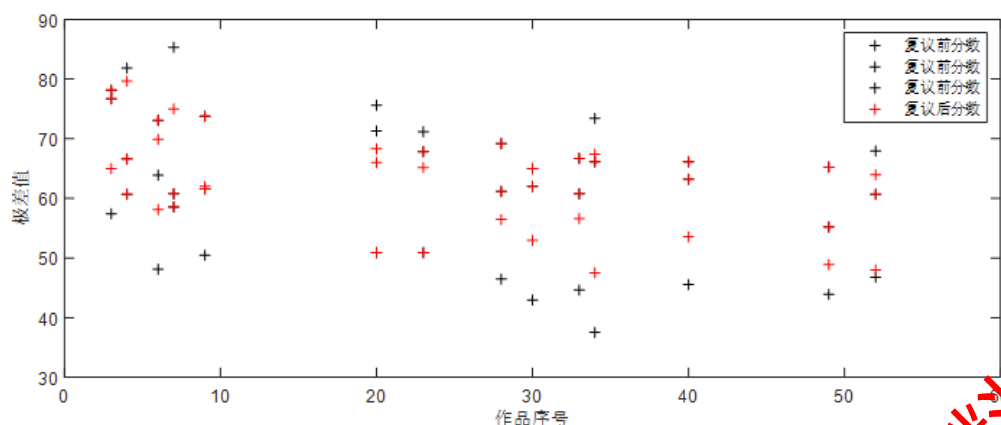
通过对散点图和具体数据表格的统计观察作品的调整模式，调整极差的模式共分为六种，如表所示，按照调整模式对复议作品进行分类，并以“0”，“1”，“2”，“3”，“4”，“5”标识，不同类别作品数据如表10所示。

通过对散点图和具体数据表格的统计观察作品的调整模式，调整极差的模式共分为六种，如表所示，按照调整模式对复议作品进行分类，并以“0”，“1”，“2”，“3”，“4”，“5”标识，不同类别作品数据如表11-15所示。

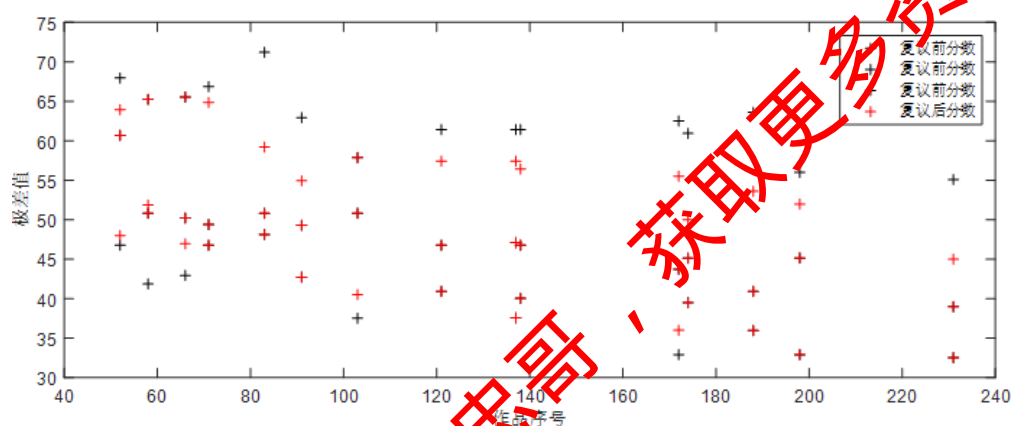
表 10: 极差调整模式列表

极差调整模式	类别
不调整	0
调整极大值	1
调整极小值	2
调整极大值和极小值	3
调整极小值和中间值	4
调整极大值和中间值	5

通过五类极差调整模式的作品空间分布，我们可以发现通过 SVM、K 均值聚类等分类器难以将作品模式识别出来并进行有效分类，且将三位评审的打分作为输入属性区分度不高，因此考虑采用相关性分析，先判断属性之间的相关关系以及属性与类别之间的关系。



(a) 前半部分复议作品复议前后分数序列图



(b) 后半部分复议作品复议前后分数序列图

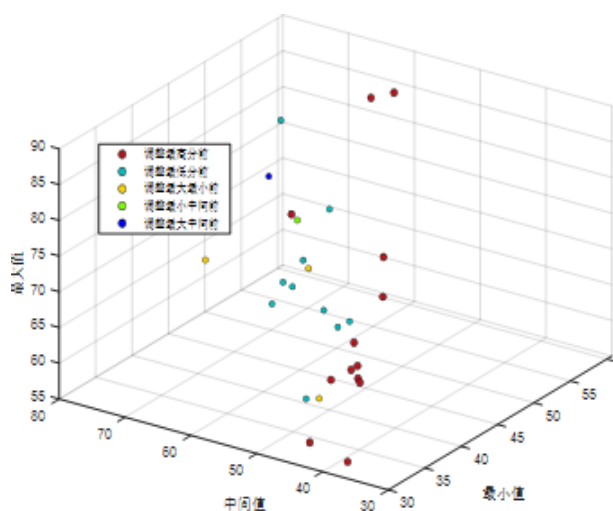
图 17 复议作品分数变化散点图

表 11: 1 类作品数据

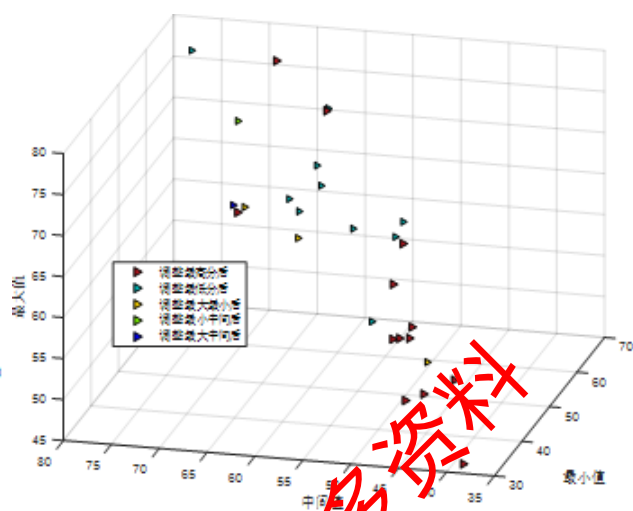
作品序号	调整前 最低分	调整后 最低分	调整前 中间分	调整后 中间分	调整前 最高分	调整后 最高分
3	57.41	65	76.69	76.69	78.11	78.11
9	50.46	62	61.58	61.58	73.75	73.75
28	46.45	56.45	61.15	61.15	69.18	69.18
30	42.95	52.95	62.01	62.01	64.98	64.98
33	44.62	56.62	60.76	60.76	66.69	66.69
40	45.54	53.54	63.21	63.21	66.13	66.13
49	43.9	48.9	55.2	55.2	65.24	65.24
58	41.87	51.87	50.82	50.82	65.24	65.24
66	42.95	46.95	50.22	50.22	65.53	65.53
103	37.52	40.52	50.82	50.82	57.87	57.87

表 12: 2 类作品数据

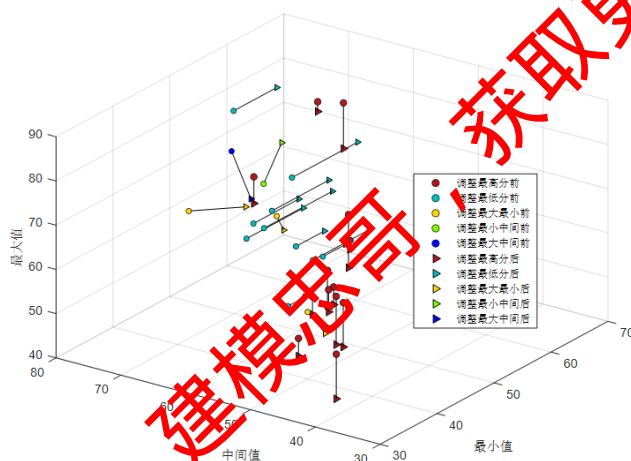
作品序号	调整前 最低分	调整后 最低分	调整前 中间分	调整后 中间分	调整前 最高分	调整后 最高分
6	48.11	58.11	63.88	69.88	73.07	73.07



(a) 极差调整前的复议作品空间分布



(b) 极差调整后的复议作品空间分布



(c) 复议作品极差调整移动路径与空间分布对比

图 18: 复议作品调整前后的三维空间分布

表 13: 3 类作品数据

作品序号	调整前 最低分	调整后 最低分	调整前 中间分	调整后 中间分	调整前 最高分	调整后 最高分
34	37.52	47.52	66.13	66.13	73.4	67.4
52	46.77	48	60.69	60.69	67.94	63.94
172	32.92	36	43.74	43.74	62.51	55.51

表 14: 4 类作品数据

作品序号	调整前 最低分	调整后 最低分	调整前 中间分	调整后 中间分	调整前 最高分	调整后 最高分
20	50.9	50.9	71.29	68.29	75.61	66

表 15: 5 类作品数据

作品序号	调整前 最低分	调整后 最低分	调整前 中间分	调整后 中间分	调整前 最高分	调整后 最高分
4	60.69	60.69	66.6	66.6	81.83	79.63
7	58.57	58.57	60.76	60.76	85.33	75
23	50.9	50.9	67.85	67.85	71.17	65.17
71	46.77	46.77	49.39	49.39	66.86	64.86
83	48.14	48.14	50.81	50.81	71.2	59.2
91	42.73	42.73	49.3	49.3	62.93	54.93
121	40.92	40.92	46.77	46.77	61.42	57.42
137	37.58	37.58	47.1	47.1	61.42	57.42
138	40.04	40.04	46.77	46.77	61.42	56.42
174	39.51	39.51	45.13	45.13	60.93	50
188	35.96	35.96	40.92	40.92	63.6	53.6
198	32.92	32.92	45.16	45.16	55.92	51.99
231	32.54	32.54	38.99	38.99	55.09	45

考虑到现有六类分数难以支撑后面的模型训练，因此我们构建了多个属性指标，共同组成作品的属性集，主要包括以下几类指标

1. 上/下偏距离 D_{up} / D_{dwn}

$$D_{up}(D_{dwn}) = s_{\max} - s_{\text{mid}} (s_{\text{mid}} - s_{\min})$$

2. 偏差距离差 Dev_d

$$Dev_d = s_{\max} - 2s_{\text{mid}} + s_{\min}$$

3. 偏差度 Dev_{degree}

$$Dev_{degree} = \frac{s_{\max} - s_{\text{mid}}}{s_{\text{mid}} - s_{\min}}$$

基于计算得到的协方差矩阵，可视化得到相关性热力图如图19所示，可以看出，三位评审的分数与其他修正分数之间呈强正相关，且与最终成绩相关性高，而基于六类分数计算得到的二级指标如上偏距离，下偏距离与模式的相关性较高。

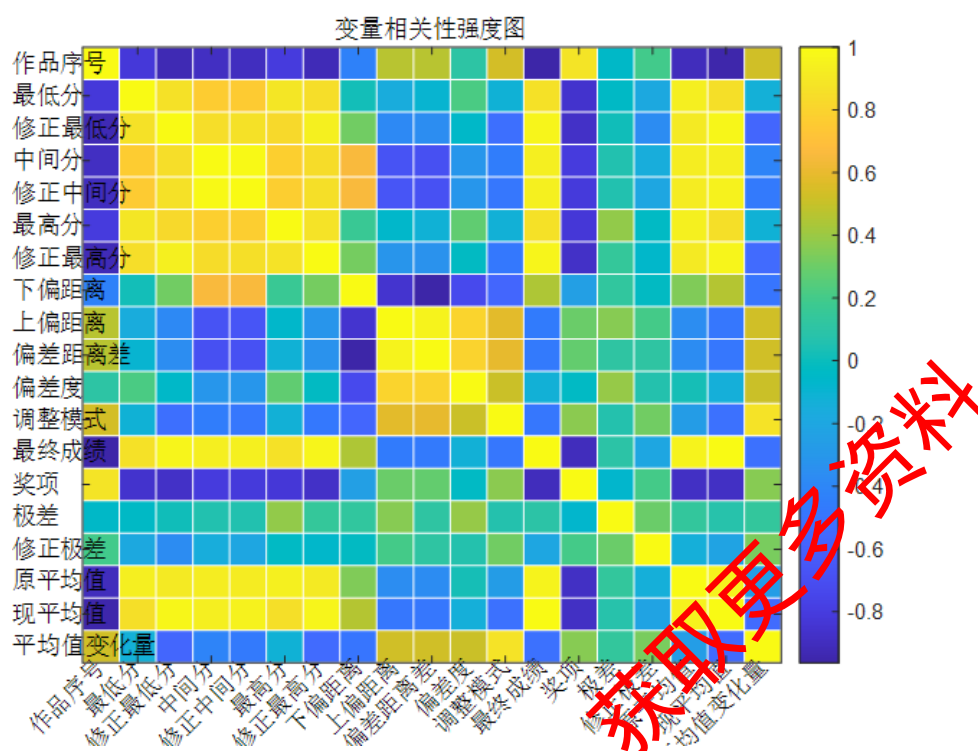


图 19: 作品属性集相关性热力图

通过算法混淆矩阵对比，基于集成的提升树模型预测准确率相对其他算法更高更稳定，因此选用作为预测模型。下面展示其训练结果，决策树、随机森林以及 KNN 算法运行结果见附录 B-D。

表 18: 超参数优化结果

	最佳点超参数	最小分类误差点
集成方法	Adaboost	Bagging
学习器数量	10	500
最大分裂数	1	2
学习率/要采样的变量数	0.059864	10

通过初步分类结果，我们发现增加属性变量对于提升模型分类准确率的提升并不高，上限基本限制在 70%-80%。通过统计分析误判的类别，我们发现误判类别主要集中在类型 3、类型 4 以及类型 2，考虑到极差调整的目的为减小极差，其最终将会以均值的变化影响作品的最终排名我们将调整模式合并简化，将模式 4 合并到模式 5 中，模式 2 合并到模式 1 中，模式 3 根据上偏距离与下偏距离之间的大小关系决定样本的模式，若上偏距离大于下偏距离，则属于模式 5，否则属于模式 1。进而得到新的模式分类训练集，具体表格见附录 E，具体空间分布如图 20(a)(b) 所示，结果见 20(c)-(e)。

将该模型应用到数据 2.2 的计算中，计算结果如，训练结果分类准确率为 88.0%，高于数据 2.1 的 82.1%。

6.2.2.4 极差幅度调整

由于实际过程中极差的调整涉及参评专家给出复议分数作为新的标准分，在实际调整过程中经过专家协商讨论，需要根据论文实际创新性进行确定调整分数大小。在利用上一小节的分类算法确定调整模式后，将该分数替换成该分数来自的评阅评委评分分布的平均

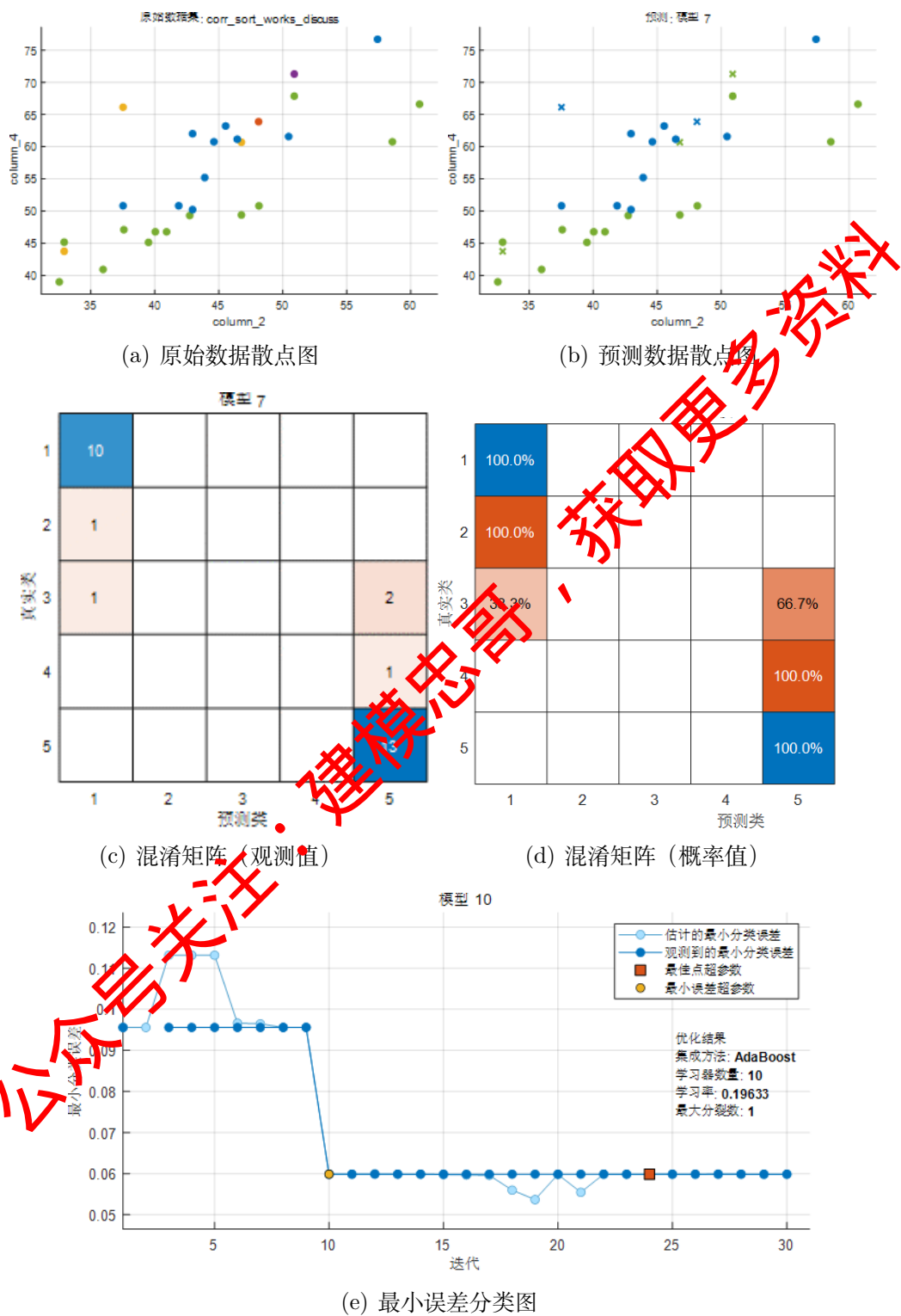


图 20: Adaboost 集成优化训练结果

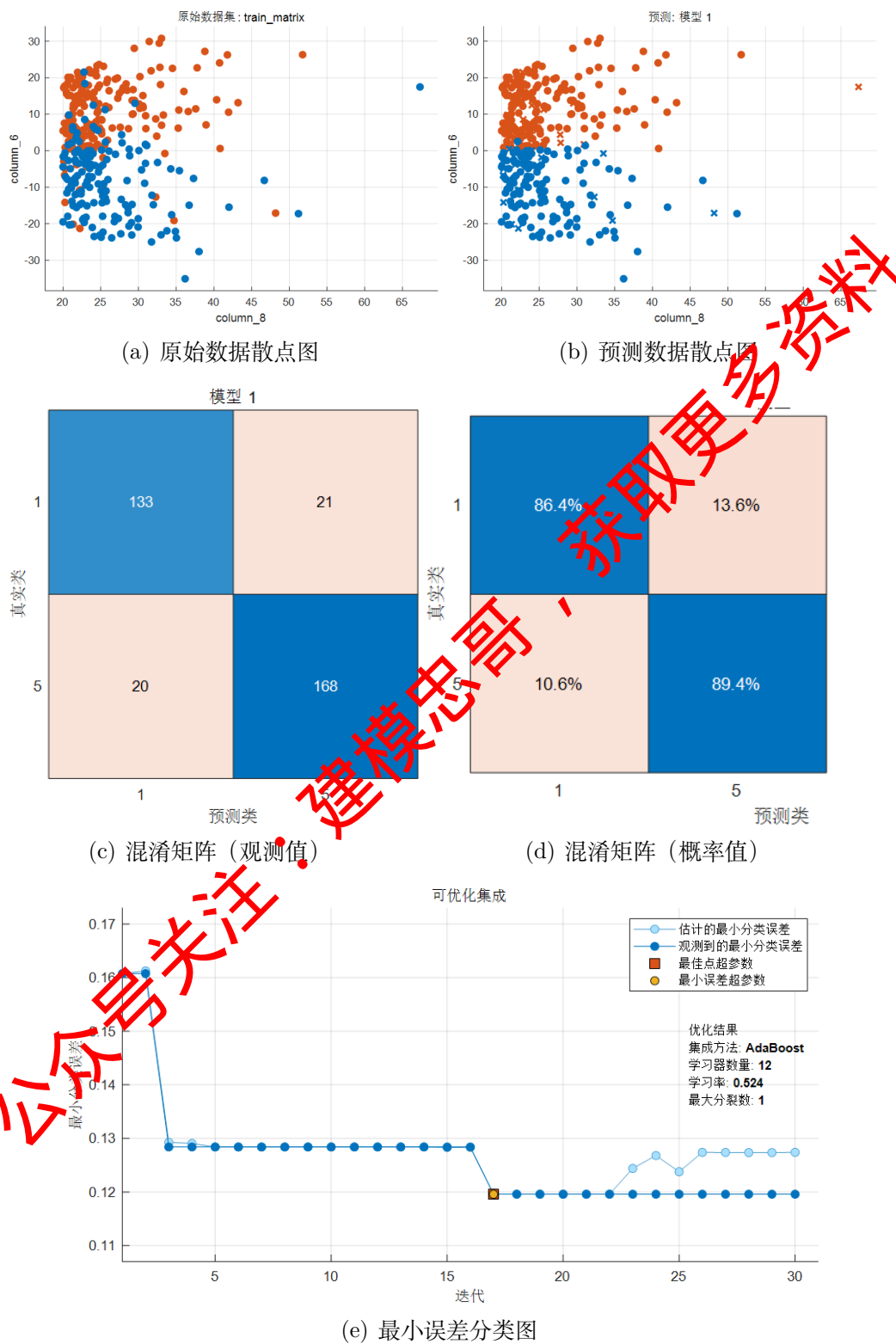


图 21: 两类模式下 Adaboost 集成优化训练结果

分，同时，为了确保将极差调整至阈值以下，取

$$\max \{|a_{ik} - \mu_k|, |a_{k_max} - a_{k_min} - 20|\}$$

作为最终的调整分数。

6.3 模型求解和分析

基于上述模型构建过程，我们以数据 2.1 的第一阶段评审数据为例，进行极差自动调整演示与创新性作品发掘探索。

首先根据数据 2.1 的第一阶段评审分数进行排序，划分非高非低分数段的作品集。根据上文定义的上下界，以数据 2.1 的 884 份作品作为总体，计算得到上界为排名第 131 号作品，下界为排名第 289 号作品。在该范围内，筛选超过极值阈值 20 的“大极值”作品集。经筛选，共有 18 份作品极值超过 20。在此基础上，构建非高非低作品集的分数列表，包含五个标准分，按照最高分至最低分排序、极值、平均分、第一阶段极差调整前分数排名等属性，作为下一步数据分析的基础，如表17所示。

表 17: 第一阶段非高非低分数段“大极差”作品

作品编号	109	155	32	221	196	51	85	235	40
最大值	78.11	66.62	65.21	71.84	69.36	71.17	64.15	66.01	66.88
最小值	47.05	43.44	43.96	50.20	40.63	40.86	43.44	44.49	45.18
极差	31.06	23.18	21.25	21.64	28.73	30.31	20.71	21.52	21.70
均值	58.25	58.06	57.95	57.95	55.55	56.82	56.31	56.04	55.69
作品编号	199	248	250	260	266	269	273	283	289
最大值	70.71	74.28	65.18	62.09	69.14	61.51	64.86	66.87	69.24
最小值	44.17	46.36	41.63	42.30	45.81	39.57	41.06	43.31	46.29
极差	26.54	27.92	23.55	20.79	23.33	21.94	23.80	23.56	22.95
均值	55.06	54.88	54.64	54.51	54.37	54.30	54.25	54.12	54.05

构建模式识别属性输入矩阵，计算并绘制相关性热力图。利用训练好的分类器进行模式识别。确定极值调整幅度，更新作品评阅分数与排名。如表18所示，据统计，在非高非低范围内的作品经过调整极差后，仅有少部分作品由未进入第二阶段评审跃升为第二阶段评审范围，并无作品由获奖范围跌至未获奖范围，说明该极差调整程序能够在控制获奖作品大幅度变动的条件下缩小入选作品的极差，从而减少第一阶段成绩对最终成绩的影响。

此外，为了发掘大极差与创新性作品的关系，我们统计了调整极差后新进入第二轮评审的作品列表，如表19所示，该部分作品名次变化较大，主要原因在于其极差过大，且得分模式基本均为存在一个较低的评分，使得其排名偏低，经过极差调整后，其能够进入第二阶段评审，具有创新型作品的潜力，但具体是否具有较高创新型，还需要根据作品实际结合其他信息进行判断。需要注意的是，我们发现编号 199 和 260 的作品仍然存在极差超过 20 的情况，分析原因，通过观察空间分布图，发现这两份作品均属于类别 2，需要同时调整最小值与接近最小值的中间值，而我们在将六类调整模式融合为两类调整模式后，未能将需要“同时调整两个以上数据”的作品正确分类，因此只调整了最小值，调整过后原来的中间值变为了最小值，其与最大值的差值仍然大于 20，下一步针对模型的该缺陷，可以通过设置多轮极差调整，实现分布调整，来克服“不能识别调整两个以上数据”的缺陷。

表 18: 部分作品调整极差前后排名变化

调整前排名	131	132	133	134	135	136	137	138	139	140	141
调整后排名	139	140	141	142	143	75	144	145	146	147	148
调整前排名	142	...	281	282	283	284	285	286	287	288	289
调整后排名	64	...	283	284	193	285	286	287	288	289	230

表 19: 调整极差后进入第二阶段的作品

作品编号	原排名	现排名	名次变化	原极差	现极差
199	242	170	72	26.54	23
248	249	186	63	27.92	8.89
250	250	161	89	23.55	13.85
260	260	175	85	30.79	22.21
266	266	198	68	23.33	11.77
269	269	169	100	21.94	9.31
273	273	176	97	23.8	15.23

7. 问题四的模型建立与求解

大规模的创新类竞赛中，如何评选出较优作品是评审过程中面临的一个难题。由于评阅资金、时间等客观条件的限制，在评阅过程中，往往会有很多的缺陷。每个作品被评阅的次数是有限的，专家的精力也不可能做到将所有论文都评阅一遍。如上文中数据分析结果可知，作品集合学术水平不一致时，由不同专家所打分数往往无法做到统一，导致多种排序方法与分数标准化后，作品的成绩波动大，无法较好的体现一致性问题。

同时，近年来随着过着对教育工作的重视，各类竞赛发展迅速，参赛规模也日趋增大，也为评阅工作带来了一定的挑战。阅卷方式也随着计算机技术的发展不断完善，各种类型的全国竞赛以网上阅卷为主。

考虑当前评审方式中的存在的部分问题，结合“创新型”竞赛的发展趋势，文章主要解决以下问题：

- 作品分配过程中，考虑作品、评委以及所属学校等方面的要求，对问题一的任务分配模型进行完善补充。
- 考虑作品数量过多，增加评审次数或者流程都会增加工作量，优化第一轮评审过程，节省资源，提高评审效率。
- 结合当前竞赛、考试等评分流程机制的发展实际，提出有利于减轻评阅压力，提高评审公平性与合理性的实际建议。

7.1 问题描述与分析

在现实情况中，评委专家往往来自于高校的关于竞赛主题的专家学者，所以存在学生作品和评委来自同一学校的可能。为了确保评审结果的公平合理，原则上都要采取回避原则。同时，也会存在部分高校学生参与竞赛积极主动性高的现象，因此，在分配论文的过程中，也要考虑同一个学校的作品要尽可能的出现在不同的评委手中。上述问题本质上还是组合优化问题，但是约束与目标函数相对问题更加复杂。因此，考虑从问题的论文分配模型基础上进行模型修正与改进。

总结各类评审的排名依据，最常用的是百分制和等级制。但是本质上两者是相同的，百分制可以看作以一分为间隔的等级制，相反，等级制度是以 5 分、10 分等跨度更大的分

数为一个等级。百分制的优势显而易见，分数给出后能够直接得到具体的排名，但是由于不同评委的喜好与尺度等原因，所得的结果可能是不准确的。而等级制度由于评价标准宽、分数跨越尺度大，一个作品的等级如好中差，往往是比较确定的结果，所以评判的确定性要更高，缺点是但无法比较出作品的最终排序。

基于上述分析与参考资料，多阶段评审是一种较为合理的评审模式。因此，在现有评审模式的基础上，引入基于序关系的筛选模式，一方面，由于好的作品的质量与较差作品质量很容易判断，所以能够通过等级制的转换筛选一部分作品，节省工作量；另一方面，所打分数能够用于后面的论文排序，确定进入第二轮评选的作品集合。

7.2 数学模型的建立

7.2.1 评审流程说明

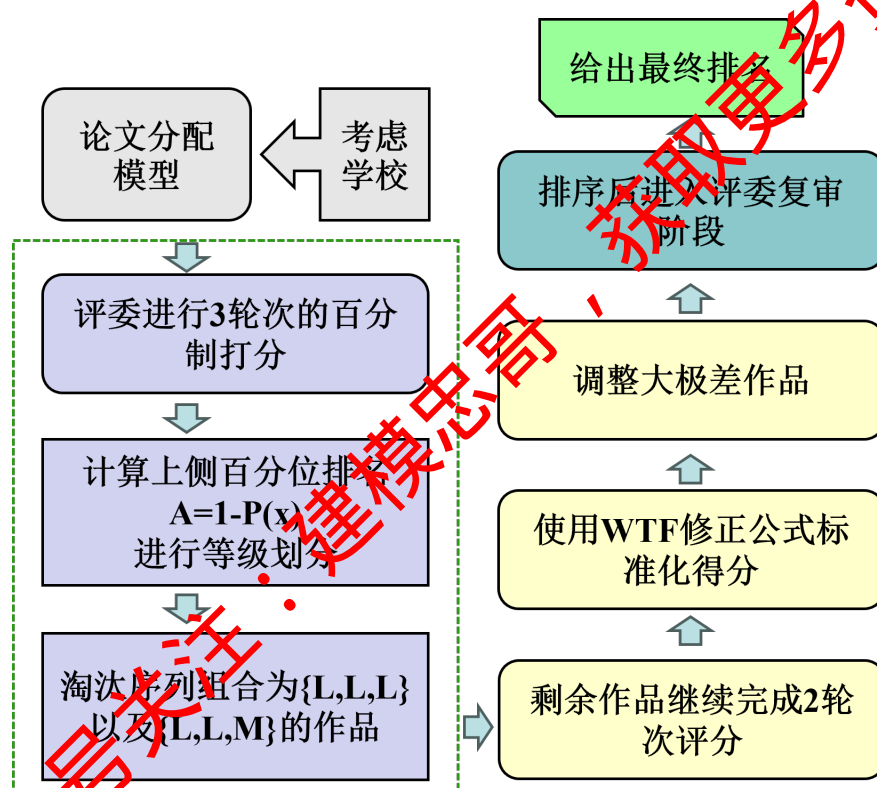


图 22: 评审流程图

图22为改进后的评审流程方案，首先对问题一的任务分配模型进行了完善修改，加入了关于学校的约束与目标；然后进行3次百分制的作品评阅，根据作品在专家的评分中的上偏百分位排名确定其等级序列，若3次评阅都为低序列作品或者2次低序列1次中序列，则将该作品淘汰；接着，将剩余作品继续2轮次的评选后，完成WTF修正公式的分数标准化以及极差处理，基于该成绩进行排序，确定最终的复审名单。

7.2.2 考虑学校的论文分配模型

作品分配的方式采用本文问题中的基于模拟退火的两阶段任务交叉算法进行求解，使得作品的分配尽可能交叉均匀。同时，在问题一的基础上，进一步考虑学校、评委、作品三者之间的从属关系，新增以下要求：

- 来自同一学校的作品和评委要采取回避原则，原则上，评委都不能评阅本校的作品
- 同一学校的作品不能集中到某一评委处，要尽可能的让不同的评委进行评阅

x_{ki} 表示作品 k 与评委 i 之间的决策关系，为 1 表示被评阅，为 0 则反之。 m_{ij} 表示评委 i 与学校 j 的关系，若是 i 来自学校 j 则为 0，否则为 1。 s_{kj} 代表了作品 k 与学校 j 之间的关系，如果作品 j 来自学校 k 则为 1，为 0 则反之。由此，所有评委不能评阅本校作品的约束如下表示：

$$\sum_{k=1}^n x_{ki} m_{ij} s_{kj} = 0 \quad \forall j = 1, 2, 3, \dots; i = 1, 2, \dots, m$$

$$m_{ij} = \begin{cases} 1 & \text{if } i \text{ comes from } j \\ 0 & \text{otherwise} \end{cases}$$

$$s_{kj} = \begin{cases} 1 & \text{if } k \text{ comes from } j \\ 0 & \text{otherwise} \end{cases}$$
(7.1)

$$Q_{ij} = \sum_{k=1}^n x_{ki} s_{kj} \quad \forall j = 1, 2, \dots, m; i = 1, 2, \dots, m$$
(7.2)

Q_{ij} 表示评委 i 评阅到学校 j 的作品总数，为了使学校 j 的作品尽可能的均匀分配，则要保证所有评委评阅到的同一学校的作品数量尽可能小，则有如下目标：

$$\min \left\{ f_3 = \max_i \left(\sum_{k=1}^n x_{ki} s_{kj} \right) \right\}$$
(7.3)

所以基于问题一中的论文分配模型，考虑学校的作品分配模型如下：

$$\min \left\{ f_1 = - \sum_{j=1}^{n-1} \sum_{k=k+1}^n d_{jk} \right\}$$

$$\min \{ f_2 = \max c_{iz} - \min c_{iz} \}$$

$$\min \left\{ f_3 = \max \left(\sum_{k=1}^n x_{ki} s_{kj} \right) \right\}$$
(7.4)

$$\begin{cases} \sum_{i=1}^m x_{ki} \leq \lceil n/m \rceil + 1 \\ \sum_{i=1}^m x_{ki} \geq \lfloor n/m \rfloor - 1 \\ \sum_{k=1}^n x_{ki} m_{ij} s_{kj} = 0 \quad \forall j = 1, 2, 3, \dots; i = 1, 2, \dots, m \\ \sum_{j=1}^n x_{ki} = 5 \\ c_{iz} = \sum_{i=1}^m (x_{ki} \wedge x_{ki}) \\ m_{jk} = \begin{cases} 1 & \text{if } i \text{ comes from } j \\ 0 & \text{otherwise} \end{cases} \\ s_{ik} = \begin{cases} 1 & \text{if } k \text{ comes from } j \\ 0 & \text{otherwise} \end{cases} \\ d_{iz} = \begin{cases} 0 & \text{if } c_{iz} == 0 \\ 1 & \text{otherwise} \end{cases} \\ x_{ki} = 0 \text{ or } 1, i = 1, 2, \dots, m; j = 1, 2, \dots, n \end{cases}$$
(7.5)

7.2.3 基于序关系的筛选流程

为了规避作品数量过多，需要投入大量人力物力的缺陷，本文提出将现有的两阶段评分拆为 3 个阶段，将原有的 5 次评分过程拆解为 3 次和 2 次评审阶段。5 次评审仍然采用百分制，但前 3 次将百分制转换为等级制，进行初步筛选，该过程若为网上评审，可有计算机自动处理完成。后 2 次评选是为了增加进入复审环节作品可信度与公平性进行的，完成后将前 5 次评审的分数标准化，依据排名进入最后的复审阶段。

基于序关系的初选流程（百分制）虽然等级制不利于最终的排名，但是通常情况下，百分制可以转换到等级制。考虑到一位专家评阅作品的数量大致是 100 份左右，因此可以在打分后将某位评委的所有作品的分数作为排序依据，计算其打分的均值与方差，然后确定作品的正态百分位。百分制到排位序的转换过程由正态分布的累计概率函数进行计算，公式如下：

$$\begin{aligned} T(y_{ki}) &= 1 - A(y_{ki}) \\ A(y_{ki}) &= P(y_{ki} \leq Y_i) \\ P(y_{ki} < Y_i) &= \int_{-\infty}^{y_i} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{(x - \bar{y}_i)^2}{2\sigma_i^2}\right\} dx \end{aligned} \quad (7.6)$$

上式 P 为正态分布的累计概率函数，为一个介于 0—1 之间的数，将其转换为百分数的形式，它的意义为作品 k 在评委 i 的总体评分中的位置式是 $100 * T_{yjk} \%$ 也称为正态上侧百分位排名 (*NormalUpperPercentile*)，简称为 NUP-排位 [7]。

一次竞赛过程中，获奖比例大致为 30%。定义在每位专家的作品集合中，前 30% 为高序列（下文以 S 代称），30%—50% 为中间序列（下文以 M 代替），50% 之后的为低序列（下文以 L 代替）。下式含义为在第 h 次评阅中，作品 k 在评委 i 评阅的作品总体中的正态上侧百分位等级的划分依据。

$$R_h(y_{ki}) = \begin{cases} S & T(y_{ki}) > 30\% \\ M & 30\% \leq T(y_{ki}) \leq 50\% \\ L & T(y_{ki}) < 50\% \end{cases} \quad (7.7)$$

经过排列组合后，所有作品可能的位置信息如下表所示：

表 20: 作品序列组合表

作品序列的排列组合	第 1 轮	第 2 轮	第 3 轮
组合 1	高序列 (S)	高序列 (S)	高序列 (S)
组合 2	高序列 (S)	高序列 (S)	中间序列 (M)
组合 3	高序列 (S)	高序列 (S)	低序列 (L)
组合 4	高序列 (S)	中间序列 (M)	低序列 (L)
组合 5	高序列 (S)	中间序列 (M)	中间序列 (M)
组合 6	高序列 (S)	低序列 (L)	低序列 (L)
组合 7	中间序列 (M)	中间序列 (M)	低序列 (L)
组合 8	中间序列 (M)	低序列 (L)	低序列 (L)
组合 9	低序列 (L)	低序列 (L)	低序列 (L)

理想条件下，假设评委能够准确的判断出一份作品的优劣，即能够判断前 50% 与后 50% 的作品之间的差别。低序列作品的划分界限为后 50% 的作品为低序列作品。经过 3 轮评选后，如果一个作品的序列为组合 8 或者组合 9，如下式，则将其淘汰。

$$\begin{aligned} R(n_i) &= \{L, L, L\} \\ R(n_i) &= \{M, L, L\} \end{aligned} \quad (7.8)$$

过程合理性讨论：每位专家评阅的作品数量假设为 100 份，假设某一次竞赛过程中参赛的优胜作品占比为 30%。则优胜作品被淘汰的情况是，连续 3 次前面 50% 都是来自高序列作品。由于样本总体较大，抽样出现 50 份优胜作品的情况的概率数量级为 $e - 15$ 以上，几乎可以忽略不计。

$$Rate_{good} = 30\% \quad (7.9)$$

假设另外一种情况，评委有可能将优胜作品误判到 50% 以后，假设该概率为

$$P_{SL} = 0.02 \quad (7.10)$$

表 21: 作品判别转移概率

实际类别\评判类别	高序列	中间序列	低序列
优胜品	0.86	0.12	0.02
中间序列	0.12	0.85	0.03
低序列	0.02	0.03	0.95

$$\begin{aligned} P_{LLL} &= P_{error} P_{error} P_{error} \\ P_{LLM} &= P_{erroe} P_{error} P_{SM} \\ P_{error} &= p_{SL} + p_{ML} \end{aligned} \quad (7.11)$$

计算可知， $P_{LLL} = 0.0125\%$ ， $P_{LLM} = 0.03\%$ 。如果考虑，在评审过程中，对于评委的某一非常低分的作品给出说明与标注，则误判的概率将会进一步缩小。

第二轮评选以及最终复议：将低序列作品剔除以后，剩余的作品继续未完成的第 4 和第 5 次评阅。完成后，将 5 次评阅的百分制成绩根据问题二中建立的 WLT 修正公式进行得分的标准化，而后调整作品的评分级差。最后选出进入复议环节的作品，根据专家复议结果确定最终排名。

7.3 模型求解和分析

7.3.1 求解结果与分析

我们采用附件 2“数据 1.xlsx.com”的 2015 个作品及其第一轮评审分数进行模拟评阅，假定五位专家分五轮进行打分，位于高序列的得 2 分，位于中间序列的得 1 分，位于低序列的得 0 分，得到 2015 份作品五轮得分如表 22 所示，

基于该表格，计算作品排名变化情况，得到

由表 23 结果可知，经过 3 轮百分制转化为 NUP 排名后的综合计算后，淘汰了 895 个末位的低序列作品，并且，通过对进入前 352 名参赛者的排名可见，该筛选没有影响优胜作品的名次。相比于题目所给的两阶段评审方法，减少了评委的评阅负担，节省了资金，换言之，拥有更多的资源能够用于后续的评审，以保证公平性。

7.3.2 改进建议

在主观意识主导的作品评选过程中，作品得分或者排序与实际水平之间必然会存在一定的偏差，且该偏差很难被衡量。针对大规模创新类竞赛的评选过程，利用科学的评审模式和数理统计知识来缩小客观排序与主观排序之间的偏差是十分有效的。同时，我们也要考虑随机评阅方式、评阅空间、软硬件措施的改进，也可以规定一些有效的举措去减少人为误判的概率。据此本文针对论文评审模式给出以下改进建议：

表 22: 作品五轮序列得分

作品序号	每一轮得分情况					前三轮得分
	一	二	三	四	五	
1	2	2	2	2	2	6
2	2	2	2	2	2	6
3	2	2	2	2	2	6
4	2	2	2	2	2	6
5	2	2	2	2	2	6
6	2	2	2	1	2	6
7	2	2	2	2	2	6
8	2	2	2	1	2	6
...
348	2	2	2	1	2	6
349	2	2	0	2	2	4
350	2	2	0	0	2	4
351	1	2	0	2	2	5
352	2	2	2	2	2	6
353	1	0	0	0	0	1
354	0	0	0	0	0	0
355	0	0	0	0	0	0
356	2	0	0	0	0	2
...
2010	1	2	0	2	0	4
2011	2	2	2	1	0	6
2012	0	0	0	0	0	0
2013	0	0	0	0	0	0
2014	0	1	0	0	0	1
2015	0	0	1	0	0	1

- 当前的评阅方式都是以网上评阅为主，并且计算机系统软件设计技术较为成熟。在论文的评阅过程中，可以考虑添加特定的标签和说明给出评分理由。尤其是针对打分特别高或者特别低的作品，不同评委看到的角度可能不同，当发现别人的评分理由十分充分，自己给低分的原因不能说服的情况中，那么可以进行分数二次调整。该机制的好处是方便处理问题三中的极差情况，不过为了防止评委出现偷懒的行为，应该等评委评分后再推断有没有必要进行说明或者分数的调整。
- 基于大语言模型对作品进行主题分析，采用计算机 + 人的筛选方式进行初选。大语言模型的文本理解能力十分出众，对于论文的结构、主题等都能够清楚的判断出来。考虑到创新类竞赛没有标准答案，解决方案多元的特点，可以对参赛作品进行初步分析。之后将作品分配到对相关领域较为了解的评委进行评阅，这样能够一定程度上避免作品评分波动性大，评委对作品内容认识不深刻的缺陷。另一方面，能够一定程度的较少评委对于论文评阅与筛查的时间消耗，不过要求初期做好评委与模型的训练工作。

表 23: NUP 排位方法的排名变化度量指标

	总作品数	前三轮淘汰数	重合度	平均乱序度	前 352 名重合度
NUP 排位	2015	895	31.4%	429	100%

8. 模型评价

在问题一中，本文通过参赛队伍与评委之间的关系构建了论文分配模型，并且基于评委之间评阅的作品集需要尽可能交叉这一需求，基于模拟退火算法设计了两个阶段的规划操作进行求解，实验证明，其结果要明显优于随机的任务分配方案。在问题四中，考虑了作品来源、评委所属学校以及新增的评阅要求等，进一步修正了问题模型。但模型变量规模也相应地爆炸式增长，非线性约束、求解难度加大。

在问题二中，本文基于描述性统计对成绩数据从专家、作品、阶段等多维度进行了分布特点提取，对比了现有方案在排名效果上的优劣，发现了标准化公式容易受到组间差异的影响，而去除极大极小值通过数据删减一定程度上克服了专家打分的差异性，基于此，我们评估了专家的可信度并将其应用到作品分数的计算中，使得更加可靠的分数所占权重相对更大。将最后的一等奖排名结果与附件 2 数据 1 中的排名相比较，发现我们提出的考虑专家综合评分的加权 T 分数修正公式在分数预测上效果更好。

在问题三中，我们通过分析极差调整的流程和机制，综合利用 SVM 支持向量机、决策树等多种方法训练模型，最终得到了自动化极差调整程序。从模型的分类效果来看，其对于相似的调整模式 1 与 2、4 与 5 的识别准确率还有待提升，但是在实现极差调整的效果上能够达到较好的效果。

在问题四中，在现有评审模型的基础上，进一步考虑了等级制对作品评估准确性高的特点。引入了基于序关系的评审淘汰机制，将前面几次评阅过程中的百分制分数转换为序列等级，优先淘汰不参与后续的作品评阅工作。未被淘汰的作品继续接下来的评审过程，采用问题二中的 WTF 分数修正模型与问题三中的极差调整机制进行标准化处理与排序，最后取较优作品进行专家复审阶段。

8.1 模型的优点

1. 论文分配模型的目标函数与约束与实际较为符合，所求解能够用于现实的论文分配；
2. 基于序关系的论文筛选机制充分利用现有数据减少了了评阅的工作量，所得结果不会导致明显偏差；
3. 考虑专家可靠性作为分数权重的策略有效减弱了专家之间的差异性带来的组间不均衡；
4. 基于机器学习算法的多阶段极差调整模型能够有效实现极差调整的自动化，并一定程度上筛选出名次变化比较大的作品，有利于进一步探究其创新性。

8.2 模型的缺点

1. 论文分配模型本质上为 NP-Hard 问题，由于非线性约束较多，无法使用精确求解器进行求解，且随着问题要求的增加，解空间也会越来越复杂
2. 基于序关系的论文筛选机制，在评阅规模较小，评委所的给评分不满足正态分布时，无法准确衡量作品的 NUP 排位

3. 进行极差调整时，对极差调整模式识别的准确率还比较低，不能准确识别极差调整的六个模式，且单次极差调整难以同时调整多个分数，存在少部分作品进行极差调整后仍然存在大极差问题

8.3 模型的改进和展望

1. 实际论文评选过程中，也要考虑评委的专业领域，满足约束条件的基础上有针对性地进行任务分配；
2. 结合主题挖掘、文本分析、大语言模型等机制进一步完善论文的评阅过程；
3. 指定更加科学的论文评判体系，尽量规避主观因素对作品结果最终成绩的影响；
4. 设计多阶段极差调整环节，同时参考更多信息，丰富训练模型的属性集，提升分类器的准确率。

参考文献

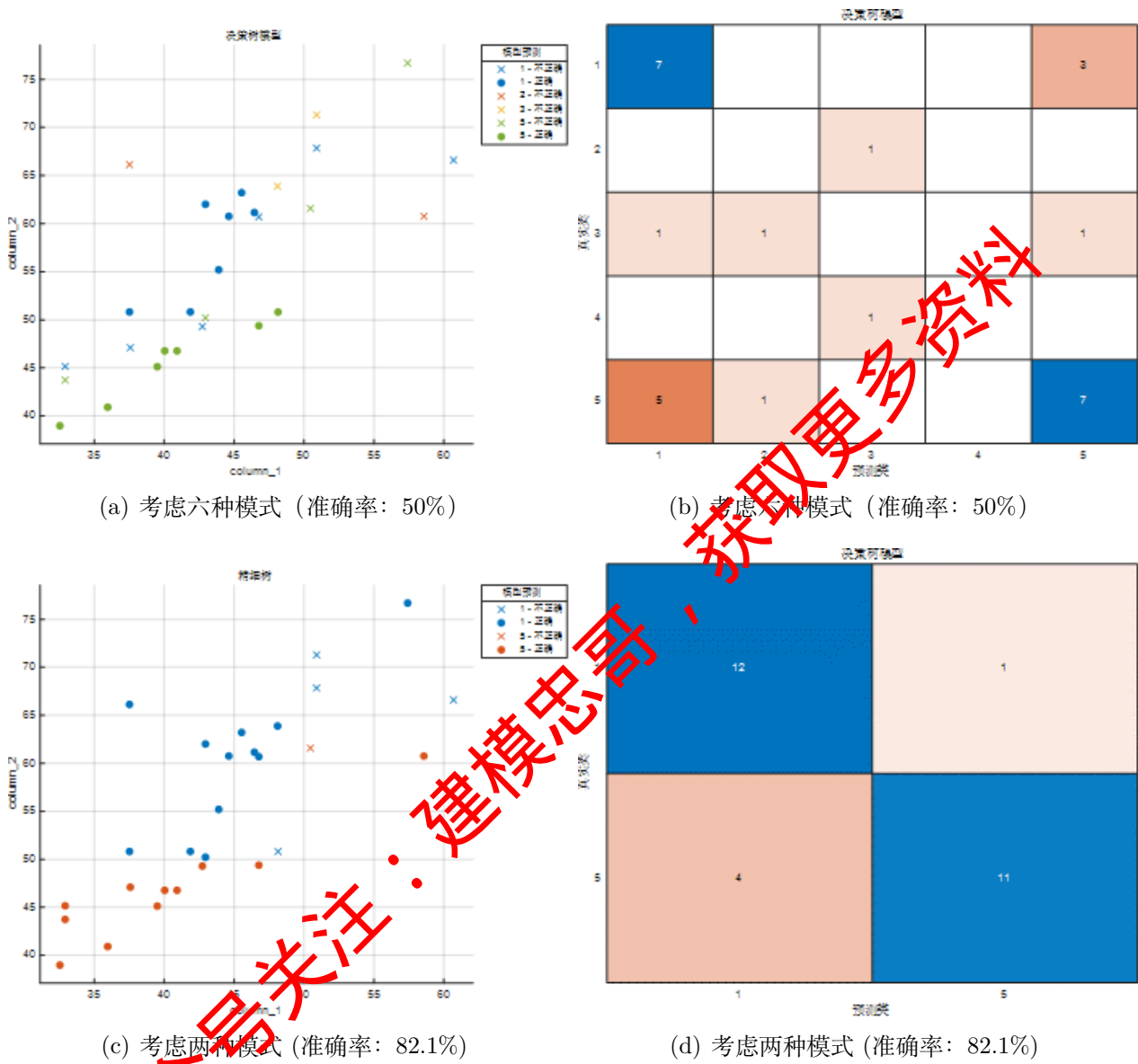
- [1] 郭东威, 丁根宏, 毛俊诚, 陈玉磊. 群决策论文型竞赛名次的加权 T 分数法 [J]. 中国科技论文, 2015, (17): 2059-2063.
- [2] 吕书龙, 梁飞豹, 刘文丽. 关于评委评分的评价模型 [J]. 福州大学学报·自然科学版, 2010, 第 38 卷 (3): 358-362.
- [3] 鄢丽. 公平的评卷系统 [D]. 哈尔滨工业大学, 2007.
- [4] 易昆南, 梁霞, 易芳. 缺损评分矩阵的论文排名 [J]. 铁道科学与工程学报, 2008(03): 93-96. DOI:10.19713/j.cnki.43-1423/X.2008.03.020.
- [5] 毛紫阳, 吴孟达. 基于序关系的竞赛评委评价方法 [J]. 数学的实践与认识, 2006, 36(10): 14-23. DOI:10.3969/j.issn.1000-0984.2006.10.003.
- [6] 田园, 丁根宏. 群组决策论文名次的集对分析模型 [J]. 安徽师范大学学报 (自然科学版), 2018, 41(01): 19-23. DOI:10.14182/J.cnki.1001-2443.2018.01.005.
- [7] 徐明. 正态分布应用于分数排位及调整的研究 [D]. 华东师范大学, 2012.
- [8] 龚千健, 王涛, 裴莹莹. 打分机制公平性评估的概率统计模型 [J]. 数学建模及其应用, 2013, 2(02): 31-41.
- [9] 刘星子, 林亮, 臧东冉. 大型竞赛的一种评卷模型 [J]. 广西科学, 2008(03): 266-268+273.
- [10] 易昆南. 残缺数据的论文名次及评委水平的评判与逆判 [J]. 湘潭大学自然科学学报, 2005(02): 39-43.
- [11] Xuesheng P, Xiwen Z. Behavioral Analysis of Judges in Subjective Evaluation Based on Statistics[C]//AEIC Academic Exchange Information Centre(China). Proceedings of 2017 2nd International Conference on Humanities Science and Society Development (ICHSSD 2017) (Advances in Social Science, Education and Humanities Research VOL.155). Atlantis Press, 2017: 5.

附录 A 复议作品模式识别训练集

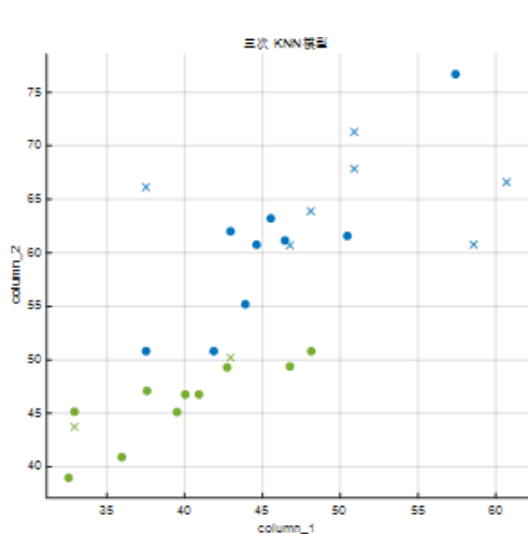
表 24: 复议作品模式识别训练集

最低分	中间分	最高分	下偏距离	上偏距离	偏差距离差	偏差度	极差	原平均值	调整
57.41	76.69	78.11	19.28	1.42	-17.86	0.07	20.70	70.74	1.00
60.69	66.60	81.83	5.91	15.23	9.32	2.58	21.14	69.71	5.00
48.11	63.88	73.07	15.77	9.19	-6.58	0.58	24.96	61.69	1.00
58.57	60.76	85.33	2.19	24.57	22.38	11.22	36.76	68.22	5.00
50.46	61.58	73.75	11.12	12.17	1.05	1.09	23.29	61.82	1.00
50.90	71.29	75.61	20.39	4.32	-16.07	0.21	24.71	65.93	5.00
50.90	67.85	71.17	16.95	3.32	-13.63	0.20	20.27	63.31	5.00
46.45	61.15	69.18	14.70	8.03	-6.67	0.55	22.13	58.93	1.00
42.95	62.01	64.98	19.06	2.97	-16.09	0.16	22.03	56.65	1.00
44.62	60.76	66.69	16.14	5.93	-10.21	0.37	22.07	57.36	1.00
37.52	66.13	73.40	28.61	7.27	-21.34	0.25	35.88	59.02	1.00
45.54	63.21	66.13	17.67	2.92	-14.75	0.17	20.59	58.29	1.00
43.90	55.20	65.24	11.30	10.04	-1.26	0.89	21.34	54.78	1.00
46.77	60.69	67.94	13.92	7.25	-6.67	0.52	21.17	58.47	1.00
41.87	50.82	65.24	8.95	14.42	5.47	1.61	23.37	52.64	1.00
42.95	50.22	65.53	7.27	15.31	8.04	2.11	22.58	52.90	1.00
46.77	49.39	66.86	2.62	17.45	14.85	6.67	20.09	54.34	5.00
48.14	50.81	71.20	2.67	18.39	17.72	7.64	23.06	56.72	5.00
42.73	49.30	62.93	6.57	12.63	7.06	2.07	20.20	51.65	5.00
37.52	50.82	57.87	13.30	7.05	-6.25	0.53	20.35	48.74	1.00
40.92	46.77	61.42	5.85	14.65	8.80	2.50	20.50	49.70	5.00
37.58	47.10	61.42	9.52	14.32	4.80	1.50	23.84	48.70	5.00
40.04	46.77	61.42	6.73	14.65	7.92	2.18	21.38	49.41	5.00
32.92	43.74	62.51	10.82	18.77	7.95	1.73	29.59	46.39	5.00
39.51	45.13	60.93	5.62	15.80	10.18	2.81	21.42	48.52	5.00
35.96	40.92	61.60	4.96	22.68	17.72	4.57	27.64	46.83	5.00
32.92	45.16	55.99	12.24	10.83	-1.41	0.88	23.07	44.69	5.00
32.54	38.99	55.09	6.45	16.10	9.65	2.50	22.55	42.21	5.00

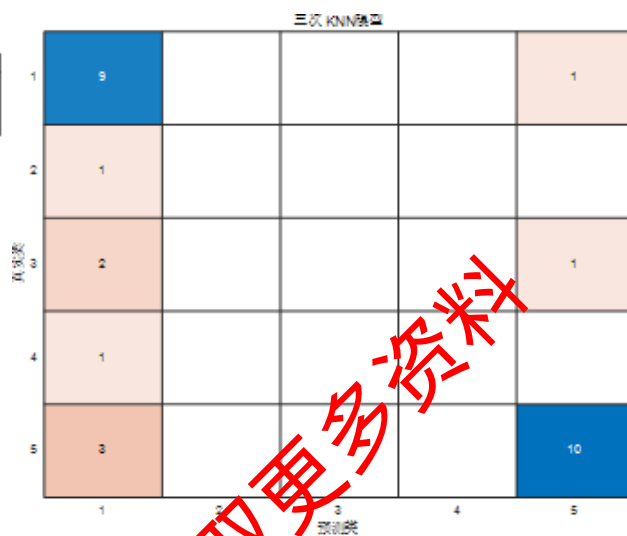
附录 B 决策树模型预测散点图与混淆矩阵结果图



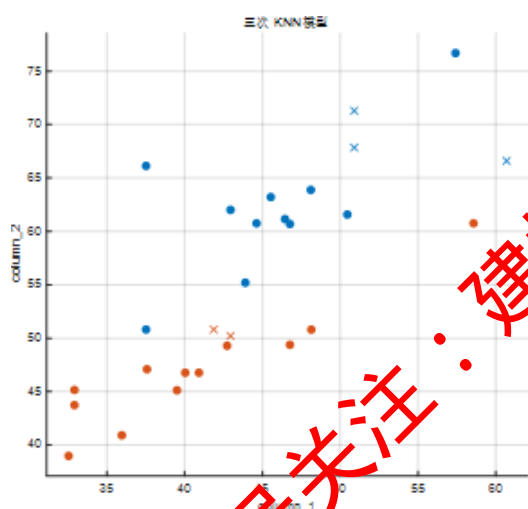
附录 C KNN（三次）模型预测散点图与混淆矩阵结果图



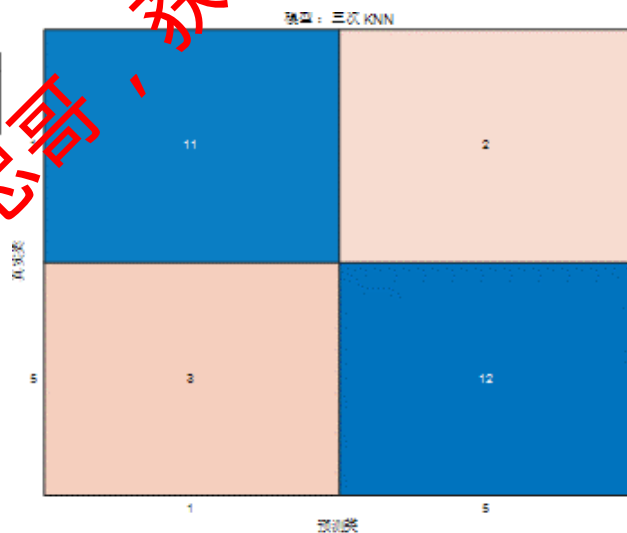
(a) 考虑六种模式 (准确率: 67.9%)



(b) 考虑六种模式 (准确率: 67.9%)



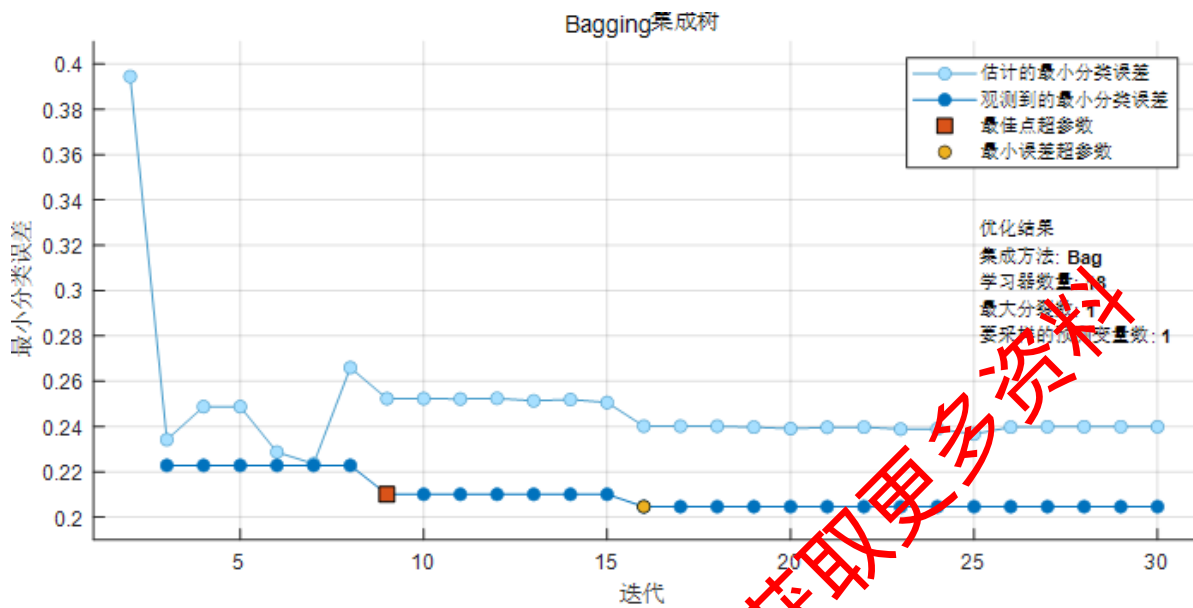
(c) 考虑两种模式 (准确率: 82.1%)



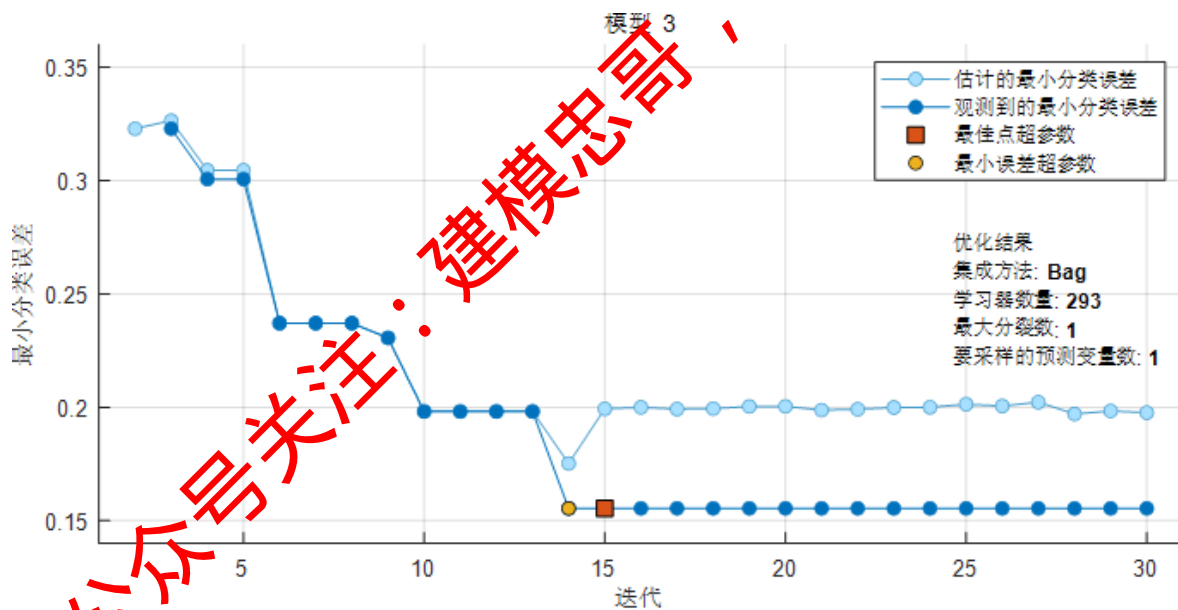
(d) 考虑两种模式 (准确率: 82.1%)

图 24: KNN（三次）模型预测散点图与混淆矩阵结果图

附录 D Bagging 集成树模型的最小分类误差图



(a) 考虑六种模式 (准确率: 67.9%)



(b) 考虑两种模式 (准确率: 82.1%)

图 25: Bagging 集成树模型的最小分类误差图

附录 E 两种模式下的样本训练属性集

表 25: 两种模式下的样本训练属性集

最低分	中间分	最高分	下偏距离	上偏距离	偏差距离差	偏差度	极差	原平均值	调整
57.41	76.69	78.11	19.28	1.42	-17.86	0.07	20.70	70.74	1.00
60.69	66.60	81.83	5.91	15.23	9.32	2.58	21.14	69.71	5.00
48.11	63.88	73.07	15.77	9.19	-6.58	0.58	24.96	61.69	1.00
58.57	60.76	85.33	2.19	24.57	22.38	11.22	36.76	68.22	5.00
50.46	61.58	73.75	11.12	12.17	1.05	1.09	23.29	61.82	1.00
50.90	71.29	75.61	20.39	4.32	-16.07	0.21	24.71	65.93	5.00
50.90	67.85	71.17	16.95	3.32	-13.63	0.20	20.27	63.31	5.00
46.45	61.15	69.18	14.70	8.03	-6.67	0.55	22.13	58.93	1.00
42.95	62.01	64.98	19.06	2.97	-16.09	0.16	22.03	56.65	1.00
44.62	60.76	66.69	16.14	5.93	-10.21	0.37	22.07	57.36	1.00
37.52	66.13	73.40	28.61	7.27	-21.34	0.25	35.88	59.02	1.00
45.54	63.21	66.13	17.67	2.92	-14.75	0.17	20.59	58.29	1.00
43.90	55.20	65.24	11.30	10.04	-1.26	0.89	21.34	54.78	1.00
46.77	60.69	67.94	13.92	7.25	-6.67	0.52	21.17	58.47	1.00
41.87	50.82	65.24	8.95	14.42	5.47	1.61	23.37	52.64	1.00
42.95	50.22	65.53	7.27	15.31	8.04	2.11	22.58	52.90	1.00
46.77	49.39	66.86	2.62	17.45	14.85	6.67	20.09	54.34	5.00
48.14	50.81	71.20	2.67	18.39	17.72	7.64	23.06	56.72	5.00
42.73	49.30	62.93	6.57	12.63	7.06	2.07	20.20	51.65	5.00
37.52	50.82	57.87	13.30	7.05	-6.25	0.53	20.35	48.74	1.00
40.92	46.77	61.42	5.85	14.65	8.80	2.50	20.50	49.70	5.00
37.58	47.10	61.42	9.52	14.32	4.80	1.50	23.84	48.70	5.00
40.04	46.77	61.42	6.73	14.65	7.92	2.18	21.38	49.41	5.00
32.92	43.74	62.51	10.82	18.77	7.95	1.73	29.59	46.39	5.00
39.51	45.13	60.93	5.62	15.80	10.18	2.81	21.42	48.52	5.00
35.96	40.92	61.60	4.96	22.68	17.72	4.57	27.64	46.83	5.00
32.92	45.16	55.99	12.24	10.83	-1.41	0.88	23.07	44.69	5.00
32.54	38.99	55.09	6.45	16.10	9.65	2.50	22.55	42.21	5.00

附录 F 问题三核心代码：程序化调整极差

```
clc;
close all;
%% 构建各阶段每个专家的评阅作品集以及其标准分
num=0;
for j =1:1:size(count,1)% j为专家的序号, num为
    for i = 1:1:count(j,2)
        professor_work_1(j,i) = professor_1(num+i,2);
    end
    num = num+count(j,2);
end
%% 构建经过复议的作品集合
%判断进入第二阶段的作品数 num_works_2
for i=1:1:size(works,1)
    if(isnan(works(i,4)))
        num_works_2 = i-1;
        break;
    end
end
num_works_reward = 0;
num_works_reward2 = 0;
for i=1:1:size(works,1)
    if(works(i,3)~=0)
        num_works_reward = num_works_reward +1;
    end
    if(works(i,3)==1 | works(i,3)==2)
        num_works_reward2 = num_works_reward2 +1;
    end
end
end
%极差矩阵deviation:
%[列: 作品序号, 第一阶段极差, 第二阶段复议前极差, 第二阶段复议后极差]
deviation = zeros(size(works,1),4);
deviation(:,1) = [1:1:size(works,1)];%第一列为作品序号
deviation(:,2) = works(:,23);%第二列为第一阶段极差
deviation(1:num_works_2,3) = works(1:num_works_2,4);%复议前极差
works_discuss=[];
for i=1:1:size(works,1)
    if(isnan(works(i,5)))
        deviation(i,4) = works(i,4);
    else
        deviation(i,4) = works(i,5);
        works_discuss=[works_discuss;[i,works(i,:)]];
    end
end
end
%% 1、绘制第一阶段极差, 第二阶段复议前极差以及第二阶段复议后极差的散点图
%1-2绘图
%1-2-1绘制折线图
figure();
```

```

plot(deviation(1:num_works_2,1),deviation(1:num_works_2,2),'b^');
hold on;
plot(deviation(1:num_works_2,1),deviation(1:num_works_2,3),'k.',
      deviation(1:num_works_2,1),deviation(1:num_works_2,4),'ro');
legend("所有作品第一阶段极差","入选作品第二阶段复议前极差","入选作品第
      二阶段复议后极差");
xlabel("作品序号");
ylabel("极差值");
%1-2-2绘制经过复议的作品前后的极差变化
figure();
plot(works_discuss(:,1),works_discuss(:,5),'ko',works_discuss(:,1),
      works_discuss(:,6),'r. ');
hold on;
%plot(works_discuss(:,1),20*ones(size(works_discuss(:,1),1)), 'b--');
legend("复议作品复议前极差","复议作品复议后极差");%,"极差调整阈值");
xlabel("作品序号");
ylabel("极差值");
%% 绘制每个复议作品的三个二轮分数以及复议分数27,31,35
figure();
plot(works_discuss(1:size(works_discuss,1)/2,1),works_discuss(1:size(
      works_discuss,1)/2,[27,31,35]),'k+');
hold on;
plot(works_discuss(1:size(works_discuss,1)/2,1),works_discuss(1:size(
      works_discuss,1)/2,[28,32,36]),'r+');
legend("复议前分数","复议前分数","复议前分数","复议后分数");
xlabel("作品序号");
ylabel("极差值");
figure();
plot(works_discuss(size(works_discuss,1)/2:end,1),works_discuss(size(
      works_discuss,1)/2:end,[27,31,35]),'k+');
hold on;
plot(works_discuss(size(works_discuss,1)/2:end,1),works_discuss(size(
      works_discuss,1)/2:end,[28,32,36]),'r+');
legend("复议前分数","复议前分数","复议前分数","复议后分数");
xlabel("作品序号");
ylabel("极差值");
%% 聚类分析确定调整模式
% 按照调整模式将works_discuss中的作品进行分类存储
%sort_works_discuss:[作品序号, 最小值, 最小值复议, 中间值, 中间值复议,
      最大值, 最大值复议]
sort_works_discuss=zeros(size(works_discuss,1),6);
mode = zeros(size(sort_works_discuss,1),1);
for i=1:1:size(works_discuss)
    [sort_works_discuss(i,[1,3,5]),index] = sort(works_discuss(i
        ,[2,4,6]));
    sort_works_discuss(i,[2,4,6])=works_discuss(i,24.+4.*index);
    %判断模式
    if sort_works_discuss(i,1)~=sort_works_discuss(i,2)
        if sort_works_discuss(i,5)~=sort_works_discuss(i,6)
            if sort_works_discuss(i,5)-sort_works_discuss(i,3)>

```

```

        sort_works_discuss(i,3)-sort_works_discuss(i,1)
        mode(i,1)=5;
    else
        mode(i,1)=1;
    end
else
    if sort_works_discuss(i,3)~=sort_works_discuss(i,4)
        mode(i,1)=1;
    else
        mode(i,1)=1;
    end
end
else
    if sort_works_discuss(i,5)~=sort_works_discuss(i,6)
        if sort_works_discuss(i,3)~=sort_works_discuss(i,4)
            mode(i,1)=5;
        else
            mode(i,1)=5;
        end
    else
        mode(i,1)=0;
    end
end
end
end
sort_works_discuss=[works_discuss(:,1),sort_works_discuss];
%% 计算协方差矩阵，构造属性集
%distance[下偏距离，上偏距离，偏差距离差值，偏差成度]
distance=[sort_works_discuss(:,4)-sort_works_discuss(:,2),
    sort_works_discuss(:,6)-sort_works_discuss(:,4),
    sort_works_discuss(:,6)-2*sort_works_discuss(:,4)+
    sort_works_discuss(:,2),
    (sort_works_discuss(:,6)-sort_works_discuss(:,4))./(
    sort_works_discuss(:,4)-sort_works_discuss(:,2)),mode];
corr_sort_works_discuss=[sort_works_discuss,distance,works_discuss
(:,[2,4,5,6])];
corr_sort_works_discuss = [corr_sort_works_discuss, mean(
    corr_sort_works_discuss(:,[2,4,6]),2)];
corr_sort_works_discuss = [corr_sort_works_discuss, mean(
    corr_sort_works_discuss(:,[3,5,7]),2)];
corr_sort_works_discuss = [corr_sort_works_discuss,
    corr_sort_works_discuss(:,end-1)-corr_sort_works_discuss(:,end)];
%%绘制变量间相关性强度图
covmat = corrcoef(corr_sort_works_discuss);
figure();
x3str={"作品序号","最低分","修正最低分","中间分","修正中间分",
    "最高分","修正最高分","下偏距离","上偏距离","偏差距离差","偏差度",
    "调整模式","最终成绩","奖项","极差","修正极差","原平均值","现平均值",
    "平均值变化量"}; %新坐标的值
imagesc(covmat);

```

```

colorbar;
xtextp = 1:19;%每个标签放置位置的横坐标, 这个自然应该和原来的一样了。
ytextp = 19.5*ones(1,length(xtextp));
text(xtextp,ytextp,x3str,'HorizontalAlignment','right','rotation',45,"
    fontsize",10)
set(gca,'XTickLabel',[]); %将原坐标(1,2,3,...)去掉
ytextp = 1:19;%每个标签放置位置的横坐标, 这个自然应该和原来的一样了。
xtextp = -1.4*ones(1,length(xtextp));
text(xtextp,ytextp,x3str,'HorizontalAlignment','left')
set(gca,'YTickLabel',[]); %将原坐标(1,2,3,...)去掉
str = {'变量相关性强度图'};
text(12,-1.6,str)
title('变量相关性强度图','fontsize',10)
hold on;
xline = 0:20;
xline=xline-0.5*ones(1,21);
yline = 0:20;
yline=yline-0.5*ones(1,21);
for i=1:1:21;
    plot(xline(1,:),yline(1,i)*ones(1,21),'w-');
    plot(xline(1,i)*ones(1,21),yline(1,:), 'w-');
end
%% 将三种模式的复议作品分别存储到不同的矩阵中, 各自第一列为作品序号
mode_1 = [sort_works_discuss(find(mode(:,1))==1,:),
    distance(find(distance(:,end)==1),1:4)];
mode_2 = [sort_works_discuss(find(mode(:,1))==2,:),
    distance(find(distance(:,end)==2),1:4)];
mode_3 = [sort_works_discuss(find(mode(:,1))==3,:),
    distance(find(distance(:,end)==3),1:4)];
mode_4 = [sort_works_discuss(find(mode(:,1))==4,:),
    distance(find(distance(:,end))==4),1:4)];
mode_5 = [sort_works_discuss(find(mode(:,1))==5,:),
    distance(find(distance(:,end))==5),1:4)];
%%
figure();
scatter3(mode_1(:,2),mode_1(:,4),mode_1(:,6),30,'MarkerEdgeColor','k'
    ,...
    'MarkerFaceColor',[0.69 0.09 0.12]);
hold on;
scatter3(mode_2(:,2),mode_2(:,4),mode_2(:,6),20,'MarkerEdgeColor','k'
    ,...
    'MarkerFaceColor',[0 0.75 0.75]);
hold on;
scatter3(mode_3(:,2),mode_3(:,4),mode_3(:,6),20,'MarkerEdgeColor','k'
    ,...
    'MarkerFaceColor',[1 0.84 0.0]);
hold on;
scatter3(mode_4(:,2),mode_4(:,4),mode_4(:,6),20,'MarkerEdgeColor','k'
    ,...
    'MarkerFaceColor',[0.49 0.99 0.0]);

```

```

hold on;
scatter3(mode_5(:,2),mode_5(:,4),mode_5(:,6),20,'MarkerEdgeColor','k'
,...
'MarkerFaceColor',[0 0 1]);
%绘制复议后的作品位置
hold on;
scatter3(mode_1(:,3),mode_1(:,5),mode_1(:,7),30,'>','MarkerEdgeColor','
k',...
'MarkerFaceColor',[0.69 0.09 0.12]);
hold on;
scatter3(mode_2(:,3),mode_2(:,5),mode_2(:,7),20,'>','MarkerEdgeColor','
k',...
'MarkerFaceColor',[0 0.75 0.75]);
hold on;
scatter3(mode_3(:,3),mode_3(:,5),mode_3(:,7),20,'>','MarkerEdgeColor','
k',...
'MarkerFaceColor',[1 0.84 0.0]);
hold on;
scatter3(mode_4(:,3),mode_4(:,5),mode_4(:,7),20,'>','MarkerEdgeColor','
k',...
'MarkerFaceColor',[0.49 0.99 0.0]);
hold on;
scatter3(mode_5(:,3),mode_5(:,5),mode_5(:,7),20,'>','MarkerEdgeColor','
k',...
'MarkerFaceColor',[0 0 1]);
%绘制位置变化三维图连线
x=[mode_1(:,2),mode_1(:,3);mode_2(:,2),mode_2(:,3);mode_3(:,2),mode_3
(:,3);
mode_4(:,2),mode_4(:,3);mode_5(:,2),mode_5(:,3)];
y=[mode_1(:,4),mode_1(:,5);mode_2(:,4),mode_2(:,5);mode_3(:,4),mode_3
(:,5);
mode_4(:,4),mode_4(:,5);mode_5(:,4),mode_5(:,5)];
z=[mode_1(:,6),mode_1(:,7);mode_2(:,6),mode_2(:,7);mode_3(:,6),mode_3
(:,7);
mode_4(:,6),mode_4(:,7);mode_5(:,6),mode_5(:,7)];
plot3(x',y',z','k');
leg_name=['调整最低分前',"调整最低中间前","调整最大最小前","调整最大
中间前",
"调整最大前","调整最高分后","调整最低分后","调整最大最小后",
"调整最小中间后","调整最大中间后"];
legend(leg_name,"location","northwest");
xlabel("最小值");
ylabel("中间值");
zlabel("最大值");
%%绘制新三维坐标的空间分布
figure();
scatter3(mode_1(:,11),mode_1(:,10),mode_1(:,8),30,'MarkerEdgeColor','k'
,...
'MarkerFaceColor',[0.69 0.09 0.12]);
hold on;

```



```

scatter3(mode_2(:,11),mode_2(:,10),mode_2(:,8),20,'MarkerEdgeColor','k'
    ,...
    'MarkerFaceColor',[0.75.75]);
hold on;
scatter3(mode_3(:,11),mode_3(:,10),mode_3(:,8),20,'MarkerEdgeColor','k'
    ,...
    'MarkerFaceColor',[1.84.0]);
hold on;
scatter3(mode_4(:,11),mode_4(:,10),mode_4(:,8),20,'MarkerEdgeColor','k'
    ,...
    'MarkerFaceColor',[0.49.99.0]);
hold on;
scatter3(mode_5(:,11),mode_5(:,10),mode_5(:,8),20,'MarkerEdgeColor','k'
    ,...
    'MarkerFaceColor',[0.0.1]);
legend("调整最低分前","调整最低中间前","调整最大最小前","调整最大中间前",
    "调整最大前","location","northwest");
xlabel("最小值");
ylabel("中间值");
zlabel("最大值");
works_discuss=[works_discuss,model];
train_matrix=[corr_sort_works_discuss(:,[2,4,6,8,9,10,11,15,17]),model];
train_colname_1=[ "最低分","中间分","最高分","下偏距离","上偏距离",
    "偏差距离差","偏差度","极差","原平均分","调整模式"];
[trainedClassifier,validationAccuracy]=trainClassifier(train_matrix)
;
%要对新预测变量列矩阵X进行预测,请使用: yfit=c.predictFcn(X)
yfit=trainedClassifier.predictFcn(train_matrix(:,1:9));
%%
%计算非高非低分数线
works_new=sortrows(works,36,"descend");%将第一轮的作品根据平均分进行
    排序
theta_1=works_new(num_works_reward2,36);
theta_0=works_new(num_works_reward,36);

works_discuss_1=works_new(num_works_reward2:num_works_reward,:);
works_discuss_1(find(works_discuss_1(:,23)<20),:)=[];
%%构建成绩列表
%sort_works_discuss:[作品序号, 最小值, 最小值复议,
%中间值, 中间值复议, 最大值, 最大值复议]
sort_works_discuss_1=zeros(size(works_discuss_1,1),5);
mode=zeros(size(sort_works_discuss_1,1),1);
for i=1:1:size(works_discuss_1)
    sort_works_discuss_1(i,[1,2,3,4,5])=sort(works_discuss_1(i
        ,[8,11,14,17,20]));
end
sort_works_discuss_1=[works_discuss_1(:,2),sort_works_discuss_1];
train_matrix_1=[sort_works_discuss_1(:,2),
    mean(sort_works_discuss_1(:,3:5),2),sort_works_discuss_1(:,6)];
train_matrix_1=[train_matrix_1,train_matrix_1(:,2)-train_matrix_1

```

```

(:,1),
train_matrix_1(:,3)-train_matrix_1(:,2),
train_matrix_1(:,3)-2*train_matrix_1(:,2)+train_matrix_1(:,1),
(train_matrix_1(:,3)-train_matrix_1(:,2)./train_matrix_1(:,2)-
train_matrix_1(:,1)),
train_matrix_1(:,3)-train_matrix_1(:,1),mean(sort_works_discuss_1
(:,2:6),2)];
%模式识别
yfit_1=trainClassifier.predictFcn(train_matrix_1);
%更新分数
for i=1:1:size(sort_works_discuss_1)
    if yfit_1(i,1)==1
        sort_works_discuss_1(i,2)=mean(sort_works_discuss_1(i,2:6));
        if (sort_works_discuss_1(i,6)-sort_works_discuss_1(i,2)>20)
            sort_works_discuss_1(:,2)=sort_works_discuss_1(i,6)-20;
        end
    end
    if yfit_1(i,1)==5
        sort_works_discuss_1(i,6)=mean(sort_works_discuss_1(i,2:6));
        if (sort_works_discuss_1(i,6)-sort_works_discuss_1(i,2)>20)
            sort_works_discuss_1(:,6)=sort_works_discuss_1(i,2)+20;
        end
    end
end
%更新平均值
for i=1:1:size(sort_works_discuss_1)
    works_new(find(works_new(:,2)==sort_works_discuss_1(i,1)),36)=mean(
        sort_works_discuss_1(i,2:6));
end
rank=[[1:1:size(works_new,1)],works_new(:,2)];
works_new=sortrows(works_new,36,"descend");
rank=[rank,works_new(:,2)];
result = zeros(size(rank));
cou =0;
for i=1:1:size(rank(:,1))
    result(i,:)= [i,find(rank(:,2)==i),find(rank(:,3)==i)];
end

```

附录 G 问题三核心代码：基于集成优化的分类器模型

```
function [trainedClassifier, validationAccuracy] = trainClassifier(
    trainingData)
% 提取预测变量和响应
% 以下代码将数据处理为合适的形状以训练模型。
%
% 将输入转换为表
inputTable = array2table(trainingData, 'VariableNames', {'column_1', '
    column_2', 'column_3', 'column_4', 'column_5', 'column_6', 'column_7
    ', 'column_8', 'column_9', 'column_10'});

predictorNames = {'column_1', 'column_2', 'column_3', 'column_4', '
    column_5', 'column_6', 'column_7', 'column_8', 'column_9'};
predictors = inputTable(:, predictorNames);
response = inputTable.column_10;
isCategoricalPredictor = [false, false, false, false, false, false,
    false, false, false];

% 训练分类器
% 以下代码指定所有分类器选项并训练分类器。
template = templateTree(...
    'MaxNumSplits', 16, ...
    'NumVariablesToSample', 9);
classificationEnsemble = fitcensemble(...
    predictors, ...
    response, ...
    'Method', 'Bag', ...
    'NumLearningCycles', 10, ...
    'Learners', template, ...
    'ClassNames', [1;5]);

% 使用预测函数创建结果结构体
predictorExtractionFcn = @(x) array2table(x, 'VariableNames',
    predictorNames);
ensemblePredictFcn = @(x) predict(classificationEnsemble, x);
trainedClassifier.predictFcn = @(x) ensemblePredictFcn(
    predictorExtractionFcn(x));

% 向结果结构体中添加字段
trainedClassifier.ClassificationEnsemble = classificationEnsemble;
trainedClassifier.About = '此结构体是从ClassificationLearnerR2020b
    导出的训练模型。';
trainedClassifier.HowToPredict = sprintf('要对新预测变量列矩阵X进行预
    测，请使用：\nyfit=c.predictFcn(X)\n将'c'替换为作为此结构体
    的变量的名称，例如'trainedModel'。\nX必须包含正好9个列，因
    为此模型是使用9个预测变量进行训练的。\nX必须仅包含与训练数据具有
    完全相同的顺序和格式的\n预测变量列。
href="matlab:helpview(fullfile(docroot,'stats','stats.map'),'
    appclassification_exportmodeltoworkspace')">How to predict using an
```

```

    exported_model</a>。');

% 提取预测变量和响应
% 以下代码将数据处理为合适的形状以训练模型。
%
% 将输入转换为表
inputTable = array2table(trainingData, 'VariableNames', {'column_1', 'column_2', 'column_3', 'column_4', 'column_5', 'column_6', 'column_7', 'column_8', 'column_9', 'column_10'});

predictorNames = {'column_1', 'column_2', 'column_3', 'column_4', 'column_5', 'column_6', 'column_7', 'column_8', 'column_9'};
predictors = inputTable(:, predictorNames);
response = inputTable.column_10;
isCategoricalPredictor = [false, false, false, false, false, false, false, false, false];

% 执行交叉验证
partitionedModel = crossval(trainedClassifier, ClassificationEnsemble, 'KFold', 5);

% 计算验证预测
[validationPredictions, validationScores] = kfoldPredict(partitionedModel);

% 计算验证准确度
validationAccuracy = 1 - kfoldloss(partitionedModel, 'LossFun', 'ClassifError');

```