

一种大型创新类比赛评审方案的研究

赵大胜, 吕园园, 金志尧

上海理工大学机械工程学院, 上海

收稿日期: 2024年1月9日; 录用日期: 2024年3月13日; 发布日期: 2024年3月20日

摘要

近年来, 大规模创新比赛越来越多, 参赛人数和参赛作品的数量也越来越多, 这就要求我们建立更加公平合理科学的大赛评审方案。现阶段这类比赛一般采用两阶段(网评、现场评审)或三阶段(网评、现场评审和答辩)评审, 然而当竞赛规模巨大并且评委评分出现较大分歧时, 现有的评审机制不能够较好地解决这类问题。为了提高大赛评分的科学性与客观性, 本文首先建立线性规划模型将现有的专家集和作品集以交集最大化进行分类完成准备工作, 其次对现有的评分标准进行改进引入总体均值和方差完成对初始标准分模型的优化, 最后建立极差模型对评分结果中“极差”较大的作品进行分类判断其是否需要下一阶段的复议。

关键词

线性规划模型, 标准分模型, 极差模型, 创新类比赛

A Study on the Evaluation Scheme of a Large-Scale Innovation Competition

Dasheng Zhao, Yuanyuan Lyu, Zhiyao Jin

School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Jan. 9th, 2024; accepted: Mar. 13th, 2024; published: Mar. 20th, 2024

Abstract

In recent years, there have been more and more large-scale innovation competitions, and the number of participants and entries has also increased, which requires us to establish a more fair, reasonable and scientific evaluation scheme for the competition. At this stage, this kind of competition generally adopts two stages (online evaluation, on-site evaluation) or three stages (online evaluation, on-site evaluation and defense) evaluation, but when the scale of the competition is huge and the judges' scores are quite different, the existing evaluation mechanism can not solve

this kind of problem well. In order to improve the scientificity and objectivity of the scoring of the competition, this paper firstly establishes a linear programming model, classifies the existing expert set and the portfolio to maximize the intersection to complete the preparatory work, then improves the existing scoring criteria, introduces the population mean and variance, and completes the optimization of the initial standard score model, and finally establishes a range model to classify the works with large “range” in the scoring results to determine whether they need the next stage of reconsideration.

Keywords

Linear Programming Model, Standard Score Model, Range Model, Innovation Competition

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

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

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



Open Access

1. 引言

随着我国高等教育的发展,近年来我国大学招生数量每年都超过 1000 万人。国务院总理李克强曾指出,大学生是实施创新驱动发展战略和推进大众创业、万众创新的生力军,要把创新创业教育融入人才培养,切实增强学生的创业意识、创新精神和创造能力[1]。在党和国家政策的引领下,出现了越来越多的创新类比赛,仅互联网+比赛就有近 1000 万人报名[2],这迫切要求我们建立一个更加公平客观的竞赛评审机制。目前对创新类比赛的研究主要集中在对学生创新能力的培养,以及大赛开展等问题上,很少聚焦在比赛评审方案上。饶柳等[3]分析了“互联网+”环境下大学生创新创业大赛开展的重要因素及瓶颈,最后提出了“互联网+”环境下大学生创新创业大赛开展的突破路径,但是并未对评审方案进行深度的研究。当前这类比赛一般采用两阶段(网评、现场评审)或三阶段(网评、现场评审和答辩)评审,然而当竞赛规模巨大并且评委评分出现较大分歧时,现有的评审机制不能够较好地解决这类问题。这类问题的核心在于如何解决作品分类问题和极差较大问题,因此我们对现有的评审标准进行改进,并对偏差程度过大的作品进行再次评审,从作品和专家出发,从资源的充分利用层面出发,提出了一种新型的大型的创新类比赛评审方案,用于解决在作品评审过程中可能出现的问题,保证评审的公平性合理性和科学性。

2. 作品分发与评分标准的建立

本文用到的符号及其含义如表 1 所示:

Table 1. Symbol description

表 1. 符号说明

符号	含义
D	离散系数
d	欧氏距离
a_k	某位专家给出的成绩
$\bar{\alpha}$	某位专家给出成绩的样本均值
s	某位专家给出成绩的样本均值

2.1. 线性规划模型

线性规划(Linear Programming, 简记 LP)则是数学规划中的一个关键领域。自从 1947 年 G.B. Dantzig 提出了解决线性规划的单纯形方法以来, 线性规划在理论上逐渐成熟, 在实践中得到了广泛而深入的应用。利用线性规划知识对不同的评审专家和评审作品进行分类, 确保在分配问题中评审专家和评审作品间的“交集”符合预期。

2.1.1. 线性规划模型的建立

确立模型的变量、目标函数和约束条件(我们以 3000 份作品, 125 位评审专家为例)。

变量:

二进制变量: x_{ij} 表示专家 i 是否评审作品 j , 其中 i 的范围是从 0 到 124, j 的范围是从 0 到 2999。

整数变量: z_{ij} 表示专家 i 和 j 之间的交集大小, 其中 i 和 j 为不同的专家, 且 $i < j$ 。整数变量: w_{ij} 表示交集大小之差的绝对值, 其中 i 和 j 为不同的专家, 且 $i < j$ 。

目标函数: 最小化交集大小之差的绝对值之和[4]:

$$\text{minimize} \sum_{i=0}^{124} \sum_{j=i+1}^{124} w_{ij} \quad (1)$$

约束条件:

每份作品必须由 5 位专家评审:

$$\sum_{i=0}^{124} x_{ij} = 5, \forall j \in \{0, 1, 2, \dots, 2999\} \quad (2)$$

每位专家必须评审至少 120 份作品:

$$\sum_{j=0}^{2999} x_{ij} = 120, \forall i \in \{0, 1, 2, \dots, 124\} \quad (3)$$

计算交集大小和绝对差值的约束:

1) 计算交集大小 z_{ij} 的约束(k 表示作品编号):

$$z_{ij} \geq x_{ik} + x_{jk} - 1, \forall i, j \text{ 且 } i < j, \forall k \in \{0, 1, 2, \dots, 2999\} \quad (4)$$

$$z_{ij} \leq x_{ik}, \forall i, j \text{ 且 } i < j, \forall k \in \{0, 1, 2, \dots, 2999\} \quad (5)$$

$$z_{ij} \leq x_{jk}, \forall i, j \text{ 且 } i < j, \forall k \in \{0, 1, 2, \dots, 2999\} \quad (6)$$

2) 计算绝对插值 w_{ij} 的约束:

$$w_{ij} \geq z_{ij} - \text{abs}(z_{ij}), \forall i, j \text{ 且 } i < j \quad (7)$$

$$w_{ij} \geq \text{abs}(z_{ij}) - z_{ij}, \forall i, j \text{ 且 } i < j \quad (8)$$

以上描述了线性规划问题的目标函数和约束条件, 目标是最小化交集大小之差的绝对值之和, 同时满足每位专家评审足够数量的作品。

2.1.2. 模型的结果

运用上述模型求解的部分专家评委分发到的作品序号如下所示(1~5 为同一个作品分发给不同评审专家, 依此类推)。利用上述模型可以在对作品与专家的分配问题上得到较好的解决, 避免了同一份作品没有被两位及以上专家评审和每两位专家没有评审到相同作品的情况, 实现了资源分配的合理化与效益

的最大化，一定程度上实现了公平性。

评审专家 1 评审的作品列表：[15, 58, 63, 70, 136, 151, 193, 201, 235, 255, 395, 420, 430, 457, 484, 493, 500, 510, 535, 539, 568, 569, 598, 617, 630, 646, 657, 664, 668, 687, 693, 701, 726, 734, 741, 747, 752, 784, 800, 806, 810, 815, 820, 821, 822, 823, 824, 825, 826, 923, 1360, 1425, 1452, 1458, 1480, 1514, 1535, 1540, 1547, 1554, 1556, 1571, 1579, 1600, 1638, 1641, 1674, 1680, 1692, 1699, 1701, 1715, 1734, 1739, 1761, 1778, 1831, 1844, 1850, 1862, 1927, 1945, 1968, 1993, 2005, 2012, 2035, 2051, 2080, 2105, 2155, 2157, 2205, 2223, 2251, 2258, 2274, 2296, 2338, 2365, 2402, 2418, 2419, 2456, 2492, 2527, 2538, 2560, 2568, 2593, 2623, 2653, 2681, 2709, 2736, 2763, 2790, 2818, 2861, 2899]。

2.2. 标准分模型

2.2.1. 数据分析

在制定新的标准分规则之前，我们需要找到评审专家评分差异的原因，究竟是不同评审专家不同的个人色彩导致的，还是由于不同评审专家的作品集合的学术水平差异导致的，即各位评审专家的作品集合学术水平是否相同。

我们确立了以下离散系数 D [5]。

$$D = \frac{\sigma}{\mu} = \sqrt{\frac{n\alpha(1-\alpha)(N-n)}{(N-1)}} \quad (9)$$

其中， σ 和 μ 分别为所选取样本的均值和标准差。

在 2015 个作品中选取 50 个样本，对比原始离散系数和所选样本的评判系数的差异，依据是否获奖来分类，如下表 2 所示：

Table 2. Comparison of discrete coefficients
表 2. 离散系数对比

获奖级别	局部离散系数	离散系数
一等奖	0.2918	0.3812
二等奖	0.2489	0.3564
三等奖	0.2142	0.2689
获奖	0.1362	0.1783

我们可以看到一等奖的离散系数最大，由于一等奖的获奖个数最少，所以不影响最后的评判结果，二、三等奖的评判系数相应减少，符合我们的预期。总体获奖的离散系数较小，反映了真实的样本之间可能存在 10% 到 20% 的误差，这是在合理的范围之内的。

结果表明不同评审专家的作品合集的学术水平是相同的，我们不用担心专家拿到的作品差异导致评判结果的改变。

2.2.2. 标准分模型的建立

在对作品进行评分之前我们必须建立统一的标准来确保评分的可靠性使其具有说服力。在这里我们引入欧氏距离的定义用来作为模型优化的评判标准，它是在 m 维空间中两个点之间的真实距离，欧式距离越小说明模型的误差越小。在二维和三维空间中的欧式距离就是两点之间的距离，二维的公式是[6]：

$$d = \sqrt{(x^1 - x^2)^2 + (y^1 - y^2)^2} \quad (10)$$

初始模型:

$$X_k = 50 + 10 \times \frac{a_k - \bar{\alpha}}{s} \quad (11)$$

其中, a_k 为某位专家给出的成绩

$\bar{\alpha}$ 为某位专家给出成绩的样本均值: $\bar{\alpha} = \frac{1}{n} \sum_{k=1}^n \alpha_k$

s 为某位专家给出成绩样本标准差: $s = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (\alpha_k - \bar{\alpha})^2}$

初始模型所给出的标准分模型中含有某位专家给出成绩的样本均值和标准差, 为了确保每位专家打分的统一性, 我们准备对这两个系数加入所有专家给出样本成绩的均值和标准差作为修正, 以改进标准差公式。

改进模型:

$$X'_k = 50 + 10 \times \frac{a_k - (\bar{\alpha} - \bar{\alpha}')/2}{(s + s')/2} \quad (12)$$

其中, $\bar{\alpha}'$ 为所有专家给出所有样本成绩的均值

s' 为所有专家给出所有样本成绩的标准差。

2.2.3. 模型的求解

我们利用上文提到的初始模型和新建立的改进模型来计算欧氏距离以此来对比两个模型的不同(表3)。

Table 3. Comparison of models before and after improvement

表 3. 改进前后模型对比

奖项	最终成绩	初始标准分 总成绩	改进标准分 总成绩	初始标准分 欧氏距离	改进标准分 欧氏距离
一等奖	257.87	240.88	238.77	16.99	19.1
一等奖	257.06	62.592	264.47	194.468	7.41
一等奖	250.62	63.592	253.2965	187.028	2.676496
一等奖	246.39	59.682	234.7571	186.708	11.63289
一等奖	246.63	234.894	236.7	11.736	9.93
一等奖	250.93	262.222	258.35	11.292	7.42
一等奖	242.67	62.33	266.34	180.34	23.67
一等奖	242.62	58.686	256.93	183.934	14.31
一等奖	243.05	233.054	223.87	9.996	19.18
一等奖	259.57	60.974	280.6	198.596	21.03
一等奖	264.15	63.056	281.279	201.094	17.12899
一等奖	250.91	236.762	242.0304	14.148	8.879587

续表

一等奖	246.06	241.876	248.2245	4.184	2.164542
一等奖	244.03	226.018	228.1863	18.012	15.84372
一等奖	268.8	252.8	274.02	16	5.22
一等奖	281.77	277.766	262.077	4.004	19.69299
一等奖	274.67	68.074	276.6097	206.596	1.939731
一等奖	254.95	249.92	263.762	5.03	8.812
一等奖	254.62	233.12	251.372	21.5	3.248
一等奖	254.2	63.596	258.7156	190.604	4.515616
一等奖	244.49	63.436	235.728	181.054	8.762022
一等奖	263.88	63.988	279.89	199.892	16.01
一等奖	247.9	61.886	258.01	186.014	10.11
一等奖	256.26	61.492	233.0368	194.768	23.2232
一等奖	248.42	62.412	240.0026	186.008	8.417392
一等奖	242.72	64.138	232.5726	178.582	10.14739
一等奖	245.88	60.592	250.29	185.288	4.41

由图1可以看到, 改正之后的标准分计算模型得到的总分与最佳评判标准的总分的欧式距离大约在0~25之间, 明显小于原文标准分计算模型(0~220), 并且由图2可以看到通过改进之后模型的欧式距离均值为11.3, 标准差为6.59, 明显小于原文模型的欧式距离均值117.6和标准差87.74, 说明改进之后的标准分计算模型评选出来的获奖论文分数具有更大的可信度。

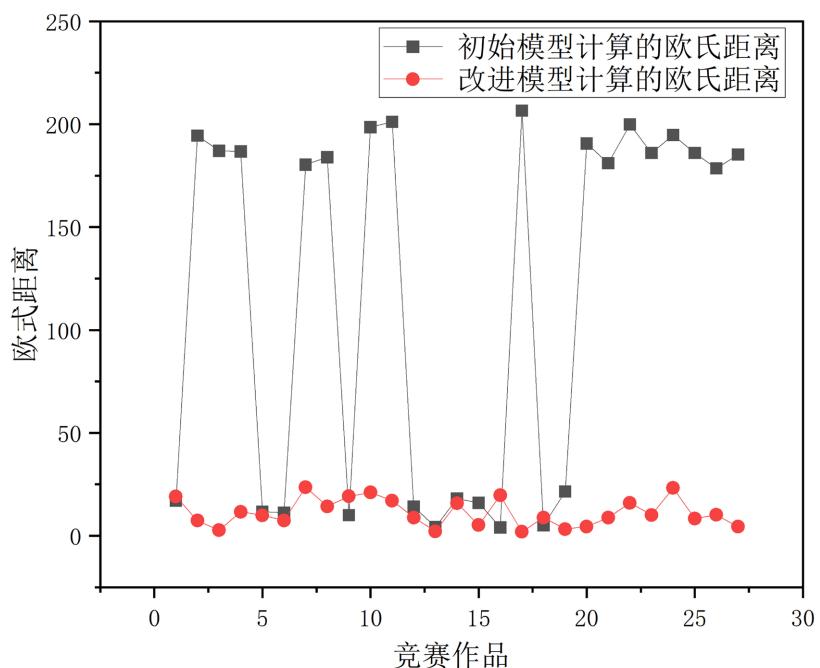


Figure 1. Euclidean distance comparison

图1. 欧氏距离对比

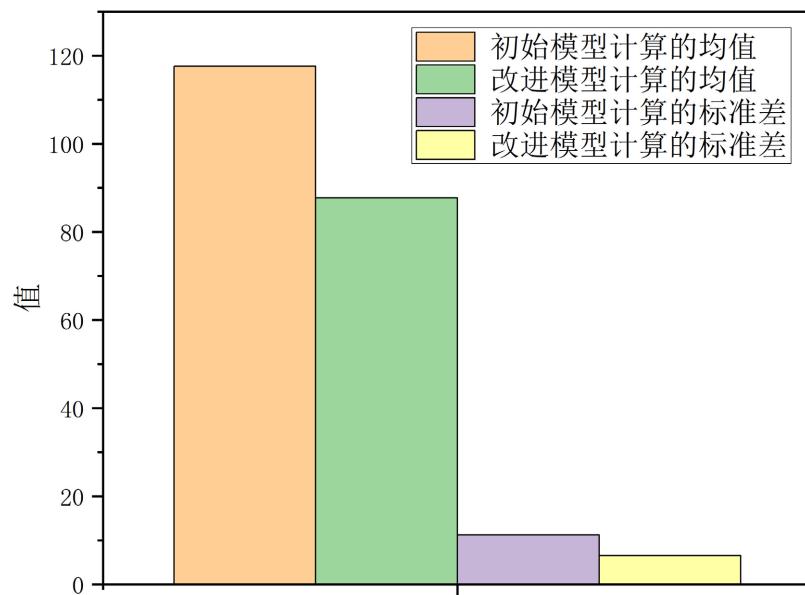


Figure 2. Euclidean distance mean and standard deviation
图 2. 欧氏距离均值和标准差

3. 极差模型的建立与求解

3.1. 极差模型

极差：是指同一组(同一评审阶段同一份作品的成绩集合)数据中最大值与最小值之差。是用来度量评审得分在评审专家之间的差异程度的指标。由于作品的创新到了什么程度，后续研究的前景如何，很难有一致看法，即使专家面对面的交流，都可能由于各持己见而无法统一。加上研究生的论文表达不到位，评审专家的视角不同，同一份作品的几位专家给出的成绩会有较大的极差。基于此我们建立“极差”模型，对上述模型在结果阶段出现的极差过大问题进行优化，同时对评审问题提出新的解决方案。

3.1.1. 数据分析

在此研究中，我们运用模拟数据，这些数据或许反映了不同评审阶段的成绩。初步，我们分别算出了各个阶段的平均分、中位数和标准差，以便综合比较这两个阶段的成绩变化情况并提取了两组数据中的整体的成绩数据与极差数据，通过绘制散点图来进一步解释这些数据所表达的规律，接着，我们推算出了各个阶段的极差，并进行了整体对比。

图 3 中的蓝色直线是最理想评审状态，即两阶段评审分数完全相同，将这一理想状态当作标准可以看出整体的成绩的分布是呈现一个下坠趋势的，这说明第一阶段的打分成绩是高于第二阶段的，从另一角度说明专家在评审第二阶段的作品时更加的谨慎仔细，而在第一阶段则比较随意。因此这也是造成后续极差问题出现的因素之一。

由图 4 可以得出极差分布的一个规律即复议后的极差时比复议之前的极差分布更加靠左的，说明复议后的极差整体是会变小，但是两者的变化差异不会过大。由图 5 可以看出极差值 20 是一个关键的点，极差值大于 20 或是小于 20 将会是“极差”模型定义的一个关键指标。另外从这两张图中可以看出进入第二阶段评审的作品大约是整体作品数的前 10%~30% 左右，因此我们可以认为这部分作品是优秀作品，而作为创新创意大赛的优秀作品则必定是具有创新性的，而这类作品的大部分必定是需要进行复议的，所以这一现象也可以作为后续模型建立的依据之一。

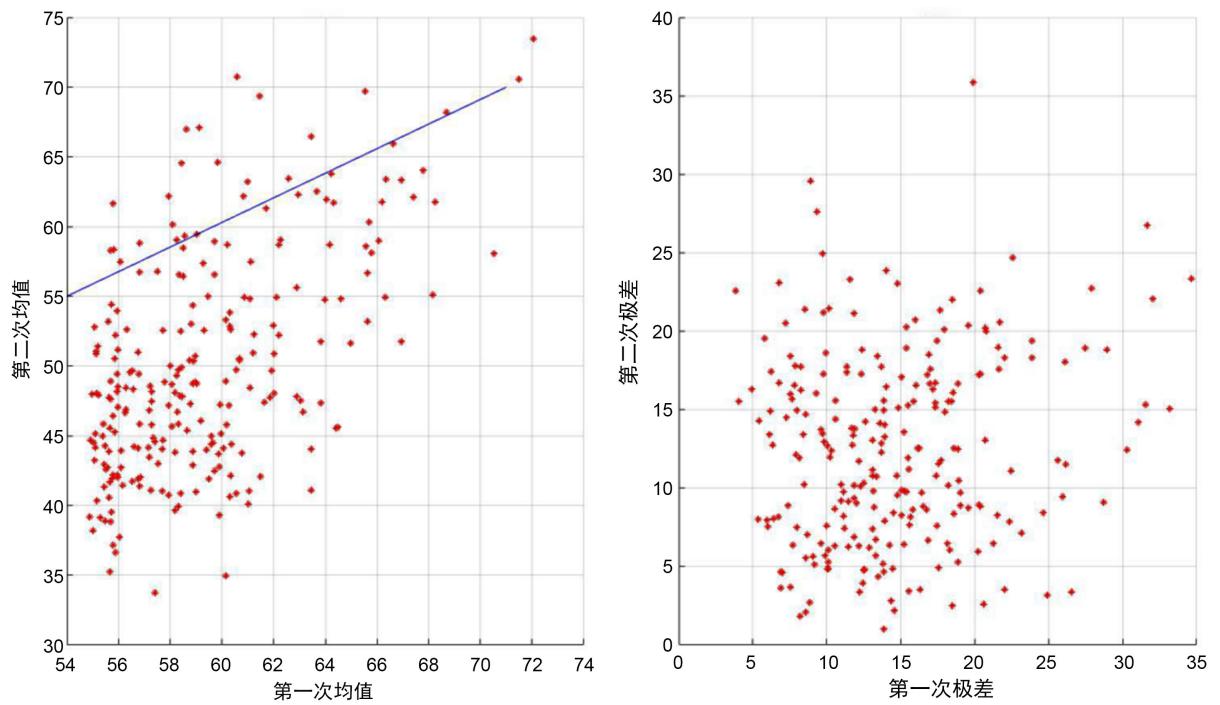


Figure 3. Scatter plot of the overall mean and range distribution of the results
图 3. 成绩整体均值与极差分布散点图

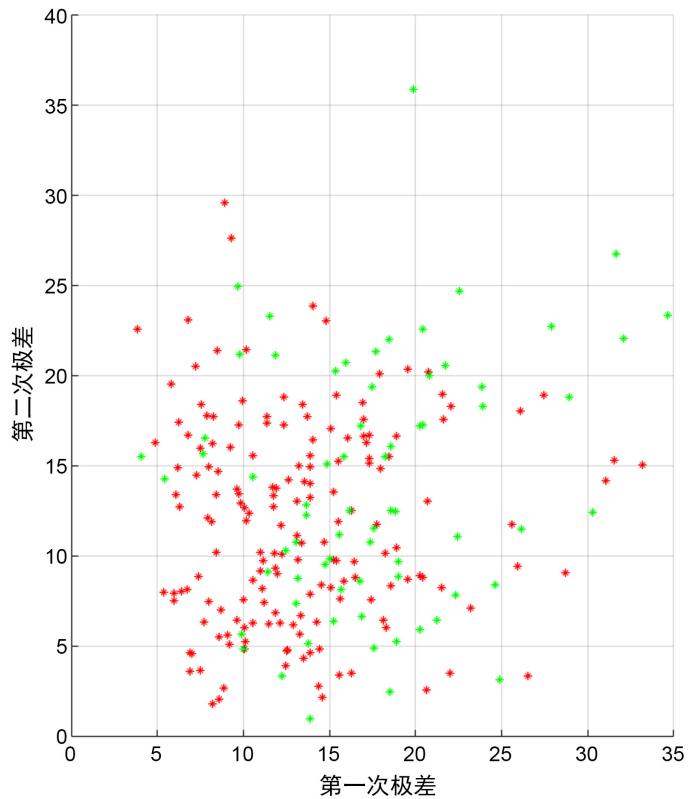


Figure 4. Scatter plot of the distribution of the data before and after the reconsideration
图 4. 极差复议前后分布散点图

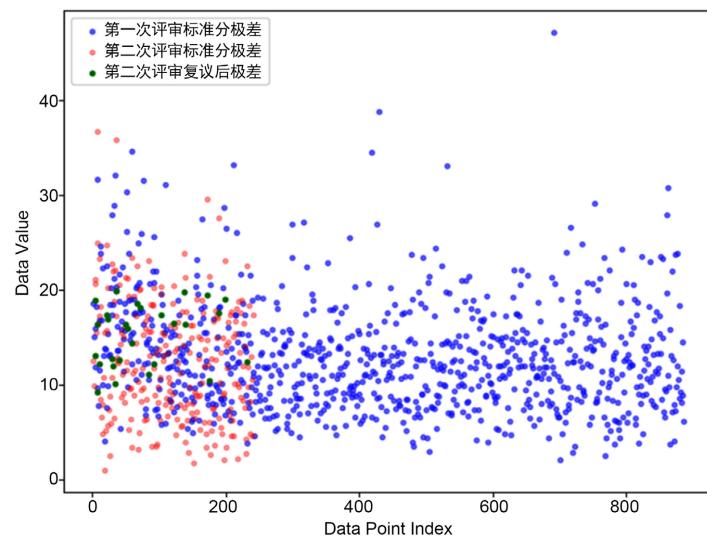


Figure 5. The two-stage range scatter plot is shown from another angle
图 5. 两阶段极差散点图另一角度展示

3.1.2. 极差模型的建立与优化

极差模型是一种分类工具，可通过数据分析提取阈值函数，将作品分为两类：优秀和非优秀。对于优秀作品中的极差部分，我们需要进行调整，因为这些作品具有潜在的优势，可以通过调整来提升它们的竞争力。而对于非优秀作品中的极差大小则不太重要，因为它们通常无法进入第二轮评审，排名较低。为了挑选出优秀作品，我们可以根据其极差特征来设计相应的调整方案。尽管有些作品的极差很大，但经过调整后，它们也有可能进入第二轮评审。在创新创意大赛中，创新是关键竞争因素，排名越高通常代表创新性越强，排名靠前的作品也更有可能提出复议请求。因此，我们可以设定前百分之三十的作品为优秀作品，然后利用它们的数据来确定极差分类的边界。

根据直方图显示的双峰图特性，在 ± 10 处出现了两个峰值，暗示着极差的调整应该保持在适度的范围内，避免过大的幅度。因此，我们可以根据图 6 的信息初步确定每次极差调整的合理数值为 10，将这个数值作为起始的调整步长。

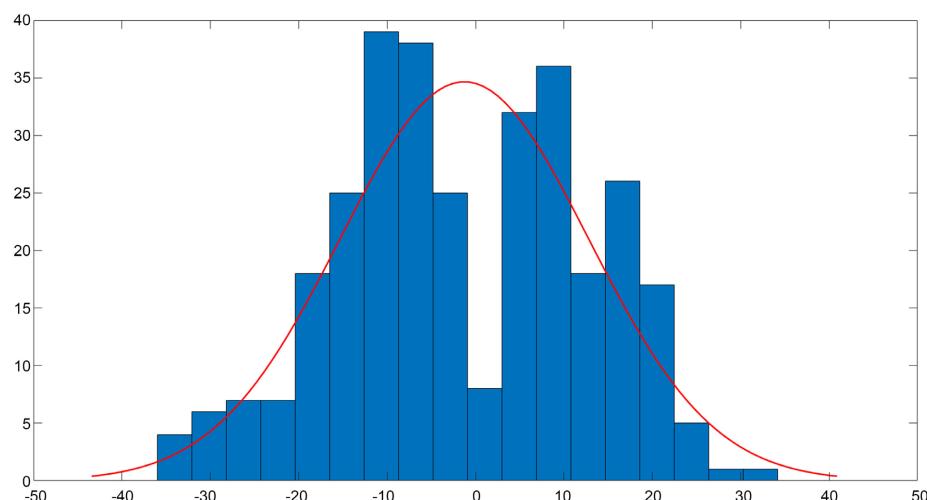


Figure 6. Histogram of the range adjustment law
图 6. 极差调整规律直方图

通过仔细整理和分析数据，我们可以确定极差值为 20 是一个关键点。在极差值为 20 之前的作品，基本上不需要复议，因此它们的极差值可以调整也可以不调整。然而，极差值达到 20 之后的作品，通常会提出复议，因为它们具有较高的极差，往往属于优秀作品，而且经过极差调整后，它们的总成绩通常会有所提升。此外，这些优秀作品也通常具有创新性。

为了自动化极差调整的过程，我们需要建立一个函数关系，将三个专家评分作为输入，并以极差值 20 作为约束条件。这个函数将首先判断作品是否需要提高、降低或保持不变的极差值，从而使极差调整过程不再需要人工干预，而可以通过公式来实现。

基于此建立可以手动调整的极差模型。根据极差阈值划分作品为两个分段：大极差和小极差；对于大极差的作品，可以采用不同的调整方法，比如平均化；计算最终得分时采用加权平均。

根据图 7 的观察结果，我们发现当 $k = 6$ 时，模型的准确率达到最大，在 20 次的取值中模型的准确率最终维持在 0.9 左右。因此我们把 6 作为模型调整的最优步长。

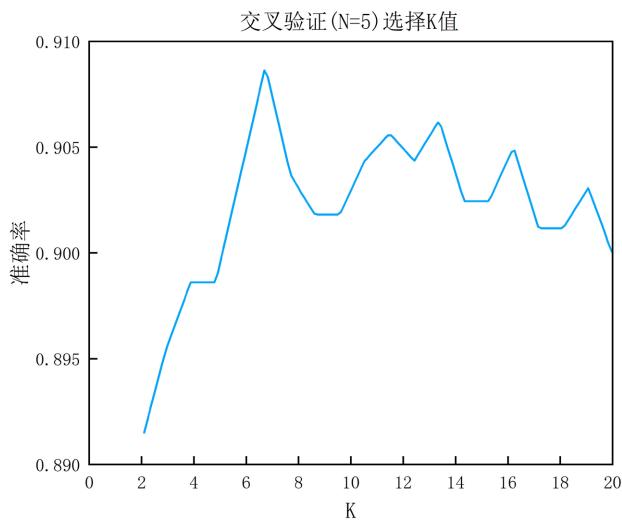


Figure 7. Verification of model accuracy
图 7. 模型精度验证

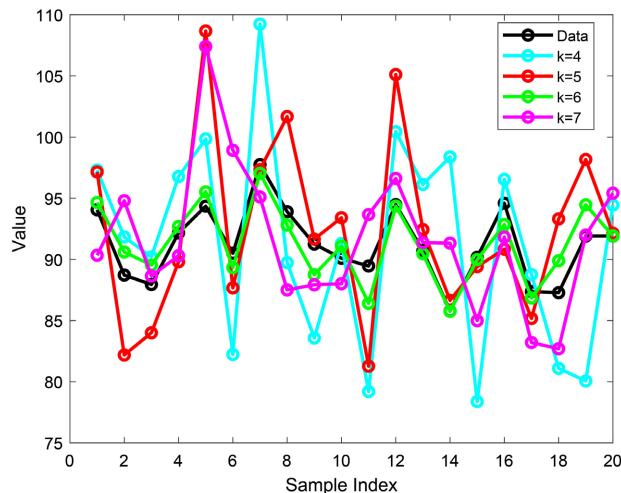


Figure 8. The selection of different K values
图 8. 不同 K 值的选择情况

3.1.3. 极差模型的应用

前文中我们对极差模型进行了研究，下面我们将建立的模型进行在同类比赛中的推广与应用，以此来确定最优步长的可靠性，以 2023 年上半年某大型创新类比赛为例选取适量样本进行不同调整步长的比较。

从图 8 中可以看出不同的 k 值选择与实际的情况相比较的结果是不一样的，随着步长的增加误差也发生了一定程度的变化，但当 $k = 6$ 时模型的误差最小大约为 5%，达到了最佳的效果，这与前文的结论相符合，因此我们认为选择 6 最为合适的调整步长是符合这类大型创新类比赛的。

4. 总结

对于创新类比赛而言我们首先通过数学分析的方法对具有一定数量的作品集和专家集按照彼此间“交集”进行分析，建立了线性规划模型用于对专家和作品进行分类提高彼此间的关联性；其次总结初步标准分模型并引入样本总体均值和反差对初始模型进行改进，使得专家们在评审作品时的评分标准得到了统一，使得评审的结果更具有客观性；最后由于在评审过程中会出现极差较大的现象，这种情况往往会在作品数量很多的时候，我们引入了极差模型，对作品的分布进行研究，确定了在前 30% 的作品中进行极差的确定，最终确定步长为 6 作为本文中模型的最优步长，以此作为极差调整的参考，判断作品是否需要进行下一阶段的复议，并运用数据进行验证，说明了方案具有较好的适应性。

参考文献

- [1] 习近平：决胜全面建成小康社会夺取新时代中国特色社会主义伟大胜利——在中国共产党第十九次全国代表大会上的报告[M]. 北京：人民出版社，2017.
- [2] 李腾龙. 大学生参加创新创业类比赛的培训指导研究——以“互联网+”大学生创新创业大赛为例[J]. 现代商贸工业, 2019, 40(25): 75-77.
- [3] 饶柳, 王天兰, 唐梦雨, 等. “互联网+”环境下大学生创新创业大赛的开展[J]. 西部素质教育, 2023, 9(23): 104-107.
- [4] 钟智研, 樊奕君, 刘昊然. 基于线性规划的供应商最优策略模型研究[J]. 企业改革与管理, 2022(13): 29-31.
- [5] 夏圣亭. 关于离散系数的探讨[J]. 工科数学, 1997(2): 144-146.
- [6] 吴晓军, 刘伟军, 王天然, 等. 改进的基于欧氏距离测度网格模型体素化算法[J]. 计算机辅助设计与图形学学报, 2004, 16(4): 592-597.