

大规模创新类竞赛评审方案研究

现在创新类竞赛很多，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。创新类竞赛的特点是没有标准答案，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。事实上，当竞赛规模大，评委的人数众多时，极差（定义见附件 1）大的问题更为突出。显然，简单地依据多位评委评分的总和进行排序并不是创新类竞赛评审的好方案。因此，探讨大规模创新类竞赛评审方案的公正性、公平性和科学性具有深远意义。

目前，各项创新类竞赛都在摸索、调整自己的评审方案。现有方案包括：

(1) 对每位评审专家的评分进行标准化（公式见附件 1），按作品将标准分相加得每件作品总分，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 (1) 或 (2) 或 (3) 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。这些方案都有一定的合理性，但也有局限性。特别是针对大规模创新类竞赛评审，现有方案偏简单，研究不多。

在大规模创新类竞赛中，增加评审每份作品的专家人数，显然有利于评审工作的公正、公平。然而，由于种种原因，参与评审工作的专家数目是受限的。评审专家少了，评审工作的误差会变大。不过，考虑到大规模创新类竞赛获奖比例通常小于 50%，有些误差并不影响是否获奖。因此，在不影响获奖等级的前提下，为了适应评审专家人数较少的现状，不少竞赛采用两阶段评审办法。

为探索大规模创新类竞赛评审的好方法，附件给出模拟大规模创新类竞赛的数据。其包含两阶段评审，第一阶段由五位专家对作品评审，取标准分后，将五位专家的标准分取均值后排序，按事先约定的比例取排名在前的作品，进入第二阶段评审。第二阶段由三位专家对作品评审，分别取标准分，并对少数极差大作品的标准分进行必要的调整后，再将第一阶段五位专家评审标准分的均值、第二阶段三位专家的评审标准分共 4 份成绩求和，依最终总成绩再排序。请利用这批数据建立数学模型，探讨建立更为合理、公平的评审方案。

问题一 在每个评审阶段，作品通常都是随机分发的，每份作品需要多位评委独立评审。为了增加不同评审专家所给成绩之间的可比性，不同专家评审的作品集合之间应有一些交集。但有的交集大了，则必然有交集小了，则可比性变弱。请针对 3000 支参赛队和 125 位评审专家，每份作品由 5 位专家评审的情况，建立数学模型确定最优的“交叉分发”方案，并讨论该方案的有关指标（自己定义）和实施细节。

问题二 在评审中采用标准分（附件 1）为基础的排序方法，其假设是不同评审专家评审的作品集合的学术水平分布相同。但在大规模创新类竞赛评审中，通常任意两位专家评审的作品只有小部分是共同的，绝大多数作品是不同的

(见问题一)，而且每位专家只看到作品集合的很小部分，因此标准分评审方案的假设可能不成立，需要探索新的评审方案。请选择两种或两种以上现有或自己设计的评审方案和题目附件数据，分析每位专家、每份作品原始成绩、调整之后（如取标准分）成绩的分布特点，按不同方案进行排序，并设法比较这些方案的优劣。进而针对大规模创新类竞赛的评审，设计新的标准分（公式）计算模型。另外，一般认为经多位专家协商一致的获奖论文具有最大的可信度，附件 2 提供的数据 1，其第二评审阶段评选出的一等奖作品排序是经专家协商取得一致的，请利用这批数据，改进你们的标准分计算模型。

问题三 “创新类”大赛的特点是“创新性”，即没有标准答案。由于这类竞赛的问题难度较大，一般需要通过创新才能在竞赛期间部分解决。而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差异（极差）。极差大是大规模创新类竞赛的特点，极差比较大的作品一般处于高分段或低分段。低分段属于淘汰范围，低分段极差大的原因是专家对违规作品或有重大失误的作品给了很低的分数，或评审专家都认同该作品质量不高，只是其中某位（些）专家更不认同该作品。故这里极差虽大，但属于不获奖范畴，一般不需要调整极差。而高分段作品还要参加权威性较高的第二阶段评审（附件数据表格同一行代表同一个作品在两个阶段的成绩，没有第二阶段评审成绩的作品只参加了第一阶段的评审）。第二阶段评审仍然存在部分极差大的作品，因为是终审，误差可能影响获奖等级，因此对部分极差大的作品，需要复议调整极差（附件的数据中有记录，复议分就是该专家最后给的标准分，用来替换原来的标准分）。第二阶段（注意两个阶段每份作品评审专家人数不同）专家调整“大极差”的规律可以作为建立极差模型的借鉴。

请根据题目所给的模拟数据 2.1 和 2.2，讨论两阶段的成绩整体的变化和两阶段极差整体的变化，分析两阶段评审方案相比不分阶段评审方案的优劣。注意到极差大和创新性强两大特点之间会有一定的关系，为了发掘创新论文，请建立“极差”模型（含分析、分类、调整等），并针对所给数据，尝试给出第一评审阶段程序化（不需要人工干预）处理非高且非低分段作品的“大极差”的办法。

问题四 对“创新类”竞赛，给出一个完整的评审模型（提示：例如优化模型），并针对所给的数据研究如何求解？也可对现行的评审方案给出改进的具体建议（包括未来还要收集哪些数据）。

附件：

1. 极差的定义及标准分的计算方法
2. 数据 1：模拟某大型创新类竞赛的评审数据，其第二阶段被评选为一等奖作品的排序经专家协商取得一致。
3. 数据 2：模拟某大型创新类竞赛两阶段的评审数据，有两组 2.1 和 2.2。