2025年太原科技大学研究生数学建模竞赛C题

**“人工智能创新应用类”竞赛评审方案研究**

当下，人工智能领域的创新应用类竞赛数量繁多。其中，具有较大影响力的竞赛，其评审流程通常分为两阶段（网评、现场评审）或者三阶段（网评、现场评审、答辩）。创新类竞赛的一大特性在于不存在固定的标准答案，需要评审专家依据命题人（组）给出的评审框架（建议）展开独立评审。如此一来，针对同一作品，不同评委给出的评分可能会有显著差别。实际上，当竞赛规模庞大且评委人数众多时，评分极差（具体定义见附件 1）较大的问题会愈发凸显。很明显，仅仅依靠多位评委评分的总和来对作品进行排序，并非创新类竞赛评审的理想方式。所以，对大规模创新类竞赛评审方案的公正性、公平性以及科学性展开探究，有着十分深远的意义。

目前，各个人工智能创新应用类竞赛均处于评审方案的探索与调整阶段。现存方案主要包含：（1）依照特定公式（详见附件 1）对每位评审专家的评分实施标准化处理，以作品为单位将标准分累加得出每件作品的总分，进而依据总分进行排序；（2）剔除同一份作品得分中的最高值与最低值，对剩余评分进行求和，最终按照总分排序；（3）若同一份作品的专家评分差异（极差）超出一定范围，则组织相关专家开展协商调整，以调整后的得分进行累加并排序；（4）当竞赛规模较大时，先采用上述方案（1）、（2）或（3）对作品进行初步筛选，再针对初选入围作品开展第二阶段专家评审或设置答辩环节以确定获奖名单。这些方案虽具备一定合理性，但也存在局限性。尤其是针对大规模创新类竞赛的评审需求，现有方案普遍存在设计偏简单、系统性研究不足的问题。

在人工智能创新应用类竞赛里，增加评审每份作品的专家数量，无疑对评审工作的公正、公平性有积极作用。但受多种因素制约，参与评审的专家人数存在限制。若评审专家数量不足，评审工作的误差可能会增大。不过，鉴于大规模创新类竞赛的获奖比例通常低于 50%，部分误差并不会对作品是否获奖产生影响。所以，在不影响获奖等级判定的前提下，为契合评审专家人数较少的实际情况，许多竞赛采用了两阶段评审方式。

为探寻适用于大规模人工智能创新应用类竞赛的优质评审方法，附件提供了模拟大规模创新类竞赛的数据。该数据涵盖两阶段评审流程：第一阶段由五位专家对作品进行评审，计算标准分后取均值并排序，按预先设定比例选取排名靠前的作品进入第二阶段；第二阶段由三位专家评审并计算标准分，针对少数评分极差较大的作品进行必要的标准分调整，最终将第一阶段五位专家的标准分均值、第二阶段三位专家的标准分共 4 项成绩求和，依据总成绩再次排序。请结合这批数据构建数学模型，研究如何制定更合理、公平的评审方案。

**问题一** 在大规模竞赛评审过程中，为确保评审结果的科学性与公正性，作品往往采用随机分配机制，且每份作品均需多位评审专家独立评分。为提升不同专家评分的可对比性，专家所评审的作品集合需存在一定交集，但作品交集分布不均会引发新问题 —— 部分交集过大，部分交集过小，反而削弱评分的可比性。现针对 3000 支参赛队伍、125 位评审专家，且每份作品由 5 位专家评审的具体场景，要求构建数学模型，以求解实现评分可比最大化的作品交叉分配策略，并对该策略的评价指标（需自主构建）及实际操作流程展开深入探讨。

**问题二** 基于标准分（见附件 1）的评审排序方法，依赖于 “不同专家所评作品学术水平分布一致” 的前提条件。然而，在大规模人工智能创新应用类竞赛中，由于评审规模庞大，实际情况与假设存在显著差异：一方面，任意两位专家所评作品仅有少量重叠，大部分作品相互独立；另一方面，每位专家仅能接触全部作品中的极小部分。这使得传统标准分评审方案的理论基础不再可靠，亟需研究替代性的评审策略。请从现有评审方案或自主设计方案中选取至少两种，结合题目附件中的数据，系统分析每位专家评分、作品原始成绩以及标准化处理后成绩的分布特征。通过运用不同方案对作品进行排序，建立科学的评价体系，对比各方案的优缺点。此外，针对大规模创新竞赛评审的特殊性，需重新构建标准分计算模型。值得注意的是，专家协商一致的评审结果通常被视为最具权威性。附件 2 中的数据 1 记录了第二评审阶段一等奖作品的协商排序结果。请充分利用这一权威数据，对新构建的标准分计算模型进行优化与验证，提升模型的准确性与实用性。

**问题三** “人工智能创新应用类竞赛类”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有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“人工智能创新应用类”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。