

科技计划项目评审专家精准评价模型研究

——以江苏省科技计划项目评审为例

李 彬¹, 李旭红¹, 应 媚²

(1.江苏省生产力促进中心,江苏 南京 210042)

(2.江苏省科学技术发展战略研究院,江苏 南京 210042)

[摘要] 评审专家是科技计划项目评审工作的关键核心,其专业能力水平、诚信等方面因素对保证评审质量至关重要. 本文以江苏省科技计划项目评审为例,研究借助大数据分析技术,探索构建了一套覆盖客观和主观要素多个维度的评审专家智能评价模型,能够精准绘制专家全息画像,对优化现有专家评价模式以及实现专家精细化管理具有一定的参考价值.

[关键词] 江苏,科技计划项目,专家画像,智能评价,科研诚信

[中图分类号] G311 [文献标志码] A [文章编号] 1001-4616(2024)01-0133-09

Research on the Accurate Evaluation Model of Experts in Science and Technology Plan Project Review: Taking Jiangsu Province's Science and Technology Plan Project Review as an Example

Li Bin¹, Li Xuhong¹, Ying Mei²

(1.Productivity Center of Jiangsu Province, Nanjing 210042, China)

2.Jiangsu Academy of Science and Technology for Development, Nanjing 210042, China)

Abstract: The review experts are the key core of the review work for science and technology plan projects, and their professional competence, integrity, and other factors are crucial to ensuring the quality of the review results. This article takes the evaluation of science and technology projects in Jiangsu Province as an example, and explores the construction of an intelligent evaluation model for evaluation experts covering multiple dimensions of objective and subjective factors using big data analysis technology. It can accurately draw expert holograms and has certain reference value for optimizing existing expert evaluation models and achieving refined expert management.

Key words: Jiangsu, science and technology plan project, expert portrait, intelligent evaluation, research integrity

科技计划项目同行评审是配置科技资源的重要依据,关系到科研人员和机构的切身利益,一直受到科技界的高度关注^[1-3]. 评审专家是科技计划项目评审工作的关键核心,其业务能力、职业道德等要素均会对项目评审结果产生直接影响^[4-6]. 江苏省科技计划项目每年的受理和评审量均达到 10 000 多项,且在逐年增加(详见图 1),是江苏省支持和引导科技创新的重要途径,为确保项目评审专家遴选工

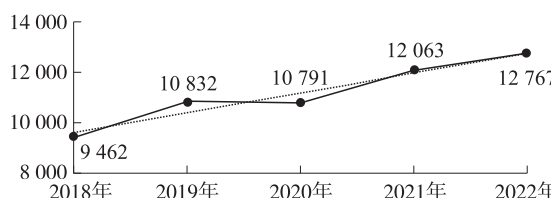


图 1 2018-2022 年项目申报数量走势图

Fig. 1 Trend chart of project application quantity
from 2018 to 2022

收稿日期:2023-09-26.

基金项目:江苏省社科应用研究精品工程科技伦理(科研诚信)专项课题项目(23SLB-12).

通讯作者:李彬,项目主管,研究方向:科技项目管理、数据分析. E-mail:025bin@163.com

作的科学性、公正性,减少人为主观因素干扰,项目管理专业机构已经开始采用通过计算机系统自动随机抽取的方式遴选评审专家,以后将继续探索“一键评审”的智能化评审模式,实现遴选、邀请等工作全程无需工作人员干预的目标,但是这些智能化的遴选方式,必须有精准的、量化的专家评价数据作为支撑,传统的专家评价模式远远达不到要求,急需利用先进的数据分析技术,突破现有评价体系瓶颈,探索构建更为科学有效的评审专家评价模型。

1 现状分析

1.1 存在问题

评审专家管理普遍缺少有效的考核评价手段,导致存在小部分专家的职业素养不够高、工作态度随意、科研信用不良等现象,在一定程度上已经影响到评审工作和评审结果质量^[7-9]。截至 2022 年底江苏省科技咨询专家库中已经汇集了各行业领域专家超过 50 000 名,其中核心专家达到 9 000 余名,为优化专家资源,提升项目评审质量,虽已经开始探索对专家进行考核评价,但现有评价方式主要存在如下三点不足:一是评价指标内容过于侧重对专家的学历、工作单位、职务、职称、获奖、头衔、履历等面上信息的评价;二是对评审专家的评价方式基本上是采用人工依照评价指标逐项对评审专家进行分级分类,存在工作量大、操作效率低、更新周期长等短板,无法保证评价结果的时效性和可信度;三是无法对评审专家工作质量进行量化,导致出现奖惩不分明,评审专家干好干坏一个样的状况,在一定程度上挫伤了专家工作积极性和影响了项目评审质量。

1.2 研究现状

经文献检索发现,目前已有专家学者开始了针对项目评审专家评价方面的研究,如:杨健安等^[10]利用相对逆序数概念和归一化方法,通过分析研究评审专家对项目评价结果的不同排序差异,提出了专家评审可信度的评价模型,研究成果适用于排序导向型的评审场景;杨晓秋等^[11]利用逆向云算法,通过分析研究专家评审数据的异常程度,提出了一种用于识别专家评议异常数据的检测方法;张尧等^[12]研究利用整体网分析方法估计专家的学术圈地位以及学术水平,以及通过对专家评审资助命中率、评分一致性等方面研究,形成了对专家专业水平的评价方法;汪建等^[13]基于元评价理论和实证数据分析,研究提出了科技项目评审专家评价模型;李旭彦等^[14]利用高斯分布函数计算以及 PageRank 算法,通过对评审结果矩阵进行标准差归一化、偏离度分析等研究,提出了度量评审专家信誉度的方法;黄进^[15]基于“能力-行为-绩效”的视角,研究了科学基金项目评审专家科研信誉综合评价体系;蔡文学等^[16]通过对不同计划项目类别评审结果的偏差度进行分析研究,提出了基于偏差度权重的评审专家绩效评价方法;张金焕^[17]通过对项目的争议性和评审结果的偏移性进行分析研究,提出了异常数据的修正模型及算法;徐林生等^[18]研究了评价和决策活动中专家个体可信度的 3 种建模方法,提出了对专家个体与群体评审能力水平排序和分类的方法;李樵等^[19]通过研究评审专家在心理层面对与评审对象有关的信息进行加工时产生的认知偏差,并以双重加工理论为基础,提出了干预评审人认知加工的技术、管理和制度策略。

综上,当前对科技计划项目评审专家的评价方式研究各有特点,主要分为两类:一是侧重对评审专家的专业水平评价,如杨健安、杨晓秋、张尧、汪建、李旭彦、蔡文学、张金焕、徐林生等人的研究;二是侧重对专家评价体系的宏观理论研究,如黄进和李樵等人的研究。上述研究对于突破现有专家评价模式瓶颈均具有重要意义,但在实践应用方面略有不足,为此本文将在借鉴上述研究思路的基础上,结合江苏省科技计划项目评审工作特点,面向实操应用场景,构建一套科学有效、实用性强的项目评审专家评价模型。

2 指标构成

本模型设计遵循科学有效、数据可用、操作简便的原则,指标设置既要覆盖客观数据也要兼顾主观评价,相关数据能够从江苏省科技计划项目现有评审工作体系中获取。综合考虑后,本评价模型将重点从评审专家的专业水平、信用状况、评审态度以及承担科技计划项目能力四个维度进行研究,模型架构如图 2。

2.1 专业水平评价

项目评审专家具备扎实的专业知识和技能,是保证评审质量最基本的条件,需要综合考虑专家的专业背景、从事工作、实践经验以及参与项目评审活动的评价等因素^[20]。本模型对专家的专业水平评价重点从

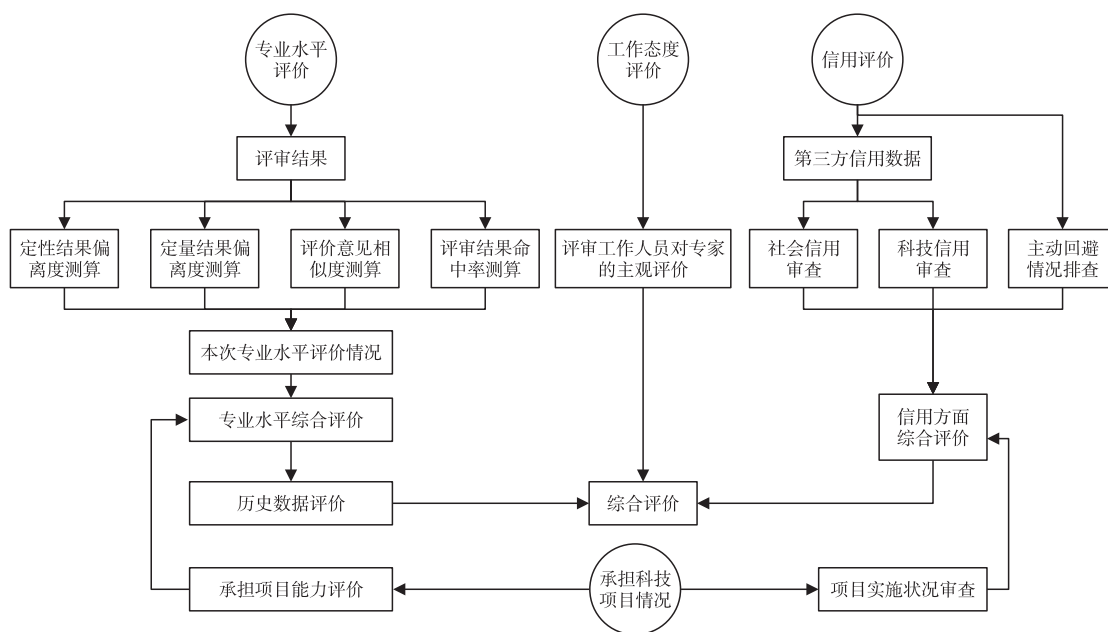


图2 模型架构图

Fig.2 Model architecture diagram

评审项目的定性结果偏离度、定量结果偏离度、评价意见相似度、评审结果命中率四个维度进行测算。

(1)定性结果偏离度. 定性结果偏离度反应被评价专家在评审中较同组其他专家定性评价结果的偏离程度,偏离值越大,在一定程度上说明被评价专家的专业能力越弱. 设评审组有项目 k 项,某评审专家对该组项目的定性评价异常情况(例如某项目的定性评价结果是 $1A4B$,该专家对该项目的定性评价是 A ,或某项目的定性评价结果是 $4A1B$,该专家对该项目的定性评价是 B)有 n 次,则该专家在此次评审工作中的定性结果偏离度 P 测算方式为:

$$P = \frac{n}{k}. \quad (1)$$

(2)定量结果偏离度. 定量结果偏离度反应被评价专家在评审中较同组其他专家定量评价结果的偏离程度,偏离值越大,在一定程度上说明被评价专家的专业能力越弱. 设评审组有项目 k 项,评审专家 s 名,项目定量得分为 f ,评审专家 a 在本次评审中的定量结果偏离度 L 测算方式为:

$$L = \frac{\sum_{i=1}^k \left(\left| f(a,i) - \frac{\sum_{j=1}^s f(j,i)}{s} \right| \div \sum_{j=1}^s f(j,i) \right)}{k}, \quad (2)$$

其中 $\frac{\sum_{j=1}^s f(j,i)}{s}$ 代表某个项目的专家组平均分,如测算结果 $L > 1$,则 $L = 1$.

(3)评价意见相似度. 科技计划项目评审过程中,需要评审专家指出推荐项目的亮点和进一步加强的方面,以便作为立项依据,同时亦要指出不推荐项目存在的问题和改进建议,以便项目团队后续改进提升,所以专家意见的质量至关重要,是评审工作的重要成果. 评价意见相似度反应被评价专家在评审中对所评项目文字意见的相似程度,本项重点是发现专家评审时给被评项目的意见存在“创新性一般,不建议支持”“创新性好,建议支持”等雷同情况,或者部分网评专家采用粘贴复制的方式录入评审意见导致出现多个项目的评审意见相似或完全相同的情况,失去了专家意见的实际意义,这种情况能反映出评审专家的专业能力和对待评审工作的态度.

本模型中对专家意见文本相似度测算采用经典的计算文本间的余弦值,也叫余弦相似度,即通过计算一个向量空间中两个向量夹角的余弦值作为衡量两个个体之间差异的大小,所有事物的相似度范围都应

该是 0~1,余弦相似度越接近 1,表明两个向量越相似. 专家评价意见相似度只针对文本字面测算,即只判断两段意见字符的相似程度,陈观林等^[21]提出的基于深度强化学习的文本相似语义计算模型以及钟桂凤等^[22]提出的基于差分进化的卷积神经网络的文本分类方法,也为深度分析专家评审意见提供了非常好的路径,我们将在后续研究中进行探索应用. 设专家在某次评审中对其中两个项目的评价意见内容分别为 A 和 B ,通过中文分词算法把两段文字分成独立的词集合,并得到两个词集合的并集,分别计算各自词集的词频^[23],并把词频向量化,即 $A=(a_1,a_2,a_3,a_4,\cdots,a_n)$ 和 $B=(b_1,b_2,b_3,b_4,\cdots,b_n)$,那么 A 和 B 的文本相似度 $\cos \theta$ 测算模型为:

$$\cos \theta(A, B) = \frac{\sum_{i=1}^n (A_i \times B_i)}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (3)$$

假设评审组有项目 k 项,那么被评价专家在本次评审中会产生 k 段评审意见,那么被评价专家在本次评审工作中的评审意见相似度 S 的测算模型为:

$$S = \frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^k \cos \theta(i, j)}{\frac{k!}{2! (k-2)!}}, \quad (4)$$

其中 $\frac{k!}{2! (k-2)!}$ 是被评价专家在本次评审中对所有项目评审意见不考虑顺序的两两组合数.

(4) 评审结果命中率. 评审结果命中率是被评价专家在某次项目评审中推荐的项目最终获得立项支持的比例,命中率越高,则说明该专家的专业水平越高. 设被评价专家在某次评审工作中推荐项目数为 k 项,经过项目立项、公示等环节后,最终确定立项支持的项目数为 x 项,则该专家在本次评审中的命中率 Z 测算模型为:

$$Z = \frac{x}{k}, \quad (5)$$

(5) 专业水平综合评价. 评审专家专业水平综合评价通过定性评价偏离度(P)、定量评价偏离度(L)、评价意见相似度(S)、评审结果命中率(Z)四个维度进行综合测算. 设被评价专家已累计参加了 n 场评审活动,专业水平综合测评模型为:

$$ZY1 = \left\{ \sum_{i=1}^n \left[\frac{(1 - P_i) + (1 - L_i) + (1 - S_i) + Z_i}{4} \right] \right\} \div n. \quad (6)$$

2.2 信用评价

科研诚信是科技创新的基石^[24]. 同行评议活动涉及到科学资源、科学信誉等各种利益的分配,因而利益冲突以及引起的伦理问题更加突出^[25]. 科技计划项目评审过程中,公平公正是最基本的要求,但因评审专家具有较大裁量权,其道德素养、社会关系、研究倾向等方面的信用因素为评审工作的公平公正性带来了诸多不确定性,所以加强对评审专家的信用评价,积极构建“负责任、讲信誉、计贡献”(简称 RCC)的评审机制^[26],对于提高同行评审公平信度具有重要意义. 2022 年 12 月江苏省科技厅印发了《江苏省科技计划项目信用管理办法》,进一步强化了对包括科技计划项目评审专家在内的各科研主体的信用管理.

(1) 信用情况评价. 专家是评审工作的核心,而信用则是专家的核心,是保证专家评审工作质量的重要前提条件,关系到项目评审结果的公平、公正,所以良好的信用记录情况是优选专家的关键. 该项评价重点关注 2 个方面:一是社会信用状况,即评审专家在社会活动中是否存在失信行为,包括法律法规、学术、道德等方面. 二是科技信用状况,重点是评审专家在作为项目负责人承担各类项目过程中是否存在失信行为.

设该项评价指标为 α ,评价方式为:

$$\alpha = \begin{cases} 1, & \text{无失信记录} \\ 0.6, & \text{一般失信} \\ 0, & \text{严重失信} \end{cases} \quad (7)$$

由工作人员依据江苏省科技咨询专家库有关管理规定进行选择配置,默认值为无失信记录.对于有失信记录的专家,待信用修复后,可由工作人员根据情况进行调整配置.

(2)主动回避情况审查.遴选专家回避制度,直接将与被评价主体有特殊关系的相关评价专家排除在评价活动之外,以此从形式上保证了评价活动的中立性^[27].科技计划项目评审在专家遴选环节,工作人员或抽取系统必须要对待选专家进行回避操作(详见图3),但因受到信息资源有限、评审准备时间仓促等客观因素的影响,对于专家和待审项目之间存在间接的、隐含的可能影响评审结果客观性和公正性的关联关系,在遴选阶段很难做到100%回避,这些情况只能希望评审专家在评审过程中主动提出回避.

对于评审专家应当回避但是没有提出回避的情况,我们可以通过分析专家对评审项目的打分是否存在明显过高或过低的情况发现线索.假设评审组中专家 a 对该组项目的定量评价平均分为 F_p ,对项目 i 的定量评价分为 $f(a,i)$,则对项目 i 的倾向值为:

$$Q(a,i)=\frac{f(a,i)-F_p}{F_p}. \quad (8)$$

暂定将 $|Q|>0.2$ 设定为异常阈值,当评审专家对某个项目的倾向值超过预设阈值后,可以提醒工作人员进一步调查,如确认无异常,则可以忽略该提醒信息,如发现评审专家与该项目存在关联关系,且未主动提出申请回避,则要扣该专家的信用分值.如在实际使用过程中发现该算法误报情况较多,可以适当调高 Q 阈值,直至相对理想的状态.

设该项信用评价指标为 β ,测算方式为:

$$\beta=\begin{cases} 1, & \text{无异常} \\ 0, & \text{应回避未回避} \end{cases} \quad (9)$$

由工作人员根据调查情况进行选择,模型对所有专家的默认值为正常.

(3)评审意见关键词审查.评审专家需要按照评审要求对被评价项目进行评议,评审意见要客观反映项目的实际情况,应避免出现过于主观、极端的评价内容.针对专家评审意见中可能出现的刺激性词语(如带有人身攻击性质的词语)进行收集整理,形成关键词库,以关键词比对的方式对专家评审意见进行审查,保障评审专家能够对被评项目进行客观、公正的评价.

该方面信用评价指标为 Ω ,测算方式为:

$$\Omega=\begin{cases} 1, & \text{发现存在刺激性词语} \\ 0, & \text{未发现存在刺激性词语} \end{cases} \quad (10)$$

本项工作由模型自动比对,如发现异常,则提醒工作人员作进一步确认处置.

(4)信用综合评价.评审专家信用综合评价得分 $XY=\alpha\times\beta\times\Omega$,信用评价采用“零容忍”模式,即只要有1项信用指标得0分,则信用方面综合评价得0分.

2.3 评审态度评价

项目评审是一项集体工作,需要多方密切协作配合推动,评审专家作为评审活动中的核心角色,其工作态度对保障评审工作顺利开展至关重要.评审活动过程中,专家需要严格遵守评审活动中的各项规定,避免影响评审工作秩序以及对其他评审专家造成干扰,以保证评审结果质量.评审态度评价用于评审工作人员根据在评审过程中专家的具体表现情况对其进行评价,主要包括参加评审活动时主动配合工作人员进行身份核验以及通讯工具集中保管、遵守会场纪律、遵守评审时间规定、不得擅自离开评审区域、不将评审资料带出评审场所等情况.

设该项评价指标为 TD ,测算方式为:

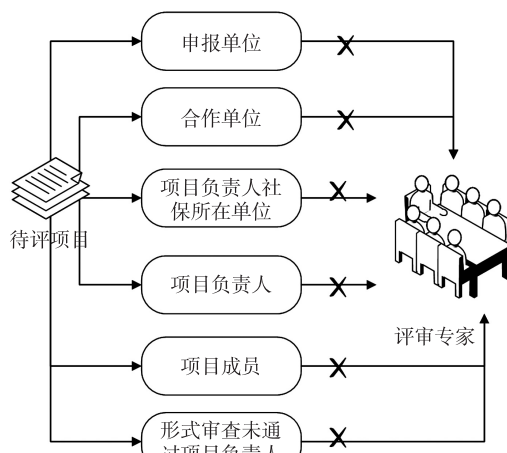


图3 专家回避要素

Fig.3 Elements of expert avoidance

$$TD=\begin{cases} 1.1\text{ 分}, & \text{优秀} \\ 1.0\text{ 分}, & \text{良好} \\ 0.7\text{ 分}, & \text{一般} \\ 0\text{ 分}, & \text{较差} \end{cases} \quad (11)$$

本模型对所有专家默认为“良好”档次,只有在评审过程中表现非常好的专家,工作人员可将其设置为“优秀”档次,对于存在评审态度恶劣且不服从工作人员管理等情况的专家,工作人员可将其设置为“较差”档次,专家抽取系统不再将其纳入遴选范围。

2.4 承担科技计划项目情况

该模型借助数据接口可实现与江苏省科技计划项目管理系统的交互,实时获得评审专家承担江苏省科技计划项目的数据资料,借助这些数据信息可进一步评价专家的专业水平和审查专家的科技信用情况。

(1)承担项目能力评价. 如评审专家作为项目负责人承担过江苏省科技计划项目,在一定程度上能够说明该专家对江苏省科技计划项目的定位、实施要求等方面有较为全面的理解,更加有利于从事项目评审工作,模型以其承担过的项目经费总额(JF)作为参考依据,适当提高该评审专家的专业水平综合评价得分,测算方式如下:

$$ZY2(ZY1)=\begin{cases} ZY1\times1.3, & JF\geq 1\,000\text{ 万元} \\ ZY1\times1.2, & 1\,000\text{ 万元}>JF\geq 500\text{ 万元} \\ ZY1\times1.1, & JF<500\text{ 万元} \end{cases} \quad (12)$$

(2)项目实施状况审查. 如发现评审专家作为项目负责人承担的江苏省科技计划项目存在强制终止、经费违规使用等严重科技失信情况,模型将采取“零容忍”^[28]的处理方式,直接将该评审专家的信用综合评价得分 XY 设置为 0。

2.5 历史数据评价

依据 2018–2020 年度江苏省科技计划项目管理系统中基础研究计划(自然科学基金)、省重点研发计划、省科技成果转化专项资金等共计 100000 多条网上评审记录,对涉及的 4300 多名评审专家的定性评价结果偏离度进行统计,计算得到平均值为 0.0628,标准差为 0.0653,详细情况见表 1,图 4。

表 1 2018–2020 年评审专家定性评价结果偏离度统计(部分)

Table 1 Statistical deviation of qualitative evaluation results of evaluation experts from 2018 to 2020(partial)

| 专家姓名 | 评审项目数 | 推荐异常数 | 不推荐异常数 | 异常比值 | 偏离度 | 偏离度绝对值 -1 个标准差 | 偏离度绝对值 -2 个标准差 |
|-------|-------|-------|--------|---------|----------|-------------------|-------------------|
| 专家 1 | 13 | 1 | 0 | 0.076 9 | 0.014 1 | -0.051 1 | -0.116 4 |
| 专家 2 | 12 | 1 | 0 | 0.083 3 | 0.020 6 | -0.044 7 | -0.110 0 |
| 专家 3 | 32 | 1 | 0 | 0.031 3 | -0.031 5 | -0.033 7 | -0.099 0 |
| 专家 4 | 12 | 1 | 1 | 0.166 7 | 0.103 9 | 0.038 6 | -0.026 6 |
| 专家 5 | 74 | 1 | 0 | 0.013 5 | -0.049 3 | -0.016 0 | -0.081 3 |
| 专家 6 | 9 | 0 | 2 | 0.222 2 | 0.159 4 | 0.094 2 | 0.028 9 |
| 专家 7 | 9 | 0 | 1 | 0.111 1 | 0.048 3 | -0.016 9 | -0.082 2 |
| 专家 8 | 9 | 0 | 1 | 0.111 1 | 0.048 3 | -0.016 9 | -0.082 2 |
| 专家 9 | 12 | 0 | 0 | 0.000 0 | -0.062 8 | -0.002 5 | -0.067 8 |
| 专家 10 | 24 | 0 | 2 | 0.083 3 | 0.020 6 | -0.044 7 | -0.110 0 |

专家评审受到各种主观客观因素的影响,导致评审质量存在波动的现象,所以对专家单次评审情况的分析研究存在一定局限性,需要从历史数据的维度对专家定性评价结果偏离度进行分析,更能准确地评价专家的专业水平,据此对评审专家的专业水平测评得分进行修订,具体如下:

$$ZY3(ZY2)=\begin{cases} ZY2\times0.7, & bz>BZ+\sigma \\ ZY2\times0.9, & BZ<bz\leq BZ+\sigma \\ ZY2\times1.0, & bz\leq BZ \end{cases} \quad (13)$$

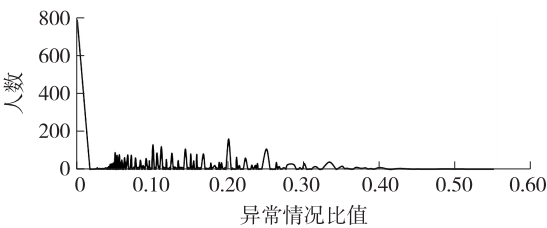


图 4 定性评价结果异常情况比值分布图
Fig. 4 Ratio distribution map of abnormal situations in qualitative evaluation results

其中: b_z 为某专家定性评价结果异常比值, BZ 为所有专家定性评价结果异常比值平均值(0.062 8), σ 为异常比值的标准差(0.065 3)。

2.6 综合得分

综合得分是对评审专家的综合量化评价,是对评审专家分级、分类、遴选、淘汰的重要依据,由专业水平综合评价、信用综合评价、工作态度评价三方面综合测算,综合得分 ZH 测算方式是:

$$ZH = ZY3 \times XY \times TD \times 100 \quad (14)$$

按照综合得分情况,划分评审专家所属等级(DJ),模型可实现对评审专家的自动分级,划分方式如下:

$$DJ(ZH) = \begin{cases} A \text{ 级,} & ZH \geq 80 \\ B \text{ 级,} & 80 > ZH \geq 60 \\ C \text{ 级,} & ZH < 60 \end{cases} \quad (15)$$

在实际应用场景中,对于每次评审活动的评价结果可以保留,避免每次更新评价数据时都要全部重新计算,以提高模型计算效率。

评审专家的综合得分并不是固定不变的,随着评审专家参与江苏省科技计划项目评审次数的增加,对专家的评价结果是动态变化的。随着江苏省科技计划项目管理系统与其他数据系统建立更多共享接口,本模型会探索将更多数据纳入对评审专家的评价范围,尤其是评审专家发表的论文数量和质量数据信息,也是体现评审专家学术水平的重要指标^[29]。

3 测算演示

本文以某次评审为例,随机选择其中一组评审结果演示该模型运行机理。该评审组共有 3 个项目,由 5 名专家分别进行评审,评审定量定性结果如下:

表 2 某组项目评审结果示例

Table 2 Example of review results for a group of projects

| 评审 项目 | 专家 1 | | 专家 2 | | 专家 3 | | 专家 4 | | 专家 5 | | 专家组 平均分 | 定性综 合评价 |
|----------|------|----|------|----|------|----|------|----|------|----|------------|------------|
| | 定量 | 定性 | 定量 | 定性 | 定量 | 定性 | 定量 | 定性 | 定量 | 定性 | | |
| 项目 1 | 60 | B | 84 | A | 90 | A | 93 | A | 84 | A | 82.2 | 4A1B |
| 项目 2 | 88 | A | 70 | B | 78 | B | 48 | B | 75 | B | 71.8 | 1A4B |
| 项目 3 | 91 | A | 83 | A | 92 | A | 93 | A | 85 | A | 88.8 | 5A0B |

本文以评审专家 1 为评价对象,具体测算步骤如下:

步骤 1:测算定性结果偏离度,该专家对项目定性评价结果与其他专家截然不同的情况有 2 次,依据公式 1 该指标测算结果为:

$$P = 2 \div 3 \approx 0.667. \quad (16)$$

步骤 2:测算定量结果偏离度,从表 2 可以得到该专家对每个项目的定量评价得分以及每个项目专家组定量平均得分,依据公式 2 该指标测算结果为:

$$L = \frac{(|60-82.2| \div 82.2 + |88-71.8| \div 71.8 + |91-88.8| \div 88.8)}{3} \approx 0.174. \quad (17)$$

步骤 3:测算评审意见相似度,为方便演示计算,本文将该专家对 3 个项目的评审意见简化为:“该项目创新性一般,不建议资助”“该项目创新性好,建议资助”“该项目技术先进,前景可期,项目方案可行”。经分词后形成词合集的并集(该,项目,创新性,一般,好,不,建议,资助,技术,先进,前景,可期,方案,可行),统计各段意见的词频后形成各段评审意见对应的向量分别为:

$$a(1,1,1,1,0,1,1,1,0,0,0,0,0,0), b(1,1,1,0,1,0,1,1,0,0,0,0,0,0), c(1,2,0,0,0,0,0,0,1,1,1,1,1,1)$$

三段评价意见不考虑顺序的两两组合一共有 3 种,即只要计算向量 a 和 b 、向量 a 和 c 以及向量 b 和 c 3 组向量之间的余弦夹角值即可。依据公式 3 计算向量 a 和向量 b 之间的余弦夹角为:

$$\cos \theta(a, b) = \frac{\sum_{i=1}^{14} (a_i \times b_i)}{\sqrt{\sum_{i=1}^{14} (a_i)^2} \times \sqrt{\sum_{i=1}^{14} (b_i)^2}} \approx 0.77, \quad (18)$$

以此类推分别计算出 $\cos \theta(a, c) \approx 0.34, \cos \theta(b, c) \approx 0.37$, 依据公式 4 测算该专家在此次评审中的评审意见相似度为:

$$S = \frac{\cos \theta(a, b) + \cos \theta(a, c) + \cos \theta(b, c)}{3} \approx 0.49. \quad (19)$$

步骤 4: 测算评审结果命中率, 本次评审中项目 1 和项目 3 最终获得了立项支持, 其中该专家命中了 1 项, 依据公式 5 测算他在本次评审的命中率 Z 为 0.33.

步骤 5: 专业水平综合评价, 综合上述测算情况, 该专家在本次评审活动中的专业水平评价得分为:

$$[(1-0.667) + (1-0.174) + (1-0.49) + 0.333] \div 4 \approx 0.5. \quad (20)$$

据统计, 专家 1 累计参加过 3 次江苏省科技计划项目评审活动, 前两次专业水平评价得分为 0.75 和 0.8, 依据公式 6 测算该专家在历次评审工作中的专业水平综合评价结果为:

$$ZY1 = (0.5 + 0.75 + 0.8) \div 3 \approx 0.683. \quad (21)$$

步骤 6: 信用方面评价, 经核对确认, 未发现该专家存在社会信用和科技信用方面的问题 ($\alpha = 1$); 主动回避情况排查方面, 依据公式 8 测算该专家对 3 个项目的倾向值分别为: 0.25、0.1、0.14, 对项目 1 的倾向值过高, 工作人员经多方确认后未发现异常, 可忽略该提醒, 主动回避情况指标正常 ($\beta = 1$); 经模型自动比对, 未发现评审意见中存在刺激性词语 ($\Omega = 1$), 则该专家信用综合评价得分 $XY = \alpha \times \beta \times \Omega = 1$.

步骤 7: 评审态度方面评价, 经向该专家参加过的 3 次评审活动现场工作人员了解, 该专家在评审工作态度方面良好 ($TD = 1$).

步骤 8: 承担科技计划项目情况方面评价, 经与江苏省科技计划项目数据库自动检索比对, 发现该评审专家承担过 2 项科技计划项目, 项目总经费为 600 万元, 项目均已正常验收, 按照公式 12, 则 $ZY2 = ZY1 \times 1.2 = 0.683 \times 1.2 \approx 0.82$.

步骤 9: 经对历史评价情况进行测算, 该专家的异常情况比值 $b_z = 0.0135$, 小于 $BZ(0.0628)$, 按照公式 13, 则 $ZY3 = ZY2 \times 1.0 = 0.82$.

步骤 10: 依据公式 14 测算该专家的综合评价得分为:

$$ZH = ZY3 \times XY \times TD \times 100 = 0.82 \times 1 \times 1 \times 100 = 82. \quad (22)$$

按照公式 15 的等级划分标准, 该评审专家等级属于 A 级. 用雷达图表示评价情况如图 5.

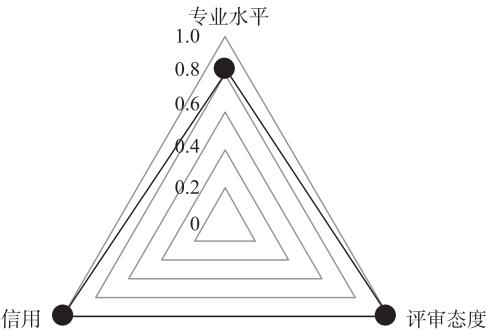


图 5 专家综合评价雷达图
Fig. 5 Expert comprehensive evaluation radar chart

4 结论

本文主要结论和创新性有: 一是可有效突破传统瓶颈, 有效解决了传统评价体系中无法量化评估专家的专业水平高低、评审工作质量高低、项目推荐是否客观公正等关键指标的瓶颈问题, 可显著提高评价结果的实际应用价值. 二是实现多维度立体评价, 采用了定量与定性相结合的方式设计指标体系, 既有基于客观信息的数学公式计算, 也有对日常评审工作表现的主观评价, 从多个维度评价专家的综合情况, 可提高评价结果的精准度. 三是实现智能化自动评价, 大部分测算可由计算机系统自动完成, 具有评价速度快、更新周期短、评价结果客观等特点, 可实现对江苏省科技计划项目评审专家的自动智能评价.

当前, 大数据、人工智能等新一代信息技术蓬勃发展, 驱动着科研管理模式快速变革^[30], 江苏省科技计划项目评审工作的组织将会更加规范化和智能化, 相关操作环节对专家评价数据的精准度、时效性要求越来越高, 本研究成果将对优化江苏省科技计划项目评审专家评价方式和提高评价效率效果具有“理论+实践”方面的参考价值, 同时对其他领域、其他地区优化评审专家管理亦有参考意义. 当然考核评价评审专家并不是最终目标, 而是期望运用科学有效的方法提高江苏省科技计划项目评审工作的质量, 所以在优化专家评价方式的同时, 更要配套制定和落实科学合理的激励机制, 强化对评审专家的正向激励, 给予其应有的认可^[31], 才能有效吸引更多优质行业专家加入江苏省科技咨询专家库, 才能为江苏省评选出更多优质的科技计划项目, 共同推动江苏科技创新高质量发展.

[参考文献]

- [1] 邵红琪,刘雪艳,张琪,等. 山东省科技计划项目评审改革现状及对策建议[J]. 中国科技资源导刊,2021,53(2):72-80.
- [2] GHORBANI A. A review of successful construction project managers' competencies and leadership profile[J]. Journal of rehabilitation in civil engineering,2023,11(1):76-95.
- [3] CHEN S,LIU J,ZHANG Q,et al. A critical review on deployment planning and risk analysis of carbon capture,utilization,and storage(CCUS)toward carbon neutrality[J]. Renewable and sustainable energy reviews,2022,167:112537.
- [4] 罗军,陈之瑶,莎薇,等. 科技计划项目评审专家工作质量元评价体系及应用研究——以广东省重点领域研发项目为例[J]. 科技管理研究,2021,41(13):65-70.
- [5] SAMPAIO S,WU Q,CORMICAN K,et al. Reach for the sky:Analysis of behavioral competencies linked to project success[J]. International journal of managing projects in business,2022,15(1):192-215.
- [6] HOU L,TAN Y,LUO W,et al. Towards a more extensive application of off-site construction:a technological review[J]. International journal of construction management,2022,22(11):2154-2165.
- [7] 江笑颜,李栋亮. 多层次、智能化的项目评审专家管理体系构建研究——基于广东科研项目评审专家管理的优化[J]. 科技管理研究,2018,38(16):212-217.
- [8] ELROD C C,STANLEY S M,CUDNEY E A,et al. Management information systems education:a systematic review[J]. Journal of information systems education,2022,33(4):357-370.
- [9] VAGALE A,OUICHEIKH R,BYE R T,et al. Path planning and collision avoidance for autonomous surface vehicles I: a review[J]. Journal of marine science and technology,2021:1-15.
- [10] 杨健安,刘传斌,余乐安. 基于相对逆序数的评审专家可信度评价方法[J]. 统计与决策,2022,38(3):184-188.
- [11] 杨晓秋,李旭彦. 同行评议中的异常数据检测方法研究——以科研项目评审为例[J]. 中国软科学,2016(5):133-142.
- [12] 张尧,林春. 专家评审特征分析及专家反评估方法研究[J]. 工业工程,2020,23(6):124-130+154.
- [13] 汪建,王裴裴,丁俊. 科技计划项目专家评审的元评价综合模型研究[J]. 科研管理,2020,41(2):183-192.
- [14] 李旭彦,宋英华,杨晓秋. 基于 PageRank 的评审专家信誉度量方法[J]. 科研管理,2016,37(3):133-142.
- [15] 黄进. 科学基金项目评审专家科研信誉综合评价体系研究:基于“能力-行为-绩效”的视角[J]. 中国科学基金,2012,26(1):57-60+64.
- [16] 蔡文学,温旖旎,郑顺潮,等. 基于偏差度权重的科技计划项目评审专家绩效评价方法研究[J]. 科技管理研究,2012,32(15):75-78,83.
- [17] 张金焕. 科技评价中专家评审粗差数据的处理及仿真[J]. 计算机仿真,2013,30(7):214-217+255.
- [18] 徐林生,王执铨,戴跃伟. 评审专家可信度评价模型及应用[J]. 南京理工大学学报(自然科学版),2010,34(1):30-34.
- [19] 李樵,王菲菲. 认知视角下创新性研究的同行评议:偏差与干预[J]. 科学学研究,2021,39(8):1354-1363.
- [20] 张琼丽,陈翼. 科技成果评价中专家评级模型研究[J]. 技术与市场,2022,29(9):136-138.
- [21] 陈观林,侍晓龙,周梁,等. 基于深度强化学习的文本相似语义计算模型[J]. 郑州大学学报(理学版),2020,52(3):1-8.
- [22] 钟桂凤,庞雄文,孙道宗. 基于差分进化的卷积神经网络的文本分类研究[J]. 南京师大学报(自然科学版),2022,45(1):136-141.
- [23] 武永亮,赵书良,李长镜,等. 基于 TF-IDF 和余弦相似度的文本分类方法[J]. 中文信息学报,2017,31(5):138-145.
- [24] 新华社. 中共中央办公厅 国务院办公厅印发《关于进一步加强科研诚信建设的若干意见》[EB/OL]. (2018-05-30) [2023-08-05]. https://www.gov.cn/zhengce/2018-05/30/content_5294886.htm.
- [25] 周颖,王蒲生. 同行评议中的利益冲突分析与治理对策[J]. 科学学研究,2003,21(3):298-302.
- [26] 吴刚,霍红,任之光,等. 管理科学部 RCC 评审机制试点效果分析[J]. 中国科学基金,2022,36(1):81-88.
- [27] 王贤慧. 探析同行评议中的学术评价回避制度[J]. 科技视界,2015(24):85.
- [28] 陈宜瑜. 健全学术不端惩戒机制 加强科研诚信建设 稳步推进科学基金监督工作[J]. 中国科学基金,2016,30(3):197-199.
- [29] 王大阜,邓志文,贾志勇,等. 多维属性视角下学者用户画像构建及合作学者推荐研究[J]. 南京师大学报(自然科学版),2023,46(3):112-122.
- [30] 李昱,赵静宜,左家平. 人工智能赋能科技管理变革的新趋向[J]. 科技智囊,2022(4):52-60.
- [31] 陈晓峰,蔡敬羽,刘永坚. 科技期刊同行评议中审稿人激励措施研究[J]. 中国科技期刊研究,2019,30(11):1157-1163.