

Penetrating the Fog of Bias in the Peer Review

Xuefeng Fu, Yafei Dai, Yan Huang, Lin Cui, Yongjun Chen*

Department of Chemical Sciences, National Natural Science Foundation of China, Beijing 100085, P. R. China.

Abstract: Peer review plays a crucial role in the fair and efficient allocation of high-quality research grants. This invited comment analyzes the fuzzy criteria and a possible bias in the peer review process, and explores the use of a potential approach based on fuzzy sets for tackling this challenge.

Key Words: NSFC; Peer review; Bias; Fuzzy sets

浅析基金评审的模糊属性

付雪峰, 戴亚飞, 黄艳, 崔琳, 陈拥军*

国家自然科学基金委员会化学科学部, 北京 100085

摘要: 同行评议是科学基金项目评审的重要环节。高质量的同行评议是实现基金精准资助的前提和保障。本文以国家自然科学基金通讯评审为例, 针对同行评议中出现的评审偏差, 探讨了基金评审的模糊属性, 提出了通过构建评审结果的模糊集合及一致性函数来解决该问题的可能性, 为进一步优化评审结果、实现基金精准资助指出了新方向。

关键词: 自然科学基金; 同行评议; 评审偏差; 模糊集合

1 引言

基础研究是整个科学体系的源头¹。国家对基础研究项目的组织、评审、管理等出台了一系列相关政策与规定。国务院印发的《关于全面加强基础科学研究的若干意见》明确提出了“完善符合基础研究规律的项目组织、申报、评审与决策机制”的要求²。国家自然科学基金委员会(以下简称基金委)是国家基础研究项目的重要组织部门, 贯彻“依靠专家、发扬民主、择优支持、公正合理”的方针, 秉持“公开、公平、公正”的评审原则, 采用同行评议的方式对项目申请进行评审。

高质量的同行评议对实现基金项目的精准资助具有重要意义³⁻⁸。然而, 评审过程中会出现同行评议结果与项目的真实质量不相符的情况。本文将国家自然科学基金项目的通讯评审(函评)为例, 针对同行评议结果与项目真实质量不相符的现象进行分析, 提出可能的解决方案。

2 评审偏差

国家自然科学基金项目在申请阶段的评审一般包括函评和会议评审(会评)两个环节。函评过程中, 每个项目均被分配给数名同行专家进行评审。评审专家针对项目的科学价值(Intellectual Merit)、影响力(Broader Impacts)、创新性(Originality)和可行性(Plausibility)进行评判。本文中, 把上述四个要素的总体完成情况称为项目的“真实质量”, 把评审专家对“真实质量”的评价称为“质量预期”。评审专家在不受非科学因素干扰的情况下, 理解所评项目的科学内涵、准确把握评审标准, 就能做出正确的“质量预期”评价, 给出合理的评审结果。

实际评审结果和理想评审预期之间有时会出现评审偏差(Bias in peer review)。评审偏差是指评审专家评审项目时, 受到各种因素干扰而未能对项目进行准确、客观、公正的评价⁹。评审偏差产生的原因很复杂¹⁰⁻¹², 可能因素包括: 受到项目申

Received: April 16, 2020; Revised: April 16, 2020; Accepted: April 16, 2020; Published online: April 20, 2020.

*Corresponding author. Email: chenyj@nsfc.gov.cn.

© Editorial office of *Acta Physico-Chimica Sinica*

请人学术地位或学缘关系等因素影响、自身科学鉴赏力不足、对项目研究内容存在个人偏好以及对评审标准把握不一等¹³。本文主要探讨由于对评审标准把握不一而产生的“质量预期误判”，从而导致评审偏差的问题。

基金委化学科学部(以下简称化学部)对历年重要人才类项目和研究类项目的评审与执行情况进行了大量的统计分析, 试图寻找评审偏差的来源, 提出有效的解决措施和办法, 完善项目评审, 提高项目遴选质量, 推动化学学科整体健康发展。通过统计分析发现, 不仅函评评价和会评评价之间存在偏差, 获资助项目的会评评价与结题评价(项目完成质量)之间也存在偏差。例如: 获资助项目会评评价一般, 结题评价很好; 获资助项目会评评价很好, 但结题评价一般。在排除项目负责人个体因素的情况下, 说明在项目函评和会评阶段, 确实有部分评审专家对项目的评价出现了评审偏差。

3 基金评审的模糊属性

从发布项目指南到项目结题验收, 国家自然科学基金项目管理过程涉及多个环节。2019年化学部在基金项目申请和评审过程中采取了“专家参与指派”、“三盲评审”等一系列新举措^{14,15}, 并提出了会评的“确认-纠偏-择优”机制¹⁶, 有效降低了由于项目申请人学术地位和学缘关系等因素导致的评审偏差。但是, 由“质量预期误判”导致的评审偏差问题仍然没有得到很好解决。

评审专家通过对项目申请真实质量的评价形成质量预期, 并对质量预期分别给出“好”或“不好”程度的判断, 再将这种判断归入“优、良、中、差”等级(对应4、3、2、1评分)。然而, 这些等级(或集合)之间的边界并不清晰。在评审过程中, 由于项目质量预期评价的复杂性和专家对真实质量认知的局限性, 评审专家在做出评价决策时可能会在项目等级归属问题上犹豫不决。例如, 质量预期“好”到什么程度是“优”? 或者, 比“优”差多少是“良”? 虽然每个等级的外在表现形式(评分值)是明确的, 但其内涵(质量预期的等级)却是不确定的、具有模糊属性(如图1所示)。基金项目评审是一种群体决策¹⁷过程, 单个评审专家对质量预期等级的模糊性可能会导致评审偏差或因叠加作用产生更大的评审偏差, 进而颠覆处于资助或不资助边缘区域项目的评审结果。因此, 有必要对基金评审过程中的模糊属性进行详细分析并进行合理处置。

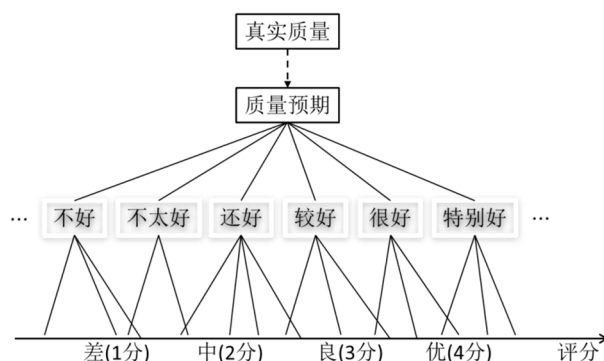


图1 基金评审的模糊属性。

4 评审结果的模糊集合表达

在经典集合(Two-valued Sets)理论中, 一个对象是否属于集合, 只有“是”或者“否”两种情况。比如定义一个集合表示“天气热”, 那么我们需要设置一个阈值(如气温超过30 °C)。当气温高于这个阈值时, 这天就属于“天气热”集合, 反之则不属于“天气热”集合。这种归类简单明了, 但在解决实际问题时可能会遇到困难: 比如气温是29 °C时, 就不属于“天气热”集合了, 但在人们主观判断上却可能觉得天气热。这表明主观评价中存在模糊属性, 经典集合无法描述这种模糊属性。在现实生活中, 人们通常不会把诸如天气、“颜值”等与人类主观感受密切相关的事务简单分成是或否、对或错的类别, 而更倾向于以“是”或“否”的程度来区分。例如, 在进行天气总体情况描述时, 天气预报一般不说: “今天热, 有云”, 而会使用一些模糊的说法, 如“今天天气较热, 部分地区多云”等等。这些“较”、“部分”、“多”等修饰程度的语言虽然是模糊的, 但在人类世界中却可以传递比数值更为丰富的信息。

模糊集合(Fuzzy Sets)由L. A. Zadeh在1965年提出, 用语言变量代替数值变量来描述系统行为, 能够较好地反映人类思维及决策过程的复杂性, 更适用于“有人参与”的系统, 提供了一种处理类似“天气热”等不确定性的有效方法¹⁸⁻²¹。模糊集合在自动控制、化工、交通、机器人、电子、家电、管理、决策、医疗、心理和经济等多方面得到了广泛的应用, 现已成为人工智能领域最为实用的理论工具之一²²。

与经典集合相比, 模糊集合在处理“天气热”问题时, 把气温作为基础变量, 构建一个值域为[0, 1]的一致性函数, 给基础变量赋予一个模糊参数(也称为一致性)。如此, 集合中的每个对象就由基础变量(气温)和模糊参数(由一致性函数生成的0到1之间的数值)组成。由这些对象构成的集合就是模糊集合。如气温为35 °C, 则可表达为(35, 1),

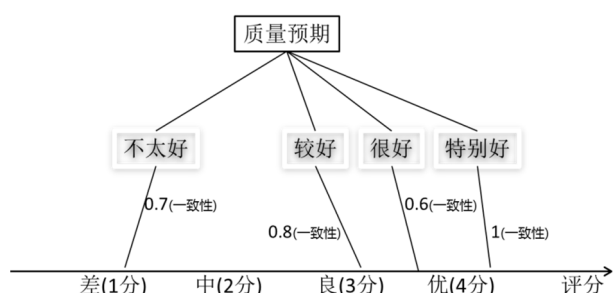


图2 基金项目申报评分等级的模糊参数。

其中“1”表示这天完全属于“天气热”模糊集合；如气温为15℃，则可表达为(15, 0)，其中“0”表示这天完全不属于“天气热”模糊集合；如气温为25℃，可表达为(25, 0.7)，其中“0.7”表示以0.7的一致性属于“天气热”模糊集合，当然也可以说是以某种“不一致性”不属于“天气热”模糊集合。通过上述事例，我们可以得到模糊集合的几个要素：一致性函数表示一个对象对于某个模糊集合的隶属程度，这种隶属程度用0到1之间的模糊参数表达，一个对象可以同时属于多于一个模糊集合。

应用模糊集合理论，我们可对项目申请函评价的每个等级或评分值赋予一致性函数，构造等级或评分值的模糊集合，以此解决“优、良、中、差”等级或“4、3、2、1”评分边界不清晰的问题。如果将一个项目申请包中的所有项目申请当作一个集合，则“优、良、中、差”等级就是这个项目申请包的四个模糊子集(项目评分等级的模糊参数如图2所示)。若某个项目申请的质量预期“特别好”，评分值就可表达为(4, 1)，其中1表示完全属于“优”这个模糊子集；若某个项目申请的质量预期“很好”，则可表达为(4, 0.6)，也就是以0.6的一致性属于“优”这个模糊子集；若某个项目申请的质量预期“较好”，则可表达为(3, 0.8)，即以0.8的一致性属于“良”这个模糊子集；若某个项目申请的质量预期“不太好”，则可表达为(1, 0.7)，即以0.7的一致性属于“差”这个模糊子集。

在实际应用中，评审专家依然是按照现有评审规则对项目申请给出“优、良、中、差”等级和对应的4、3、2、1评分，在专家完成评审之后，由系统通过一致性函数的算法进行相应的模糊参数赋值。我们希望运用模糊集合理论和一致性函数减小评审结果的质量预期误判，提升群体决策的共识程度，从而进一步提高评审质量。

5 总结和展望

本文针对国家自然科学基金项目函评环节的同行评议进行分析，讨论了基金评审结果的模糊

属性，阐述了如何利用模糊集合更精准地描述评审结果。评价等级/评分的模糊集合在形式上比较容易实现。但是，描述和设置模糊参数、构造一致性函数、将一组评分的模糊集合转化为一个确定的结论(或单一分值表达)是一个较为复杂的问题。今后，我们将针对正确构建一致性函数开展深入研究，努力解决“质量预期误判”导致的评审偏差问题，以助力基金委“负责任、讲信誉、计贡献”的评审机制改革，为实现科学基金项目精准资助的目标提供理论依据。

References

- (1) 习近平. 在中国科学院第十九次院士大会、中国工程院第十四次院士大会上的讲话. 北京, 2018.
- (2) 国务院关于全面加强基础科学研究的若干意见, 国发〔2018〕4号, 2018年01月31日.
- (3) Wade, N. *Science* **1973**, 179 (4069), 158. doi: 10.1126/science.179.4069.158
- (4) MacLane, S. *Science* **1975**, 190 (4215), 617. doi: 10.1126/science.190.4215.617
- (5) Cole, S.; Rubin, L. J.; Cole, R. *Sci. Am.* **1977**, 237 (4), 34. doi: 10.1038/scientificamerican1077-34
- (6) McCullough, J. *Sci. Technol. Hum. Values* **1989**, 14 (1), 78. doi: 10.1177/016224398901400107
- (7) Absher, J. R. *JAMA* **1990**, 264 (24), 31. doi: 10.1001/jama.1990.03450240045020
- (8) Scott, A. *Futures* **2007**, 39 (7), 827. doi: 10.1016/j.futures.2006.12.009
- (9) Rosenthal, R. *Behav. Brain Sci.* **1982**, 5 (2), 235. doi: 10.1017/S0140525X00011614
- (10) Cole, S.; Cole, J. R.; Simon, G. *Science* **1981**, 214, 881. doi: 10.1126/science.7302566
- (11) Blackburn, J. L.; Hakel, M. D. *Psychol. Sci.* **2006**, 17 (5), 378. doi: 10.1111/j.1467-9280.2006.01715.x
- (12) Lee, C. J. *Philos. Sci.* **2015**, 82 (5), 1272. doi: 10.1086/683652
- (13) Lee, C. J.; Sugimoto, C. R.; Zhang, G.; Cronin, B. *J. Assoc. Inf. Sci. Technol.* **2013**, 64 (1), 2. doi: 10.1002/asi.22784
- (14) Dai, Y. F.; Gao, F. X.; Wang, C. X.; Chen, Y. J. *Acta Phys. -Chim. Sin.* **2020**, 36 (8), 2003034. [戴亚飞, 高飞雪, 王翠霞, 陈拥军. 物理化学学报, **2020**, 36 (8), 2003034.] doi: 10.3866/PKU.WHXB202003034
- (15) Fu, X. F.; Huang, Y.; Cui, L.; Chen, Y. J. *Acta Phys. -Chim. Sin.* **2020**, 36 (8), 2004002. [付雪峰, 黄艳, 崔琳, 陈拥军. 物理化学学报, **2020**, 36 (8), 2004002.] doi: 10.3866/PKU.WHXB202004002
- (16) Zhang, G. J.; Fu, X. F.; Dai, Y. F.; Chen, Y. J. *Acta Phys. -Chim. Sin.* **2020**, 36 (8), 2003051. [张国俊, 付雪峰, 陈拥军. 物理化学学报, **2020**, 36 (8), 2003051.] doi: 10.3866/PKU.WHXB202003051

- (17) Ellis, D. G.; Fisher, A. B. *Small Group Decision Making: Communication and the Group Process*, 4th ed.; McGraw Hill Higher Education: New York, NY, USA, 1993.
- (18) Zadeh, L.A. *Inf. Control* **1965**, 8 (3), 338.
doi: 10.1016/S0019-9958(65)90241-X
- (19) Zadeh, L. A. *Inf. Sci.* **1975**, 8 (3), 199.
doi: 10.1016/0020-0255(75)90036-5
- (20) Zadeh, L. A. *Inf. Sci.* **1975**, 8 (4), 301.
doi: 10.1016/0020-0255(75)90046-8
- (21) Zadeh, L. A. *Inf. Sci.* **1975**, 9 (1), 43.
doi: 10.1016/0020-0255(75)90017-1
- (22) Zimmermann, H. J. *Fuzzy Set Theory—and Its Applications*, 4th ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2001.