



武汉大学科教管理与评价中心
Center for Science, Technology & Education Assessment



武汉大学
WUHAN UNIVERSITY

2023年中国情报学年会暨情报学与情报工作发展论坛

2023年7月13日

打开同行评议的“黑匣子”：专家评审行为特征分析

涂子依，周凯静，孙梦婷，张琳

涂子依

武汉大学信息管理学院

武汉大学科教管理与评价中心

OUTLINE

01

研究背景

02

数据与方法

03

结果与分析

04

结论与讨论



- **同行评议**是一种由从事该领域或接近该领域的专家对研究工作进行学术水平或重要性评价的方法，被广泛的应用于**论文评审、项目评估和科技管理等领域**。
- 尽管**同行评议**被普遍认为是评估科学知识的**黄金标准**，能够帮助期刊遴选高质量的、有价值的论文，但其过程中存在的**不公平、偏见**等现象也一直饱受学术界的诟病。
- 评审专家是同行评议制度的主体，其**评审行为的公正性**是保证同行评议评审效果的关键。
- 传统同行评议模式下的**评议报告、评审决议等过程数据皆不公开**，整个评审过程也并不受外界的监督。而评审过程的不透明可能会让审稿人利用匿名性给出更多的批评性评论，或者写出缺乏严谨性的意见，从而影响同行评议的质量，造成审稿人和作者之间的信任缺失等问题。



传统同行评议就像一个黑匣子，遮蔽了审稿人的身份和行为。

- 随着开放科学运动的不断发展，**开放同行评议**以其**公正、透明**的优势受到关注，并被国内外出版平台或期刊采纳。**同行评议报告的开放获取**为揭示同行评议的运行机制和探究专家评审行为提供了数据基础。

主要研究内容

本研究基于开放同行评议数据，结合定性和定量的方法，**探究开放同行评议背景下专家的评审行为特征**，对比其在不同自信度、不同评审模式和不同评议结果下**评审打分**和**评审意见**的行为特征上存在的差异，有助于更好了解同行评议的运行机制，加深对专家评审行为的理解。

OUTLINE

01

研究背景

02

数据与方法

03

结果与分析

04

结论与讨论



□ 数据源选择

OpenReview是一个秉持着公开同行评审、公开发表、公开来源、公开讨论、公开引导、公开推荐、公开API和开源八大原则的会议论文开放评审平台。其中**国际表征学习大会（International Conference on Learning Representations，以下简称“ICLR”）**是深度学习领域的顶级会议，自2013年以来一直采用公开评审机制，其公开的同行评议数据包括**论文基本信息、评审意见、最终评议结果**等信息，为国内外开放同行评议实证研究提供了大量的、良好的数据集。

通常情况下，一篇ICLR会议论文的评议报告包括以下三个部分：

- 专家评审打分（Rating）；
- 评审专家对自己评审意见准确性的自信度（Confidence）；
- 具体评审意见（Review）。

□ 概念界定

➤ 评审专家评审行为

在本研究中，将评审专家**评审行为**定义为评审专家参加各类评审活动时，在评审过程中所展现出的**行为活动的总和**。具体分为以下两个方面：

- 专家评审打分：在ICLR会议中每位评审专家会对其评审的论文给出1-10的评审得分；
- 专家评审意见：评审专家依据论文内容在不同方面的表现给出的具体文本评价。

➤ 论文评审得分和评议结果

- 论文评审得分：ICLR会议的每篇论文通常由3位评审专家对其进行评审并给出评审打分。本研究将每篇论文获得的专家评审打分的平均值作为该论文的评审得分；
- 论文评议结果：会议主席会综合考虑评审专家的评审打分、评审意见等，将论文划分为优秀程度最高、录用率最低的**口头汇报论文（Oral Presentation Paper，以下简称“OP论文”）**、次优秀的**海报展示论文（Poster Presentation Papers，以下简称“PP论文”）**和质量较差的**拒收论文（Rejected Papers，以下简称“RP论文”）**。

□ 概念界定

➤ 评审专家自信度

评审专家对自己评审意见正确性的自信程度，评审专家在给出评审打分和评审意见的同时，对自己的评议报告给出1-5分的自信度评分，自信度分值越高代表评审专家越肯定自己评审打分和评审意见的准确性。

自信度分值	定义
1	The reviewer's evaluation is an educated guess. 审稿人的评价是一种有根据的猜测。
2	The reviewer is willing to defend the evaluation, but it is quite likely that the reviewer did not understand central parts of the paper. 审稿人愿意为自己的评价辩护，但很可能审稿人没有理解论文的核心部分。
3	The reviewer is fairly confident that the evaluation is correct. 审稿人对自己评价的正确性比较有信心。
4	The reviewer is confident but not absolutely certain that the evaluation is correct. 审稿人对评价的正确性有信心但并非绝对肯定。
5	The reviewer is absolutely certain that the evaluation is correct and very familiar with the relevant literature. 审稿人非常肯定其评价的正确性，并且非常熟悉相关文献。

OUTLINE

01

研究背景

02

数据与方法

03

结果与分析

04

结论与讨论



□ 数据样本

本研究选取ICLR会议2017-2019年共计2534篇论文以及7602条评审数据，并从不同自信度、不同评审模式和不同评议结果三个维度对评审专家评审行为展开分析。

表 1 2017—2019 ICLR 各年份不同类型论文数量

年份	OP 论文	PP 论文	RP 论文	总计
2017	12	163	214	389
2018	23	306	479	808
2019	23	456	858	1337

注：ICLR会议2018年由单盲评审模式改为双盲评审模式，因此，本研究将ICLR会议2017年的数据作为单盲评审模式的同行评议数据集，ICLR会议2018年的数据则作为双盲评审模式的同行评议数据集，并基于此对比分析单盲和双盲两种不同评审模式下专家评审行为特征分布。

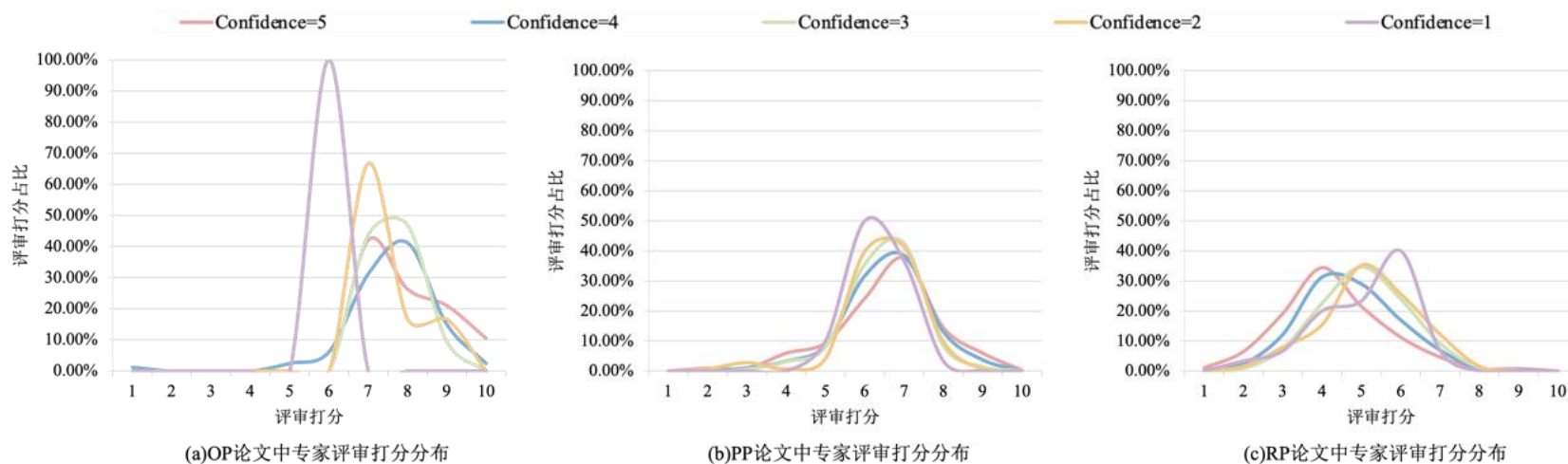
□ 不同自信度下评审专家评审行为特征差异

- 不同自信度下评审专家**评审打分**分布

表 2 不同自信度评审专家平均评审打分

ICLR会议2018-2019

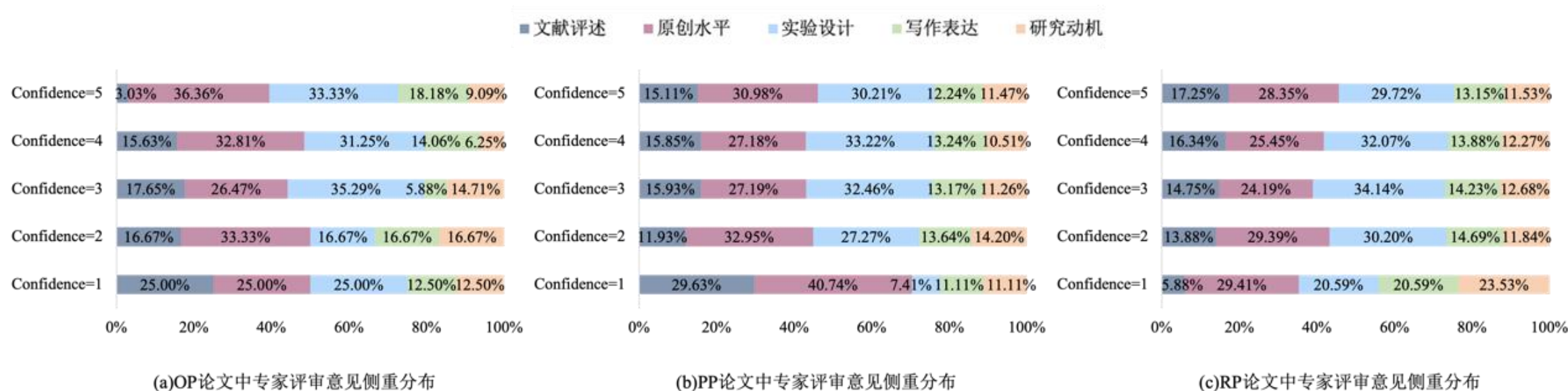
自信度	平均评审打分		
	OP 论文	PP 论文	RP 论文
Confidence=1	6.00	6.33	5.10
Confidence=2	7.50	6.46	5.19
Confidence=3	7.66	6.48	5.06
Confidence=4	7.60	6.57	4.72
Confidence=5	8.00	6.58	4.29



结果与分析-不同自信度

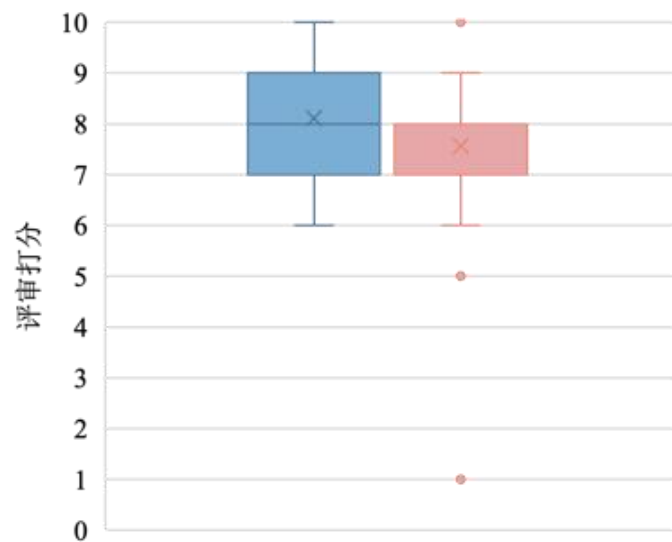
不同自信度下评审专家评审行为特征差异

- 不同自信度下评审专家**评审意见**侧重角度分布

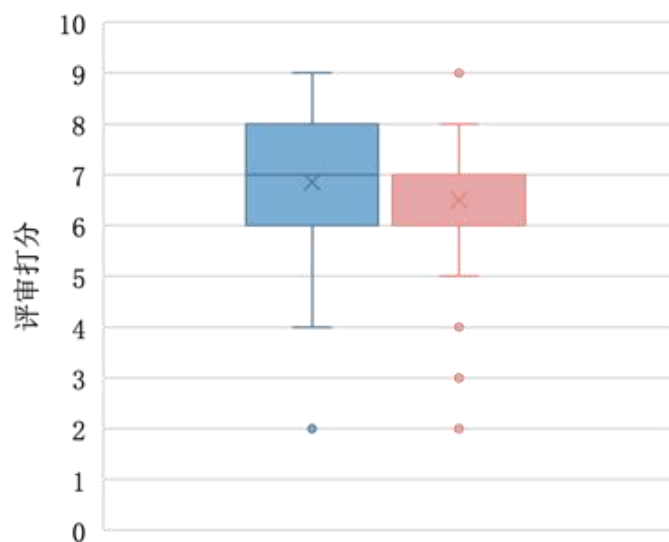


□ 不同评审模式下评审专家评审行为特征差异

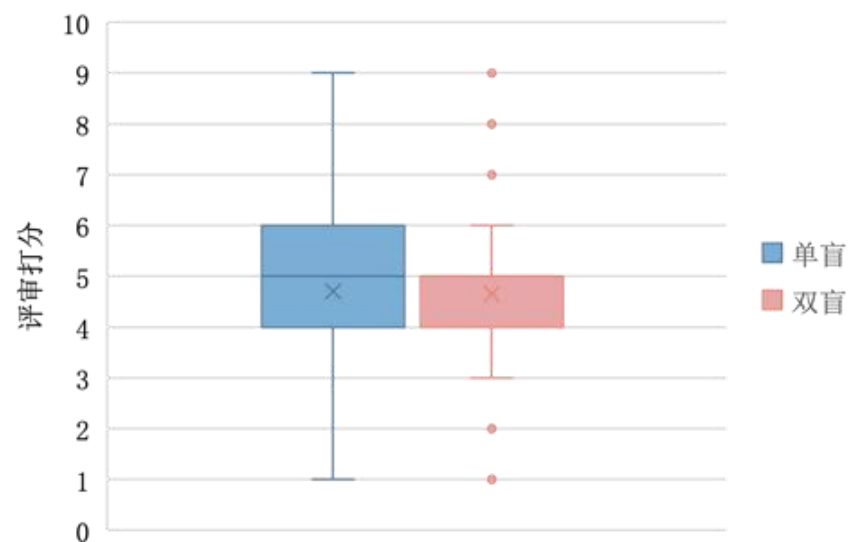
- 不同评审模式下评审专家**评审打分**分布



(a)OP论文中专家评审打分分布



(b)PP论文中专家评审打分分布

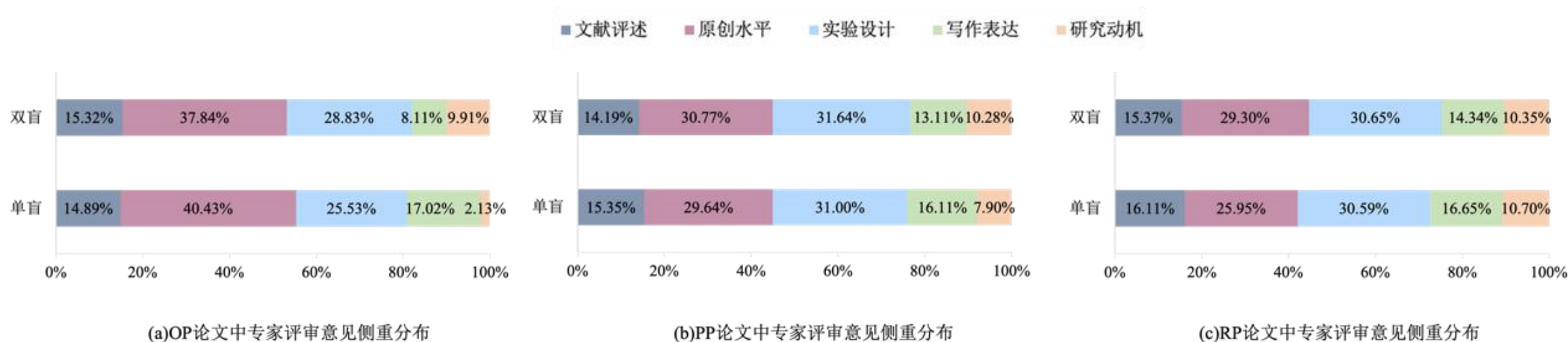


(c)RP论文中专家评审打分分布

ICLR会议2017-2018

□ 不同评审模式下评审专家评审行为特征差异

- 不同评审模式下评审专家**评审意见**侧重角度分布



□ 不同评议结果下评审专家评审行为特征差异

- 评审得分一致但最终评议结果不同的论文

表 3 不同评审得分区间三类论文的数量统计表 ICLR会议2018-2019

评审得分	OP	PP	RP
[9,10]	1	0	0
[8,9)	18	14	0
[7,8)	26	230	2
[6,7)	0	435	123
[5,6)	1	78	529
[4,5)	0	4	485
[3,4)	0	1	166
[2,3)	0	0	32

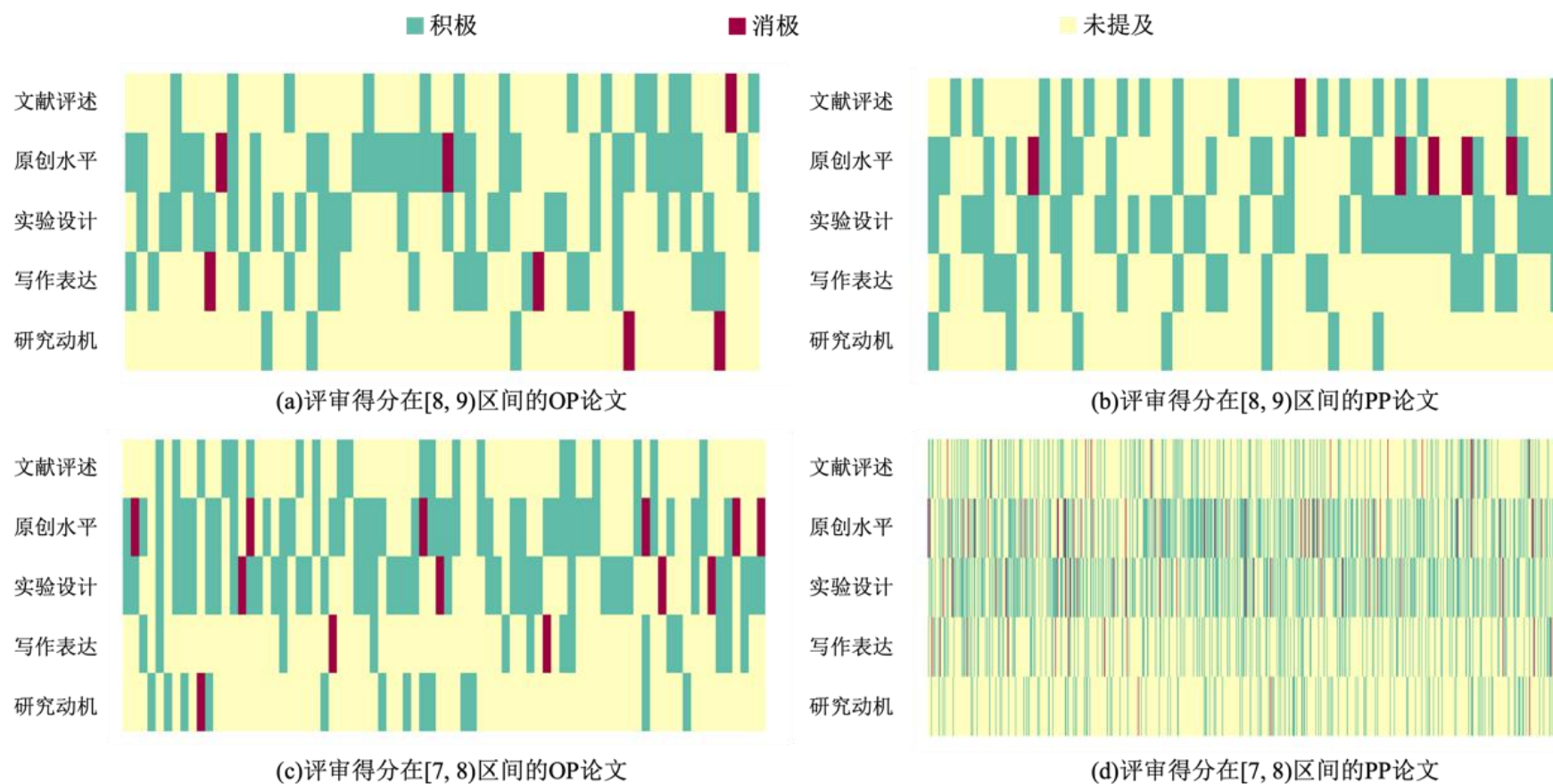
- 本研究在控制了评审得分相近的情况下，通过比较不同评审结果下专家评审行为的差异性，即评审意见的侧重分布和情感分布差异，来探究专家评审意见对于论文最终评审结果的影响。



结果与分析-不同评议结果

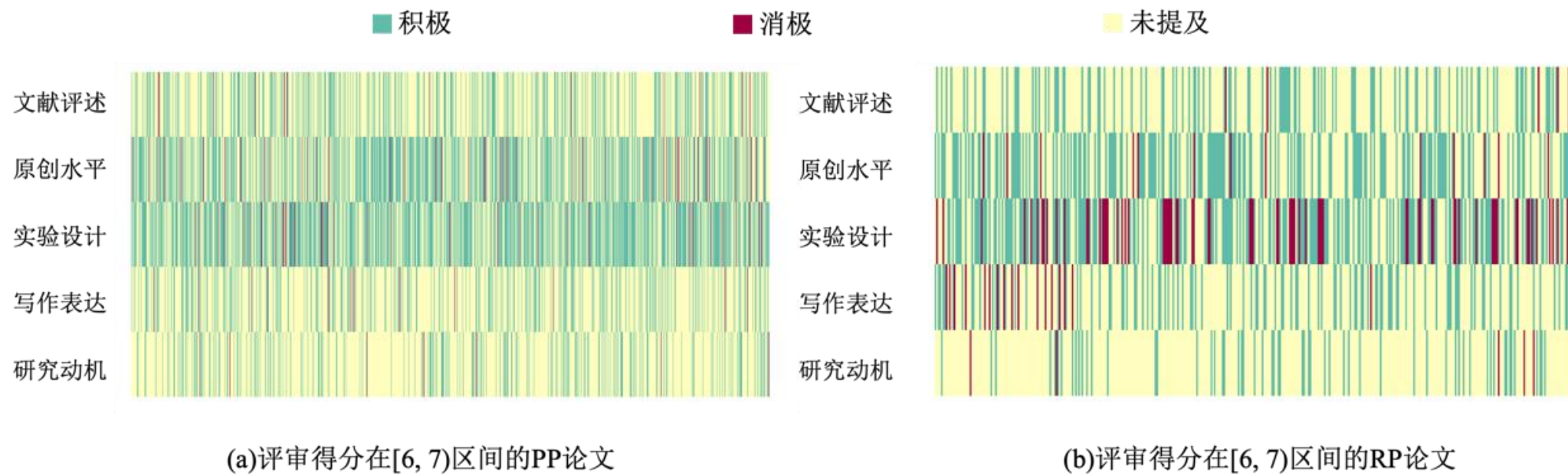
不同评议结果下评审专家评审行为特征差异

- OP论文和PP论文比较下专家评审行为特征



□ 不同评议结果下评审专家评审行为特征差异

- PP论文和RP论文比较下专家评审行为特征



结果与分析-不同评议结果

不同评议结果下评审专家评审行为特征差异

- OP论文和RP论文比较下专家评审行为特征 得分区间[7, 8)

RP 论文 A	专家 1	3	(a) The discussion on related work seems to be good.	文献评述	积极	OP 论文 A	专家 1	5	(a) Interesting and potentially very useful novelties are presented.	原创水平	积极
			(b) The quality is high and the approach seems to me to be new and worth studying.	原创水平	积极				(b) The writing is excellent	写作表达	积极
			(c) English must be checked.	写作表达	消极				(c) Experiments can be improved.	实验设计	消极
	专家 2	3	(a) Needs to provide more details.	文献评述	消极		专家 2	3	(a) The Gaussian approximations and moment Propagation approximations are clever and highly original.	原创水平	积极
			(b) In the experiment, the authors only compare their work with NMNs.	实验设计	消极				(b) The derivation is rather heuristic.	实验设计	积极
	专家 3	5	(a) The proposed model not only get the correct answer but also ...	实验设计	积极		专家 3		(c) The paper is generally well written.	写作表达	积极
									(d) The paper is clearly motivated in the context of the existing literature.	研究动机	积极
									(a) This paper is interesting since most of the existing works focus on Monte Carlo variational inference.	原创水平	积极
									(b) I would like to see some empirical comparison on real-world datasets.	实验设计	消极

OUTLINE

01

研究背景

02

数据与方法

03

结果与分析

04

结论与讨论



□ 主要结论

- 在**专家自信度层面**，自信度高的评审专家对论文研究内容更了解，能给出更具针对性的评审打分，在对论文进行评价时能从论文的核心内容，即实验设计方面给出更多的、具体的意见；
- 在**评审组织形式层面**，双盲评审模式下评审专家能更好克服单盲评审时作者身份带来的影响，其评审打分更加严格，更敢于给出低分的评价，且更少提及论文语言表达方面的不足；
- 在**论文的评审决议层面**，会议主席除了考虑论文评审得分，还会重点参考评审专家对论文核心内容方面的点评，例如，在评审得分相近的ICLR论文中，评审专家在论文实验设计和原创水平方面给出更对正面评价的论文，更有可能获得更好的评审决议。

□ 政策意见

- （1）**在遴选评审专家时，应该优先选择对论文所属领域更熟悉的“小同行”专家。**专家的认知和理解能力在很大程度上影响评审质量，对论文的研究内容更了解的“小同行”专家能给出更具针对性和明确性的评审打分和评审意见，能够为作者提供明确的论文修改方向，有助于论文学术水平的提升。
- （2）**在同行评审过程中引入双盲评审模式。**相较于单盲评审，双盲评审模式下的评审专家评审打分更为严格，且较少的关注于论文的语言表达等方面，能够更公正、客观的对论文进行评审，弥补单盲评审常见的由作者身份信息带来的偏见等缺陷。
- （3）**规范专家行为，加强对评审指标的完善和细化。**因此同行评议组织方在邀请评审专家进行评审时，应该根据所属研究领域和评价标准，细化专家评审的评分指标，规范专家从不同的方面分点对论文进行评价，提高评议报告的合理性和完整性，同时也有助于提高最终决策的效率和准确性。

研究局限与展望

- ✓ 本研究使用的ICLR数据集样本量较小，可能会限制研究的深入开展，在进行结论推断时支撑力不够充足；
- ✓ 受限于文章所使用的技术方法，本研究仅抽取了评审意见中的侧重角度和情感态度两种特征，对于蕴含了大量评审信息的评审意见文本，未能做到充分的利用和挖掘；
- ✓ 本研究没有对论文的所属研究领域进行分类，控制论文研究主题对于结果的影响，在后续研究中会进一步补充对该分析维度的探讨。





武汉大学科教管理与评价中心
Center for Science, Technology & Education Assessment



武汉大学
WUHAN UNIVERSITY

敬请批评指正！

打开同行评议的“黑匣子”：专家评审行为特征分析

涂子依

武汉大学信息管理学院

武汉大学科教管理与评价中心