

数学论文学术不端的表现形式及防范措施*

黄 颖

《重庆师范大学学报(自然科学版)》编辑部,401331,重庆

摘 要 由于学术不端查重系统对数学公式查重的局限性,导致数学论文查重结果不理想,因此需要编辑结合实际情况对数学稿件可能存在的学术不端行为进行排查。本文结合实例分析了数学类论文学术不端行为的几种类型,并提出相应的防范措施。

关键词 学术不端文献检测系统;数学论文;学术不端;防范措施

Academic misconducts in mathematical papers and precautionary measure//HUANG Ying

Abstract Due to the limitation of duplicate checking system for academic misconducts on mathematical formulas, the result of the duplicate checking of mathematical papers is not ideal. Therefore, it is necessary for editors to check the possible academic misconduct in mathematical manuscripts according to the actual situation. This article analyzes several types of academic misconduct in mathematical papers, and puts forward corresponding precautionary measures.

Keywords academic misconduct document detection system; mathematical papers; academic misconduct; precautionary measure

Author's address Editorial Office of Journal of Chongqing Normal University (Natural Science), 401331, Chongqing, China

DOI:10.16811/j.cnki.1001-4314.2021.03.009

为了有效防范学术不端行为,国内各大数据库早在10余年前就纷纷推出了各自的论文查重检测系统。例如,中国学术期刊(光盘版)电子杂志社与万方知网技术有限公司共同研发了科技期刊学术不端文献检测系统(AMLC);万方数据公司也开发了万方论文相似性检测系统;维普公司继后也推出了维普一通达论文引用检测系统;超星数据库也推出了大雅相似性分析系统^[1]。这些论文检测系统在实际操作中为期刊编辑人员带来很大便利,使用者上传待测稿件后,可得到一份详细的检测报告,内容包括疑似重复部分、重复比例、重合文献、是否提前检测等等。这些检测系统的使用在一定程度上对学术不端行为起到了震慑作用,使之得到了遏制。由于人工智能技术的局限性,这些检测软件对数学论文中的公式却一筹莫展。打开数学论文的学术不端检测报告,不难发现有公式的地方自动跳过了。如果有疑似重复的内容,一定是以文字居多,偶尔也会有未使用专业软件编辑而直接在Word文档

中录入的公式、字母被检测到。而结合数学学科特点,不难发现定理、公式等的证明往往才是论文的核心内容。也就是说,查重检测系统有可能无法对数学类论文的核心内容进行检测。

2020年6月15日,德国独立科学记者Schneider在他主办的科学新闻网站“For Better Science”(http://forbetterscience.com)上发布了65篇数学领域论文涉嫌造假的消息^[1],涉及国内44家高校,署名作者多达77名^[2]。与前几年披露的学术不端行为所涉及的生物学、医学等领域不同^[3-4],数学领域极少被报道存在学术不端行为,因而此次曝光也为国内学术出版再次敲响警钟,学术编辑需要在稿件筛选中重视数学稿件学术不端问题。编辑在实际操作中需要加强责任意识,发挥专业优势,对数学稿件采取检测系统查重与人工查重相结合的方法达到甄别的目的。本文主要结合笔者的实践经验,对数学论文的学术不端检测中存在的问题进行分析归类,并提出相应的改进或防范措施。本文的查重结果参考AMLC的检测结果。

1 数学论文学术不端行为的表现形式

1.1 一稿多投

数学作为基础学科,在大部分自然科学综合类期刊中都占有一席之地。中国知网搜索O1、O2中图分类号,结果累计有601311条文献,其中有226311条发表在各类学报。笔者选取了《中国学术期刊影响因子年报(自然科学与工程技术·2019年版)》中自然科学与工程技术综合类下前100本大学学报作为调查对象^[5],通过对各刊在中国知网近10年载文学科数据统计发现,有63本刊物的数学论文数量排在该刊学科载文量首位,并且所发表的数学论文没有特定的研究方向。同时,《中国科技期刊发展蓝皮书(2020)》统计数据显示,数学学科的论文有94.26%的作者来自高校^[6],而高校作者往往迫于职称压力而追求论文发表数量。因此,大学学报的同质化办刊以及庞大的数学论文发表需求量,直接导致了数学论文易出现一稿多投现象。而地方高校学报在送审稿件时通常会优先选择所在地高校的研究人员作为审稿专家,当作者将同一稿件投往同级别但不同省市的高校学报时,面对的是不同的评审专家,所以这样的行为也不容易被发现。

* 重庆市期刊出版专项资助项目(2020)

1.2 替换变量符号或用不同表述

1.2.1 简单的字母替换 数学公式的变量、函数符号通常是作者自定义,因此,有的作者会通过将同一论文中的函数用不同字母表示来规避学术不端查重系统。

例如, $f(x) = y$ 用 $g(a) = b$ 来替换。再如图1所示,为同一作者发表在不同期刊的论文,在对同一个问题的表述中,将其中一些字母进行了替换,但公式的基本形式、实质不变,而 AMLC 查重结果却显示为0。

考察如下 Rössler 混沌系统

$$\begin{cases} \dot{x}_1 = -(x_2 + x_3), \\ \dot{x}_2 = x_1 + ax_2, \\ \dot{x}_3 = b + x_3(x_1 - c), \end{cases} \quad (1)$$

其中 x_1, x_2, x_3 为系统的状态量, $a = 0.2, b = 0.2, c = 5$ 时系统 (1) 出现混沌吸引子, 以 (1) 作为主系统, 从系统设计为

(a) 论文一

Rössler 混沌系统为

$$\begin{cases} \dot{x}_1 = -(y_1 + z_1), \\ \dot{y}_1 = x_1 + ay_1, \\ \dot{z}_1 = b + z_1(x_1 - c), \end{cases}$$

其中: x_1, y_1, z_1 为系统状态量.

当 $a=0.2, b=0.2, c=5.0$ 时, 出现混沌吸引子. 初值 $(x_1(0), y_1(0), z_1(0))$ 取 $(2, 2, 2)$ 时, 系统状态量 x_1 的时域波形图和系统相图如图 1, 2 所示.

(b) 论文二

删除	生成报告单	转移至	导出Excel	查询到346篇文章	首页	上一页	1	2	3	4	5	...	下一页	尾页
<input type="checkbox"/>				参数未知的分数阶不确定Rössler混沌系统的自适应滑模同步	ssler混沌系统的自适应滑模同步	0%	-	2021/3/15	17:46:40					选择报告单
				去除引用复制比: 0% 总复制比: 0% 表格: 0 公式: 0 ocr图片: 0										
<input checked="" type="checkbox"/>				不确定整数阶和分数阶Rössler混沌系统的自适应滑模同步	ssler混沌系统的自适应滑模同步	0%	-	2021/3/15	17:46:38					选择报告单
				去除引用复制比: 0% 总复制比: 0% 表格: 0 公式: 0 ocr图片: 0										

(c) 2 篇论文查重结果对比

图1 字母简单替换的公式对比及查重结果

1.2.2 中文表述、不同的数学公式互相替换 数学也是一门语言,用数学公式可以简单明了地表达一些数学关系,但有的作者却利用简洁的数学语言与中文表述相互转换来规避查重系统的检测。例如,将“ a_1, a_2, \dots, a_p 不同时为0”表述为“ $a_i (i = 1, 2, \dots, p)$ 不同时为0”;将“反比例函数”用“函数 $f(x) = 1/y$ ”来表示,有时也会用“ $g(a) = 1/b$ ”来表示;“函数 $f(x)$ 的导数”用“ $f'(x)$ ”或“ dx/dy ”表示;函数的指数形式 $x^{p/q}$ 与根式形式 $\sqrt[q]{x^p}$ 互相替换。

1.2.3 利用等量关系替换 数学学科中往往会存在一些等量代换、等量关系,因此有的作者会投机取巧,将发表过的论文中的某一概念用不同的表述表示,再

重新投稿。例如,指数函数 $a^x = N$ 与对数函数 $x = \log_a N (N > 0)$ 互换;将 $\log_a (MN)$ 写成 $\log_a M + \log_a N$;用 $\exp(x)$ 与 e^x 互相替换。由于这些公式的基本写法不同,所以即使 AMLC 中能对公式进行检查,但由于替换可能不止一处,且公式可能比这些基本形式更加复杂,使得检测系统很难检测到。

1.3 简单修改论文内容或定理条件

有的作者会将论文作一些细小的修改,但修改的内容在创新性方面不足以支撑一篇新论文发表。例如,《整数阶分数阶 Rucklidge 混沌系统的自适应滑模同步》《不确定分数阶 Rucklidge 系统自适应滑模同步》与《不确定分数阶 Rucklidge 混沌系统的两种滑模

同步方案》,这3篇论文为同一研究团队撰写,同年发表于不同期刊,但研究内容相近。经过比对发现他们对所研究问题的描述一模一样,仿真图形也为相同图片。尤其是前2篇论文的证明思路、证明过程高度相似,而实际查重结果却显示仅有1篇与原文重复率为30.2%,与其余2篇没有表现出重复,并且另2篇已发表论文再次查重结果竟然为0。

1.4 不当署名

不同于实验学科,一篇数学论文撰写往往是基于文献的个人思考结果,不会有太多的人参与,即使多人合作,在专业上也不会相差太多。若数学稿件中存在多人署名,且其中有作者以往研究方向与投稿方向相差较大或存在多个不相关方向作者署名的情况时,可能就会存在不当署名的学术不端行为。

另外,稿件状态为外审通过或录用待刊发时,作者要求增加作者数量甚至要求更换第一作者,这时大概率存在不当署名情况。笔者曾遇到一篇数学论文录用后要求增加作者,而作者单位为某高校外语学院英语教研室,明显属于不当署名范畴,因此予以严词拒绝。笔者所在编辑部因此制定了相应的规定,所有论文一经投稿,不再允许修改第一作者、通信作者,不得增加与论文无关的作者。

2 数学论文学术不端行为的防范措施

基于上述分析,数学论文学术不端行为表现形式多样,AMLC查重结果又不能全面地反映实际情况。因此,笔者基于以往工作经验,建议对于数学论文学术不端行为采取查重系统与人工查重相结合的方式。

2.1 初审

2.1.1 AMLC查重报告解读 查重报告结果分为2类:一是重复率超过编辑部规定范围,一是重复率未超过编辑部规定范围。但鉴于前述AMLC对数学论文查重的缺陷,编辑不应该简单地根据查重结果一刀切,而应根据查重报告的情况进行处理。处理流程见图2。

从图2可以看出,在初审阶段的AMLC查重结果显示为高时,如确定大部分内容重复,则毫无疑问直接退稿;但如果仅是引言、定义、引理这些内容重复时,根据数学学科特点,任何一项定理及证明都不是凭空而来,而是以已有研究成果和相关定义为基础,这些内容都是数学论文所必需的,因此,这种情况下就还需转编辑进行人工排查。当AMLC查重结果显示为低时,如果查重“全文(标明引文)报告单”中无乱码,则基本上可判断稿件在当时无学术不端行为,可通过该项审查;若该报告单中公式或文字大量显示乱码,则说明查重

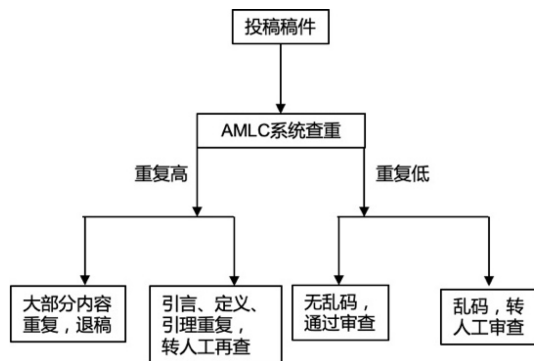


图2 初审阶段AMLC查重报告解读流程

结果失效,应该转编辑进行人工排查。其中,根据笔者经验,由LaTeX生成的PDF文件查重时,由于LaTeX文件的编辑会使用编程语言,因此在查重系统中会因无法解析还原,所有的公式均显示为乱码从而影响查重,这一情况也同样存在于Word转换的PDF文件中。因此,最好能要求作者提交Word格式的稿件。

2.1.2 编辑人工排查 编辑进行人工排查时有2种情况,一是查重报告结果较高,但显示重复内容为非定理、结论部分。对这部分稿件先在CNKI中搜索第一作者、通信作者,再根据检索情况排查。主要看题目重合程度,定理、结论部分是否有替换类的表述,公式结构上有没有本质区别。如果都没有问题,则可通过检测,进行下一步审稿流程;如果有疑问,可就疑问处与作者沟通,请作者予以口头或书面说明区别或创新之处,再进入下一步审稿流程,同时将存疑的内容和作者说明提交给外审专家。还有一种情况是查重报告结果因为论文文档格式问题显示乱码,从而无法获得准确的查重结果。可请作者按要求提交Word文档再按流程重新进行查重。

在人工排查的时候,还应将论文关键词、作者姓名、单位等信息在CNKI上检索,筛选出相关论文或同一研究团队发表的相似论文。若筛选结果与查重报告中同时出现相同的文献,则应予以高度关注,从前文提到的几类学术不端表现形式进行人工比对,如果明显重复,则显示存在学术不端行为,可直接退稿;若不确定则可留待下一流程处理。

2.2 利用同行评议防范学术不端行为

前面提到,在人工排查时,由于数学学科的研究方向众多且编辑自身水平有限,可能还是会存在有疑问但又无法直接判定是否有学术不端行为的稿件。这时,编辑可请作者说明自己论文的创新之处及与疑似文献的区别。编辑在将稿件送同行评议时可说明情况,并同时附上投稿论文、疑似文献和作者说明,以便审稿专家判断。例如,本刊曾前后相隔3个月收到

《公共工期指派和可控处理时间资源约束下最大费用最小化单机排序问题》《公共工期窗口指派和可控处理时间资源约束下最大费用最小化单机排序问题》2 篇论文,第 1 篇通过同行评议确定录用后,作者投了第 2 篇,虽然查重顺利通过,但由于 2 篇论文题目高度相似且作者为同一人,因此编辑将 2 篇论文同时再送同行评议专家评审,而专家经仔细核对后得出结论,后一篇论文包括前一篇的内容,前一篇事实上为第 2 篇稿件的特例。这一结论后来与作者沟通中也得到了确认。因此,编辑部将第 1 篇论文作退稿处理,刊发了第 2 篇论文。

2.3 编辑注意积累易出现学术不端的关键词

虽然数学学科的研究方向比较多,但数学编辑应该对热门的研究方向、较成熟的研究成果等有所了解,因为通常成熟的数学理论更容易产生修改重发这类学术不端行为。因此,编辑需要通过对同类期刊数学论文的浏览来获取出现频次较高的论文主题、关键词及作者名称等内容,对经常出现的内容可建立编辑部的小型数据库,以便随时查阅。例如,前面提到的 Schneider 爆料的 65 篇涉嫌抄袭的论文中有 34 篇论文仅标题就出现了薛定谔(Schrödinger)。本刊曾收到关于“Hermite-Hadamard 不等式”的论文,经 CNKI 查询,其中 1 篇稿件的第一作者发表相关论文 160 余篇,2015 年以来该作者发表近 90 篇论文,因此,编辑在此后处理该作者稿件时尤其注意查重问题。其中“Schrödinger”“Hermite-Hadamard 不等式”等词语就可作为论文易出现学术不端行为的关键词收集起来。

同时,由于数学学科方向众多,一个数学编辑不可能掌握所有数学学科内容,这就需要大家群策群力,集众家之长。建议在期刊集群的模式下建立 QQ 群、微信群,将数学编辑聚集在一个平台,形成学科编辑联盟,以实现数学不端相关信息共享。例如:中国高校科技期刊研究会下师范类期刊、数学稿件占比较多的综合大学学报等期刊的数学编辑可建立数学公式类的学术不端联盟,就相关问题、防范措施进行交流、探讨。

2.4 调整期刊定位,专业化办刊

前面已经分析,正是由于各高校学报的同质化办刊、学科分散,因而更容易产生数学稿件学术不端行为。为了改善这一情况,高校期刊应尽快调整办刊方向,确定期刊报道的优势学科。即便是报道数学学科,也应尽量集中在 1 个或 2~3 个方向,而不应来者不拒,尤其是对讨论比较成熟的数学理论论文,更应该着

重审查其创新性。

《重庆师范大学学报(自然科学版)》自 2007 年开设“运筹学与控制论”栏目以来,累计载文 297 篇,报道的数学内容越来越集中在该领域,2019 年刊登的 53 篇数学稿件中运筹学稿件占绝对优势,共 27 篇;并在 2020、2021 年出版 2 期“运筹学与控制论”专辑,累计刊发 36 篇论文。这 2 期专辑中第一作者具有高级职称或博士学位的比例达到 55.6%,国家自然科学基金资助率为 91.7%。集中报道使得越来越多的优秀作者愿意投稿,从而来稿质量也得到了明显提升。

3 结束语

近些年来,《重庆师范大学学报(自然科学版)》编辑部将上述措施应用于对数学类稿件的学术不端行为的防范,并取得了较好的成绩。CNKI 的个刊分析报告显示:2018 年本刊已刊载的数学论文中涉及学术不端行为稿件量仅 1 篇,2019 年则降为 0 篇。因此,数学稿件的学术不端行为不能简单地依靠学术不端查重系统来判断,数学编辑必须要提高数学修养,掌握一定的查重技巧,才能尽可能地防范学术不端行为,为提高稿件质量、提升期刊影响力、净化学术氛围做出应有的贡献。

4 参考文献

- [1] SCHNEIDER L. Begger's test for Schrodingerean predator-prey system[EB/OL]. (2020-06-15)[2021-02-15]. <https://forbetterscience.com/2020/06/15/beggers-test-for-schrodingerean-predator-prey-system/>
- [2] SCHNEIDER L. 论文中文列表[EB/OL]. (2020-06-15)[2021-02-15]. <https://forbetterscience.files.wordpress.com/2020/06/public-copy-of-math-papermill-chn.pdf>
- [3] 李伟,陈新石,游苏宁. 医学期刊编辑在预防学术不端中的辅助作用[J]. 编辑学报, 2012, 24(3): 254
- [4] 张维,吴培红,汪勤俭,等. 国内外生物医学期刊撤稿规范分析及应用学术不端行为的防范策略[J]. 编辑学报, 2020, 32(3): 251
- [5] 中国科学文献计量评价研究中心,清华大学图书馆. 中国学术期刊影响因子年报(2019年版)[M]. 北京:《中国学术期刊(光盘版)》电子杂志社有限公司, 2019: 115
- [6] 中国科学技术协会. 中国科技期刊发展蓝皮书(2020)[M]. 北京:科学出版社, 2020: 122
(2021-03-16收稿;2021-04-24修回)