

# 基于裁判文书的法学实证研究之审视

屈茂辉

(湖南大学,长沙 410000)

**摘要:**知识图谱分析结果表明,基于裁判文书的法学实证研究成果数量总体上呈现增长趋势,但在近两年出现明显下降;基于裁判文书法学实证研究学术群体规模正在逐步缓慢扩大,但研究成果在研究主体间的分布呈较为集中趋势;研究成果所涉及学科领域具有较为明显的局限性,且具体研究内容伴随热点问题获得短暂爆发之后开始呈退缩之势。在方法转型背景之下,基于裁判文书法学实证研究的演变,受到研究方法在数据可得性、数据完整性以及数据客观性等方面所存在之优势与缺陷影响。研究人员自身应当在研究过程中通过丰富研究数据、规范研究方法之适用以及推动数据处理技术的发展等方式寻求基于裁判文书法学实证研究的长远发展之道。

**关键词:**知识图谱;裁判文书;实证研究;内容分析法;数据处理

中图分类号:DF920.0 文献标志码:A

DOI:10.3969/j.issn.1001-2397.2020.03.03 开放科学(资源服务)标识码(OSID):



近年来,随着我国裁判文书公开制度的不断完善,以裁判文书法律大数据作为分析对象的法学实证研究越来越受到学者的青睐。然而,无论是基于裁判文书实证研究所对应的研究方法,还是作为具体数据来源的裁判文书本身,都存在可能阻碍法学实证研究整体发展的固有缺陷。<sup>①</sup>因此,我国法学实证研究是否真如学界所期许的那样,借助裁判文书可得性的提升,实现其自身的迅猛全面发展,还需要予以全面审视。长期以来,关于法学实证研究之审视的理论研究成果并不缺乏,遑论研究者对实证研究是秉持支持还是反对态度,但这些研究成果中,一则缺少将基于裁判文书法学实证研究作为独立研究对象的审视,二则对于法学实证研究实然状态的评判缺乏客观数据的支持。因

收稿日期:2020-01-15

**基金项目:**本文得到湖南大学数理-计量法学研究中心“数量法学理论与应用研究”(900226103)。曾作为第六届数理-计量法论坛的主旨论文进行讨论,从白建军教授、曾贇教授、熊谋林副教授等诸多学者的意见中获益良多,特此致谢。

**作者简介:**屈茂辉(1962),男,湖南新宁人,湖南大学法学院教授、博士生导师,法学博士。

<sup>①</sup>参见程金华:《迈向科学的法律实证研究》,载《清华法学》2018年第4期,第149页。

<sup>②</sup>以citespace软件为工具进行知识图谱分析,系受同济大学上海知识产权学院博士研究生刘星同学在第六届数理-计量法论坛中所作报告之启发。

此,为明确裁判文书在法学实证研究中的具体应用,本文以 citespace 软件为工具<sup>②</sup>,拟以从中国知网检索得到的 612 篇以裁判文书为数据来源的实证研究期刊文献为分析对象,<sup>①</sup>对我国基于裁判文书法学实证研究中的研究成果产出、研究参与者以及研究视域等三个主要方面进行实证考察,进而在剖析其演变内因的基础上对如何利用裁判文书进行实证研究的科学径路陈一孔之见。

## 一、基于裁判文书法学实证研究发展状况检视

### (一)研究成果产出

裁判文书作为司法过程的主要产物,其中承载着可为法学实证研究所用的丰富数据信息,其网络公开数量在近年来的迅猛增长,则被诸多法学实证研究学者视为本领域发展的重要契机。<sup>②</sup>而理论研究成果作为学科发展的主要表征之一,也必然随着裁判文书网络公开数量的增长而增多。申言之,基于裁判文书法学实证研究作为以裁判文书为基础数据来源的具体研究形式,其所产出之研究成果数量,除受法学研究方法转型之影响外,还应与裁判文书网络公开数量在一定程度上呈现出正相关关系。

通过对中国知网中检索所得的期刊文献按年度进行统计,结果如图 1 所示。根据发文量的年度分布数据所构建的线性方程系数为正,R 方值为 0.7539。这一结果表明,基于裁判文书法学实证研究成果的数量与时间序列成正相关,就总体趋势而言,随着时间的推移,基于裁判文书的实证研究成果数量也随之增长。然而,如若对发文量在具体年份的分布情况进行观测,在总体增长趋势之下,基于裁判文书法学实证研究成果数量变化仍可以划分出较为明显的三个阶段:第一阶段为 2000 年至 2011 年,该阶段中,研究成果数量呈缓慢增长趋势,年度发文量较少,且年度增长率较低;第二阶段为 2012 年至 2017 年,与第一阶段末的 2011 年发文量相比,本阶段初 2012 年的发文量得到明显提升,并且,本阶段中的 2013 至 2016 年之间的发文量以较高的年度增长率明显高于第一阶段;第三阶段,即 2019 年,本年度发文量与第二阶段末的 2018 年发文量相比急剧下降。

期刊文献年度发文数量在前两个不同阶段的界限划分,即 2011 年至 2012 年间发文量的明显增长,与该阶段我国裁判文书网络公开数量的发展变化存在高度契合。<sup>③</sup>然而,单就发文数量而言,第三阶段的研究成果数量急剧下降,与之前较长时期内的研究成果总体增长趋势明显相悖,进而表明该阶段研究成果数量变化与我国裁判文书网络公开裁判文书总量不断增长趋势明显不相符合。<sup>④</sup>据此,综合比较各阶段研究成果数量的变化趋势,可以得知,基于裁判文书法学实证研究的发展确实与裁判文书网络公开数量增长之间存在较为明显的相关性,但该相关性本身可能受到其他因素

<sup>①</sup> 检索方案为:以“实证研究”及“实证分析”为主题,并且全文包含“裁判文书”与“判决书”为关键字,期刊来源为“CSSCI”,检索日期截止 2019 年 12 月 31 日,并人工将无关检索结果剔除。

<sup>②</sup> 参见程金华:《迈向科学的法律实证研究》,载《清华法学》2018 年第 4 期,第 149 页。

<sup>③</sup> 2011 至 2013 年,互联网裁判文书公开数量得到成倍增长。参见马超、于晓虹、何海波:《大数据分析:中国司法裁判文书上网公开报告》,载《中国法律评论》2016 年第 4 期,第 200-210 页。

<sup>④</sup> 时至今日,裁判文书上网工作仍处于不断推进与完善进程之中,裁判文书上网率同样也在不断上升过程中。参见杨金晶、覃慧、何海波:《裁判文书上网公开的中国实践——进展、问题与完善》,载《中国法律评论》2019 年第 6 期,第 125-147 页。

的反向影响与调节。

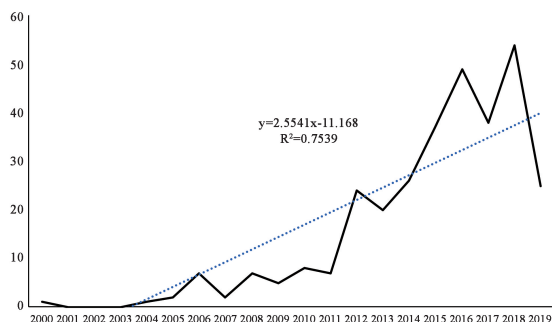


图1 年度发文量趋势图

## (二) 研究参与主体

基于本文所选取的期刊文献样本,通过使用 citespace 软件对文献中所涉及研究机构进行分析,首先得到研究机构的总体分析情况。<sup>①</sup> 如图 2 所示,西南政法大学、中国政法大学、山东大学法学院、吉林大学法学院、四川大学法学院、北京大学法学院、浙江大学光华法学院、中国人民政法学院、中南财经政法大学法学院以及上海交通大学法学院等机构研究成果突出,其他研究机构所产出的成果数量则十分有限。<sup>②</sup> 即从总体上来看,虽然已有一定数量的研究机构开展基于裁判文书法学实证研究,但研究成果的产出则集中来自于以“五院四系”为典型的少数法学研究机构。对比当前我国法学院系及专门科研机构总体数量,可以得知,基于裁判文书开展法学实证研究的学术群体规模依然较小。

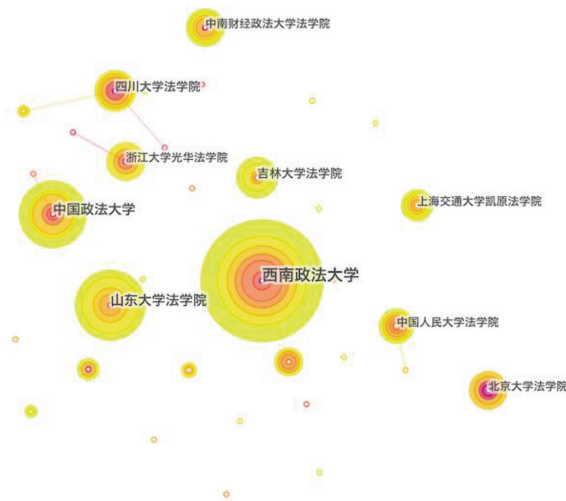


图2 发文作者所属机构分布图

与此同时,图中所展示的研究机构共现关系也表明,到目前为止,从事实证研究的科研机构之

<sup>①</sup> 由于相同研究机构中的研究人员之间联系较为紧密,方法的转换与适用通常可能受到彼此影响,对于研究方法转变的独立自主意识观察,应以研究机构为单位更加准确。由于文献中所涉及的具体研究机构,大部分均为隶属于相同机构,如所属同一高校的具体研究中心,为有效明确机遇裁判文书法学实证研究主体分布,我们对实属同一机构的不同下属机构进行了合并。

<sup>②</sup> 每个节点对弈一个研究机构,研究机构所对应的节点越大,说明该机构在研究成果中出现的频次越高。

间的相互合作甚少。<sup>①</sup> 对于法学实证研究而言,在大数据与人工智能理论与技术不断发展的背景之下,由关起门来做研究向合作共享的转变应是其实现迅猛发展并真正实现自身价值的正确趋向。<sup>②</sup> 而图中所揭示的“单兵作战”为主的法学实证研究现状,也在一定程度上进一步证实,实证研究方法在法学领域的应用尚处于初级阶段,研究人员对于法学实证研究发展趋向尚未形成准确而深刻的认识。

如图3所示,不同于研究成果数量的分布,开展基于裁判文书法学实证研究的机构并未随着裁判文书可得性的提升而出现急剧增长,而是在基于裁判文书法学实证研究兴起后较长一段时间内学术群体规模成缓慢增长之势。<sup>③</sup>

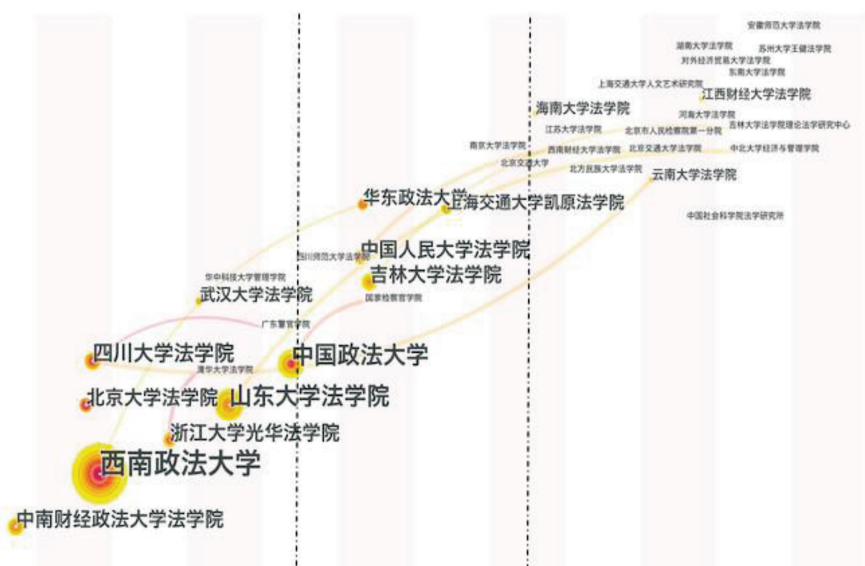


图3 研究机构时区分布图

具体而言,基于裁判文书法学实证研究的学术群体规模变化,同样可以分为较为明显的三个阶段。在第一阶段,参与研究之主体主要为西南政法大学、中南财经政法大学以及山东大学等总体分布结果所显示的本领域主要研究机构。结合总体分布结果可以得知,该部分研究机构不仅在本领域研究成果丰富,其开展本领域研究的时间跨度同样长于其他研究机构。该阶段中虽然也存在少量数量的其他研究机构,如广东警官学院、四川师范大学等,但从其所对应的节点大小来判断,该部分研究机构总体研究成果极少,说明该部分研究机构并没有持续在本领域开展研究。在第二阶段,新增研究机构数量总体较少,但在此阶段中所新增的研究机构,如吉林大学、中国人民大学以及华东政法大学等,都在总体上产出了丰富的研究成果,其他新增机构则在本领域的研究成果有限,从而未体现出研究的持续性。而在第三阶段中,即最近几年以来,新增研究机构数量较第二阶段明显增多。本阶段中,已有部分研究机构,如湖南大学、江西财经大学、云南大学等,以较其他新增机构

① 图中节点间的连接线表示机构之间存在共现关系,即两个机构研究人员曾共同署名发表论文。

② 参见左卫民:《迈向大数据法律研究》,载《法学研究》2018年第4期,第140-145页。

③ 时区分布图中的节点均表示所对应机构首次出现的年份,如该机构在后续年份再次出现,则仅在首次出现的位置增加相应频次,即表明,每个节点所对应机构均属于在时间变化过程中本领域新增研究机构。

在研究成果数量方面显现出一定的优势。<sup>①</sup>

相比于研究成果数量,学术群体的规模发展能更为准确地揭示法学研究者对于实证研究方法的认可与接纳状况。综合基于裁判文书法学实证研究机构数量增长的三个阶段来看,与研究成果数量随着裁判文书可得性增强而得到明显提升不同,本领域学术群体的扩展经历了一个较为漫长的过程。并且,结合研究机构总体分布所呈现出的集中趋势,可以证实,裁判文书网络公开数量的增长主要促成了原有研究机构的成果产出数量,而非其他研究机构对于实证研究方法的认可接纳,进而尝试方法转型。申言之,不同于研究成果数量在特定阶段与裁判文书网络公开数量之间明显相关,实证分析方法在法学领域的传播与发展表现得更加独立且缓慢。

### (三) 研究领域

在通过 citespace 软件所得出的关键词共现关系的基础上,<sup>②</sup>进一步进行关键词聚类分析,聚类结果显示 Q 值为 0.6556 且 S 值为 0.5119,说明分析结果中的结构划分是显著的、聚类结果具有可靠性。<sup>③</sup>如图 4 所示,依据所占比重大小,聚类分析共得到实证研究、量刑建议、法定赔偿、法治、司法改革以及司法鉴定在内的 6 个主要类别。单从类别名称来看,以学科划分为标准,基于裁判文书法学实证研究成果中涉及最广泛的“量刑建议”相关内容应属刑事诉讼领域,而广泛程度紧随其后的“法定赔偿”相关内容则属于知识产权领域。关于其他类别,法治、司法改革以及司法鉴定属于全局性或者说普遍性关键词,则需要通过分析本类别中的关键词共现关系,在明确本类研究成果论证基点的基础上,来明确其所在研究成果涉及之具体研究领域。故此,为明确各类别研究成果所归属的具体学科领域,我们继续结合各类别中所包含的主要关键词进行探讨。



图 4 关键词聚类图

① 由于该部分机构进入本领域的时间尚短,暂不宜评判其在本领域的研究是否具有持续性。

② 对于学术论文而言,关键词能够高度概括文中的研究内容和观点,因此,对于所选文献的关键词进行统计分析,可以有效地考察基于裁判文书的法学实证研究成果中所涉及的具体学科领域。

③ 一般认为,Q 值大于 0.3,则说明结构划分是显著的,S 值大于 0.5,则说明聚类结果具有可靠性。参见陈悦、陈超美、刘则渊、胡志刚、王贤文:《CiteSpace 知识图谱的方法论功能》,载《科学学研究》2015 年第 2 期。

如表一所示,综合聚类分析结果及对各类别中所包含主要关键词的解读,可以得知,基于裁判文书的实证研究成果中涉及最广泛的问题是刑事诉讼领域中的量刑问题,其次是知识产权侵权赔偿与法治问题,最后是司法体制改革及环境纠纷等阶段性热点问题。而且,对于法治问题以及司法改革问题的研究也主要基于刑事诉讼程序来展开,对于知识产权的研究中同样涉及刑事诉讼方面的具体问题,因此,完全可以确定,我国基于裁判文书法学实证研究所涉学科领域之中心在于刑事诉讼,对于其他领域的研究则极其缺乏。换言之,从总体分布上来看当前我国基于裁判文书法学实证研究之视域难谓宽阔。当然,我国法学实证研究尚处于发展进程中较为初级的阶段,研究者视域需要随着本领域的发展而不断开拓。

表1 聚类关键词表

聚类名称	主要关键词
量刑建议	量刑(9)、量刑建议(8)、缓刑(5)、死刑(4)、量刑规范化(4)、司法裁判(3)、污染环境罪(3)、量刑均衡(2)、量刑证据(2)、宣告刑(2)、罚金(2)、明确性(2)、速裁程序(2)、认罪态度(2)、故意杀人(2)、累犯从严(2)
法定赔偿	知识产权(9)、损害赔偿(7)、著作权(6)、刑事司法(6)、指导性案例(5)、法定赔偿(4)、案例指导制度(4)、侵权(2)、判例(2)、证据制度(2)、知识产权判例(2)、诉前调解(2)、专利侵权(2)
法治	调解(5)、受贿罪(5)、违法(3)、量刑基准(3)、法治(3)、诉讼调解(2)、瑕疵(2)、合同(2)、腐败(2)、标准(2)
司法改革	司法改革(8)、法院(6)、庭审实质化(6)、审判委员会(4)、以审判为中心(3)、量刑程序(2)、量刑说理(2)、检察院(2)、独立审判(2)、试点(2)
司法鉴定	司法鉴定(4)、环境司法(3)、环境纠纷(2)、专家辅助人(2)、鉴定人出庭(2)、新《环境保护法》(2)

注:括号内为对应关键词词频。

为进一步明确基于裁判文书法学实证研究所涉及的具体研究内容之发展,本文通过 citespace 软件得出关键词时区分布图。如图 5 所示,自基于裁判文书法学实证研究成果得到显著增长以来,<sup>①</sup>研究成果中所出现关键词的时区分布同样呈现出较为明确的三个阶段。2011 年至 2015 年之间,研究成果中所出现关键词的总量与年度增量均不大,甚至可以称之为匮乏。并且,本阶段所出现的为数不多的关键词中,半数以上均可明确被归类于刑事诉讼领域,如量刑程序、量刑、量刑建议等。2016 年与 2017 年,关键词数量急剧增长,呈爆发状态。关键词类型同样得到一定程度的丰富,知识产权、证券等领域归属明确的关键词出现,也说明基于裁判文书法学实证研究所涉及之领域得到拓展。由于在第一阶段末期,即司法改革的总体规划和具体方案陆续出台之际,<sup>②</sup>司法改革作为研究成果关键词出现,且其与第二阶段关键词存在密切的共现关系,即可认为,第二阶段研究关键

<sup>①</sup> 时区分布起始时间为 2011 年,即本次时区分布分析以研究成果数量得到本质提升后。

<sup>②</sup> 参见徐昕、黄艳好、汪小棠:《中国司法改革年度报告(2014)》,载《政法论坛》2015 年第 3 期,第 125-141 页。

词数量急剧增长在一定程度上得益于当前研究热点与基于裁判文书法学实证研究的密切关联性。<sup>①</sup>当然,与前一发展阶段一致,在本阶段之中,刑事诉讼领域关键词仍然大量存在,如刑事司法、死刑以及受贿罪等。2018 年至今,关键词数量急转直下,且呈逐年下降之势。由于 2018 年所出现之关键词主要与司法改革相关,如庭审实质化、以审判为中心等,且该相关性在 2019 年随之减弱,即可认为,该阶段关键词数量下降同样可在一定程度上归结于司法改革作为研究热点的消退。同时,这也可以在一定程度上对近年来参与本领域研究的研究机构数量增多而研究成果数量反而下降之现象予以解释。即随着研究热点消退,研究者们的个体产量逐渐下降,即便有新的研究机构加入,也无法阻止研究成果总量的下降。

根据对关键词时区分布的分析,可以得知,基于裁判文书法学实证研究所涉之学科领域在较长的发展进程中确实在一段时间内得到明显拓展,然而,由于该拓展本身具有一定的偶然性,学科领域的拓展进程随着机遇的丧失而停滞,进而表现出退缩之势。与此同时,结合前文关键词总体分布结果,由于在基于裁判文书法学实证研究发展的各个阶段均绕不开对于刑事诉讼领域问题的专门研究,以及以刑事诉讼实践为基点对相关问题的分析探讨,故难以认定基于裁判文书法学实证研究者的视域得到了本质性拓展。

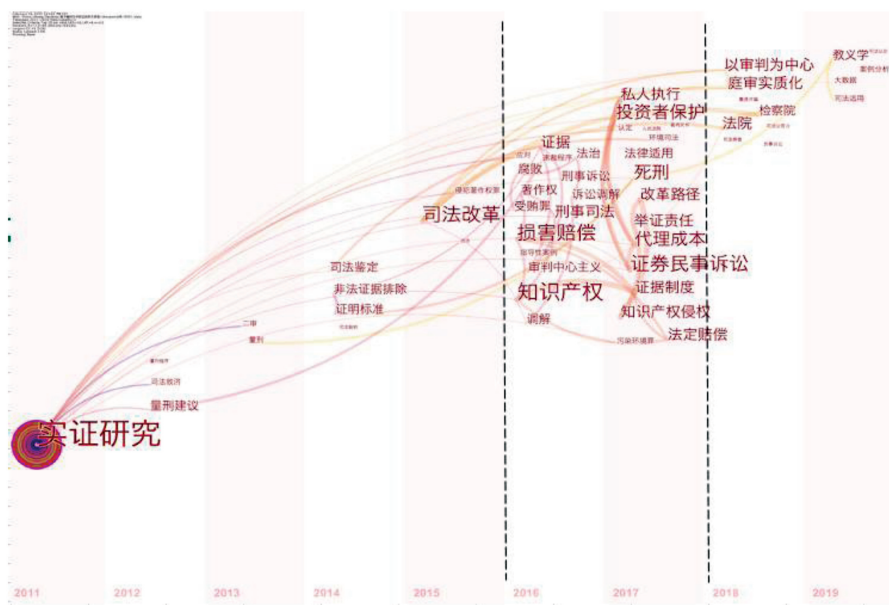


图 5 关键词时区分布图

而从理论上来说,实证研究作为一种独立的研究方法,其所针对的研究对象具有广泛性,即便是将其严格限定在以具有数量变化关系的法律现象为主要研究对象的定量研究范围之内,该广泛性依然存在。<sup>②</sup>实证研究方法要得到发展并实现其自身价值,需要立足于习惯传统研究方法的学者

<sup>①</sup> 裁判文书作为司法过程的产物,基于裁判文书所开展的实证研究所主要能揭示与论证的即为司法实践相关之问题。

<sup>②</sup> 屈茂辉:《计量法学基本问题四论》,载《太平洋学报》2012年第1期,第26-33页。

对其自身的认可与接纳,进而主动对实证研究方法的研习。<sup>①</sup>然而,对于实证方法在本领域的适用,不同法学研究领域学者所持坚守教义学阵地之信念有所差异。<sup>②</sup>实证研究方法的适用最早开始于刑事司法领域,其在该领域的认可与接纳程度较高。而以民法领域为代表的其他领域学者,对于实证研究方法在本领域的适用均在一定程度上持保守态度,即便是认为实证方法对本领域研究有所裨益的学者,也认为法律实证研究方法只能针对个别问题的、点对点的零碎研究方式,只能为法教义学提供体系建构、融合之素材,无法担负起法律体系建构与融合的任务,“扛大旗者”非教义学方法莫属。<sup>③</sup>故而,法学实证研究的研究领域,可能受传统研究理念的影响,在学科分布上存在一定的局限性。与此同时,在当前学界推崇使用统计分析模型进行定量研究的趋势下,具有数量关系的法律现象是我国法学实证研究的重要研究对象。该类法律现象中,部分数量关系表现得较为直观,可在相关的资料信息载体中直接体现出来,研究人员可通过直接观测来实现对于数量关系的萃取与分析;另一部分数量关系,则并在资料信息载体中得以直观显现,需要研究人员投入更多的成本,并借助一定的技术方法来实现对于数量关系的抽象与提取,即对研究人员对于方法的理解及方法适用能力要求高于直观数量关系。<sup>④</sup>而数量关系直观程度的差异,在裁判文书之中同样客观存在。其中刑事审判结果中所体现的法定刑与宣告刑之间数量关系,即是直观数量关系的典型。对于意欲从事实证研究以及刚刚步入法学实证研究领域的学者而言,成本投入大小以及方法适用的难以程度是其开展实证研究的重要考量因素。申言之,对于裁判文书之中数量关系抽象的难易程度,也在一定程度上影响基于裁判文书法学实证研究所涉足的具体学科领域。

## 二、基于裁判文书法学实证研究演变内因剖析

尽管被视为法学实证研究发展契机的裁判文书公开数量提升,能够在一定程度上对特定时期研究成果数量的发展变化之缘由予以解答,但对于基于裁判文书法学实证研究的总体概况,尤其是发展过程中所存在的具体问题,则明显缺乏解释力。故此,在当前法学方法转型的外部环境之下,对于基于裁判文书法学实证研究发展演变之内因,还需进一步基于方法论视阈予以阐释。具体而言,依据以数据资料收集方法为标准所进行的社会科学研究方法划分,基于裁判文书法学实证研究中所适用之主要方法为内容分析法,即对裁判文书中所承载的内容做客观的、系统的量化描述和分析。<sup>⑤</sup>对基于裁判文书法学实证研究发展演变内因之阐释,应围绕内容分析法自身的优势与缺陷来展开。

### (一)从数据可得性方面的分析

适用内容分析法开展实证研究,需要在获取原始数据载体的基础上,进一步提取可为分析所用

<sup>①</sup> 在当前法学教学模式下,法学生所接受的研究方法教育仍然以教义学方法为主,且其开展研究学术研究必然受师长之影响。因而意欲从事学术研究的法学生对于实证方法的理解与认知同样具有局限性。

<sup>②</sup> 参见陈柏峰:《法律实证研究的兴起与分化》,载《中国法学》2018年第3期,第132-149页。

<sup>③</sup> 参见金可可:《民法实证研究方法与民法教义学》,载《法学研究》2012年第1期,第48-50页。

<sup>④</sup> 参见屈茂辉:《民法实证研究中的计量方法》,载《法学研究》2012年第1期,第52-53页。

<sup>⑤</sup> Earl R. Babbie, *The practice of social research*, Nelson Education, 2015, p. 323-324.

的标准化数据。因此,对于数据可得性,需要从原始数据可得性与标准化数据可得性两个层面予以阐述。

### 1. 原始数据可得性

实地调查、访谈、问卷调查等其他较为常用的社会科学研究方法,<sup>①</sup>均需要预先制定较为复杂的数据获取方案,即如何对研究所需数据进行获取并得以客观记录,并需要组建相对庞大的研究团队,<sup>②</sup>才能够切实有效地获得开展后续研究所需要的原始数据,进而对原始数据进行编辑整理获得研究所需的最终数据。对于内容分析法而言,在开展数据整理与分析之前,研究所需的具体数据已经被客观地记录于相应载体之中,研究人员仅需要直接对相关数据载体进行获取,即可在一定程度上被认为已经实现对于研究所需原始数据的获取。申言之,通过比较内容分析法与其他类研究方法的数据获取过程可知,内容分析方法在原始数据获取方面具有较为明显的便捷性。当然,在方法适用过程中,内容分析法较之于其他分类方法的便捷性优势能否得以体现,还需取决于数据载体获取的便捷程度。如若数据载体获取足够便捷,内容分析法的实施甚至可以无需组建研究团队,而由一人独立完成数据获取工作。<sup>③</sup>

对于仍然处于探索起步阶段的法学实证研究学者而言,研究成本的降低,即意味着研究开展可行性提升。据此,前文概况检视部分所揭示的,基于裁判文书法学研究成果数量在裁判文书公开制度施行初期的显著增长得以解释。即该阶段基于裁判文书法学研究成果数量,源自于裁判文书网络公开对于研究开展便捷性的本质提升。

### 2. 标准化数据可得性

对于大部分研究方法而言,在开展具体的数据分析之前,都需要经历由原始数据向标准化数据的转化。内容分析的基本做法是对数据载体中的信息进行编码,即根据特定的概念框架,对载体中的信息进行分类记录,以获得标准化数据。<sup>④</sup>普遍意义上而言,内容分析法与其他方法在标准化数据可得性方面并不存在明确差异。然而,随着计算机与信息科学等相关领域的迅速发展,借助于技术手段的应用,社会科学研究中数据提取的效率得以显著提升,但基于待提取数据类型之间的差异,技术手段应用所产生的效果也并不一致,并最终在一定程度上导致研究进程中数据提取效率的高下之分。

就数据结构类型而言,数据形式主要可分为结构化数据、非结构化数据与半结构化数据三个类型。<sup>⑤</sup>其中结构化数据,即包括预定义的数据类型、格式和结构的数据,通常以二维表形式保存;半结构化数据系指具有可识别的模式并可以解析的文本数据文件,通常情况下是指 XML 数据文件;而非结构化数据则是没有固定结构的数据,通常保存为包括文本文档、图像和视频在内的不同类型

<sup>①</sup> 现阶段,我国法学实证研究中主要适用的研究方法,除内容分析法外,主要为实地调查、访谈、问卷调查以及二手数据分析等。(参见陈柏峰:《法律实证研究的兴起与分化》,载《中国法学》2018年第3期。)

<sup>②</sup> 当然,借助于互联网技术的发展,访谈、问卷调查法等研究方法对于研究团队人数规模的要求正在逐步降低。

<sup>③</sup> Earl R. Babbie, *The practice of social research*, Nelson Education, 2015, p. 334.

<sup>④</sup> Earl R. Babbie, *The practice of social research*, Nelson Education, 2015, p. 328-329.

<sup>⑤</sup> 也有学者认为还存在第4种数据结构类型“准”结构化数据,即通过使用工具可以使之格式化的具有不规则数据格式的文本数据。(参见周苏、王文编著:《大数据导论》,清华大学出版社2016年版,第10-11页。)

的文件。由于结构化数据自身具备固定的类型、格式与结构,在进行数据分析之前并不需要对结构化数据进行过多处理,甚至可直接被用于数据分析与挖掘。但其他两种类型数据,尤其是非结构化数据,在进行数据分析与挖掘之前,往往需要对其进行深入处理使其转变为符合分析要求的结构化数据。现阶段,裁判文书基本以纯文本形式公开,即属于典型的非结构化数据。

而实际上,在我们现阶段所处的大数据时代,作为数据处理的重要内容,如何将非结构化数据与半结构化数据转化为可供使用的结构化数据不仅是大数据技术领域的关键技术之一,同时是大数据技术所面临的关键问题之一。具体到计算机技术实现层面,数据处理,尤其是文本内容抽取的有效实现,在极大程度上依赖于自然语言处理技术的发展与应用。<sup>①</sup>通过让计算机“理解”自然语言,从而实现对以文本数据为典型非结构化数据的信息提取,进而实现非结构化向结构化的转变。然而,就当前研究与应用成果来看,自然语言处理技术对于法学研究中的数据处理能够提供的支持十分有限。<sup>②</sup>裁判文书作为典型的非结构化数据,当然也面临着数据转化方面的技术难题。而当缺乏有效技术支持时,数据量级的增长,即意味着人力成本增加,<sup>③</sup>很显然,这是目前以裁判文书为实证研究数据来源的一个重要缺陷。

因此,对基于裁判文书法学实证研究成果数量在2019年的急剧下降,我们可以寻得部分合理解释。<sup>④</sup>随着计算科学等相关领域的快速发展,大数据不断被各领域所推崇,法学领域也不例外。近年来,法学领域不断有学者强调在大数据在法学领域的适用,并在强调全样本优越性的基础上提倡获取全面多样数据。<sup>⑤</sup>在当前裁判文书公开数量不断攀升的情势之下,获取全面多样数据,即意味着数据量级的增长,进而使得研究成本大幅提升。在研究方法转型的背景之下,相对于传统研究方法对于文献资料的梳理,大数据获取对于人力物力需求的明显增加,可能使得基于裁判文书法律实证研究的开展面临困境,也可能使得意欲尝试研究方法转型的研究人员望而却步。在“小数据”研究成果认可程度降低,“大数据”研究开展受阻的情境之下,实证研究成果的下降不可避免。

## (二)从数据完整性方面的分析

数据分析之目的,在于从直观数据之中,提取出有用的信息和知识。毋庸置疑,数据的完整性可以影响到所提取信息和知识的价值与厚度。涉及到目的性较强的数据分析,如果关键性数据的缺失,甚至可能导致分析目的无法实现,从而进一步导致整体研究方案的可行性丧失。对于完全依赖于既有数据资料的内容分析法而言,由于可获取数据的完整性不由研究人员所掌控,<sup>⑥</sup>可获取数据的完整性程度不仅会影响到研究结果的可靠性及价值,甚至导致研究目的难以实现。而具体到

① 参见姜岩:《大数据技术概述》,清华大学出版社2017年版,第10-11页。

② 即便是使用计算机辅助技术进行数据处理,其所提取出来的也局限于结构性较强的字段。(参见深圳国泰安教育技术股份有限公司大数据事业部群、中科院深圳先进技术研究院——国泰安金融大数据研究中心:《大数据导论关键技术与行业应用最佳实践》,清华大学出版社2015年版,第5-6页。)

③ 据了解,国内几乎所有的研究者均以人力逐份阅读裁判文书,美国哥伦比亚大学李本(Benjamin L. Liebman)教授在进行基于中国裁判文书的研究中则与其大学的计算机方面的教授团队合作,希望借助计算机对裁判文书进行分析。

④ 另一个可能在原因恐怕在于近些年尤其是2018年以来,绝大多数法学期刊原则上不刊发合作作品,而实证研究几乎都是多人合作才能完成的工作,如果据实写明作者则难以发表。

⑤ 参见左卫民:《迈向大数据法律研究》,载《法学研究》2018年第4期,第139-150页。

⑥ 如调查研究法,则具有较大的灵活性,研究者可以通过在研究主题范围之内增加问题的数量来提升所获取数据的完整性。(See Earl R. Babbie, *The practice of social research*, Nelson Education, 2015, p. 279.)

基于裁判文书的实证研究中,由于现阶段我国裁判文书公开所存在的缺陷,使得数据获取缺乏完整性成为内容分析方法相对于其他方法所表现出的主要劣势之一。

从我国现阶段数据公开来看,公开数据的完整性依然存在较大的问题。基于裁判文书法学实证研究自身,则在原始数据与标准化数据获取两个层面都可能面临数据完整性问题。从原始数据层面来看,目前裁判文书网公开的裁判文书数量与当年实际审结案件数量仍然存在较大差距,<sup>①</sup>更为重要的是,裁判文书公开比例与案件类型、审判层级以及地域等因素均存在极为明显的相关性。<sup>②</sup>而从标准化数据层面来看,依据裁判文书制作规范,裁判文书主要对案件审理相关之要素进行记载,对于案件审理过程中所涉及的部分案件信息,在裁判文书制作过程中不可避免地需要进行简化甚至省略。并且,虽然现行规范对于裁判文书的撰写进行了较为明确的规定,但基于法院即审判人员自身的主观原因,不同审判人员、不同法院之间出具裁判文书中所记载的具体内容可能存在差异,部分审判人员依据其主观意愿进行裁判文书制作时,即便是必需承载的案件要素信息同样可能被其进一步简化甚至省略。

裁判文书在总体上公开的不完整性,导致公开数据与真实数据总体存在非一致性,从而进一步导致基于不完整公开数据进行全样本或抽样分析所得出的研究结果与真实情况存在偏差,从而降低总体研究的可靠性。与此同时,裁判文书公开比率过低,会在一定程度上实证研究可用数据总量造成限制。<sup>③</sup>而公开比率与案件类型等之间所存在的明显相关性,则会导致拟适用具体案件类型裁判文书开展实证研究的可行性差异。回归到前文所论述的研究领域中,由于案件类型与学科领域分布密切相关,公开比率与案件类型相关性所导致的可行性差异,最终将表现为研究成果在学科领域部分上的差异性。申言之,对于前文所揭示的研究领域分布集中趋势,除教义学方法坚守信念以及方法适用难易程度外,还可能受当前裁判文书公开缺陷之影响。

### (三)从数据客观性方面的分析

内容分析法还具有另一个重要特性,即其无需对研究对象进行介入,无需与研究对象交流与互动。实际上,基于内容分析法无需介入研究对象的特质,其也被划归为“非介入性研究法”。而非介入性研究法的优势,则在于,可以避免来自研究对象的主观因素干扰,进而确保所获取之数据的客观性。<sup>④</sup>

具体到基于裁判文书法学实证研究之中,其研究对象所涉及之个体主要以法官为主。任何职业都会或多或少地伴随着一定的职业风险,而法官职业风险则来源于裁判权运行中对外或者对内发挥作用是对法官自身产生的负面影响,这种负面影响则可以表现为对法官的心理压力、身体损害、各种形式的责任承担等等。在裁判权行使过程中,事实认定、适用法律、裁判可改变性及社会质

<sup>①</sup> 在涉及隐私、国家安全等事项时,裁判文书便不公开,因此,在实际操作中,各地法院对于裁判文书的公开与否都各有自己的做法。

<sup>②</sup> 参见杨金晶、覃慧、何海波:《裁判文书上网公开的中国实践——进展、问题与完善》,载《中国法律评论》2019年第6期,第125-147页。

<sup>③</sup> 有学者在其研究成果中明确表述了目前裁判文书网数据公开缺陷对其开展研究,尤其是数据获取方面造成的限制与影响。参见乔仕彤、毛文铮:《行政征收的司法控制之道:基于各高级人民法院裁判文书的分析》,载《清华法学》2018年第4期,第68-70页。

<sup>④</sup> Earl R. Babbie, *The practice of social research*, Nelson Education, 2015, p. 334.

疑等都可能引发相应的法官职业风险。<sup>①</sup> 纵观我国司法体制改革历程,制度设计所体现出的司法体制改革理念则主要在于防范权力滥用,而非在注重法官职业保障的基础上保障权力实施。<sup>②</sup> 正确行使裁判权可以有效降低法官所面临的执业风险,但基于司法裁判本质,裁判过程中的事实认定与法律适用均具有一定的不确定性,裁判结果并不能完全避免错误出现。故此,对于我国法官而言,面对来自司法责任制与内部考核制度方面的压力,以及裁判错误的不可避免性,出于规避风险之目的,其必然不希望裁判行为被过分研究和解读,以产生对其自身的不利的负面影响。申言之,对于任何研究对象的介入,都可能受主观因素之干扰,但在我国法官所处的制度环境之下,介入性研究受到干扰的可能性及程度相对更高。

为尽可能避免受到来自研究对象的主观因素干扰,研究人员在适用介入性研究方法获取数据时,需要在全面了解可能出现之干扰因素的基础上,不断调整与完善数据采集方案。以问卷调查法为例,为尽量保证数据客观性,问题本身是否清晰、问题长度是否适宜、问题本身是否包含偏见都需要根据研究目的及研究对象的具体情况进行设计与调整。<sup>③</sup> 对于知识体系相对单一的国内法学研究者而言,在缺乏系统实证研究方法教育的情形下,基于介入性研究方法所获取数据的客观性必然更加容易受到质疑。质言之,适用内容分析法所获取数据的客观性更容易得到认可,基于客观数据所得出之研究结论必然更易为评阅者所接受。据此,研究成果数量与裁判文书网络公开数量增长的高度契合,除裁判文书可得性外,还应当受到研究成果接受程度之影响。即,较高的接受程度,使得基于裁判文书法学实证研究的直接产出更易转化为公开发表的研究成果,从而缩短裁判文书网络公开数量增长与研究成果数量增加之间的时滞。

当然,在当前强调“大数据”,且数据提取技术尚未成熟之时,基于裁判文书法学实证研究同样可能面临数据客观性方面的问题。具体而言,研究人员在进行编码的过程中,对于编码所需的特定概念框架,需要通过理解数据载体所承载的具体内容进行抽象与确定,而基于主观认知的差异,以及阅读顺序差异等外在因素,对于具体信息内容的理解,在不同的研究人员之间则可能会存在一定的差异,进而影响到编码结果的统一性与客观性,当数据载体所承载的内容主观性非常强时,此类问题则会更加凸显。<sup>④</sup> 申言之,当一项适用内容分析法的研究需要由多个研究人员同时进行数据编码工作时,同样可能引发数据客观性问题。

### 三、基于裁判文书法学实证研究推进路径的思考

囿于传统研究方法在研究对象方面所存在的局限性,以法律现象中的数量关系为主要研究对

<sup>①</sup> 参见尹茂国:《裁判权让渡:一个不得已的选择——基于刑事法官职业风险化解视角的思考》,载《法律科学(西北政法大学学报)》2016年第5期,第90-98页。尹茂国:《裁判权让渡:一个不得已的选择——基于刑事法官职业风险化解视角的思考》,载《法律科学(西北政法大学学报)》2016年第5期。

<sup>②</sup> 参见武健:《惩戒制度设计中的法官权益保障》,载《法律适用》2017年第7期,第26-30页。

<sup>③</sup> Earl R. Babbie, *The practice of social research*, Nelson Education, 2015, p. 248-254.

<sup>④</sup> 参见林聚任、刘玉安主编:《社会科学研究方法》,山东人民出版社2004年版,第147页。

象的实证研究方法,因有助于全面把握法学研究对象,故有其在法学领域存在与发展之价值与必要。<sup>①</sup>因此,在明确发展概况及与其相关影响因素的基础上,为谋得基于裁判文书法学实证研究之长远发展,除针对数据公开及技术发展方面的宏观政策与制度支持外,研究人员同样应当以研究方法为基点,在研究过程中积极主动寻求针对当前所处困境的破局之道。

### (一) 丰富数据来源

现阶段,我国法学实证研究水平尚待提高,尤其是在团队建设观念及团队协作能力都亟待提高的情况下,利用公开数据进行实证研究仍然是最为行之有效且可靠性最强的研究方式。具体到基于裁判文书法学实证研究中,基于公开数据对研究方法的完善,即可主要表现为,在坚持适用内容分析法的基础上,以裁判文书所承载内容为依据,通过广泛搜集其他公开数据载体来充实可用于实现研究目的的数据类型与总量。以基于裁判文书法学实证研究成果中所涉及主要研究内容中的司法鉴定而言,裁判文书中通常仅对出具鉴定意见的司法鉴定机构以及司法鉴定人名称进行简单记载,而更为详细的信息则均由司法鉴定行政管理部门进行登记与公开。如若需要研究司法鉴定机构以及司法鉴定人所具有特性对鉴定意见审查结果的影响时,除去裁判文书中对于鉴定意见审查程序及结果的记载,司法鉴定行政管理部门所公开的司法鉴定机构及司法鉴定人登记信息也是实现分析目的的重要数据来源。并且,由于该部分登记信息均得以通过互联网渠道公开,对于项目开展所需人力成本的影响也十分微小,因而该种研究方法完善途径具有较高的可能性。

当然,基于内容分析法自身的局限性,以及我国现阶段的信息公开制度,并非所有与裁判文书相关数据均能从公开渠道获得。此时,对于研究方法的完善则意味着对于其他研究方法的引入。例如,当研究目的的实现需要分析审判人员背景信息,尤其是涉及审判人员性格特征等非公开资料能获取的信息时,裁判文书不会对该部分信息予以记载,而对于该部分信息的获取需要对审判人员进行介入,此时,内容分析法作为非介入性研究方法,当然无法再有效满足分析需求。在此种情况下,研究人员则可以借用问卷调查法或观察法,对于该部分数据资料进行补充。具体而言,研究人员可通过向具体审判人员发放问卷,或对审判人员进行观察访谈,获取审判人员的背景特征信息,然后再结合承载该部分审判人员具体审判过程的裁判文书实现最终的研究目的。<sup>②</sup>诚然,由于其他研究方法的数据获取并没有相对于裁判文书可通过统一平台进行获取的便捷性,其他方法的引入,对于整个研究项目的影响并非简单的导致研究成本增加,而是直接影响抽样方案的制定,进而造成整个研究方案的变化。<sup>③</sup>

综上,对于裁判文书内容承载局限性所造成的阻碍,研究人员可以通过完善研究方法的形式予以弥补,对于研究方法的完善,以扩充数据类型与总量为目的,可以内容分析法为基础,通过公开渠道对研究所需数据进行扩充,也可以内容分析法为中心,引入其他研究方法。而具体完善方案的选择,则需要对研究目的实现、成本投入等多方面因素进行考量。

<sup>①</sup> 参见屈茂辉、张杰、张彪:《论计量方法在法学研究中的运用》,载《浙江社会科学》2009年第3期,第21-27页。

<sup>②</sup> 已有学者在实践中使用此弥补方法。(参见高海:《农村宅基地上房屋买卖司法实证研究》,载《法律科学(西北政法大学学报)》2017年第4期,第180-189页。)

<sup>③</sup> 在单纯以围绕裁判文书获取公开数据开展实证研究时,由于数据获取的便捷性,往往可以实现对于全国范围内的数据抽样,甚至是全样本分析,但如要引入实证调查或问卷调查等方式,基于研究方案可行性方面的考量,抽样范围往往需要进行限缩。

## (二) 科学适用实证研究方法

对于适用内容分析法的实证研究而言,除定量分析中对于统计分析方法的理解与适用较易出现问题外,<sup>①</sup>如前文所述,数据的采集与提取更可谓“雷区”。故此,谋求基于裁判文书法学实证研究的长远发展,必然无法绕开对于研究方法适用的规范。与此同时,由于知识体系集体单一,所接受的研究方法教育均以教义学方法为主,法学研究者关于实证研究方法的知识储备匮乏,缺乏系统化研究思维。对于研究方法适用的规范,应以提升研究人员对于方法本身的理解与认知,并实现期思维从定性到定量之转换为前提。

### 1. 规范研究方案的制定。

任何研究项目的顺利开展,都需要事先制定一个完整的研究方案,对研究实施的细节问题和实施步骤进行明确。对于基于裁判文书的实证研究而言,如上文所述,如何正确采集裁判文书,以及如何准确地从裁判文书中抽象与提取数据,都需要在项目实施前予以明确。尤其对于需要团队协作的项目,事先制定详细的研究方案对于提升项目成员执行项目的规范性与统一性尤为重要。具体而言,基于裁判文书的实证研究方案,除研究方案普遍应当具备之内容外,还应着重强调以下几点。第一,裁判文书抽样方案,包括抽样范围与比例的确定,以及在裁判文书总体不可得的情况下,<sup>②</sup>通过制定适当的抽样办法以确保抽样的随机性与准确性。第二,数据编码规则,如前文所述,内容分析法最为基本的做法即为对载体中的信息进行编码,事先制定详细的数据编码规则有利于研究人员在对裁判文书内容进行抽象的过程中概念统一,从而提升数据提取的有效性。而在量化实证研究之中,数据编码即定性内容向定量数据转变之关键。第三,明确团队成员分工,尤其在数据采集与提取过程中,应当明确各个研究人员的职责范围,以避免造成数据采集与提取的重复或遗漏,提升项目执行之效率。

### 2. 提升团队协作能力与效率。

如前文所述,由于辅助技术对于法学实证研究,尤其是对于数据处理方面的支持十分有限,现阶段基于裁判文书的实证研究开展,仍然需要以人工阅读与整理的形式完成裁判文书内容的提取,以及将定性内容转化为定量数据。然而,传统的法教义学研究,对于团队运行的形式并不十分依赖,通常只需要较少的研究人员,甚至是单独的研究人员即可形成具体的研究成果。从传统研究方法向定量的实证研究方法的转变,尤其是向需要耗费大量人力资源的裁判文书分析研究来说,如何适应团队协作,并提升团队协作与效率,同样是现阶段谋求基于裁判文书的法学实证研究发展所需解决的问题。具体而言,实质开展研究项目前,在研究方案中对团队成员分工进行明确的基础上,还需要在项目开展前对研究人员进行培训,并在项目开展过程中注重研究团队之间的沟通交流。而从长远发展来看,组建较为固定的研究团队,则是提升协作能力与效率的重要基础。

### 3. 重视研究结果的检验。

社会科学研究中任何一种精确的、系统的收集资料的方法,实际上都是一种特定形式的社会测量,而为了衡量收集资料的准确性,以及依据该部分数据资料分析所得的研究结论是否具有稳定

<sup>①</sup> 参见张永健:《量化法律实证研究的因果革命》,载《中国法律评论》2019年第2期,第135-141页。

<sup>②</sup> “中国裁判文书网”对于裁判文书的下载数量进行了限制。

性,任何研究成果均需接受信度与效度的检验。<sup>①</sup> 法学作为一门社会科学,内容分析法作为一种特定的资料搜集方法,基于裁判文书的实证研究结果必然也需要接受信度与效度检验。尤其是在多人协作进行裁判文书内容提取的时候,为确保研究结果尽量少的受到数据提取瑕疵或错误的影响,对于研究结果的检验显得更加必不可少。对于使用通过抽样所获得小数据进行分析的实证研究而言,抽样方法的合理性则为影响最终研究结果稳健型的关键因素之一。在当前国内法学研究群体普遍缺乏实证研究系统知识背景及方法适用经验的背景下,研究人员不应对自身所开展的研究及所获得之结果过分自信,结果的稳健型必然应以检验结果为依据。研究成果受众,无论是期刊编辑还是政策制定者,对于研究结果可靠性的判断同样应以检验结果为标准。然而,从现阶段已有的基于裁判文书的研究成果来看,至少在期刊论文之中,极少存在研究人员对研究结果进行稳健型检验。<sup>②</sup> 而基于前文对研究成果数量变化趋势的分析,可以推测,当前法学实证研究领域对与研究结果检验的忽视,可能导致了不利于本领域发展的极端现象,即将研究数据的“量”与研究结果的“质”进行绝对性关联,进而导致学界对于“量”的过分推崇,并最终导致近年来法学实证研究发展步伐放缓。为此,为提升基于裁判文书的实证研究成果在法学研究领域的认可度,以谋求法学实证研究的长远发展,研究人员应重视对于研究成果信度与效度方面的检验,无论是全样本“大数据”,还是抽样“小数据”,从纯粹的学术研究角度而言,<sup>③</sup>只要方法适用得当且研究结果符合稳健型检验要求,即应当被认可。

### (三)推动数据处理技术的发展

从本质上来讲,提升裁判文书内容提取效率的根本之策应在于加速相关数据处理辅助技术的发展。对于裁判文书内容提取而言,核心的数据处理辅助技术即为自然语言处理技术。而自然语言处理技术的发展,及其在法学研究领域的有效应用,则是法学领域研究人员可以切身参与的。

数据处理技术的核心在于自然语言处理,而自然语言处理作为一项跨学科研究领域,其在特定领域范围内的研究与应用必然离不开本领域专家学者的参与。自然语言处理研究早期,基于规则的自然语言处理方法,需要依靠本领域专家耗时数月或者更长时间总结并参与编写语法规则,而机器学习以及现阶段推动了自然语言处理技术飞速发展的深度学习方法则需要基于大量本领域数据进行长时间地训练才能不断提高数据处理的准确率,机器训练的过程同样需要本领域专家的参与,以更好地引导机器做出涉及本领域专业知识的正确判断。<sup>④</sup>因而,要实现自然语言处理技术在特定领域的应用,则不可避免地需要本领域专家参与应用程序研制及开发。对于研发机构而言,在应用层面,基于训练过程的大量投入,以及应用的领域性,研发人员不得不将产出作为应用程序开发的一个重要考量因素。相对于诸如金融等高收益、高回报领域,机器训练阶段较大的人力成本投入与预期产出的失调,则正是阻却自然语言处理领域研究成果在法律行业有效应用的主要原因之一。<sup>⑤</sup>

<sup>①</sup> 参见林聚任、刘玉安主编:《社会科学研究方法》,山东人民出版社2004年版,第95-96页。

<sup>②</sup> 在实证分析较为成熟的其他社会科学研究中,如经济学等,已发表文章中基本均详细记载了研究结果检验过程。而在法学实证研究成果中,基本没有专门对此部分内容予以陈述。

<sup>③</sup> 我们并不否认大数据对工业社会发展所起到的巨大推动作用。

<sup>④</sup> 参见王飞、陈立、易绵竹等:《新技术驱动的自然语言处理进展》,载《武汉大学学报(工学版)》2018年第8期,第669-678页。

<sup>⑤</sup> 参见左卫民:《关于法律人工智能在中国运用前景的若干思考》,载《清华法学》2018年第2期,第108-124页。

因而,法学研究学者作为本领域的专业人员,如果能通过其自身对于自然语言处理技术与应用产品研发过程的参与来降低法律行业应用程序研制开发成本,则能有效降低研发机构在投入产出方面的顾虑,进而促进研发人员在从事法律行业领域自然语言处理技术与应用产品研发方面的积极性,并最终导致法律行业领域自然语言处理技术的提升及相应应用产品的出现。

通俗来讲,单纯的基于自然语言处理技术的法学研究辅助工具及软件研发并不能为投资者及研发人员带来可观的经济收益,但自然语言处理技术本身确是目前蓬勃发展且商业利益巨大的高阶人工智能产品开发之基石。由于基于自然语言处理技术的法学研究辅助工具及软件的应用可以间接地吸纳较多地本领域专家对于机器训练过程的参与,<sup>①</sup>如果可以创建一种双赢模式,使法学研究人员及软件开发主体均能在技术应用中获利,即可有效促进数据处理技术在法律行业的有效利用,进而推动基于裁判文书法学实证研究的进一步提升与发展。具体而言,软件开发主体可先投入少量成本研发出具备一定数据处理能力的软件产品,并以开源形式供本领域专家学者做学术研究之用,从而使得专家学者可以在学术研究过程中获得计算机技术的辅助与支持,而软件开发人员则可以通过专家反馈实现对于机器的训练,训练完成之后则可进一步用于开发更加高阶的人工智能产品。

## Review on Judgment Documents Based Empirical Legal Research

*QU Mao-hui*

(Hunan University, Changsha 410082, China)

**Abstract:** The knowledge map analysis results show that the outcome number of judgment documents based empirical legal research has generally increased, but has declined significantly in the past two years. And the scale of academic community is gradually expanding, but the distribution of research results among research groups is relatively concentrated. Also, the subject areas involved in the research results have obvious limitations, and the specific research content has begun to shrink after a brief outbreak of hot issues. In the context of method transformation, the evolution of judgment documents based empirical legal research is affected by both the advantages and disadvantages of research methods in terms of data availability, data integrity, and data objectivity. Researchers themselves should seek for long-term development of judgment documents based empirical legal research through enriching research data, standardizing the application of research methods, and promoting the development of data processing technologies.

**Key Words:** knowledge map; judgment documents; empirical research; content analysis method; data processing

本文责任编辑:董彦斌

青年学术编辑:周力

<sup>①</sup> 即便是该部分辅助软件在数据提取准确性方面可能存在一定的欠缺,但在辅助工具获取数据的基础上对数据进行补正完善,也能从本质上减轻法学实证研究开展过程中的人力成本投入。