

## 大数据时代司法裁判的路径探索 ——以大数据分析证明方式的提出与规范为视角

李 茜\*

---

**内容提要：**大数据分析证明方式既能拓宽法官知识治理的路径，又能从司法运行角度反哺证据规则的设计与完善，还能倡导公正司法的时代理念，提升法官对证明机理的综合运用。大数据分析证明方式在司法裁判中不断试错，总体思维、容错思维和相关思维的缺失是大数据分析证明失范乱象的主要成因。以大数据分析证明方式的关联性为焦点，可从三方面规范其在事实认定中的运用路径：一是从人、事、物、时、空的五维向度来规制载体的关联性，二是从法官对证明方式的合理评价与科学采信层面来设计内容的关联性，三是从大数据预测的有限引入角度来探析大数据相关性的司法运用，以期为法官精准认定事实和科学高效裁判提供智识支撑。

**关键词：**大数据分析证明方式 载体关联性 内容关联性

---

“证据领域不过是知识领域。”多知识学科交叉已成为了证据法学的研究理论品格，新的证据或证明方式如果能有助于查明案件事实，即能成为证据法研究的关注视点。在DT时代，以大数据为代表的智能力量为司法裁判注入了智慧因素，也进一步充实了证据科学与证明机理的理论品格。伴随着“证据学第三浪潮”，<sup>[1]</sup>大数据分析证明方式在司法裁判领域纵深延展，并在事实认定场域中不断地试错论证。大数据是数据、技术与思维的要素联动，然而囿于总体、容错、相关思维的缺失，大数据分析证明方式在法官事实认定过程中乱象丛生。本文以人工智能时代的司法裁判为背景，检讨法官在事实认定中的思维范式转型，并以关联性为焦点探析大数据分析证明在法官事实认定中的路径构建，以期为法官精准裁判提供智识支撑。

\* 湖南省永州市零陵区人民法院审判员。

本文为国家社会科学基金项目“司法裁判过程中的人工智能应用研究”（18BFX008）的阶段性成果。

[1] 王跃、易旻：《迈向“证据科学”——法庭科学学科建设模式的“大证据学”视野》，载《法制与社会发展》2011年第3期。

## 一、启示：大数据对司法裁判发展的多维聚力

大数据在司法裁判中的核心作用主要有两个向度：一是历史向度，大数据分析证明方式可对基础数据进行梳理分析以论证过往事件的生产缘由、变化方向或事件之间的因果关联，故可作为事实认定的证明方式；二是未来向度，以相关性为基点的大数据分析可对实时数据进行趋势捕捉以预测未来情势的发展。大数据相关性预测虽尚难被纳入司法证明的程序，但却能多维推进司法裁判的深远发展。

### （一）智识拓展：裁判思维的多重锤炼

现阶段的司法实践中，同案不同判的乱象难以禁绝。深究裁判尺度偏差的缘由，剥离掉经济、文化、地域等外因影响，由于法律规范本身的表意不明或法官对大前提理解偏差所导致的差异裁判其实不多，真正能够撼动正当裁判基础的是法官对事实认定的思维偏差。最高人民法院虽积极发布了一系列的指导案例，要求法官既针对规范调适事实，又针对事实调适规范，但这种自上而下的类案指导规制并未能跳脱出类推思维的局限，加之指导案例的存量有限以及类推的大、小前提的贴合有限等因素的制约，裁判尺度的规制范围也极其有限。

司法裁判是法官以事实推理为载体，以演绎推理为形式，通过对法律规范的逻辑涵摄及正当性理由解释以完成法律适用的综合过程。<sup>〔2〕</sup>从客观事实到裁判事实，法官在事实认定中的思维模式通常表现为顺向、逆向、类推、往返四种（见表1），指导案例仅仅提供了类推思维，而大数据分析证明方式则锤炼了法官事实认定的多重模式。

表 1 法官事实认定的思维模式分类表

思维模式	认定过程	运用场域	模式缺陷
顺向模式	审查当事人的主张及证据→确定案件事实→确定法律规范并涵摄剪裁要件事实→形成裁判结论 小前提→大前提→结论	事实清楚的简单案件	难以应对复杂案情的需要
逆向模式	初步了解案情→预设结论及裁判规范→将法律规范的行为模式与当事人的主张及证据对应涵摄→确定裁判事实 结论→大前提→小前提	法官对法律关系有通透的掌握	可能引发先入为主的错误判断
类推模式	从生效裁判的案件中提炼事实要素→与在审案件的各项事实要素进行对比→通过类案事实与法律的类比涵摄→确定裁判事实 以案定案，包含案件事实与法院规范的双重类推比对	审判经验丰富的法官，通过总结类型化案件的裁判要素，形成共识性判断；审判经验尚少的法官，通过类比案件事实要素，理顺对事实的认定及法律规范的适用	受到查无先例的局限

〔2〕 参见张浩：《大数据与法律思维的转变——基于相关性分析的视角》，载《北方法学》2015年第5期。

续前表

思维模式	认定过程	运用场域	模式缺陷
往返模式	预设多种初步的结论→探求印证该结论的事实前提→如能印证该结论则采信相关事实；如不能印证则寻求另一结论再进行印证 法律规范与案件事实往返流转、彼此接近	疑难复杂案件，需要在事实与规范之间不断探索	缺乏清晰的逻辑主线，事实认定效率难以保障

大数据分析证明是数据信息收集→科学模型运算→小前提剪裁成形→大前提涵摄对应→法律适用的综合过程。基于特定证明目的而对全量性数据进行相关性筛选是逆向模式，大数据信息的预设结果推导是顺向模式，大数据分析的事实剪裁及与法律大前提的涵摄修正是往返模式，类案参考中事实与规范的双轨式大数据分析是类推模式。因此，大数据分析证明涵盖了科学知识及逻辑经验的综合运用，也承载了举证人与法官对论证思维的双重检验，具有极强的可靠性与证明力，为裁判尺度的统一贡献了自下而上的强大内生力量。

大数据分析证明过程中所积攒的“大数据经验”也是司法裁判语境知识的重要渊源，它能在法官认定事实初期或真伪不明之时，消解法官的个人偏见和信息不对称，为法官提供客观数据而非个人经验的采信参考，增进法官对举证责任的合理分配及对证据证明力的量化评估，提高法官事实认定的准确性和效率。<sup>〔3〕</sup>

## （二）法度之治：证据规则的臻于完善

“证据乃诉讼之灵魂，一国证据制度的完善程度，直接影响到该国诉讼制度的文明和理性程度。”<sup>〔4〕</sup> 证据规则的完善是推进以审判为中心的诉讼制度改革的重要环节，而我国当前的证据规则体系的完备之路道阻且长。以刑事证据规则为例：一是立法层面缺乏专门的证据法典，刑事证据制度与证据规则多呈凌乱分散态势，加之“按需设法”的倾向，致使刑事证据规则缺乏系统性、完整性与长远性；二是立法内容不仅缺少了证据法原则，在证据规则方面特别是证据资格方面的规范少之又少；三是公检法三机关在司法实践中多以“利益共同体”抱团取暖、分工配合、通融处理案件，三机关往往配合有余而制约不足，极易出现法律规定的正当程序的虚置和证据审查的形式化。<sup>〔5〕</sup>

由于公检法三机关对案件的证据标准都是“案件事实清楚、证据确实充分”，大数据分析对司法裁判证据规则的形态提取，一方面可从源头规范侦查取证的思维方式，助力建构新型的侦诉、诉辩和审辩的制约关系；另一方面还能以司法实践的经验反哺立法证据规则的完善，<sup>〔6〕</sup> 为统一刑事司法标准和证据规则提供科技支持。如对现场目击型、现场留痕型、认罪供述印证型、

〔3〕 参见周蔚：《大数据在事实认定中作用机制分析》，载《中国政法大学学报》2015年第6期。

〔4〕 沈德咏、宋随军主编：《刑事证据制度与理论》（上），人民法院出版社2006年版，第5页。

〔5〕 参见兰跃军：《刑事证据规则体系的建构》，载《中国刑事法杂志》2015年第6期。

〔6〕 现阶段，上海高院研发试用的“上海刑事案件智能辅助办案系统”、贵州高院开发运用的“办案偏离度分析工作系统”，都是以大数据、云计算和人工智能为技术内核，通过解构大量同类案件的数据关联与裁判关键，为统一证据规则的构建提供大数据经验。

拒不认罪型四种命案，通过大数据分析可知，四种命案不论在证据种类、数量、证明力度等层面都呈现出从少到多、从简到繁、从强到更强的递进证明要求，通过辨识命案证据链条的共同点与不同点，可提炼总结出规范证据能力、证明力和证据运用等一系列证据规则。

### （三）公正司法：证明机理的综合运用

尽管大数据分析证明方兴未艾，但审判终究是人类的审判，〔7〕案件事实的认定终究是法官逻辑经验判断所形成的内心确信。从大数据筛选运算到事实构建证成，大数据分析证明是多学科知识的理论融合与运算推导，是证明机理在时代语境下的综合运用。〔8〕面对大数据分析证明方式的新生力量，全体法官更应担当起公正司法之使命，主动探索完善新型证明方式，积极推进法治社会的发展。

如果说公正司法的崇尚与证明机理的弘扬相辅相成，那么判例法国家的开放心态则是二者繁茂的土壤。如“穆勒诉俄勒冈州案”中，美国联邦最高法院积极肯定了基于大体量数据论证的新型方式。〔9〕在“布朗诉教育委员会案”及“ZIPPO v. ROGER 商标侵权案”中，举证人对心理学、社会学、统计学知识证明机理的综合运用及大体量数据论证方式，再度得到了大法官们的首肯。

回望国内，“王老吉与加多宝红罐凉茶包装装潢纠纷案”〔10〕中，法官们慎重考察了涉争凉茶及中国饮料行业市场运行的大数据分析，同时也通过对品牌历史考察、公众认知判断、公平原则阐释等证明机理的综合运用，证成了红罐装潢权益完全判归任一品牌都将显失公平，并将损及社会公众利益，故包装装潢权益应由两大品牌商共有的裁判立场。“快播案”〔11〕中，两级法院审慎考察了P2P网络平台上淫秽视频传播行为的刑事责任，认为现有证据仅能证实快播公司明知其网络平台上存在淫秽电子信息而放任传播，不能证实明知是淫秽电子信息而（放任）传播，此种明知不同于司法解释的相关规定，故直接适用了刑法对“情节严重”的规定，作出了顺应时代发展的公正裁判。〔12〕

## 二、检视：大数据分析证明在司法实践中的乱象

作为大数据时代的新兴科技，〔13〕大数据在未来向度层面为司法裁判的纵深发展描绘了美好蓝图。历史向度层面，大数据分析证明方式在事实证明与认定的场域不断地试错论证，既展现出

〔7〕 参见〔美〕理查德·伦伯特编：《证据故事》，魏晓娜译，中国人民大学出版社2012年版，第39页。

〔8〕 威格莫尔创建的“证明机理”（principles of judicial proof）理论，将逻辑学、心理学、一般经验、叙事修辞等多维方法融入到了证据论证与事实认定之中。详见〔英〕威廉·特文宁：《证据理论：边沁与威格莫尔》，吴洪淇、杜国栋译，中国人民大学出版社2015年版，第189-239页。

〔9〕 参见孙光宁：《大数据时代对司法审判的冲击及其应对》，载《湖北社会科学》2016年第5期。

〔10〕 参见人民法院新闻传媒总社：《最高人民法院公开宣判王老吉与加多宝红罐凉茶包装装潢纠纷案》，载 <http://www.court.gov.cn/zixun-xiangqing-56152.html>，最后访问时间：2017年8月16日。

〔11〕 参见北京市海淀区人民法院（2015）海刑初字第512号刑事判决书、北京市第一中级人民法院（2016）京01刑终592号刑事裁定书。

〔12〕 参见范君、游涛：《P2P网络平台上淫秽视频传播行为的刑事责任》，载《人民司法》2017年第11期。

〔13〕 人类社会的信息化进程可归类为三个时代，即计算机时代、互联网时代（Information Technology）和大数据时代（Data Technology）。参见徐继华、冯启娜、陈贞汝：《智慧政府——大数据治国时代的来临》，中信出版社2014年版，第11页。

了对成为新型证明方式的渴望，亦留下了探寻自身合理存在的实践乱象。

#### （一）取证失严：载体保全的困境与证据偏在的压制

大数据信息的原始载体保全直接决定了基础信息的真实性与完整性，进而也影响了大数据分析证明的合法性与关联性。大数据原始载体的缺失与证据偏在的困境，不仅给法官认定案件事实带来了巨大冲击，也使得法律规范的预测评价指引功能处于不确定的危险状态。Y市L区和J省H县法院审理的两起关联案件，均因大数据的取证失严而导致了法官对事实的认定经不起重复回溯检验。具体情况请参见表2。

表2 大数据载体保全失严的裁判困境

序号	案例	犯罪事由	裁判困境
例一	(2016)湘1102刑初418号被告人周某、王某犯传播淫秽物品牟利罪案	2013至2015年间，周某通过租赁境外服务器创建网站并添加淫秽视频网址，编写可观看淫秽视频的电视棒程序及密码卡向全国各下线进行批量贩卖，用户只要配合密码卡在电脑上运行电视棒程序，便可观看到色情电影等淫秽视频。周某的下线王某对外贩卖的电视棒，经鉴定可观看138部淫秽视频	周某被抓获后，域外服务器因未续费而致使网站永久关闭，因域外服务器未能被公安机关控制，现已无法重新鉴定淫秽视频的加载及可供观看的数量，致使法官对事实的认定无法经受重复回溯的检验
例二	(2016)苏0621刑初161号被告人许某犯传播淫秽物品牟利罪案	周某下线许某对外贩卖的电视棒，经鉴定可观看55部淫秽视频	

1. 大数据原始载体证据保全的多重困境。上例中，淫秽视频数量鉴定的差异虽不排除大数据流变特性的干扰可能，但根本原因还在于原始载体的保全缺位。域外服务器作为淫秽视频的原始载体，由于地域管辖、取证权限等多重原因而未能被保全封存及数据还原，致使涉淫秽视频鉴定无法从原始载体或其副本展开，只能从电视棒用户端切入，然而电视棒只是淫秽视频的下级或分级载体，并不包含完整的数据文件，由此引发了原貌恢复的鉴定偏差。同时，域外服务器租用权的失效，也使得鉴定结论无法得到重复检验。

2. 大数据持有人证据偏在的优势压制。大数据的控制者拥有着数据信息的所有权和管理权限，即便是面对公安机关等国家侦查力量，也处于证据偏在的优势地位。<sup>〔14〕</sup>加之大数据信息通常依赖云计算的分布式计算架构，<sup>〔15〕</sup>大数据被碎片化地存储于动态的群体网络之中，这又进一步加剧了大数据信息完整采集的困难程度。

#### （二）论证失实：目的扭曲的论证与错误隐蔽的计算

大数据分析证明须借助于科学知识的专业运算，科学知识对数据“内部证成”起到了决定性作用，而科学知识本身的“外部证成”也将影响大数据分析证明的效力，表3中的例三即是科学知识周延欠缺的典型。

〔14〕 参见高波：《大数据：电子数据证据的挑战与机遇》，载《重庆大学学报》（社会科学版）2014年第3期。

〔15〕 参见何晓行、王剑虹：《云计算环境下的取证问题研究》，载《计算机科学》2012年第9期。

表3 大数据分析证明失严的概率悖论

序号	案例	指挥罪名	自我辩护意见	涉及知识	论证悖论
例三	一审：江西省南昌市中级人民法院（2014）洪刑初字第29号刑事判决 二审：江西省高级人民法院（2016）赣刑终33号刑事判决书	受贿罪、挪用公款罪	被告人周文斌认为：行贿人和受贿人在最初审讯中都交代犯罪行为发生在5月份，后来均发现出现了记忆错误的概率为1/140；在纠正此错误后，双方又同时将犯罪发生时间（错误的）改为10月，其概率为1/20700；从5月到10月的绝对误差为5个月，相对误差为50%；从1号到10号的绝对误差为10天，相对误差为2.8%，两个人分开审讯，误差率同时发生同样的巨大变化说明侦查人员存在非法取证行为	概率论等数理知识 基准率频率 发生率频率	被告人试图混淆基准率频率与发生率频率的概念，通过对时间错误特征的放大及对背景因素的修改，以转移法官的注意力，从而达到受贿指控的彻底失实的目的

1. 预设目的扭曲了概率运算的方向。大数据分析证明常表现为论证事件存在、发展及变化的绝对概率，具有概率论证的特性内核。例三中，被告人虽采用了概率分析的大数据论证，但却暗中偷换了基准率频率与发生率频率的运用。通过对记录误差的放大及行贿时间背景因素的冲淡，剔除了重要数据并进行了选择性运算，从而达到所谓的“非法证据排除”的主张。<sup>[16]</sup> 例三的概率悖论在于，5月或10月可能是行贿人及受贿人特定的碰头交易期，或存在其他来往接触的频繁期，故双方对于上述两个日期具有特别的记忆；这些潜伏着特定意义的时间节点因被选择性地忽略，故而更能实现特定目的的偏差论证。

2. 数据选择性处理暗含了错误风险。大数据分析证明的话语优势显而易见，它迎合了法官对事实认定精确性和可操作性的需求，使得非科学性的、非统计学证据所具有的同等证明力黯然失色。<sup>[17]</sup> 但例三也暴露了大数据分析证明方式的巨大风险——不同主体基于特定目的可以对大数据进行筛选、对算法进行操纵，形成不同甚至相互冲突的结果。大数据分析证明方式具有极强的专业性，数据信息的周延及算法模式的错误将变得极其隐蔽，即使是面对法官一般性的数理逻辑及经验审查，这种披着科学外衣的隐秘错误计算也不难遁形。

（三）关联失度：相关趋势的充斥与因果程度的缺乏

运用大数据分析证明的司法实例通常具有以下特征：待证事实往往需要结合时间与空间的大数据资料，通过对地势地貌、气候温差、金融变革等大事件的性质成因、变化缘由进行分析，以论证事件的存在、发展和变化的情况。表4中的“涉农案件”即采用了大数据分析证明方式，二案均试图通过大数据分析论证气候、气温是制种成功的关键。而法院一正一反的判决，虽都表明了对大数据论证方法的肯定，却均以气温不是唯一决定因素为由，未采信大数据论证一方的最终观点。

[16] 参见巩寒冰：《概率性证据研究中的认识悖论》，载《证据科学》2016年第2期。

[17] 参见前引[16]，巩寒冰文。

表 4 未证成因果关系的大数据分析证明案例

序号	案例	纠纷缘起	证明核心	论证方法	裁判结果
例四	江苏省高级人民法院(2007)苏民三终字第0010号二审民事判决书	江苏省农业科学院与江苏苏农种业科技有限公司签订制种技术专利实施许可协议,许可苏农公司以制种技术于2003年制种3000亩,并对专利关键问题提供解决方案。2004年6月,苏农公司仅有约300亩制种合格,苏农公司遂诉请江苏农科院承担违约赔偿责任	涉案的专利技术采用低温敏不育系作母本,母本抽穗期控制在每年的8月16日左右,如遇上日平均气温连续三天低于23.5℃,则会导致纯度下降,关键在于对抽穗日期及温度的证明	苏农公司提交了其制种地区盐城市、滨海县、金湖县三地气象局出具的2003年5月下旬至8月平均气温和日照时数以及常年旬平均气温和日照时数,并申请了袁隆平教授作为专家辅助人出庭说明制种要求;江苏农科院提供了专利技术申请资料,资料中包含了对过往50年的气象观测数据及一系列的制种配套材料,并申请了气象学教授姚克敏作为专家辅助人,就涉案专利技术中的播种期和亲本孕育期等问题说明情况	两级法院经审理均认为“导致抽穗期延迟存在诸多因素,除气候外,还要考虑是否做好氮肥施用、病虫害防治、割叶是否适当等”,苏农公司因提交的证据不足以证明气候是导致制种不合格的原因,未尽完全的举证责任而被驳回诉请
例五	安徽省高级人民法院(2016)皖民终163号二审民事判决书	朗溪县建平镇陈福林粮食种植家庭农场与常州溧里粮销有限公司签订种子购销合同,由溧里粮销公司向陈福林农场提供香稻种进行生产并回购粮食,陈福林农场在种植过程中因严重减产诉请溧里粮销公司承担违约赔偿责任	涉案“香稻种”是否因患水稻颈瘟而减产,是否具有严重缺陷,因而不适合在朗溪区域种植	溧里粮销公司提交了朗溪县种子管理站的田间现场鉴定、政府文件、媒体报道、气象数据及专家辅助证人的意见,拟证明2014年因天气原因导致全国范围内有多款种子大面积爆发稻瘟病,并在二审中申请专家出庭佐证每年8至9月中旬下雨易引发稻瘟病,病发主要受气候、品种抗性、病菌突变、管理措施等因素影响	两级法院经审理均认为,虽然气候条件系引发稻瘟病的原因之一,但未排除香稻品种自身易感稻瘟病的情形,故而支持了陈福林农场的合理诉请

法律上的因果关系多采以相当因果关系通说,或采以必要条件规则(but for规则),<sup>[18]</sup>是一种强关联或完全关联关系。由于事实认定要求的是强因果或完全因果关系,故大数据分析证明不应只涉及相关性的浅象,还需直击因果性的深核。当前司法实例中,大数据分析证明大多仍停留在相关性分析的弱关联程度,仅仅证成了一个变量的变动可能引发相关变量不特定变化,而往往没有证成涉争因素之间变与不变的绝对趋势。上述两例中,举证人虽论证了气温是制种成功的影响因素,但未能证成气温系决定制种成功的唯一、排他性的因素,这种对不确定的弱关联关系而非法律上的因果关系的论证,无法形成充分证明力,亦无法说服法官采信。

[18] 相当因果关系说认为,在一般认知的角度上观察到哪些因素能够引起一定结果的发生时,即可判断其存在因果关系。事实上的因果关系常用于侵权法领域,即没有被告的侵权行为,就没有原告的损失事实。参见前引〔2〕,张浩文。

### 三、探因：大数据分析证明在事实认定中的失范透析

大数据分析证明方式虽能助力于法官对案件事实的认定，但面对新型证明方式的乱象风险，法官应保有更多的理性思考，从而更好地厘清大数据在事实认定中的偏差。

#### （一）总体思维之缺：载体漏洞与关联阙如

大数据在形式上常表现为电子证据，故在载体形式的证据采集层面应具有整体思维。大数据信息采集的完整程度直接决定了其证据能力的有无和证明力的大小，也直接界定了案件事实认定的可能范围。大数据分析证明的形式缺憾源于两方面：

一是大数据本身的特质决定了其形式缺憾的必然性。大数据信息具有高度复杂性与不确定性，即使人们收集及处理数据的能力不断提高，但相对于大规模的数据增长而言依然滞后，因此，人们对大数据信息范围或内容的“全貌”采集也只能是相对的。“北京奇虎科技有限公司与腾讯科技（深圳）有限公司滥用市场支配地位纠纷案”（以下简称“奇虎诉腾讯案”），<sup>〔19〕</sup>奇虎提供的大数据分析只监测了PC端口的即时通讯软件数据，遗漏了对手机和平板电脑等移动端口及微博或SNS产品的数据采集，而终审法官通过对CNNIC、艾瑞咨询的大数据复合性统计，最终将移动端即时通信服务纳入了相关商品市场的审查范围。<sup>〔20〕</sup>

二是大数据的信息收集及载体固定方式带来了形式缺憾的可能性。由于载体保全等因素影响，大数据分析证明在关联性层面极易产生多重偏差：即大数据的信息采集将因病毒攻击、污染篡改而产生介质关联性偏差；大数据的运算处理将因特定目的的筛选控制而产生行为关联性偏差；大数据收集的时间滞后、IP地址不符将产生时空关联性偏差；基于当事人特定主张的大数据论证可能导致身份关联性偏差。如例一中，淫秽视频的原始载体的保全失严，引发了大数据载体介质的关联性漏洞，进而造成了同案不同判的司法窘境。

#### （二）容错思维之缺：样本偏差与全量精准

传统的概率统计学是一种抽样范式的精确思维。这种对随机抽样加以统计概率的运算，受样本随机性抽取的影响，极有可能产生结果的偏差。由于如何抽样、如何选择样本范围等计算方式缺乏科学标准，抽样统计的概率更像是统计人员主观见诸客观的演绎活动，它需要通过统计员的直觉、经验甚至是臆想来填补样本之间的信息空缺，一些可能影响到事物定性的关键数据往往会被遗漏，故传统的概率统计论证在真实性层面上存在巨大风险。如例三中，周文斌采用了定向抽取的样本来进行概率论证，但未剔除5月、10月对行贿人和受贿人可能存在的特定含义，而是以普通人的视角去进行大概率的分析，故其自我辩护也嬗变为了假借科学外观的谬论。

大数据分析证明须有全量数据的容错思维。大数据除拥有信息密度较高的结构化数据外，还存有95%非结构化的混杂信息，这些信息往往会因价值密度过低而未被纳入概率统计的抽样范围，但非结构化数据往往蕴含着巨大的潜在价值，它被大数据有“容”乃“大”地涵盖，并成就

〔19〕 参见最高人民法院（2013）民三终字第4号民事判决书。

〔20〕 参见李慧：《大数据成为证据的新标准》，载《人民法院报》2016年12月7日，第2版。



了其论证精髓。如例三中5月、10月的特定含义就是一种信息含量较低的非结构化数据。这些数据因含义模糊而极易被传统的概率统计所忽略，但却能使事实认定更客观、更真实。从传统概率统计的样本精确模式到大数据容错模式，人类恰似经历了正一反一合的思维转变，或是肯定—否定—否定之否定的涅槃重塑。<sup>〔21〕</sup>

### （三）相关思维之缺：相关到因果的转化困境

小数据时代的因果关系具有一定的历史局限。受制于技术手段限制、信息来源缺乏、统计成本过大等因素，小数据时代的信息采集被局限于明显可见的几个维度，而那些暗含了关联信息的剩余维度则往往被人为地忽略。<sup>〔22〕</sup>大数据分析证明则保留了高维数据的全貌，其混杂的原生态信息涵盖了运算对象的多维参数及相关关系，通过有效整合看似无关联的信息碎片，使得事实证明更加精准，论证方式也由经验驱动转向数据驱动。

因果关系是司法裁判的基础，如何从相关关系中推断出因果关系，才是大数据分析证明在事实认定中的深义所在。<sup>〔23〕</sup>在前述“涉农案件”中，举证人虽论证了气温与制种成功具有相关性，但单一的气温变量在何时、何地、何种条件下能成为影响制种成功的决定因素，才是大数据因果论证的关键。在“奇虎诉腾讯案”中，奇虎虽有新意地提出了腾讯在盈利能力、专利数量和布局等非机构因素中具有优势，遗憾的是未能证明这些非结构因素是如何绝对影响市场支配地位的，也未能完成非结构因素对市场支配地位影响的因果证成。<sup>〔24〕</sup>

相关思维虽在案件事实的认定中无法取代因果思维，但却能避免数据收集、判断及利用的漫无目的，同时也开启了因果关系界定的另一种可能——传统模式只是把原因等同于自然界中“原子式”的必然性存在，而非视为对事物相互作用的过程描述。在大数据时代，“因果关系”应描述为因素相互作用过程与其效应之间的联系；而因素尚未相互作用时的状态，则体现了一种与潜在结果密切相连的“相关关系”，这种“相关关系”实质是“因果派生关系”的体现。<sup>〔25〕</sup>

## 四、破冰：关联性视角下大数据分析证明的路径探析

大数据分析证明若被笼统置于传统的证明方式之中，其特殊的证明规则必然会被忽视。而鉴于大数据分析证明方式在真实性、合法性层面涉及的科技知识及法律规定较多而法官逻辑经验判断较少，故本文仅从关联性视角探析大数据分析证明方式引入事实认定的机制建构，具体包括两大方面：一是对负载了大数据信息的载体及系统的关联性审查；二是对大数据分析证明目的及内容的关联性审查。<sup>〔26〕</sup>

### （一）载体关联性：人、事、物、时、空的五维一体设计

在载体的关联性层面，大数据分析证明可比照适用电子证据的相关规定。大数据载体关联性

〔21〕 参见张弛：《大数据思维范畴探究》，载《华中科技大学学报》2015年第2期。

〔22〕 参见阳杰、应里孟：《大数据时代的审计证据与审计取证研究》，载《财会月刊》2017年第1期。

〔23〕 参见姜奇平：《因果推断与大数据》，载《互联网周刊》2014年第18期。

〔24〕 参见杨文明：《论互联网企业市场支配地位认定的非结构因素》，载《河北法学》2014年第12期。

〔25〕 参见王天思：《大数据中的因果关系及其哲学内涵》，载《中国社会科学》2016年第5期。

〔26〕 参见陈瑞华：《刑事证据法学》，北京大学出版社2012年版，第63页。

的分析证明核心在于通过类型化分析以提取有价值的信息，它建立在对基础数据全面容错性的收集之上。通过相关性的指引对基础数据归类整合，挑选出与人、事、物、时、空相对应的载体关联项，排除干扰项或无关选项。本文以“快播案”为例（见表 5），探讨大数据证明方式的载体关联性构建：

表 5 大数据证明对载体关联性的五重要求

大数据分析证明的载体关联性审查	内容		含义	以“快播案”为例
	人	身份关联性	诉讼参与人在大数据中具有特定身份	虽然光×公司提供了涉案服务器，但快播公司在实际上是四台服务器的远程控制者和日常维护者
事	行为关联性	与大数据交互的诉讼参与人的责任承担	缓存的视频文件以快播公司特有文件格式和特征码文件名的方式存储。公安机关从服务器里提取了 29 841 个视频文件进行鉴定，认定其中属于淫秽视频的文件为 21 251 个	
物	介质关联性	承载大数据内容的电子载体与诉讼参与人的关系	国家信息中心电子数据司法鉴定中心对四台服务器及存储内容进行检验表明：公安机关扣押的四台服务器及其所存储的视频文件没有遭到破坏或污染	
时	时间关联性	大数据所反映的焦点时间与涉案时间的关系	2013 年 8 月，光×公司提供四台服务器开始上线测试，快播公司为四台服务器安装了快播公司的缓存服务器系统软件，并通过账号和密码远程登录进行维护；2012 年 8 月与 2013 年 10 月快播公司因对淫秽视频等不良信息怠于监管而受到二次行政处罚，快播公司明知其网络平台上存在淫秽电子信息而仍然放任传播	
空	地址关联性	大数据反映的虚拟地址与涉案地址的关系	快播网站网址包括：www.kuaibo.com、www.qvod.com 等；执法人员登录 www.kuaibo.com，进入快播超级雷达，很快便找到了可播放的淫秽视频	

1. 身份关联性。辨明被告人或其他诉讼参与人（以下统称诉讼参与人）在大数据虚拟空间中的特定身份，取证时应查明电子设备、存储介质、虚拟账号的所有人、保管人、使用人等信息。如快播公司系涉案四台服务器的远程控制者及日常维护者。

2. 行为关联性。厘清诉讼参与人因大数据信息与现实行为的交互所应负的法律責任，取证时应全面提取电子设备、存储介质、密钥等密码设置情况、运行状态、特殊要求等信息，查明虚拟空间中的操作行为、信息传输行为、黑客攻击行为等是否系诉讼参与人所为。如快播客户端可见的视频中有 2 万余个淫秽视频，且淫秽视频系以快播公司特有的文件格式和特征密码进行存储。

3. 介质关联性。查实承载大数据信息的载体是否为诉讼参与人占有、所有或共有等情况，并查明是否为原始载体或复制小样，信息收集是否完整、有无删减或人为控制等情况。如涉案四台服务器及视频文件并未受到污染、篡改或攻击。

4. 时间关联性。区分虚拟时间或机器时间与真实的物理时间是一致还是存在时差等。如从快播公司对服务器的运行监控及维护的时间节点、怠于监管淫秽视频而受到两次行政处罚的时间跨度可知，该公司明知其网络平台上存在淫秽电子信息而仍然放任传播。

5. 空间关系性。由于大数据的载体多具有内置或外置的地址信息，故需证明虚拟空间的独特地址、基础信息的存储位置或定位等。如从快播客户端进入即可搜索到淫秽视频。

6. 载体之保全。坚持介质优先原则，应尽可能保全原始数据的存在载体，同时也应保全基

础数据的应用环境，审查负载基础数据的设备、系统是否处在正常、稳定的状态之中。同时考虑到大数据的流变特性，可考虑加入指纹签名的载体保全技术，如对原始数据进行复制并生成数字指纹，数字指纹交扣押方、被扣押方和独立第三方进行保存，基础数据一旦改变则指纹签名也将发生变化；所有的分析鉴定工作都在复制件上进行，以避免对原件产生污染与破坏。<sup>〔27〕</sup>

综上，大数据分析证明方式的载体关联性审查规则可概括如下：（1）大数据分析证明的内容对待证事实能产生实质性影响，载体保全完好且涵盖的数据信息在身份、行为、介质、时间、空间上同诉讼参与人存在联系，应认定该证明方式具有关联性；（2）法官应结合待证事实及质证意见来审查大数据分析证明方式的关联性，并运用经验法则进行判断，必要时可通过司法鉴定予以查实；（3）如上述方法无法确定大数据分析证明的关联性，则可由法官根据举证责任来判定相应的证明责任及后果。

## （二）信息关联性：法官对专业知识的合理评价

大数据分析证明在事实认定中将接受两次法官的认知。第一次是“大数据分析证明”本身表达的内容，类似于科学认知；第二次是“大数据分析证明的采信”所表达的内容，需交由法官直接评价并形成诉讼认知。诉讼认知是证据裁判思维，是法官运用经验法则对大数据分析证明的采信判断。本文以“奇虎诉腾讯案”为例，从诉讼认知角度来探讨其关联性的构建。大数据分析证明的采信需经过两次审查：一是大数据被整合分析为证明主张时，需要排除计算偏差或主观控制的可能，主要体现为大数据分析证明的采纳规则；二是法律层面对大数据证明目的之评价，主要体现为大数据分析证明的排除规则。

### 1. 采纳规则：论辩相关性与修辞相关性

大数据分析证明的内容关联性是其可采性的前提，关联性关注的是证据主张的运算与形成，须审查举证人、鉴定人或专家辅助证人在大数据运算过程中是否存在数据筛选、计算错误、主观控制等干扰因素。大数据分析证明采信的知识问题，应参考科学证据的审查标准，并从论辩相关性和修辞相关性两方面进行“外部证成”。具体内容参见表 6。

表 6 大数据分析证明的采纳规则

	维度	指标	主要针对的问题	主要评价指标
大数据证据采信的知识问题	论辩相关性	有效性	大数据证据是建立在科学的理论和分析计算方法的基础之上	逻辑的有效性 科学的有效性
		可靠性	具备有效性的大数据证据被合理地运用到了案件事实发现的过程之中	检验方法—误差控制 检验结论—叶贝斯估计
	修辞相关性	证明力维度	大数据证据在不同类型案件中应当达到什么样的证明标准	微弱证据标准 优势证据标准 清晰且令人信服证据标准 确信无疑证据标准
		实质性维度	证据提交人应从哪些方面回答裁判者的内心疑惑	最佳解释原则（IBE 模式）

论辩相关性要求大数据证明必须经过科学运算并合理地运用于事实认定之中，运算方式应符合科学逻辑，同时也对算法误差提出了叶贝斯模式的控制。如：“奇虎诉腾讯案”中，一审法院

〔27〕 参见戴士剑、刘品新主编：《电子证据调查指南》，中国检察出版社 2014 年版，第 209-215 页。

在进行即时通讯市场分析的假定垄断者测试（HMT）中采用了价格上涨法（SSNIP），但此法却在二审中被质量下降法（SSNDQ）所取代。这两种测试方式虽均基于市场运行的专业大数据分析，但由于 SSNIP 不适用于免费或以非价格为主要竞争力的商品测评，如 QQ 软件糅杂了（付费）广告推广群体和普通交流群体，具有双边市场特性，不能以 SSNIP 大数据分析进行简单判定，否则将导致相关市场界定过宽，不具备有效性及可靠性，故二审法院最终采用 SSNDQ 方法进行定性分析，这也反映了法官对大数据分析证明的论辩相关性的严格要求。<sup>〔28〕</sup>

修辞相关性则需要举证人、鉴定人或专家辅助人对证明的过程及方式进行充分解释，对事实建构形成最佳解释，以达到说服法官的目的，它强调了一种优势证明或排除合理怀疑的证明标准。如：腾讯虽提出了相关市场应界定为全球市场的抗辩，由于并未“主要考虑多数需求者选择商品的实际区域、法律法规的规定、境外竞争者的现状及其进入相关地域市场的及时性等因素”，未能达到清晰且令人信服的证明标准，缺乏了修辞相关性的证明力因素，故未能说服二审法官改变将即时通讯服务市场界定为中国大陆的观点。

## 2. 排除规则：法官经验逻辑与法律规制

大数据分析证明的排除规则，与其他证明方式的排除规则内容相近，均可借助于法官的经验逻辑及法律规则进行判断。排除规则通常采用图 1 中的相关标准来控制证明目的与待证事实的关联程度。

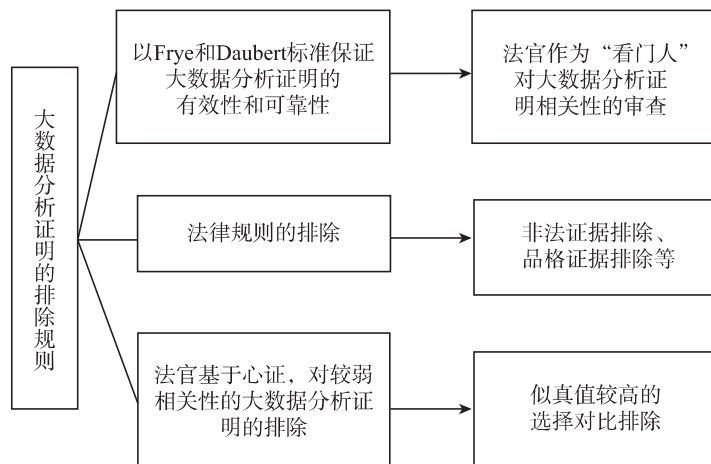


图 1 大数据分析证明的排除规则

“Frye 标准”（“普遍接受”标准），指科学检验必须达到“其所属的特定领域获得普遍接受”的程度才能被法官采信。然而，由于部分前沿的科学知识在得到同行普遍认同前可能会经历漫长过程，此标准很可能将新型证明方式拒之门外。<sup>〔29〕</sup>“Daubert 标准”则从四个方面对大数据分析证明严格把关：一是大数据分析理论是否得到了检验；二是大数据分析原理或证明方式是否已得

〔28〕 参见丁春燕：《论我国反垄断法适用中关于“相关市场”确定方法的完善——兼论 SSNIP 方法界定网络相关市场的局限性》，载《政治与法律》2015 年第 3 期。

〔29〕 See David W. Louisell & Christopher B. Mueller, *Federal Evidence: Civil and Criminal*, Lawyers Co-operative Publishing Company, 1993, p. 290.

到了公开发表或者专家同行的认可；三是大数据计算的错误率是否已得到了统计明确并有规范的操作标准；四是大数据分析所依附的理论是否已达到所在领域“普遍接受”的程度。<sup>〔30〕</sup>

“似真性理论”解决的是大数据分析证明存在冲突结果时的取舍问题。结果冲突的背后是举证人证明目的的两造对抗，或者是鉴定人、专家辅助人的观点隔阂。似真性理论要求法官结合大数据分析证明与待证事实的关联程度，选择出相对似真值较高的论证方式，并将冲突主张进行有效整合。“奇虎诉腾讯案”中，假定垄断者测试是普遍适用的相关市场的界定思路，SSNIP 或 SSNDQ 等方法亦都得到了 Frye 及 Daubert 标准的认同。法官对 SSNDQ 方法的选定及对 SSNIP 方法的排除，则体现了大数据分析证明排除规则的实践运用；而法官将即时通信市场从全球市场缩减至中国大陆，则是受到了似真性理论的影响。

### （三）相关性限制：大数据预测的运用与规制

大数据预测与品格证据具有一定的相似性，二者对法官认定事实均能产生间接效力；但大数据预测是基于数据相关性的未来趋势判断，而品格证据则是通过对被告人一般特性的概括性表述，<sup>〔31〕</sup>判断已然发生的某个事件与被告人过往行为是否具有一致性。品格证据并非一味地被排除，而是具有有限的可采性。美国《联邦证据规则》第 404（b）条规定，品格证据可用于证明动机、机会、意图、准备、计划、知识、身份、无过失或意外事件；<sup>〔32〕</sup>而这与大数据相关性预测的司法功能又存在一定的重合。

结合品格证据的相关特质可知，大数据预测分析尤其是基于被告人行为数据的历史分析，在刑事诉讼法亦具有有限的适用基础，如判断累犯、惯犯、缓刑、减刑、假释的条件；需特定能力才能完成的犯罪（性犯罪、暴力犯罪、毒品犯罪等）；未成年人的社会调查等。在这些裁判场域，公诉方往往会主动提出对被告人过往历史大数据的分析并提出量刑建议，法官也享有自由裁量权以判断被告人历史大数据的相关性及证明价值。

从法律因果关系的角度看，基于被告人行为数据分析而得出大数据预测仍属于弱关联性的间接证明，但却能对被告人的权利产生限制或剥夺，故法官在审查大数据的预测分析时，应坚持谨慎区分的严格采纳原则，且不能作为裁判的唯一依据，法官可采纳大数据预测分析的场域为：（1）基于特定身份、能力、知识或技能的犯罪，如与性、毒品、专业知识相关的犯罪，或贪污、受贿等特殊身份的犯罪；（2）累犯、惯犯、缓刑、减刑、假释的适用；（3）未成年人犯罪案或对自首、立功的认定，应采纳对被告人有利的大数据预测分析。

## 五、结语：裁判道路的深远延展

大数据分析证明方式作为人工智能时代的一种更为客观与科学的新型证明方式，正全方位地渗入证据科学与证明机理的理论品格之中，为司法裁判的深远发展增添了科技动力。尽管大数据分析证明方式在发展初期还存在着一定的实践乱象，但通过对其载体形式与信息内容的二元规

〔30〕 参见陈邦达：《美国科学证据采信规则的嬗变及启示》，载《比较法学研究》2014年第3期。

〔31〕 参见陈琳、刘晓清：《品性证据规则一般禁止条款的文本分析》，载《海峡法学》2015年第3期。

〔32〕 参见何家弘、张卫平主编：《外国证据法选译》（下卷），人民法院出版社2000年版，第600页。

制，法官们可以拓展裁判思维的实践疆界，拥有驾驭大数据分析证明的坚实力量，进而对案件事实进行愈发精准与科学的认定。

附件一：

## 《刑事案件大数据分析证明方式审查采信意见（建议稿）》

为规范法官对刑事案件中大数据分析证明方式的审查判断，确保犯罪事实认定的准确性与客观性，根据《中华人民共和国刑事诉讼法》等有关法律规定，结合刑事审判经验及司法实际，制定本规则。

### 一、大数据分析证明载体的关联性

第一条【载体的身份关联性】大数据载体所承载的内容应涵盖被告人在大数据虚拟空间中的特定身份，法官须查明大数据所附着的电子设备、存储介质、虚拟账号的所有人、保管人、使用人等信息。

第二条【载体的行为关联性】大数据载体所承载的内容应涵盖被告人因大数据信息与现实行为的交互所应负的法律責任，法官须审查侦查机关是否全面提取了电子设备、存储介质、密钥等密码设置情况、运行状态、特殊要求等信息，并审查虚拟空间中的操作行为、信息传输行为、黑客攻击行为等是否系被告人或其关联行为人所为。

第三条【载体的介质关联性】法官须查明大数据载体所承载的内容是否为被告人或其关联行为人占有、所有或共有等情况，查明大数据的原始载体的收集情况及现有载体是否为原始载体或复制小样，查明复制的大数据载体在基础数据的信息收集方面是否完整、有无人为控制等情况。

第四条【载体的时间关联性】法官须查明大数据信息所存在的虚拟时间或机器时间与真实的物理时间是否一致或存在时差等，查明特定领域大数据的拥有人、操控人对数据流的操控时间节点等情况。

第五条【载体的空间关联性】由于大数据的载体多存在内置或外置的地址信息，法官须查明大数据在虚拟空间中的独特地址、基础信息的存储位置或定位等情况。

第六条【载体的保全】坚持介质优先原则，应尽可能地保全原始数据的存在载体，同时也应保全基础数据的应用环境，审查负载基础数据的设备、系统是否处在正常、稳定的状态。能够获取原始存储介质的，应当封存原始存储介质，并制作笔录，记录原始存储介质的封存状态，由侦查人员、原始存储介质持有人签名或者盖章；持有人无法签名或者拒绝签名的，应当在笔录中注明，由见证人签名或者盖章。如有条件，侦查人员应对相关活动进行录像并加入原始载体数字指纹签名进行保全，数字指纹需与物理实体一并记录，数字指纹应同时交由扣押方、被扣押方和独立第三方进行保存。

### 二、大数据分析证明内容的关联性

第七条【内容的有效性】大数据分析证明所采用的计算方法、分析模式或科学理论必须建立

在业界普遍认可之上，法官既须审查举证人计算方式的科学性，又须审查大数据运算的逻辑推导的有效性。

第八条【内容的可靠性】法官在认定大数据分析证明的内容具有有效性之后，仍须审查该有效的证明内容是否被合理运用到了案件事实的证明过程之中，并框定大数据分析证明的误差范围。

第九条【大数据分析证明的证明力】举证人对自身提出的大数据分析证明主张，应保障充分的证明力；如大数据分析证明系公诉机关或自诉案件的原告人提出，则应达到与全案证据综合印证吻合，且能使得法官排除合理怀疑的标准；如大数据分析证明系公诉案件的被告人及其辩护人提出，则应达到能够引起法官合理怀疑的证明标准。

第十条【大数据分析证明的说服力】大数据分析证明的举证人应围绕待证事实的争议焦点进行充分论证，全面回应事实间的冲突或疑点，同时举证人应向法庭陈述其所欲论证的案件事实。

### 三、大数据分析证明的关联性排除规则

第十一条【关联性的排除规则】法官在对大数据证明的关联性进行审查时，应循序如下顺序：

（一）大数据分析证明的内容对待证事实能产生实质性影响，且载体涵盖的数据信息在身份、行为、介质、时间、空间上同被告人相关，应认定该证明方式具有关联性；

（二）法官应结合待证事实及质证意见综合审查大数据分析证明方式的关联性，并运用经验法则进行判断，必要时可通过司法鉴定予以查实；

（三）法官须审查大数据分析原理或证明方式是否已经公开发表或者得到专家同行的认可，审查大数据的分析理论是否得到了检验并达到所在领域“普遍接受”的程度，审查大数据计算的错误率是否已经得到统计明确并有规范的操作标准；

（四）法官在面对大数据分析证明的结果冲突时，须结合似真性理论审查大数据分析证明与待证事实的关联程度，选择出相对似真值较高的论证方式，并对冲突主张进行有效整合；

（五）如上述方法无法确定大数据分析证明的关联性，可由法官根据举证责任的规则及后果进行责任判定。

### 四、大数据预测相关性的合理规制

第十二条【大数据预测的限制规则】由于大数据预测分析可能会对被告人的权利产生限制或剥夺，故法官在审查大数据的预测分析时，可采纳大数据预测分析的场域如下：

（一）法官在审理基于特定身份、能力、知识或技能的犯罪，如与性、毒品、专业知识相关的犯罪或贪污、受贿等特殊身份的犯罪时，可采纳大数据对被告人刑期可能加重或减轻的综合性利弊预测；

（二）法官在审查累犯、惯犯、缓刑、减刑、假释的情节适用时，可采纳大数据对被告人刑期可能加重或减轻的综合性利弊预测；

（三）法官在审理未成年人、限制行为能力人犯罪案件时，或审查被告人是否存在自首、立功等情节时，应采纳对被告人有利的大数据预测分析。

附件二：

附件一《刑事案件大数据分析证明方式审查采信意见（建议稿）》，系笔者在梳理了现行电子证据关联性规范的基础之上综合拟定而成，具体规范的统计如表7：

表7 当前法律规范对电子证据载体关联性的规制梳理

序号	规范名称	分布单位	身份关联性	行为关联性	介质关联性	时间关联性	地址关联性
1	《人民检察院电子证据工作细则（试行·审议稿）》	最高检	√	√	√	√	√
2	《关于办理网络犯罪案件适用刑事诉讼程序若干问题的意见》	最高法 最高检 公安部	×	√	√	√	×
3	《关于审理证券行政处罚案件证据若干问题的座谈会纪要》	最高法	×	×	√	√	√
4	《关于办理刑事案件收集提取和审查判断电子数据若干问题的规定》	最高法 最高检 公安部	√	×	√	×	×
5	《关于办理网络赌博犯罪案件适用法律若干问题的意见》	最高法 最高检 公安部	√	×	√	×	×
6	《计算机犯罪现场勘验与电子证据检查规则》	公安部	×	√	×	√	×
7	《人民检察院电子证据鉴定程序规则（试行）》	最高检	√	×	√	×	×
8	《人民检察院刑事诉讼规则（试行）》	最高检	×	×	×	√	×

**Abstract:** The method of big data analysis proves that it can broaden the path of judges' knowledge management, and improve the rules of evidence from the perspective of judicial operation. It can also advocate the concept of fair and enhance the comprehensive application of judges to the principles of judicial proof. The method of big data analysis proves itself through trial and error constantly. The lack of overall thinking, fault-tolerant thinking and relevance thinking are the main causes of the abnormality of big data analysis. Focusing on the relevance of big data analysis and proof method, we can regulate its application path in fact recognition from three aspects, Firstly, regulate the relevance of the carrier from five aspects of people, object, things, time and space. Secondly, design the relevance of the content from the judge's reasonable evaluation of the way of proof. Thirdly, analyze the judicial application of the relevance of big data from the perspective of the limited introduction of big data forecast, in order to accurately determine the facts and scientific efficiency for judicial judgment.

**Key Words:** big data analysis proof method, carrier relevance, content relevance

(责任编辑：刘君博 赵建蕊)