



The Impact of Big Data on Criminal Justice

Li Xiaokai

China University of Political Science and Law, Beijing

Abstract: As a method different from traditional data analysis, big data can provide powerful information retrieval, information processing and information judgment tools for criminal investigation, crime prevention and criminal trial, thus positively influencing the maintenance of a good criminal operating environment in many aspects, such as promoting judicial openness, grasping and guiding public opinion, understanding judicial operation, objectively evaluating judicial staff and rationally allocating judicial resources, and providing unprecedented empirical analysis and empirical data for theoretical research. At the same time, however, it is difficult to strike a balance between the fight against crime and the protection of the rights of individual citizens, and to see how machines can be used in predicting crime, and sentencing, and how to objectively confront the empirical patterns of criminal justice derived from big data analysis in theoretical research. The new issues raised by big data in criminal justice, such as the development and innovation of criminal justice theory itself, can likewise not be ignored.

Key words: Criminal justice; Big data; Data analysis

Received: 2020-06-12; Accepted: 2020-06-27; Published: 2020-07-08

大数据对刑事司法的影响

李小恺

中国政法大学，北京

邮箱: lixiaokai@cupl.edu.cn

摘要: 作为一种有别于传统的数据分析方法，大数据能够为刑事案件侦查、犯罪预防和刑事案件审判等实务工作提供非常强大的信息检索、信息处理和信息研判工具，从而促进司法公开、掌握并引导社会舆论、了解司法运行情况、对司法工作人员进行客观评价以及合理配置司法资源等多个方面对维护良好的刑事运行环境产生积极影响，并为理论研究提供前所未有的实证分析和经验数据资料。但是与此同时，如何平衡打击犯罪和保护公民个人权利之间的关系、如何正确看待机器在犯罪预测和定罪量刑方面的作用，以及如何在理论研究中客观面对大数据分析得出的刑事司法经验规律并借此完成对刑事司法理论自身的发展和创新等刑事司法大数据所引发的新问题，也同样是不可忽视的。

关键词: 刑事司法；大数据；数据分析

收稿日期：2020-06-12；录用日期：2020-06-27；发表日期：2020-07-08

Copyright © 2020 by author(s) and SciScan Publishing Limited

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

<https://creativecommons.org/licenses/by/4.0/>



1 引言

进行数据分析工作的主要目的就是为了让“数据说话”，从即有的各种信息中找到客观规律，

从而对未来事物的分析提供依据。大数据不同于传统的社会学统计分析,其对待数据的新思维以及对现代信息技术的灵活运用,使经过分析得到的经验规律更加贴近真实,从而对于未知事物的预测和判断更加准确、高效,这也正是大数据能够吸引众人的魅力所在。毫无疑问,将大数据应用到刑事司法领域,会带来十分显著的积极影响,甚至能够对改变刑事司法制度提供强大的推力。然而与此同时也应当注意的是,大数据并不是“万金油”,更不是“灵丹妙药”,作为一种数据分析方法,它在很多方面依然具有局限性,不能过度夸大其功能,否则将陷入对技术的迷信和盲从,进而将刑事司法大数据的发展方向引入歧途。因此,本文分别从实务工作、刑事司法的运行环境以及刑事司法理论研究三个方面,对大数据给刑事司法带来的影响进行探讨,以明确大数据在刑事司法领域的效用空间,同时提出在发展和运用大数据时应当注意的问题。

2 大数据对实务工作的影响

一直以来,刑事司法实践也经常需要进行数据统计和数据分析,并且这些工作对于刑事司法实践也具有十分重要的价值。例如,通过对破案率的统计可以反映侦查机关在过去一段时期的办案效率;对某法院几年内审理案件的数量进行分析,有利于了解该辖区出现法律纠纷的情况;对社会公众关于刑事判决的态度进行抽样调查,可以了解社会公众对于法院的认可度情况。而相比于以往这些分析方法,大数据在数据分析方面带来的作用更加强大,因此也能够给刑事司法实践带来许多不同的影响。

2.1 大数据对刑事案件侦查的影响

在所有的刑事司法实践活动中,侦查活动对数据分析的依赖程度是最高的,因为数据规律往往能够在刑事案件的侦办过程中发挥非常重要的作用。

实际上从很早以前,犯罪侦查就已经开始使用数据分析作为拓展线索和分析案情的工具。例如英国为侦破连环杀人案,就利用已有案件的信息对犯罪地理数据进行分析,并根据分析结果判断杀人现场、抛尸地点以及犯罪嫌疑人藏身处之间的关系,从而确定侦查的方向,并最终破获几起著名的连环杀人案,从而开创了犯罪地理学分析的先河[1]。除了这些著名的应用实例之外,在一般的侦查活动中,侦查人员结合以往案件侦破的经验信息,根据当前案件的现场痕迹以及其他证据信息,就可以从体貌特征、职业、心理等方面对犯罪行为制作犯罪心理画像,并利用此画像确定嫌疑人的范围,这也是对数据分析的典型应用[2]。

从刑事案件侦查对于数据的使用方式来看,传统的许多分析方法实际上已经体现出大数据思维的雏形。信息碰撞、摸底排队、拉网排查等常用的侦查措施都是与当前所说的大数据有异曲同工之处。此外,对足迹数据库、指纹数据、油漆数据库、DNA数据库以及串并案系统的开

发和利用,也都体现出对数据价值的充分重视。应当说,在对数据的分析和使用方面,无论是思维上还是技术上,侦查领域都要比其他刑事司法实践部门领先很多。只是在这些传统的分析活动中,这些方法都是在案件办理过程中不断摸索形成的,并未如当前的大数据一样经过系统的研究和开发。

随着大数据的兴起,对于这种数据分析方法和具体应用的开发逐渐成为显学,并成为信息技术领域的独立学科。基于对于数据分析的现实需求,侦查机关也率先开始对大数据进行应用。目前,大数据在侦查领域已经产生了许多应用实例,并且对案件侦破带来了很大的促进作用。例如在2013年美国波士顿爆炸案的侦查过程中,当地警方将收集附近12个街区广大居民私人拍摄、录像终端的图片视频资料,通过对与案件有关的海量数据以及现场数据采集进行比对鉴别,短时间内就迅速缩小范围并确定了犯罪嫌疑人的图像特征[3]。

在我国,大数据在侦查中也正在逐渐发挥着作用。例如在对涉嫌伪造、套用车牌的车辆进行筛查时,就是利用大数据的分析方法从海量的车牌中将关键特征分析出来,在几分钟时间内就可以锁定特征车辆以供进一步排查和确认。各级检察机关也正在加紧研究,如何运用大数据技术和思维,对来自各领域与金融往来有关的数据进行分析,从而获得侦破职务犯罪的重要线索。

大数据的理念能够第一时间被侦查工作接受并转化为破案利器,其中最主要的原因是侦查活动天生就对数据有敏锐的“嗅觉”。对于侦查人员而言,每一个待侦破案件都是一道结果未知的题目,所有可被利用的信息就是解决这些题目的已知条件。在对案件的侦办过程中,侦查人员需要对各种信息加以搜集、提取、固定和保全,并通过这些信息完成对案件原貌的“拼图”。为了完成这一工作,单纯依靠各种信息直接表达的内容是远远不够的,需要侦查人员发挥主观能动性,充分运用分析能力和判断力。当然,侦查人员对信息的分析和判断并不是主观臆测,而是依据信息与案件事实之间以及信息与信息之间所具有的客观联系。因此,以数据相关性为挖掘对象的大数据分析方法,与侦查过程中的信息处理方式有内在的契合之处。

侦查工作本身就是建立在对信息的收集与分析基础之上,而大数据的引入能够在数据分析的广度、数据分析的速度以及数据分析的思路等多方面给侦查数据分析工作带来更大的提升。目前所展现出还只是初步效果,随着大数据与侦查工作的进一步深入结合,它很可能会给侦查工作带来更大的变化。

然而,在意识到这些积极影响的同时,也不能忽视大数据在侦查工作中运用时可能产生的一些负面效果。

诚然,从分析案情和拓展线索的角度来看,大数据无疑会成为“破案利器”,但是侦查机关所能接触到的数据有很大的特殊性,不仅有与案件事实和相关嫌疑人有关的信息,同时也有大量与此无关的“无辜者”的信息,甚至包括许多敏感的个人信息。虽然侦查活动在法律允许

的范围内,可以根据案件侦查的需要,合理的对公民信息进行使用,但是这种使用行为必须受到法律的严格限制。如果突破这些限制,虽然对于侦破案件是有利的,但是对于公民的合法权利以及基本人权则可能构成严重的侵犯。例如,随着移动通讯设备的普及,通讯数据已经成为十分重要的个人信息来源,如果能够采用大数据的方法对海量的通讯数据进行分析 and 碰撞,必然能够对侦查工作提供大量的线索,但是这种作法也将涉及到对普通公民的通讯进行普遍监控和监听,否则就无法得到大量数据来源,这就涉及到对公民通信自由和通信秘密的侵犯。因此,类似这样的大数据运用方式是应当受到严格限制的,即便真的有必要使用,也需要有相应的法律和配套制度对其进行约束。2014年,美国联邦第四上诉巡回法院在 *United States v. Quartavious Davis* 一案中作出判决,认定警方在没有获得搜查令的情况下随意获取公民手机的位置信息违反了宪法第四修正案^①。该判决作出后,对于美国警方通过广泛搜集并使用公民位置信息进行大数据侦查的手段形成了限制。

由此可见,由于刑事案件侦查活动的特殊性,法律赋予侦查机关较大的数据使用权限,因而导致侦查机关往往会占有数量十分可观的数据资源。在这种条件下,合理的运用大数据必然会成为刑事案件侦查的有利工具;但是,如果大数据被滥用,也同样会带来非常严重的危害。

2.2 大数据对犯罪预防工作的影响

从狭义上看,对犯罪行为的预防应当属于社会治安工作的一部分,因此许多工作并不属于刑事诉讼程序的范畴。但是,在通过侦查和审理刑事案件对犯罪行为进行打击的同时,刑事司法也承担着教育公众和预防犯罪的职能。并且,犯罪预防的对象是各种潜在的犯罪活动,与纯粹的治安管理也有所区别,在工作上与侦查工作有诸多的交叉。因此,犯罪预防也属于刑事司法所应考虑的一项重要工作内容。

犯罪预防的对象是潜在的犯罪行为,这些犯罪行为尚未实际发生,只是存在不同程度的可能性而已,这是犯罪预防与犯罪案件侦查相比最本质的区别。例如,对某类人群是否可能从事犯罪或者成为某些犯罪行为潜在的侵害对象、某类财物是否可能成为犯罪目标。甚至可以通过分析,判断出特定情形下,潜在的犯罪行为人可能使用的犯罪手段以及可能带来的突发性后果,从而使犯罪预防工作可以更加有针对性,同时兼顾对可能发生的危害的预防和损失控制。除此之外,对一个区域甚至全国范围内存在何种犯罪趋势、犯罪动向进行预判,这也是犯罪预防工作的任务。

犯罪预防实际上是对犯罪行为和犯罪趋势的预测。具体而言,就是根据已有的犯罪案件对犯罪人的行为特点、诱发犯罪的影响因素以及其他影响犯罪行为实施和发生变化的情况进行规

^① See *United States v. Davis*, 573 Fed. Appx. 925 (11th Cir. 2014).

律总结,再针对当前行为主体的特征和行为习惯以及经济、文化、环境等因素,从而对可能发生的行为或者变化趋势进行预测。这种预测的思路与商业领域通过数据分析对消费动向的预测基本一致,这充分说明在对待数据的思维方面,犯罪预防与大数据是完全契合的。

实际上,利用朴素的大数据思维对潜在犯罪行为进行预测,在犯罪预防领域已经有了比较成熟的研究和运用。比较典型的就是根据已有犯罪案件信息,针对地理环境因素与犯罪行为之间的相关关系,实现对特定区域犯罪行为规律和犯罪行为发展趋势的预测,这种分析方法被称为“犯罪地理学”,目前已经发展成为一门独立的学科,在各国犯罪预防研究领域都发挥着重要的作用[1]。

由此可见,对于犯罪预防工作而言,大数据的发展和运用无疑是一次良好的契机。借助大数据对数据资源的整合,以及其在全样本分析和挖掘方面所体现出的能力,对于犯罪行为规律分析也会更加的全面和细致,从而也会使对潜在犯罪行为的预测更加准确、更加高效。因此,仅从犯罪预防的效果来看,大数据势必会带来巨大的促进作用。

但是在关注到这一有利影响的同时也应当注意到,围绕着犯罪预测产生的争议也是十分明显的。

技术研究领域普遍认为,只要有足够的行为数据,大数据就可以达到深入人思想的效果,从而实现对广泛人群的未来行为进行预测,这对于犯罪预测而言无疑是一个巨大的诱惑。如果实现了这样的功能,那么犯罪预防工作将变得空前高效,甚至可以如电影《少数派报告》^①中描绘的那样,在犯罪行为发生之前就能准确的将全部犯罪事实以及即将犯罪的行为主体预测出来,届时犯罪预防也将成为整个刑事司法的核心工作。

然而,这样的结果真的是我们所希望的吗?这种发展思路实际上是对19世纪以来针对个体的“天生犯罪人”思潮的延续,而这种预测行为是十分危险的,可能导致将人身危险性与实际犯罪行为的割裂,进而造成仅凭借人身危险而非实际的犯罪行为对特定主体进行限制甚至处罚[4]。大数据所提供的只是相关性,而这种预测效果实际上是利用经验的相关性替代了法律依据人的客观行为与相应法律后果之间的因果关系。“如果我们能够在某些犯罪发生之前就预测(而非证实)其发生,可能会颠覆作为个体承担法律责任之基础的自由意志假定。”[5]即便最终并未真正因为预测的结果就对“危险个体”施加刑罚,仅仅是因为其潜在的犯罪可能性,就对其采用监控措施甚至对人身自由以及其他合法权利进行限制,那么都会使犯罪预防措施本身就带有了侵害性。

综上所述,由于在思维和方法上的高度契合,大数据必然能够很直接的被运用到犯罪预防

^① 《少数派报告》是改编自菲利普·迪克的短篇小说《少数派报告》,由史蒂文·斯皮尔伯格执导,汤姆·克鲁斯、柯林·法瑞尔、萨曼莎·莫顿等主演的科幻悬疑电影。

工作之中并发挥巨大的功效。但是,当大数据将犯罪预测的能力大幅提升时,我们也有必要思考,对于这种基于经验规律得出的预测结果,其应用应当如何被限制在一个合理的范围之内,以避免在预防犯罪的同时造成新的侵害。

2.3 大数据对刑事案件审判的影响

目前,对于大数据在刑事案件审判中所能发挥的作用可谓是备受期待,无论是在各实务部门还是学术界,都纷纷对此进行了诸多的预期和列举。特别是在2013年之后,随着全国裁判文书上网公开,中国裁判文书网正式建立,海量的裁判文书首次成为触手可及的数据资源。通过裁判文书对我国法院在案件审判时的态度和趋势进行分析,已经被许多人视作大数据在司法审判中发挥作用的美好前景。

从数据规模的角度来看,我国法院早已拥有了海量数据的基础。新中国建立的60多年时间里,全国法院审理的案件总量已经有近30亿之多,在每一个案件中又都包含着当事人信息、案件事实信息、证据信息以及相关的大量案卷材料[6]。将这些信息汇集在一起,其规模是十分庞大的。在过去一直没有找到合适的利用途径时,这些信息不过就是深埋于档案馆中的资料,随着时间的流逝而慢慢变为历史。而在大数据的视野下,这些数据的价值就显现了出来,成为深入了解审判工作规律和发展趋势的宝贵资源。

大数据刑事案件审判中有以下几个方面的作用是值得期待的:

首先是在事实认定环节,大数据能够给法官带来更加客观、准确和高效的分析结果。在刑事案件的审理过程中,事实认定是核心任务之一。准确的认定案件事实,是最终对被告人定罪量刑的前提和基础,并且对刑事司法的客观和公正有着决定性的影响。在经历了神示证据和法定证据等历史阶段之后,目前无论是大陆法系还是英美法系,都要求事实认定者在依据证据的前提下,需要凭借自身的经验、理性和良知对案件事实作出认定;而在具体适用证明标准时,无论是排除合理怀疑还是内心确信,其最终也需要依靠经验理性来进行判断。因此,在对案件事实进行认定的过程中,事实认定者基于经验事实得到的认知在其中发挥着关键的作用。从分析和判断的过程来看,大数据的分析结果也是基于海量的既有事实信息形成的经验性结论,这一点与事实认定过程有很大的相似之处。然而人的脑力以及精力和阅历都是有限的,大数据却能够凭借强大的计算能力对海量信息进行高效、全面的分析,这显然在学习和认知的能力上远远超越了人所能做到的程度。并且,由于不同的人有不同的学习和认知经历,并且在判断过程中不可避免的要掺杂部分主观因素,导致其对相应事实形成的经验判断往往会有所差异,这就可能形成对事实认定的偏差,或者是出现难以做出判断的情况;而与之相比,大数据则不受这些因素的影响,只要依据相同的数据,其结果只会忠实于数据的真实情况,形成更加稳定、客观的结论。

其次是在定罪量刑方面,大数据也能发挥积极的作用。在准确认定案件事实的基础上,判断相关涉案行为及涉案事实是否构成犯罪,以及应当对被告人处于何种刑罚,这也是刑事审判工作的重要任务。然而在司法实践中,存在着对于相似性质的案件事实最终得到的判决结果却有很大差异的现象,即“同案不同判”。这种现象的出现,与法律适用的统一性原则形成矛盾,同时造成社会公众对于刑事审判的权威性产生怀疑。这种现象存在的一个主要原因是法官对法律适用的理解存在偏差。尽管法律规定是全国统一的,但是对法律的理解和认知却与适用法律的主体有密切的关系,无法避免受到审判主体的主观倾向性以及经验认知上存在偏差所带来的影响。而大数据是基于客观数据得出的分析结论,可以完全避免主观因素干扰,同时也可以为定罪量刑提供了统一的经验规律。因此,如果能够将大数据运用到定罪量刑之中,则可以提高法律适用的客观性和统一性,大大减少甚至避免“同案不同判”现象的发生。

由上可知,无论是在事实认定方面还是在定罪量刑方面,大数据之所以能够给刑事审判带来积极影响,主要原因都在于其在经验数据的分析方面体现出的客观性和高效性都是人所无法做到的。大数据能够消除人的主观因素对审判活动带来的影响,同时也能够在客观上展现出比人更为强大的分析和判断能力。

因此不难形成这样的假设:如果能够提供足够数量以及足够全面的数据,并且有可靠的、准确的分析模型和算法,大数据在审判的准确性和合理性方面完全可能超越人的能力。那么,大数据真的能够代替人来完成刑事审判工作吗?刑事判决是否真的能达到“自动售货机”^①一样的效果呢? [5] 对于这样的设想,需要从以下几个方面进行认真的思考:

首先,足够数量以及足够全面的数据是大数据能够得出准确的经验规律的前提。从目前的情况看,虽然裁判文书的量已经积累到很大,但是其全面性显然是不够的,至少还应当有包括全部证据材料在内的详细的案卷材料以及以往审理过程中对事实认定的具体理由,这些对于相关性的分析尤为重要。在没有这些信息的情况下,是无法形成满足事实认定所需要的经验规律的。即便根据裁判文书数据分析出一些结论,那也只能是在某些特定的方面能够提供预测和判断作用,并不能完成全部的事实认定任务。

其次,即便数据质量满足了要求,还需要有可靠的、准确的模型和算法来实现分析和挖掘。这可能是目前面临的障碍。将事实认定以及法律适用的逻辑完全转变为程序模型和算法,这在技术上还需要经历大量的研究过程,并且需要大量验证之后才能确定其是否准确可靠。

最后,即便我们拥有了这样的模型和算法,我们真的就要将审判交给“代码”来完成吗?十多年前就有人提出网络空间中“代码就是法律”的断言 [5]。一些技术领域的研究人员也多

^① 德国思想家马克斯·韦伯提出过“自动售货机”的比喻,意思是主要输入案件证据材料,法院就会自动吐出相应的判决,这一设想曾经引起计算机能否定罪量刑的讨论。

次表示,程序比人更可靠,其出错的概率要远远低于人的出错概率。但是,如果利用大数据对经验事实的分析结果作为认定事实的依据和标准,那么就等同于将“描述性的遍历性事实作为一种规范性规则施加给每一个人”[5],这实际上是取代了立法的作用。同时,随着数据分析的复杂化,程序和算法也会变得越来越复杂难懂,除了技术人员之外,普遍的社会公众是很难明白这其中究竟是怎样的逻辑,这就导致整个过程已经脱离了有效的监督。当算法复杂到一定程度而无法被人直观理解,形成“黑箱”的时候,这将是危险的[7]。

由此可见,就目前的情况来看,大数据还不能替代人对刑事案件进行审判。当然,在刑事案件的审理过程中,大数据可以提供基于数据分析得出的材料以供参考,这已经能够给审判工作带来极大的帮助了。“鉴于中国地方法院往往会深入介入本地政治经济社会的发展,因此在司法统计基础上发展起来的大数据分析更有可能加强各级法院整体上的能动性,而非在个案中提供指引。”[8]而对于应当如何在个案中对事实和法律问题作出最终的评断,则依然需要由人来完成。

3 大数据对刑事司法运行环境的影响

大数据除了能够为打击犯罪、预防犯罪和刑事审判工作提供有力的数据支持以外,在促进司法公开、掌握并引导社会舆论、了解司法运行情况、对司法工作人员进行客观评价以及合理配置司法资源等方面也能够发挥积极的作用。这些作用对于维护良好的刑事运行环境都有着重要的影响。

第一,大数据有助于提升刑事司法的公开度和透明度。司法公开一直是宪法和法律确定的基本原则。司法权的行使关系到社会正义和公民权益,应当和其他公权力一样在阳关下运行。司法公开是促进司法民主的基础,也是实现司法公正的保障,同时也是树立司法公信力的有效途径。正如丹尼勋爵所言:“正义不仅要实现,而且要以看得见的方式实现”。近几年,在刑事司法信息化的不断推动下,我国在侦查、起诉和审判等环节的信息公开程度正在日益提升。相比于普通的信息化,大数据在信息共享和信息分析的速度上会有质的飞跃,从其他领域的应用实例可以看出,大数据往往可以做到对信息的实时呈现,即在信息的产生和流转过程中,相关的信息内容以及信息所反映的情况都可以即时表现出来,这也意味着司法信息的公开可以做到与案件办理过程同步。此外,大数据不仅仅是对信息内容的反映,其主要作用在于能够将信息的变化趋势和变化规律也都呈现出来,这样就会使公众对于刑事司法的运行情况有更深层次的了解。因此,随着大数据的发展和运用,刑事司法公开无论是在效率上还是程度上,都会有很大的提升。

第二,大数据在网络舆情分析方面的作用,能够使司法行政机关更好的了解社会公众对个

案的反馈以及对整个刑事司法工作的评价。“舆情”是近几年比较火热的词汇之一，这个词最早出现在《旧唐书》中：“朕采于群议，询彼舆情，有冀小康，遂登大用。”[9]发展至今，虽然对于舆情已经形成了很多不同的理解[10]。但是依然可以将其简单的理解为根据社会舆论了解到的社会公众对相关问题的态度。由于互联网的普及，网络已经成为大众舆论的主阵地，网络舆论也就成为了反映舆情信息的主要来源，这就是网络舆情的分析。舆情所反映的内容是多种多样的，既有对政治的态度，也有对经济形势的态度，当然也有对刑事司法的看法和态度。这其中比较明显的是对热门案件办理情况的反馈，例如在徐玉玉案、药家鑫案、快播案等案件诉讼过程中，就有大量的舆情信息产生，反映了社会公众对这几个案件的不同态度。除此之外，对于我国刑事司法工作的整体运行情况，社会公众也会有相应的评价，形成相应的舆情信息，例如对某项侦查强制措施的看法等。舆情就好比一面镜子，能够从社会公众的视角将刑事司法的运行是否良好以及存在什么样的问题清晰的映射出来，了解这些舆情对于刑事司法改革具有十分重要的指导意义。当然，也有一些舆情表现的是社会公众对刑事司法机关以及相应活动的误解，主要是因为社会上存在一些不负责任的恶意造谣、抹黑和煽动性言论等在误导公众，对于这些舆情信息也需要及时进行收集，并采取相应的措施对社会公众进行引导，防止这些不利影响造成更恶劣的后果。因此，无论是对于何种情形下产生的舆情，各级刑事司法机关都应当予以充分的重视，尽早加以掌握和妥善引导。在网络舆情的收集和分析方面，大数据具有得天独厚的优势。大数据最初的发展就是针对互联网信息的分析和挖掘，这一优势恰好可以在网络舆情分析方面得以充分发挥。并且，大数据不仅仅是信息收集工具，同时也能够对相关信息的发展变化趋势以及各种影响舆情的因素之间存在的相关关系进行更深入的揭示，这种高效、深入的分析对于掌控舆情引导社会舆论都具有非常重要的作用。

第三，运用大数据，能够有利于及时发现刑事司法运行过程中存在的问题，并深入全面的分析出其中相关的影响因素，从而可以有针对性的对相关制度、政策和法律规范进行调整。刑事司法是一个复杂的程序体系，所有刑事诉讼活动都必须在遵守《刑事诉讼法》以及相关法律的规定下进行。然而，法律并不能预先就对所有事项都规划得尽善尽美，在刑事诉讼程序的实际运行当中，必然会有许多的问题暴露出来。这些问题有的是法律执行不严造成的司法人员滥用权力所导致的，而有些则是因为制度设计上存在不合理之处所造成的，必须予以及时的发现和解决，否则这些程序和制度上的问题可能最终影响到刑事诉讼的整体效率，甚至会影响到刑事审判结果的客观、公正，危机司法的公信力和权威性。而这方面的要求恰恰属于大数据的可实现功能范畴。每一起案件的相关数据不仅会反映案件的事实和证据信息，同时也会将相关的程序信息记录其中。对这些分散于每个案件中的碎片化的信息进行统计和分析，就能够发现哪些问题是普遍存在的。而基于大数据在数据相关性上的挖掘作用，发现问题的同时也能够将与

这些问题密切相关的事项、程序和因素都分析出来,这样就可以据此进一步采取有针对性的措施以解决问题。

第四,大数据的分析结果可以被作为评价办案人员是否具备业务能力以及是否符合职业规范的评价依据。刑事诉讼程序并不是一台自动运行的机器,每一个环节的运行都是在相关办案人员的主导下进行的。因此,这些办案人员必须在业务能力上达到相应的要求,保证能够胜任并准确完成相应的任务;同时,办案人员的行为需要符合职业道德和职业规范的要求,不能恣意妄为,更不能枉法裁判胡作非为。对于办案人员是否符合这些要求,可以通过其在日常工作中的表现进行判断,而这种判断过程则可以利用大数据的方法从其以往办案过程中形成的各类数据中分析得出。例如,对于法官办理的所有案件数据进行分析,宏观上的统计可以得出其错案率、上诉率等信息,而深入的挖掘则可以进一步发现其在判决中是否会体现出特定的倾向性、对于那些问题的判断总是会容易出现偏差等。相比于主观评价的模式,基于数据分析得出的结果更加客观,能够更加真实、准确的反映办案人员在各方面的表现。因此,大数据的分析结果可以作为一种对办案人员进行评价和考核的可靠依据。例如我国目前正在推行法官员额制改革,其中就涉及到对相关人员在业务能力和职业操守方面是否符合对法官的要求进行评价,这种时候就需要有类似于大数据这样的客观依据。当然,除了对特定司法人员的评价提供依据之外,大数据还能够从整体上分析我国司法人员职业队伍的业务素质情况。这种分析结果对于改革司法人员选聘机制、调整司法职业队伍人才培养方案等方面也都能够起到非常大的作用。

第五,大数据在合理配置司法资源方面也能够发挥积极的作用。在我国,因为不同地区的实际情况不同,对于司法资源的需求也会有所差异。例如,在北京上海等城市,虽然法官和检察官的人数配置也要远远多于许多边远省份,但是这些办案人员的工作负担依然十分繁重,人均年办案量也排在全国前列。司法资源是一种公共资源,在保障我国刑事司法质量同时,也需要节约资源,优化配置,提高单位资源的利用效率。然而,为了实现合理配置司法资源,必须以了解各地对于刑事司法资源的需求情况为前提,从而作到按需分配,即不造成司法资源的短缺,同时也不能无端的浪费。在这方面可以依靠大数据对现有各地区的数据进行分析,从而全面细致的了解各地区过去对司法资源的使用情况,并对几种调整资源配置的方案所对应的效果进行预测,从而避免依据片面认识导致的无功而返和资源浪费。当然,基于大数据对司法资源进行配置,其作用不仅仅是体现在这些传统的区域性配置模式上。由于大数据在数据分析方面能够作到比传统分析更加细致且更加高效,因此在利用大数据进行资源配置时,可以将配置方案更加细化,即根据更加具体的情形划分来规划资源的使用。例如,通过大数据对一个地区治安和犯罪现象的分析,可以预测出未来不同的时间段以及不同事件发生时社会治安需要面对的风险大小以及潜在风险的可能来源,依据这些信息就可以进一步精细化的调整警力配置,比如哪些

时间可以进行轮休、什么特定情况需要借调武警或者其他辖区的警力等。这种“精细化”的资源配置模式是大数据带来的新思路和新方法,这种模式会极大的改变现有的司法资源配置思维,也许能够使我国有限的司法资源发挥出更大的效能。

4 大数据对理论研究的影响

众所周知,理论研究并不是凭空产生的,需要以客观现实为基础。因此实证研究一直是理论研究的重要方法,脱离实证的理论研究往往会给人以空谈的印象,不能对实践产生指导意义,并且也很难在理论上有进一步的发展。在刑事司法领域里,许多著名的理论成果实际上都是建立在长期的实证分析和经验总结的基础上才形成的。而大数据可能是目前最为先进的实证分析方法。与大数据基于海量数据和全样本分析为基础的分析思路相比,传统的社会统计学分析就如同摆在计算机面前的算盘一般简陋。

以往的实证调查和实证分析都是采用抽样统计的方法进行的。对于从事法学研究的人都清楚,在这种抽样分析中,样本的数量对于分析结果是否符合客观实际有很大的影响。例如在对某地区基层法院进行实证调研时,抽取十个案例进行分析与抽取一千个案例进行分析,其结果的丰富性和可信度显然都是不同的;那么如果能够对该法院历史上所有的案卷进行分析,其结果将会是最全面和可信的,只是对类似这种案卷分析的工作往往需要依靠人力来完成,而受到思维能力和生理机能的限制,这种分析一直以来只能是不切实际“奢望”。但是这种想法是与大数据的全样本分析的思维完全一致的。因此大数据对于刑事司法理论研究而言无疑是一种不可抵挡的诱惑。

当前,我国在刑事法律的理论研究方面面临的重大问题,就是理论与实际相脱节。许多研究不了解实际情况,只是根据媒体的报道或者其他学者的言论就“人云亦云”。当理论脱离实际时,理论研究就会变成空中楼阁。法律是一门实践科学,正如美国最高法院霍姆斯大法官的名言:“法律的生命不在于逻辑,而在于经验。”^①因此,刑事司法的各项理论也都是由实践中产生并最终用于指导司法实践的。基于客观数据的经验分析,可以有效的预测出各种因素对法律运行情况的影响,将问题揭示出来。对问题进行有针对性的研究,方能使研究真正的有的放矢。然而,在导致我国刑事司法理论研究普遍缺少充分实证分析的各种原因中,排除研究者个

^① 霍姆斯关于“法律的生命不在于逻辑,而在于经验”的观点最早出现在1880年他对兰德尔论合同法的书所撰写的评论中,随后,在其同年出版的巨著《普通法》开篇,他再次重申了这个观点:“法律的生命不在于逻辑,而在于经验。对时代需要的感知,流行的道德和政治理论,对公共政策的直觉,不管你承认与否,甚至法官和他的同胞所共有的偏见对人们决定是否遵守规则所起的作用都远远大于三段论。法律包含了一个民族许多世纪的发展历史。它不能被当作由公理和推论组成的数学书。”如果不只是单纯的将“大数据”视作某种机器或算法,而是将其视作是一种只关注相关性而不去总结因果律的信息分析方法的话,其信息录入和信息产出的形式,就与我们通常所说的“经验”有很大的相似性,尽管它是机器和算法得出的。

人的习惯之外,还有两个重要的客观原因:其一是我国关于刑事司法的实证资料公开程度较差,大量与实证有关的一手资料很难被学者轻易获得,造成实证数据资源匮乏的局面;二是分析能力受限,刑事司法资料大多为文本材料,通常只能依靠人工分析,而且每个案件资料中都包含大量的信息,仅凭人的能力,即使有足够量的数据也很难完成分析工作。

因此,大数据的发展也许能够给刑事司法理论研究带来新的实证分析思路和分析工具,其主要作用可以归结为以下几个方面:

首先,能够引发在实证分析思维上的转变。随着大数据的产生和发展,在法学研究领域已经掀起了一场关注实证数据分析的浪潮。众多学者都对数据所能够揭示出的信息和规律寄予厚望,并期待由此可以带来理论研究上的深入和创新。大数据在数据处理思维方面的转变,也会带动刑事司法理论研究开始关注数据的广度能够带来的价值。也许未来对基础文本资料的结构和数据化可能会成为新的研究起点,因为这些工作将使大量的法律文本信息转变为有价值的数据,从而进一步增加可供大数据使用的基础数据资源。

其次,发展大数据能够带来海量数据的汇集,这会给传统法学研究带来前所未有的丰富资料。由于刑事司法实践同样对大数据分析有所倚重,海量的案件资料也会因此从“档案”变成“数据”。这些数据的公开和共享,也会同时给理论研究带来海量的一手资料。例如在2013年中国裁判文书网建成之后,已经有众多的学者开始将实证研究的目光几种到这个新兴的数据资源上。

最后,大数据对现代信息技术的灵活运用,能够给刑事司法理论研究带来强大的数据分析能力,从而使全样本分析在刑事司法实证分析中也有可能从理想变为现实。当然,这种分析能力的提升并不只是通过借用现代计算机技术就可以实现的,还需要在软件开发结合刑事司法数据的特点和实证分析的具体需求进行专门的定制和开发,而对于这种软件的研发属于技术和法学的交叉研究,实际上相当于一个新的学科领域,从目前来看,这项工作还任重而道远。

在为大数据能够给刑事司法实证研究带来强大作用而欣喜的同时,我们也应当注意到大数据只关注数据相关性而非因果性的特点。因此,大数据给刑事司法理论研究带来的只能是基于实证数据的经验规律,这种经验分析也并不是万能的。

首先,经验分析除了有海量数据作为支撑之外,还需要有适当的分析模型。在数据分析中,数据扮演者原材料的作用,计算机的运算能力是纯粹的工具,而分析模型则是大脑。缺少大脑的指导,只有其他两个方面也无法得出准确、可靠的分析结果。正如前文所言,要对刑事司法数据进行分析,就需要在软件上进行专门的定制和研发,其核心就是对分析模型的构建。在分析模型的构建中,关键是对刑事司法信息的解读逻辑,这是信息技术所无法提供的,只能由刑事司法理论研究来完成,并且最终还要将之转化为计算机可以解读和运行的逻辑,实现两个学科领域在思维逻辑上的深度融合。只有在这项复杂的任务完成之后,大数据才可能在经验分析

方面带来我们所期待的最佳效果。

其次,经验分析只是对现象的客观呈现,但是对于这些现象产生的原因,依然需要在这些经验结论的基础上进行认真细致的剖析。法学的理论与商业领域追逐经济利益是不同的,仅仅知道现象的相关性和未来的可能性并不是研究的最终任务。把握现象背后的规律,将这些规律转化为对刑事司法基本理论和基本原则的理解和创新,这才是刑事司法理论研究所要实现的价值。对于法律精神和法律价值的追寻,这项极具人文精神的探索性工作是无法被程序和代码所替代的。

因此,大数据能够为理论研究提供丰富、客观的实证资料,理应成为理论研究的有利工具。但是,理论研究不能止步于此,深入的分析 and 研究无论对于刑事司法大数据的发展,还是对于刑事司法理论自身的发展和创新,都是极为重要的。

5 结语

自从大数据在商业领域发挥出强大的分析和预测功能并带来以往所无法想象的巨大收益以来,各领域都纷纷掀起了发展大数据的狂潮。人们追捧大数据,甚至将大数据视作能够给传统行业带来革命性发展的灵丹妙药。这一现象也延续到了司法领域,许多司法人员和法学的研究人员好像一夜之间都成了工程师或者产品经理,纷纷围绕大数据在司法领域中的应用以及其可能得到的效果做出各种设想,似乎一旦将大数据应用于司法,就可以彻底革除目前司法制度和司法实务中所有难以根治的顽疾,打造出一个人类历史上从未有过的、极致公正和高效的司法机器。然而,这样极端的想法显然是不切实际的,甚至是带有浓厚“科幻”色彩的。只有当我们冷静下来,细致的分析了大数据究竟在司法领域中有哪些是可为的,哪些是不可为的,以及还有哪些是根本不属于大数据的功能范畴而是需要法律和司法制度自身加以改变的,我们才有可能真正看清司法大数据的发展方向。

从前文的分析中不难看出,直到目前为止,发展刑事司法大数据最主要的动机,依然是希望大数据技术在其他领域中所发挥出的显著优势,能够在刑事司法领域中得到“复现”。那么,大数据的“魅力”究竟体现在哪里呢?

首先,强大的数据检索和数据分析能力。基于由海量、全面的数据构成的数据仓库,大数据的应用可以为人们提供精准的数据检索功能,同时还能够依托于完善的算法和强大的算力,对这些数据进行统计分析、碰撞分析等。刑事司法领域的“数据”,主要是以文本的形式存在的。而事实上,从目前大数据在技术方面的发展情况来看,已经可以胜任这种针对纯文本性数据的分析工作。因此,从理论上讲,大数据在其他领域可以发挥出的数据检索和数据分析能力,完全可以移植到刑事司法之中。

但是,由于我国刑事司法领域一直处于较低的信息化水平,全行业对“司法信息化”的认识,实际上还停留在“数字化办公”“电子卷宗”的水平上。在大数据分析和计算机程度的视野下,对于一个刑事案件或者是具体的刑事案卷而言,究竟什么才叫数据?怎样才是结构化的数据?目前无论是理论界还是实务界,多数人还缺乏必要的认知,更不用说是实现案件信息数据化了。那么,在这种缺少司法数据基础、司法数据意识以及司法数据理论的情况下,拿什么给大数据分析提供基础的数据仓库?没有数据仓库,哪来的数据分析结果呢?因此,要发挥大数据对刑事司法的积极影响,必须先从理论和实务层面都完成刑事司法“信息化”和“数据化”的建设,让刑事司法的方方面面都能够通过数据呈现出来。否则,片面追逐大数据技术应用,无异于缘木求鱼。

其次,神奇的预测能力。虽然从本质上看,大数据也属于数据分析,但是由于其分析对象是全样本和海量的数据,因此,往往可以依据过去数据所呈现出的相关性,实现对未来数据变化趋势的预测——这项能力已经在商业领域和流行病预防领域获得了非常明显的成功。理论上讲,只要能够给大数据模型提供充足、全面的历史数据,就能够对未来遵循同样规律发展的事物作出预测。对于刑事司法而言,虽然也需要经常对相关法律法规和政策进行调整,但是从整体上看,无论是办案规律还是法律原则,都是以稳定为主旋律。一个国家的刑事法律制度和刑事政策不可能总是在发生颠覆性变化。因此,如果有全面的刑事司法历史数据,实现对犯罪趋势的预测、裁判结果的预测甚至是对具备某些特征的主体是否具有社会危害性和刑事犯罪风险的预测,从单纯的数据分析角度来看,都是可行的。

但是,正如前文在“大数据对犯罪预防工作”部分对“个体犯罪可能性预测”以及在“大数据对刑事案件审判的影响”部分对“自动贩卖机”理论这两个问题的阐述,如果要在这些方面使用大数据的预测功能,并且要以此替代人的内心确信的话,必须要面临巨大的障碍:我们在数据质量上,能否满足大数据的“全面性”要求?至少目前,国内正在开发和运行的刑事大数据项目中,还没有任何一个可以获得完整的刑事案件卷宗材料。如果数据不满足“全面性”,大数据是无法给出准确预测的。此外,即便每个历史案件都能拿出完整的案卷材料,那么这是不是就可以认为是“全面数据”呢?人的内心选择、伦理道德、价值判断是否有所反映?即使有反映这些信息的数据,我们又能否认为其他案件中其他人的数据对当前案件中的主体以及主体的犯罪行为也是有“相关性”的?就现有的法律原则和规则看来,显然不能。所以可以得出这样的结论:至少在现阶段,大数据在刑事司法中针对个案的预测是技术上不可行的,也是不精确、不客观的。至于如果真的技术上可以实现时,它是不是能够实现用机器替代人来办案、审案?笔者认为,这个问题可以留待条件成熟时再详细讨论不迟。

综上所述,随着以“互联网+刑事司法”为导向的刑事司法信息化不断发展,越来越多的

大数据应用出现在刑事司法领域将会成为必然的趋势，我们也都共同期待刑事司法大数据在未来的发展能够成为引领刑事司法改革的主要推动力。但是，实现这一目的不能依靠对大数据技术的“迷信”和“盲从”，更不能简单奉行从其他领域的“拿来主义”，而是要加强刑事司法的数据化建设，至少要按照符合数据分析要求的条件开始积累和建立能够全面反映刑事司法各方面信息的数据仓库；同时，加强理论研究，逐渐形成关于刑事司法的数据化理论和数据分析方法理论。只有如此，方能使刑事司法大数据走上理性的发展轨道。

基金项目

教育部人文社科项目“刑事证据信息管理的理论和方法研究”（项目编号：19YJC820029）；霍英东青年教师基金项目“刑事诉讼电子数据证据规则”（项目编号：161083）；中国政法大学青年教师学术创新团队支持计划资助。

参考文献

- [1] 李小恺，于冲．犯罪地理学模型在毒品犯罪区域防治中的构建及应用[J]．北京警察学院学报，2013，6.
- [2] 李玫瑾．侦查中犯罪心理画像的实质与价值[J]．中国公安大学学报（社会科学版），2007，4.
- [3] 美国联邦机构利用大数据的五个例子（Five Examples of How Federal Agencies Use Big Data）[EB/OL]．[2016-09-10]．https://www.sohu.com/a/114071795_465985.
- [4] 陈兴良．论人身危险性及其刑法意义[J]．法学研究，1993，2.
- [5] 胡凌．大数据兴起对法律实践与理论研究的影响[J]．新疆师范大学学报（哲学社会科学版），2015，4.
- [6] 严戈．人民法院发展“大数据”战略的思考——以司法统计工作为视角[EB/OL]．[2014-08-29]．<http://www.chinacourt.org/article/detail/2014/08/id/1429348.shtml>.
- [7] Pasquale F. The Black Box Society: The Secret Algorithms That Control Money and Information [M]．Cambridge, MA: Harvard University Press, 2015.
- [8] 张化冰．加强大数据的司法应用[J]．人民法院报，2013（10）.
- [9] 刘毅．网络舆情研究概论[M]．天津：天津人民出版社，2007.
- [10] 王高飞，李明．我国网络舆情研究的回顾与展望[J]．现代情报，2016，5.