

智能辅助:AI下民商事办案系统的建构

——以裁判思维与要件标注为切入点

蔡一博*

内容摘要:目前AI与法律结合的研究处于蹒跚起步阶段,该领域文献综述与研究滞后于系统研发,为更好地对司法科技进行有效地能动回应,立足实证研究,总结现有系统研发的瓶颈问题,并立足于法院系统内的开发者、裁判者、使用者的三重身份提出以法官的裁判思维与要件标注的方法来审视现阶段的应用成果。首先,以场景需求为导向,在法学、计算机的理论支撑下构建AI下民商事办案系统,并补强实践短板,使办案系统后期具备深度学习能力。在路径依赖上,运用“开放结构”法学理论对类案场景模式下的裁判思维进行建模。其次,采用专家经验提取案件要素和证据规则,通过大量的人工化的要素标注建立起标签体系夯实数据基础,提高案件自动标注能力。最后,通过数据的机器学习、优化、迭代,最终实现机器对法律各领域的自主学习功能,实现民商事办案系统的智能化。

关键词:人工智能 民商事系统 裁判思维 要件标注

数据由法官创造,AI办案系统的开发应以法官的需求侧为设计进路,法官的亲历性决定其深度参与研发才能让AI更好地服务于审判。

——题记

信息化、智能化建设系智慧法院建设中牵一发而动全身的“牛鼻子工程”,是进一步推进和落实司法体制综合配套改革精神的“现代”助推器。过去数十年来,S市法院的信息化建设工作将案件基本信息、流程管理信息等进行有机统合,而且在AI^[1]与法律结合的辅助办案系统方面进行大胆探索,扎实推进,比如“以审判为中心的刑事诉讼制度改革软件”的开发取得初步成效,并得到中央政法委的肯定。同时,G省模式和Z省模式在智能辅助办案系统方面也卓有创新。但是,目前三省市的智能辅助办案系统基本限于刑事领域或极特定的民事领域,鲜有大范围、标准化、系统性地对民商事案件智能辅助办案系统进行深度研发。笔者立足于推进智慧法院这一重大课题的背景下,利用好现有大数据资源、已有的法学理论研究成果和日趋完善的AI模型算法,通过横向对案件标准化要件提取与标注、纵向对裁判思维的梳理与整合等进行多重维度的深挖以构建民商事案件智能办案系统。为该领域具备深度学习功能打下坚实的基础,以此促进审判体系和审判能力现代化、破解案多人少的矛盾、实现简单案件智能力。

*上海市高级人民法院研究室法官助理。

[1]AI:(Artificial Intelligence)英文缩写“AI”是研究使计算机来模拟人的某些思维过程和智能行为(如学习、推理、思考、规划等)的学科。笔者仅是学术讨论,不代表任何机构的观点和态度。

一、现象观察:系统的实证分析

国内关于AI与法律融合的各种研究孵化成果并不丰富,^[2]甚至有应用成果倒逼理论研究之势。相关内容的模型实用性不强、建模系数提取的合理性不够,真正将计算机技术、裁判思维与法学理论结合起来的模型不多。^[3]虽然S市和G省模式下的刑事领域已进行系统性的研发,但是民商事领域因覆盖面广、类型复合高、类案差异性大等特点鲜有深入研发,目前呈碎片化、分散式研发态势。比如S市高级人民法院开发的知识产权民事侵权案件赔偿数额专题分析系统、G省研发的道路交通事故损害赔偿系统等。我们通过实证分析、走访调研和理论分析三个角度对已有系统进行述评以明确研发难点和重点,为下一步大范围、深度化研究指明方向。

(一)宏观视角:系统述评

1.供给层面的述评。目前而言,法院系统内部的S市模式和G省模式在办案系统方面最具代表性,但是两个系统的特色各有不同,S市偏向于办案系统在贯彻证据规则,统一法律适用上起到智能辅助作用,而G省偏向于办案流程智能化管控。G省刑事模式的优势在于各单位专网整合早,数据利用的整合能力强,但是系统适用范围相对较窄,仅针对十余类刑事案件的裁判文书进行梳理分析,并且主要提取量刑要素分析处理;应用效果和部门协调应用有待提升;辅助法官制作裁判文书方面功能较弱。^[4]上述研发虽是刑事领域的内容,但是对于民商事案件智能辅助办案系统在办案流程规范、网络建设平台、建模算法等方面具有启发性。比如,通过刑事领域的建设,S市法院系统在数据采集、存储、分析、可视化等多个基础性技术领域已经取得较大的突破,形成了实用性强、稳定度高的技术能力,整体技术体系已初步构建完成。

虽然国内外部分软件已取得较好成效,但是存在研究领域过于集中(刑事案件领域、民事案件领域中的合同文本自动生成及部分内容可视化表达、智能类推等)、研发深度浅、实际应用少、覆盖面窄等不足。同时,国外的系统大多系服务市场的律师事务所研发,与法院系统的需求点有所不同,所以应扩大研发领域,并明确研究方向,只有将占案件总量85%的民商事案件进行深度研发,明确用途为辅助办案系统,才能切实做到为法官减负、达到服务诉讼当事人的美好愿景。但在研发的过程中应立足民商事案件的本质规律进行科技能动的系统研究,将产品目标琢磨透,从而有利于办案系统的深入研究和持续应用。

2.需求层面的问题。(1)开发应用与实际需求存在偏差的主要原因系需求分析的产品负责人对于法律和一线办案人员所需不甚了解。以笔者经验,如果该项工作的起步存在偏差,后期研发难度和应用适宜度将大打折扣,我们必须重视此项工作。所以笔者提出数据由法官创造,“智能辅助”的办案系统开发应以法官的需求侧为改革进路,让法官深度参与研发才能让AI更好地服务于审判。我们在提出产品需求的同时一定要抛弃传统的纯文字表达方式,争取利用文字说明与思维导图相结合的方式进行充分的可视化表达。(2)关于研发数据的数量和质量问题系事关AI能否有效开展研究的基础性工作,因为现有法律数据大多是非结构化数据,将此类数据有效的转化为准确和可理解的数据需要标注工作科学化。尤其是民商事案件诉请内容复杂、证据形式多样,该部分既是供给需求层面问题,也是运行中的最大障碍之一。未来的发展方向将主要集中在非结构化数据的标准、快速的提取方面,

[2]截至2017年5月30日,据不完全统计,通过中国知网、google学术等网络关键词或者字符段搜索相关领域内容,涉及大数据的研究论文352篇,大量文献集中于图书情报学、管理学、经济学界,但关于法律大数据及智能辅助等学术研讨类论文全文库共计9篇文章,并且该9篇文章的研究重点各不相同。

[3]张妮、杨遂全、蒲亦非:《国外与法律研究进展述评》,《法律方法》(第16卷),山东人民出版社2016年版,第458页。

[4]最高人民法院司法办课题组:《关于目前司法大数据的评价》,《上海市第一中级人民法院关于大数据、与智慧法院研讨会纪实整理》,2017年2月。

大量存在的非结构化数据和半结构化数据带来的技术和应用领域是大数据的蓝海,多样类型的数据分析、复杂的数据组合、多源的数据融合等问题将成为创新的重要聚焦点。上述两部分内容的解决方法后文会予以实操演示。

(二)微观聚焦:运行问题

现阶段各项办案系统的运行远没有达到人们预想的实用结果,从法律角度看其主要原因是系统的开发者对法律问题还不太熟悉,没有将法学理论的方法引入这些系统开发。^[5]同时对数据未进行有效法标化、推理建模算法不成熟等问题是给我们法律工作者和技术研发者提出的棘手课题。

1.系统数据表达的非标准化。研发系统的成功很大程度上需依赖于自然语言理解研究工作的突破。之所以目前困难重重,首要原因是法律实务中专业术语使用混乱。^[6]术语非标化问题的根源在于理论与实践在某些表达方面存在脱轨,这种脱轨不是不可以避免的,通过同类案件中对高频词汇进行统计后统一同一内涵的语词作为备注库可解决技术分辨的问题。其次除术语非标化外,法律文本或者裁判文书的系统数据标注化建设严重滞后,针对此种非结构化的内容我们要时常进行类案语言的标注工作。更重要的是希望有关部门能够贯彻、落实好“关于进一步完善与推广最高法院2015年法标规范工作”的指导意见。

2.系统逻辑建构的不确定性。现有各种系统难以突破的热点问题,即法律推理的认知学习难以确定的设定造成逻辑结构的不确定性。这个原因要素包括法理基础不够完善、^[7]实证研究成果未予应用“跑测数据”、价值判断和社会效应很难纳入推理等。比如,在案件智能类推方面,因数据采集的体量大、对象层次复杂,所以传统的数据采集只是将众多已公开的裁判文书按类别放到统一的“数据池”,然后根据搜索引擎的关键词逐步切换到下一个模块,最终查找所需内容。具体如体系检索法、关键词检索、案例倒查检索,这些背后的原理基本相同,基本做到了数据的有效整合和智能搜索,但缺陷是因未能建构类似于人脑的审判思维的数据提取和智能推理,所以导致智能类推的内容存在全而杂的问题。智能推理模块的编写需前期对非结构化内容进行人工的大样本分析和知识图谱设计,然后根据不同的功能进行以程序语言和人类表述语言相结合的信息逻辑设计。如果这个问题不进行艰难尝试,并且进行长期有效探索,真正的AI下的专家系统很难有效建构。

3.系统场景设计高度同质化。民事案件案由共424个,因覆盖面广、类型复合高、类案差异性大等特点,所以单一的场景设计路径难以满足多元化、复杂性的案件类型要求。目前,大部分智能办案系统的研发路径依靠模仿类似“审判要件九步法”“法律适用五步法”等抽象思维模式或者办案流程程序来设计场景要素的方向是对的,但是这样设计的弊端在于专家知识模型越通用,就越会包容噪音,带来高拟合性和重复性,更带来智能辅助的不精确性。^[8]反之,专家知识模型的设计路径越体现案件的个性化特征,就越无法夹杂无用的信息;适用的场景越有限,提供有条件检索或者智能辅助就更精确化。目前办案系统并不具备知识获取能力和对复杂场景的自适应能力,只有在通用场景下强化在特定场景下的设计和学习,才能避免系统场景设计的高度同质化,最大化地接近案件自身的真实,并且对于真理的可证伪性保持谦抑。

二、理论支撑:AI系统的建构路径

为解决办案系统数据采集非标准化、案件模型建构同质化的问题,应积极建构符合法理依据和

[5]黄良洪、曹旭东、刘树铭:《法律领域的专家系统》,《计算机工程与科学》2001年第1期。

[6]前引[3],张妮等文,第459页。

[7]P. Wahlgren 指出,“许多人工智能技术在法律领域的开发项目之所以失败,就是因为许多潜在的法理学原则没有在系统开发的开始阶段被遵守或给予有效的注意”。

[8]汪承昊:智能法律的两大风险:数据缺陷与知识的僭妄,载<https://zhuanlan.zhihu.com/p/32092347>,2017年12月17日。

满足技术要求的应用性强的民商事案件智能辅助办案系统。笔者在分析、研讨上述问题的基础上,认为在研发技术载体上应坚持专家系统和人工神经网络技术并存发展的路径,这对民商事案件智能辅助办案系统的初期建构非常必要。在初期的研发路径上分三步:第一步是仿照裁判方法成果进行树状结构图梳理,再通过排除选项和人工提取法律文本中的审判要件进行建构;第二步是立足审判要件、庭审笔录和裁判文书(非结构化文本)等,对案件进行要件提取与数据标注,进而归纳各种可能出现的情形进行模块化建构。第三步是将裁判思维提取与审判要件标注通过技术领域的算法进行研发整合,并在数据跑测下完善对大量案例和文档的深度学习和分析。

(一)民商事专家系统的破题基础

1.研发对象的选择原则。AI下的民商事智能辅助办案系统的研发按照对象类别可分为简易型和复杂型,初期研究对象的选取上应考虑简易类型为主,复杂案件为辅,但是最好两类案件可以在体系上一脉相承。研究对象的选取一定要科学论证,通过大量的试错与有益尝试,初期研究对象的选取最好满足以下原则。第一,选择的领域应是一个相对自我完备的法学分支领域,比如法律架构完整的合同法,但是合同法项下子项目过多,需要进行限缩。第二,该法律领域的确需要法律专业知识,比如合同损害赔偿中涉及与有过失、损益相抵等专业词汇。第三,最好选择一个窄而深的领域,而不是广而浅的领域,这样的角度方便日后从简易型扩充到复杂型。第四,最好选择法律专家争论较少的领域,避免逻辑不清或者适法不统一的情形。第五,选择的领域不需要大量知识性常识。^[9]比如,S市高级法院开发的知识产权民事侵权案件赔偿数额专题分析系统、该市某中级人民法院研发的金融诈骗类犯罪分析系统和该市F法院开发的道交案件大数据分析系统均在研发的选取对象上遵循上述原则,并在辅助审判环节中取得较好效果。

2.研发建模的法理基础。民商事案件智能辅助办案系统的模型建构需要提出关于法律推理性质的假设,该假设不仅需要基础性的法律知识,更需要法律推理的完整理论,即法学规范理论、法学推理理论、描述法律科学的理论。^[10]尽管法律推理十分复杂,但它具有相对稳定对象(案件)、相对明确的前提(法律规则、法律事实)及严格的程序规则,且须得出确定的判决结论,这为人工智能模拟提供了极为有利的条件。^[11]历经法学理论的变迁演化和人工智能技术的迭代进步,基础性的法学理论依旧是人工智能建立数学模型并编制计算机应用程序的基础。如法律形式主义、法律现实主义、开放结构等理论为办案系统建模提供了强大的法学理论基础。

法律形式主义^[12]遵循三段论的推理逻辑模式,最基本的两个要素为机械的演绎推理和封闭的规则体系,其认为将法律化成简单的几何公式是完全可能的。^[13]同时以图尔敏和佩雷尔曼为代表的逻辑学家们开始把注意力转向实践推理特别是法律推理领域,开辟了法律逻辑研究的新领域。^[14]该法律逻辑理论的进一步发展为人工智能的司法应用提供了可能的理论前提,20世纪70年代开发了律师推理系统,使机器法律推理第一次从理论变为现实。^[15]但法官所从事的法律活动不可能不受到其社会体验和思维结构的影响,尤其是民商事案件夹杂了大量社会利益的衡量问题。因此,法律形式主义忽视了推理主体的社会性,造成系统设计僵化性的致命弱点。

此时法律现实主义^[16]的理论运用促使人工智能研究从模拟法律推理的外在逻辑形式,进一步转

[9] R.E.Susskind, *Expert Systems in Law: A Jurisprudential Inquiry*, Oxford University D.Phil.thesis, 2016.

[10] Gold and Susskind, *Expert Systems in Law: A Jurisprudential and Formal Specification Approach*, pp.307—309.

[11] 张保生:《人工智能法律系统的法理学思考》,《法学评论》2001年第5期。

[12] 法律形式主义:即以法律规范为大前提,以具体的案件事实为小前提,进而推导出裁判结果,其核心主张就是坚信法律制度是一个封闭的逻辑自足的概念体系。

[13] [美]史蒂文·J.伯顿:《法律和法律推理导论》,张志铭、解兴权译,中国政法大学出版社1998年版,第3页。

[14] [荷兰]亨利·柏肯:《建模法律论证的逻辑工具——法律可废止推理研究》,熊明辉译,中国政法大学出版社2014年版,第12页。

[15] 前引[11],张保生文。

[16] 法律现实主义:20世纪上半叶在美国兴起了一场法律变革的运动,反对传统的法律形式主义等法学方法,倡导法律方法必须把社会利益的衡量引入到规范的法律论证中。