

向通过探求法官的内在思维结构,总结裁判经验中的规律性和普适性问题。在此背景下,大规模知识库系统(KBS)^[17]开发就注意了思维结构的整合作用,又通过联想程序被有机联系起来,构成具有法律推理整体功能的系统。^[18]

“开放结构”理论^[19]为人工智能在司法领域的深度学习应用提供了理论基础,打开了疑难案件自下而上的法律推理模拟的思路。^[20]在这一理论的指导下,人工智能在司法领域可以进行更具深度和广度的应用,一方面是将简易问题从疑难问题中筛选出来,运用基于规则的技术来解决;另一方面是将疑难问题先用非案例知识,如规则、控辩双方的陈述、社会常识来获得初步答案,再运用案例来比对,检查案件的正确性。上海市高级人民法院目前承担研发“推进以审判为中心的诉讼制度改革软件”的任务正是对这一理论最好的实践和诠释。^[21]

3.研发路径的技术支持。早期的法律智能辅助系统研发路径主要依靠专家系统理论^[22]来构建,如国外具有代表性的美国D.沃特曼和M.皮特森1981年研发的法律判决辅助系统、1993年武汉大学开发《实用刑法专家系统》等均是大量“如果-就”(If-Then)规则定义进行“自上而下”的设计。专家系统理论是在产生式系统的基础上发展起来的,通过知识库和推理机两部分来描述从一个基本概念演绎出的系统。^[23]虽然专家系统存在欠缺自我学习的能力等缺点,但现阶段关于利用人工智能进行法律推理、建立司法裁量模式,专家系统仍是人工智能与法律结合研究的重点之一。^[24]

目前,神经网络是AI研究的另一个重要方向,属于“自下而上”的机器学习思路。它的基本特点是试图模仿大脑的神经元之间传递、处理信息的模式来解构法律问题。其技术研发的路径和步骤:通过小样本数据进行对偶研究,然后经过监督学习、人工打标签、强化学习、交叉验证、原型开发、模型测试、业务试用、模型调整等阶段完成初期研发任务。

根据上述两个技术研发路径和理论支持,民商事案件智能辅助办案系统的架构建设在技术开发角度看具有较高的同质性,基本均采用依靠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签建立起办案系统的标签体系,再通过逻辑建模、物理建模等对数据进行机器学习训练,同时形成数据支撑下的神经网络再造和发育,通过数据的优化、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。综上所述,根据需求正确地引入不同的研发技术路径可以有效地将审判要件进行系统归类和裁判思维的构建进行有机契合,使其达到研发目标。

4.研发步骤的注意事项。细节决定成败,蹄疾步稳中坚持法律问题标准化、技术问题简单化的原则。法律问题标准化建设包括术语表达统一内涵、裁判依据的归纳概括具有明确性等,为解决好此方面的问题,需要经验丰富的法官在实践的过程中有计划地逐渐统一术语表述,并通过完善类案制度统一裁判依据。涉及技术问题简单化方面,主要针对法律工作者与技术人员的知识嫁接“鸿沟”问题,该内容的清障是各环节工作开展的基础,通过总结优化各机构的研发经验,涉及审判逻辑方面的内容可

[17]知识库系统。它通常需要人们的智力与经验来扩展知识面,并建立相关的函数关系。相关的术语“专家系统(ES)”。参见注释[22]。

[18]P.Wahlgren, *Automation of Legal Reasoning: A Study on Artificial Intelligence and Law*, Computer Law Series 11. Kluwer Law and Taxation Publishers. Deventer Boston 1992. Chapter 7.

[19]开放结构:使基于规则的法律推理模拟在受到概念封闭性的限制而对疑难案件无能为力时,找到了新的立足点。在此基础上,运用开放结构概念的疑难案件法律推理模型,通过逻辑程序工具和联想技术而建立起来。转引自前引[11],张保生文。

[20]折中的“开放结构”理论既承认逻辑的局限性又强调其重要性;既否认法官完全按照自己的直觉经验来随意判案的见解,又承认心证的存在。

[21]崔亚东:《人工智能让司法更加公正高效权威——人工智能在司法领域应用的理论分析与实践探索》,《上海智库报告——司法决策与参考》2017年第13期。

[22]专家系统:使用于相关的更高特定领域的知识库系统,用来提供建议并使用于特定的目的上。KBS系统具有内在联系的小规模子系统,在分别模仿法律推理要素功能(法律查询、法律解释、法律适用、法律评价、理由阐述)的基础上。张扬武:《基于产生式的民事法律专家系统的研究》,《电脑知识与技术》2013年第15期。

[23]前引[5],黄良洪等文。

[24]前引[3],张妮等文。

以通过思维导图方式进行图示化展示,可尽可能的将双方沟通的内容精准化。涉及技术语言的理解方面,首先,要确定文本语义相似度比较的对象,并实现语义相似度计算。讨论的对象包括篇幅划分(如词组、段落还是全文)和结构划分(如证据、法院认定和法律依据)两个方面。其次,采用何种自然语言描述推理规则的提取关系到规则库、领域库和法律库(例如储蓄存款合同纠纷案由相关的法律)的设计和构建技术。最后,输入系统的语义概念和逻辑体系必须自洽,不能含糊不清甚至互相矛盾。

(二)民商事案件智能辅助办案系统建构的方法初探

民商事案件办案系统的智能化是个宏观而复杂的问题。主要以裁判思维的成果整理和类案要件提取标准化建设作为逻辑起点的两个方面进行办案系统建构初探。

1.裁判思维的整理与提取——数据结构树的逻辑起点。按照诉讼请求、法律关系、法律规范、案件事实、法律结论的逻辑思路建构裁判思维。在这套系统工程中,我们要善于归纳、利用已有的裁判思维的研究成果,运用到数据的清洗、整理过程中。典型的民商事案件讲究以请求权为基础作为审判逻辑的起点,在经过充分学理研究且有强大实践大数据作为支撑的情况下,已有的较成熟的裁判思维方法的模型研究可否有效运用到审判数据建构基础值得尝试,为办案系统中的模型建构建设奠定扎实基础。

(1)诉讼请求表达固定化。诉请基础的固定事关整个诉讼,所以要在系统建构中先决的预设完整、精准的请求权词语表达。在这方面,S市某中级人民法院已卓有成效的构建了诉讼请求主题词库和裁判文书主题词库,在诉讼请求主题词库方面进行全面的、标准化的整理编撰,细分300项案由,1483项诉讼请求。^[25]这项诉请的整理基于最高人民法院关于案由的规定,但又创新的结合司法实践中常用术语表达进行升华。同时对提出各项诉请所依据的请求权基础及法条的原文进行整理和编纂。这一步的跨出实质地解决了诉讼中请求权基础词语混杂、机器学习难以识别的问题,增强了案件信息间的聚合度,有效回应了案件信息的非结构化难题,为办案系统的建构提供了标准化、体系化的审判数据模型。

(2)推理模型选择与场景深化设计:诉请固定后如何选用适合的裁判方法或者要件方法系办案系统设计中的核心问题。目前而言,实证表明所谓的逻辑三段论在实践中套用情形并不足以支撑细化的案件事实,所以三段论只能作为检验模型设计是否合理的工具或者语言编写的程序,但不适宜作为案件推理模型进行整体设计。如果能够将所有的法律论证置于演绎论证框架之内,就可以将所有的法律推理和论证交给计算机处理。但是司法实践中的法律概念、规则往往可以作不同解释,而且往往是诉诸例外的,如果遵循演绎主义,每增加一次例外就重写一次公式,无疑会增加一阶公式表达的复杂性,因此需要另寻途径和方法。法律实践表明,将法律推理处理为一种非单调推理^[26]是有效的解决办法。^[27]目前,从实践中看案例法律推理和规则法律推理现在已聚合了,这个框架提供了一个能够组合规则推理和案例推理的语境。人脑的逻辑基础是非形式逻辑,而电脑的逻辑基础是形式逻辑。如果说形式逻辑对论证的分析与评价仅仅是建立在语义和句法维度之上的话,那么非形式逻辑显然在形式逻辑框架基础上引入了一个语用维度。

在场景模型的选择上,经过反复研讨并与技术人员沟通,笔者认为有两种进路选择可供参考。一种是以引入德国成熟的裁判模型,即法律适用方法中常用的归入法或者相关性分析法。优势在于裁判思维模式成熟,简单案件适用归入法,复杂案件适用相关性分析法,同时技术层面的树状图制作也不复杂。劣势在于模型较简单和我国实践中较少应用此方法,需要一线法官逐渐在案件中丰富并进

[25]张能:《重构案件信息模型助推司法》,《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》,2016年12月。

[26]经典逻辑是单调的,推理的单调性是指,如果一个命题可以从前提集合推出,那么加入任意命题到这个前提集合中都不会使结论发生改变。非单调逻辑是前提集合与结论之间的推论关系不是单调的那一类非经典逻辑,换言之,这类逻辑的前提集合的某些扩充会导致结论的改变,比如可废止逻辑和缺省逻辑都是典型的非单调逻辑。法律推理是可废止的,因为这种推理所包含的推论关系是非单调的,证据前提集合的某些扩充会使原有结论发生改变。非单调推理是人工智能中的一种重要的推理方式。梁庆寅、魏斌:《法律论证适用的人工智能模型》,《中山大学学报》(社会科学版)2013年第5期。

[27]前引[26],梁庆寅等文。

行深化设计,但受限于目前我国法官队伍掌握此裁判方法的人并不多,制约了该方法的快速系统地应用。另一种是尝试直接试用我国理论学界或者法院实证已受关注的民商事裁判思维。其优势在于裁判规则逆向总结贴合司法实践,同时丰富的类案模型设计可以满足基本的案件需要并方便机器深度学习,尤其现阶段的研究进展使研发者越来越清晰地认识到不同类型的法律问题要求不同的推理形式,如证据推理不同于法律解释,而法律解释又不同于法律规则适用或立法辩论。^[28]在这样的法学理论和技术要求的背景下,通常场景模式下的特定化场景构建就是一个非常重要的内容。比如,四川省郝廷婷法官实证研究归纳七种典型的民商事裁判思维方法就非常符合特定的场景化构建要求,分别是回流型、往返型、平衡型、循环型、同步型、逆向型、顺向型。^[29]这七种类型是在请求权为基础的裁判思维方法下,通过司法实践的指引细化特定场景下不同案件的审理步骤等。这也只是实证分析中认可的一种思维方式,此类问题值得深入研究和扩展应用。

以上两种进路无论选择哪种作为模型基础都是初步的框架性逻辑设定,需要在数据跑测中逐渐丰富和完善,更离不开我们审判实务部门在日常的审判活动中积极整理各种疑难案例的裁判思路。

(3)思维导图的充分运用:思维导图是一种将放射性思考具体化的表达方法,优势在于清晰展现案件全部裁判思维流程,具有点线面结合的特点,易于将复杂的问题简单化,在数据建模的始端概念建模阶段通过思维导图易于明确产品需求以及业务规则,避免法律人口述或者文字表达后技术人员理解偏差的问题。然后落实到实体,^[30]这些信息将来会成为数据字典中非常重要的一部分,也就是所谓的元数据。比如,结合司法实践的研发实例,借记卡盗刷类案件通过尽可能的贴合程序员的设计路径,来运用法律思维导图进行各研发要素的表达,达到了较好的效果,这种单向性的逻辑表达设计具有简明性特点。在工具选择上以PROCESS ON或者XMIND等导图工具进行操作,建议裁判者可以学习这方面的知识,既方便审判思维的归纳整理又利于日常工作思维记录。在制作的过程中需要注意的是,尽可能使用单一化的线条表示,避免导图发散过多。如果说案件非常复杂,包含往返穿梭式的涵摄,具有极其复杂性和抽象性的话,建议使用正面或负面清单的表述进行来备注。

2.要件标准化和标注化建设——数据标准化的建构过程。类案要件提取标准化、标注科学化、问答交互化处理可以帮助我们构建各类型案件的审判要素模型,形成从生产到加工再到产出的这样一个司法大数据产业,为构建裁判数据模型暨办案系统的知识库建设奠定基础。

(1)要件提取“标准”化。国外在进行应用之前,通常将已有的类案研究成果进行标准化提取,然后再根据各类案常用的不同审判思维进行模拟编程,这种将理论研究成果与法律科技创新相结合的方式已得到绝大部分实务人士和技术人员的认可。但是问题就在于一般的标准化提取要件如何实现结构化和科学性。从宏观上看,法官的智慧是最核心的财富,审判业务部门应配合好要件提取工作,通过类案审判要件的大数据分析总结案件具体特征、事实情况、诉请与抗辩内容等填充到树状结构图,通过样本的机器学习最终有效得出所需结果。笔者正在尝试利用审判要件标准化的提取来构建AI知识图谱。如当事人的诉讼请求为“查阅、复制会计账簿”可以快速定位其请求权基础及其构成要件,并且通过对应被告的主要抗辩理由,可以快速对案件法律关系予以整体性的把握。

从微观实证上看,每类案件具有其特殊性,需要分类整理证据内容与证明目的、通常争议点及对应的法律依据。比如,德国Hollatz教授采用高斯发散神经网络的方法对200个非物质损害案例进行类案提炼,利用数学模糊性的评价系统将数据推理具体化为:伤害类型、持续时间长短、造成的后果的严重性、性别、职业的伤害、特别严重情节和医疗损害,最终获取非物质损害的裁量模式,^[31]该模型良

[28]前引[14],柏肯书,第10页。

[29]郝廷婷:《民事审判思维方法实证研究——“三段论”逻辑的中国式续造与验算》,载万鄂湘主编:《探索社会主义司法规律与完善民事法律制度研究》(下),人民法院出版社2011年版,第763页。

[30]当一个大型程序交由数个不同的程序人员开发时,用到相同的类名是很有可能,在程序开发中为了避免上述事件,提供了一个包的概念(package),使用方法很简单,只需要在写的程序第一行使用package关键字来声明一个包。

[31]前引[11],张保生等文。

好地运用于非物质损害的赔偿系统软件中。目前,S市民事、行政案件智能辅助办案系统通过人工提取大量卷宗和整理各类材料,已建立包括证据规则、证据指引库、办案要件(六类八个案由)、立案审查关键要素库、诉请库、抗辩库,案例库等12项大数据库,为办案提供信息支撑和保障。

(2)类案“标注”科学化。标注系技术用语,应用于模型和复杂神经网络建构,如标注的数据越准确、数量越多,模型的效果越好。所以类案标准化确定后标注内容表达要素提取要全面化、科学化,S市某中级人民法院在两库建设中的做法值得我们关注。运用技术手段对裁判文书主题库方面中各类案由所含的高频词汇,关涉事实认定和法律适用难点的关键词,以及文书中所含的包括但不限于行业经济发展、社会综合治理人员结构等重要要素高亮提示,以“贴标签”的形式予以设置标注,共梳理2654个主题词。^[32]比如,在民事领域该法院选取近三年案件量较大的机动车交通事故责任纠纷中3000份文书进行高频词汇分析并标注,得出案件特点、诉请内容、证据类型、胜负趋势等。在刑事领域,案情预测系统“法狗狗”设计思路正是:对于人类说不清楚的问题,通过裁判文书整理后输入大量的标注数据,通过机器学习等算法得到测量值最终得到同案类判的匹配结果。这种主动标注的方法,整体上看类似于百度后台有强大的数据标注员一样,为后续机器智能深度学习打下基础。目前,S市民事、行政案件智能辅助办案系统已建成8组标签体系支撑办案系统的前期运行,为后续的迭代开发奠定坚实的基础。

(3)类案“问答式”交互化。深度学习技术将改革人们与技术的交互方式,同样智能辅助办案系统如何在设计应用中体现这种交互性非常重要,也事关能否实现法律AI化的关键。笔者也正在尝试利用在类案“庭审笔录”整理与“思维导图”的指引下,通过EXCEL表格细化法官审理思维,实现当事人求问思维和技术人员程序思维的问答交互,进而在程序编写过程中可初步跑测数据的合理性。现在外界的研发机构的基础数据主要为对外公布的司法文书,不包括庭审笔录、专家研讨会议笔录或者其他不可对外的秘密内容。但是笔者认为这些内容对于办案系统建设至关重要。比如,庭审笔录系法官根据案件事实情况逐一将要点查明的思维方法的书面化表述,也是对于同一类型问题的庭审确认和争议问题的庭审辩白。

三、实务指南:AI示范模本的应用

在上述问题意识的指引下,下面以银行借记卡被盗刷的储蓄存款合同纠纷为例。进行标准化的审判要件提取和裁判结构树的建构,做到智能深度学习与服务诉讼当事人的作用。以下内容系S市M法院与H省某大学机构按照数据的来源侧、数据的处理侧、数据的产业侧暨挖掘、分析、应用等三个方面进行共同研发的微观展示。

(一)研究对象的选取与破题技术的基础

1.研究对象的符合性。借记卡盗刷类案件在研究对象上具有有限缩性特征,避免出现研究面过大的易遗漏问题。研究的立足点基于违约或者侵权损害赔偿的要件体系具有完整性,符合智能推理的逻辑自洽性。更重要的是该类问题的诉请、证据形式、争议焦点都比较明确和统一,有利于率先尝试简单案件智能化。同时该类案件符合热点问题的回应性和智能研究领域的空白性特点。

2.技术性目标的设定。第一步,以“伪卡盗刷”类型为案由的判决书收集和分析(法律研习者为主);第二步,研究基于结构和语义的判决书文本提取和语义匹配方法(高频词汇整理外以技术人员为主);第三步,研究基于规则的判决结果诊断方法,设计和实现基于规则的判决结果推理服务;第四步,研究基于模板的判决书生成方法,设计和实现基于模板的判决书生成向导服务。上述第三、第四步主要以技术人员为主导。

[32]前引[3],张妮等文。

(二)裁判思维的整理与提取

1.诉讼请求表达的确定系源泉的固定。设定程序后电脑自动下载1600份伪卡盗刷案件的判决后对诉请部分进行结构化文本的提取、理解和分析。整理诉请表达如下,案由为储蓄存款合同纠纷,请求权表达通常为原告向某银行主张的赔偿盗刷具体金额。附加内容为原告主张的赔偿利息等损失的金额、计算方式及起止日期。极个别情况下有些原告的诉讼请求中包含解除合同、要求赔礼道歉等内容。

2.模型选择与思维导图深化表达。根据上述第二部分的七大常用裁判思维模型分析来看,储蓄存款合同纠纷常用的裁判思维属于逆向性逻辑,逆向性的民商事审判思维通常适用于信用卡纠纷、储蓄存款合同纠纷、金融借款纠纷案件。在模型设计上即对已有书面材料或者问题的审查完成后,很大程度上对结论的形成起了决定性作用。庭审查明或者问题互答只是为了排除例外事实以及法律规范的寻觅,主要作用则在对于预设结论的进一步验证。^[33]按照以上的“逆向式民事思维”检索顺序制作逻辑树和思维导图。步骤演示,第一步:诉请与证据类型;第二步:得出预设结论;第三步:排除例外事实;第四步:寻找法律规范;第五步:作出裁判。最终通过制作思维导图,暨与技术人员沟通较为流畅的导图化表达。

(三)要件标准化提取和标注科学化建设

按照要件标准化提取伪卡盗刷类案件,首先审查原告请求权基础并将其固定化,如审查原告向发卡行或特约商户的主张、原告诉请是否明确等;其次审查被告的抗辩权基础,如被告对民刑竞合情形下该案件受理是否提出异议、对诉讼时效是否异议、伪卡交易过错责任抗辩等;同时进一步归纳争议焦点,如一般而言,伪卡交易中发卡行有无尽到注意义务,在伪卡交易中有无过错对损失的产生有因果关系等;再次根据规则分配举证责任,最后明确归责原则,具体不再一一赘述。

同时要件标准化下关键词提取系重要任务,即高频词汇整理。第一步,选取200份借记卡盗刷案件的裁判文书,即“伪卡盗刷”类型案件的判决书大范围采样,采样过程中需注意判决书采样的审理地区、审理审级等问题,建议采用法律发达地区或者该类型分布较密集地区为主。第二步,对高频有用词汇进行标注细化,涉及的技术问题如分词、词典建立、词性标注和相似度计算(文本匹配)内容等由技术人员根据大数据自动高频检索进行分析。在此过程中除却传统类案研究的书面报告整理外,笔者整理出一套行之有效的新技术。在第一步的基础上,先挑选200份相关案由的判决,全部下载下来。然后阅读3-5份判决,寻找20个左右大概出现的特殊词,接下来将所有文档进行合并,创建一个文档,word中有插入文档选项进行批量(多选)插入文档。最后在合并后的文档对20个词进行分别查找,统计每个词出现的频率(一般100次),并按照频率进行排序。

(四)人机对话的互动基础设计

立足当事人咨询、法官询问和银行抗辩三者的角度构建智能辅助化的咨询系统或者类案结果智推系统。选取大样本的庭审笔录是先行之举,通过对庭审笔录的充分整理和高度概括明晰审理此类案件的事由和问题。比如,笔者通过庭审笔录为基础,结合判决书等材料标准提取和裁判树状设计问题取得良好效果。只有以持续不断的数据记录,数据分析模型形成支持事实还原的证据;以用户为中心的司法流程设计依靠高度结构化的改造,才能大大降低当事人进入司法通道的成本,方便法律服务的获取和起到智能交互作用

[33]前引[29],郝延婷文。