



关于法律人工智能数据和算法问题的若干思考

丛颖男, 王兆毓, 朱金清

引用本文

丛颖男, 王兆毓, 朱金清. [关于法律人工智能数据和算法问题的若干思考](#)[J]. 计算机科学, 2022, 49(4): 74-79.

CONG Ying-nan, WANG Zhao-yu, ZHU Jin-qing. [Insights into Dataset and Algorithm Related Problems in Artificial Intelligence for Law](#)[J]. Computer Science, 2022, 49(4): 74-79.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于物理信息的神经网络:最新进展与展望](#)

Physics-informed Neural Networks:Recent Advances and Prospects

计算机科学, 2022, 49(4): 254-262. <https://doi.org/10.11896/jsjcx.210500158>

[非均衡数据分类经典方法综述与面向医疗领域的实验分析](#)

Imbalanced Data Classification:A Survey and Experiments in Medical Domain

计算机科学, 2022, 49(1): 80-88. <https://doi.org/10.11896/jsjcx.210200124>

[人工智能治理理论及系统的现状与趋势](#)

AI Governance and System:Current Situation and Trend

计算机科学, 2021, 48(9): 1-8. <https://doi.org/10.11896/jsjcx.210600034>

[面向大数据分析的智能交互向导系统](#)

Smart Interactive Guide System for Big Data Analytics

计算机科学, 2021, 48(9): 110-117. <https://doi.org/10.11896/jsjcx.200900083>

[人工智能安全框架](#)

Artificial Intelligence Security Framework

计算机科学, 2021, 48(7): 1-8. <https://doi.org/10.11896/jsjcx.210300306>

关于法律人工智能数据和算法问题的若干思考

丛颖男¹ 王兆毓² 朱金清³

1 中国政法大学商学院 北京 100088

2 中国政法大学法治信息管理学院 北京 102249

3 北京字节跳动网络技术有限公司 北京 100043

(cyn_2010@163.com)

摘要 人工智能技术的不断发展使其在司法方面的应用逐渐增多,并引起广泛关注。具体来说,人工智能已经在合同审查、智慧法院等应用场景中崭露头角,相比传统人工,人工智能的高效率表现展示了其在司法领域的巨大应用潜力。但在其他应用场景,如智能司法裁判,虽然国内外有一定尝试,并取得了一些成果,但仍面临着数据样本量不足、算法与待解决实际问题匹配度不够的问题,以及算法过程不够透明等方面的质疑。文中围绕现有法律人工智能的相关工作,探索了人工智能可能带来的司法流程上的巨大变革,并对人工智能目前在智能裁判中遇到的数据和算法方面的问题是否会对司法的公正性产生影响进行了探讨,最后对上述问题的解决方案以及司法人工智能的未来发展路线略抒拙见,以期人工智能技术在我国司法领域有更为系统性的应用,助力社会主义法治建设。

关键词: 人工智能;法律;数据分析;AI 算法

中图法分类号 TP182

Insights into Dataset and Algorithm Related Problems in Artificial Intelligence for Law

CONG Ying-nan¹, WANG Zhao-yu² and ZHU Jin-qing³

1 Business School, China University of Political Science and Law, Beijing 100088, China

2 School of Information Management for Law, China University of Political Science and Law, Beijing 102249, China

3 Beijing Bytedance Network Technology Co., Ltd, Beijing 100043, China

Abstract With the rapid development of artificial intelligence (AI) technology, the application of AI-related technologies in law is increased and attracts extensive attention. Specifically, AI has emerged in multiple legal scenarios such as automatic contract review and smart courts, compared with traditional artificial intelligence, its high efficiency shows its great application potential in the judicial field. However, in other scenarios such as legal judgement prediction (LJP), AI faces challenges and doubts in data analysis and algorithms, although some attempts have been made. Through analysis of the work related to legal AI, this paper summarizes the potential problems in datasets and algorithms in intelligent referees, investigates the changes in judicial progress that AI may bring and discusses whether the problems encountered by AI will affect the justice of law. Finally, this paper briefly expresses the potential solutions to the above problems, and provides insights into its future development, in the hope that AI technology will have a more systematic application in China's judicial field and contribute to the construction of socialist rule of law.

Keywords AI, Law, Data analysis, AI algorithm

1 引言

随着信息技术的进步,人工智能在法律行业的应用快速发展。早在 20 世纪 70 年代,国外便着手研究人工智能与法律的结合,至今已经取得了不少成果。具体而言,在民商事

领域,摩根大通利用人工智能技术,开发了商业贷款合同审查系统,该系统可以在几秒内完成人工 36 万小时才能完成的信贷审查工作^[1],大大提高了律师对合同的审查效率。在刑事司法领域,美国部分州的法庭采用人工智能来计算被告人对社会的“风险指数”,并以此为依据决定是否对被告人进行

到稿日期:2021-09-23 返修日期:2021-12-22

基金项目:北京市教改项目“法商大数据分析创新型人才培养模式研究”(京教函[2020]427 号);中国政法大学新兴学科培育建设计划

This work was supported by the Beijing Education Reform Project “Research on the Training Mode of Innovative Talents for Law and Business Big Data Analysis” (Jingjiaohan [2020] No. 427) and Cultivation and Construction Plan of Emerging Disciplines of China University of Political Science and Law.

通信作者:朱金清(zhujinqing@bytedance.com)

保释或假释^[2]。此外,据 BBC 报道,人工智能律师 Case Cruncher Alpha(CCA)与伦敦的 100 名人类律师就“基于数百个付款保护保险(Payment Protection Insurance,PPI)错误销售案例事实来判断索赔与否”的案由进行竞赛,CCA 最终以 86.6% 的准确率战胜了上百名人类律师,后者的平均成绩为 66.3%^[3]。而在司法裁判领域,虽然存在许多争议,但也有部分 AI 裁判的尝试。美国芝加哥伊利诺理工大学等机构利用美国最高法院数据库(1791—2015),通过算法重现了从 1816 年起法官的 28000 项决定和 240000 次投票,正确率高达 70% 以上^[4]。在知识产权领域,Lex Machina 公司利用人工智能技术进行裁判预测,获得了和人类律师类似的结果^[5]。2016 年 11 月 24 日,据英国《卫报》报道,伦敦大学学院、谢菲尔德大学和宾夕法尼亚大学的科学家表示,人工智能已经可以分析法律证据与道德问题,进而预测审判结果。在研究中,人工智能程序能够对相关信息予以分析,并在此基础上做出判决,且 79% 的判决结果与当时法庭的判决一致。

与国外人工智能在法律行业的发展相比,我国法律人工智能呈现出起步晚、发展速度快、应用潜力大的特点。随着国家政策的出台和具体技术的实现,我国人工智能与法律不断碰撞、融合,催生了司法领域的大变革。在顶层设计方面,国务院于 2017 年 7 月发布《新一代人工智能发展规划》,提出“三步走”战略目标,旨在抢抓人工智能发展的重大战略机遇,构筑我国人工智能发展的先发优势;最高人民检察院于 2018 年 1 月发布《关于深化智慧检务建设的意见》,提出了建设全业务智慧办案、全要素智慧管理、全方位智慧服务、全领域智慧支撑的智慧检务总体构架;最高人民法院于 2020 年 11 月发布《人民法院信息化建设五年发展规划(2021—2025)》,以期加快智慧法院的建设,推进我国审判体系和审判能力的现代化发展。公安部也在不断修改和完善《公安部大数据智能化建设规划》,希冀不断提升公安工作的整体效能与核心战斗力。政策积极信号的释放为法律人工智能在实践领域的深入探索注入了新动力。

在实践中,目前我国已有多项与人工智能相关的应用落地于司法实践场景。截至 2017 年 6 月,最高人民法院、上海高院、广州中院和苏州中院等 100 逾家法院部署了科大讯飞公司所研发的智慧法院庭审系统,通过提高庭审记录的准确率来促进庭审高效化。又如,最高人民法院委托苏州中院研发的庭审语音识别系统,不仅可以在庭审过程中实现语音到文字的转换,并能自动区分庭审发言对象及发言内容,法官、当事人和其他参与人均能实时看见转录文字^[6]。在系统试用中,语音识别的正确率已达到 90% 以上,书记员只需进行少量修改即可实现庭审的完整记录。经对比测试,庭审时间平均缩短 20%—30%,复杂庭审时间缩短超 50%,庭审笔录的完整度达到 100%。河北高院历时近一年组织研发了“智慧审判支持”系统,通过对电子文件的要素化管理和自动分析,辅助法官进行法律文书的撰写。在河北的 194 个法院,有 6961 名法官应用了该系统。截至 2016 年底,共处理案件逾 11.1 万件,辅助生成文书逾 78.4 万份^[7]。

随着人工智能在法律行业的应用的不断深入,实践经验不断累积,更多的问题也暴露出来。想要让人工智能更好地

为法律服务,必须深入分析其应用的现实困境,进而对未来人工智能在司法中的应用提出切实可行的举措。如在自动量刑等应用场景下,人工智能面临大量问题亟待解决;尤其在数据处理和算法设计方面,如何将人工智能嵌入量刑环节还存在较大的理论争议与实践障碍,需要更多理论和技术上的突破作为支撑。本文从数据量、数据内容和数据结构化等数据方面的困境和算法可解释性、算法透明度等算法的需求来阐述目前法律人工智能开发中面临的困境,对部分学者的疑问作了浅显的探讨,并给出可能的解决路径。

2 数据:从裁判文书到知识图谱

在模型研发和测试阶段,许多科研人员会采用中国法律技术智能评测(China AI and Law Challenge,CAIL)等数据集来进行实验^[8],该测评提供相应的公开数据集。此类数据集通常针对的是法律人工智能中的特定应用场景,如量刑预测、裁判文书摘要和司法问答等。数据集专业性较强,针对的领域切口较小。此类数据集结构化程度较高,数据量较小,比较适合进行模型的训练,但不适宜进行模型的验证。相比通用的自然语言处理数据集,CAIL 等数据集的数据内容不够丰富、数据量不够庞大,即使采用迁移学习^[9]等技术手段来减少数据数量对实验结果产生的副作用,仍然只是让模型无限逼近其效能的“天花板”,但这一“天花板”可能仍远远低于实际应用的要求。

为了获得数据量更大、数据内容更丰富的司法相关数据,降低此类数据集的不足所带来的不利影响,在实验前期的数据准备过程中,研究人员常常对裁判文书的文本进行预处理,以获得训练所需要的个性化的数据集。中国政法大学的 Wang 等基于中国裁判文书网中公开的文书数据,开发了由 10000 篇裁判文书组成的民事裁判数据集。该团队利用 Transformer 模型处理该数据集并对其进行优化,得到了高效的民事案件判决结果分析模型^[10]。

据统计,我国裁判文书网中公开的案件数量约占全部审结案件的 50%^[11]。也就是说,仍然有大约一半的裁判文书并未上网公开。据此,部分学者认为裁判文书的数据量是不充分的^[12]。单纯从数量上来说,目前中国裁判文书网已经拥有超过一亿的文书保有量^[13]。从人工智能开发的角度来看,虽然可获得的数据仅占全部数据的 50%,但相比其他行业的人工智能训练数据集,该数据集的体量较大。这样的数据量是否仍然不足以支撑人工智能算法训练?笔者认为可以尝试从以下角度进行研判。

在传统机器学习模型的构造过程中,为了保证训练得到的结果具有较高的准确率和召回率,通常依据以下两个假设进行操作:1)训练集与测试集的样本点为独立同分布;2)必须有足够多的训练样本才能保证训练模型的效果。

从上述两个基本假设出发,根据相关实践经验,只要训练样本的统计学分布“合理”,50% 的案件数据就可以训练出效果相当优越的模型。但数据的分布是否“合理”实际上是一个难以界定的问题。《最高人民法院关于人民法院在互联网公布裁判文书的规定》(以下简称《规定》)中规定了裁判文书不公开的 5 种情形。此类特定数据的缺省是否会导致数据分布

不符合预期假设？各地法院对《规定》的执行情况究竟如何？是否会刻意或无意不上传某一类型本应上传的文书？这些问题或许会导致裁判文书数据的分布总体上产生变化，因此在部分裁判文书根据法定情形不予以公开的前提下，剩余的数据能否支撑相关模型的训练，笔者认为，针对应用场景的不同，这一问题仍需要更多的实验结果加以佐证，而不是简单地对其持否定的态度。例如，对于某些通用的量刑预测模型来说，基于此类数据进行训练或许对模型产生的效果微乎其微，但在某些特定领域下，如针对未成年人的相关量刑预测模型，由于《规定》中明确了涉及未成年人的相关文书依法不上网，此类数据显然是难以支撑相关模型的训练的。

总的来说，对于法律数据，CAIL、裁判文书网等数据集只适用于部分应用场景下的模型开发。在通用场景下，显然需要基于更全面的数据集进行训练，但部分学者的担忧或许仍为时尚早，许多疑惑仍需要通过后续更多的实验与实践来解开。

2.1 数据内容

除了裁判文书的数量，部分法律学者对裁判文书反映法律信息的能力也存在疑惑。“除了对外公布的法律裁判文书，法律决策过程中的关键行为(如形成决策的内部讨论)往往是高度非文字化、非数据化的。无论是中国和大陆法系国家的法律决策组织如合议庭、审委会的讨论、决策过程，还是英美法系国家的法官、陪审团审判的内部讨论、裁判过程，基本上都是非公开或无记录甚至不允许记录与公开。这使得我们对法律决策过程中的博弈过程、裁判目的、考虑条件和心证形成等这些影响决策最为关键的因素缺乏普遍、充分的记载与掌握。”^[12]事实上，监督学习的本质过程就是在输入和输出中建立一种函数(映射)关系，即通过机器学习来拟合 $\hat{f}$ 并实际逼近真实的映射关系 f 。这样的映射关系在计算机中可以表达为算法训练得到的映射关系，而从司法角度来说，往往就是法官等法律工作者逻辑思路的抽象形式。因而笔者认为，裁判文书中缺少裁判过程中的相关决策信息，对法律人工智能算法的总体效率在大多数场景下并不会产生显著的影响。因为法律人工智能算法的训练本质上是拟合法律事实到法律结果的映射，也就是拟合法律推理的过程，所以对于这一过程相关数据的确定，似乎并不是法律人工智能应当关注的重点。

2.2 数据结构化

除数据量和数据内容外，目前法律人工智能开发的最大挑战源于法律文书的结构化程度。在自然语言处理的开发中，模型训练的第一步通常就是对文本进行向量化或结构化。而在文本向量化的过程中，如果能将同义词进行相同的标记，显然可以大大降低文本向量的维度，进而提高数据的质量^[14]。尽管法律裁判文书对法律要素、文书格式等进行了基本的规定，但并未对文本内容中的同义词和近义词予以统一，事实上也几乎不可能达到完全的统一化。以丧葬费为例，在不同的法律文书中，“丧葬”可能被表述为“安葬”“殡葬”；“保存尸体”的同义词有“冷冻尸体”“冷藏尸体”等。这种个性化表述的成因多为各地惯用语的差别，以及法官等法律文书写者的个人习惯。这样个性化的表达降低了裁判文书的结构化程度，对人工智能的应用造成了极大的障碍。目前在业内，

解决这一问题的思路通常是采用知识图谱技术来获得同义词之间的逻辑关联。在中国法律人工智能应用领域，特别是预测裁判方面，知识图谱搭配深度学习已成为话语层面的主流解决方案。

所谓知识图谱，是显示知识发展进程与结构关系的一系列不同的图形，并用可视化技术来描述知识资源及其载体，挖掘、分析、构建、绘制和显示知识以及它们之间的相互联系^[15]。具体来说，知识图谱是把复杂的领域知识通过数据挖掘、信息提取等技术处理后，绘制出的富含概念、实体和关系的可视化的大规模语义网络^[16]。知识图谱对数据和模型的颗粒化程度要求较高，模型越精细，数据越详细，知识图谱的效果就越好。以危险驾驶类案件为例，尽管此类案件案由简单、数量庞大，但在知识图谱上的颗粒化程度仍然极高，需要考虑的因素也相当多，如酒后逃逸致人死亡等。如果忽略了这一点，以颗粒化程度低的知识图谱进行案件结果预测，难以获得期望的效果。

知识图谱技术可以有效地对法律数据进行结构化，这也是目前许多科研人员研究司法数据结构化的主流方向。如中国科学院大学的 Huang 通过知识图谱和深度学习、迁移学习等技术的结合开发了法律内容问答模型，并有效地解决了先前相关模型忽视行业知识的缺陷^[17]。但知识图谱的开发涉及法律知识与计算机知识的交叉，并且开发前期需要依赖大量的人工标注。目前相关机构在开发法律知识图谱时，往往需要招募大量具有法律背景的人员进行手工标注，这也是目前法律知识图谱技术难以实现跨越式发展的症结。倘若未来能进行更科学有效的研究，采取更加高效的标注方法，有理由相信知识图谱技术能为法律数据的结构化赋能，从而推进法律人工智能得到跨越式的发展。

3 算法：坚持可解释性和透明度

在开发算法的过程中，研究人员通常优先考虑提高算法的准确性与合理性，或者说开发思路是“效果主导”的。这一做法的正当性在于：如果在司法实践中贸然使用与实际问题匹配度相差较大(如数据集不匹配)的人工智能算法，最终训练得到的可能是一个“欠拟合”的不准确的模型。这样的人工智能不仅不能提高司法实践的效率，反而有可能干扰正常的司法审判流程，影响司法公正。另一方面，准确性高并不意味着效果一定好，而仅仅表明训练模型的参数较为适配测试集数据的特点。由于现有的人工智能效果评价标准多依赖事先设计好的测试集，因此在更为通用的场景下，模型的实际表现可能与测试结果相差甚远，甚至产生错误，即可能会产生模型的“过拟合”。在法律人工智能的应用场景中，模型的“过拟合”意味着模型拟合出的映射规律远比法律专家所熟悉的规律更复杂，这导致输入与输出之间的相关关系难以直观显现^[18]。如何在模型的准确性与合理性之间获得平衡，以得到理想的结果，是司法领域人工智能的开发者在大规模应用算法之前应关注的问题。

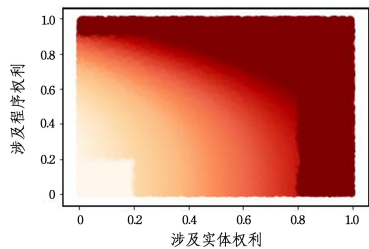
除准确性和合理性外，法律问题的特殊性还对法律人工智能的开发提出了更多要求。有学者提出，当前法律人工智能开发者与服务使用者之间不平衡的风险负担和责任分配，

对法律人工智能的算法解释权治理提出了全新的要求,算法解释权可消弭法律实然权利与应然权利的鸿沟^[19]。笔者认为,从技术实现的角度出发,算法的可解释性和透明度是算法解释权的底层基础。虽然这两个方面在实际开发过程中常常被开发人员忽视,但是在法律人工智能应用中,提高算法可解释性和透明度是走出人工智能伦理冲突与责任风险困境的重要途径。

3.1 可解释性

2017 年,ICML (International Conference on Machine Learning) 的 Tutorial 中对可解释性 (Interpretation) 给出了如下定义:“Interpretation is the process of giving explanations to human.”简单来说,可解释性就是算法可以被人类解释的能力。Sutskever 等^[20]发现,层次更深的神经网络可以更好地逼近任意函数,进而获得更强的拟合能力。但在深度学习系统中,分类器和特征模块都是自动学习的,因此神经网络就有了一个灰色区域:可解释性的缺失^[21]。深度神经网络学习输入输出映射关系是相当不连续的,微小的变化都可能会引起映射关系总体上发生相当大的变化^[22]。但在当今以人工智能开发效率为导向的背景下,开发者为了获得更好的算法性能,常常执着于更深的神经网络和更多的参数设置,这导致模型的可解释性被进一步弱化。在法律人工智能中,如果算法不具备可解释性,也就很难对其正当性进行论证,难以达到许多场景下法律实践的需求;而当算法出现错误甚至歧视时,也难以确定相应的责任分配。目前越来越多的学者意识到了法律人工智能中可解释性的重要性。例如,Tao 等从算法可解释性出发,详细阐述了司法场景中的算法治理问题,从程序保障和实体规范两个维度提出了可解释性结构的具体要求^[23]。

需要注意的是,法律人工智能对可解释性的客观需求并不代表总是坚持可解释性优先,而忽略算法本身的效率。这是因为在不同的应用场景之下,对于可解释性要求的强弱程度是不一致的。笔者认为,越是涉及当事人实体(或程序)权利核心的应用场景,对可解释性的就要求越高。而随着应用领域与实体权利核心的偏移,可解释性逐渐成为算法的次要目标,算法回归其效率导向的本源,甚至在许多应用场景下实际上是没有可解释性需求的。而当不同的应用场景在实体权利和程序性权利的范畴下对当事人权力的影响基本一致时,相比应用于程序性权利的法律人工智能算法,应用于实体权利的法律人工智能算法对可解释性的要求应当更高。因此在不同应用场景下法律人工智能对可解释性的要求强弱可以归结如图 1 所示。



注:颜色由浅入深代表对其要求不断增强

图 1 不同应用场景对可解释性的要求

Fig. 1 Demand of interpretation in different application scenario

因此,在进行法律人工智能相关算法的开发时,首先可以根据笔者提出的上述方法,对其可解释性进行判断,然后将该应用对可解释性的需求分为如下 3 个类别:

(1)强可解释性需求。该类应用是由可解释性驱动的,通常涉及当事人实体权利或程序性权利的核心内容。甚至在某些场景下,需要牺牲算法的效率来实现可解释性的目标。主要的应用如罪名、量刑预测等。

(2)中可解释性需求。该类应用涉及当事人一定的实体权利或程序性权利,但应用并不涉及该类权力实现的核心内容。此类应用开发时对可解释性有一定需求,但同时兼顾算法效率,二者之间的如何平衡在不同应用场景下常常需要不同的方式。主要的应用如司法智能问答、裁判文书的自动生成等。

(3)弱可解释性需求(无可解释性需求)。该类应用通常对当事人权利影响不大,或者在实践中需要大量的人工干预来保障其效率,算法本身只起到提高司法效率的目标。此类应用开发时通常是由效率驱动的,开发人员无须考虑其可解释性需求。应用场景如法庭中当事人的语音识别等。

在明确了法律人工智能应用中坚持可解释性的必要性和可解释性的客观要求之后,最重要的问题在于,如何实现法律人工智能中的可解释性。笔者认为,有如下两种可行的路径,可以在法律人工智能开发的过程中,在符合可解释要求的前提下,尽可能提高算法的效率。

3.1.1 算法的改制

众所周知,模型的弹性越高,其拟合能力则越强,虽然可能会出现过拟合等情况,但在应用中也可以通过许多方式来避免。而神经网络正是由于其高复杂度、高弹性和强拟合能力的特点,目前在人工智能的开发中应用广泛。通常来说,高弹性则意味着低可解释性,这是一对不可调和的矛盾。因而在一些场景下,如果可解释性需求必须被满足,并且算法对应的应用场景并不需要算法本身过强的拟合能力,那么许多相对而言弹性较低、拟合能力较弱的算法也有其“用武之地”。东南大学的 Zhang 等对传统的贝叶斯网络进行优化,对裁判文书进行分类,最后取得了比较优秀的实验结果。相比深度神经网络,朴素贝叶斯、决策树等算法显然具有更好的可解释性^[24]。此类模型因其拟合能力有限,在当今人工智能效率导向的发展浪潮下似乎已经逐渐淡出人们的视野。但在法律人工智能领域,随着对算法可解释性要求观念的深入人心,相信此类算法的改进算法、衍生算法和融合算法都将在一定场景下发挥作用。

3.1.2 模型的改制

在许多场景下,除了对可解释性的需求,应用场景的复杂度对模型的拟合能力也提出了非常高的要求,上述算法已经不能满足实践的需要。Yang 等^[8]的工作为这一问题提供了一种全新的思路。他们认为,在量刑预测 (Legal Judgement Prediction, LJP) 的任务中,可以将总体的任务分解为法条预测、罪名预测和刑期预测 3 个子任务。虽然在对每一个子任务进行预测时采用了 MPBFN (Multi-Perspective Bi-Feedback Network) 网络进行实现,但该算法本身的可解释性较弱。通过子任务的分割,最后藉由子任务之间的拓扑关系 (Topolo-

gical Dependencies)来实现总任务的目标,这使得模型总体上具有了一定的可解释性。这一思路的创新性在于,通过模型的重构使原本不具有可解释性的算法在模型总体层面获得了一定的可解释性。这样一种子任务分解,再藉由拓扑关系融合的思路,也是赋予法律人工智能可解释性,并保证模型拟合能力行之有效的做法。而如何对这一子任务进行划分,Zhong 等^[25]在进行量刑预测时提出了相似的思路,并将其模型命名为 TopJudge,该模型为了实现其“Judge”(裁判)的目的,实际上是以“Judge”(法官)的方式进行思考,并划分相关的子任务。因而在对子任务进行划分并确定其拓扑关系时,相关研究人员可以从法律人的角度出发,从法律人推理的过程来思考,将总体任务分解为法律人思考相应问题时的逻辑步骤,并发现这些逻辑过程之间的依赖关系。如此,便可以在模型层面,在保证模型效率的前提下,“人工构建”出模型的可解释性。

3.2 算法透明度

国内法律人工智能的实际应用过程中,算法对于司法工作者来说通常是“黑盒子”状态。对于算法的正确率,即使在不同应用场景下,开发人员会通过 F1 Score^[26],Rouge 等^[27]指标对其模型进行评估,但似乎总能找到合适的方式证明其模型的高效性。一个模型的泛化能力究竟如何,需要实践来检验。基于过去的案件开发的模型对现在以及未来的案件是否适用,存在着许多未知数。对于算法本身,开发人员常常对算法的描述语焉不详,或者作出过高评价。出于对其技术知识的保护,开发人员常常不会将算法完全公开,而激烈的市场竞争更导致了法律科技公司常常对其核心算法或主要内容进行保密处理,对算法的保密甚至还可能得到法律的支持。在美国威斯康星州诉卢米斯一案中,被告人认为 COMPAS 系统的算法存在算法歧视,并要求公开算法。但最后联邦最高法院的裁判结果认为该系统具有中立性,并免除了算法开发者进行公开的义务。

由于人工智能算法中通常存在随机过程,即使对算法进行完全透明的公开,研究人员在对算法进行复现时,不同的数据集、不同的参数和不同的训练环境都有可能使最终的模型结果产生许多不可预知的变化。即使采用统一的数据集、参数集和完全相同的训练环境,由于算法内部存在不确定性,得到的结果也可能不一致。因此有人认为,算法透明度的要求是多此一举。但笔者认为,上述实践中算法复现的客观困难不应当成为算法不公开的正当基础。目前算法透明度已经成为人工智能实践中许多场景下的客观要求。为了规制算法权力,使算法的开发者和使用者承担相对均衡的风险,透明性原则在算法开发中越来越受到人们的重视。因为开发者和使用者之间的信息不对称难以彻底消除,而透明性原则能够弥补决策者与相对人之间形成的“数字鸿沟”,一定程度上减小信息不对称产生的算法垄断和信息垄断,进而保护使用者的权利。

实现算法透明主要包括 3 个方面的内容:公开算法源代码;公布用于算法模型的原始输入数据和输出结果;公开输入到输出的中间处理过程的算法流程,以便被决策对象充分知悉并认同算法的正当合理性等内容^[28]。2017 年,美国计算机学会公众政策委员会针对算法透明度问题,对其进行细化,并提出了知情原则、质询和申诉原则、算法责任认定原则、

解释原则、数据来源披露原则与可审计原则等更细节的规定。而欧盟也正在该领域积极探索,根据《通用数据保护条例》(General Data Protection Regulation, GDPR)第 71 条解释(Recital 71),“在任何情况下,应该采取适当的保障,包括向数据主体提供具体信息,以及获得人为干预的权利,以表达数据主体的观点,在评估后获得决定解释权并质疑该决定。”^[29]可见,如何在实践中保障算法透明与算法公开已然成为各国的关注重点,而以追求公平正义为首要价值的法律必然会对人工智能的算法透明度提出更加严苛的要求。

如果说可解释性是法律人工智能算法公信力的基础保障,那么算法的透明度则是其公信力的实现方式。对法院等公权力机构来说,算法的透明度是其进行确定责任风险分配的客观要求。对研究人员来说,即使对模型完全的复现只存在理论上的可能,但这并不意味着算法透明度的需求是无病呻吟。在算法的复现之外,算法透明度给他人论证算法可用性提供了切实的依据,也方便其他研究人员在此基础上进行更深入的研究。对于普罗大众来说,算法透明度显然并不能让他们理解模型采用的具体方法以及方法的优劣,但在公共领域所应用的法律人工智能技术,作为其用户,显然对其算法具有知情权。因此人工智能裁判算法的透明度是其获得大众、法院和研究人员的支持并不断更新迭代的关键因素。

结束语 本文详细分析了目前法律人工智能在数据和算法中遇到的困境,针对部分学者的疑问进行了思考,并对部分困境提出了可行的解决方法。针对目前业内主流的知识图谱搭配深度神经网络的开发模式,在数据方面,笔者认为知识图谱的技术具有较高的实践运用价值,能够有效提高法律数据的结构化程度。但知识图谱的开发工程量较大,且需要法律专家和计算机专家之间大量的沟通,研究之道路阻且长。此外,法律数据的收集和获取还需要注意数据量和数据内容等问题,在保证数据分布的前提下采用更大量、更优质的数据,可以使法律人工智能模型性能的提升事半功倍。但对算法而言,即使部分深度神经网络表现出了比较好的性能,也未必是最佳的答案。因为法律人工智能算法不仅追求准确度与科学性,更需具备可解释性与透明性,否则会对司法的公正价值产生影响。这不仅关乎算法性能及其公信力,甚至会阻碍人工智能在司法领域的发展与推行。

因此,笔者认为,在上述条件尚未成熟之际,人工智能适合担当法律场景中的辅助者,而不是决定者。尤其在司法裁判等场景中,需要大量人类经验和价值判断,当下的人工智能显然没有达到这样的高度。目前人工智能在司法裁判中切实的任务,应当是整合先前案例中法官裁判的成功经验,提出指导性的意见供法官参考,并且对于算法生成的裁判结果,有必要对相应的计算过程进行解释。在其他许多应用场景中亦是如此,而不是在算法、数据的可靠性不能得到保证前,盲目进行法院“智能化”,这会在一定程度上消解法院的权威性和主体性。公平正义是人类社会所追求的首要价值目标,是司法的灵魂。法律界在积极拥抱新技术的同时,更应当谨慎思量司法被技术异化的可能及风险。我们可以相信技术进步给司法运行带来的优化,但不能轻易踏入技术陷阱的泥潭之中。人工智能司法裁判的实现,还需要更多法律专家和人工智能研究人员的共同努力。

参 考 文 献

- [1] RAISCH S,KRAKOWSKI S. Artificial intelligence and management: The automation-augmentation paradox[J]. Academy of Management Review,2021,46(1):192-210.
- [2] DOAA A E. Bail or Jail? Judicial Versus Algorithmic Decision-Making in the Pretrial System[J]. The Columbia Science & Technology Law Review,2019,21:376-446.
- [3] CAMPBELL R W. Artificial intelligence in the courtroom: The delivery of justice in the age of machine learning[J]. Colo. Tech. LJ,2020,18:323-350.
- [4] KATZ D M,BOMMARITO II M J,BLACKMAN J. Predicting the behavior of the supreme court of the United States:A general approach[J]. arXiv:1407. 6333,2014.
- [5] RAPPA M,EISENBERG T,DOHERTY J W,et al. The Evolving Role of the Corporate Counsel:How Information Technology is Reinventing Legal Practice[J]. Campbell L. Rev. ,2013,36:383,446-447.
- [6] DING G F. ‘Smart Court’ injects momentum into the modernization of judicial capacity[N]. Legal Daily,2017.
- [7] LIU Z Y. The modernization of judicial system and judicial capacity by the integration of advanced technology and trial implementation[N]. Legal Daily,2016.
- [8] YANG W,JIA W,ZHOU X,et al. Legal Judgment Prediction via Multi-Perspective Bi-Feedback Network [J]. arXiv:1905. 03969,2019.
- [9] PAN S J,QIANG Y. A Survey on Transfer Learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10):1345-1359.
- [10] WANG L M,ZHU X G,WANG D J,et al. Study on Judicial Data Classification Method Based on Natural Language Processing Technologies[J]. Computer Science,2021,48(8):80-85.
- [11] MA C,YU X H,HE H B. Big Data Analysis: The Report about Chinese Online Judicial Judgment Documents [J]. China Law Review,2016(4):195-246.
- [12] ZUO W M. Insights on the prospect of using legal AI in China [J]. Tsinghua University Law Journal,2018,12(2):109-125.
- [13] HUA L. Ten key words of judicial reform in 2020, the public disclosure of adjudication documents broke 100 million[J]. China Trial,2021,3:30-31.
- [14] XIANG L,YANG X,ZHANG J,et al. A word-frequency-preserving steganographic method based on synonym substitution [J]. International Journal of Computational Science and Engineering,2019,19(1):132-139.
- [15] WANG Z,ZHANG J,FENG J,et al. Knowledge graph embedding by translating on hyperplanes [C] // Proceedings of the AAAI Conference on Artificial Intelligence, 2014:1112-1119.
- [16] QIN C J,HOU H Q. Mapping Knowledge Domain—A New Field of Information Management and Knowledge Management [J]. Journal of Academic Libraries,2009,27(1):30-37.
- [17] HUANG W Y. Deep Neural Networks for Legal Question Answering based on Knowledge Graph[D]. Hefei: University of Chinese Academy of Sciences(Shenzhen Institutes of Advanced Technology,Chinese Academy of Sciences),2020.
- [18] MOSES L B,CHAN J. Using big data for legal and law enforcement decisions: Testing the new tools[J]. The University of New South Wales Law Journal,2014,37(2):643-678.
- [19] ZHANG L H. A study of algorithmic interpretation rights for automated business decisions[J]. Science of Law (Journal of Northwest University of Political Science and Law), 2018, 36(3):65-74.
- [20] SUTSKEVER I,HINTON G E. Deep,narrow sigmoid belief networks are universal approximators[J]. Neural Computation, 2008,20(11):2629-2636.
- [21] CHAZETTE L,SCHNEIDER K. Explainability as a non-functional requirement: challenges and recommendations [J]. Requirements Engineering,2020,25(4):493-514.
- [22] SZEGEDY C,ZAREMBA W,SUTSKEVER I,et al. Intriguing properties of neural networks[J]. arXiv:1312. 6199,2013.
- [23] TAO HC,XU Y C. Construction of the Mechanism of Algorithm Interpretation—Algorithmic Logic Regulation on Judicature[J]. Journal of Wuhan Vocational College of Transportation, 2021,23(3):50-57.
- [24] ZHANG W T. Research on Judicial Documents Classification Method based on Bayesian Network [D]. Nanjing: Southeast University,2019.
- [25] ZHONG H,GUO Z,TU C,et al. Legal judgment prediction via topological learning[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018:3540-3549.
- [26] SALTON G,HARMAN D. Information retrieval[M]. Encyclopedia of Computer Science, 2003:858-863.
- [27] LIN C Y. Rouge: A package for automatic evaluation of summaries[C]// Proceedings of the Workshop on Text Summarization Branches Out, 2004:74-81.
- [28] ZHENG Z H. The Ethical Crisis and Legal Regulation of Artificial Intelligence Algorithms[J]. Social Sciences Digest,2021(4):74-76.
- [29] LEPRI B,OLIVER N,LETOUZÉ E,et al. Fair, transparent, and accountable algorithmic decision-making processes[J]. Philosophy & Technology,2018,31(4):611-627.



CONG Ying-nan, born in 1985, Ph.D, senior lecturer, master supervisor, is a member of China Computer Federation and Chinese Association for Artificial Intelligence. His main research interests include big data on business and law, artificial intelligence, blockchain, Fin-tech, Reg-tech and complex system.



ZHU Jin-qing, born in 1984, postgraduate, engineer, is a member of China Computer Federation and Chinese Association for Artificial Intelligence. His main research interests include database systems, content data analysis, artificial intelligence and knowledge graphs.