

DOI:10.13288/j.11-2166/r.2022.06.002

论智能中医辨证可解释研究的挑战和研究思路

张婷婷¹, 王亚强², 温川飙¹, 叶莹¹✉

1. 成都中医药大学, 四川省成都市金牛区十二桥路37号, 610075; 2. 成都信息工程大学

[摘要] 智能中医辨证的研究已有数十年的历史,但由于不理解机器做出决策的原理,决策者难以信任机器的辨证结果,导致至今智能辨证的临床应用仍十分有限。通过分析中医智能辨证的发展,认为其局限主要包括有限的可用数据、语料库的缺乏、评价方法的局限,且中医辨证本身的解释难度大和非规范化等因素,致使中医辨证可解释研究的开展面临挑战。提出智能辨证可解释研究具有必要性和可行性,未来的可解释研究需要多学科合作,充分利用已有的解释资源,构建辨证的解释体系和大型语料库,以支持对可解释智能中医辨证的探索。

[关键词] 智能辨证; 可解释; 人工智能; 自然语言处理

辨证是中医认识疾病的基本原则,是中医对疾病的一种特殊的研究和处理方法,也是中医学区别于其他医学的重要特征。早在东汉时期的《伤寒论》中“观其脉证,知犯何逆,随证治之”,强调了“证”在疾病诊疗过程中的重要性。辨证是中医诊疗过程中的关键环节,对治疗方案乃至治疗效果有决定性的影响。近40年来,我国先后研制出数百种辅助辨证的智能系统,包括专家系统、智能诊断系统、辅助诊疗系统等^[1-3]。然而,智能中医辨证的临床应用至今仍然十分有限,重要原因之一是智能辨证解释性的不足。由于不理解机器做出决策的原理,决策者和患者均难以信任机器的辨证结果。基于此,辨证的可解释性成为智能中医辨证从科研到临床应用的重要条件。本文回顾智能中医辨证的发展和局限,梳理可解释性的定义和概况,从文本挖掘的角度提出智能中医辨证可解释研究面临的挑战和初步研究思路,以期为未来的研究提供参考。

1 中医智能辨证论治的发展和局限

中医数字化辨证一直都是人工智能在中医领域

的热点研究课题。20世纪70年代中后期以来,我国研发了许多以中医数字辨证论治相关的系统^[4],也尝试了基于知识图谱开展中医数字辨证知识表示与推理研究^[5]。近年来,利用深度学习模型研究中中医辨证论治成为热门^[6-8],在复杂症状中提取、归纳中医证型,分析症状、证型之间潜在的关联规则^[6,9-12],辅助临床诊疗工作^[13-14]等方面均取得了一定成果。然而,目前基于深度学习的端到端的辨证模型解释性很低,“黑箱”问题导致机器与人之间的信任难以建立。用户会产生疑问,比如为什么是这个结果?为什么不是那个结果?如何信任这个结果?什么情况下结果是错误的?在这样的背景下,中医领域迫切需要从更加详细和具象的角度理解得到结论的原因,解决“为什么是”“为什么不是”等关键问题,以获得用户信任并支撑临床应用。与此同时,较强的可解释性也有利于确保辨证模型的公平性、隐私保护性能、鲁棒性,提升中医决策者对模型的信任程度。

2 “可解释”的定义和发展概况

在人工智能领域,具有解释能力的人工智能模型被称为“可解释的人工智能”。在机器学习场景中,可解释性(interpretability)被定义为向人类解释或呈现可理解的术语的能力,即将抽象概念映射到人类可以理解的领域中^[14]。美国国防高级研究

基金项目:国家自然科学基金(61801058);中国博士后科学基金(2021M693785);成都中医药大学“杏林学者”学科人才科研提升计划(BSH2020031);成都中医药大学校教育教学改革基金(XJG2019-11982)

✉ 通讯作者:yeying@cudtc.edu.cn

计划局(Defense Advanced Research Projects Agency, DARPA)在2016年8月提出了可解释的人工智能(explainable artificial intelligence, XAI)计划^[15],目的是建设一套全新、可解释的深度学习模型。该计划的四大原则草案由美国国家标准局于2020年8月公开,分别是提供解释、解释有意义、解释的精确度、系统的知识边界。神经信息处理系统(neural information processing systems, NIPS)2017的工作组曾针对“可解释性在机器学习中是否必要”这一问题展开讨论^[16],强调可解释性在医疗行业的重要性和必要性。

近年来,可解释成为人工智能领域的热点研究话题,深度学习^[17]的出现让各大领域对可解释的需求更加迫切,越来越多的解释方法和模型被提出用于解释深度学习的模型和结果^[15,16-20]。在医疗行业,基于现有的可解释技术,已经有学者开始研究辅助诊断的可解释。如有研究团队^[21]研发了一种可解释的心血管疾病的诊断系统,采用注意力机制自动生成热力图,显示对诊断结果有显著影响的像素,利用视网膜图像为参考来预测和心血管疾病相关的各种风险因素,具有70%的准确率。还有学者^[22]研究出具备可解释性的精确诊断致盲视网膜疾病的肺炎医疗诊断系统,可以在30 s内确定患者是否需要治疗,准确率高达95%。另有研究者^[23]以肺癌为研究对象,提出一种面向TNM分期的中文影像报告肿瘤信息提取方案,利用命名实体识别和关系抽取组成的信息提取技术,从报告中自动获得与分期相关的文字表达,再利用规则对获取文字进行分析推断,得到布尔型或数值型的分期信息,可直接服务于分期决策,并且提取信息具有报告文字作为证据支撑,具有一定的解释性。可见,现代医学已经开始进行临床诊疗的可解释研究,并初具成果。

3 中医智能辨证可解释研究的可行性和必要性

中医辨证是一个复杂的思维过程,既有依据中医辨证先验知识的逻辑推理,也有基于临床大量医案的归纳演绎。中医辨证方法丰富,临床常用的包括脏腑经络辨证、气血辨证、三焦辨证、八纲辨证、六经辨证等多种辨证系统^[24]。在复杂的中医基础理论和辨证方法的支撑下,尽管整个辨证过程非常复杂但却有根有据。例如,失眠患者出现细数脉和心肾不交证之间的逻辑关系是:失眠,脉细

数→细提示阴虚,数提示阳亢→肾水不能上济心阴,心阴虚而难以抑上炎之心火,心火不能下降以温暖肾水→心火旺→扰动心神→神难安→失眠→诊断为心肾不交证。可见,在中医的先验知识和医生经验的支撑下,中医辨证过程是可解释的,利用人工智能技术解释中医智能化辨证决策在理论上是可行的。中医临床对安全性的要求非常高,错误的决策是难以被接受的。在这样的背景下,“可解释”给中医辨证智能化从科研到临床实际应用带来了可能,并成为人工智能技术在中医临床辨证诊断中受到信任和运用的重要条件。

近年来,中医药领域也开始研究深度学习的可解释,如骆明楠^[25]在舌诊的深度学习模型中引入注意力机制,以改善神经网络对图像的建模性能,提高智能舌诊的病性诊断效果,以及增加智能舌诊的可解释性。Pang等^[26]在辨证过程中引入注意力机制,在症状特征之间为不同的症状分配不同的权重,突出了主要意义的症状,具有一定的解释性。然而,虽然注意力机制具有一定的解释作用,但是对辨证过程的解释并不充分,没有呈现真实临床的中医思维。在中医智能辨证领域,可解释研究尚处于起步阶段。一方面,中医决策者不理解和认可目前的以评价指标衡量模型性能的解释方法;另一方面,辨证是中医诊断知识的综合体现,没有中医学背景的技术人员对复杂的辨证过程无法充分理解,导致可解释性的研究在中医智能辨证领域难以开展。未来的可解释研究,不仅仅是字符和数值的比对,更需要从语义的层面入手,不仅理清解释症与证之间的对应关系,更要进一步说明对应关系中所蕴涵的辨证思维和逻辑推理。

4 中医智能辨证可解释的挑战

4.1 可利用的解释数据不足且质量参差不齐

由于大多数临床记录缺少解释,而来自非作者的解释又难以得到广泛认可,因此目前可用于辨证解释的资源并不多,主要包括住院病历、门诊医案以及论文。住院病历大多由一线住院医生记录,其解释内容质量参差不齐。相比而言,由经验丰富的中医专家或其徒弟整理和记录的医案和按语能够更大程度地体现中医专家的学术经验,其讨论部分的内容对可解释性的研究有很好的参考价值。然而,还有部分医案记录的解释非常简略,按语部分寥寥数语,让人难以完全掌握中医专家的辨证思维过

程。此外,论文也是可用于解释的数据,主要是个案报道,但作者分散,整理起来相对困难。从现存的可用于解释研究的数据来看,略者多、详者少,尽管专家本人医术精湛,但是其辨证治疗的思维过程难以从医案按语中完全体现,因此研究者难以通过医案的内容充分继承其学术思想。

4.2 中医辨证及解释的差异性

中医学数千年历史中呈现出一源多流的现象,中医学学术呈现百家争鸣的繁荣景象,也导致辨证观念存在了显著的差异性。中医学属于经验医学,除了丰富的中医理论,中医辨证过程还体现了专家多年的诊疗经验和习惯。即使面对相同的患者,不同地域(如南方和北方)、不同学科领域(如针灸和内科)、不同流派(如川派和新安医学)的专家的辨证思维和结果也可能不同。而辨证思维的不同,导致辨证的解释难以统一,甚至存在争议。

4.3 解释语料库的缺乏

目前,中医领域自然语言处理的研究很大程度上依赖于“金标准”的语料库,即受到领域广泛认可的带标记数据库。但中医领域语料库极度缺乏,目前已发表的少数语料库^[9,27-30]也都是针对某个特定任务的语料库,且规模较小。要进行辨证的解释,需要构建用于解释研究的金标准语料库,用以训练和测试模型,同时也能用于模型性能的测评。目前,中医领域的解释研究的语料库的构建面临巨大的挑战,数据的标注需要耗费大量的时间和精力,且权威专家也很难有时间来参与标注工作,导致目前中医领域解释语料的标注非常困难。

4.4 智能辨证结果评价的局限

在当下的深度学习浪潮中,许多新发表的工作都声称自己可以在目标任务上取得良好的性能。然而,由于缺少对决策和预测的解释,人们无法获知所有结论的内在逻辑,所以只能用准确率、错误率、回归等评价标准来代替评估模型的性能和结果的可信度^[31]。显然,这样的评价是不足的,只能解释模型预测决策的概况。在中医应用场景中,需要结合中医特点,制定中医甚至患者能够理解的评价方式,来说明决策是否可信。

5 中医智能辨证可解释的研究思路

5.1 多领域专家合作

尽管可解释性是人工智能技术在中医临床受到认可和运用的重要条件,但是由于中医领域知识的

专业性非常强,中医辨证过程拥有复杂的内部逻辑和层次递进关系,导致技术人员不清楚真实中医辨证过程的“解释”是什么样的,中医需要的“解释”是什么,如何评价“解释”是合理的、可信的。因此,技术人员难以利用计算机技术对智能中医临床的研究结果进行合理的、可信的、直观的解释。换言之,中医决策者和技术人员之间缺乏有效的连接。因此,中医临床辨证论治的可解释研究需要技术人员和中医专家,甚至语言学专家的合作。这种合作应该是深入的,且参与合作的专家都应具有一定的交叉学科背景,对任务有充分的认识。

5.2 充分利用已有的解释资源

目前可用于中医辨证解释研究的数据资源不多,因此一定要充分利用目前已有的可用于解释研究的资源。这些资源包括:1)中医基础理论知识。即获取中医辨证的先验知识,具体包括中医基础理论知识、方剂学基础知识、诊断学基础知识等中医临床初学者必须掌握的基本知识。2)有解释内容的临床记录。筛选出“辨证依据”可用的病历和按语可用的医案,以及部分案例报道的论文。在数据筛选过程中,可以利用大数据技术,通过数据清洗等方法,减少劣质数据对结果的影响。

5.3 构建辨证解释体系

构建辨证的解释体系即将基于真实的中医临床辨证过程,从更为详细和具象的角度,去研究中医决策者和技术人员都能明白的辨证解释要点和表达方式,以及中医专家理解和认可的评价方式,并形成体系。构建辨证解释体系是为了让技术人员能够充分理解中医辨证的解释原貌,让更多的计算机专家能够参与到中医辨证的解释研究中,同时让中医专家能够理解和认可解释的内容。

5.4 构建大规模解释语料库

中医领域目前主流的数据挖掘技术是有监督的深度学习模型,需要一定规模的标注语料为机器提供训练数据和测试数据。目前我国公开的医学领域的语料库很少,而散在的语料库相关的研究都是以国外经典标注语料库为参考。在中医领域,目前尚未有与解释相关的语料库发布。构建大规模的语料库是今后进行解释研究的重要途径之一。

6 结语

“可解释”的提出给中医智能辨证技术的临床应用带来了新的可能。然而,由于可利用数据的不

足、数据质量的参差不齐、语料库的欠缺、评价方法的局限等,辨证的可解释研究仍未取得长足的进展并面临巨大的挑战。据此,我们提出辨证可解释的初步研究思路,包括多学科合作、充分利用已有的解释资源、构建辨证的解释体系和大型语料库,以支持开展智能中医辨证可解释的研究。此外,中医数据挖掘的基础研究目前还比较薄弱,如中文分词、实体识别、关系抽取、数据标准化、指代消解等,加强基础研究,解决术语不规范、语义模糊、指代不明等关键问题,可为辨证的解释研究奠定坚实的基础。概言之,“可解释”是未来中医智能辨证能为临床应用的必要课题,目前面临巨大挑战,希望更多的学者重视并参与其中。

参考文献

- [1]白春清. 中医专家系统三十年[J]. 医学信息(上旬刊), 2011, 24(2): 550-552.
- [2]杨涛, 朱学芳. 中医辨证智能化研究现状及发展趋势[J]. 南京中医药大学学报, 2021, 37(4): 597-601.
- [3]石英杰, 李宗友, 赵攀, 等. 我国中医计算机辅助诊断研究热点与趋势的知识图谱分析[J]. 中国中医药图书情报杂志, 2021, 45(2): 11-18.
- [4]韦昌法, 晏峻峰. 从知识表示与推理方法探讨中医数字辨证发展[J]. 中华中医药杂志, 2019, 34(10): 4471-4473.
- [5]李新龙, 刘岩, 何丽云, 等. 知识图谱研究概况及其在中医药领域的应用[J]. 中国中医药信息杂志, 2017, 24(7): 129-132.
- [6]HU Q, YU T, LI J, et al. End-to-End syndrome differentiation of Yin deficiency and Yang deficiency in traditional Chinese medicine [J]. Comput Meth Prog Bio, 2019, 174: 9-15. doi 10.1016/j.cmpb.2018.10.011.
- [7]李佳佳. 基于深度学习的高血压病肝火亢盛证面部色诊研究[D]. 北京: 中国中医科学院, 2020.
- [8]张阳, 凌娜, 张国华, 等. 基于深度学习多囊卵巢综合征辨证分型的构建与实现[J]. 辽宁中医药大学学报, 2019, 21(5): 13-16.
- [9]LIANG Z, LIU J, OU A, et al. Deep generative learning for automated EHR diagnosis of traditional Chinese medicine[J]. Comput Meth Prog Bio, 2019, 174: 17-23. doi: 10.1016/j.cmpb.2018.05.008.
- [10]ZHANG T, LIU M. The causality research between syndrome elements by attribute topology [J]. Comput Math Methods Med, 2018: 9707581. doi: org/10.1155/2018/9707581.
- [11]WAN H, MOENS M, LUYTEN W, et al. Extracting relations from traditional Chinese medicine literature via heterogeneous entity networks [J]. J Am Med Inform Assoc, 2016, 23(2): 356-365.
- [12]ZHANG H, NI W. Artificial intelligence-based traditional Chinese medicine assistive diagnostic system: Validation study[J]. JMIR Medical Informatics, 2020, 8(6): e17608. doi:10.2196/17608.
- [13]ZHOU X, CHEN S, LIU B, et al. Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support [J]. Artif Intell Med, 2010, 48(2/3): 139-152.
- [14]DOSHI-VELEZ F, KIM B. Towards a rigorous science of interpretable machine learning[J]. arXiv, 2017: 1702.08608.
- [15]GUNNING D, AHA DW. DARPA's Explainable Artificial Intelligence Program [J]. AI Magazine, 2019, 40(2): 44-58.
- [16]SAMEK W, MONTAVON G, VEDALDI A, et al. Explainable AI: interpreting, explaining and visualizing deep learning[M]. Switzerland :Springer Nature, 2019.
- [17]LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [18]LINARDATOS P, PAPASTEFANOPOULOS V, KOTSIANTIS S. Explainable AI: A review of machine learning interpretability methods [J]. Entropy (Basel), 2020, 23(1): 18.
- [19]GUIDOTTI R, MONREALE A, RUGGIERI S, et al. A survey of methods for explaining black box models [J]. ACM Comput Surv (CSUR), 2018, 51(5): 1-42.
- [20]ADADI A, BERRADA M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI) [J]. IEEE access, 2018, 6: 52138-52160. doi: 10.1109/ACCESS.2018.2870052.
- [21]POPIN R, VARADARAJAN A, KATYBLUMER, et al. Predicting cardiovascular risk factors from retinal fundus photographs using deep learning [J]. Nat Biomed Eng, 2018: 158-164. doi: 10.1038/s41551-018-0195-0.
- [22]KERMANY DS, GOLDBAUM M, CAI W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning [J]. Cell, 2018, 172(5): 1122-1131. e9.
- [23]王雨虹. 面向TNM分期的中文影像报告肿瘤信息提取研究[D]. 杭州: 浙江大学, 2020.
- [24]朱文锋. 中医诊断学[M]. 北京: 中国中医药出版社, 2007.
- [25]骆明楠. 面向智能舌诊的注意力卷积网络研究[D]. 广州: 华南理工大学, 2020.
- [26]PANG H, WEI S, ZHAO Y, et al. Effective attention-

- based network for syndrome differentiation of AIDS [J]. BMC Med Inform Decis, 2020, 20(1): 1-10.
- [27] CAO C, SUN M, WANG S. Extracting terms from clinical records of traditional Chinese medicine [J]. Front Med Pre, 2014, 8(3): 347-351.
- [28] LI Y, ZHOU X, ZHANG R, et al. Detection of herb-symptom associations from traditional Chinese medicine clinical data [J]. Evid Based Compl Alt, 2015: 1-11. doi: 10.1155/2015/270450.
- [29] XU D, ZHANG M, ZHAO T, et al. Data-Driven Information extraction from Chinese electronic medical records [J]. PLoS One, 2015, 10(8): e0136270. doi: 10.1371/journal.pone.0136270.
- [30] ZHANG T, WANG Y, WANG X, et al. Constructing fine-grained entity recognition corpora based on clinical records of traditional Chinese medicine [J]. BMC Med Inform Decis, 2020, 20(1): 2-17.
- [31] JOSÉ HO. Evaluation in artificial intelligence: from task-oriented to ability-oriented measurement [J]. Artif Intell Rev, 2017, 48(3): 397-447.

Challenges and Ideas in the Interpretability of Intelligent Syndrome Differentiation of Traditional Chinese Medicine

ZHANG Tingting¹, WANG Yaqiang², WEN Chuanbiao¹, YE Ying¹

1. Chengdu University of Traditional Chinese Medicine, Chengdu, 610075; 2. Chengdu University of Information Technology

ABSTRACT Although it has been decades since the intelligent syndrome differentiation was produced in traditional Chinese medicine (TCM), its clinical application has been limited due to the difficulty in understanding how machines make decisions and thereafter the distrust of decision makers in the machine results. The development of intelligent TCM syndrome differentiation is mainly restricted by limited available data, lack of corpus, insufficient evaluation methods, as well as the challenges in interpretability and standardization of TCM syndrome differentiation. Therefore, it is necessary and feasible to conduct researches to explore the interpretability of intelligent TCM syndrome differentiation by integrating multi-disciplines, making full use of existing interpretability-related resources, and building syndrome-differentiated interpretable system and large corpora.

Keywords intelligent syndrome differentiation; interpretability; artificial intelligence; natural language processing

(收稿日期: 2021-11-04; 修回日期: 2021-11-29)

[编辑: 侯建春]

广告

《中医杂志(英文版)》欢迎投稿

Journal of Traditional Chinese Medicine [《中医杂志(英文版)》] (ISSN 0255-2922, CN 11-2167/R, 双月刊)由中华中医药学会与中国中医科学院联合主办,是我国中医药界具有权威的中医药英文科技期刊,现发行和交换90多个国家和地区。

《中医杂志(英文版)》自1981年创刊以来一直坚持正确的办刊宗旨,国际影响力不断提升。2010年入选SCI-E(《科学引文索引》扩刊版),2011年获得中国中医科学院自主选题项目资助,2013年获得“中国科技期刊国际影响力提升计划”资助,自2012年起连续被评为“中国最具国际影响力学术期刊”。本刊先后被国际知名文献数据库MEDLINE、《化学文摘》(CA)、《医学文摘》(EM)等收录,是MEDLINE收录的第一份中国出版的中医药期刊,是SCI收录的30余种全科医学和替代医学期刊中的中国传统医学综合性学术期刊之一。

《中医杂志(英文版)》主要栏目包括循证研究、理论研究、基础研究、临床观察、中药研究、文献研究,思路与方法、综述等。内容以报道中医药基础、临床研究成果和科研前沿进展为主,覆盖中医药学科各个领域。欢迎广大专家学者积极投稿。作者可直接登录本刊网站<http://www.journaltcm.com>投稿。

联系邮箱: jtcmen@126.com。电话: 010-64089190。

本社地址:北京市东城区东直门内南小街16号,邮政编码:100700。