

DOI:10.13288/j.11-2166/r.2022.09.001

学术探讨

构建知识与数据双驱动中医诊疗系统的可行性及路径分析

许鸿本^{1,2}, 路熹雅¹, 刘怡馨¹, 高灵婧¹, 周常思^{1,2}, 杨朝阳^{1,2}✉

1. 福建中医药大学中医学院, 福建省福州市闽侯上街邱阳路1号, 350122; 2. 福建省中医健康状态辨识重点实验室

[摘要] 在梳理知识驱动及知识工程、数据驱动及大数据等人工智能技术要素的基础上, 根据基于知识驱动与基于数据驱动的中医诊疗系统各自的优势与不足, 提出基于知识与数据双驱动的人工智能来构建中医诊疗系统, 以适用于中医临床。从中医学学科属性、基于逻辑演绎的中医药理论、基于归纳总结的实践经验角度分析构建双驱动诊疗系统的可行性, 同时从中医药领域知识图谱、中医药医案大数据、知识图谱与大数据融合方面构建双驱动诊疗系统的实现路径, 以期为中医诊疗系统的发展提供思路。

[关键词] 中医诊疗系统; 知识驱动; 数据驱动; 知识图谱; 神经网络

人工智能 (artificial intelligence, AI) 作为计算机学科的重要分支, 自 1956 年诞生以来, 一直存在着两大相互竞争的范式, 即符号主义与连接主义^[1]。虽然关于人工智能的定义至今尚未统一, 但总体上可以概括为研究人类智能活动的规律, 构造具有一定智能行为的人工系统^[2]。有研究^[3]认为, 人类的智能主要包括逻辑演绎和归纳总结, 对应着人工智能中的符号主义和连接主义。符号主义在 20 世纪 90 年代以前主导着人工智能的发展, 而连接主义是 20 世纪 90 年代后才逐渐发展起来, 并在 21 世纪得到快速发展^[1]。符号主义和连接主义的先后发展, 也间接导致中医诊疗系统历经了以知识驱动为主和以数据驱动为主的两个不同发展阶段。知识驱动中医诊疗系统以专家经验、知识规则为主, 而数据驱动中医诊疗系统则以诊疗规律、大数据为主。中医诊疗过程既需要理论知识也需要诊疗经验, 故无论是知识驱动还是数据驱动都无法完全满足临床实际, 无法充分体现中医思维。当前, AI 领域正掀起一场变革, 即通过符号 AI 与连接 AI 的融合, 构建知识与数据双驱动的 AI, 这与中医诊疗过程契合, 因此中医诊疗系统有望在该技术的帮

助下突破现阶段的困境。

1 人工智能技术要素

1.1 知识驱动及知识工程

符号主义学派的科学家认为, 实现 AI 必须使用逻辑和符号系统^[4]。根据此观点, 科学家提出了基于知识与经验的推理模型。而这需要在机器里面建立知识库和推理机制, 利用这两者对人类的推理和思考行为进行模拟, 此为逻辑演绎, 因此符号 AI 又称为知识驱动方法。

知识是用于解决问题的具有指导意义的信息。斯坦福大学教授费根堡姆在第五届国际人工智能会议中提出“知识工程”这一概念, 认为知识尤其是特定领域的知识才是人类智能的基础, 对专家知识的获取、表达和推理是设计基于知识系统的重要技术问题^[5]。知识工程代表的知识驱动方法主要过程包括知识发现、知识表示、知识推理等, 这同时也是构成专家系统的核心要素^[6]。一个典型的专家系统主要由知识库、数据库、推理机、解释机制、知识获取和用户界面六部分组成^[2]。基于知识驱动的中医诊疗系统大多为中医专家系统。

1.2 数据驱动及大数据

连接主义学派的科学家认为, 模拟人类大脑中的神经网络可以实现 AI^[4]。基于此观点, 科学家提出了人工神经网络模型。人工神经网络模型是由

基金项目: 福建省科技计划高校产学研合作项目(2020Y4017); 福建中医药大学校管科研课题(X2019001-财政专项)

✉ 通讯作者: yzy813@126.com

大量的节点之间相互联接而构成,按不同的连接方式组成不同的模型,利用模型对数据进行学习,此为归纳总结,因此连接 AI 又被称为数据驱动方法。

信息技术飞速发展,数据呈现指数级的增长,大数据时代已经到来。而大数据的存储、计算和分析是大数据处理中的关键技术,其中数据分析环节是大数据处理流程的核心和关键^[7]。人脑是天然的信息处理引擎,大数据时代下,传统的数据处理方法已经无法解决海量的信息增长,而人工神经网络是目前大数据分析中最成功的方法之一^[8]。“大数据+神经网络”能够推动各行业的发展。在中医领域,“大数据+神经网络”所带来的思维方式变革分别从整体性、模糊性、多样性和相关性方面推动了中医药现代化的发展^[9],近年来在中医证候研究、中医学术经验传承等方面取得了一定的成果^[10]。

2 中医诊疗系统研究进展及不足

2.1 基于知识驱动的中医诊疗系统

1974 年,斯坦福大学基于知识驱动开发了 MYCIN 专家系统,该系统能够协助内科医生诊断、治疗细菌感染疾病^[11]。受此启发,国内中医界有不少学者加入到专家系统的研究中,这是中医诊疗系统的第一个发展阶段,也是计算机人工智能在中医诊疗领域应用的起步时期^[12]。在该阶段共有 180 多个专家系统面世,囊括大部分中医类病种,如外感病、脾胃病、月经病等,不仅实现了对当代名老中医学术经验的模拟和传承,还复现了古代名医的诊疗经验^[13]。其中最具代表性的是关幼波肝炎诊断治疗程序,作为第一个中医专家系统,该程序以概率论、模糊集合论等作为研究中医诊断数学模型的工具,将关幼波诊疗肝炎的经验和过程转化为诊疗程序^[14]。

随着对知识工程和专家系统的深入研究,基于知识驱动的人工智能暴露了自身的局限性。虽然通过建立知识库和推理机制可以模拟人类的理性智能行为,具有可解释性和与人类一致的显式推理过程等优势,但与此同时也存在如下 4 点问题:1) 专家知识稀缺、需要人工编程、费时且昂贵;2) 以自然语言形式表示的知识难以处理;3) 知识表示与推理困难,尤其是不确定性知识表示与推理;4) 难以从数据中自动获取知识。

因此,基于知识驱动的中医诊疗系统仅能作为案例推理参考,不能进行自我学习,其智能化程度偏低。同时,人工编程工作量大,即使是单病种系统仍然需要大量的医案来覆盖临床常见症状,而后

期的优化更需要人力对知识库的不断补充。此外,基于知识库推理的中医诊疗系统的诊疗有效性过于依赖最初建立的知识库,而当患者的病情在知识库之外,便很难得出准确或者接近的结果。

2.2 基于数据驱动的中医诊疗系统

由于符号 AI 的局限性,中医诊疗系统的研究于 20 世纪 90 年代进入一段低谷时期,而后随着连接 AI 开始兴起,中医诊疗系统也迎来第二个发展阶段。21 世纪以来,AI 在中医诊疗领域进入深入应用阶段,该阶段主要基于人工神经网络、贝叶斯网络、本体语言、支持向量机、粗糙集等更为先进的人工智能方法构建诊疗模型。如名老中医方证医学经验的自动诊疗处方系统^[15]、BP 神经网络中医辨证模型^[16]、中医体质辨识模型^[17]等。该阶段诊疗系统以大数据为特点,通过不同的神经网络算法对大量的医案进行学习,从而构建不同诊疗思维的中医诊疗系统,不仅可以处理更多数据,还有利于搭建平台,对实现中医诊疗规范化、标准化起到了推进作用。

连接 AI 发展到现阶段同样也暴露了自身的局限性。虽然连接 AI 可视为一种通用的工具,即不需要领域知识及专家知识,使用的技术门槛低,同时还可以构造一个巨大的多层次与多维的空间,能够处理大量的数据,但与此同时也存在不可解释性和抗干扰能力弱等问题。不可解释性主要是指人还不能理解连接 AI 是如何进行决策的,即所谓的“黑箱理论”。不同于符号 AI 模拟的是人类的逻辑演绎,连接 AI 模拟的是人类的归纳总结,而这个归纳总结是如何进行以及如何得出结论的,目前尚未明确。抗干扰能力弱主要体现在对数据样本的依赖性过高。由于中医学历史悠久,数据受到多种方面的影响,加之目前术语规范化推广困难^[18],导致中医数据样本质量不佳,也影响了系统的诊疗效果。

虽然基于数据驱动的中医诊疗系统对于挖掘医案背后的诊疗规律有独特的优势,连接 AI 作为一种通用工具也为构建诊疗系统提供了极大的便利,但同时也存在诸多问题。首先,连接 AI 的不可解释性弱化了中医理论的指导,基于数据驱动的中医诊疗系统多数是通过从医案学习构建而成的,而如何明确诊断和如何基于诊断对患者实施个性化的治疗,这部分内容是被忽略的。其次,连接 AI 抗干扰能力弱的问题给基于数据驱动的中医诊疗系统带来了从医案数据样本依赖性过高的弊端,如医案数据来源广泛而繁杂容易导致模型失真、准确度差,

而医案数据来源单一纯粹又会导致模型泛化能力差、难以满足临床实际需求。

由上分析可知,无论是基于知识驱动的中医诊疗系统还是基于数据驱动的中医诊疗系统都有其不足之处,离实际应用相差甚远,通过基于知识与数据双驱动的人工智能来构建中医诊疗系统则有望摆脱当前的困境。

3 知识与数据双驱动的中医诊疗系统的可行性及应用前景分析

虽然基于数据驱动的人工智能在一定程度上推动了各行业的发展,但由于其不可解释性的缺点,在某些行业逐渐呈现应用效果不佳、与实际结果不符等表现。这些行业都有一个特点,即既需要专业领域知识的指导,也需要大量的样本学习。从人脑的角度而言,则为既需要逻辑演绎,也需要归纳总结。中医学在几千年的发展过程中,不仅形成了一套成熟的理论体系作为指导,同时还有大量的医案可供学习,因此从中医学学科属性角度分析,中医学是适合通过基于知识与数据双驱动的人工智能来构建中医诊疗系统的。

3.1 中医学学科属性

中医学是以中医药理论与实践经验为主体,研究人类生命活动中健康与疾病转化规律及其预防、诊断、治疗、康复和保健的综合性科学^[19]。也就是说,中医学既包含了以中医药理论作为逻辑演绎,比如“象思维”“阴阳学说”“五行学说”,又包含了实践经验作为归纳总结,比如“治湿不利小便非其治也”“灵机记性不在心在脑”。

3.2 中医药理论——逻辑演绎

中医学在发展过程中,吸纳了中国古代哲学和传统文化,又在长期的临床实践中不断丰富和完善,逐渐形成了自己独特的思维理论,用于指导临床诊疗活动,主要包括阴阳五行学说、藏象学说、气血津液学说、病因学说、病机学说等。现以藏象学说为例,介绍中医药理论作为知识驱动的适用性。

藏象学说是研究人体脏腑生理功能、病理变化规律及相互关系的学说,在中医药临床诊疗中起到了重要的指导作用。“五脏一体观”是藏象学说的主要内容,说明五脏之间既有各自一套生理功能系统,如肺系统(肺-大肠-皮-鼻-毛-涕),同时五脏之间又可通过经络相互联络,在功能上相互配合,在病理上相互影响。根据藏象学说“肺与大肠相表里”,大肠归属于肺系统,当大肠功能失调,从肺论治往往可以取得不错的效果;而肝失疏泄或肝气

郁结可导致脾胃升降失常,气机壅滞,此为“木郁土壅”,临床上常有从肝论治脾胃病的例子。这些理论都可作为领域知识用于知识驱动。

3.3 实践经验——归纳总结

“实践是检验真理的唯一标准”,从古代名医提出的“六气皆从火化”^[20]^[8]、“五志过极皆能化火”^[20]^[27]、“内伤脾胃,百病由生”^[21]等,到现代中医专家提出的“糖尿病肾病以虚、瘀、浊为核心病机”^[22]、“冠心病属本虚标实之证,正虚为内因,而痰、瘀为继发因素”^[23]等,以及在新型冠状病毒肺炎疫情中通过总结临床实践而提出的中医药治疗方案^[24],均体现了中医临床实践的重要性。可见,中医药是在长期的医疗实践中不断归纳总结、升华理论而持续发展的。数据挖掘技术为挖掘医案背后的实践经验提供了帮助,而这有助于深入、全面、准确地研究临床问题,以更接近于真实世界的研究成果来优化或革新临床诊疗模式^[25]。而这些实践经验就属于归纳总结,可作为数据驱动的数据来源。

通过对中医药理论和实践经验的分析可以看出,中医诊疗过程是在中医药理论指导下,不断总结临证经验的过程,这也是中医历久弥新的基石^[26]。因此,单纯的知识驱动或单纯的数据驱动都不能满足中医诊疗系统的临床实际应用,故知识与数据双驱动的中医诊疗系统具有可行性及广阔的应用前景。

4 知识与数据双驱动的中医诊疗系统的实现路径

基于知识驱动的人工智能优势在于对事物全局属性的把握,而基于数据驱动的人工智能在于对事物局部特征的把握,知识与数据双驱动能够将局部与全局、机器的隐式直觉与人的常识判断有机结合起来^[27]。然而符号 AI 与连接 AI 有不同的模型框架,如何将两者结合起来,目前常用的方法是通过构建专业领域知识图谱并将其嵌入神经网络进行训练学习而实现,如以检验相关领域知识为知识驱动、以医院检验大数据为数据驱动的知识与数据双驱动的检验 AI 系统^[28]。

4.1 中医药领域知识图谱

知识图谱是构建基于知识与数据双驱动的人工智能研究中的一个重要环节。知识图谱本质上为具有语义网络的知识库,即具有有向图结构的一个知识库,其中图的结点代表实体或者概念,而图的边代表实体或是概念之间的语义关系^[29]。简单来说就是将知识结构化,变成可被计算机学习的知识

图。一个完整的知识图谱构建过程大致需要三个步骤,即知识抽取、知识融合以及知识加工。其中知识抽取是构建知识图谱的第一步,涉及实体、关系及属性的抽取;知识融合包括实体链接和知识合并,这是对获取的知识进行清洗和梳理的过程,是保证知识质量的步骤;知识加工涉及本体构建、知识推理和质量评估,这是使知识完成结构化、网络化的关键步骤^[30]。

从知识图谱的构建方法来看,要构建中医知识图谱,首先要对中医理论和诊疗经验进行知识抽取,包括抽取知识实体本身以及知识实体间的关系;其次是对抽取的理论和经验进行清洗和梳理,以保证其质量;最后是对融合过后的理论和经验进行加工,使其具有结构化和网络化的属性,以方便计算机学习。

以藏象学说为例,藏象学说中的每条理论知识为实体,理论知识之间的联系为实体间的关系,数据来源可包含藏象理论相关古籍、现代教材等资料;之后对纳入的数据进行融合处理,对释义不清的理论知识予以补充或删除,并按照脏腑、生理、病理、病机等分类对纳入的数据进行梳理;最后对纳入的数据进行加工,通过计算机程序绘制具有语义网络的藏象理论知识图,使藏象理论具有结构化和网络化,如构建的脾脏象理论知识图谱^[31]。

4.2 中医药医案大数据

完成中医药领域知识图谱构建后,需将其嵌入神经网络与医案数据样本进行训练学习,故对医案数据样本质量的把控对模型的构建具有重要的意义。

在长期的医疗实践中累积的大量医案是中医药的宝藏,在大数据时代运用计算机技术对医案进行挖掘,探索相关规律已成为中医药的研究热点。基于对目前数据驱动中医诊疗系统的思考,我们认为,一方面,应对用于纳入的医案数据来源进行严格的范围界定,为避免数据驱动模型失真和泛化能力差,纳入的医案数据应广而不杂。首先是范围应广,应充分纳入不同地区、不同医家以及患者处于不同性别和不同年龄段的医案,并应呈现出平均分配的情况;其次是排除标准、纳入标准应统一,以避免数据杂乱的情况,对于描述不清、没有明确诊断的医案应予以排除,同时多病种、症状复杂的医案也应予以排除。另一方面,应对纳入的数据进行充分的清洗和梳理。首先应对医案进行拆分,按照病名、证型、症状、治疗等内容进行梳理,构建医案数据库;其次应参照标准对纳入的数据进行规范化处理,以保证数据一致性。严格进行以上步骤处理医

案数据,可有效提升数据驱动模型的效果。

4.3 知识图谱与大数据融合

在构建中医药领域知识图谱以及对医案大数据样本处理后,应对两者进行融合,而如何更好地将知识图谱嵌入神经网络仍处于研究阶段。目前常用的方法是将具有语义网络的知识图嵌入神经网络进行融合学习,而图神经网络(graph neural networks, GNN)的诞生为实现融合研究提供了极大的便利。GNN能够从图数据中提取各个节点的邻居特征,并将节点的邻居特征和自身融合形成新的特征表示^[32]。中医学主张“整体观念”和“个性化治疗”,一种疾病往往提示多脏腑的病变,而不同性别和年龄段的患者发病情况也不尽相同。因此,在临证时往往需综合多方面因素的考虑,每一个考虑的因素都可算是一个权重,最终众多权重指向的目标即为诊断结果,这些权重可视为特征。从中医学角度来看,这些特征便是症状与证型之间的关系,而这些关系可借助GNN进行信息提取与模型构建。因此我们认为,对中医药知识图谱和医案大数据样本进行融合的算法模型可采用GNN来构建。

5 结语

基于知识与数据双驱动的人工智能给中医诊疗系统的发展带来希望,但目前仍然面临众多问题,如中医理论利用率低、中医数据样本质量参差不齐以及中医理论与中医实践结合困难等。因此我们提出以图神经网络对知识实体进行提取学习并结合医案数据进行融合训练的初步方案,旨在解决当前中医诊疗系统面临的问题。总之,基于知识与数据双驱动的中医诊疗系统是中医诊疗系统未来的发展方向,目前处于初步发展阶段,诚盼更多的科研工作者参与其中。

参考文献

- [1]张钊,朱军,苏航. 迈向第三代人工智能[J]. 中国科学:信息科学, 2020,50(9): 1281-1302.
- [2]邹蕾,张先锋. 人工智能及其发展应用[J]. 信息网络安全, 2012,12(2): 11-13.
- [3]顾险峰. 人工智能中的联结主义和符号主义[J]. 科技导报, 2016,34(7): 20-25.
- [4]尼克. 人工智能简史[M]. 北京:人民邮电出版社, 2017:97.
- [5]袁国铭,李洪奇,樊波. 关于知识工程的发展综述[J]. 计算技术与自动化, 2011,30(1): 138-143.
- [6]万超. 知识驱动与数据驱动系统构建综述[J]. 长江信息通信, 2021,34(4): 145-147.

- [7]余滨,李绍滋,徐素霞,等.深度学习:开启大数据时代的钥匙[J].工程研究-跨学科视野中的工程,2014,6(3):233-243.
- [8]章毅,郭泉,王建勇.大数据分析的神经网络方法[J].工程科学与技术,2017,49(1):9-18.
- [9]黄欣荣,张艳朋.大数据技术与中医现代化[J].中医杂志,2014,55(19):1621-1625.
- [10]刘艳飞,孙明月,姚贺之,等.大数据技术在中医药领域中的应用现状及思考[J].中国循证医学杂志,2018,18(11):1180-1185.
- [11]赵文,陈启亮,李永,等.四诊合参现代化发展研究[J].天津中医药,2020,37(10):1090-1094.
- [12]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.
- [13]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.
- [14]郭荣江,马斌荣,谢敏,等.中医计算机诊断治疗系统与模糊数学模型[J].计算机学报,1981,4(4):260-266.
- [15]陈肇文.基于人工神经网络的中医证治模型探析[J].中华中医药学刊,2009,27(7):1517-1520.
- [16]徐亮,陈守强,侯建辉,等.基于BP神经网络的中医辨证模型构建方法探讨[J].世界中医药,2016,11(2):335-338.
- [17]严玲,周作建,宋懿花,等.基于ML-kNN多标记学习的中医体质辨识模型研究[J].世界科学技术-中医药现代化,2020,22(10):3558-3562.
- [18]杜含光,李灿东.中医诊断学研究的瓶颈与对策[J].中华中医药杂志,2015,30(4):971-974.
- [19]郑洪新.中医基础理论专论[M].10版.北京:中国中医药出版社,2016:1.
- [20]刘完素.素问玄机原病式[M].北京:人民卫生出版社,1983.
- [21]李东垣.脾胃论[M].北京:中国中医药出版社,2007:8.
- [22]吴群励,杨丹.梁晓春补益脾肾、活血泄浊治疗糖尿病肾病肾功能不全经验[J].中医杂志,2020,61(1):17-19.
- [23]吴焕林,林晓忠,邹旭.邓铁涛治疗冠心病临床经验探析[J].辽宁中医学院学报,2005,7(4):312-313.
- [24]薛伯寿,姚魁武,薛燕星.“清肺排毒汤”快速有效治疗新型冠状病毒肺炎的中医理论分析[J].中医杂志,2020,61(6):461-462.
- [25]梁文娜,林雪娟,俞洁,等.真实世界的大数据助推中医健康管理进入人工智能时代[J].中华中医药杂志,2018,33(4):1213-1215.
- [26]张伯礼.中医思维与实践养成[J].中国中医基础医学杂志,2017,23(5):593-594.
- [27]朱小燕.知识与数据的融合推动人工智能发展[J].信息通信技术,2019,13(1):4-6.
- [28]井杰.知识和数据双驱动的检验人工智能疾病诊断系统的构建[D].上海:中国人民解放军海军军医大学,2021.
- [29]漆桂林,高恒,吴天星.知识图谱研究进展[J].情报工程,2017,3(1):4-25.
- [30]刘屹,李杨,段宏,等.知识图谱构建技术综述[J].计算机研究与发展,2016,53(3):582-600.
- [31]崔家鹏,王彩霞,袁东超,等.基于本体的脾脏象理论知识体系构建研究[J].中华中医药学刊,2018,36(2):388-391.
- [32]WU Z, PAN S, CHEN F, et al. A comprehensive survey on graph neural networks[J]. IEEE Trans Neural Netw Learn Syst, 2021, 32(1): 4-24.

Analysis on Feasibility and Path of Building a Knowledge-Data Dual-driven Traditional Chinese Medicine Diagnosis and Treatment System

XU Hongben^{1,2}, LU Xiya¹, LIU Yixin¹, GAO Lingjing¹, ZHOU Chang'en^{1,2}, YANG Chaoyang^{1,2}

1. School of Traditional Chinese Medicine, Fujian University of Traditional Chinese Medicine, Fuzhou, 350122; 2. Fujian Key Laboratory of Traditional Chinese Medicine Health State, Fujian University of Traditional Chinese Medicine

ABSTRACT By summarizing the technical characteristics of artificial intelligence (AI) such as knowledge-driven and knowledge engineering, data-driven and big data, it is found that both knowledge-driven and data-driven traditional Chinese medicine (TCM) diagnosis and treatment system has its strengths and weaknesses. Therefore, it is put forward to build a TCM diagnosis and treatment system driven by both knowledge and data to promote the suitability in TCM clinical practice. The feasibility of the dual driven system was analyzed from the perspectives of the disciplinary attributes of TCM, logical deduction-based TCM theory, and generalization and summarization-based practical experience. Moreover, the pathway of this dual driven system could be built through TCM knowledge graph, big data of TCM cases, and fusion of knowledge graph and big data, so as to provide ideas for the development of TCM diagnosis and treatment system.

Keywords traditional Chinese medicine diagnosis and treatment system; knowledge-driven; data-driven; knowledge graph; neural network

(收稿日期: 2021-10-10; 修回日期: 2022-01-28)

[编辑: 崔京艳]