

文章编号: 1000-8934(2022)2-0023-07

DOI:10.19484/j.cnki.1000-8934.2022.02.004

符号主义与联结主义人工智能的融合路径分析

魏 斌

(浙江大学 数字法治研究院 杭州 310008)

摘要: 人工智能的演化形成了符号主义与联结主义人工智能两条代表性进路。符号主义人工智能的优点在于推理过程透明、可解释,但存在不完备、框架问题和知识接收瓶颈等问题。联结主义人工智能的泛化能力和可移植能力强,但缺陷是算法不可解释性、过拟合等问题。符号主义与联结主义人工智能有着显著的融合可能性,在新一代人工智能背景下,两条进路的融合又产生了新的路径,在基于神经网络的逻辑推理、符号表示的机器学习、认知图谱及其推理、机器学习与论辩挖掘等交叉领域表现出强劲的融合潜力。

关键词: 符号主义; 联结主义; 人工智能; 融合路径

中图分类号: N031 **文献标识码:** A

人工智能在过去六十多年当中形成了两条脉络清晰的进路——符号主义(symbolism)和联结主义(connectionism),两条进路贯穿于整个人工智能的发展史,各自都形成了鲜明的特色,表现出不同的理论和应用竞争力。符号主义人工智能是第一代人工智能,它主张人类思维的基本单元是符号,人类认知的过程是符号运算,表现为知识表示和推理,主要通过逻辑进路来研究。联结主义人工智能是第二代人工智能,它源于仿生学,研究上模拟人脑的神经系统,主要聚焦神经网络及其连接机制,由此来探索大数据驱动下的学习方法。

符号主义人工智能擅长研究人类的逻辑思维活动,学习过程是从经验或规则出发,根据前提推出结论的过程,采用从一般到特殊的演绎式推理,是一种“自上而下”的模式。符号主义人工智能的基础是公理化理论和数理逻辑,特点在于逻辑推理过程是可解释和错误可追溯的。与之不同,联结主义人工智能采用的是由特殊到一般的归纳式推理,依赖的是假设检验、相关性分析、线性预测、时间序列分析等数学方法,学习过程是从大数据中发现隐藏的特征和规律,拟合反映特征的正面数据,排除

特征异常的反面数据,这种学习方式是一种“自下而上”的模式。一直以来,科学家都在探索两种路径融合的技术和方法,第三代人工智能的发展思路是把第一代和第二代人工智能结合起来,同时利用知识、数据、算法和算力要素,构造安全、可信赖、可靠和可扩展的 AI 技术。^{[1]1285}

一、符号主义与联结主义 人工智能之比较

符号主义和联结主义之间的路线竞争推动了人工智能的快速发展,形成了“符号派”和“联结派”两个阵营,两个学派之间经常发生碰撞,甚至一度发生过激烈的“对抗”。在当前新一代人工智能背景下,以深度学习为代表的联结派占据了绝对优势,但联结主义人工智能也存在明显的缺陷。基于机器学习的方法存在不同程度的不可解释性问题,因而难以阐明推理的因果关联,从而导致解释能力弱,并且鲁棒性较差。联结主义的方法同样还缺乏坚实的理论基础,即使在仿生学和脑科学理论有所

收稿日期: 2021-10-16

基金项目: 国家重点研发计划“基层社会网格治理数字化关键技术研究与应用示范”(2021YFC3300300); 国家社会科学基金重大项目“新一代人工智能背景下的逻辑学研究”(20&ZD047); 国家社科基金青年项目“智慧司法中法律推理与法律解析的融合路径研究”(21CFX006)。

作者简介: 魏斌(1986—),江西南昌人,哲学博士,浙江大学百人计划研究员,博士生导师,主要研究方向:形式认识论、人工智能逻辑、人工智能与法律。

突破,但仍然难以透彻解释决策的过程。人类的思维过程是综合演绎思维和归纳思维的过程,人工智能模拟人脑的思维必须综合两种思维模式,通过分析两种路径的优缺点,做到优势互补、协同融合。

1. 符号主义人工智能的优缺点

符号主义人工智能的早期探索主要是逻辑推理的研究,即如何使用逻辑符号系统来表达逻辑思维的过程,因而又被称为“逻辑主义”。在符号主义看来,计算机是一个可解释的自动化形式系统,或者说是一个符号操纵的机器。这种定义包含两层含义:一是形式系统的参与主体是自动化的,它们是根据规则来自动化执行的;二是符号是可以被解释的,这使得其行为是有效且合理的。⁽²⁾¹⁰⁶符号主义的特点在于推理的过程是可解释的,如果推理过程出现错误,那么是可以追溯和定位的。符号主义还研究知识的表示、推理和应用,因为符号化方法对于知识的显示性和表达能力强,能够应对较复杂的知识推理,被广泛地运用于自动定理证明和专家系统的构建。

然而,符号主义人工智能也面临困境,首先,符号系统是不完备的,希尔伯特纲领(Hilbert programme)提出构造一个满足完备性、协调性和可判定性的数学系统,哥德尔不完全性定理(Gödel's incomplete theorems)证明了对于任何一个公理化系统,都存在此公理体系所无法判定真伪的命题。这表明符号系统构建的逻辑形式系统不可能是无所不包的,它无法保证在完备的形式系统中推出所有为真的结论,即不可能同时满足“完备性”和“无矛盾性”两个特征。因而,不存在一个无所不能的符号化人工智能系统,这也显示了符号化方法的局限性。

符号主义人工智能的第二个局限是知识获取瓶颈。知识引导的人工智能之所以能展示知识推理等智能化行为,是因为它运用了常识或专家知识。人工智能根据专家知识来执行专家的任务,其推理引擎高度依赖专家知识获取,但这些知识库是封闭的,并且需要由领域专家来构建专家知识,因而,符号主义人工智能无法自主获取知识,这也是专家系统难以维持的主要原因。除了专家知识,人工智能同样难以理解常识,甚至包含一些儿童都能辨别的常识,这是由于人类学习常识往往是通过生活实践来学习的,而机器难以从实践中去学习知识。

符号主义人工智能的第三个问题是框架问题(frame problem),该问题最初是由麦卡锡(J. McCarthy)和哈耶斯(P. J. Hayes)提出的⁽³⁾⁴⁶³,当人工智能试图模拟人类适应环境中相关变化的能力时,就会出现框架问题。通常来说,人类很容易掌握周围发生的事情,这从他们快速预测、反应或调整特定事件的重要后果的能力中可以明显看出。⁽⁴⁾¹⁶¹然而,人工智能却难以用符号来刻画人类周围环境中的知识。换言之,经典人工智能直接可用的显性符号表征由于难以实时、自动地更新,也无法捕捉大量的隐性知识,从而无力表征现实世界,面临框架问题的困扰。⁽⁵⁾³例如,符号主义人工智能就很难用符号去表征图片、语音或视频等感知数据,而这恰是联结主义人工智能所擅长的。

2. 联结主义人工智能的优缺点

由数据驱动的联结主义人工智能以机器学习为主要驱动技术,以深度学习为代表的算法和模型通过组织和拟合参数,学习特征表示,有着良好的泛化能力和可移植性。然而,机器学习运用于大数据分析还存在着若干缺陷:第一,机器学习对大数据的容量和质量有较高的要求,数据的质量直接决定了模型训练的效果,“垃圾数据”输入必然会造成“垃圾结果”输出。第二,数据来源不明确、不完整、可靠性不高、缺乏针对性和准确性,甚至出现数据之间的冲突和矛盾,这些都导致数据的可实用性不高。第三,数据通常要求是满足独立同分布,这在诸多实践应用中难以得到满足。第四,在监督学习当中,数据还需要被精确标注,机器学习模型表现力很大程度上取决于数据标记质量的好坏。

联结主义人工智能还面临难以克服的技术困难。一方面,联结主义人工智能构建“特征工程”困难,如果设置的特征与数据样本真实的特征出入过大,那么机器所习得的特征与人所设置的特征将相差甚远。另一方面,联结主义人工智能所采用的深度学习容易造成过拟合(overfitting)的问题。直观上,过拟合指的是模型学习能力过于强大,将训练数据集的特点看作是所有数据集的属性,从而导致模型的泛化能力下降。随着模型训练复杂度不断增强,模型在验证数据集上的误差会随之增大,导致模型不能很好地拟合训练集之外的其他数据,由此便发生了过拟合问题,导致降低了预测的准确性。

联结主义人工智能的算法还面临不可解释性

问题,尤其是基于多隐层神经网络的深度学习算法会造成“黑箱算法”,算法的不透明成为深度学习饱受诟病的问题。原因在于深度学习高度依赖数学和统计学,机器学习本质上不能解决语义理解的问题。最典型的表现是人工智能本身不能够理解因果关联,因为所学习到的统计频率上的相关性不能够代表因果关联,而表达因果关联恰是符号主义人工智能所擅长的。因而,深度学习虽然有强劲的结果表现能力,但存在的问题仍不足以独立应对所有的学习和决策。此外,从目前深度学习应用的落地场景来看,大量还是采用监督学习的方式,仍然在数据标记等工作上还高度依赖于人的参与,因而,深度学习看似是“机器的学习”,但在实践中并非完全依靠机器,大量的应用仍然停留在“有多少人工就有多少智能”的状态。

综上,符号主义与联结主义人工智能的比较如表1:

表1 符号主义与联结主义人工智能之比较

	符号主义人工智能	联结主义人工智能
数据/知识	知识引导	数据驱动
推理/学习	知识表达、逻辑推理	机器学习、深度学习
优点	算法透明、可解释性	泛化能力、可移植性
缺点	不完备性、知识接收瓶颈、框架问题	不可解释、过拟合、语义理解问题

二、符号主义与联结主义人工智能的融合可能

从模拟人类思维的过程来看,符号主义与联结主义人工智能的融合具备现实可行性。在推理方向上,符号主义人工智能采用的是“演绎式”的推理模式,它是由前提根据推论规则得到结论的模式。反之,联结主义人工智能采用的是“归纳式”的推理模式,它是从已经具有某种特征或规律的数据当中归纳或类比推出相似的新数据也具有某种特征的推理模式。人类的思维往往是综合了演绎思维和归纳思维,推理过程也往往混合两种思维,不存在绝对的分野。

比较而言,联结主义人工智能受到极大的认可和关注,但因此就推崇联结主义而否定符号主义显得过于片面,美国俄亥俄大学的詹德拉(G. Chan-

drasekaran)就指出“最近几年,人工智能在很大程度上集中于统计学和大数据。我同意由于计算能力的大幅度提高,这些技术曾经取得过某些令人印象深刻的成果。但是我们完全有理由相信,虽然这些技术还会继续改进、提高,总有一天这个领域会对它们说再见,并转向更加基本的认知科学研究。尽管摆钟摆回去还需要一段时间,我相信定有必要把统计技术和对认知结构的深刻理解结合起来。”⁽⁶⁾ⁱ⁻ⁱⁱ融合的路径已经成为人工智能领域的共识,也逐渐成为人工智能的研究热点。

符号主义进路和联结主义进路的人工智能没有绝对的优劣,只是两者适用的科学和社会领域不同,它们对应的是人类理性思维活动的不同向度,擅长解决不同类型的问题。有些领域适用符号主义进路的人工智能,它更加侧重算法的透明性和可解释性,对算法黑箱和不可解释性的容忍度较低。有些领域则适用联结主义的人工智能,这类领域往往要求能够满足较强的算法表现力,能够为决策的结果提供直接和有效的建议,但却对算法的可解释性没有强制或过高的要求,这意味着这些领域对风险和错误的容忍程度较高,允许算法存在一定的风险。

符号主义进路和联结主义进路的人工智能都存在局限性,不可能存在一种在任何问题和领域当中都统一适用的人工智能方法,尽管“深度学习+大数据”已经在较大范围的科学和社会活动中发挥了重要作用,但是在其他很多问题当中仍然作用有限。人工智能技术的优劣是相对于具体的问题和领域而言的,就某个问题适用何种人工智能,有的人工智能在可解释性和表现力方面能够达到符合既定目标和要求的某种平衡,那么它就是相对更优的方法,人工智能科学家一直在这样的研究方向上发展更优的人工智能算法。

两种路径融合的优势在于两者能够优势互补,符号主义的可解释性可以弥补联结主义的算法可解释性和推理能力弱等问题,联结主义可以避免符号主义的知识接收瓶颈困境等。融合的观点越来越受到人工智能科学家的认同,融合神经网络和符号过程的混合系统(hybrid system)、“神经-符号机”等理论相继被提出,常见的一阶逻辑、多值逻辑、模态逻辑都被尝试嵌入到神经网络当中。随着新一代人工智能的发展,符号主义与联结主义人工智能又各自取得了新进展,使得两种路径的融合又出现了新的生命力。

三、新一代人工智能背景下的融合路径

符号主义与联结主义人工智能的融合产生了两种融合的思路,第一种是非对称的单向融合模型,这又包含两种方法,其一是将符号主义方法融入到联结主义人工智能,利用逻辑符号来表达神经网络等研究;其二是将联结主义方法融入到符号主义人工智能,利用机器学习的特点加深因果关系及其推理等研究。第二种是双向模型,即结合双方的特征产生新的应用场景,如认知图谱及其推理的研究。神经网络能够与逻辑结合的原因在于,它能够构建神经网络中不同隐层之间的关联结构与逻辑推理的规则结构之间的映射。换言之,神经网络中原本难以解释神经网络之间的关联性,可以用可解释的逻辑规则来加以说明。张钹院士还提出了第三代AI的三空间融合模型,即融合双空间和单空间两种模型,其中,单空间模型是以深度学习为基础,将所有的处理都放在亚符号(向量)空间,双空间模型模仿了大脑的工作机制,但对其工作机制知之甚少。⁽¹⁾¹²⁹⁰

1. 基于神经网络的逻辑推理

在符号主义与联结主义融合的早期,科学家尝试借助神经网络的优点来提升逻辑推理能力。明斯基(M. Minsky)总结到:符号知识和联结主义各有优缺点,我们需要一个系统能够将它们的优点集成起来。⁽⁷⁾³⁴自1980年以来,人工智能科学家就通过各种方法将神经网络和符号学相结合,形成了神经符号主义(neural-symbolism)。平卡斯(G. Pinkas)就提出了一个能够表示和学习命题知识的联结主义框架,开发了命题演算的扩展版本,并且证明了可用于非单调推理、处理相互冲突的信念和知识。新框架能够实现快速学习,扩展逻辑可用作联结主义网络的高级规范语言,能够映射主要的符号系统。⁽⁸⁾²⁰³

20世纪90年代初期,托维尔(G. Towell)和沙弗里克(J. Shavlik)给出了一种基于知识的人工神经网络(KBANN)这种模型研究如何从神经网络中插入和提取规则,为使用反向传播训练的基于知识的神经网络学习算法提供了一种学习背景知识的有效方法。⁽⁹⁾¹¹⁹人工神经网络后来还启发了联结主

义归纳学习和逻辑程序(CILP),这种模型使用逻辑程序作为知识表示语言,在神经网络中进行知识精炼,从受过训练的神经网络中提取知识,并从例子中归纳得到新的知识。加尔塞斯(A. Garcez)等提出了一种在已训练后的神经网络中抽取逻辑知识的方法,帮助增强神经网络的可解释性。⁽¹⁰⁾¹⁵⁵理查德森(M. Richardson)等以一阶符号逻辑和概率图模型结合的方式进行了探索,提出了马尔科夫逻辑网络⁽¹¹⁾¹⁰⁷,该网络是一阶符号逻辑的知识库,其中每一个公式都有对应的权重。受限于当时机器学习技术和自然语言处理技术的制约,这些探索并不能充分利用神经网络的优势,因而没有取得更进一步的成功。未来的研究将推动以深度学习为代表的神经网络在模态逻辑、直觉主义逻辑等更多逻辑分支中发挥作用。

2. 基于符号表示的机器学习

联结主义人工智能所采用的深度学习算法不可解释的原因在于,神经网络包含了多层难以被表达的隐藏层,导致无法解释输出是如何由输入所得到的。解决这个问题的一种思路是利用逻辑符号来建立神经网络中隐藏层之间的映射关系。早在1990年,深度学习技术的代表人物辛顿(G. E. Hinton)就提出了将一种“部分-整体”层次结构映射到联结主义网络的方法,用符号来表达整体结构中的最小单元,并使用符号来解释数据结构的类型,实现局部和整体的一致性表示。⁽¹²⁾⁴⁷BoltzCONS是一种表达动态符号结构的联结主义模型,能够动态创建和操作复合符号结构。这些结构是使用链表的功能模拟来实现的,但BoltzCONS使用分布式表示和关联检索来代替传统的内存组织。⁽¹³⁾⁵

符号逻辑的作用是帮助神经网络训练、学习和推理,通过将符号规则转化为神经网络可以理解的向量。有学者提出了一种知识蒸馏的师生(teacher-student)网络,使神经网络拟合符号规则的结果,在规则的指导下训练学习。有研究将深度神经网络与结构化逻辑规则相结合,比较直接的方法是将一阶逻辑规则用于增强各种类型的神经网络(CNN和RNN等),思路是将逻辑规则的结构化信息转换为神经网络的权重。⁽¹⁴⁾²⁴¹⁰联结主义和符号主义结合的困难在于,深度神经网络中隐节点上发生的事情是不可言说的,因为隐节点可能并不表达我们使用的任何概念或概念组合,可能只有把认知过程分解成远比我们的概念体系细得多的碎片,再按另一种方式重新组合才能

得到一点语义的蛛丝马迹。⁽¹⁵⁾⁴⁹

第三代人工智能的研究动向之一是可解释的人工智能,其中一个方向是通过因果关系来解释推理的过程。在最新的研究中,“图灵奖”得主本吉奥(Y. Bengio)研究了因果表示的机器学习,他将因果推理与机器学习迁移与泛化联系起来,进而分析因果推理可能对机器的价值、鲁棒性和泛化性,以及现有的共性比如半监督学习、自我监督学习、数据增强和预训练。⁽¹⁶⁾⁶¹²周志华结合机器学习和逻辑编程实现了溯因学习,提出了一种新的方法来联合优化机器学习模型和逻辑推理模型,其中逻辑推理利用符号域知识和纠正错误感知的事实来改进机器学习模型,通过溯因学习使机器同时学习识别手写数字和求解数学运算。⁽¹⁷⁾²²⁰⁻²²²机器学习借助符号主义路径的思路是利用逻辑推理的简洁性和可解释性来弥补机器的不可解释性,但如何充分地使逻辑与机器学习算法相融合,还需要深度挖掘逻辑的推理机制与机器的预测功能相耦合的新路径。

3. 认知图谱与推理研究

知识工程在语义网络之后进入了知识图谱的时代,谷歌于2012年正式提出了知识图谱这一新概念,其最初的目的是为了提高搜索引擎检索的效率,通过知识图谱可以将相关联的信息共同提供给使用者。谷歌的知识图谱不仅基于Freebase、维基百科等知识图谱,而且还致力于全面和深入的知识图谱建构。该知识图谱包含超过5亿个对象,超过35亿个这个对象之间的关系。^①除了构建和应用知识图谱的技术之外,知识图谱研究还包含融合认知计算、自然语言处理、语义网络技术和机器学习等人工智能技术。近些年,随着知识图谱推理和图神经网络技术的应用,知识图谱研究开始走向认知图谱。

认知图谱是以实现知识引导和数据驱动相结合的知识表示和推理的认知引擎为目标,研究支持鲁棒可解释人工智能的大规模知识的表示、获取、推理与计算的基础理论和方法。⁽¹⁸⁾¹⁹现有的认知图谱的研究方法主要是基于知识图谱结合逻辑推理的方法来实现认知层面上的图谱推理。在知识图谱推理方面,主要是基于逻辑推理的方法,常见的

方法包括基于一阶逻辑、模态逻辑、描述逻辑、分布式表示推理、神经网络推理等方法。另一方面,在知识图谱的逻辑表达方面,主要的研究方法是基于自然语言生成(NLG),常见的模型包括Transformer模型和Bert模型等。然而,神经网络模型难以解释的问题仍然存在,还需要从认知层面来寻找解释神经网络推理过程的新方案。

认知图谱研究不能只停留在“知识图谱+逻辑推理”的层面,而应当深入发掘知识图谱和认知科学的结合。这需要交叉可解释性理论、反事实因果推理、持续学习等方法,建立高效且稳定的知识获取、表达和推理机制,让知识能够被机器理解和运用,进而进行可解释的推理。最新的研究动态指向多种路径:一是知识图谱与图神经网络的结合,利用循环神经网络(RNN)和注意力机制(attention)等来构建知识图谱。二是大规模知识图谱和知识计算研究,这需要从海量、多源、异构、碎片化的大数据当中大规模自动化地获取知识,并进行动态融合。三是从人机知识交互的角度,研究人机交互的知识表征体系,实现高效的智能交互。四是从高阶认知的视角来研究,这需要从形式认识论、认知心理学、脑机融合等学科进行交叉性探索。

4. 基于机器学习的论辩研究

符号主义人工智能的最新研究进展是可计算论辩模型与机器学习的结合,实现了自动化辩论。2018年IBM的人工智能辩论机器人(Project Debater)在与人类专业辩手对局的国际辩论比赛中表现出色。2020年《自然》杂志发表了人工智能论辩机器人的工作原理,辩论机器人结合了计算辩论理论和论辩的数据挖掘等多种技术,使得机器人具备了从海量文本数据中获取知识来构建论证以支持论点,反驳对方论点等多种论辩能力。辩论机器人由四个主要模块组成:(1)论证挖掘,机器通过处理LexisNexis语料库中的大约4亿篇报纸文章来检索与论点相关的主张和证据。(2)论证知识库(AKB),机器获取不同辩论之间的共性,通过特征的分离器来判定哪些与论点相关,从而获取与论点相关的知识。(3)论证反驳,机器通过论证挖掘来编译可能被对手提及的线索,AKB同样也可以挖掘反驳的论点。(4)论证构建,机器通过聚类分析构建一个基

^① Singhal A. Introducing the Knowledge Graph: things, not strings [EB/OL]. [2012-05-19] [2021-8-31]. <https://www.blog.google/products/search/introducing-knowledge-graph-things-not/>.

于规则的论辩系统,该系统根据与论证的语义相似性来聚类论证,获取一组能够反映聚类主体的高质量论证集,以供系统选择论辩的素材。⁽¹⁹⁾³⁸⁰⁻³⁸¹

基于机器学习的方法,人工智能已经实现了论辩挖掘和论辩文本的自动化生成。论辩挖掘是在海量文本大数据当中根据机器学习算法来预测文本中所包含的不同类型的论证。首先,论辩挖掘需要构建特征工程,这要先定义完全的论证类型,明确论证的组成要素。其次,论辩要根据这些要素组成不同论证类型的方式,采用有监督学习的方式进行人工标注,构建相对应的特征工程。第三,再根据特征工程,采用深度学习、朴素贝叶斯分类器、统计学习等算法来训练模型,通过不断调整模型参数,使模型获得良好的泛化能力。第四,采用模型检验论辩挖掘的准确率,对目标文本进行预测,准确挖掘文本中所包含的论证,而后再采用论辩语义的高效求解方法来进行评估。

尽管论辩研究为两种路径的融合提供了巨大的推动力,它不仅提供了从海量文本大数据中挖掘和生成论辩文本的技术,还推动了论辩本体论的研究和语料库的开发。然而,基于机器学习的论辩研究还有待进一步完善,其一,最突出的问题在于人工智能挖掘和生成论辩并不能像人类一样理解上下文的含义,也就无法理解论证文本的语义,生成的文本也无法保证论点在语义上是有意义的,甚至论证当中的简单因果关系都难以识别。其二,论辩研究目前只限于辩论赛式的说服型对话,没有涵盖日常生活中其他类型的辩论,比如信息寻求、科学探索、谈判等对话类型。其三,论辩研究难以获得人类在辩论中的修辞技巧,这往往和语用情境相关,人类理解观点是需要具体的论辩情境中去理解其含义,脱离情境极可能导致含混。符号主义人工智能的新趋势之一是研究非形式逻辑和新修辞学的形式化⁽²⁰⁾融合路径下的论辩研究蕴含着无穷的潜力。

四、结语

人工智能是在不断向人类学习的过程中发展起来的,符号主义人工智能的思维对应于人类的演

绎式思维,而联结主义人工智能的思维则对应于人类的归纳式思维。演绎式思维注重从人类的经验当中提炼和获取知识,归纳式思维则倾向于从个案当中总结和归纳知识,人类的思维活动往往是融合两种思维模式。本吉奥就指出人工智能研究应当从直觉的、快速的、无意识的、非语言的系统(system 1)转向有意识的、逻辑性的、有规划的、推理性以及可以用语言表达的认知系统(system 2)。^①符号主义和联结主义路径的融合不能只停留在以一方的技术优势来弥补另一方的缺陷,而是应当真正理解对方的技术逻辑,找到最契合的融合点,探索新的理论和技术增长点,创造具备协作(collaborative)能力、自适应(adaptive)能力、负责任和可解释的人工智能。

参考文献

- (1) 张钹,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) Haugeland J. *AI: The Very Idea* [M]. MIT Press, 1985.
- (3) McCarthy J, Hayes P J. Some Philosophical Problems from the Standpoint of Artificial Intelligence [C]//Meltzer B, Michie D. (eds.) *Machine Intelligence*. Edinburgh: Edinburgh University Press, 1969: 463-502.
- (4) Haselager W, Rappard J. Connectionism, Systematicity and the Frame Problem [J]. *Minds and Machines*, 1998, 8(2): 161-179.
- (5) 夏永红, 李建业. 人工智能的框架问题及其解决策略[J]. 自然辩证法研究, 2018, 34(5): 3-9.
- (6) 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- (7) Minsky M. Logical versus Analogical or Symbolic versus Connectionist or Neat versus Scruffy [J]. *AI magazine*, 1991, 12(2): 34-51.
- (8) Pinkas G. Reasoning, Nonmonotonicity and Learning in Connectionist Networks that Capture Propositional Knowledge [J]. *Artificial Intelligence*, 1995, 77(2): 203-247.
- (9) Towell G, Shavlik J. Knowledge-based Artificial Neural Networks [J]. *Artificial Intelligence*, 1994, 70(1): 119-165.
- (10) Garcez A, Krysiak B, Gabbay D. Symbolic Knowledge Extraction from Trained Neural Networks: A Sound Approach [J]. *Artificial Intelligence*, 2001, 125(1-2): 155-207.
- (11) Richardson M, Domingos P. Markov Logic Networks [J]. *Machine Learning*, 2006, 62(1-2): 107-136.
- (12) Hinton G. Mapping Part-whole Hierarchies into Connectionist Networks [J]. *Artificial Intelligence*, 1990, 46(1-2): 47-75.

^① Bengio Y. From System 1 Deep Learning to System 2 Deep Learning (2019) [EB/OL]. <https://nips.cc/Conferences/2019/Schedule?showEvent=15488>.

- (13) Touretzky D. BoltzCONS: Dynamic Symbol Structures in a Connectionist Network [J]. *Artificial Intelligence*, 1990, 46(1-2): 5-46.
- (14) Hu Z, Ma X, Liu Z, et al. Harnessing Deep Neural Networks with Logic Rules [C] // *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2016.
- (15) 李国杰. 有关人工智能的若干认识问题 [J]. 中国计算机学会通讯, 2021, 15(7): 44-50.
- (16) Schlkopf B, Locatello F, Bauer S, et al. Towards Causal Representation Learning [J]. *Proceedings of the IEEE*, 2021, 109(5): 612-634.
- (17) Zhi - Hua Z. Abductive Learning: Towards Bridging Machine Learning and Logical Reasoning [J]. *Science China (Information Sciences)* 2019, 62(7): 220-222.
- (18) 唐杰. 认知图谱——人工智能的下一个瑰宝 [J]. 中国计算机学会通讯, 2020, 16(8): 8-10.
- (19) Slonim N, Bilul Y, Alzate C, et al. An Autonomous Debating System [J]. *Nature*, 2021, 591: 379-385.
- (20) Walton D, Gordon T. Formalizing Informal Logic [J]. *Informal Logic*, 2015, 35(4): 508-538.

Analysis of the Integration Path of Symbolism and Connectionism of Artificial Intelligence

WEI Bin

(Institute of Digital Jurisprudence, Zhejiang University, Hangzhou 310008, China)

Abstract: The evolution of AI has formed two representative approaches of symbolism and connectionism of AI. The advantage of symbolic AI is that the reasoning process is transparent and interpretable, but there are problems such as incompleteness, framework problems, and knowledge reception bottlenecks. Connectionist AI has strong abilities of generalization and transportability, but its drawbacks are the inexplicability and overfitting of the algorithm. Symbolic and connectionist AI have significant integration possibilities. Under the background of the new generation of AI, the integration of the two approaches has produced new paths, which have shown strong integrating potential in neural networks with logical reasoning, machine learning based on the symbolic representation, cognitive graph and its reasoning, machine learning and argumentation mining, and other cross-fields.

Key words: symbolism; connectionism; Artificial Intelligence; integration path

(本文责任编辑: 董春雨 赵月刚)