人工智能拥有概念吗?

李 珮

(清华大学 人文学院科学史系,北京 100084)

摘 要: 概念学习原本是认知科学领域研究人获得概念的方法,但是现在已经成为人工智能进行自我学习的新领域。文章梳理了有关人工智能概念学习的内容和方法,并由此引发对什么是概念,人工智能是否拥有概念这些问题的思考。对于什么是概念主要有两种范式的解读,这两种范式分别是: 认知科学、某种哲学。该研究发现人工智能在认知科学范式下拥有概念,但在某种哲学范式下不拥有概念,并启发了以下问题: 我们如何看待认知科学与某种哲学之间的巨大鸿沟。

关键词: 概念; 人工智能; 认知科学; 哲学

中图分类号: N02 **文献标识码**: A **文章编号**: 1674 - 7062(2020) 05 - 0009 - 06

"概念"是一个常用的词,但对于什么是"概念"人们却有不同的看法。有从认知科学领域研究"概念"的,也有从哲学领域研究"概念"的。近年来,人工智能领域的人们也开始研究与"概念"相关的"概念学习"。人工智能中的"概念学习"是什么? 人工智能真的能拥有人类所拥有的"概念"吗? 围绕这些问题,本文将分别对"人工智能的概念学习"和"概念"进行分析,从而初步回答"人工智能拥有概念吗"这个问题。

本文在宽泛的意义上使用人工智能这个词语,即,人类制造的智能,包括当前热门的机器学习、深度学习等领域,以区别于人自身所具有的智能。

一 人工智能领域的概念学习

机器为什么需要概念理解以及创造概念? 这涉及人工智能与人类之间极其复杂的关系,由于人工智能机器在执行各种"类人"任务时,需要理解认知及其道德评价的一系列概念,特别是为了成为一个人工的道德行动者,并能够在特定情况下决定道德正确的行为时,人工智能需要对与道德相关的概念有一定的理解[1]。卡伊·苏塔拉(Kaj Sotala) 在其论文"安全自主人工智能的概念学习"中,就试图将"权利""幸福"等道德概念植入自主人工智能中。其策略是研究人类认知概念的机制,进而模仿人类的认知概念的机制并将其应用到人工智能领域中。

> "我们的假设是,人类大脑利用一套相对有限的规则和机制来产生自己的概念,我们在逆向工程方面正取得良好进展。"[2]

苏塔拉介绍了有关人类概念研究的不同理论。例如,一种理论从语法的角度研究概念。根据这种理解,我们可以把学习"有效英语句子"这个概念的过程等同于学习语法。有的理论则是建立在表征的一般理论基础上的,如加登福斯(Gardenfors) 认为概念可以被表征为多维空间中的几何结构。总之,通过将概念进行形式化和结构化,人们可以找到概念学习的机制,并将其应用于人工智能之中。上述方法是人工智能学习概念的可行框架,但苏塔拉在其文章中没有具体说明如何学习概念的技术细节。

布伦登姆·雷克(Brenden M. Lake) 在其论文"概率程序归纳法的人类层次上的概念学习"中提出了贝叶斯程序学习。在布伦登姆看来,机器学习和人类学习最大的区别是机器学习概念需要大量的数据和例子,而人类学习概念只需要很少的数据或

【收稿日期】 2020 - 06 - 13

【作者简介】 李 珮(1990 -) ,女,山东莘县人,清华大学人文学院博士研究生,研究方向为科技哲学、心灵哲学。

仅需一个例子。此外，人类可以通过概念学习产生额外的学习功能，例如使用现有的类别创建新模板和新类别，这是机器学习无法做到的。所以布伦登姆想要模仿人类的学习经验并将其应用于人工智能之中。其要解决的核心问题是如何从稀疏的数据中总结出丰富的概念，生成丰富的表征。它的贝叶斯程序学习（BPL: Bayesian program learning，以下简称BPL）框架允许机器从一个例子中学习大量的视觉概念，并以与人类难以区分的方式对其进行概括。

在BPL中，概念被表征为简单的概率性程序——即，用抽象描述语言将概率性生成模型表示为结构化过程，以此来最好的解释被观察到的例子。这种学习程序是如何实施的呢？

> “BPL定义了一个生成模型，这种生成模型可以通过以新的组合部件和子部件的方式来对新类型的概念进行采样。每一种新类型也被表征为一个生成模型，而这个较低层次的生成模型产生了概念的新示例（或标记），使得BPL成为生成模型的生成模型。最后一步以原始数据的格式呈现标记级变量。”[3]1333

这种贝叶斯学习框架汇集了三种关键的理念——合成性（compositionality）、因果性（causality）和学会学习（learning to learn），这三个关键的理念曾分别对认知科学和机器学习都产生了很大影响。这三种理念是如何体现在BPL中的呢？

> “在BPL中丰富的概念可以从更为简单的原语（primitives）中‘合成’构建。其概率语义（probabilistic semantics）以一种过程形式来处理噪音并支持创造性的概括（creative generalizations），从而自然地捕获了产生类别（category）示例的现实过程的抽象‘因果’结构。模型‘学会学习’则是通过开发分层先验（hierarchical priors）来实现的，这些先验代表了一个学习的归纳偏差，它抽象化了既定概念中变化的关键规律和维度，这种先验允许使用以前的相关概念经验来简化对新概念的学习。简而言之，BPL可以通过重新使用现有程序的片段，通过捕获在多个维度上运行的真实世界的生成过程的因果关系和合成属性来构造新程序。”[3]1333

布伦登姆使用现有的有关手写字符的数据库开发出一个程序，这个程序可以做到只看一眼就能书写。这个程序通过了视觉图灵测试，使得人们难以区分什么是机器写的字符什么是人写的字符。这种尝试使得人工智能能够学习有关书写字符的概念。

通过以上有关人工智能进行概念学习的例子可以看出概念有多种用法，概念可以是苏塔拉研究中的一种道德概念、一种语法或者一种多维空间中的几何结构，也可以是布伦登姆研究中的手写符号或者概率性程序。二者共同的工作前提是要将概念表征成什么，而这个什么本身却是各不相同的。概念学习是人与机器进行自我学习的重要方法，但我们在多种维度上使用概念这个词使得概念本身有些扑朔迷离。

二　认知科学意义上的概念

上文谈及了人工智能的概念学习，但是对什么是概念还不清楚，亦不清楚人工智能是否真的拥有类人的概念。对这两个问题的回答涉及如何理解概念。在日常生活中，我们通常从常识的角度去理解概念。从常识的角度看，“概念是一个名称或标签，它将抽象看成好像拥有具体的、物质的实存物，如一个人，一个地方，或一个东西……自由、平等、科学、幸福等抽象的观点和知识领域也象征着概念”[4]。

简而言之，常识的意义上概念只是一种符号，一种抽象的表征。在这里有实指的名称、标签以及没有实指的抽象观点都被看作是符号和表征。任何表征或符号都可以称为概念，这是对概念最宽泛的理解。由于人工智能离不开这种意义上的符号或表征，基于这种对于概念的常识性的理解，人工智能是有概念的。

在学术研究中我们则主要从认知科学、哲学这两种范式进行概念研究。在认知科学中主要存在三种概念，即原型（prototypes）、范例（exemplars）和理论（theories）[5]76-77。在爱德华·马歇瑞（Edouard Machery）看来，这三种概念是三种实体，它们毫无共同之处，是三个截然不同的概念。这三个实体已被证明存在于大脑中，并经常用于不同的认知过程。本文将分别解释这三个概念，然后说明人工智能是否拥有这三种概念。

原型理论认为概念是原型，原型是“关于本类别成员所拥有的属性的统计知识体系”[5]83-84。概念的原型理论取决于它们如何描述原型中存储的统计知识的性质。该原型不仅可以表示类别的典型属性（the typical properties of categories），而且也可以表示类别的线索有效属性（cue - valid properties of

categories)。例如,四条腿是狗这种类别的典型属性,吠叫是狗这种类别的线索有效属性。根据原型理论,人们可以通过检查原型和事物之间的相似性来判断一个事物是否属于一个类别。原型体现的是一物体最典型的具有统计意义上的抽象特征,以“苹果”为例,苹果的原型就是从一组苹果中抽象出来的共同的某些特征,比如,不规则圆形,红色、绿色或黄色,具有什么程度的硬度和重量等等。人们可以通过查看一物体是否具有以上特征而决定该物体是否是苹果。

范例理论认为概念就是范例,范例“是关于一个类的特定成员所拥有的属性的知识体系”[5]93。与原型相似,人们可以通过查看范例和事物之间的相似性来判断一个事物是否属于一个类别。范例和原型的区别在于,范例是从记忆中提取出来的类别的实际成员,而原型是类别成员的抽象平均值。范例体现的是一物体或多个物体的典型的实际样例表征,其存储的是某些对象或某个对象的真实样例。以“苹果”为例,苹果的样例就是人们今天看到的苹果或者昨天吃掉的苹果,苹果的样例体现了人们在经验苹果时所知觉到的有关苹果的具体信息。因此在范例理论看来,有关苹果的概念不是那些具有“不规则圆形,红色、绿色或黄色,具有什么程度的硬度和重量等等”的抽象原型特征,而是人们所经验过的所有有关苹果的例子。原型理论与范例理论各有优缺点,人们可以使用原型进行快速判断,但是当人们遇到非典型类别成员时,使用范例进行判断则要容易得多。

理论说认为概念就是理论,该理论认为“概念之间的关系就像科学理论的术语一样,分类是一个与科学论证非常相似的过程”[6]。与原型和范例不同,一个概念在理论中的内容是由它在理论中的使用决定的,而不是仅仅由它的组成部分决定的。理论代表了知识的背景,它包含着隐藏的本质、因果规律和功能[7]。与原型和范例相比,理论更有可能表征道德概念,因为它包含更多的背景知识。

总之,原型、范例和理论是三种完全不同的异质性概念。原型概念存储统计知识,涉及基于线性相似度计算的认知过程。范例概念存储关于特定个体属性的知识,并涉及基于非线性相似性计算的认知过程。理论概念储存有关因果、规律和功能的知识,类似于最优推理或因果推理[8]。虽然这三种概念是完全不同的异质性概念,但其并不是相互排斥的,它们可以成功的解释不同的现象;虽然这三种概念代表了不同的认知过程,但它们都有一个共同点:它们都有各自的认知机制,这种机制是可以用数学、逻辑或抽象知识所表征的。

布伦登姆所研究的概念学习主要涉及原型概念与范例概念,BPL 的 3 个关键性成分合成性、因果性以及学会学习可以体现这一点。合成性是使用范例的过程,在合成性中该程序将一个特定的符号范例分解成部分笔画;因果性则在抽象层次上捕获了产生类别实例的真实因果过程的各个方面,该因果性构成了产生新例子的概率性运动程序,BPL 的因果过程是一个构成原型概念的过程;而学会学习则是将之前获得的有关概念的参数、约束条件应用到学习新概念的任务中,让该程序可以自己生成新的手写符号,这个过程既涉及范例的先验知识,又涉及抽象化的原型规律。但是基于 BPL 的这种只看一眼就能学会的程序只能学习最简单的类似于手写符号的表征类型,还不能学习类似于结构描述、语法等复杂的表征类型,因此该程序还不涉及理论概念。苏塔拉所研究的有关安全自主人工智能的伦理概念是更为复杂的概念,这种伦理概念涉及更多的背景知识,更需要理论概念的参与。总而言之,由于人工智能是要模拟人类的各种认知机制,而原型、范例、理论这三种概念恰恰代表了人类的认知机制,所以人工智能在认知科学范式中是可以拥有概念的。

人工智能学习的概念与认知科学所研究的概念之间的渊源可以追溯到人工智能早期的发展阶段。1950 年图灵所写的有关图灵测试的论文拉开了符号人工智能发展的序幕,图灵认为人工智能的工作就是要模拟以生理为基础的心智所发生的过程,并做心智才能完成的事情。从其模拟心智的目标来看人工智能的发展就与认知科学密切关联。图灵的这一信念在精神病学家沃伦·麦卡洛克(Warren McCulloch)与数学家沃尔特·皮兹(Walter Pitts)合作的“神经活动中内在思想的逻辑演算”论文中得到了支持,这篇论文将图灵的观点与命题逻辑和神经突触理论结合在了一起[9]。因此早期符号人工智能的发展本身就是与认知科学紧密相连的,二者共同建基于二进制的逻辑运算之上。当前的人工智能已经从符号人工智能过渡到联结主义人工智能,但联结主义也是认知科学的成果,其描述了作为神经网络的大脑是如何工作的。苏塔拉以及布伦登姆的上述研究就是一种基于联结主义的工作,这种联结主义用概率代替了逻辑,用自下而上的处理代替了自上而下的控制。总之人工智能的概念学习与认知

科学密不可分，二者的工作都是在相似的范式下进行的，因此从认知科学的意义上看，人工智能是拥有概念的。

三　某种哲学意义上的概念

从某种哲学的角度看，“概念”则具有先天或先验的色彩。本文将以三位哲学家——杰瑞·福多(Jerry A. Fodor)，黑格尔(Georg Wilhelm Friedrich Hegel)，约翰·麦克道威尔(John H. McDowell)论述概念的思想为例来阐述什么是概念，人工智能是否拥有概念。

福多认为几乎我们所有的单词都是先天的，如萝卜或伞这种概念都是对应于自然语言的单词。“其概念先天论的主要观点为：词汇概念(lexical concepts)是原初的(primitive)——它们缺乏结构——而原初概念是无法学习的”[10]。

福多的概念先天论是与概念经验论相对的，概念经验论是认知科学所采用的一种范式，这种范式认为词汇概念本身拥有内在结构，这种概念可以通过类似归纳式的概念学习所获得，因此福多的概念先天论反对认知科学的这种研究概念的范式。概念先天论与概念经验论主要对立的地方在于词汇概念是否具有内在结构这个问题。在论述二者之间的分歧之前有必要澄清其共同点，福多的概念先天论和其所反对的概念经验论都认可概念可以分为以下几种：原初概念、词汇概念、复杂概念或短语概念。二者都认可原初概念是内在的非结构性的概念，原初概念是通过刺激物刺激知觉器官而被获得的，也都认可复杂概念或短语概念由词汇概念组成。其不同之处在于，在概念经验论看来，所有的概念都是由原初概念通过组合器官构成。原初概念组成词汇概念，并且词汇概念的归纳的组合式的构成方式使得词汇概念本身具有内在结构。而概念先天论则认为词汇概念与原初概念一样，都是由刺激物刺激知觉器官而获得，其本身并没有内在结构[11]。从概念先天论的观点出发，既然词汇概念并没有内在结构，那么其本身就不能被学习，因此从这点看人工智能是不能拥有概念的。

除此之外，福多的概念先天论由信息原子论(Informational Atomism)所支持，信息原子论包含两部分内容，一部分是概念原子论(Conceptual atomism)，另一部分是信息语义学(Informational semantics)。在概念原子论中，大部分的词汇概念是没有内在结构的，这也就支持了福多的上述思想。在信息语义学中，概念内容是由某种惯常的(nomic)、具有心灵－世界的关系所构成的，因此，拥有一个概念至少部分的由处于某种惯常的、心灵－世界的关系所构成[12]。这种惯常的、心灵与世界之间的关系也就是命题态度。进一步而言“人们有信仰、欲望、观点、愿望——这在哲学中被称为命题态度”[5]32。最终在福多看来，拥有x的概念就是能够将x理解为x的命题态度。既然只有人类才具有信仰、欲望和愿望，只有人类具有命题态度，那么概念也只属于人所拥有。总而言之从福多的思想看，人工智能是没有概念的，无论是布伦登姆的贝叶斯学习程序，还是苏塔拉将道德概念植入人工智能的尝试，都不能使人工智能拥有命题态度。此外，由于我们的词汇概念是先天的，是没有内在结构的，是不能被学习的，因此即便人工智能发展的再强大也无法拥有概念。

在黑格尔看来概念具有两个层次。一个层次是“确定概念”(determinate concepts)，即“基层经验和实践概念”(ground－level empirical and practical concepts)，其主要用于描述和解释经验的活动。另一个层次是“元概念”，即康德所说的“范畴”，其工作是用来表达有关“基层经验和实践概念”的内容及对其使用的关键特征，其对阐明框架特征使框架具有描述和解释的可能性发挥着独特的表达作用。“元概念”是先验的，独立于基层概念的任何特定使用。除此之外元概念是自我意识的表达器官，是自我意识的前提[13]。从对概念的两个层次的划分来看，由于“元概念”是先验的，是人类所独有的，因此人工智能学习不到“元概念”；而“基层经验和实践概念”是与经验相关的，似乎是人工智能可以学习的概念，那么人工智能真的可以拥有“基层经验和实践概念”吗？

从罗伯特·布兰顿(Robert Brandom)对黑格尔的“基层经验和实践概念”解读来看，概念是被社会实践建构出来的，是具有规范性的东西。在布兰顿看来概念是用于判断的规范，“当我们判断“这辆汽车是红色的”时，我们使用了“汽车”和“红色”两个概念。然而，这些概念不能被认为是物体的抽象图像或物体的属性。它们构成了规范，规定了什么是“汽车”或什么是“红色”。也就是说，这些规范决定了什么东西应该被理解为什么”[14]139－140。按照这个思路，只要社会规定好了什么是汽车那么就可以将这种规定的内容表达出来，进而机器就可以学习这种规定好的内容。但是概念内容能够被完全确定下来吗？答案是否定的，因为一个特定概念的内容

并不是简单的所与的东西。它包括那些使用它的人所采取的推理联系(inferential connections) 。但是不同的人对同一概念所采用的推理联系是不同的,人与人之间的对话是一直在进行的,因此概念内容永远不会最终确定,而只能通过谈判的过程得到暂时的妥协。因此从“基层经验和实践概念”的视角出发,概念的内容是在发生变化的,人工智能只能部分的学习概念的内容,而不可能拥有一个完整的固定的概念。

总而言之,按照黑格尔对概念的两个层次的划分,人工智能只能部分的学习“基层经验和实践概念”,而不能拥有完整的“基层经验和实践概念”,更不能学习或拥有“元概念”。但是,由于“元概念”是“基层经验和实践概念”的基础,因此,仅仅学习“基层经验和实践概念”而不学习元概念就不是学习真正的概念。正如黑格尔所认为的那样,经验主义的概念性规范或许在很大程度上是社会性的成就,但我们所服从的或者应该服从的基本范畴规范却植根于思维本身的本质之中,它们没有任何超越或超自然的理由,但它们也不只是社会和历史协商的产物[14]149。正是在这个意义上,人工智能是不可能拥有类人的概念的。

麦克道威尔是持有概念论观点的代表人物,在其著作《心灵与世界》这本书中他并没有给概念下一个定义,但我们还是可以通过其文字描述去理解概念。在书中麦克道威尔写道“我们不应当将康德称作直观的东西——经验接纳——理解为一个概念之外的所予的赤裸的获得(a bare getting of an extra - conceptual Given) ,而是应当将其理解为一种已然具有了概念内容的发生过程或状态。”[15] 从这段文字中我们可以看出我们的经验本身已经具有了概念内容,我们的经验内容本身已经具有了概念性。单从这点出发,我们可以说既然人工智能也是我们的经验内容,那么人工智能本身也具有了概念性。我们如何去理解这种概念性呢? 麦克道威尔又写道“我利用康德的自发性思想的方式,迫使我对‘概念’和‘概念性的’这样的词进行严格的解释。从严格的意义上说,概念能力的关键在于,它们可以被用于主动思维,这种思维是开放的,是可以反思自身的理性凭据。当我说经验的内容是概念性的,这就是我所说的‘概念性的’。”[16]47 这种概念性是离不开人的自发性的,综合第一段引用,我们可以看出来概念性源于自发性与接受性的结合,概念性既受到来自世界的有关经验的冲击,又依赖于我们自身的理性能力。当麦克道威尔谈及概念性、概念能力时都是围绕人进行谈论的,其对概念的重视程度也彰显了他对人作为理性的动物的这种独一无二的特征的强调。人工智能与人有着本质的区别,其并没有自发性,理性,脱离人本身去谈论人工智能是否拥有概念是没有意义的。按照麦克道威尔的理解我们可以说作为我们的经验内容人工智能是具有概念性的,但人工智能本身并不拥有概念能力,并不拥有概念。

四　结论与进一步的讨论

人工智能领域中的概念学习引发了本文对于人工智能是否拥有概念这个问题的探讨。对于此问题进行回答的关键在于如何理解概念。在对人工智能领域概念学习的讨论中我们可以看出其学习的概念是一种认知机制,或仅仅是一种文字符号,这种概念观与对概念的常识性理解相吻合,也与认知科学对概念的理解相吻合。在认知科学中概念是原型、范例、理论,这三种不同的概念表现了三种不同的归类方式和学习方式,而人工智能,比如机器学习就是在模仿认知科学中所研究的归类方式与学习方式,因此在认知科学范式中人工智能是拥有概念的。

但是人工智能概念学习中的这种概念与上述所讨论的哲学中的概念却是南辕北辙。人工智能中的概念是可以学习的,而福多的概念先天论认为词汇概念是不能学习的,除此之外黑格尔以及麦克道威尔又是从理性主义角度理解概念,其将概念理解为人作为理性动物的标志,因而人工智能中的概念以及认知科学中的概念与上述哲学家所理解的概念不在一个层面上,在本文所引述的三位哲学家的哲学中概念是具有先天或先验色彩的,是人所独有的,人工智能作为人类制造的智能是不拥有概念的。

通过以上论述产生了一个有意思的局面,即: 认知科学和上述所讨论的哲学都是对人的概念的研究与反思,但其对概念的研究方向或者研究内容却大相径庭。在认知科学中概念被表征为外在的特征,而在上述哲学中概念是不能被外在表征的,概念是人的内部的状态或属性,这样在认知科学与上述哲学中产生了一个巨大的鸿沟。就像麦克道威尔所说的那样“从哲学的扶手椅上否认认知心理学是一门在智力上受人尊敬的学科是危险的,至少只要它保持在其适当的范围内。很难理解,如果不将(概念) 内容归因于内部状态和事件,并且这种(概念) 内容又不受概念能力(如果有的话) 的限制的话,认知心理学是如何运作的,如若认知心理学试图使生

命变得可理解。”[16]55那么我们如何去理解这种鸿沟呢,如何去理解这两种范式对于概念所做的完全不同的研究和反思呢? 是向科学靠拢? 还是执行某种哲学的判定?

其实,最基本的做法,应该向科学学习,并且注意哲学的观念,特别是不要以某种哲学作为科学思想与科学实践的裁决者。如果哲学不把概念看作是先验的呢? 有这样的哲学吗? 事实上,科学实践哲学和大多数经验主义的哲学,都认为概念是后天形成的,是与经验相关的。基于经验的哲学概念研究可以与认知科学的概念研究齐头并进,共同促进人工智能的概念学习。

【参 考 文 献】

[1] ALLEN C, VARNER G, ZINSER J. Prolegomena to any future artificial agent [J]. Journal of experimental & theoretical artificial intelligence, 2000, 12(3): 251 – 261.

[2] SOTALA K. Concept learning for safe autonomous AI [C]// WALSH T. Artificial intelligence and ethics, papers from the 2015 AAAI workshop. Austin: AAAI Press, 2015: 83 – 84.

[3] LAKE B M, SALAKHUTDINOV R, TENENBAUM J B. Human – level concept learning through probabilistic program induction [J]. Science, 2015, 350(6266): 1332 – 1338.

[4] WIKIPEDIA. Concept [EB/OL]. (2020 – 03 – 14) [2020 – 03 – 16]. https://en.wikipedia.org/w/index.php? title = Concept&oldid = 971136729.

[5] MACHERY E. Doing without concepts [M]. New York: Oxford University Press, 2009.

[6] MARGOLIS E, LAURENCE S. Concepts [EB/OL]. (2019 – 06 – 21) [2020 – 03 – 14]. https://plato.stanford.edu/archives/sum2019/entries/concepts.

[7] PARK J J. The theory – theory of moral concepts. [J]. Cognition and neuroethics, 2015, 3(1): 122.

[8] 向必灯,李平. 概念的异质性学说剖析[J]. 自然辩证法通讯,2018,040(004): 29.

[9] BODEN M A. AI: it's nature and future [M]. Oxford: Oxford University Press, 2016: 9.

[10] MCCAFFREY J, MACHERY E. Philosophical issues about concepts [J]. Wiley interdisciplinary reviews: cognitive science, 2012, 3(2): 274.

[11] FODOR J A. Representations: philosophical essays on the foundations of cognitive science [M]. Sussex: The Harvester Press, 1981: 263 – 265.

[12] FODOR J A. Concepts: where cognitive science went wrong [M]. New York: Oxford University Press, 1998: 121.

[13] BRANDOM R. A spirit of trust: a semantic reading of Hegel's phenomenology [M]. Cambridge: Harvard University Press, 2019: 4 – 5.

[14] HOULGATE S. Hegel and Brandom on norms, concepts and logical categories [C]//HAMMER E. German idealism contemporary perspectives. London: Routledge, 2007: 139 – 149.

[15] 麦克道威尔. 心灵与世界: 新译本 [M]. 韩林合,译. 北京: 中国人民大学出版社,2014: 29.

[16] MCDOWELL J. Mind and world [M]. Cambridge: Harvard University Press, 1996: 47 – 55.

Does AI have concept?

LI Pei

(*Department of the History of Science*, *Tsinghua University*, *Beijing* 100084, *China*)

Abstract: Concept learning was originally a method to study people's acquisition of concepts in the field of cognitive science, but now it has become a new field for artificial intelligence to conduct self – learning. This study summarizes the content and methods of concept learning related to artificial intelligence, and thus leads to thinking about what is concept and whether artificial intelligence has concept. There are two main paradigms for understanding what a concept is, namely, cognitive science, and a certain philosophy. The analysis shows that AI has concept under cognitive science paradigms, but not under certain philosophical paradigms, which raises questions about how we treat the vast gap between cognitive science and a certain philosophy.

Key words: concept; artificial intelligence; cognitive science; philosophy

(责任编辑 殷 杰)