

纯粹理性批判与人工智能^{*}

TP18

林作铨 (国家智能计算机研究开发中心, 100080)

(汕头大学计算机与信息工程系, 515063)

戴汝为 (中国科学院自动化研究所, 100080)

摘 要

Drew McDermott has criticized the logicism in Artificial Intelligence which caused reconsidering the fundamental issues of AI about commonsense knowledge. The forum centered around the question of whether logic is the best method for representing and reasoning about commonsense. McDermott's conclusion is that logic cannot justify the inference a program makes. This essay addresses a number of questions about commonsense by formalizing nonmonotonic reasoning. We present AI formalism program for challenging further fundamental studies for AI.

一、前 言

非单调推理的提出及其进展^[1]被认为是AI的一个重大进步^[2], 十几年来其研究所揭示出来的一些AI基本问题引起了AI界的深刻反思。1987年CI杂志出版了一个专集^[3], 耶鲁大学的McDermott教授写了一篇题为“纯粹理性批判”的檄文, 以非单调推理研究的现状作为主要依据, 强烈地批评了AI中的“纯粹理性”——AI逻辑主义, 从而引起一场AI中传统的逻辑主义与过程主义的新论争; 继后, CI杂志在1988年又刊出另一个专集^[4], 由Cheesman著文“计算机理解质疑”, 就包括不确定推理问题作进一步讨论, AI杂志也刊登了AI基础论题的更广泛论辨^[5], 包括逻辑, 知识, 过程, 联接与进化主义诸

方面^[6], 此外, 近年的AI会议上都有关于非单调推理的热烈讨论^[7]。一时间纷纷扬扬, 前后有上百名研究者参加, 可谓老中青结合, 不乏有代表性的AI知名学者, 形成了关于AI基础问题争论的一次高潮。在国内, 也有许多类似的讨论。

论争的焦点仍然是逻辑在AI中的作用与地位问题, 特别是从非单调推理到常识推理的一般基础问题。这场论争与由Dreyfus挑起的哲学与AI的“外部”论争不同, 也比以往AI中描述与过程、符号与联接诸派的“内部”论争更富挑战性; 它是由AI逻辑主义者阵营中引发的论争, 特别具有技术性, 并将持续下去, 直接牵动着每个研究者进一步的工作方向。

这场论争自然引起AI研究者的广泛关注与兴趣, 可以说是带动AI研究进展的一种源

^{*} 八六三计划基金资助课题

动力。例如，法国9名主要的AI逻辑研究者拟名Lea Sombe对非单调推理问题进行认真的分析比较^[8]，美国MCC公司中Lenat等人的CYC项目充分地运用了形式化常识带来的积极结果^[9]。对非单调推理的批评迫使研究者们必须进行更加深入与广泛的研究，由此产生了许多新的思想与技术，这些工作在最近的IJCAI, AAAI, KR等AI会议中已呈现为主流，非单调推理作为反绎推理(Abduction)，归纳推理与机器学习等的形式化途径也出现了一些成功的作品，说明了演绎逻辑还有潜在的广阔发展前景。

但是，片面地批评或捍卫非单调推理研究都是没有根据的^[10]，这里，有必要充分地运用辩证法。然而，本文不准备作取长补短之类的综述，相反，企图以维护理性为基点，主要就McDermott提出的批评与挑战作一番介绍与剖析，供读者参考与批评，并简要地说明AI需要建设更为“纯粹理性”的形式化体系的基本意义。

二、纯粹理性批判

2.1 McDermott

Drew McDermott是一位卓有成就的AI逻辑主义者，非单调逻辑的提出者之一，在他的AI著作与他领导的AI研究小组中都明确地运用逻辑作为研究AI的基本途径，他的时态规划等逻辑理论与系统影响广泛，因此，McDermott从AI逻辑学家一下子转向程序员的观点，足以引起普遍注意。

McDermott[1982]证明了非单调模态语义并不比单调模态系统强^[11]（但后来已有一系列改进工作克服这个困难），导致McDermott观点改变的是他与Hanks提出的“Yale枪击问题”^[12]，使他得出类似非单调逻辑的做法是毫无出路的结论，1985年在美国国内的一次AI会议上，McDermott批评了AI逻辑主义学派，这就是后来出现在CI杂志上的结果。

2.2 纯粹理性

McDermott引用Kant的哲学名著指称AI中的逻辑主义是一种纯粹理性。McCarthy是第一个倡导用逻辑作为AI的一般基础，Hayes是逻辑主义的坚定捍卫者，在McCarthy, Hayes[1969]的经典文献中已提出了“框架问题”这类AI基础问题^[13]，特别是，1977—1978年间McCarthy与Hayes分别给出形式化非单调推理与定性推理（“自然物理”）的结果，从两个相关的角度明确地提出使用符号逻辑形式化常识及其推理的目标。

McDermott文中首先给出一个AI逻辑主义者的描述。简明扼要地说，AI逻辑主义有如下的基本纲领：

1. 智能程序拥有关于世界的各种知识。
2. 知识用某种方式表达出来。
3. 在程序使用知识前，必须显式地描述知识。
4. 数理逻辑（特别是一阶谓词演算与简单清晰的真值语义）是一种最好的知识表示符号。
5. 智能主体具有在所归属世界中进行推理的能力，这些推理是基于演绎或其扩展的推理模式。

例如，Hayes的自然物理（或更确切地称为常识物理）是把常识翻译成一阶逻辑公式，McCarthy的限制逻辑是基于—阶逻辑的扩展来处理不完全知识下常识推理的模式。

2.3 批判

McDermott正确地指出，AI逻辑主义有两个基本前提：

1. 所有推理都是演绎或其扩展。
2. 描述知识是独立于程序的。

然而，大多数常识推理模式都是非演绎的，这就有一个问题：基于演绎推理及其扩展的技术能否处理常识推理？例如，从P, P→Q推出Q是演绎推理；而从Q, P→Q找出P是反绎，这是直接用演绎技术无法处理的；同样，从P, Q发现P→Q是归纳，也是非演绎的。

McDermott认为从常识通过演绎逻辑作出常识推理是不可能的, 陈述知识的公理离不开使用知识的程序。

McDermott列举了AI中各种捍卫逻辑主义的主张, 并逐一加以批驳, 这些主张与批驳都联系着非单调推理研究的背景。

1. “理想化”: 理想化观点主张逻辑是现实世界解题的一种理想化的基本模型; 但是, 许多通过观察得到知识的领域问题根本找不到理想化模型, 大多数近似处理成演绎的问题都难以得到合理的结论。

2. “基本词汇”: 基本词汇观点主张可以选择一个所需的谓词集作为基本词汇来解决某一类基本问题; 但是, 基本词汇的变动难以规定一类问题的充要条件, 而且把复杂问题变成简单平凡的问题来处理有时无助于问题的最终解决。

3. “科学皇后”: 科学皇后观点主张逻辑是发现演绎与非演绎之间的纽带, 演绎是基础; 但是, 如果把反绎推理直接看成演绎推理的反面来处理, 那么科学解释显然就是没用的, 因为解释与被解释之间的关系既不是充分也不是必要的。

4. “元理论”: 元理论观点主张存在一个演绎的元理论足以保护元级推理模式; 但是, 这样一种强大能力的元理论在计算上是难于保证的。

5. “演绎技术”: 演绎技术观点主张许多常识推理问题基本上是演绎的, 特别是借助于逻辑程序设计技术与工具; 但是, 存在许多问题, 如非有限公理化理论, 不可能在归结证明中做到, 对构造性问题证明器是无能为力的。

6. “非单调”: 非单调观点主张通过扩展演绎推理到非单调推理能捕捉很大一类常识推理; 但是, 非单调推理存在一些无法解决的困难。

McDermott通过具体例证指出AI逻辑主义的尝试都是失败的。以上六种主张都坚信一阶逻辑及其扩展或限定可以很好地处理常

识问题, 都围绕非单调推理研究为主要内容。McDermott通过“枪击问题”否定了非单调逻辑的作用。枪击问题其实是现存非单调逻辑存在本质困难的一个生动例子, 即“扩张(模型)存在性”问题, 各种非单调逻辑(如默认逻辑或限制逻辑)都可能存在多个扩张(或模型)或没有扩张(或模型)的情形, 从而非单调逻辑或太弱或无法进行优先选择或导致悖论, 枪击问题存在多个(两个)扩张, 因此可能作出反常的结论。非单调逻辑的另一个基本问题是不可判定性, 这使得非单调推理难于应用, 也是AI逻辑的一个共性问题。

McDermott指出几乎所有计算机程序中的推理都包含有非单调与非演绎的部分, 非单调逻辑面临的困难在算法中都很容易解决, 例如, 解决优先选择问题的所有处理方案都包含某种算法的成分, 现有的形式理论还没能很好地说明程序中所蕴含的这类非演绎性质, 特别是, 逻辑只是作了过程推理的一些简单重复工作, 对理解智能机制没什么新贡献, 寻找一种如非单调性的演绎系统是注定不可能成功的, 不管用多少条公理来陈述多少知识, 用演绎技术难以做出大多数常识推理形式, 而且还需要一个程序来保证, 换句话说, 没有一个“过程模型”, 仍然无法得出一种“描述模型”。McDermott的结论是:

1. 大多数常识推理本质上是非演绎的, 不能用演绎技术来处理;
2. 不能独立于程序描述知识, 而只能更好地设计程序。

2.4 论争

McDermott 批评的问题是围绕什么是关于常识表示及其推理的最好方法, 这也是AI研究者面临的选择问题, 自然引起强烈的反应与论争。

论争中有拥护者也有反对者, 拥护McDermott的各方有来自传统的哲学或新近的联接主义的观点, 也有过程主义的典型观点, 这些论点或者更仔细地强调如启发式搜

索技术的重要性，或者从实用立场考虑问题，或者追求逻辑与程序的有机结合。

坚定的逻辑主义者被 McDermott 的批评激起了强烈的反响，他们或者分析 McDermott 论证中的错误，或者指责 McDermott 断言的轻率，大多数反对者的反批评是技术性的。其实，在逻辑主义纲领中，对“过程模型”与“描述模型”有一定的界限，就是如 McCarthy 对认识合适性与启发合适性的区分^[13]，或 Newell 的知识级与操作级的作用，或 Feigenbaum 的“知识阈值”或“智能阈值”^[15]等等，但是，McDermott 的结论命题是彻底的过程主义，这就伤了逻辑主义者的心。

2.5 纯粹反理性批判

A. 逻辑主义者

McDermott 自己并摆脱不开追求 AI 逻辑基础的理性，他所指出的逻辑对智能机制的困难也是研究 AI 的其它途径所共有的，因此强调了逻辑在实现方面的困难。非单调逻辑一般是不可计算的，实即逻辑的先天问题。如所周知，一阶逻辑是 AI 逻辑主义的基础，只是半可判定，即使命题逻辑计算也是 NP 性的，但这并不影响寻找一段足够描述知识的可计算部分。如果 McDermott 再写一个批评 AI 中过程主义的文章，也会反现同样的困难，经过过程实现的模型最终还是希望能得到一个描述模型，以便说明过程的机制。

逻辑主义者基本上同意 McDermott 所作出的分析，但不同意 McDermott 的结论，问题的关键是关于常识的理解与所运用的技术。但是，大多数逻辑主义者的反批评都认为逻辑是理解智能程序的强有力工具，而且认为 McDermott 所勾勒出的逻辑主义太狭窄了，而应区分逻辑与逻辑主义，逻辑只是在 AI 中使用，不可能要求逻辑自动地提供所陈述公理的理论基础。而且，真正的逻辑主义者也接受 McDermott 提出的挑战。

McDermott 论证中的两个核心问题是

演绎的作用与非单调推理面临的困难：a. 演绎问题。演绎逻辑是成熟的技术，简单清晰的真值语义对理解程序是必要的，而且也是常识推理中有实质意义的基础部分，问题是用什么技术可以把常识推理处理成演绎的，非单调推理的成功的一例。b. 非单调问题。非单调推理是经过长期努力才能提出来的，短短几年的发展已经带来了明显的成效，例如，奠定了逻辑程序设计(如 Prolog)的理论基础，不能说面临的困难是无法解决的，例如，就“扩张存在性”问题近年来已有几十种之多的解决方案，这些方案不是说已最终解决了问题，但使得对问题的理解更深入了。非单调推理研究甚至还会揭示出更大的困难，例如，McCarthy 提出了形式化环境概念的目标，至今仍无动静^[16]。

B. AI 概率主义

Cheeseman 是 AI 概率主义的主要倡导者，他继承 McDermott 对演绎逻辑的批评，但反对从逻辑主义走向过程主义，提出不确定推理可以很好地处理常识表示及其推理问题，例如，容易克服非单调推理无法描述环境的困难；进一步，Cheeseman 论证了 Bayes 推理可以作为不确定推理的一般基础^[4]，基于概率之外的不确定推理方法都是不必要的。Cheeseman 论题引起进一步的论争，包括两个主题：关于不确定推理（特别是概率论）在 AI 中的作用问题，以及不确定推理与逻辑的关系问题。大多数研究者都寻找一种不确定推理与逻辑结合的基础，并且付诸实践。

三、人工智能的逻辑基础

3.1 关于非单调推理的反思

非单调推理研究者强调建立描述常识的一般机制的重要性，批评者认为太一般化的形式化对了解某一类问题的智能机制没什么帮助。关于非单调推理的论争需要研究者作出有说服力的技术性结果，例如，应用非单调

推理来形式化如反绎推理, 归纳推理等已有积极的结果, 说明了运用合适的技术来扩展演绎推理的能力是完全可以做到的。

实际上, 人的常识推理能力包括诸多方面, 非单调推理只是在不完全知识情形下进行常识推理的一般基础, 而非协调推理是在不相容知识状态下常识推理的一般基础, 定性推理是通过常识物理描述来刻画常识推理的基本性质, 不确定推理则以常识推理普遍的非确定性为出发点。此外, 归纳推理、反绎推理、信念推理等等都从不同方面研究常识推理的形式化问题。研究者们常常从各自方面武断某一种推理为一般的常识推理基础, 各种推理之间的内在联系尚缺乏研究, 这就有如瞎子摸象。

其实, 各种推理的研究都面临共同的常识问题, 下一步是建立它们之间存在的联系。例如, 尽管从矛盾观点看, 非单调推理与非协调推理具有类似的功能, 但大多数非单调逻辑是一种协调理论, 不能处理含矛盾的推理。形式化非协调推理必须使矛盾局部化或孤立化, 解决演绎逻辑中从矛盾可推出一切命题的平凡性问题。从非单调推理方面, 如果能容忍矛盾而且使矛盾传播局部化, 也就无所谓扩张存在性问题了。非单调性与非协调性都是常识不完全性的具体特征, 通过结合两种推理可以做到在更一般意义下形式化常识推理。此外, 如定性推理一般须建立在非单调推理之上, 非单调推理中结合不确定推理可以进行各种优先选择, 等等, 这些联系都有待进一步研究, 这是从非单调推理到常识推理的必由之路。

3.2 AI 基础研究

非单调推理研究的一大贡献是由它才能揭示出 AI 面临的关于常识的一些根本困难, 这是没有深刻的理论研究所不能做到的, 这种从实践到理论的过程是 AI 成熟的标志。另一个生动的例子是, 软件工程中的许多程序设计专家现在都在研究程序的逻辑基础。

实际上, AI 逻辑主义的最终目标就是试图建立智能程序的严密数学理论基础, 但是, AI 逻辑基础研究的进展相当缓慢, 非单调推理的进展只是很小的一步, 可以说, AI 逻辑主义者仍然没有一个明确的基础纲领, 也没有建立一种统一的基础体系。

AI 通过机器智能来模拟人的智能, 而衡量一个智能程序是否达到人的智能程度的一般准则是看它是否具有人的知识 (特别是常识) 行为, 因此, 关于常识表示及其推理是 AI 的基础研究, 正如对算法的研究是计算机科学的基础研究。

3.3 常识逻辑

AI 面临的常识问题要求跳出基于一阶逻辑及其计算理论的框架, 这是 AI 面临困难的根源。但是, 要给常识下一个精确的定义决非易事, 人的常识的某些方面并不适宜从逻辑进行研究, 这里有一个关于常识的常识问题。

A. McCarthy 的追求

John McCarthy 的工作是 AI 逻辑主义的基本纲领和典范^[16], 极富启发性, 在他补交的 Turing 奖演讲词中对 AI 逻辑基础作出了最好的总结^[18]。

McCarthy 看到通过自动机来定义人的智能是不可能成功的^[17]。1956年在 Dartmouth 出现的 AI 工作已开始了新的尝试, Turing 测试说明了对智能的定义是难于从数学上直接做到的, 在 Teddington 会议上 McCarthy 首先明确地提出常识问题^[16], 通过建造具有常识的程序作为智能的标准, 从而开拓了 AI 描述途径^[18]。

McCarthy 指出建立一般用途的独立于程序的常识库的必要性, 而且希望有一种具有足够丰富的表达能力且能适当控制的表示语言。沿着 McCarthy 的路线, AI 逻辑将有一条漫长的历程。

B. CYC 的目标

与日本人的第五代机不同, MCC 中 Lenat 等人的 CYC 系统的目标是建造一个足

够大的常识库,以供智能程序共享,正如微机中必备的DOS。CYC 发展一个基于一阶逻辑具有很强的表达能力的表示语言,并在认识级与启发级的不同层次上容纳各种推理模式。CYC 充分运用了形式化常识及其推理的积极结果^[1],但 CYC 没能形成一个内在统一的常识理论来保证系统的常识行为,也很难看到从 CYC 中可以得出这样一个理论。然而,CYC 将提供给我们最可靠的经验材料。

C. AI 形式主义

尽管数学中的逻辑主义不可能完全成功,但逻辑主义坚实地奠定了数学的逻辑基础,数学基础研究至少形成了什么是数学的一个很大的严密统一体系。AI 逻辑主义也有类似的有限目标,那就是 McCarthy 所指的 AI 认识论问题^[10]。然而,AI 逻辑主义者更多地把逻辑作为一种分析程序的工具来使用,还没有形成一种严密统一的智能基础体系。在这个意义上说,AI 需要类似数学中形式主义的更为“纯粹理性”的 Hilbert 规划^[20],AI 形式主义的目标是用逻辑定义出一类“机器智能”或“逻辑智能”,使得什么是智能有一个坚实的理论基础,或作为接受批评的基础。机器智能不一定完全等同于人的智能,但至于给出一个很大程度上已是智能的人工仿物。

AI 形式主义者的一个具体目标是:a. 常识的逻辑基础。建立一个常识表示及其推理的形式化体系,从而在一定意义下定义什么是常识;b. 具有常识的程序系统。建造一个独立的常识库使得使用它的程序体现智能行为,从而在一定意义下定义什么是智能。

我们可以对人的常识推理做一些正常合理的假设,从而得到一个基本的常识模型,在这个模型中充分地考虑常识推理的诸方面特性,使得模型中主体的行为符合常识规范。我们希望所建立的常识模型是完全内在独立的统一体系,它不依赖于任何外在的控

制机制,具有足够丰富的表达能力,而且在大多数情况下是可计算的,所建造常识库的常识系统是常识逻辑的实现形式。

从常识观点看,常识逻辑可以超脱悖论或 Godel 不完全性定理或 Hofstadter 泛悖论主义的束缚。从逻辑观点看,这也是数理逻辑面临的挑战。

McDermott 对 AI 逻辑主义的批评是严肃中肯的,特别是 AI 逻辑主义至今未能形成一种坚实的基础研究,但 McDermott 悲观的结论是不可取的。AI 逻辑存在的困难并不是无法解决的,而是如何寻找解决问题的方法的技术问题,这也不是能一蹴而就的。AI 需要类似数学中形式主义的更加明确的基础研究来形成严密统一的智能理论体系。我们看到 McDermott 仍旧是 AI 逻辑主义者,从事“纯粹理性”工作^[21]。

应当指出,这场论争在 1987 年形成气候,现在已经不能单纯地从哲学上进行议论,而需要研究者从技术上给出实质性结果来说明问题,只有这样才能促进 AI 的进步。值得一提的是,关于 AI 争鸣已大约有 500 篇的研究文献,结合技术性工作全面细致地讨论关于 AI 与常识问题引起的诸子百家的争鸣历史、现状与发展是一项待做的有益工作,它会帮助我们更好地理解 AI,我们希望不久能见到一个这样的工作。

我们主要介绍了 McDermott 对 AI 逻辑主义的批评引起的论争,以及由非单调推理研究揭示出来的 AI 关于常识的基础问题,而更多的问题未能讨论,所讨论的问题与想法限于篇幅都是相当简略的,相信读者会有自己的主张而且引导自己进一步的工作,这样我们的目的就达到了。

最后,注意到,本文之外的另一方面内容是 AI 中的“非理性”部分,如上所述,关于常识的许多方面并不适合于从逻辑进行研究,这需要另文专论。

参考文献

- [1] 林作铨,石纯一,非单调推理十年进展,计算机科学,1990年6期
- [2] Bobrow, D, Hayes, P. (Eds.), What We Are, AI 25 (1985)
- [3] Levesque, H. (Ed.), Tacking Issue/Forum: A Critique of Pure Reason, and

- Commentaries, CI 3,3 (1987), 151-237
/CI 4,4 (1988), 401-414
- [4] McLeish, M. (Ed.), Tacking Issue/Forum: An Inquiry into Computer Understanding, and Commentaries, CI 4,1 (1988), 57-142/CI 6,3 (1990) 179-192
- [5] Krish, D. (Ed.), Special Issue on Foundations of Artificial Intelligence, AI 48 (1991); 人工智能基础, 石纯一, 吴轶华 (编辑) (国家智能计算机研究开发中心技术资料, No. 1, 1991)
- [6] 石纯一, 李国杰, 吴轶华, 林作铨, 人工智能基础的辩论, 人工智能基础, 石纯一, 吴轶华 (编辑), (国家智能计算机研究开发中心技术资料, No. 1, 1991)
- [7] Etherington, D., et al., Critical Issue in Nonmonotonic Reasoning, KR-89, 500-504
- [8] Lea Sombe, A Comparative Study of Reasoning under Incomplete Information with a Simple Example, Int. J. Intell. Syst., 5,2 (1990); (AW Press, 1991)
- [9] Lenat, D., et al., CYC: Toward Programs with Common Sense, CACM 33 (1990), 30-40
- [10] 王珏, 王培, 白硕, 赵致琢, 吴轶华, 私人交谈, "863"理论 AI 讨论班, 1990
- [11] McDermott, D., Nonmonotonic Logic I, Nonmonotonic Modal Theories, JACM 29 (1982)
- [12] Hanks, S., McDermott, D., Default Reasoning and Temporal Logics, AAAI-86
- [13] McCarthy, J., Hayes, P., Some Philosophical Problems from the Standpoint of Artificial Intelligence, MI 4(1969)
- [14] Newell, A., The Knowledge Level, AI 18 (1982)
- [15] Lifschitz, V. (Ed.), Formalizing Common Sense Reasoning, Papers by John McCarthy, (Alex, 1990)
- [16] McCarthy, J., Generality in Artificial Intelligence, CACM 30 (1987), 中译文见《计算机科学》, 1988, No.5
- [17] Shannon, C., McCarthy, J. (Eds.), Automata Studies, (Princeton Uni. Press, 1956)
- [18] McCarthy, J., Program with Common Sense, The Teddington Conf. on the Mechanization of Thought Process, 1958
- [19] McCarthy, J., Epistemological Problems of Artificial Intelligence, IJCAI -77
- [20] 林作铨, 人工智能形式主义, 人工智能与智能机, (国防工业出版社, 1991)
- [21] McDermott, D., A General Framework for Reason Maintenance, AI 50 (1990)