

John McCarthy谈人工智能研究途径

—Reid G. Hoffman 访问记

R. G. Hoffmam (RGH): 并行分布式处理近来很热门, 有些人认为, 它可以解决“传统的”基于符号的 AI 方法久攻不下的问题。对此您有何看法?

J. McCarthy (JM): 我看过 Rumelhart 等人著的《并行分布式处理》这一两卷长的 PDP 经典著作, 觉得它仍有五十年代和六十年代神经模式识别研究中所存在的许多缺点。确切地说, 它注重一阶谓词。

不妨举一个例子, Rumelhart 等人写的第一篇论文是“PDP模型概要及其顺序思维过程”(载于第二卷), 文中有一个(神经)网络按某些家具陈设与否, 对一幢房子的各个房间学会识别或学会卧室、厨房、等等的归类。这个网络仅对单个房间有效。

然而, 假如我们想把毗邻卧室且仅在其旁边的小房间归类为卫生间, 或者把位于厨房与餐室之间的房间归类为配膳室, 那么就需设定整幢房子所有房间的特征及各房间的相互关系(如毗邻关系), 这将使输入的数目大大增加。我觉得, 假如他们想要解决这个问题, 他们或许能解决, 但这势必会将他们原先训练过的网络实质上加以抛弃。

某些搞并行分布式处理的人认识到这些问题。他们认识到与符号结构有关的各种问题。但我最近未曾留意他们的工作, 也未听到他们宣称已解决了这些问题, 所以, 目前我觉得并行分布式处理是一条局限性很大的途径。

RGH: 最近, 在《脑与行为科学》这本杂志上有一篇“论正确对待连接论”的文章, Paul Smolensky 假设: 连接论的体系结构或设计

处于子符号一级, 或者说位于“符号的”认知科学或一种人工智能方法的底层。从这一基础出发, 或许可以这样说: 这些特征起因于一个巨大的神经网络——例如, 它们无需通过编程。

JM: 他们认为自己正确, 这未免有点乐观。

RGH: 有人断言, 并行分布式处理是解决感知或学习问题的正确途径, 研究一种介于并行分布式处理方法和传统的人工智能方法之间的混合式机器会克服人工智能面临的某些问题, 对此您有何评论?

JM: 学习有多种方式, 通过增强逐步掌握仅是其中一种。占相当数量的是一学便会的方式, 并行分布式处理看来与此无关。

例如, 举 Nettalk 网络为例, 它会调整 24,000 个权值, 当该网络考虑了英文文本中的某一字母及其左右各三个字母, 会知道如何读出第一个字母的音。现在假如我们附加下述关于现代汉语拼音的说明, 如 q 发音为“ch”, x 发音为“sh”, 那么, 若将此告诉一个人, 他(或她)便能朗读夹有汉字的英文文本, 他(或她)当然不是通过即刻调整 24,000 个权值来做到这一点的。所以, 我觉得, 即便对于学习, 并行分布式处理也是一种有局限性的途径。

我对连接论如何与符号的人工智能方法结合并无定见, 当然这值得深思。

RGH: 人工智能与计算机科学的联系较之计算机科学与其它子学科的联系可能更加疏远, 因为人工智能在很大程度上涉及对知识、解题或学习的思索, 这往往是哲学家而非计算机科学家的主题。您对人工智能与计算机

科学、哲学、心理学的关系有何评论？

JM: 我们可考虑通过两种途径来实现人工智能。一是生物学的方法，二是计算机科学的方法。生物学方法得益于这一事实：智能是人类的存在所体现的，因此，在神经生理学或心理学方面对人进行研究，可望知道一些关于智能的奥秘。计算机科学方法认为：我们真正所要研究的是世界以及智能的策略，只要这些策略能使一个系统实现世界上的目标，便不管其是生物的还是人工智能的。

我觉得人工智能方面的大部分工作沿用计算机科学的方法。现在，它与心理学相互之间已有相当的影响。不过在很大程度上，这种相互影响沿着另一方向进行，即：心理学家一贯致力于信息处理模型。在某种意义上，这正是他们一直应做的事。不过在1910年前后，他们曾误入行为主义的死胡同，直到他们注意到计算机时方才钻了出来。人工智能研究对此有何特殊贡献我尚不清楚。

说到哲学，人工智能也有哲学面临的一些问题，因为如果我们想搞一个智能计算机程序(如机器人控制程序)，那么该程序对世界要有一个总的概念。要编制这一程序，我们必须理解关于知识的一些概念：各类知识是什么？应考虑的各类对象又是什么？等等。这些基本上也是哲学家在认识论和本体论方面研究的问题。我现在觉得有些东西人工智能可取自哲学，也有些东西哲学可取自人工智能，因为人工智能学者考虑这些问题的方式比起哲学家来更为具体。

RGH: 您对人工智能研究的生物学/脑/神经学方法和计算机科学方法作了区别。关于这两条途径，您认为相互间是否会有一些有益的影响？

JM: 最终将会，但这取决于神经学研究或许是遗传学方面的某些进展。换句话说，我觉得一旦使用神经学方法的那些人能说出在人脑中事实和行为是如何表示的某些奥秘，或者在基因中是如何的，那就会有相当强烈的

相互影响。举例来说，某种动物具有一种相当固定而精巧的行为，象蜜蜂或苍蝇，由于这种行为是固定的，因而必定是以某种遗传方式表现出来的。我相信：对于人工智能来说，学习一些与此有关的东西是极其有益的。我想这需采用生物学的方法。

RGH: 在以Edward Feigenbaum为代表的工程人工智能和以您为代表的科学人工智能之间，有时看来有些分歧。有些人断言：科学的大部分工作已经做完，所剩下的不过是建立一个足够大的专家系统，另一些人则主张：基础科学依然留下许多工作要做，那些技术专家的断言未免言之过早。

JM: 我不知道是否有人采取这种极端工程的观点。这可有点象稻草人一样不堪一击。我不知道有谁会断言科学上已无所作为，如果有人，我将十分乐意与他辩论。现在，且让我们暂时设想有这么一个稻草人，可举MYCIN这个用于诊断血液细菌感染的专家系统为例，MYCIN具有一个很大的规则集，这些规则涉及病症、诊断测试结果和可能的治疗方案。但MYCIN并不知道病人，你不能对MYCIN讲：“我昨天按你的意见治疗一个病人，结果他(或她)死了。我今天又来了一个相同症状的病人。”MYCIN只是说：“无法识别的响应。”MYCIN不知道细菌是一种侵入人体后便能在体内生长繁殖的生物，MYCIN甚至对病程的最终发展一无所知，也无法对它讲述这些东西。如果你希望有一个具备这些性能的系统，那么只能丢弃MYCIN，一切从头开始。虽然并非所有专家系统都有那样的局限性，但处理一般信息并推出一般结论并非为目前一类专家系统所能胜任。我想某些专家系统将需要这种一般知识，赋予它这些知识依赖于科学的进步。

RGH: 显然，科学上某些东西仍有待解决，那么，顺利构造一个人工智能机器，其最大障碍是什么？

JM: 唔，自1958年以来我一直认为：人工智能最关键的问题是常识和推理，这里，常

识意味着我们的工作并不限于一个固定的模型之内，相反，我们随时可把新现象考虑进去。这一点截然不同于诸如运筹学或其它形式化科学理论，在这些领域中，科学家首先确定要考虑哪些现象及其关系，然后是在此理论内工作的程序。而常识和推理并不受此约束。附带说一句，或许常识一词表达此意并非是一个最贴切的字眼，因为“常识”(Common sense)含有因人看法而异的意思。但我现在尚未想出一个更好的词。此外，常识推理与解微分方程无关。正如事实所表明的，它与逻辑推理也完全无关。近十年来，在常识和推理研究方面，形式化非单调推理的发展是重要的一步，这种推理采用数理逻辑，但在推理过程中超越了早先数理逻辑系统的范围。

RGH: 近10年至20年来，您认为人工智能方面最重要的进展是什么？根据人工智能的目前状况，您或许愿意详细说一下，下一个10年至20年内您认为将有哪些进展？

JM: 唔，我认为专家系统的发展已取得了令人瞩目的进步。我过去认为一种技术难以在当前有限的科学基础上得到发展，现在我觉得我们正在从专家系统方面学到相当多的东西。目前，在严格科学的领域内，形式化非单调推理是关键——加上其各种外延（人们说的逻辑程序设计和真值维护系统）。

RGH: 您认为人工智能的目前状况如何？下一个20年会给我们带来什么？

JM: 这很难说，因为我的观点是：要达到人类水平的智能，不能只沿一个预期的领域继续进展下去，非有某些根本的突破不可。举例来说，我想，在七十年代中期之前，没有人曾看出形式化非单调推理是值得干的事。我怀疑我们现在所干的事亦很难断定其意义。不过，有两样东西我正在继续研究并且抱有很大希望，那就是上下文概念的形式化和用语音行为描述的系统。

RGH: 关于Allen Newell 的著名物理符号系统假说，通俗地说就是：任何机器，若具

有充分强大的符号处理装置，则具有了足够的、甚至必要的实现智力行为的手段，对此您有何看法？这种假说是否产生于人工智能研究？

JM: 唔，我不能肯定你对这一假说的阐述是否正确。我也不能肯定我是否确切地理解了Newell 的意思。不过，我可把我所认为的一种说法讲一下，我想它符合Newell 的意思，那就是：任何达到人类智能水准的智能系统至少在本质上按其特性可看作为一个符号处理器。

RGH: 人工智能是否已改变了这一观点。我觉得以前有一种强烈的看法：充分有力可实现的符号处理是实现人工智能的手段。

JM: 唔，实现人工智能单凭符号处理的观念还不够。凡是计算机所做的都可看作为符号处理，今天我想没有人会加以反驳。但纯粹就我阐述过的形式而言，这有点象老生常谈。

RGH: 还有一种假说（即知识表达假说）认为：任何机械的智能装置需要有两样东西，第一，具有可为我们（作为外部观察者）看成命题知识的内部结构，第二，这些结构与此机器的行为具有因果关系。您是否认为这确切地阐述了知识表达的基本原则？您对此有何评论？

JM: 唔……这相当于Daniel Dennett在1971年所称的“意向态势”和我在1979年在“赋予机器智能品性”那篇论文中提出的观念以及Newell称之为“逻辑层”的某种意思。或多或少用Dennett的话来说就是：把许多机器、许多人和其它系统看作知识、信条、意向的主体是理解其行为的最佳途径。（用Newell的话来说），对其理解往往是根据一种理性原则，即：行为受思想指导将达到目标。

RGH: 您的回答很有趣，但似乎不那么切题。我想我上面提问的原意是指一个带哲学意味的问题，即推动知识表达方面大部分研究工作的动因是这样一种清晰的语义：知识用结构来表示，容易在讨论中作为有关世界

的命题。(令人想起J. A. Fodor的智能机的“思维语言”。)而且,这些结构与行为直接有关。您是否认为这是一条很好的途径?您是否认为在计算机内知识确应以这种方式构造?

JM: 是的。

RGH: 由于受过各种不同研究方法训练的研究人员后来去了其他一些学校,这就有点普及的味道。但人工智能的某些方法学或途径已与某些学校挂起钩来。一个例子便是斯坦福大学采用逻辑,而卡内基·梅隆大学搞问题求解/认知模型。您能否对方法的这些差异谈一下看法?

JM: 唔,人人都走同一条途径未免糟糕,即使那是一条最佳途径,探索其它许多种不同的途径很有必要。

RGH: 您以前说过:人工智能的背景思想是要实现机器智能而非人类智能。然而您所举的大部分例子,即证明某些设计为合理的论据都涉及到人,例如您反驳 *Nettalk* 是因为一个人不可能即刻调整24,000个权值。这类例子在人工智能中比比皆是。您是否认为这有点矛盾?

JM: 唔,我真正的意思是我们思想上尚未认识到许多智能机理,它们在本质上与我们目前对自身的观察结果不同。我推测:当我们实现了人类水准的智能时,一些主要的概念将显而易见,因为这些概念将通过我们对自身的观察变得明白无误,而且这些概念将由一些大智者加以公式化。让我再加上一点:在人工智能方面,我们确实在实际上利用了我们人类自身并不具备的一些条件:容量更大的快速存储器、快得多的计算速度以及诸

如散列表的各种技巧,所以我相信在观察人类自身的基础上,我们还将发现更多的诸如此类的东西。

RGH: 有些人对人工智能持批判态度。*Winograd*和*Dreyfus*断言人工智能不可能。*Searle*断言人工智能无法创造脑,至多只能搞出思维的模型。而*Weizenbaum*则断言人工智能是不道德的。您可否对此谈点您的看法?

JM: 唔,我觉得:你所提及的一切都归因于我认为已过时的一些哲学观点。我认为影响*Dreyfus*以及*Winograd*和*Flores*的唯象学即使不涉及人工智能而仅局限在哲学范畴内,最终亦将被淘汰。

我认为*Searle*关于意向性的一些观念建立在自省及来自普通语言哲学的进展的基础上,并把过多的一般性原则归于自省自察。我们可以在*Searle*论语音行为和学习语言的语音行为思考一书中明白地看到这一点。我们需要*Searle*将其归于语音行为的某些特性,但不要另外一些特性。看一下我们不想要的那些特性,我们会怀疑这些特性是否正确地反映了人类语音行为这一观察结果。我正在专门思索他的有关许诺的特性表示。

*Weizenbaum*反对人工智能与当前反对技术的总趋势有关,其表现是设想任何专门的科学或技术方面的进步由于可能被人们滥用会如何如何地变成坏事。这可以说是与六十年代的政治激进主义和环境保护主义有关。他们评说过去甚欠公允,我不明白他们怎么能妄言将来。

[崔良沂译自《*Computing Futures*》1989
-90, 冬季刊 孙永强校]