王培专栏｜人工智能：何为“智”？

原创 2015-08-31 王培 赛先生

一个智能系统不必在内部结构或外部行为上和人脑“形似”，但必须在理性原则上与其“神似”。一个计算机系统是否有智能，不在于某一个时刻它能解决什么实际问题，而在于它提供的解是否依赖于系统的历史和处境。

- 王培（美国天普大学计算机与信息科学系）

在《当你谈论人工智能时，到底在谈论什么?》一文中，我解释了在人工智能这个领域中目前对“智能”的理解有非常不同的几派。这种差别直接表现在他们给“智能”下的工作定义上。

所谓“工作定义”，是指在一个理论中使用一个本来有歧义的词汇时，对其在该理论中的意义的界定。这就像是说“我知道这个词有不同的意思，但在这个理论中它是下面的意思……”一个词汇的工作定义当然要尽量符合其在一般使用时的意义，但这不是唯一的考量，尤其是对一个本来就有歧义的词汇来说。实际上，科学史上的一些重大进展恰恰是从对某个常用词汇的重新解释开始的。

在前文中已经提到，我把“智能”定义成“在知识和资源相对不足的条件下的适应能力”，而本文的目的，就是解释我为什么采用了这样一个独特的工作定义。

在与“智能”（及其相关概念，如“认知”“思维”“意识”等等）有关的诸多研究方向中，我所感兴趣的是，通过对人类智能的研究发现智能的一般规律，并将其在计算机中实现。这个说法听上去平淡无奇，但仔细探究起来，它隐含着下列三条基本预设：

（1）人类是有智能的（尽管不同人的智能可能在程度和特征上有所差别）；
（2）人类智能不是智能的唯一可能形态（智能系统并非在所有方面和人完全一样）；
（3）现存的计算机系统基本是没有智能的（否则人工智能早已经实现了）。

下面让我们逐条分析这些基本预设，并考察它与智能的各种工作定义的关系。

第一预设看上去不言自明，但实际上排除了许多“原则派”的智能定义。如前文所述，这一派认为智能代表着某种理性原则，“智能的”意味着按某种标准衡量是最好的。传统的理性标准包括经典逻辑模型和概率论模型及其变种，但人的现实思维活动往往背离这些模型，以至于

有不少人认为人的思维是“非理性”的。如果只有符合经典逻辑或概率论才算有智能，那么普通人也要被排除在智能系统之外了。

第二预设要求智能定义不能只包括人类智能。即使在一个“广义智能”理论完全建立之前，我们也有理由要求它至少涵盖下列智能形态：

（1）人类智能（第一预设）；
（2）人工智能，或者叫计算机智能，至少作为一种理论可能性；
（3）动物智能：某些动物公认比另一些动物“聪明”，而一个广义智能理论应当能解释这种差别。另一方面，既然我们认为智能是进化的产物，那么说其它动物一点智能都没有，这似乎也说不过去，尽管可以说它们的智能比人类低很多；
（4）群体智能：把一个人类或动物群体看成一个智能系统绝不仅仅是一种比喻或拟人化的修辞手段，而是有深刻的合理性。蚁群和蜂群的内在整体性和行为协调性已广为人知，而一个人类组织（如政府、军队、公司、社团等）常常可以像一个人一样被分析其“聪明”或“愚蠢”之处，尽管群体和个体确有各种差别；
（5）外星智能：尽管尚无其存在的证据，起码没有人否认“智能外星人”是个有意义的概念。

一旦第二预设的上述解释被接受，“结构派”和“行为派”的智能定义就显然太“窄”，太“人类中心主义”了。即使撇开人工智能不谈，后三种智能显然既未必基于和人脑一样的内部结构，也未必产生和人类一样的外部行为。

当图灵提出他著名的“图灵测试”时，是将其做为“思维”的**充分**条件，而非**充分必要**条件的，因为他明确承认一个机器可以表现得不像人但仍被认为能思维。因此，他并没有给智能或思维一个“行为派”的**定义**，而这一点被大多数后人误会了。沿“结构派”或“行为派”的路线仍可能造出智能系统来，但由于它们的目标和途径附加了只对人类智能来说是必要的限制，因此它们不是导向人工智能的合适路径。就像在研究制造飞行器时完全“以鸟为师”一样，其问题不在可能性，而在必要性和一般性。

第三预设是基于下面的直觉：尽管今天的计算机已经可以解决很多复杂的问题，我们仍常常觉得它们缺乏人类思维的某些本质特征。这里的差别主要不是在速度、容量、复杂性、可靠性等方面，而是在适应性、灵活性、创造性、自主性等方面。在解决一个具体问题的时候，计算机往往是依赖于一个事先给定的程序，而缺乏变通的能力，也无法应对那些系统设计者没有预料到的情况。而这种工作方式恰恰是传统计算机科学所要求的，即把“问题”看成一个固定的输入—输出关系，把问题的“解法”看成一个固定的逐步将输入变换成输出的“算法”，而“程序”则是算法的计算机可以理解和执行的形态。

根据第三预设，如果一个系统在解决问题时完全依靠预先给定的程序，那就不算有智能，不论问题在人看来有多难，或程序相应于哪种人类认知功能。由此说来，“能力派”和“方法

派”的智能定义就太“宽”了，以至于包含了很多大家直觉上不认为有智能的系统。比如说，现在大概没有人会认为一个排序程序有智能，但在计算机出现之前，“把一组任意对象按某种次序排列”的确是只有人脑才能解决的问题，而其中也涉及了若干认知功能。

综上所述，如果上面三条预设及其解释被接受，那么智能的工作定义只剩下“原则派”一条路，且不能采用基于经典逻辑或概率论的理性原则。

下面让我们一起来理解，为什么“在知识和资源相对不足的条件下的适应能力”符合对智能的工作定义的上述要求。

首先，“知识和资源相对不足”需要进一步澄清。“资源不足”是说，尽管系统的信息加工能力（处理器数量、速度，存储器容量，等等）是有限的，它却必须实时工作，即新任务可以随时出现，且带有时间要求。因此时间不足（没时间想）、空间不足（没地方装）是常态。“知识不足”是说新任务常常超出已有知识的范围，而且所有已有知识都可能被新知识挑战。

这些不足是以往的理论模型（经典逻辑、概率论、图灵机）所未完全考虑的。尽管每个具体系统都仅有有限的能力、知识、资源，但这些模型都假定，它们已经足以完成系统所需应对的任务。以计算模型为例，系统只负责解决那些它已有算法，而且能满足其时空资源要求的问题。对一个超出系统知识和资源范围的问题，它基本上不能提供有价值的答复。如果你问了计算机这样一个问题而它答不上来，那是你的错，不是它的错。

这不是一个智能系统应该做的。我们经常遇到新问题，而且一般没有时间去考虑一个问题的所有相关因素，我们的智能恰恰是这种情形下才得以展现的，而不是体现在那些我们已经预知详细解法，且有足够的时间来依此行事的问题上。对后一种问题，我们会说其解决是“机械的”或“本能的”，是不需要“动脑子”的。

能够在“知识和资源相对不足”的条件下工作只是这个智能定义的一半，而另一半是关于如何在这种条件下体现“理性”。这里的关键概念是“适应”。具体说来包含两点：以过去的知识应对未来的情况；以有限的资源应对无限的需求。有适应能力的系统依据经验决定自己的行为，即使在遇到前所未见的问题时，也会比照已知的类似情况定下对策。当需要考虑的事情超出系统的思考能力时，适应性意味着集中精力于最重要的事情，而什么是“重要”的，也是根据过去经验确定的。

既然系统对未来经验抱开放态度，那就是承认了未来经验和过去经验可能不同，因此也就无法保证自己行为的绝对正确性。在这里，“适应”是一种“尽力而为”的努力，其合理性是不以结果的成败而论的。因为目前的应对都会成为未来的经验，即使失败的尝试也会对未来的决策做出贡献。在一个相对稳定但仍充满变化的环境中，“尽人事，听天命”比“以不变应万变”和“既然天意难测，何妨任意为之”更体现适应性，因此更合乎理性。这种理性是相对于系

统的以往经验和当前资源供应的，因此可以解释系统的错误。“智能”绝不是“全知全能”，但就算错，也要错得情有可原。

“在知识和资源相对不足的条件下的适应能力”作为智能定义是符合前面的三条预设的。当我们考察一个系统的智能时，其着眼点一般不是看它能解决什么问题，而是看其能力是事先确定并无法改变的，还是在经验中逐渐形成并保持一定的可塑性的。这也正是传统的计算机系统和人脑的主要差别。即使是现有的各种机器学习算法，也仍然是基于各种知识资源相对充足的假设，并且其学习结果往往收敛于一个确定的输入—输出映射。按照这些传统观点，每个问题都有一个“正确的”解，和系统的经验与处境无关。但根据上述“相对理性”，什么样的解是“正确的”常常（尽管并非永远）是和系统的经验与处境有关的。

如此一来，本文所描述的这个智能定义和其它智能定义的关系也就清楚了。按照这个标准，一个智能系统不必在内部结构或外部行为上和人脑“形似”，但必须在理性原则上与其“神似”。一个计算机系统是否有智能，不在于某一个时刻它能解决什么实际问题，而在于它提供的解是否依赖于系统的历史和处境。

这个定义的另一个优点是，它为人工智能和一般智能理论划定了一个独特的研究领域，而其它定义则很大程度上把人工智能归结于一个现存的学科：“结构派”主要贡献于脑科学，“行为派”主要贡献于心理学，“能力派”基本就是计算机应用，“方法派”主要贡献于计算机科学，而“原则派”中基于传统理性模型的工作主要贡献于数学和逻辑学。与它们相反，对“在知识和资源相对不足的条件下的适应能力”的研究不完全属于现有的任何一个学科，尽管和很多学科有关系。人工智能系统最终是要通过计算机技术来实现的，但这不意味着智能理论一定要以计算理论为基础。

这个智能定义为人工智能指示了一条和主流观点很不一样的道路，并且可以用来解释人类智能中的很多现象。详细讨论其引申结论及实现途径则超出了本文的范围。

参考资料：

[1] Stuart Russell, Rationality and intelligence, Artificial Intelligence, 94(1-2): 57-77, 1997.
[2] Pei Wang, The assumptions on knowledge and resources in models of rationality, International Journal of Machine Consciousness, 3(1):193-218, 2011.

欢迎个人转发分享，刊物和机构如需转载，请联系授权事宜：iscientists@126.com。

更多精彩文章：您可以回复“目录”，接收往期文章目录和每一篇的获取方式，也可以返回主页点击屏幕下方子菜单获取最新文章、往期文章或直达赛先生微博。谢谢！

关于我们

《赛先生》由百人传媒投资和主办，全球范围内邀请顶尖科学家轮值担任主编。关注请加微信号：iscientists 或扫描下方二维码。

微信扫一扫
关注该公众号