**AI需要理解真实世界**

刘逸川

2020-06-26

理解是从认知对象到智能体的映射。具体而言，理解意味着把握认知对象的本质，正确反映认知对象的规律和模拟认知对象已经或可能发生的事，在各种复杂的情形中也能做出正确判断，而不会被表象所迷惑或误导。真正的人工智能必须理解真实世界，将真实世界映射到自身，以此为基础做出正确的选择、控制、预测等智能活动。

用语言来判断AI是否理解真实世界是常见的方式，AI对语言的理解程度能够反映其对真实世界的理解程度。图灵测试是行为主义的方式，根据机器的表现来判断机器是否理解，这种方式存在局限性。如果机器没有理解真实世界而只是模仿人类的语言表达，那么这不是在测试机器的智能水平，而是在测试提问者的提问水平。如果问题设计得足够好，那么总能设计出问题将机器和人类区分开。同样的，“中文屋”假设的语言手册是有限的，真实的对话是无法列举完全的，总能找到在语言手册之外的问题，而这些问题只要理解真实世界就不难回答。正因如此，自然语言理解被认为是AI完全（AI-Complete）问题[1]，即机器能够完全理解自然语言也达到了人类的智能水平，能够理解真实世界的一切。机器不可能脱离真实世界、操纵文字符号也能表现出人类水平的语言理解能力。

目前的AI能够在某些领域超过人类不代表它们对这些领域的理解比人类更深刻，它们只不过通过强大的计算能力和巨大的存储容量弥补了理解能力的不足。AlphaGo战胜了人类顶尖大师，但是它对围棋的理解甚至不如初学者，它不能解释自己走每一步的意图和战略思想。沃森存储了相当于2亿页书的内容，每秒可处理100万页的信息，凭借着远超人脑的记忆力和检索能力战胜了人类，但是它对问题的理解是非常肤浅的。面对稍微需要理解的问题，沃森仅通过搜索很难找到答案。例如，面对问题“如果把它打进去，你就输掉了比赛”，沃森很难得出“台球的白球”这个答案。[2] 谷歌大脑能够识别猫的图片，准确率和人类不相上下，但它真的理解什么是猫吗？似乎是的，因为它能够把握猫的本质特征。然而它只是理解了猫的外观，而猫还有很多真实世界的常识，例如，猫会抓老鼠；很多人养猫做宠物……只有对这些常识有了清晰的认识，才能说理解了猫。面对下面这张图片[3]，人类能够轻易地描述：这只猫在喝水，但是目前的计算机视觉系统却很难做到，因为它不理解猫是一种生物，会感到口渴，需要找水喝。



侯世达对还没有实现理解的AI嗤之以鼻[4]，他认为现在的AI不过是一种高级的统计方法，和真正的智能相去甚远。例如机器学习本质上是通过分析大量数据找到使得误差最小的模板，但是机器并没有理解学习的对象。我们不必像侯世达那么激进，AI固然需要理解，但是在现阶段，基于统计的AI仍然有着不可替代的作用，这不是误入歧途，而是“两条腿走路”的一条，另一条是等到我们对智能有了更深刻的理解时开发理解真实世界的AI。

**参考资料**

1 Yampolskiy, Roman. AI-Complete, AI-Hard, or AI-Easy: Classification of Problems in Artificial Intelligence[A]. The 23rd Midwest Artificial Intelligence and Cognitive Science Conference, 2012.

2 [美]加里·史密斯. 错觉：AI如何通过数据挖掘误导我们[M]. 钟欣奕,译. 北京：中信出版集团, 2019:10

3 图片来源：麻省理工大学公开课：人工智能. 第9集

4 https://www.sohu.com/a/197702621\_99985415