
人工社会智能——迈向通用人工智能的关键挑战

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/22695.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

人工社会智能——迈向通用人工智能的关键挑战。

本文经授权转载自北京通用人工智能研究院(BIGAI)。

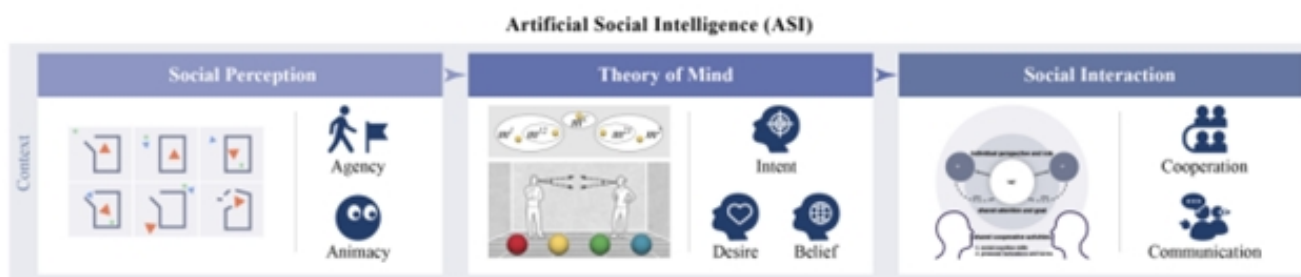
近期，通用人工智能受到了广泛关注，很多人认为通用人工智能的实现似乎近在眼前，然而这其中还有很多问题没有得到解决，其中的一个关键挑战就是人工社会智能(ASI)除了对世界的物理理解，人类还拥有高度的社会智能，一种感知社会事件、推断他人目标和意图并促进社交互动的智能。这种社会智能是人类与最接近的灵长类亲戚区分开来的关键特质。然而，将这种独特智能应用于人工智能领域是一项极具挑战性的任务。

3月10日，北京通用人工智能研究院认知计算与常识推理实验室(<https://www.bigai.ai/department/>)联合北京大学人工智能研究院在CAAI Artificial Intelligence Research期刊上发表综述论文"Artificial

Social Intelligence: A Comparative and Holistic View", 呼吁对人工社会智能(Artificial Social Intelligence, 简称ASI)领域的关注。综述深入研究社会智能的认知机制,并探索如何将这些机制应用于人工智能领域,以创造更加智能、人性化的机器,迈向通用人工智能。

论文链接：<https://www.sciopen.com/article/10.26599/AIR.2022.9150010>

本文工作围绕着社会智能展开,探索人类独特的社会认知能力,并将其应用于人工智能领域。具体地,本综述从认知科学和计算建模两个大的方面了人工社会智能,包括社会感知、心智理论(ToM)和社会交互等子领域,并讨论了研究人工社会智能的重要难点问题是什么、领域内现有资源有哪些、以及未来发展趋势是什么等关键问题。



本文作者为范丽凤(北京通用人工智能研究院)、徐满杰(北京通用人工智能研究院,北京理工大学)、曹智昊(北京通用人工智能研究院,清华大学)、朱毅鑫(北京大学人工智能研究院)、朱松纯(北京通用人工智能研究院)。

人工社会智能的定义和应用

社会智能是指人类在适应更为复杂的社会情境中所展现的社会认知能力。从进化的角度来看,社会智能的发展对于人类的适应至关重要。研究人类社会智能有助于我们设计具有人类特征的交互智能体。

社会智能具有密不可分的三方面社会感知、心智理论和社会交互。

社会感知是指人类从输入刺激中感知到其社会属性(例如生命力、能动性)的能力。影响社会感知的一个基本的因素是动力学结构,人类可以从简单几何形状的运动中感知出复杂的社会信息。

社会感知能力可能处于人类感知和认知处理的交界，通过社会感知，基本的信号刺激被转换成因果的、有生命的、甚至是有意图的实体，从而有助于更高阶更复杂的认知推理过程。

心智理论(Theory of Mind, ToM)是指能够理解自己以及他人的心理状态(包括情绪、信念、意图、愿望等)，并认识到不同个体的视角和心理认知可能不同的能力。它最早在心理学和认知科学中被研究，现在已辐射到人工智能领域。心智理论在多智能体和人机交互环境中尤为重要，因为每个智能体都要理解其他智能体的认知状态才能更好地进行反应、完成交互和合作任务。

Sally-Anne测试是心智理论中的一个经典实验，通过研究儿童在认知发展早期时如何理解他人的信念和错误信念，来测试儿童的心智认知发展水平。在发展心理学研究中，心智理论的形成是早期童年最重要的发展里程碑之一。在婴儿的第一年中，他们开始表现出注视追踪行为，能够将自己和他人识别为能够执行有意识行动的智能体，并能够通过主观体验环境来感知世界。在约14个月时，婴儿开始通过移动获取视觉信息，并能通过眼神来回检查去确认其他智能体和自己看到了一样的现实环境。大约14-18个月时，婴儿开始通过眼神方向理解诸如愿望、意图、情绪等认知状态。到3至4岁，儿童开始理解自己与他在信念和知识上的不同，因此开始理解虚假信念。但是，这种能力直到5至6岁才能完全稳定下来。在发展后期，婴儿才建立了第二阶心智理论，即预测一个人对另一个人的思想或感受的想法，这涉及到更复杂的认知过程。

尽管有很多用来研究心智理论的方法，比如行为分析、神经影像学和神经信号分析等，但是我们亟需一个关于心智理论是什么、人类如何运用它、以及神经系统对其功能有何贡献的全景图像。

社会交互是人类社会智能中的一个关键概念。在研究社会交互时，我们需要关注社交提示、现象、规则和机制等方面，这些方面可以为人工社会智能带来更加复杂、类人化的交流和协作能力。社会提示是社会交互中一个非常重要的方面。一些重要的社会交互包括眼神交流(Gaze communication，通过眼神交流来传达信息)，共同注意(Joint attention，两个或多个人共同注意同一个物体或事件)，指涉(Pointing，用手或其他身体部位指向某个目标或位置)，合作(Cooperation，两个或多个智能体协作以实现共同的目标或任务)。

通院人工社会智能的相关研究成果登上Science子刊

2022年7月，北京通用人工智能研究院一项关于社会智能的研究成果实时双向人机价值对齐Bidirectional human-robot value alignment发表在国际顶级学术期刊 Science Robotics上，并被Science头条报道。详细报道见：朱松纯团队研究成果登上Science头条!为机器立心、实现通用AI迈出重要一步

该研究论文地址：<https://www.science.org/doi/10.1126/scirobotics.abm4183>

这项研究提出了一个基于即时双向价值对齐模型的可解释人工智能系统。在该系统中，一组机器人通过与人类的即时交互并通过人类的反馈来推断人类用户的价值目标，同时通过解释将其决策过程传达给用户，让用户了解机器人做出判断的价值依据。此外，该系统通过推测用户的内在价值偏好，并预测最佳的解释方式，生成人类更容易理解的解释。结果表明，该学习模型基于心智理论，可以在复杂协作任务中提高人机协作的效率，进而提升人机信赖关系，实现真正的自主智能。

人工社会智能难在哪？

综述认为，人工社会智能(ASI) 相比于物理智能更具有挑战性，其中一个很大的挑战就是，它高度依赖于情境(context)。这里的情境可以是大到人群的文化背景和一般常识，也可以是小到两个人的共同经历。这里举一个例子：有两个朋友A和B在图书馆外面看到一辆自行车，其中一个女

生A指向自行车，而另一个女生B对于这个指涉动作背后意图的理解取决于她们各自认知信念和共同信念的情境,如果两个人都知道这个自行车是B的男朋友C的，那么A的意图可能是说进图书馆去找C，而如果两个人都知道B和C刚刚分手了，那A的意图可能是说别去图书馆了。

Context	Intent
A and B mutually know that the bicycle belongs to B's boyfriend C.	"C is in the library. Let's go to find him!"
A and B mutually know that B broke up with C yesterday.	"C is already in the library, so perhaps we should skip it."



现实环境是复杂的、充满歧义的、随机动态多变的，涉及到多智能体交互，且环境只是部分可观察。这一挑战使得一般的标准算法在解决人工社会智能问题时可能会遇到极大的问题。因此，实现人工社会智能(ASI)需要更为全面的方法，针对特定局部模块进行改进，往往不能如愿提升ASI的整体性能。

多学科研究为人工社会智能(ASI)的实现提供了辅助和灵感，研究人类社会智能有助于我们深入了解开发具有类人特征的人工社会智能(ASI)所需的基本原理、纲要、参照系统以及测试基准。

综述还提出，研究发展人工社会智能(ASI)最好的方法是全面、整体、系统地模仿人类与他人以及周围环境的互动模式，让模型像人类一样进行多种方式的学习(如终身学习、多任务学习、单样本/小样本学习、元学习等)，需要一个开放交互式环境，以及思考如何更好地将类人偏差引入

到ASI模型中。

最后，综述表示，尽管人工智能研究取得了重大进展，但我们距离人类水平的智能还有很长的路要走。人工社会智能是实现通用人工智能的关键挑战。我们需要将人工社会智能作为一个独立的研究领域，鼓励AI领域学者讨论和研究这个大课题下的挑战性问题，定义新的问题，创建新的环境和数据集，建立新的评价方式，并建立新的计算模型，做出更多理论和计算方面的开创性工作。最终目标是让人工智能拥有高水平的社会智能，并借助人工社会智能提升人类福祉。

来源：北京通用人工智能研究院(BIGAI)

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发