

智能与计算

李 航

字节跳动科技有限公司

特邀专栏作家

关键词：脑科学 认知科学 心智计算理论 体验认知理论

前言

1950年，图灵发表论文《计算机器与智能》(Computing machinery and intelligence)，提出著名的图灵测试。这段时间里，图灵关注的主要问题是，在计算机上是否可以实现人的思考(thinking)^[1]。他的基本观点是，只要进行适当的编程，计算机可以像人脑一样工作。我们不需要给思考一个严格定义¹，可以通过图灵测试判断计算机的“思考”能力是否达到了人的水平。

1957年，冯·诺伊曼去世，次年他的遗作《计算机与人脑》(The computer and the brain)出版。该书是他在离世前的两年时间里准备的演讲草稿，讨论他当时最关心的研究课题：计算机和人脑。冯·诺伊曼把计算机和人脑都看作是计算机器(automata)，对两者进行了比较，试图为建立统一的计算机器理论奠定基础。

人的思考是不是计算，是怎样的计算？计算机是否可以实现人的思考？这个问题是认知科学、人工智能的一个核心问题，这一点从计算机领域两位巨人对这个问题的关注程度就可见一斑。

本文对计算与思考（或智能）这个话题进行简单综述与讨论。必须申明，笔者是计算机科学家，对脑科学、认知科学等是外行。因为人工智能的目标是要构建能够“思考”和“行动”的机器，所以

作为人工智能的研究人员又不能不对这些问题进行关注与思考，进而斗胆执笔，写出本篇文章，希望能抛砖引玉，引发大家的思索与辩论。

脑科学告诉我们的

人脑是由千亿级的数百种神经元（神经细胞）通过千万亿级的突触连接形成的神经网络，能够实现各种智能性功能，包括感知、认知、语言、情感、创造、意识。脑科学研究虽然取得了一定的成果，但离探明人脑的工作机理还相差甚远^[2]。

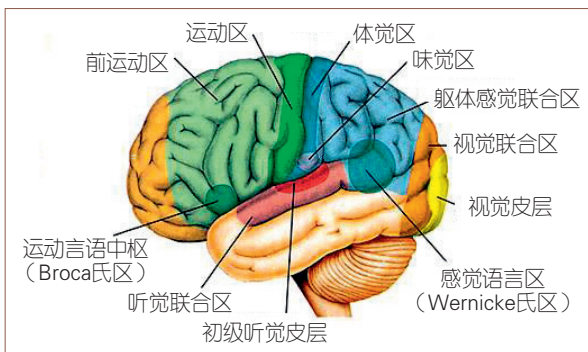


图1 大脑主要区域

在宏观层面，脑科学研究对大脑各个脑区的结构与功能有一定的认识。人脑由大脑、小脑、脑干组成。大脑最重要的部位是大脑皮层，人类与动物的主要区别在于人类拥有极其发达的大脑皮层，可

¹ 这里说的“思考”并没有严格定义，一般包括认知和感知。

以说大脑皮层造就了人类的智慧。大脑皮层不同区域掌管不同的功能,包括视觉皮层、听觉皮层、味觉皮层、体感皮层、运动皮层、语言区等(见图1)。

在微观层面,脑科学研究对神经元的处理机制有比较清楚的了解^[3]。神经元通常由一个细胞体、一个轴突和多个树突组成。树突接入信号,轴突接出信号,神经元与神经元之间由突触连接(见图2)。现神经元从多个前神经元得到输入信号,当输入信号超过一定阈值时被激活,产生输出信号,传递到多个后神经元。神经元之间的信号传递通过突触进行。前神经元在轴突末梢释放化学物质,通过突触传到现神经元的树突,打开现神经元的离子通道(ion channel),促使其细胞内外离子流动,形成现神经元的输入信号。现神经元的输出信号通过轴突以离子流的形式传递到轴突末梢,继续向后神经元传递。

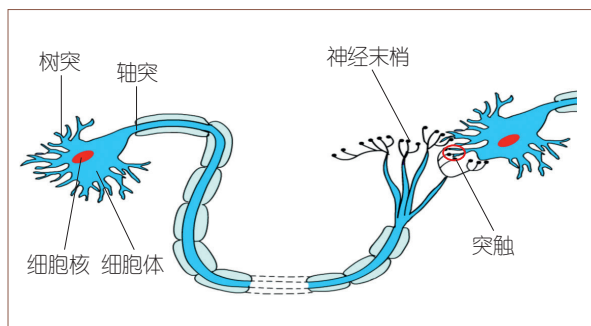


图2 神经元

在介观层面,脑科学研究对神经环路的信息处理原理有一些认识。神经可塑性是神经网络的重要特点,有所谓的赫伯法则(Hebbian rule),认为同时被激活的神经元之间的连接被强化,产生新的链路,形成新的记忆(fire together, wire together)。对概念的记忆存储于由密切连接的神经元组成的细胞群中,激活其中的部分神经元可以唤起对整个概念的记忆。

可以看出人脑是一个由庞大复杂网络组成的信息处理系统。它通过神经元之间的信号传递实现信息处理,具有以下特点:处理速度并不很快,进行的是并行处理,计算与存储融合在一起,拥有自学习能力。

心智的计算理论

心智的计算理论(computational theory of mind)认为,人的思考是计算,人脑或心智是计算系统。这里说的计算不是比喻,而是实质上的^[4]。这个认知科学、脑科学、人工智能等领域的理论,在20世纪60~70年代占据主流地位,代表人物包括认知科学家福多(Jerry Fodor)和平克(Steven Pinker)、脑科学家马尔(David Marr)、哲学家丹奈特(Daniel Dennett)等。

计算系统

马尔提出了计算的层次概念,认为无论是计算机还是心智都是计算系统,需要从三个不同且相关的层次理解,包括计算层、表征层、实现层。计算层决定系统的输入与输出,对应计算的功能;表征层决定系统内部的表征与算法,对应计算的软件;实现层决定系统的物理实现,对应计算的硬件。

心智的计算理论把心智看作是图灵式计算机(Turing style machine),认为人的思考(感知、认知等)是这种机器上的计算。这一点与图灵和冯·诺伊曼的观点一脉相承。有许多理由让人相信这个想法的正确性。给定一个输入,产生一个输出,至少从功能的角度,心智做的是信息处理,可以把心智看作是一种计算系统。神经元对输入的多个信号进行处理,输出一个信号,进而传递信息,从实现的角度,是一种计算器件。

心智的计算理论中,心智的表征理论是重要的一个分支,从表征的角度进一步推进心智是计算系统的想法。

心智的表征理论

心智的表征理论(representational theory of mind)认为思考是在心智中(图灵式计算机上)的符号操作^[5,6]。人的思考和行动是基于常识的,由信念或愿望驱动。信念是对事实的描述,愿望是对目标的描述,常识是对世界的描述,而这些描述是通过内心的语言进行的,称为“心智语言”(mentalese)。也就是说,心智中的符号操作基于心智语言。

心智语言同自然语言一样,由符号和语法组成。符号有简单的,也有复杂的,语法规则决定符号的组合方式以及产生的语义。听别人讲一段话,人一般不能复述原话,但可以把内容讲述出来,对这个现象的解释是,人理解自然语言时把它转化成了心智语言。自然语言有歧义(多个语义),但心智语言没有,原因是人能够区别自然语言的歧义,说明人用不同的心智语言表达了不同的语义。

有一些认知学实验支持心智语言存在的假说。比如,让受验者坐在电脑屏幕前,屏幕上瞬间闪过两个英文字母,根据内容快速按下两个按钮中的一个。如果两个英文字母相同,按其中的一个,如果两个字母不相同,按另外一个。有时出现的是同一个字母且大小写相同(如“A A”“a a”),有时出现的是同一个字母但大小写不同(如“A a”“a A”)。结果发现,大小写相同时,受验者按按钮的速度更快,准确率更高。说明在第二种情况,人需要做某种处理把视觉中的符号转换成心智语言中的符号。

中文房间



图3 中文房间

哲学家塞尔(John Searle)用著名的中文房间(Chinese room)思想实验,对“心智是计算系统,思考是符号操作”的想法提出质疑^[7]。

中文房间思想实验是说,有一个不懂中文的人被放到一个房间里,其他人从房间外塞进写着中文的纸条。房间里有一本书,写着中文会话的规则。他根据书上的规则,对着纸条上的中文符号,找出

相应的中文符号画在纸条上,把纸条塞出房间外(见图3)。从房间外的人看,这个人能够用中文对话,会说中文,但是事实上他完全不会。基于符号操作的计算机,和中文房间里的人一样,看似在使用语言,其实完全没有理解语言。说明语言理解乃至思考,不是计算和符号操作。

中文房间的论点引起了极大的反响,各种支持和反对的意见接踵而至。比如有一个代表性的反对意见是:确实这个人不会讲中文,但是整个房间会讲中文。因为从功能的角度来说这个房间整体可以完成中文的对话,这个人只是会讲中文的系统的一部分。塞尔对此的反驳是:这个人可以把所有的规则都记住,也可以离开这个房间,但是只要他不能把语义附加到符号上,就不能认为他会讲中文。塞尔的主要论点是符号操作只能代表语法,不能代表语义。

体验认知理论

体验认知(embodied cognition)理论²是近二十年来兴起的理论,认为生命体(包括人和其他动物)的身体是感知和认知的基础,身体的体验对感知和认知起着决定性的作用^[8]。代表人物包括认知科学家雷可夫(George Lakoff)、脑科学家达马西奥(Antonio Damasio)、哲学家克拉克(Andy Clark)等。可以说,体验认知理论对心智的计算理论提出了一定的挑战。

脑科学的假说

达马西奥认为,思考是能够在意识中产生表象(image)的,在下意识中进行的对神经表征(neural representation)的操作^[9]。神经表征是人脑的神经活动(神经网络中的信号传递)产生的状态。表象是指人的意识中对事物形象的认识,包括视觉、听觉、体感等的表象。比如,提到“黄色的帽子”,我们会在脑海里联想到黄色的帽子,这就是它的视觉表象。

脑和身体是不可分割的有机体(这里说的身体

² 也有人译作“具身认知”理论。

指除去脑之外的身体部位)。脑和身体的相互作用,形成一个整体,与外界相互作用,产生人的行为。通过神经系统,外界信号可以从身体器官传到大脑,指令信号也可以从大脑传到身体器官。大脑发出的指令未必都经过思考,有很多属于被动的反应。经过思考的指令,会在意识中产生表象,成为人的主动的命令。达马西奥指出“我们未必是思考机器,其实我们是思考的感觉机器(We are not necessarily thinking machines; we are feeling machines that think)”。

思考也使用单词和符号。单词和符号作为表征被记忆,人在说出或写出一句话之前,单词和符号相关的听觉表象、视觉表象等浮现于意识中。人的逻辑和数学思维也基于表象,而不是符号。一个证据是,许多数学家、物理学家,包括爱因斯坦,都将自己的抽象思维过程描述为表象的操作过程。

这里谈到意识,这也是认知科学、脑科学和哲学关注的一个重要问题,至今仍是一个很大的疑团。因为涉及的内容较多,本文不作讨论。

体验模拟假说

体验模拟假说(embodied simulation hypothesis)是关于语言理解的体验认知理论,认为人的语言理解是在心智中进行的,基于自己过去的视觉、听觉、运动等体验的模拟^[10,11]。

人进行语言理解时既使用语言相关的大脑部位,又使用感知和运动相关的大脑部位。理解语言描述的概念时,会联想到概念相关的图像,这时大脑视觉皮层变得活跃;会联想到概念相关的声音,这时大脑听觉皮层变得活跃;会联想到概念相关的运动,这时大脑运动皮层变得活跃。语言理解的过程就是,唤起大脑各个部位相关体验的记忆,基于这些记忆在心智中生成语言所描述的内容的过程。

语言理解大多发生在下意识,在意识层面,会产生相关的表象。比如,问:“大猩猩有没有鼻子?”

要回答这个问题,我们会在脑里先浮现出大猩猩的视觉表象,然后根据这个表象去回答问题。再比如,听到:“flying pig(飞猪)”,不同的人会根据自己对飞的概念的理解(飞的表象),以及对猪的概念的理解(猪的表象)组合成不同的新的表象。

如果认为语言理解不是基于符号,而是基于体验模拟,那么中文房间中的人确实没有理解语言,塞尔的观点可能是正确的。语义不是由符号定义出来的,而是从人与外界交互的体验中积累抽象出来的。

有很多认知学实验证明体验模拟假说的正确性。有这样的实验,让受验者先听一句话,然后看一张图片,之后快速按下两个按钮中的一个。如果图片中出现了句子中描述的物体,按其中的一个按钮,否则按另一按钮。例如,句子有“木匠把钉子钉进墙里”(常识中这时钉子的方向是水平的),“木匠把钉子钉进地板”(常识中这时钉子的方向是垂直的),图片中显示的物体有水平方向的钉子,也有垂直方向的。结果发现句子中钉子的方向和图片中钉子的方向一致时受验者的反应速度更快,判断准确率更高。更一般地,语言中描述的和图像中显示的同种物体,当方向、形状、颜色相同时³,人能更快地判断其同一性。说明人在理解语言时,根据自己的经验在视觉上想象出了对应的场景。

比较与评论

两个理论

心智的计算理论与体验认知理论在思考即计算问题上有相似的观点,但在思考是怎样的计算问题上观点完全不同⁴。从近年的研究成果来看,体验认知理论对人的感知与认知机制能够给出更好的解释,有很多理由让人相信这个理论的正确性,虽然现在还不能完全否定心智的计算理论。

心智的计算理论以意识为主要对象,基本不考

³ 形状:“天空中飞翔的老鹰”与“躲在巢中的老鹰”。颜色:“放在橱柜中的牛排”与“放在餐盘上的牛排”。

⁴ 其实这两个学派都有不同的学者,他们对具体问题的观点不尽相同。

虑下意识；只关心人脑或心智，而不关心身体，对这个理论来说，身心是可以分开的，智能可以独立于身体而存在。体验认知理论关注的是人脑和身体的统一体，强调下意识对意识的影响，身体对人脑或心智的影响；对这个理论来说，身心是不能分开的，（人的）智能不可能独立于身体而存在。心智的计算理论中的计算是意识中的符号特征的操作。体验认知理论中的计算是下意识中的神经表征的操作，其结果浮现于意识中成为表象。图4给出了两个理论的对比。

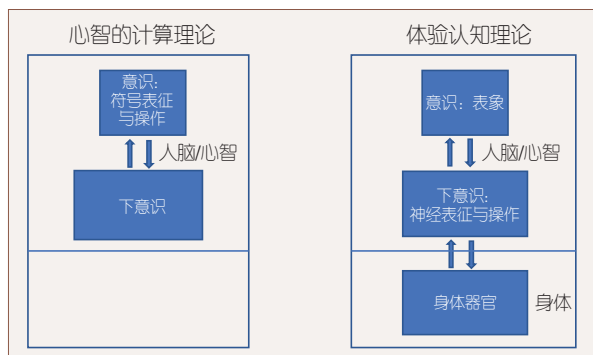


图4 心智的计算理论与体验认知理论的对比

人工智能

在人工智能60多年的历史中，一直有符号主义(symbolism)和连接主义(connectionism)之争。前30年研究的重点是知识与推理，占主导地位的是符号主义，后30年特别是近十年研究的重点是机器学习，特别是深度学习，占主导地位的是连接主义。这与心智的计算理论与体验认知理论的发展在时间上有一定的对应关系。认知科学和人工智能本来就是相互影响的两个学科。

深度学习的重要概念是人工神经网络和神经表征。神经表征将图像、语音和语言的内容都表示为实数向量。人工神经网络是对各种表征进行操作，从而完成各种感知、认知的模式识别任务的机器学习模型。深度学习的“神经表征”与体验感知理论的“神经表征”不尽相同，但也有相通之处。人工神经网络和生物神经网络具有不同的机制，后者借鉴了前者的原理。

结语

思考即计算这一命题是认知科学与人工智能的核心问题。图灵和冯·诺伊曼时代以来，其正确性就不断被一些事实所佐证。计算机在数值计算上早已超过人类，近年在智力竞赛、围棋上又完胜人类，在图像分类、语音识别、机器翻译上也接近人类。这些说明对人类来说属于思考的问题，在计算机上都可以实现。由此看来，图像理解、语言理解等现在看来还非常困难的问题，未来将有希望在计算机上实现或部分实现，原因是这些问题的本质也是人的思考。

心智的计算理论和体验认知理论从不同角度对“思考是怎样的计算”给出了答案。前者认为思考是符号计算，后者认为思考是神经计算。虽然现在没有确定性的结论，但体验认知理论似乎对人的思考机制给出了一个令人信服的解释。近年深度学习的巨大成功说明基于神经计算（人工神经网络）的信息处理能更好地实现人的感知与认知能力。希望体验感知理论的研究取得更大的进展，为人工智能提供更多的启发和引导。

如果智能和身体不可分割的假说成立，那么构建像人一样的智能系统就需要从开发智能系统的“身体”入手，让它们在与环境的互动中获得智能，这似乎意味着要走一条非常遥远而艰难的路径。但现实中往往并不需要构建像人一样的智能系统，很多情况下能得到辅助人的智能工具就足矣，所以问题可以被简化，这时体验认知理论仍然具有借鉴意义。



李航

CCF 高级会员，CCCF 特邀专栏作家。字节跳动科技有限公司人工智能实验室总监 (Director of AI Lab)。主要研究方向为自然语言处理、信息检索、机器学习等。
lihang.lh@bytedance.com

参考文献

- [1] Proudfoot D. What Turing himself said about the imitation game[J]. *IEEE Spectrum*, 2015, 52(7):42-47.
- [2] 蒲慕明, 脑科学给人工智能带来了什么? [OL].

[2016-08-27]. <https://blog.csdn.net/happytofly/article/details/80125000>.

- [3] The Neuron and Nerve System, Khan Academy.
- [4] Zalta E N, ed. Computational Theory of Mind[M]// Stanford Encyclopedia of Philosophy, 2015.
- [5] Pinker S.The Language Instinct[M].William Morrow and Company, 1994.
- [6] Pinker S.How the Mind Works[M]. W. W. Norton & Company, 1997.
- [7] Zalta E N, ed. The Chinese Room Argument[M]// Stanford Encyclopedia of Philosophy, 2014.
- [8] Zalta E N, ed. Embodied Cognition[M]//Stanford Encyclopedia of Philosophy, 2015.
- [9] Damasio A. Descartes' Error: Emotion, Reason, and the Human Brain[M].Penguin Books, 2005.
- [10]Lakoff G. Cascade Theory: Embodied Cognition and Language from a Neural Perspective[R]. Central European University, 2013.
- [11]Bergen B.Louder Than Words: The New Science of How the Mind Makes Meaning[M].Basic Books, 2012.