

McCulloch-Pitts 人工神经元模型 80 周年纪念：思想、方法与意义

缪青海 王雨桐 吕宜生 那晓翔 王飞跃

In celebration of McCulloch-Pitts ANN model's 80th anniversary: its origin, principle, and influence

MIAO Qinghai, WANG Yutong, LV Yisheng, NA Xiaoxiang, WANG Fei-Yue

引用本文：

缪青海, 王雨桐, 吕宜生, 等. McCulloch-Pitts 人工神经元模型 80 周年纪念：思想、方法与意义[J]. 智能科学与技术学报, 2023, 5(2): 133-142.

Qinghai MIAO, Yutong WANG, Yisheng LV, et al. In celebration of McCulloch-Pitts ANN model's 80th anniversary: its origin, principle, and influence[J]. Chinese Journal of Intelligent Science and Technology, 2023, 5(2): 133-142.

您可能感兴趣的其他文章

平行科研院所：从数字化转型到智能化变革

Parallel scientific research institutes: from digital transformation to intelligent revolution

智能科学与技术学报. 2023, 5(2): 212-221

平行车间调度模型构建和系统设计

Parallel workshop scheduling model and system design

智能科学与技术学报. 2023, 5(2): 254-266

探讨 AI for Science 的影响与意义：现状与展望

The DAOs to AI for Science by DeSci: the state of the art and perspective

智能科学与技术学报. 2023, 5(1): 1-6

元宇宙下的平行交通系统

Parallel transportation systems in era of metaverse

智能科学与技术学报. 2023, 5(1): 32-40

系统农业：结合农业社会经济属性的建模和控制

Systems agriculture: modeling and control based on social and economic attributes of agriculture

智能科学与技术学报. 2023, 5(1): 41-50

平行推理：一种基于 ACP 方法的虚实互动的知识协同框架

Parallel reasoning: a virtual-real interactive knowledge collaboration framework based on ACP approach

智能科学与技术学报. 2023, 5(1): 69-82

McCulloch-Pitts 人工神经元模型 80 周年纪念：思想、方法与意义

缪青海¹, 王雨桐², 吕宜生^{1,2}, 那晓翔³, 王飞跃^{1,2}

(1. 中国科学院大学人工智能学院, 北京 100049;

2. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室, 北京 100190;

3. 剑桥大学工程系, 英国 剑桥 CB2 1PZ)

摘要: 1943 年, Warren McCulloch 和 Walter Pitts 发表论文《神经活动内在思想的逻辑演算》, 证明可以使用逻辑演算描述神经网络的运行机理, 拓展了计算神经科学的领域前沿, 奠定了人工神经网络的发展基础。恰逢 M-P 论文发表 80 周年之际, 结合时代背景和作者学术生涯的变迁, 探索 M-P 理论的思想源泉; 借助原文中的示例, 概述 M-P 模型的基本原理; 以作者对循环因果和反馈机制的研究为主线, 探讨 M-P 模型对信息科学发展的影响、对控制论发展的贡献, 及其在当今人工智能大模型时代的意义。

关键词: M-P 模型; 人工神经网络; 循环因果; 控制论; 平行系统; 人工智能

中图分类号: TP13

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202314

In celebration of McCulloch-Pitts ANN model's 80th anniversary: its origin, principle, and influence

MIAO Qinghai¹, WANG Yutong², LV Yisheng^{1,2}, NA Xiaoxiang³, WANG Fei-Yue^{1,2}

1. School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China

2. State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

3. Cambridge University, Cambridge CB2 1PZ, UK

Abstract: In 1943, Warren McCulloch and Walter Pitts published a paper titled “A logical calculus of the ideas immanent in nervous activity”, which demonstrated that the functioning of neural networks could be described using logical calculus. This expanded the field of computational neuroscience and laid the foundation for the development of artificial neural networks. On the occasion of the 80th anniversary of the publication of the M-P paper, the intellectual origins of the M-P theory was explored, taking into account the historical context and the authors' career trajectories. With the help of examples from the original paper, the basic principles of the M-P model were outlined, and its advantages and limitations were summarized. Furthermore, the impact of the M-P theory on the development of information science was discussed, focusing on the authors' contributions to cybernetics, including their work on circular causality and feedback mechanisms, which provided a foundation for the development of modern artificial intelligence technologies such as parallel intelligence and large-scale models.

Key words: M-P model, neural network, circular causality, cybernetics, parallel system, artificial intelligence

收稿日期: 2023-05-31; 修回日期: 2023-06-10

通信作者: 吕宜生, yisheng.lv@ia.ac.cn

基金项目: 国家自然科学基金项目 (No.62271485, No.61903363, No.U1811463)

Foundation Items: The National Natural Science Foundation of China (No.62271485, No.61903363, No.U1811463)

0 引言

人工智能历史上经历了3次繁荣：第一次是1957年罗森布莱特（Frank Rosenblatt）提出感知器 Perceptron 模型，首次通过神经网络体现出了机器学习的能力；第二次是20世纪80年代初，霍普菲尔德（John Hopfield）引领神经网络的复兴，以罗梅尔哈特（David Rumelhart）等人为代表的PDP学派在1986年重新提出神经网络的反向传播（BP）学习算法，人工智能迎来第二个春天；第三次是2010年前后由辛顿（Geoffrey Hinton）开启的深度学习时代，从2012年的 AlexNet，到2017年谷歌提出 Transformer，再到2022年年底 OpenAI 推出 ChatGPT，人工神经网络经历寒冬，两落两起，终于将人工智能带到了历史的转折点。回顾人工智能曲折的发展历程，人工神经网络在人工智能的每一次萌芽、复苏、繁荣中，都扮演着至关重要的角色。而追溯人工神经网络的源头，是由神经生理学家麦卡洛克（Warren S. McCulloch）与自学成才的数学家皮茨（Walter Pitts）提出的神经网络逻辑演算模型。1943年，McCulloch 和 Pitts 一同发表论文《神经活动内在思想的逻辑演算》^[1-2]（A logical calculus of the ideas immanent in nervous activity）（以下简称 M-P 论文），首次提出通过逻辑演算来解释神经机制如何实现心理功能。人们将这篇论文中提出的神经网络模型，以作者姓氏首字母命名为“M-P 模型”。M-P 模型不但是人工神经网络的源头，而且也启发了计算机、控制论和人工智能等多个学科的开创，打开了现代信息科学殿堂的大门。

然而从今天的学术视角来看，M-P 论文晦涩难懂，对论文的深入分析也不多见^[3]。其原因主要有3点：首先，论文本身是神经生理学和逻辑的融合，学科交叉特点明显，读者需要具备两个学科领域的基础知识；其次，论文中使用的逻辑符号、神经网络的图标均与今天人们所熟悉的形式有较大区别，造成阅读和理解上的困难；第三，M-P 论文只是作者一个时期的工作总结，只有全面了解两位作者的整个研究生涯，以及作者所处的科学研究时代背景，才能对 M-P 模型思想和意义有较为深入的理解。

正值 M-P 论文发表 80 周年之际，本文探索 M-P 模型的思想源头，浅析 M-P 模型的基本原理，讨

论 M-P 模型对后世信息科学的深远影响，希望能引人思考，启发对智能本质的未来探索。

1 M-P 模型的起源

M-P 模型源于 McCulloch 对用神经机制解释心智现象的追求与探索，在与 Pitts 合作不到两年的时间内，便发表了 M-P 论文。图 1 给出了 McCulloch 和 Pitts 职业生涯的主要经历。本节通过介绍两位作者所处的时代背景和关键人物，帮助读者理解 McCulloch 的核心思想以及 Pitts 在其中的贡献。



图1 McCulloch 和 Pitts 职业生涯概览

1.1 麦卡洛克（McCulloch）

McCulloch 出生于 1898 年，幼年开始接受良好的教育，职业生涯相对平稳顺利。在攻读神经生理学和精神病学学位期间，McCulloch 开始思考有形的生物神经系统如何产生无形的心智，即“身体-心智”（body-mind）问题，这在当时是一个几乎空白的领域。McCulloch 提出了“精神原子”（psychon）的概念，它类似于物理学中的原子，McCulloch 认为精神原子是心智的最基本元素。精神原子包含的信息是命题式的（真或假），并且能够在不同精神原子之间按照因果关系时序传递。

1927 年，McCulloch 获得医学博士学位（MD），开始神经生理学职业道路。随后的一年，McCulloch 在纽约一家医院实习，开始对神经系统中的闭环回路（closed loop）产生了兴趣。McCulloch 当时认

为, 人类某些疾病 (例如帕金森病) 的颤抖症状, 可能是由神经闭环回路造成的。神经系统中闭环回路的功能和机理也成为 McCulloch 职业生涯中长期关注的问题。

自 1928 年起, McCulloch 在罗克兰精神病医院工作。在这期间, McCulloch 想到, 每个神经元向其相邻神经元传递的“全有或全无”(all-or-none) 电脉冲信号, 可能对应于他所提出的精神原子; 而神经元之间通过电脉冲信号实现的“兴奋或抑制”相互作用, 对应基于精神原子的命题逻辑推理演算。也就是说, McCulloch 认为大脑是按照逻辑演算方式工作的, 进而实现信息感知能力、数学和抽象思维能力。

1934 年, McCulloch 进入耶鲁大学神经生理学实验室, 跟随著名神经生理学家 Johannes Dussier de Barenne, 完成了关于大脑皮层中闭环回路、皮层功能定位等重要研究。尽管当时的生物学研究进展迅速, 但大多数研究集中在揭示纯生物学意义上的物理化学现象, 而对大脑更高级的活动 (例如学习) 知之甚少, 没有任何理论可以解释心智是如何出现的。

在耶鲁大学, McCulloch 积极学习逻辑, 参加了一系列哲学家的讨论班。但是他所掌握的逻辑知识还不足以支撑他的理论设想, 特别是神经系统中的闭环回路工作机理尚不清楚, 通过逻辑演算难以处理。直到 1941 年他搬去芝加哥, 与 Pitts 合作, 这一问题才有了解决方案。

1.2 拉舍夫斯基 (Rashevsky)

20 世纪 30 年代, 芝加哥大学理论物理学家拉舍夫斯基 (Nicolas Rashevsky) 创建了数学生物学和数学生物物理学。与经典生物学研究侧重实验研究不同, 这一新领域仅借助数学从理论上研究生物行为。Rashevsky 希望将理论物理学的方式应用到生物学领域, 对高度理想化的生物系统进行数学描述。尽管经典生物学家指责这种方法不切实际, 但 Rashevsky 坚持认为理想模型仍然具有了与真实模型一样的核心属性, 通过对问题的简化, 可以逐步找到解决方案, 进而解决现实问题。

Rashevsky 同样关注 McCulloch 考虑的“身体-意识”问题。1936 年, Rashevsky 发表《Mathematical biophysics and psychology》^[4]一文, 尝试基于抽象化的神经元来建立神经网络的数学模型。Rashevsky 的神经网络模型基于以下假设。

- 神经元有两种类型: 兴奋型或抑制型。大脑是由相互联系的神经元组成的, 神经元在受到刺激时会产生兴奋性或抑制性物质, 从而影响网络中与之相连的神经元状态。

- 神经元遵循全有或全无法则: 当接受的兴奋性物质浓度与抑制性物质浓度之比大于 1 时, 神经元才会兴奋。浓度由微分方程基于输入信号推导得出。

- 神经的阈值因先前的兴奋而永久降低 (体现学习能力)。

从以上假设可以看出, Rashevsky 的论文与 7 年后发表的 M-P 论文有许多相似之处。例如两者都采用神经元状态“all-or-none”准则, 神经元根据所受外部输入 (激励或抑制) 的累加结果, 与阈值相比较后决定输出。

不同之处主要有两点: 首先, Rashevsky 模型中, 神经元输出信号的兴奋或抑制性质是固定的, 而在 M-P 模型中, 神经元输出信号是兴奋或抑制, 仅取决于建立连接的位置。这表明 McCulloch 和 Pitts 比 Rashevsky 更远离现实, 在这一点上, 他们的处理方式更简化、更抽象; 其次, 更重要的一点是两者采用的数学工具不同。Rashevsky 模型需要建立并求解微分方程, M-P 模型是基于命题表示与逻辑演算的。从 Rashevsky 模型到 M-P 模型, 微分方程被更简单的逻辑命题所取代, 是从“连续”到“离散 (二值)”、从“模拟”到“数字”的转变。基于逻辑的模型适合用图灵提出的计算理论进行分析, 可与图灵讨论的智能关联起来, 这也是 M-P 模型成为人工智能起源的重要原因。

1939 年, Rashevsky 创建了世界上第一份数学生物学国际期刊, 即《数学生物物理学会刊》(Bulletin of Mathematical Biophysics)。4 年之后, McCulloch 和 Pitts 的论文正是在此期刊上发表的。

1.3 皮茨 (Pitts)

1923 年, Pitts 出生于底特律一个并不重视教育的家庭。性格决定命运, Pitts 的一生常被人们叹息。Pitts 从小自学逻辑与数学, 通晓多门语言。12 岁时, Pitts 自学《数学原理》(Russell B 和 Whitehead A N 著)^[5], 他在图书馆花 3 天时间通读这部数学巨著, 并写信给作者罗素 (Bertrand Russell), 指出其中存在的错误。1938 年, 15 岁的 Pitts 离家出走去芝加哥。受当时在芝加哥大学做访问教授的 Russell 的推荐, Pitts 跟随芝加哥大学的逻辑学家卡尔纳普

(Rudolf Carnap)深造学习。Pitts在课程答疑时间到Carnap的办公室,展示了一本Carnap刚出版不久的逻辑学著作《语言的逻辑语法》^[6](The logical syntax of language),书中写满笔记和批注。这部著作对Pitts影响深远,1943年发表的M-P论文共有3篇参考文献,Carnap的著作是其中的第一篇。Carnap在随后几年时间里指导Pitts继续学习逻辑,并且在大学里为Pitts争取到一份薪水微薄的工作,帮助Pitts开启了逻辑学家之路。

1940年,Pitts在芝加哥大学认识了数学生物学创始人Rashevsky。Pitts定期参加Rashevsky的讨论班,并与其团队一起研究简单神经元回路中的兴奋性和抑制性活动的数学分析方法。由此,Pitts也积累了神经生理学领域的研究基础。

1942年,Pitts与来到芝加哥的McCulloch相识,以深厚的逻辑学功力给后者留下了深刻印象。McCulloch收留了居无定所的Pitts,一同研究神经网络的逻辑模型。Pitts借助《语言的逻辑语法》和《数学原理》中的知识,帮助McCulloch建立了神经原子理论的逻辑表述,并解决了困扰McCulloch多年的神经系统中闭环回路的逻辑演算问题。1943年,这项工作发表在Rashevsky创办的《数学生物物理学会刊》上,文章取名《神经活动内在思想的逻辑演算》(A logical calculus of the ideas immanent in nervous activity),当时Pitts只有20岁。

1.4 图灵(Turing)

1936年,艾伦·麦席森·图灵(Alan Mathison Turing)发表了计算机科学和计算理论领域的开创性论文《论可计算数及其在判定问题上的应用》(On computable numbers, with an application to the entscheidungsproblem)^[7]。论文研究的核心问题是由数学家希尔伯特(David Hilbert)提出的判定问题(entscheidungsproblem),即是否存在一个算法可以判定涉及自然数的逻辑命题的真实性(“真”或“假”)。在论文中,图灵基于人类有限计算和记忆能力的类比,提出一个重要概念——计算机器(computing machine):一种能够执行任何可以用算法描述的计算的理论设备,即“图灵机”。

关于图灵机,有以下几个特点。

- 如果一台自动机器打印两种符号,其中第一种完全由0和1组成(称为数字),第二种包括其他符号(非数字),则该机器被称为“计算机器”。
- 机器从初始配置开始,输出的0和1子序列

被称为机器计算的序列。在这个数列前加一个小数点得到的以二进制表示的实数,被称为机器计算的数。

- 如果一台计算机永远不会写下超过有限数量的第一类符号,则被称为循环的(circular),否则是无环的(circle-free)。如果机器到达无法运行(move)的配置,或者如果它继续运行但只能打印第二种符号(不能再打印第一种符号),则机器将是循环的。

- 可计算序列和数字。如果一个序列可以由无环机器计算,则称该序列是可计算的。如果一个数字与无环机器计算的数字相差一个整数,则该数字是可计算的。

图灵定义了“可计算性”的概念,提供了无法由任何图灵机计算的数学函数的示例,表明可以通过算法计算的内容是有限的。图灵进一步提出了“通用图灵机”(universal computing machine)的概念,即可用于计算任何可计算序列的机器。图灵建立了计算、逻辑和机器之间清晰而严格的联系,对后来的数字计算机和人工智能等信息科学领域的发展起到了至关重要的作用。

可以看出,图灵的工作深刻影响了McCulloch对“body-mind”问题的研究思路。M-P模型中的命题逻辑演算、闭环和无环神经网络,与图灵机中的逻辑表示和计算、循环和无环计算机器的概念可以类比对。1948年的一次公开讨论中,McCulloch认为图灵的论文将他引向了“正确的方向”。正是这个原因,基于逻辑演算的M-P模型成为现代人工智能的源头,而基于微分方程的Rashevsky神经网络模型则逐渐隐匿于文献海洋的深处。

图2总结了M-P模型的重要思想源泉。M-P模型的成功源于McCulloch对“body-mind”这一科学问题的不懈追求,汇聚了Rashevsky、Turing、Russell、Carnap、Pitts等神经生理学家、数学家的智慧,是建立在前人工作基础上的创新探索。M-P模型也对计算机科学、控制论的发展起到关键促进作用。

2 M-P模型原理

M-P模型以神经生理学为基础,以逻辑为数学工具,描述无环和有环神经网络的行为,在有形的大脑与无形的意识之间建立联系。

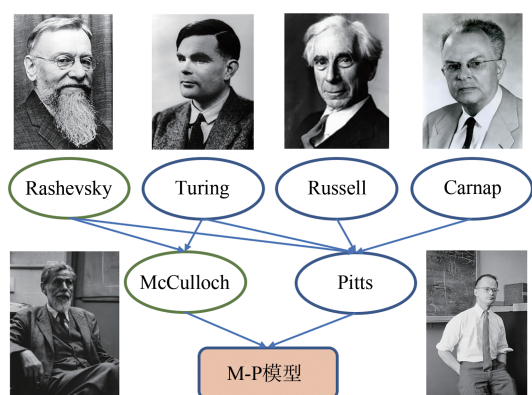


图2 M-P模型思想探源

2.1 M-P模型的神经生理学基础

20世纪40年代,神经生理学家虽然对神经细胞的内部工作机理仍不明确,但他们通过解剖实验和测量方法对神经网络的行为有了直观的了解,形成了较为统一的认识,具体如下。

- 每个神经元的主体结构是胞体和轴突。一个神经元的轴突和另一个神经元的胞体之间,通过突触形成连接。
- 神经系统是一张由众多神经元相互连接构成的网络。
- 在任何时刻,神经元都有某个阈值,外部刺激必须超过该阈值才能引起该神经元的激发产生输出。激发阈值是由神经元自身决定的,而不是取决于外部的刺激。

在由众多神经细胞相互连接组成的网络中,某一神经细胞通过突触接收来自外界的脉冲信号,这些信号在细胞体内累积,当超过一定阈值时,神经细胞通过轴突向外输出脉冲,并通过轴突末端的神经末梢将脉冲传导至与其相连的其他神经细胞。虽然单个神经元细胞的行为相对简单,但众多神经元细胞构成的网络却能产生高级的思想意识,这正是神经网络的神奇之处。

如何用数学来描述神经网络工作的机理,是理论神经生理学家多年来努力探索的问题。McCulloch认为神经元的行为等价于逻辑命题(proposition)的断言,并和Pitts一起首次使用形式语言来建立神经元的模型,运用逻辑演算来刻画神经网络的工作机理。因此,为方便使用逻辑表示与计算,McCulloch和Pitts提出了5条假设:

(1) 神经元的活动是一个“全有或全无”的过程;

(2) 一个神经元的激发(fire),充要条件是有固定数目的突触受到外部刺激,这个数目与此神经元上一次活动模式及受到刺激的突触位置无关;

(3) 神经系统中唯一明显的延迟是突触延迟;

(4) 任何时刻,抑制性突触的活动会绝对阻止神经元的兴奋;

(5) 网络的结构不随时间变化。

第1条假设,本质上是对神经元状态的二值化。Rashevsky模型也是基于同样的假设,是当时神经生理学领域的普遍认识。“all-or-none”原则即神经细胞在受到少量外部刺激时并不会有任何输出,而当外部刺激积累到一定程度并超过某个阈值时,神经细胞转为激发态对外输出脉冲,随后又恢复到正常状态。经过“all-or-none”法则,神经元的状态输出与逻辑命题的断言取值相一致(0或1,真或假),从而为使用逻辑语言表示神经元行为奠定了基础。

第2条假设是对神经元激活条件的简化。实际的神经末梢与突触连接具有可塑性,在外部的刺激下可能会产生临时或永久的改变(对应于网络的学习)。McCulloch和Pitts认为这些时变特性可以通过等效网络来解决,不影响神经活动等价于命题推理这一理论的有效性。

第3条假设可看作神经元激发时间上的离散化。根据神经生理学的发现,不同神经元的轴突长度和直径差异较大,神经元之间的脉冲传导速度也各不相同。但一般来讲,较短的轴突一般直径较小,脉冲传导速度慢;较长的轴突一般直径较大,脉冲传导速度快。因此,McCulloch认为对于不同距离的源和目标点,脉冲通过轴突传导的时间差异并不重要。McCulloch由此认为时间可以离散化,即所有神经元在每个时间步同时更新,任何信号都需要在一个时间步完成从输入到输出的传送过程。基于这个假设,McCulloch和Pitts使用时序命题表达式(temporal propositional expression, TPE)将神经网络输出形式化,进而构建了神经网络模型中逻辑演算的基础。

第4条假设定义了信号的优先级,第5条假设固定了网络结构,方便讨论网络结构与逻辑函数的双向映射。虽经简化,M-P模型在广义上与真实神经网络等价,即M-P模型只保留了两之间等价不变的网络属性,其目的是为逻辑演算提供最大的

便利。

2.2 M-P模型的逻辑符号

在论文中, McCulloch 和 Pitts 采用了 Carnap《语言的逻辑语法》中的符号体系, 并基于 Russell 和 Whitehead 的《数学原理》一书中的各种符号进行了扩充。经过大约一个世纪的发展, 这些经典著作中的逻辑符号已被新的写法取代, 给今天的初学者带来巨大的阅读障碍。因此, 本节对 M-P 论文中的符号和现代常用符号进行了对比 (见表1), 方便进一步对 M-P 模型的基本原理进行分析。

表1 部分 M-P 论文符号与常用逻辑符合对照表

M-P 论文符号	现代常用符号	意义	描述	中文
$\sim$	$\neg$	negation	not	否定、非
$\cdot$	$\wedge$	conjunction	and	合取、与
$\vee$	$\vee$	disjunction	or	析取、或
$\rightarrow$	$\rightarrow$	implication	if ... then ...	蕴含
$\equiv$	$\equiv$	equivalence	if and only if	等价
E	$\exists$			存在量词
S	/	backwards in time	/	时步回退
$\cdot \vdots \vdots \vdots$			() [] { }	分隔符

在 M-P 论文中, 逻辑合取操作符使用 “.”, 而现在常写为 “^”。此外, M-P 论文采用了《数学原理》一书中规定的计算顺序分级表示形式, 即用不同数量的 “.” 表示不同层级分隔符, 例如第一级用 “.” 分隔, 第二级用 “:” 分隔, 第三级用 “:.” 分隔, 以此类推。容易让读者困惑的是, 第一级分隔符与逻辑合取操作符都是 “.”, 而现代逻辑算式中不同层级的区分常用不同形式的括号 (例如 () [] { }) 来表示。S 是作者额外引入的操作符, 表示神经元状态的时间步回退, 即 $S[N_i(t)] = N_i(t-1)$, 用于辅助时序命题表示。

2.3 无环网络

M-P 论文对神经网络模型理论的介绍分为两个部分: 一个是没有闭环回路的无环网络 (nets without circles), 另一个是有环网络 (nets with circles)^[1]。可以看到, 这里使用的术语与图灵 1936 年论文中描述 “图灵机” 时使用的词语一致, 即 circle-free 和 circular computing machine。

M-P 论文首先介绍没有循环回路的网络。McCulloch 和 Pitts 设计了一系列基本的神经网络结构, 如图3所示 (即原文中的图 1a~d), 对应于逻辑运

算中的合取 (与)、析取 (或)、否定 (非) 等基本操作。通过这些基本运算结构的组合, 可以实现更复杂的网络用于表达更复杂的逻辑功能。这些基本网络结构对应于数字计算机中的逻辑门电路。

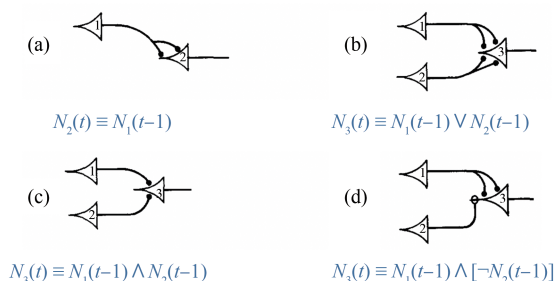


图3 4种基本网络结构与对应的命题逻辑表达

M-P 原文中的图 1-e 是一个相对复杂且特殊的例子, 网络的结构如图4所示。McCulloch 和 Pitts 通过此例说明 M-P 神经网络可以解释有趣的精神现象: 当用一个冰块短时接触皮肤后立刻移开时, 人的感觉是热的; 而当冰块持续接触皮肤时, 人才会产生冰冷的感觉。

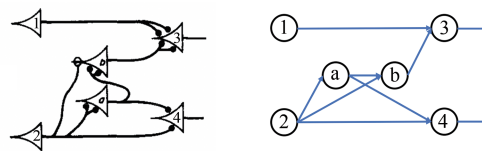


图4 可以解释精神现象的网络结构

对应图4左侧的网络 (原文图 1-e), 原文中的逻辑表达式如下:

$$N_3(t) := N_1(t-1) \cdot \vee \cdot N_2(t-3) \cdot \sim N_2(t-2) \quad (1)$$

$$N_4(t) \cdot \equiv \cdot N_2(t-2) \cdot N_2(t-1) \quad (2)$$

按照表1所示, 用括号替换 “.” 号分隔符, 以上两式可重新表述为:

$$N_3(t) \equiv N_1(t-1) \vee \{ N_2(t-3) \wedge [\neg N_2(t-2)] \} \quad (3)$$

$$N_4(t) \equiv N_2(t-2) \wedge N_2(t-1) \quad (4)$$

图4中, 神经元1和2分别是热接收和冷接收神经元, 神经元3和4分别是热感和冷感神经元。当冰块短时接触皮肤时, 如 $N_3(t)$ 所示, 即热接收或冷接收神经元发射一次然后立即停止, 热感神经元3满足触发条件, 人皮肤感觉是热的。而当冰块持续接触皮肤时, 如 $N_4(t)$ 所示, 神经元2连续激发两次, 隐层神经元会激发冷感神经元4, 而不是热感神经元3, 从而产生冷感。

图4右侧所示为该网络在结构上等价的现代神经网络表示形式,可以看出,变换后的网络结构有3层:神经元1和2是输入层,神经元a和b是隐藏层,神经元3和4是输出层。与ResNet类似,网络连接有跨层(神经元1到神经元3)。

2.4 有环网络

当神经网络中存在闭合回路时,上述逻辑演算(因果推理)不再适用。根据McCulloch和Pitts的分析,脉冲信号将在环路中不停回响(reverberating),逻辑命题的判定则需要依赖过去无穷远的事件^[1]。Pitts凭借其深厚的逻辑功底,在论文中用4篇幅给出了解决方法。这套方法是出了名的晦涩难懂,并未在后续得到普遍认可,有逻辑基础的读者可深入探索Pitts的思想和技巧。

值得注意的是,在介绍有环网络的最后一段,作者将M-P网络与图灵机联系起来,认为M-P模型为图灵可计算数理论提供了神经生理学层面的验证。

2.5 M-P 论文的总结

在论文的结论部分,作者从因果关系的角度阐述了M-P模型的价值和意义。因果关系是在多个学科领域以不同形式体现出来的共有规律。因果关系基于两个要素:事件的状态描述和事件之间关联规则的定义。在M-P模型中,每个神经元都是“全有或全无”(决定了状态)的事件,神经网络的具体形式定义了事件之间联系的法则。因此,M-P神经网络实际上是一个因果推理系统,并且具有单向不可逆的特性。也就是说,给定任一状态,可以通过M-P模型计算出其后续状态,但不能从当前状态反向推知前题(causal,即导致当前状态的前序状态)。这一方面是因为M-P逻辑中存在析取操作(逻辑或 $\vee$),推理的前题存在不唯一性;另一方面是因为神经网络中存在循环,推理的前题存在不确定性。这一结论说明M-P模型能力的有限性,作者认为这与人类在空间和时间上的认知能力的有限性是一致的,并且在现实中有大量的例子佐证。作者进一步分析,这一结论可以为精神病诊疗和心理学研究带来启示。

此外,McCulloch和Pitts指出,生物体中普遍存在一些系统:对输入信息与回路内部特定活动之间的差异做出反应以减小这种差异。这种有目的的行为是维持稳态、欲望和注意力的关键。McCulloch和Pitts在1947年的论文中^[7],进一步深

入研究了这类系统在听觉和视觉感知中的作用机理,并且使用“负反馈”(negative feedback)这一术语来描述这种机制。McCulloch对循环结构、因果关系、反馈机制的持续关注和探索,在1946年开始的梅西控制论会议(Macy Conferences on Cybernetics)的主题上得到了充分的体现,也标志着McCulloch的研究从神经生理学到控制论的转变。

3 M-P 模型的后续影响

M-P论文发表后,并没有在神经生理学领域引起广泛的关注。经典神经生理学家们普遍认为M-P模型过于简化,对于揭示神经系统的工作机理没有价值。反而是在萌芽期的信息科学领域,M-P理论得到了重视,McCulloch作为梅西控制论会议的主席,为M-P理论的推广起到了关键作用。

冯·诺依曼(John von Neumann)将M-P网络中逻辑函数的设计思想推广到计算机体系结构的设计中,奠定了计算机科学的基础。美国数学家丘奇(Alonzo Church)的学生克莱因(Stephen Cole Kleene)深入研究了M-P模型中的有环网络,创建了递归理论(recursion theory)这一数理逻辑分支,发展出有限自动机(finite automata)理论,为理论计算机科学做出了贡献。

维纳(Norbert Wiener)对McCulloch和Pitts的影响更为深刻。1943年,Pitts在好友Lettvin的介绍下来到麻省理工学院(MIT)面见当时著名的数学家维纳。初次见面,Pitts与维纳讨论了遍历定理的证明,给后者留下深刻的印象。随后,Pitts来到波士顿,在维纳的资助下开展研究工作。1950年前后,在维纳的倡导下,MIT电子学研究实验室教授Jerome Wiesner建立了一支神经生理学交叉研究团队,McCulloch、Pitts、Lettvin等都是团队成员。遗憾的是,1952年,研究团队组建之初,维纳与McCulloch关系破裂,McCulloch的团队成员也不得不离开维纳,原本不善社会交际的Pitts受到了巨大的打击。

3.1 Macy 会议

Macy会议是1941年至1960年在纽约召开的一系列会议的统称,由梅西基金会(Josiah Macy Jr. Foundation)资助,旨在促进多学科之间的交叉融合。Macy会议主题广泛,包含但不限于医学、生物、健康等领域,19年间举行了160余场。Macy

会议形式活泼自由,鼓励学者交流正在进行中的研究,因此成为学者分享前沿进展、相互启发灵感的理想场所。其中,从1946年到1953年,由McCulloch作为主席组织了10次会议,通常被称为“梅西控制论会议”,目标是为研究人类思维产生机理奠定科学基础。

第一次梅西控制论会议于1946年3月举行,会议名称为“生物和社会系统中的循环因果和反馈机制”(Circular Causal and Feedback Mechanism in Biological and Social Systems)。参加会议的学者来自数学、物理学、生态学、心理学、人类学、社会学等领域,其中包括冯诺依曼、维纳、香农等今天为人们所熟知的科学家。

非常遗憾的是,由于会议形式灵活(介于自由交谈和学术报告之间),发言没有固定的顺序,前5次梅西控制论会议都没有论文集,甚至没有留下任何文字记录。但根据奥地利裔美国控制论学者福尔斯特(Heinz von Foerster)回忆,前5次会议的主要思想已经体现在维纳1948年出版的经典著作《控制论》之中。

在1949年举行的第6次会议期间,母语为非英语的Foerster向会议委员会提议:用“Cybernetics”一词代替原来冗长拗口的会议名称。Foerster的建议被采纳,并且被任命为会议文集的编辑之一,与人类学家米德(Margaret Mead)女士一同,为人们留下了后5次会议论文集“Cybernetics: Circular Causal and Feedback Mechanism in Biological and Social Systems”,记录了当时参会学者们的闪光思想。论文集成为控制论领域的重要文献。

3.2 循环因果

McCulloch从学生时代开始关注神经系统的循环回路,此后一生都在探索神经闭环的工作机理。1939年的论文中,McCulloch和Dusser de Barenne通过实验发现在感觉皮层和丘脑之间存在神经元闭环。1943年的M-P论文中,McCulloch将其逻辑神经网络等同于因果推理系统,但是如何表示神经循环中的因果关系(circular causality)仍是困扰他的难题。

同是1943年,Rosenbluth、Wiener和Bigelow发表论文《行为、目的和目的论》(Behavior, purpose and teleology)^[9-10]。作者探讨了目的论的概念,即对自然系统中的目标、目的和设计的研究。他们认为,专注于因果关系的传统机械方法不足以完

全理解复杂的生物和行为现象,理解行为不仅需要考虑“输入-输出”关系,还需要考虑形成有机体行为的“反馈机制”和目标导向过程。作者讨论了目的论概念如何应用于广泛的现象,从简单的反应到复杂的决策过程。他们强调了将有机体作为一个整体来考虑的重要性,其内部状态和外部环境会影响其行为。通过引入反馈回路和控制机制的概念,作者为控制论和系统论的研究奠定了基础。

1947年,McCulloch和Pitts合作发表另一篇论文《How we know universals: the perception of auditory and visual forms》^[8],研究人类听觉和视觉中的普适(universals)感知机制。他们发现,视觉系统会根据视角中心和图像中心的差异来调整眼睛位置。然而,该系统的神经网络部分并不会简单地向眼部肌肉发送一组脉冲,从而立即使物体在视野里居中。相反,视觉系统使用不断的“反馈”来缩短图像中心和视野中心之间的距离。McCulloch意识到,这正是一个循环因果系统,“反馈”的概念是一种生物体中普遍存在的重要机制,这比1943年的M-P论文中关于神经循环的认识更进一步。

梅西系列会议拓展了McCulloch关于循环因果的认识。从会议的名称“Circular Causal and Feedback Mechanism in Biological and Social Systems”可以看出,人们已经认识到循环因果和反馈是生物和社会系统中普遍存在的机制,是重要的研究课题。

1964年,美国控制论学会(American Society for Cybernetics, ASC)在华盛顿特区成立,致力于促进控制论这一跨学科领域的新发展,McCulloch当选为第一任ASC主席。

循环因果和反馈机制是控制论的核心思想,图5给出了形象化表示。其中图5(a)是具有古老历史的衔尾蛇,在控制论中表示信息闭环反馈。图5(b)是美国控制论学会时任主席米德设计的



(a) 衔尾蛇 (b) 米德为美国控制论学会设计的图标

图5 循环因果和反馈机制的形象化表示

学会图标,旨在展示两个不同领域之间的循环,是对一般闭环反馈的扩展。其中飞龙代表着以生物学为基础的有形世界,蝾螈代表无形的思想世界。两者相互影响,互为因果。

3.3 平行系统

控制论在 20 世纪七八十年代进入低谷,但循环因果和反馈机制的思想不会磨灭。1994 年,王飞跃提出影子系统 (shadow system)^[11],并于 1999 年在中国科学院创建复杂系统与复杂性科学实验室,研究控制论、社会与物理系统^[12-13]。2004 年由王飞跃提出的平行系统 (parallel systems) 理论^[14-15],体现了循环因果和反馈机制在新时代的延伸与扩展。如图 6 所示,平行系统是指由某一个自然的现实系统和对应的一个或多个虚拟或理想的人工系统组成的共同系统。其中,人工系统是对实际系统的软件化定义,不仅是实际系统的数字化“仿真”,也为实际系统运行提供可替代版本。平行系统的主要目的:通过实际系统与人工系统的相互连接,对二者之间的行为进行实时的动态对比与分析,以虚实互动的方式,完成对各自未来的状况的“借鉴”和“预估”,人工引导实际,实际逼近人工,达到有效解决方案以及学习和培训的目的。

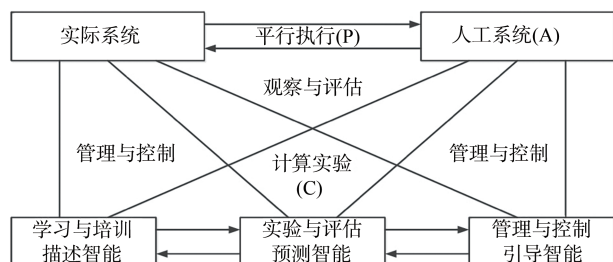


图6 平行系统原理

3.4 ChatGPT与RLHF

当前,ChatGPT的成功推动人工智能进入大模型时代,循环因果和反馈机制是ChatGPT中指令学习和对齐训练的核心思想。具体来说,以ChatGPT为代表的大模型具备了一定的通用人工智能的能力,然而要想大模型的输出符合人类的伦理、道德和法律规范,必须对大模型的输出进行引导和控制^[16-17]。一方面,基于人类反馈的强化学习 (reinforcement learning from human feedback, RLHF),本质上是在大模型和人类之间建立语言上的闭环回路,通过负反馈校正大模型,使其生成的内容符合

人类社会的规范。另一方面,大语言模型基于人类海量文本资料训练而来,是对现实世界的描述,与现实世界一同构成平行系统。平行智能体现出的循环因果和反馈控制的思想,为大模型时代的人工智能研究提供了研究框架^[18-19]。

4 结束语

虽然 80 年已经过去, M-P 论文对当前人工智能的研究仍然有很多启示。

M-P 模型是对神经生理学机制的简化, McCulloch 和 Pitts 设计的神经网络甚至还不具备学习的能力,但其价值在于神经元数学建模方法的拓展,以及通过人工神经网络的设计可以实现逻辑运算的论证。这项工作的意义在于,它为理解神经网络如何处理信息提供了理论基础,为神经计算和人工智能领域的进一步发展铺平了道路。

同时,我们敬佩作者(特别是 McCulloch)在“body-mind”这一人类智能本质问题上的勇敢探索和不懈努力,他们汇聚罗素、卡尔纳普、图灵、拉舍夫斯基等众多先驱的智慧,形成了 M-P 模型这一里程碑式的工作。McCulloch 对神经网络中闭环回路机理的不断研究和探索,以及作为主席召开的 10 届梅西会议,为控制论的创建做出了突出贡献。即使在今天看来,循环因果和反馈机制仍然是人工智能研究的核心思想,对于新时期的人工智能发展具有不可忽视的重要意义。

参考文献:

- [1] MCCULLOCH W S, PITTS W. A logical calculus of the ideas immanent in nervous activity[J]. The Bulletin of Mathematical Biophysics, 1943, 5(4): 115-133.
- [2] 王飞跃. 从叙述科学到叙述哲学: 王浩学术生涯的启示与意义[C]// 哲学与科学论坛: 逻辑、数学与哲学——王浩诞辰百年纪念. 出版地不详: 出版者不详, 2021.
WANG F Y. From narrative science to narrative philosophy: the inspiration and significance of Wang Hao's academic career[C]// Proceedings of Philosophy and Science Forum: Logic, Mathematics, and Philosophy-Commemorating the 100th Anniversary of Wang Hao's Birth. [S.l.: s.n.], 2021.
- [3] PICCININI G. The first computational theory of mind and brain: a close look at mcculloch and pitts's "logical calculus of ideas immanent in nervous activity"[J]. Synthese, 2004, 141(2): 175-215.
- [4] RASHEVSKY N. Mathematical biophysics and psychology[J]. Psychometrika, 1936, 1(1): 1-26.
- [5] WHITEHEAD A N, RUSSELL B. Principia mathematica[M]. Array Cambridge, Eng.: The University Press, 1925.
- [6] USHENKO A P, CARNAP R. The logical syntax of language[J]. The

Philosophical Review, 1936, 46(5): 549-553.

- [7] TURING A. On computable numbers, with an application to the entscheidungsproblem[M]//The Essential Turing. Oxford University Press, 1936.
- [8] PITTS W, MCCULLOCH W S. How we know universals the perception of auditory and visual forms[J]. The Bulletin of Mathematical Biophysics, 1947, 9(3): 127-147.
- [9] ROSENBLUETH A, WIENER N, BIGELOW J. Behavior, purpose and teleology[J]. Philosophy of Science, 1943, 10 (1): 18-24.
- [10] WANG F Y. From norbert wiener to karl popper: a journey of parallel cybernetics in three worlds[C]// Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC2015). Piscataway: IEEE Press, 2015.
- [11] WANG F Y. Shadow systems: a new concept for nested and embedded co-simulation for intelligent systems[D]. Tucson, Arizona State: University of Arizona, 1994.
- [12] WANG F Y. CAST Lab: a cyber-social-physical approach for traffic control and transportation management[R]. 1999.
- [13] WANG F Y. Forward to the past: CASTLab's cyber-social-physical approach for ITS in 1999[J]. IEEE Intelligent Transportation Systems Magazine, 2023, 16(4): 1-4.
- [14] 王飞跃. 人工社会、计算实验、平行系统: 关于复杂社会经济系统计算研究的讨论[J]. 复杂系统与复杂性科学, 2004, 1(4): 25-35.
WANG F Y. Artificial societies, computational experiments, and parallel systems: a discussion on computational theory of complex social-economic systems[J]. Complex Systems and Complexity Science, 2004, 1(4): 25-35.
- [15] 王飞跃. 平行系统方法与复杂系统的管理和控制[J]. 控制与决策, 2004, 19(5): 485-489, 514.
WANG F Y. Parallel system methods for management and control of complex systems[J]. Control and Decision, 2004, 19(5): 485-489, 514.
- [16] WANG F Y, ZHANG J J, ZHENG X H, et al. Where does AlphaGo go: from church-Turing thesis to AlphaGo thesis and beyond[J]. IEEE/CAA Journal of Automatica Sinica, 2016, 3(2): 113-120.
- [17] WANG F Y, MIAO Q H, LI X, et al. What does ChatGPT say: the DAO from algorithmic intelligence to linguistic intelligence[J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10(3): 575-579.
- [18] WANG F Y. Parallel intelligence in metaverses: welcome to Hanoi![J]. IEEE Intelligent Systems, 2022, 37(1): 16-20.
- [19] MIAO Q H, ZHENG W B, LV Y S, et al. DAO to HANOI via DeSci: AI paradigm shifts from AlphaGo to ChatGPT[J]. IEEE/CAA Journal

of Automatica Sinica, 2023, 10(4): 877-897.

[作者简介]



缪青海 (1979-), 男, 中国科学院大学人工智能学院副教授, 主要研究方向为智能系统、平行智能和机器学习。



王雨桐 (1994-), 女, 中国科学院自动化研究所复杂系统管理与控制国家重点实验室助理研究员, 主要研究方向为计算机视觉、图像异常检测。



吕宜生 (1983-), 男, 中国科学院自动化研究所研究员, 主要研究方向为人工智能、机器学习、深度学习、智能驾驶、智能交通和交通大数据。



那晓翔 (1984-), 男, 英国剑桥大学工程系长聘助理教授, 博士生导师。主要研究方向为驾驶员—车辆闭环系统动力学、智能车载信息系统的设计与开发、重型商用车节能减排特性的分析与评价。



王飞跃 (1961-), 男, 中国科学院自动化研究所复杂系统管理与控制国家重点实验室主任, 主要研究方向为平行系统的方法与应用、社会计算、平行智能和知识自动化。