

人工智能联结主义学派的产生与发展

张菁岑 陈洁

天津师范大学教育学部

摘要：联结主义又称仿生学或生理学派，是一种整体论的哲学观点，将计算机视为人脑，基于神经网络和联结机制的学习算法，属于非符号化的人工智能学派。本文追溯了联结主义的孕育时期，从Hebb规则出发，描述了在心灵哲学和认知科学领域的兴起。同时，回顾了20世纪80年代的联结主义研究，以及在心理学和认知科学中的应用。在形成时期，联结主义的前身可追溯到休谟哲学，其思想与现代联结主义相契合。文章从休谟的心智理论中解析了联结主义的根源，探讨了人工神经网络的发展历程，以及隐藏层的关键作用。在发展时期，深度神经网络的崛起使联结主义在21世纪迎来新的高潮，尤其在深度学习领域取得显著成就。文章还强调了计算能力、大数据、硬件支持等因素对联结主义发展的促进作用。

关键词：联结主义；神经网络；深度学习；计算能力；大数据

【DOI】 10.12252/j.issn.2096-627X.2023.07.074

一、孕育时期

在符号人工智能取得显著进展之前，众多研究人员深受大脑复杂结构的启发，努力构建能够模拟智能行为的数学模型。他们认为，如果以大脑为蓝本，那么通向创造人工大脑的合理途径就是尽可能地模拟其工作机制，从而观察能够创造出怎样的人工智能。这一研究方向的基本前提是，神经元是大脑处理信息的最基本单元，因此，只要能构建出一个合理模拟神经元间相互连接的模型，就有可能实现人工智能。这种思想后来发展成为人工智能领域中的联结主义，为其奠定了理论基础。

1949年，神经心理学家D. O. Hebb发表了《the Organization of Behavior》一书，他在书中认为心理过程背后有两个主要的神经生理过程。根据Hebb的观点，神经元动作电位中的电活动负责认知过程，例如语言识别、语言生成、运动技能、感知和记忆回忆。他还假设，相邻神经元的同时或连续活动会增加一个神经元的电活动对另一个神经元的影响。这一观点现在被称为“希布规则”。这种神经生理学上的促进作用在心理学上的体现就是学习或记忆的形成，哲学和神经科学发展的结果就是当前我们所熟知的人工智能的联结主义学派^[1]。

20世纪80年代，在心灵哲学与认知科学领域出现了一个十分重要的概念——联结主义（Connectionism）。豪杰兰德（Haugeland）认为，联结主义一直“占据着兴奋点，所有闪光都在这里。”如今，联结主义已经被教条式地广泛应用于心理学、认知科学等领域^[2]。同时，它也被认为是人工智能研究的新曙光。但是，正如著名心灵哲学家威廉·贝希特尔（W·Bechtel）所认为的那样，“联结主义的诞生构成

了认知科学发展中的重大事件，我希望人们都明白：它也提出了许多适合于哲学考察的问题。”

联结主义主张，智能源自人脑的独特结构。它深入研究人类大脑的神经活动方式、结构特性，以及大脑处理信息的具体过程和机制，旨在构建能够模拟人类智能的系统。通过模仿人脑的结构和功能，联结主义期望人工智能能够达到与人类相似的智能水平。这种模式强调整体性思维，通过技能建模来模拟人脑的特性，并利用统计学方法来模拟神经网络及其网络间的连接机制，从而实现人工智能对人类智能的模拟和逼近。

二、形成时期

最早从事联结主义的是神经生物学家麦卡洛克（W·S·McCulloch）和数理逻辑学家皮茨（W·Pitts），1943年在《数学生物物理学通报》（Bulletin of mathematical biophysics）上合作发表了《神经系统中蕴涵的思想的逻辑演算》（A Logical Calculus of Ideas Immanent in Nervous Activity）一文，该文被称为联结主义研究开端，并首次提出了“神经网络”（Neural Network）的概念。他们在这一文中证明，任何逻辑表达都可以通过适当的简化神经网络来“实现”。他们假设每个“神经元”都是二进制的，并具有有限的阈值，每个“突触”都是兴奋性或者抑制性的，并引起有限的延迟（一个周期），而且可以在任意一对节点之间去构建具有多个突触的网络^[3]。神经网络最早的模型是以神经元为最小单位的数学模型（M-P模型）。该模型把神经元的工作过程简化为非常直接、基础的运算模型。M-P模型激发了人们生物大脑通过运用逻辑机械地解释人类大脑思考过程的可能。M-P模型开启了神经科学理论研究的时代。20世纪60年代至70年代，由于神经网络研究自身的局限性致使

其研究陷入低谷。以明斯基（Minsky）等代表人物在1969年编写的《感知器》，以批评的观点和强大的影响力，直接导致神经网络的研究进入萧条期。到20世纪80年代，神经网络进入了一个新的发展时期，代表人物成果包括：鲁梅尔哈特（Rumelhart）和麦克莱兰（McClelland）等人提出的多层前向神经网络的BP学习算法，以及霍普菲尔德（J·Hopfield）提出的霍普菲尔德神经网络模型，这些都成为神经网络向前发展的重要推动力量。

在20世纪80年代末至90年代初，人工神经网络的研究达到了一个高峰，但随后却迅速陷入了低谷。这一衰落的主要原因之一是人工神经网络在发展过程中遭遇了严重的挫折。研究人员发现，随着网络层数的不断增加，初始几层的参数对最终网络输出结果的影响变得微乎其微，这导致了整个网络训练过程的收敛性无法得到保证。然而，在2006年，人工神经网络深度学习领域迎来了重大的突破，成功解决了网络训练过程中的收敛性问题。这一突破主要归功于两个方面的因素：首先，计算能力得到了前所未有的增强，这使得处理更深层次、更复杂的网络结构成为可能；其次，数据的数量也获得了空前的扩张，为深度学习提供了丰富的学习材料。这些进步共同推动了人工神经网络深度学习的复苏和快速发展，为后来的研究和应用奠定了坚实的基础。

人工神经网络是一种基于生物神经网络（动物大脑）的结构和功能的AI算法，可以应用于高级监督式、非监督式或强化式学习。这种网络通过模拟生物神经网络的工作方式，能够处理大量的输入数据，并通过复杂的计算过程产生相应的输出^[4]。

在生物大脑中，记忆和学习是通过神经元之间的连接和交互实现的。人工神经网络也借鉴了这一机制，通过调整神经元之间的连接权重来存储和更新信息。这些权重在训练过程中根据输入数据和学习目标进行不断调整，使得人工神经网络能够逐渐适应不同的任务和环境。

尽管人工神经网络在处理某些任务方面已经展现出了惊人的能力，比如在图像识别、语音识别和自然语言处理等领域，但是与高级动物的大脑相比，它们仍然显得相对简单和原始。目前的人工神经网络通常只涉及几千个到几百万个神经元，而人类大脑则由数十亿个神经元组成，每个神经元都与其他神经元有着广泛的连接。

人工神经网络通常包括三个不同的层级：输入层、隐藏层和输出层。输入层负责接收外部输入的数据或刺激，隐藏层通过一系列复杂的计算过程对输入进行处理，并最终输出结果。这些隐藏层通常由多个不同的神

经元组成，它们共同参与计算并产生输出。

在机器学习领域，人工神经网络通过不断调整其连接权重来优化性能，这一过程依据输入数据和学习目标的变化而进行。学习方式可以是监督式的或非监督式的。在监督式学习中，网络通过学习带有标签的样本数据特征来精细调整权重；而在非监督式学习中，网络则在没有先验标签的情况下，依靠自我学习和数据聚类来揭示潜在的数据结构和模式。

隐藏层构成了人工神经网络的核心组成部分，它们在计算中扮演着关键角色，并负责从输入数据中萃取和存储关键特征和模式。然而，隐藏层的非透明性也引发了一系列问题和挑战。具体来说，由于隐藏层的内部工作机制对人类来说是不透明的，这导致我们难以洞悉网络如何做出具体决策。这种不透明性不仅增加了人工神经网络决策过程的不可预测性，还削弱了其决策的可靠性。

为了应对这些挑战，研究人员正在积极探索提升人工神经网络透明度和可解释性的新技术和方法。例如，一些研究致力于构建可解释的AI模型，这些模型能够提供更加直观的解释和可视化工具，从而帮助人类更深入地理解网络的决策逻辑和依据。

综上所述，人工神经网络作为一种模拟生物神经网络结构和功能的AI算法，在众多领域已经取得了显著进展。尽管它们在处理复杂任务和大规模数据方面展现出了强大能力，但与高级动物的大脑相比，仍存在显著差距。未来的研究将致力于进一步提升人工神经网络的性能、可解释性和可靠性，以更好地应对现实世界的挑战和问题。简单人工神经网络由图1所示。

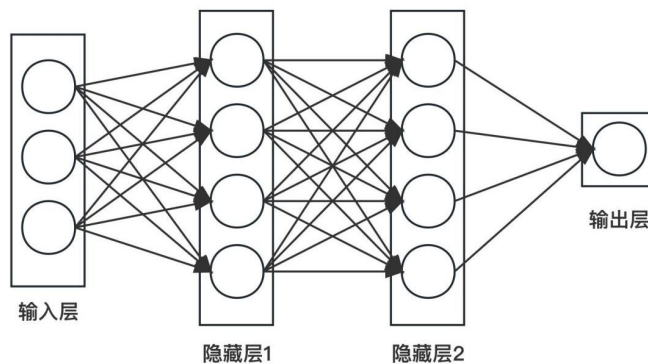


图1 包含两个隐藏层的典型的简单人工神经网络图示

三、发展时期

在21世纪，深度神经网络和深度卷积网络在ImageNet识别实验中的成功，为模式识别和图像处理等领域带来了前所未有的优势，从而掀起了联结主义的新浪潮，为人工智能的发展指明了新的方向。如今，深度人

工神经网络的层次不断加深，涉及的参数数量已高达数十亿，所需的训练样本也动辄数千万甚至上亿。这种庞大的计算需求，使得深度学习必须依赖大规模、高性能的计算机集群，并可能需要数月的时间来完成训练。

随着计算机芯片技术，特别是图形处理单元（GPU）的显著进步，深度学习得到了前所未有的硬件支持。这种技术的迅猛发展不仅为复杂的计算任务提供了动力，而且为人工智能（AI）领域的创新铺平了道路。深度学习的效能与可用数据的数量、多样性和完整性紧密相连。互联网的广泛普及和智能手机的普遍使用已经产生了海量的结构化、半结构化和非结构化数据。这些数据被有效地采集、上传至云端，并进行集中存储和处理，从而为深度学习提供了稳固的数据基础，极大地推动了AI应用的发展和进步。

在深度学习的框架下，人工神经网络的学习方法可以大致分为有监督学习和无监督学习两种。有监督学习依赖于训练数据与其对应的理想输出之间的偏差来调整网络中神经元的权重，以使得实际输出逐渐逼近理想状态，它们专注于解决特定领域的问题，而不具备跨领域的推理能力和通用性。

相比之下，无监督学习则赋予AI系统更多的自主性。在这种学习方法中，AI被赋予一定的初始能力，然后通过与环境的交互进行自主学习和适应。研究者已经开发出了多种无监督深度学习算法，包括强化学习、进化计算和深化算法等。这些算法的目标是让机器学习技能而非单纯的知识，从而使AI具备解决未知问题的能力和一定程度的通用性。基于神经网络的多目标演化优化算法就是这一方向上的重要成果之一。

而目前有一种神经网络模型，它的名字叫作分布式并行处理（PDP），平时我们所说的神经网络大多都是以这个模型为例的。这个模型有四个主要的优点。第一，它只需要通过显示实例，而不需要通过精确的编程来学习模式以及各个模式之间的关系。第二，它能够容忍那些模糊、混乱的迹象，解决约束满足问题，理解部分冲突的迹象。它不需要被严格定义，而是处理那些具有家族相似性的、重叠的集合，这也是人类概念的一个主要特征。第三，它具有识别不完整或部分损坏模式的本领。也就是说，它具有根据记忆内容进行识别的能力。例如，人们可以通过听到几个音符来识别一段旋律，或者通过听到一段旋律的几个音符来识别整段旋律，即使这段旋律有几个音符被弹错了。第四，这种神经网络模型非常健壮。即使是一些节点被丢失，它也不会完全失效，它的性能可能会适度地退化，但在这个过程中，它的性能随着损害的增加而逐渐变差。与符号程序

相比，它们更加健壮^[5]。

深度学习是一种机器学习方法，它的发展经历了多个重要阶段，从早期的概念形成到现代广泛的应用。深度学习的历史可以追溯到20世纪40年代和50年代的早期神经网络研究。在这个时期，研究者开始尝试构建神经网络模型，尤其是感知器，用于模拟生物大脑的神经元工作。然而，这一时期的神经网络受到了严重限制，因为它们只能处理浅层的网络结构。

四、未来与展望

尽管联结主义在人工智能领域已经取得了令人瞩目的成就，但它仍然面临一系列挑战。其中，神经网络的可解释性仍然是一个亟待解决的问题。由于深度学习模型的黑盒特性，我们对其决策过程的理解受到了限制。此外，随着计算能力的不断提升，虽然推动了联结主义的快速发展，但也带来了能源消耗和硬件成本方面的挑战。

在未来的研究中，研究者们不仅需要关注模型性能的提升，还需要更加重视模型的可解释性和可持续性。只有这样，我们才能更好地理解和信任人工智能系统，从而更广泛地应用它们来解决实际问题。

总的来说，联结主义作为一种整体论的哲学观点，在人工智能领域的发展中起到了至关重要的作用。从早期的Hebb规则到神经网络模型的形成，再到如今深度学习技术的蓬勃发展，联结主义一直在不断演进和进步。随着人工智能需求的不断增长，我们相信联结主义有望在未来更多领域中展现其巨大潜力，并为人类社会带来更多的智能化应用和价值。

参考文献

[1] Pollack, J. B. Connectionism: Past, Present, and Future[J]. *Artificial Intelligence Review*, 1989, 3(1), 3-20.

[2] 赵泽林. 联结主义范式的论证及其反思[J]. *科学技术哲学研究*, 2011, 28(02): 61-65.

[3] AIZAWA, K. Connectionism and artificial intelligence: history and philosophical interpretation[J]. *Journal of Experimental & Theoretical Artificial Intelligence*, 1992, 4(4), 295-313.

[4] 韦恩·霍姆斯, 玛雅·比利亚克等著. 教育中的人工智能: 前景与启示[M]. 冯建超, 译. 上海: 华东师范大学出版社, 2021: 96-98.

[5] 章钧津, 田永红, 宋哲煜, 郝宇峰. 神经机器翻译综述[J]. *计算机工程与应用*, 1-19.