



人工智能是否终将超越人类智能

——基于机器学习与人脑认知基本原理的探讨

西安交通大学人工智能与机器人研究所教授 龚怡宏

【摘要】谷歌公司的AlphaGo与韩国棋圣李世石的人机围棋巅峰对决，以AlphaGo 4比1的压倒性胜利落下帷幕。这个比赛结果不仅向全世界展示了人工智能的强大实力与巨大应用潜力，也在人类社会引起了不小的恐慌和忧虑。人们开始认真思考：人工智能将会给人类社会带来哪些进步及挑战？机器智能最终会否超越人类智能？要想回答这些问题，我们首先需要了解人工智能的本质及其基本原理。本文首先对人工智能领域里最受世人瞩目的研究成果——深度学习卷积神经网络——做一个简单描述，进而围绕机器学习的本质及其基本原理进行探讨。接下来，通过对人脑认知机理最新研究成果的概括介绍，揭示机器智能与人类智能的本质差异，比较两种智能的优劣。通过优劣势比较，试图找出针对上述问题的答案。

关键词】人工智能 AlphaGo 神经突触 机器学习 模式识别

【中图分类号】TP18 **【文献标识码】**A

【DOI】10.16619/j.cnki.rmltxsqy.2016.07.002

2016年3月9~15日，谷歌公司研发的AlphaGo围棋软件与韩国棋圣李世石进行了五场人机对决，AlphaGo以4比1的比分取得了压倒性的胜利。这个比赛结果不仅震惊了整个围棋界，也让人工智能领域的许多专家学者跌破眼镜，更让人工智能走出象牙塔，成为许多普通百姓茶余饭后的热点话题。这场人机围棋巅峰对决不仅向全世界展示了人工智能的强大实力与巨大应用潜力，也在人类社会引起了不小的恐慌和忧虑。人们开始认真思考以下这些问题：机器智能最终会超越人类智能吗？人工智能将会如何改变人类社会？未来的智能机器会像电影《终结者》里所描述的那样试图主宰人类、甚至消灭人类吗？要想回答这些问题，我们首先需要了解人工智能的本质及其基本原理，进而讨论其发展的规律和前景。当前，人工智能领域最前沿的分支学科当属机器学习分支。本文首先对机器学习分支中最受世人瞩目的研究成果——深度学习卷积神经网络——做一个简单综述，进而围绕机器学习的本质及其基本原理进行探讨。接下来，通过对人脑认知机理最新研究成果的概括介绍，揭示机器智能与人类智能的本质差异，比较两种智能的优势与劣势。通过机器智能与人类智能的优劣势比较，试图找出上述几个问题的答案。

龚怡宏，西安交通大学人工智能与机器人研究所教授、博导，中组部“千人计划”教授，国家973项目首席科学家，视觉信息处理国家工程实验室首席科学家。研究方向为多媒体内容分析、机器学习、模式识别。主要著作有主要著作有《Intelligent Image Databases - Towards Advanced Image Retrieval》《Machine Learning for Multimedia Content Analysis》等。

深度学习卷积神经网络

过去几年里，深度学习卷积神经网络所取得的成就足以使它成为人工智能王冠上最光彩夺目的明珠。基于深度学习卷积神经网络的语音识别系统把语音识别的精度提高到了产品级的精度，从而为人类与计算机及各种智能终端之间提供了一种崭新的、更为便捷的交互方式。将深度学习卷积神经网络应用于图像内容及人脸的识别，科学家们取得了能够与人类视觉系统相媲美的识别精度。战胜韩国棋圣李世石的谷歌围棋软件AlphaGo能够取得如此辉煌的战绩，深度学习卷积神经网络也发挥了关键性的作用。接下来，我们对深度学习卷积神经网络的起源及其原理做一个简单介绍。

脑神经科学领域的大量研究表明，人脑由大约 10^{11} 个神经细胞及 10^{15} 个神经突触组成，这些神经细胞及其突触构成一个庞大的生物神经网络。每个神经细胞通过突触与其它神经细胞进行连接与信息传递。当通过突触所接收到的信号强度超过某个阈值时，神经细胞便会进入激活状态，并通过突触向上层神经细胞发送激活信号。人类所有与意识及智能有关的活动，

都是通过特定区域神经细胞之间的相互激活与协同工作而实现的。

早于1943年，美国心理学家W.S. McCulloch和数学家W. A. Pitts就在他们的论文中提出了生物神经元的计算模型(简称M-P^①模型)，为后续人工神经网络的研究奠定了基础。M-P模型的结构如图1(a)所示，它包含n个带有权重的输入，一个输出，一个偏置b和一个激活函数 $f(\cdot)$ 组成。n个输入代表来自下层n个神经突触的信息，每个权重 W_i 代表对应突触的连接强度，激活函数 $f(\cdot)$ 通常采用拥有S-型曲线的sigmoid函数(参见图1(b))，用来模拟神经细胞的激活模式。

早期的人工神经网络大都是基于M-P神经元的全连接网络。如图2所示，此类网络的特点是，属于同一层的神经元之间不存在连接；当前层的某个神经元与上一层的所有神经元都有连接。然而，人们很快发现，这种全连接神经网络在应用于各种识别任务时不但识别精度不高，而且还不容易训练。当神经网络的层数超过4层时，用传统的反向传递算法(Back Propagation)训练已经无法收敛。

1983年，日本学者福岛教授基于Hubel-Wiese的视觉认知模型提出了卷积神经网络计

图1 M-P神经元计算模型与激活函数

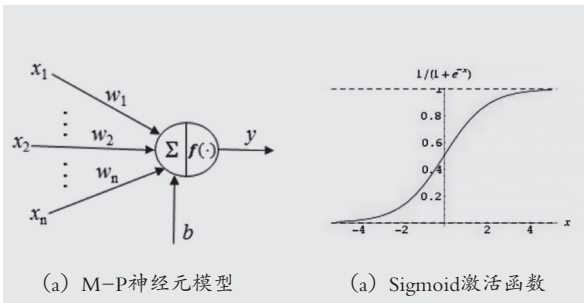
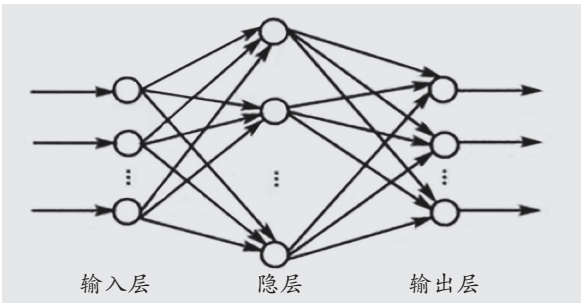


图2 全连接神经网络架构





算模型 (Convolution Neural Network, 简称 CNN)。早在1962年, Hubel和Wiesel通过对猫视觉皮层细胞的深入研究, 提出高级动物视觉神经网络由简单细胞和复杂细胞构成 (如图3所示)。神经网络底层的简单细胞的感受野只对应视网膜的某个特定区域, 并只对该区域中特定方向的边界线产生反应。复杂细胞通过对具有特定取向的简单细胞进行聚类, 拥有较大感受野, 并获得具有一定不变性的特征。上层简单细胞对共生概率较高的复杂细胞进行聚类, 产生更为复杂的边界特征。通过简单细胞和复杂细胞的逐层交替出现, 视觉神经网络实现了提取高度抽象性及不变性图像特征的能力。

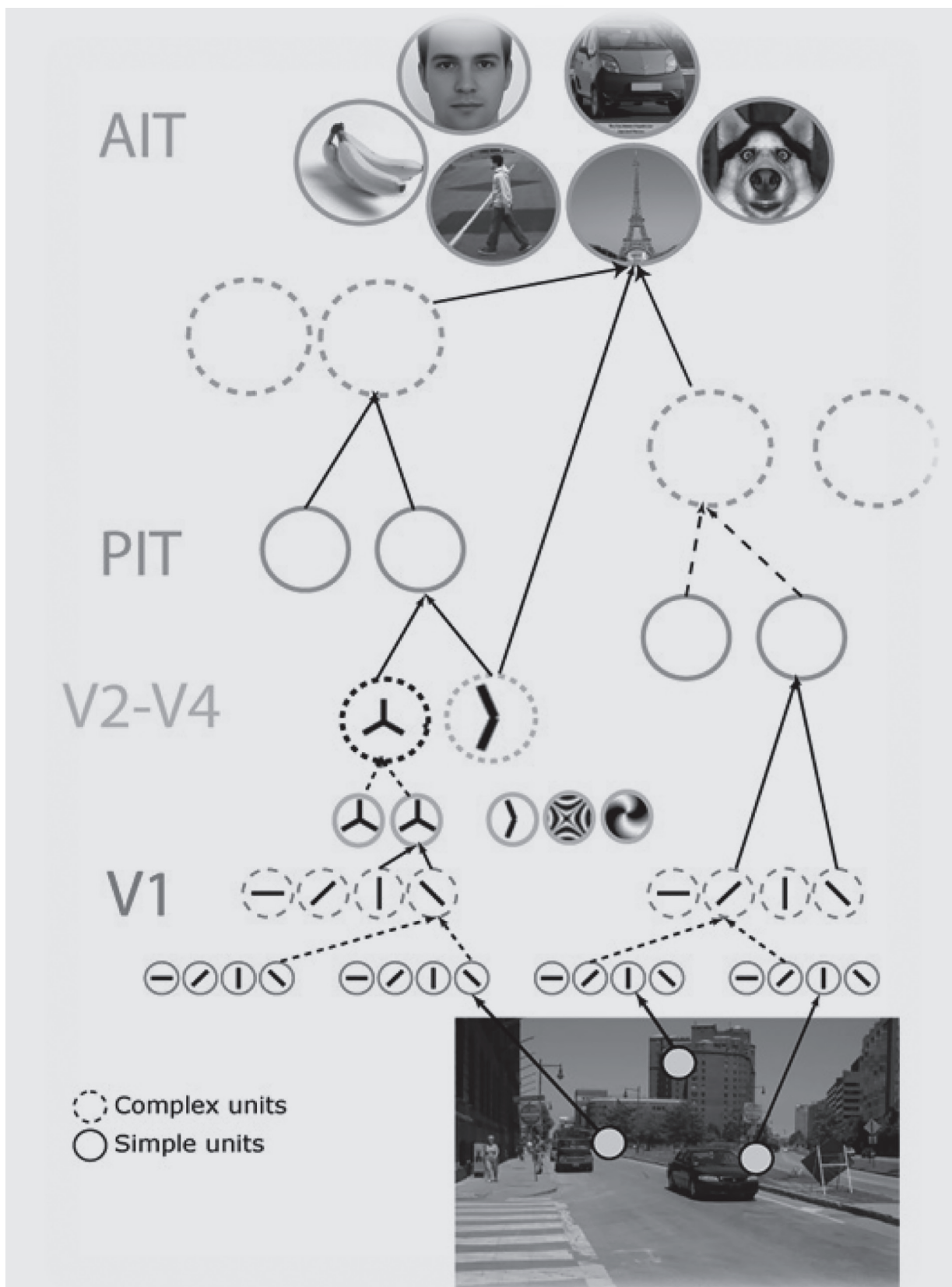
卷积神经网络可以看作是实现上述 Hubel-Wiesel视觉认知模型的第一个网络计算模型。如图4所示, 卷积神经网络是由卷积层 (Convolution Layer) 与降采样层 (Sampling Layer) 交替出现的多层神经网络, 每层由多个将神经元排列成二维平面的子层组成 (称为特征图, Feature Map)。每个卷积层和上层降采样层通常拥有相同数量的特征图。构成卷积层 x 的每个神经元负责对输入图像 (如果 $x=1$) 或者 $x-1$ 降采样层的特征图的特定小区域施行卷积运算, 而降采样层 y 的每个神经元则负责对 $y-1$ 卷积层的对应特征图的特定小区域进行Max Pooling (只保留该区域神经元的最大输出值)。卷积运算中所使用的卷积核系数都是通过学习训练自动获取的。卷积层中属于同一个特征图的神经元都共享一个卷积核, 负责学习和提取同一种图像特征, 对应 Hubel-Wiesel模型中某种特定取向的简单细

胞。卷积层中不同的特征图负责学习和提取不同的图像特征, 对应Hubel-Wiesel模型中不同类型的简单细胞。而降采样层 y 中神经元的Max Pooling操作等同于Hubel-Wiesel模型中复杂细胞对同类型简单细胞的聚类, 是对人脑视觉皮层复杂细胞的简化模拟。

上世纪90年代初期, 贝尔实验室的Yann LeCun等人成功应用卷积神经网络实现了高精度手写数字识别算法, 所提出的系列LeNet, 都达到商用级识别精度, 被当时美国邮政局和许多大银行用来识别信封上的手写邮政编码及支票上面的手写数字。然而, 受制于90年代计算机有限的内存和弱小的运算能力, LeNet网络采用了较浅的网络结构, 每层使用的特征图数目也很少。尽管它在小规模图像识别问题上取得了较好的效果, 但与传统机器学习算法 (如SVM, AdaBoost等) 相比, 优势并不十分明显。此外, 由于卷积神经网络拥有很高的自由度, 设计出一款性能优异的网络需要灵感并配合丰富的经验积累, 是一项极具挑战性的工作。因此卷积神经网络在被提出后的很长一段时间里并未得到足够的重视和广泛的应用。

2012年, 加拿大多伦多大学Geoffrey Hinton教授的团队提出了一个规模比传统CNN大许多的深度卷积神经网络 (简称 AlexNet)。该网络拥有5个卷积与降采样层、3个全连接层, 每个卷积与降采样层拥有96384个特征图, 网络参数达到6000多万个。利用AlexNet, Hinton团队在国际上最具影响力的图像内容分类比赛 (2012 ImageNet ILSVRC) 中取得了压倒性胜利, 将1000类图像的Top-5分类错误率降低到15.315%。在这

图3 人脑视觉通道神经网络



次比赛中，获得第二、三、四名的团队均采用了传统机器学习算法。三个团队的Top-5图像分类错误率分别是26.17%、26.98%和27.06%，相差不到1个百分点，而他们的成绩和第一名相比却低了超过10个百分点，差距十分明显。当前，深度卷积神经网络（Deep CNN）相对传统机器学习算法的优势还在不断扩大，传统学习方法在多个领域已经完全无法与Deep CNN相抗衡。

机器学习算法的基本原理及其本质

在几千年的科学探索与研究中，科学家们提出了许多描述自然界及人类社会中各种事物与现象的数学模型。这些模型主要可以被归纳为以下三大类别。

归纳模型：由少数几个参数（变量）构成，每个变量都具有明确的物理意义。这类模型能够真正揭示被描述对象的本质及规律，许多数学和物理定律都是典型的归纳模型。

预测模型：用一个拥有大量参数的万能函数来拟合用户所提供的训练样本。万能函数的参数一般不具备任何物理意义，模型本身往往只能用来模拟或预测某个特定事物或现象，

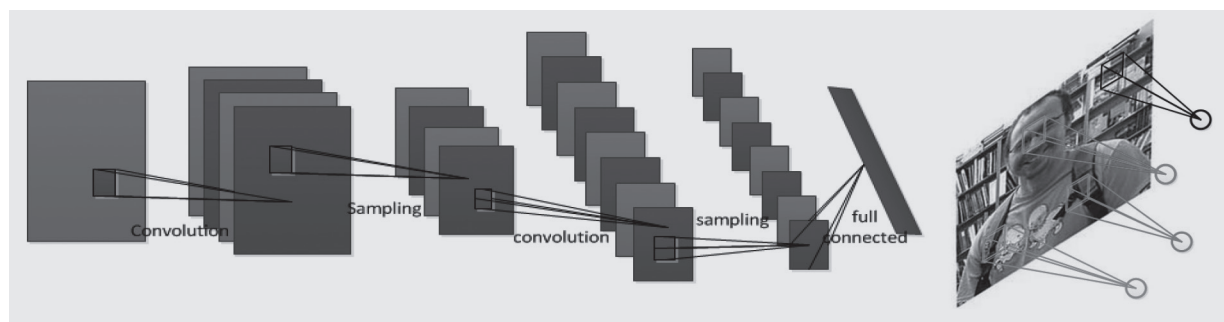
并不能揭示被描述事物或现象的本质及内在规律。当代的大多数机器学习算法都是构建于预测模型之上的。例如，单隐层全连接神经网络所使用的数学模型是：

$$F(\mathbf{x}; \Theta) = \sum_{m=1}^M \beta_m \tanh(\mathbf{w}_m \cdot \mathbf{x} + b_m),$$

上式中， \mathbf{x} 代表神经网络的输入， $\Theta = \{\beta_m, \mathbf{w}_m, b_m\}_{m=1}^M$ 代表神经网络的参数集， M 是隐层神经元的个数。这个数学模型如同一个橡皮泥，可以通过变换它的参数集 Θ 被塑造成任何形状。给定一个训练样本集 $\mathbf{T} = \{\mathbf{x}_i, y_i\}_{i=1}^N$ ，其中 \mathbf{x}_i, y_i 分别代表训练样本 i 以及人工赋予该样本的标签（标签表示样本的类别或某种属性），通过利用 \mathbf{T} 进行训练，我们就能够得到一个优化的参数集 Θ^* ，使神经网络能够很好地拟合训练样本集 \mathbf{T} 。当新的未知样本 \mathbf{x} 出现时，我们就能够利用训练好的神经网络预测出它的标签 y 。显而易见，神经网络的参数集规模与神经元的数目及输入 \mathbf{x} 的维数成正比，所有参数没有任何物理意义，模型本身也不具备揭示被描述对象的本质及内在规律的能力。

直推模型：没有明确的数学函数，利用所采集的大数据预测特定输入的标签。此类模型

图4 深度卷积神经网络构造图



认为针对某个事物或现象所采集的大数据就是对该事物或现象的客观描述。大数据的规模越大,对事物或现象的描述就越全面和准确。当新的未知样本 x 出现时,我们可以在大数据中找到 x 的 K 近邻,根据 K 近邻的标签或属性来决定 x 的标签或属性。显而易见,由于不需要定义明确的数学模型,与其它模型相比,直推模型最简单直接,但因为依靠大数据来决定未知样本的标签,直推模型往往需要较高的计算量及使用成本。同样,直推模型也不能被用来揭示事物或现象的本质及内在规律。

应当指出,随着互联网用户数量的不断增长以及互联网技术的快速进步,利用互联网获取内容或用户大数据变得越来越简单廉价,利用直推模型来预测某个事物或现象也变得越来越普及。例如,许多互联网搜索引擎利用每个网页的用户点击率来改进搜索网页的排序精度,就是直推模型在互联网内容搜索领域的一个成功应用。

综上所述,机器学习算法的本质就是选择一个万能函数建立预测模型。利用用户提供的训练样本对模型进行训练的目的,就是选择最优的参数集,使模型能够很好地拟合训练样本集的空间分布。通过训练得到的预测模型,实际上把训练样本集的空间分布提取出来并编码到其庞大的参数集中。利用这个训练好的预测模型,我们就能够预测新的未知样本 x 的标签或属性。当今大多数机器学习算法都是基于这个原理,谷歌公司的AlphaGo也不例外。

针对某个事物或现象所采集的训练样本,是对该事物或现象的直观描述,蕴藏着大量与之相关的先验知识。例如,ImageNet

ILSVRC国际图像内容分类比赛所提供的训练样本集拥有1000类、总共一百多万张彩色图像。每一类都对应自然界中的一种常见物体,如汽车、飞机、狗、鸟,等等,包含大约1000张从不同场景及不同角度拍摄的该种物体的彩色图像。利用这个训练样本集训练出来的深度卷积神经网络,实际上是将每类物体的共性特征及个体差异等进行信息提取与编码,并记忆到其庞大的参数集中。当新的未知图像出现时,神经网络就能够利用已编码到参数集中的这些先验知识,对输入图像进行准确的识别与分类。

同样,谷歌公司在训练AlphaGo时,收集了20万个职业围棋高手的对局,再利用AlphaGo不同版本间的自我对弈生成了3000多万个对局。3000多万个围棋对局包含了人类在围棋领域所积累的最为丰富和全面的知识与经验。当新的棋局出现时,AlphaGo利用被编码于其庞大参数集中的这些先验知识,预测出胜率最高的一步棋,以及这步棋所产生的最终胜率。由于AlphaGo针对3000多万个对局进行了学习与编码,它对每一步棋的胜负判定甚至比九段棋手还要准,人类棋圣输给AlphaGo也就不足为奇了。

人类智能的本质与特性

对于人脑及其高度复杂的智能,人类至今还所知甚少。关于“智能”这个名词的科学定义,学术文献中就存在着许多个版本。即使是少数几个被深入研究的认知功能(如人脑的视觉认知功能)的工作机理,也还存在着各种各



样的假说和争议。在这里，我们列出若干较具代表性、认可度相对较高的关于人脑智能的假说及阐述。

人类智能的本质是什么？这是认知科学的基本任务，也是基础科学面临的四大难题（Simon）中最后、最难解决的一个。每门基础科学都有其特定的基本单元，例如高能物理学的基本粒子，遗传学的基因、计算理论的符号、信息论的比特等。因此，“人类智能的本质是什么”这个问题在某种程度上取决于“什么是认知基本单元”。众所周知，适合描述物质世界的变量并不一定适合描述精神世界。因此，认知基本单元是什么这个问题，不能靠物理的推理或计算的分析来解决，根本上只有通过认知科学的实验来回答。大量实验结果显示，认知基本单元不是计算理论的符号，也不是信息论的比特，而是知觉组织形成的“知觉物体”。例如，实验表明，当人的视觉系统注意一只飞鸟的时候，它所注意的是整只鸟（即一个知觉物体），而不是鸟的某个特性（形状、大小、位置等）。尽管在飞行过程中鸟的各种特征性质在改变，但它是同一个知觉物体的性质始终保持不变。诺奖得主Kahneman认为，知觉物体概念的直觉定义正是在形状等特征性质改变下保持不变的同一性。中科院陈霖院士领导的团队在发展了30多年的拓扑性质知觉理论的基础上，提出大范围首先的知觉物体拓扑学定义：知觉物体的核心含义，即在变换下保持不变的整体同一性，可以被科学准确地定义为大范围拓扑不变性质。应当指出，上述大范围首先知觉物体的概念，与人工智能领域广为认同与采纳的由局部到整体，由特征到物

体，由具体到抽象的认知计算模型是完全背道而驰的，因而在人工智能领域并没有得到足够的重视及应用。

大量认知科学领域的实验研究表明，人类智能具有以下几个特性。

人类智能的目标不是准确。人类智能并不追求在精神世界里客观准确地再现物理世界。上帝设计人类智能时，不假思索地直奔“生存”这一终极目标而去：用最合理的代价，获取最大的生存优势。人类大脑的平均能耗大约只有20瓦，相对于庞大的计算机系统来说只是九牛一毛。尽管人脑的重量只有1400克左右，约占人体重量的2.3%，但它的血液供应量却占到了全身的15.20%，耗氧量超过全身的20%，对于人类已经接近其生理可以负担的极限。在这种资源极其有限的条件下，人脑通过以下几种方式实现了最有效的资源调配，由此来保障最有意义的生理和智能活动。

第一，主观能动的选择性。精神世界不是对物理世界的简单映射，而是非常扭曲和失真的。体积相对较小的手指、舌头等重点区域，在感觉运动中枢里却占据大部分的皮层区域。同样，在视觉上只有对应中央视野的视网膜具有很高的空间、颜色分辨率，而更广泛的外周视野只对物体的突然出现或消失，以及物体的运动更敏感。人类视觉处理的通常方式是，外周视野的显著变化会在第一时间被捕获，做出应激反应，然后再把中央视野移动到目标上进行后续的处理。

人类通过知觉组织的选择性注意机制，直接感知输入信号中的大范围不变性质，而忽略大量的局部特征性质。大量视而不见的现象，

在实验室研究中表现为注意瞬脱、变化盲视等等。比如，尽管可以清晰地分辨出霓虹灯中的色块颜色、形状各不相同，甚至在空间和时间上都不连续，人脑仍然把这些色块看成是同一个物体，从而产生运动的感觉。研究表明，这种运动错觉本质上不是运动，其生态意义在于对知觉对象进行不变性抽提。另一方面，人脑会主动把忽略的部分补充回来。而通过经验知识，上下文关系等补充回来的信息，难免有错。所谓错觉就是精神世界和物理世界的错位。这些错觉的生态意义在于在有限资源条件下，快速直接地形成稳定的感知。这种机制既是人类天马行空的联想能力和创造力的源泉，同时也是各种精神心理疾患的生物学基础。

第二，模块化的层次结构和分布式表征。当前认知科学越来越依赖于脑成像技术的发展。功能模块化假设认为，大脑是由结构和功能相对独立、专司特定认知功能的多个脑区组成。这些模块组成复杂的层次结构，通过层次间的传递和反馈实现对输入信号的主动调节。大量脑成像的研究实验也支持了这一假设，特别是视觉研究发现了非常详细而复杂的功能模块及其层次结构。另一方面，分布式表征的假说认为，认知功能的神经机制是相对大范围的分布式脑状态，而不是特定脑区的激活与否。当前研究认为，人脑是模块化和分布式表达共存的自能系统。

第三，反应性活动和内生性活动。人脑不是一个简单的刺激—反应系统，大量的内生性活动甚至比反应性活动还多。人脑在所谓的静息状态下的耗氧量与任务状态下相比差别很小。然而几乎所有的经典认知科学研究都是建

立在刺激反应实验范式的基础之上。这种实验范式是让实验对象在特定的条件下完成特定的认知任务，收集并分析实验对象的行为或生理反应，通过对实验数据的充分比照，建立人脑某种活动模式或认知机理的假设。内生性活动因其往往只能通过内省的方式进行研究，而被长期排除在认知科学的研究主流之外。随着脑成像技术的发展，功能连接成为分析静息态大脑自发活动的有力工具。特别是默认网络的发现，创立了强调内生性活动的全新脑功能成像研究范式。默认网络被认为涉及警觉状态、自我意识、注意调控以及学习记忆等心理认知过程，已被广泛应用于社会认知、自我、注意、学习、发育、衰老机制的研究，有力推动了各种脑生物指标的完善和脑疾病的治疗，这些疾病包括阿尔兹海默病、帕金森病、抑郁症、精神分裂症和自闭症等等。

因此，整合现有研究中有关分布式表达和内生性活动的最新研究成果，可能会带来对人脑活动模式（人类智能的物质基础）一种全新的理解。

人类智能的本质不是计算。人类智能体现在对外部环境的感知、认知、对所观察事物或现象的抽象、记忆、判断、决策等。然而，这些智能并不是人类所独有。许多高等动物，如狗、猴子、猩猩，也或多或少具有类似的能力。同时，计算并不是人类智能的强项。真正将人类与其它动物区分开来的，是人类的逻辑推理能力、想象力、创造力以及自我意识。人类利用这类能力能够想象并且创造出自然界中不存在的东西，如汽车、飞机、电视、计算机、手机，互联网。这类能力是推动人类社会



不断发展与进步的源泉，是生物智能的圣杯。

而对代表生物智能最高水平的上述能力，人类目前还所知甚少，对其机理的研究还处于启蒙阶段。研究表明，这些能力不是依靠计算得来的，而似乎是与联想记忆及人类丰富的精神世界有关。基于脑信号的分析实验发现，人脑的海马回、海马旁回、杏仁核等脑区中存在着大量专司特定联想记忆的神经细胞。例如，上述脑区中存在单个或一小簇神经细胞，会被与美国前总统克林顿相关的所有刺激信号所激活，无论刺激信号是关于克林顿的图片，还是Clinton这个英语单词，还是克林顿本人的语音回放。显然，这些神经细胞并不是被某个模态的特定特征所激活，它们所对应的是克林顿这个抽象概念。此外，脑成像研究表明，围棋专业棋手相对于业余棋手更多的是依赖联想记忆系统，而非逻辑推理来下棋。实际上，围棋界训练棋手的最常用方法就是将高手对局中的关键部分拆解成许多死活题，棋手通过大量死活题的解题训练来提升自己联想记忆的经验 and 效率。

机器智能与人类智能的优势与劣势

当代的计算机拥有强大的存储与运算能力。伴随着计算技术的不断发展与进步，这些能力的增长似乎还远没有到达尽头。早在1997年，IBM的“深蓝”超级电脑就战胜了国际象棋冠军卡斯帕罗夫。但这次胜利在人工智能领域并没有产生太大的反响，原因在于，“深蓝”几乎纯粹是依靠强大的运算能力遍历所有的可能性，利用“蛮力”取胜的。“深蓝”所遵循的，就

是“人工智能即是计算加记忆”这个简单法则。

由于围棋的搜索空间比国际象棋大很多，“深蓝”的这种制胜策略针对围棋是行不通的。与“深蓝”相比，AlphaGo的最大进步就是从“计算加记忆”进化到“拟合加记忆”法则。它利用深度卷积神经网络这个万能函数，通过学习来拟合两千多年来人类所积累的全部经验及制胜模式，并将其编码到神经网络的庞大参数集中。对于当前棋局的任何一个可能的落子，训练好的神经网络都能够预测出它的优劣，并通过有限数量的模拟搜索，计算出最终的获胜概率。这样的战略不需要对棋局的所有可能性做遍历搜索，更像人类棋手所使用的策略。然而，由于AlphaGo对每个落子以及最终胜率的预测，是建立在围棋界两千多年来所形成的完整知识库之上的，它的预测比人类最优秀的棋手更准确。与其说李世石输给了机器系统，不如说输给了人类棋艺的集大成者。由此推断，AlphaGo取胜也是情理之中的事。

与机器相比，人类智能的最大优势当属它的逻辑推理能力、想象力、创造力及其高效性。人脑功耗只有20多瓦，处理许多感知及认知任务（如图像识别、人脸识别、语音识别等）的精度与拥有庞大内存、运算速度达到万亿次的超级电脑相比却毫不逊色。尽管机器智能很可能在不远的将来在棋牌类竞赛中全面超越人类，但现有的机器学习框架并不能模拟出人类的想象力和创造力。因此，在当前情况下，机器智能全面超越人类智能的预测是不会成为现实的。

随着机器学习算法的不断发展与进步，计算机借助强大的存储与运算能力，学习人类几

千年来发展与进化过程中所积累的完整知识的能力越来越强，借助完整知识库对复杂事务进行预测与判断的准确度将会全面超越人类。由此推断，在未来几十年里，不仅是那些简单重复性的体力劳动将会全面被机器取代，而且那些需要对复杂事务进行评估与判断的工作，如金融投资、企业管理、军事指挥等，也有可能被让位于机器智能。甚至大到整个国家，也可能越来越依靠机器智能预测政治、经济、外交发展趋势，制定最优的政策方针及发展规划。实际上，许多发达国家的智囊机构已经在利用各种评估及预测模型为政府提供对各种事物的预测与判断，提出政策建议或解决方案。

然而，当前的机器学习框架无法模拟人类

的想象力及创造力，科学研究与发明创造仍将是人类的优势所在。不难预测，在未来人类社会的发展进程中，将有越来越多的人从事科学研究以及新产品的设计研发工作。社会对每个人的知识能力、智慧以及发明创造力的要求将会越来越高，不具备这些能力的人们将会无法找到满意的工作，逐渐成为处于社会底层的贫困阶层。了解并解决科技迅速发展所带来的社会挑战，仍然是人类需要面对的任务，而机器是无法替代人类解决这些问题的。

（中科院生物物理所脑与认知国家重点实验室周天罡、西安交通大学电信学院韩劲松对本文亦有贡献）

责编 / 凌肖汉

Will Artificial Intelligence Surpass Human Intelligence Ultimately? ——Discussion Based on Basic Principle of Machine Learning and Human Cognition Gong Yihong

Abstract: Recently, the smashing 4-1 win of Google's AlphaGo over the South Korean Go Master Lee Sedol has not only shown the power and great potentials of artificial intelligence, but also caused panic and worries in the human society. People begin to think carefully about the progress and challenges that artificial intelligence will bring to the human society and whether artificial intelligence will ultimately surpass human intelligence. To find the answer, we need to first understand the essence and basic principles of artificial intelligence. In the beginning, this article briefly described the deep Convolutional Neural Networks (deep CNN)—the most popular research result in the field of machine intelligence to date, and then discussed the essence and basic principles of machine learning. After that, with an overview of the latest research discoveries about human brain's cognitive mechanism, this article tried to reveal the essential differences between machine intelligence and human intelligence, and compared their advantages and disadvantages in order to find the answers to the foregoing questions.

Keywords: artificial intelligence, AlphaGo, neural synapse, machine learning, pattern recognition