

人工智能的历史回顾和发展现状

顾险峰[†]

纽约州立大学石溪分校计算机系, 纽约 11794

摘要 简略地回顾了人工智能的历史和发展现状。分析比较了人工智能两大领域: 符号主义和连接主义, 同时介绍了各个领域的主要原理和方法。着重回顾了深度学习的历史、复兴的原因和主要的应用。

关键词 人工智能; 连接主义; 符号主义; 深度学习; 图像识别; 语音识别; 神经网络

最近, 谷歌的阿尔法狗击败了围棋九段李世石, 举世震惊。有为人工智能的发展欢呼雀跃者, 也有为人类前途命运忧心忡忡者; 有对机器蛮力不屑一顾者, 也有对人类失去优越感而沮丧彷徨者。目前, 人工智能的浪潮汹涌澎湃, 在视觉图像识别、语音识别、文本处理等诸多方面人工智能已经达到或超越人类水平, 在视觉艺术、程序设计方面也开始崭露头角, 令人惊叹不已。人们已经相信, 在个人电脑时代、网络时代、手机时代之后, 整个社会已经进入人工智能时代。

这里, 我们考察人工智能发展的简要历史、目前的局限和未来的潜力, 特别是将人类脑神经认知和人工神经网络认知进行对比, 从而对人工智能有一个公正客观, 而又与时俱进的认识。

从历史上看, 人类的智能主要包括归纳总结和逻辑演绎, 对应着人工智能中的联结主义(如人工神经网络)和符号主义(如吴文俊方法)。人类大量的视觉听觉信号的感知处理都是下意识的, 是基于大脑皮层神经网络的学习方法; 大量的数学推导、定理证明是有强烈主观意识的, 是基于公理系统的符号演算方法。

1 符号主义

古希腊人将欧几里得几何归纳整理成欧几里得公理体系, 整个宏伟的理论大厦奠基几条不言自明的公理, 整个大厦完全由逻辑构造出来, 美轮美奂, 无懈可击。这为整个人类科学发展提供了一套标准的范式。后来, 牛顿编撰他的鸿篇巨著《自然哲学的数学原理》也遵循公理体系的范式, 由公理到定义、引理、定理再到推论。人类的现代数学和物理知识最终都被系统化整理成公理体系, 比如爱因斯坦的广义相对论也是遵循公理体系的范式。当然也存在例外。例如, 虽然量子理论已经为人类科技带来天翻地覆的革命, 但是量子理论的公理体系目前还没有建立起来。符号主义的主要思想就是应用逻辑推理法则, 从公理出发推演整个理论体系。

人工智能中, 符号主义的一个代表就是机器定理证明, 吴文俊先生创立的吴文俊方法是其巅峰之一。目前基于符号计算的机器定理证明的理论根基是希尔伯特定理: 多元多项式环中的理想都是有限生成的。我们首先将一个几何命题的条件转换成代数多项式, 同时把结论也转换成多

[†]通信作者, 顾险峰与丘成桐先生等合作开创了计算共形几何这一交叉学科, 他们合著出版了该领域的权威专著《计算共形几何》(Computational Conformal Geometry)。E-mail: gu@cs.stonybrook.edu

项式，然后证明条件多项式生成的根理想包含结论对应的多项式，即将定理证明转换为根理想成员判定问题。一般而言，多项式理想的基底并不唯一，Groebner基方法和吴方法可以生成满足特定条件的理想基底，从而都可以自动判定理想成员问题。因此理论上代数范畴的机器定理证明可以被完成，但是实践中这种方法有重重困难。

首先，从哲学层面上讲，希尔伯特希望用公理化方法彻底严密化数学基础。哥德尔证明了对于任何一个包含算术系统的公理体系，都存在一个命题，其真伪无法在此公理体系中判定。换言之，这一命题的成立与否都与此公理体系相容。一方面，这意味着我们无法建立包罗万象的公理体系，无论如何，总存在真理游离在有限公理体系之外；另一方面，这也意味着对于真理的探索过程永无止境。

其次，从计算角度而言，Groebner基方法和吴方法所要解决的问题的本质复杂度都是超指数级别的，即便对于简单的几何命题，其机器证明过程都可能引发存储空间的指数爆炸，这揭示了机器证明的本质难度。吴方法的成功有赖于大多数几何定理所涉及的代数计算问题是有结构的，因而可以快速求解。

第三，能够用理想生成的框架证明的数学命题，其本身应该是已经被代数化了。如所有的欧几里得几何命题，初等的解析几何命题。微分几何中许多问题的代数化，本身就非常具有挑战性。例如黎曼流形的陈省身-高斯-博内定理：流形的总曲率是拓扑不变量。如果没有嘉当发明的外微分和活动标架法，这一定理的证明就无法被代数化。拓扑学中的许多命题的代数化本身也是非常困难的，比如众所周知的布劳威尔不动点定理：我们用咖啡勺缓慢均匀搅拌咖啡，然后抽离咖啡勺，待咖啡静止后，必有一个分子，其搅拌前和搅拌后的位置重合。这一命题的严格代数化是一个非常困难的问题。吴先生的高足高山研究员突破性的微分结式理论，系统地将这种机器证明方法从代数范畴推广到微分范畴^[1]。

最后，机器定理证明过程中推导出的大量符号公式，人类无法理解其内在的几何含义，无

法建立几何直觉。而几何直觉和审美，实际上是指导数学家在几何天地中开疆拓土的最主要的原则。机器无法抽象出几何直觉，也无法建立审美观念，因此虽然机器定理证明经常对于已知的定理给出令人匪夷所思的新颖证明方法，但是迄今为止，机器并没有自行发现深刻的未知数学定理。

比如，人类借助计算机完成了地图四色定理的证明，但是对于这一证明的意义一直富有争议。首先，这种暴力证明方法没有提出新的概念、新的方法；其次，这个证明没有将这个问题和其他数学分支发生深刻内在的联系。数学中，命题猜测的证明本身并不重要，真正重要的是证明所引发的概念思想、内在联系和理论体系。因此，许多人认为地图四色定理的证明实际上“验证”了一个事实，而非“证明”了一个定理。目前，机器定理证明的主流逐渐演变成机器验证。因此，和人类智慧相比，人工智能的符号主义方法依然处于相对幼稚的阶段。

即便如此，人工智能在某些方面的表现已经超越人类。例如，基于符号主义的人工智能专家系统IBM的沃森，在电视知识竞赛Jeopardy中表现出色，击败人类对手，赢得冠军。目前，IBM进一步发展沃森认知计算平台，结合深度卷积神经网络后获得了更强的数据分析与挖掘能力，在某些细分疾病领域已能达到顶级医生的医疗诊断水平。

2 联结主义

人工智能中的联结主义的基本思想是模拟人类大脑的神经网络。David Hunter Hubel 和 Torsen Wiesel(图1)共同获得了1981年的诺贝尔生理学或医学奖。1959年，Hubel和Wiesel在麻醉的猫的视觉中枢上插入了微电极，然后在猫的眼前投影各种简单模式，同时观察猫的视觉神经元的反应。他们发现：猫的视觉中枢中有些神经元对于某种方向的直线敏感，另外一些神经元对于另外一种方向的直线敏感；某些初等的神经元对于简单模式敏感，而另外一些高级的神经元对于复杂模式敏感，并且其敏感度和复杂模式的位置



图1 1981年的诺贝尔生理学或医学奖得主
David Hunter Hubel 和Torsen Wiesel

与定向无关。这证明了视觉中枢系统具有由简单模式构成复杂模式的功能，也启发了计算机科学家发明人工神经网络。

后来通过对猴子的视觉中枢的解剖，将猴子的大脑皮层曲面平展在手术台表面上，人们发现从视网膜到第一级视觉中枢的大脑皮层曲面的映射(retinotopic mapping)是保角映射(conformal mapping)^[2]。保角变换的最大特点是局部保持形状，但是忽略面积大小(图2)。这说明视觉处理对于局部形状非常敏感。



图2 三维曲面到平面的保角映射

人们逐步发现，人类具有多个视觉中枢，并且这些视觉中枢是阶梯级联，具有层次结构。人类的视觉计算是一个非常复杂的过程。在大脑皮层上有多个视觉功能区域(v1 至 v5等)，低级区域的输出成为高级区域的输入。低级区域识别图像中像素级别的局部的特征，例如边缘折角结构，高级区域将低级特征组合成全局特征，形成复杂的模式，模式的抽象程度逐渐提高，直至语义级别。

如图3所示，毕加索的名画《格尔尼卡》

(Guernica)中充满了抽象的牛头马面、痛苦嚎哭的人脸、扭曲破碎的肢体。我们却可以毫不费力地辨认出这些夸张的几何形体。其实，尽管图中大量信息丢失，但是提供了足够的整体模式。由此可见，视觉高级中枢忽略色彩、纹理、光照等局部细节，侧重整体模式匹配和上下文关系，并可以主动补充大量缺失信息。

这启发计算机科学家将人工神经网络设计成多级结构，低级的输出作为高级的输入。最近，深度学习技术的发展，使得人们能够模拟视



图3 毕加索的名画《格尔尼卡》

觉中枢的层级结构,考察每一级神经网络形成的概念。图4显示一个用于人脸识别的人工神经网络经过训练后习得的各层特征。底层网络总结出各种边缘结构,中层网络归纳出眼睛、鼻子、嘴

巴等局部特征,高层网络将局部特征组合,得到各种人脸特征。这样,人工神经网络佐证了视觉中枢的层次特征结构。



图4 深度学习神经网络经学习得到的不同层次的特征 (作图: Andrew Ng)

3 深度学习的兴起

人工神经网络在20世纪80年代末和90年代初达到巅峰,随后迅速衰落,其中一个重要原因是因为神经网络的发展严重受挫。人们发现,如果网络的层数加深,那么最终网络的输出结果对于初始几层的参数影响微乎其微,整个网络的训练过程无法保证收敛。同时,人们发现大脑具有不同的功能区域,每个区域专门负责同一类的任务,例如视觉图像识别、语音信号处理和文字处理等等。而且,在不同的个体上,这些功能中枢在大脑皮层上的位置大致相同。在这一阶段,计算机科学家为不同的任务发展出不同的算法。例如:为了语音识别,人们发展了隐马尔科夫链模型;为了人脸识别,发展了Gabor滤波器、SIFT特征提取算子、马尔科夫随机场的图模型。因此,在这个阶段人们倾向于发展专用算法。

但是,脑神经科学的几个突破性进展使人们彻底改变了看法。在2000年, Jitendra Sharma在《自然》上撰文^[3],汇报了他们的一个令人耳目一新的实验。Sharma把幼年鼯鼠的视觉神经和听觉神经剪断,交换后接合,眼睛接到了听觉中枢,耳朵接到了视觉中枢。鼯鼠长大后,依然发展出了视觉和听觉。这意味着大脑中视觉和听觉的计算方法是通用的。在2009年, Vuilleme和Cuisinier为盲人发明了一套装置^[4],将摄像机的输出表示成二维微电极矩阵,放在舌头表面。盲人经过一段时间的学习训练,可以用舌头“看

到”障碍物。在2011年,人们发现许多盲人独自发展出一套“声纳”技术,他们可以通过回声来探测并规避大的障碍物。Thaler等人的研究表明,他们的“声纳”技术用的并不是听觉中枢,而是原来被废置的视觉中枢。

种种研究表明,大脑实际上是一台“万用学习机器”(universal learning machine),同样的学习机制可以用于完全不同的应用。人类的DNA并不提供各种用途的算法,而只提供基本的普适的学习机制。人的思维功能主要是依赖于学习所得,而后天的文化和环境决定了一个人的思想和能力。换言之,学习的机制人人相同,但是学习的内容决定了人的思维(mind)。

人的大脑具有极强的可塑性,许多功能取决于后天的训练。例如,不同民族语言具有不同的元音和辅音,阿拉伯语最为复杂,日语相对简单。出生不久的婴儿可以辨别听出人类能够发出的所有元音和辅音,但是在5岁左右,日本幼儿已经听不出很多阿拉伯语中的音素了。同样,欧洲人可以非常容易地辨认本民族面孔,但是非常容易混淆亚洲人面孔。人们发现,如果大脑某个半球的一个区域受损并产生功能障碍,随着时间流逝,另一半球的对称区域就会“接替”受损区域,掌管相应功能。这些都表明大脑神经网络具有极强的可塑性。

大脑学习算法的普适性和可塑性一直激励着计算机科学家不懈地努力探索。历史性的突破发生在2006年左右,计算机科学家Geoffrey

Hinton、Yann Lecun和Yoshua Bengio突破深度学习的技术瓶颈,进而引领深度学习的浪潮。

与传统神经网络相比,深度学习的最大特色在于神经网络的层数大为增加。深度网络难以收敛的技术瓶颈最终被打破,主要的技术突破在于以下几点:首先是计算能力的空前增强。目前深度网络动辄上百层,联接参数数十亿,训练样本经常数千万直至上亿,训练算法需要在大规模计算机集群上运行数月。这些训练过程需要非常庞大的计算资源。计算机计算能力的提升,特别是GPU的迅猛发展,为深度学习提供了强有力的硬件保障。其次是数据的积累。特别是互联网的大规模普及,智能手机的广泛使用,使得规模庞大的图像数据集能够被采集,上传到云端,集中存储处理。深度学习需要使用越来越大的数据集,大数据的积累提供数据保障。再就是深度学习网络初始化的选择。传统神经网络随机初始化,学习过程漫长,并且容易陷入局部最优而无法达到性能要求。目前的方法使用非监督数据来训练模型以达到特征自动提取,有针对性地初始化网络,加速了学习过程的收敛,提高了学习效率。更为关键的是优化方法的改进。目前的技术采用更加简单的优化方法,特别是随机梯度下降方法的应用提高了收敛速率和系统稳定性。

4 神经网络简史

4.1 第一次浪潮

在1943年,科学家Warren McCulloch和Walter Pitts提出了神经网络作为一个计算模型的理论。1957年,康内尔大学教授Frank Rosenblatt提出了“感知器”(perceptron)模型。感知器是第一个用算法来精确定义的神经网络,第一个具有自组织自学习能力的数学模型,是日后许多新的神经网络模型的始祖。感知器的技术在20世纪60年代带来人工智能的第一个高潮。

1969年,Marvin Minsky和Seymour Papert^[5]在出版的《感知器:计算几何简介》一书中强烈地批判了感知器模型:首先,单层的神经网络无法解决不可线性分割的问题,典型例子如异或门;其次,当时的计算能力低下无法支持神经网络

模型所需的计算量。此后的十几年,以神经网络为基础的人工智能研究进入低潮。

4.2 第二次浪潮

Minsky提出的尖锐问题后来被逐步解决。传统的感知器用所谓“梯度下降”的算法纠错时,其运算量和神经元数目的平方成正比,因而计算量巨大。1986年7月,Hinton和David Rumelhart^[6]合作在《自然》发表论文,系统地提出了应用反向传播算法,把纠错的运算量下降到只和神经元数目成正比。同时,通过在神经网络里增加一个所谓隐层(hidden layer),反向传播算法同时也解决了感知器无法解决的异或门难题。

Hinton的博士后Yann Lecun于1989年发表了论文《反向传播算法在手写邮政编码上的应用》^[7]。他用美国邮政系统提供的近万个手写数字的样本来训练神经网络系统,在独立的测试样本中错误率低至5%,达到实用水准。他进一步运用“卷积神经网络”(convoluted neural networks)的技术,开发出商业软件,用于读取银行支票上的手写数字,这个支票识别系统在20世纪90年代末占据了美国接近20%的市场。

贝尔实验室的Vladimir Vapnik在1963年提出了支持向量机(support vector machine, SVM)的算法。在数据样本线性不可分的时候,支持向量机使用所谓“核机制”(kernel trick)的非线性映射算法,将线性不可分的样本转化到高维特征空间(high-dimensional feature space),使其线性可分。作为一种分类算法,从20世纪90年代初开始,SVM在图像和语音识别上找到了广泛的用途。在手写邮政编码的识别问题上,SVM技术在1998年错误率降至0.8%,2002年最低达到了0.56%,远远超越同期的传统神经网络。

这时,传统神经网络的反向传播算法遇到了本质难题——梯度消失(vanishing gradient problem)。这个问题在1991年被德国学者Sepp Hochreiter第一次清晰提出并阐明原因。简单地说,就是成本函数(cost function)从输出层反向传播时,每经过一层,梯度衰减速度极快,学习速度变得极慢,神经网络很容易停滞于局部最优

解而无法自拔。同时, 算法训练时间过长会出现过度拟合(overfit), 把噪音当成有效信号。SVM理论完备、机理简单、容易重复, 从而得到主流的追捧。SVM技术在图像和语音识别方面的成功使得神经网络的研究重新陷入低潮。

4.3 第三次浪潮

(1) 改进算法

2006年, Hinton 和合作者^[8]发表论文《深信度网络的一种快速算法》。在这篇论文里, Hinton 在算法上的核心是借用了统计力学里的“玻尔兹曼分布”的概念, 使用所谓的“限制玻尔兹曼机”(RBM)来学习(图5)。

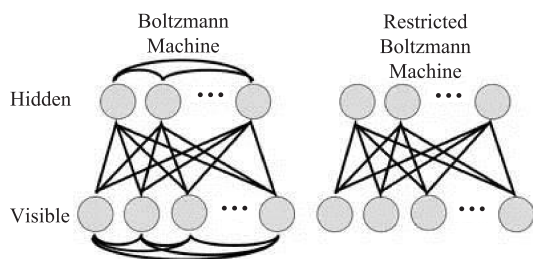


图5 玻尔兹曼机与限制玻尔兹曼机

RBM 相当于一个两层网络, 可以对神经网络实现“没有监督的训练”(unsupervised training)。深信度网络就是几层 RBM 叠加在一起, RBM可以从输入数据中进行预先训练, 自行发现重要特征, 对神经网络连接的权重进行有效的初始化。经过 RBM 预先训练初始化后的神经网络, 再用反向传播算法微调, 效果得到大幅度提升。

2011 年, 加拿大的蒙特利尔大学学者 Xavier Glorot和Yoshua Bengio发表论文《深而稀疏的修正神经网络》^[9]。论文的算法中使用一种称为“修正线性单元”(rectified linear unit, RELU)的激励函数。和使用别的激励函数的模型相比, RELU识别错误率更低, 而且其有效性对于神经网络是否进行“预先训练”并不敏感。RELU 的导数是常数, 非零即一, 不存在传统激励函数在反向传播计算中的“梯度消失问题”。由于统计上约一半的神经元在计算过程中输出为零, 使用 RELU 的模型计算效率更高, 而

且自然而然地形成了所谓“稀疏表征”(sparse representation), 用少量的神经元可以高效、灵活、稳健地表达抽象复杂的概念。

2012年7月, Hinton发表论文《通过阻止特征检测器的共同作用来改进神经网络》^[10]。为了解决过度拟合的问题, 论文中采用了一种新的被称为“丢弃”(dropout)的算法。丢弃算法的具体实施是在每次培训中给每个神经元一定的几率(比如 50%), 假装它不存在, 计算中忽略不计。使用丢弃算法的神经网络被强迫用不同的、独立的神经元的子集来接受学习训练。这样网络更强健, 避免了过度拟合, 不会因为外在输入的很小噪音导致输出质量的很大差异(图6)。

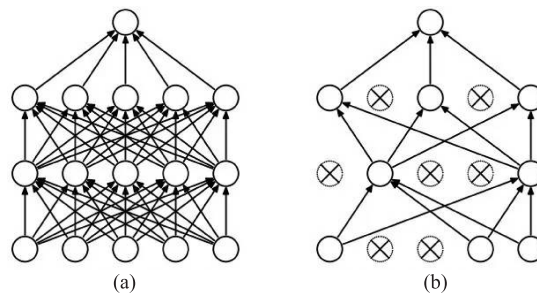


图6 标准神经网络(a)与使用丢弃算法后的神经网络(b)

(2) 使用GPU提高计算能力

2009年6月, 斯坦福大学的Rajat Raina 和吴恩达(Andrew Ng)^[11]合作发表论文《用GPU大规模无监督深度学习》, 论文模型里的参数总数(就是各层不同神经元之间链接的总数)达到1亿。与之相比, Hinton在2006年的论文里用到的参数数目只有170万。论文结果显示, 使用GPU的运行速度和用传统双核CPU相比, 最快时要快近70倍。在一个四层、1亿个参数的深信度网络上, 使用GPU把程序运行时间从几周降到一天。

2010年瑞士学者 Dan Ciresan和合作者发表论文《Deep big simple neural nets excel on handwritten digit recognition》^[12], 其中使用的还是20世纪80年代的反向传播计算方法, 但是计算搬移到GPU上实现, 在反向传播计算时速度比传统CPU快了40倍。

2012 年还在斯坦福大学做研究生的黎越国(Quoc Viet Le) 领衔, 和他的导师吴恩达, 以及

众多谷歌的科学家联合发表论文《用大规模无监督学习建造高层次特征》^[13]。黎越国的文章中使用了九层神经网络，网络的参数数量高达10亿，是Ciresan 2010年论文中的模型的100倍，是2009年Raina 论文模型的10倍。

(3) 海量的训练数据

在黎越国文章中，用于训练这个神经网络的图像都是从谷歌的录像网站youtube上截屏获得。1 000万个原始录像，每个录像只截取一张图片，每张图片有4万个像素。与之相比，先前大部分论文使用的训练图像，原始图像的数目大多在10万以下，图片的像素大多不到1 000。黎越国的计算模型分布式地在1 000台机器（每台机器有16个CPU内核）上运行，花了三天三夜才完成培训。互联网的大规模普及，智能手机的广泛使用，使得规模庞大的图像数据集能够被采集，并在云端集中存储处理。大数据的积累为深度学习提供了数据保障。

5 全面超越

5.1 图像识别

2009年，普林斯顿大学计算机系的华人学者（第一作者为Jia Deng）发表论文《ImageNet: A large scale hierarchical image database》，宣布建立第一个超大型图像数据库供计算机视觉研究者使用^[14]。2010年，以ImageNet为基础的大型图像识别竞赛ImageNet Large Scale Visual Recognition Challenge 2010 (ILSVRC2010) 第一次举办。竞赛最初的规则是以数据库内120万个图像为训练样本，这些图像从属于1 000多个不同的类别，都被手工标志。经过培训过的程序，再用于5万个测试图像评估，看看它对图像的分类是否准确。

2012年，Hinton 教授和他的两个研究生 Alex Krizhevsky、Illya Sutskever将深度学习的最新技术用到 ImageNet 的问题上。他们的模型是一个总共八层的卷积神经网络，有65万个神经元，6 000万个自由参数。这个神经网络使用了前面两篇文章介绍过的丢弃算法和修正线性单元 (RELU) 的激励函数。Hinton 教授的团队使用两

个GPU，让程序接受120万个图像训练，花了接近6天时间。经过训练的模型，面对15万个测试图像，预测的头五个类别的错误率只有 15.3%，在有30个团体参与的2012年 ImageNet的竞赛中，测试结果稳居第一。排名第二的来自日本团队的模型，相应的错误率则高达 26.2%。这标志着神经网络在图像识别领域大幅度超越其他技术，成为人工智能技术突破的一个转折点。

2015年12月的Imagenet图像识别的竞赛中，来自微软亚洲研究院(Microsoft Research Asia, MSRA)的团队夺冠。网络深度增加，学习的效率反而下降。为了解决有效信息在层层传递中衰减的问题，MSRA团队尝试了一种称为“深度残余学习”(Deep Residual Learning) 的算法。MSRA 的深度残余学习模型，使用深达 152层的神经网络，头五个类别的识别错误率创造了 3.57% 的新低，这个数字已经低于一个正常人的大约 5% 的错误率。

5.2 语音识别

RNN (recurrent neural network) 也称循环神经网络或多层反馈神经网络，则是另一类非常重要的神经网络。本质上，RNN 和前馈网络的区别是，它可以保留一个内存状态的记忆来处理一个序列的输入，这对手写字的识别、语音识别和自然语言处理尤为重要。

2012年10月，Geoffrey Hinton、邓力和其他几位代表四个不同机构（多伦多大学、微软、谷歌、IBM）的研究者，联合发表论文《深度神经网络在语音识别的声学模型中的应用：四个研究小组的共同观点》^[15]。研究者们借用了Hinton使用的“限制玻尔兹曼机”(RBM) 的算法对神经网络进行了“预培训”。深度神经网络模型(DNN)被用来估算识别文字的几率。在谷歌的一个语音输入基准测试中，单词错误率 (word error rate) 最低达到了 12.3%。

2013年3月，多伦多大学的 Alex Graves 领衔发表论文《深度循环神经网络用于语音识别》^[16]。论文中使用 RNN/LSTM 的技术——一个包含三个隐层、430万个自由参数的网络，在一个叫做 TIMIT 的基准测试中“音位错误率”达到

17.7%，优于同期的其他所有技术的表现水准。

2015年5月谷歌宣布依靠 RNN/LSTM 相关的技术，谷歌语音 (Google Voice) 的单词错误率降到了8% (正常人大约 4%)。

2015年12月，百度 AI 实验室的 Dario Amodei 领衔发表论文《英语和汉语的端对端的语音识别》^[17]。论文的模型使用的是 LSTM 的一个简化的变种，叫做“封闭循环单元” (gated recurrent unit)。百度的英文语音识别系统接受了将近12 000小时的语音训练，在 16个GPU上完成训练需要 3~5 天。在一个叫 WSJ Eval'92 的基准测试中，其单词错误率低至3.1%，已经超过正常人的识别能力(5%)。在另外一个小型汉语基准测试中，机器的识别错误率只有3.7%，而一个五人小组的集体识别错误率则为4%。

依照这个趋势，机器在语音识别的各种基准测试上的准确度很快将全面赶上并且超过普通人了。这是在图像识别之后人工智能即将攻克的一个难关。

循环神经网络 (RNN) 的本质是可以处理一个长度变化的序列的输出和输入 (多对多)。广义地看，如果传统的前馈神经网络做的事，是对一个函数的优化 (比如图像识别)，那么循环神经网络做的事，则是对一个程序的优化，应用空间宽阔得多。

5.3 艺术创作

很久以来，人们倾向于认为机器可以理解人类的逻辑思维，却无法理解人类的丰富感情，更无法理解人类的美学价值，当然机器也就无法产生具有美学价值的作品。事实胜于雄辩，阿尔法狗对局李世石下出石破天惊的一步，棋圣聂卫平先生向阿尔法狗的下法脱帽致敬，这说明深度学习算法已经能够自发创造美学价值。许多棋手在棋盘方寸间纵横一生，所追寻的就是美轮美奂的神机妙手。如此深邃优美，玄奥抽象，一夜间变成了枯燥平淡的神经元参数，这令许多人心生幻灭。

其实，在视觉艺术领域，人工神经网络已经可以将一幅作品的内容和风格分开，同时向艺术大师学习艺术风格，并把艺术风格转移到另外

的作品中，用不同艺术家的风格来渲染同样的内容(图7)^[18]。

这意味着人工神经网络可以精确量化原本许多人文科学中模糊含混的概念，例如特定领域中的“艺术风格”，博弈中的“棋风”，并且使这些只可意会、无法言传的技巧风格变得朴实无华，容易复制和推广。

5.4 其他方面

在游戏博弈方面，谷歌DeepMind团队开发的深度Q-网络DQN在49种Atari像素游戏中，29种达到乃至超过人类职业选手的水平。阿尔法狗更是完胜人类围棋顶级高手。

2016年5月，来自谷歌的 AI 实验室报道，研究者用2 865部英文言情小说培训机器，让机器学习言情小说的叙事和用词风格。从程序的演化过程看，机器模型先领悟了单词之间的空格的结构，然后慢慢认识了更多单词，由短到长，标点符号的规则也慢慢掌握，一些有更多长期相关性的语句结构，慢慢地也被机器掌握。

2016年5月，谷歌的DeepMind团队撰文他们开发了一个“神经编程解释器”(NPI)，这个神经网络能够自己学习并且编辑简单的程序，可以取代部分初级程序员的工作了。

6 人工智能商业化浪潮

Hinton 教授和他的两个研究生 Alex Krizhevsky 和 Ilya Sutskever 于2012 年底成立了一个名叫“深度神经网络研究”(DNN research)的公司，3个月后就谷歌以500万美元收购。Hinton 从此一半时间留在多伦多大学，另外一半时间在硅谷。两位研究生则成为谷歌的全职雇员。原来在纽约大学教书的Yann Lecun，2013 年底被脸书(Facebook)聘请为人工智能研究院的总管。曾在斯坦福大学和谷歌工作的吴恩达，2012年创立了网上教育公司 Coursera，2014年5月被百度聘任为首席科学家负责百度大脑的计划。

2015年，谷歌公布开源机器学习平台 TensorFlow；FaceBook 打造其专属机器学习平台 FBLearnerFlow，大幅提高员工效率；2015年5月，

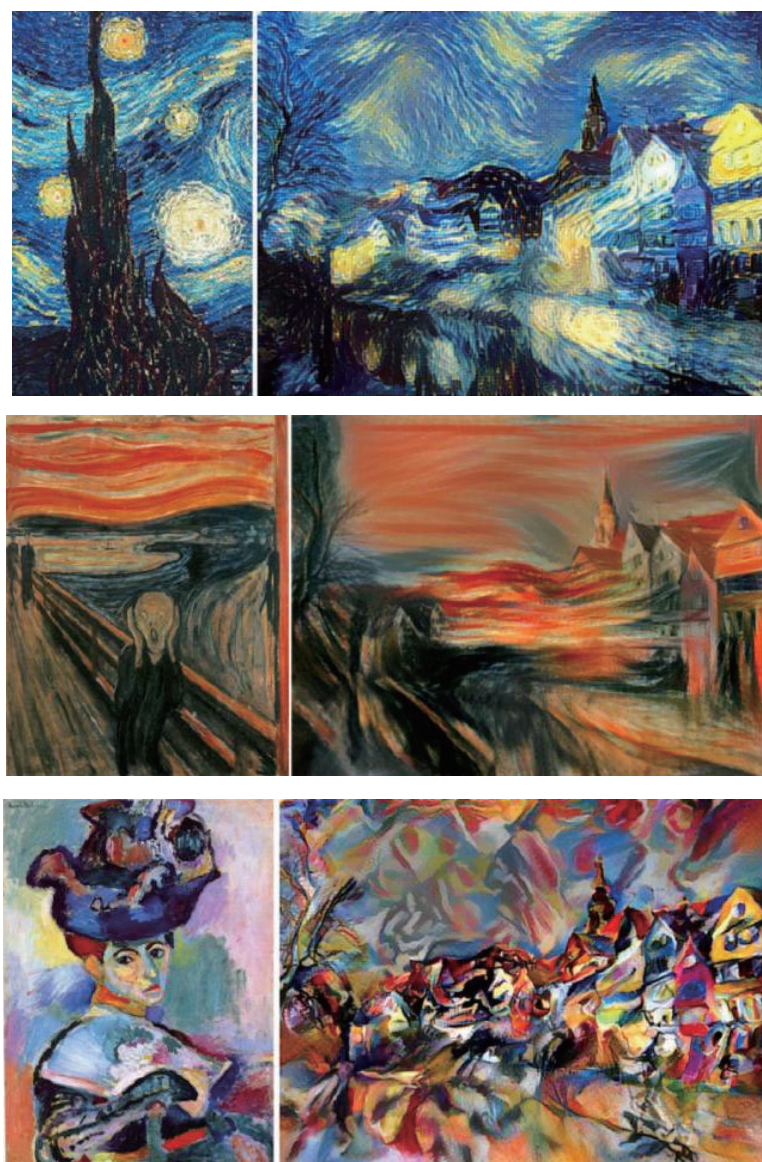


图7 神经网络能够自动学习艺术风格并用不同的风格渲染同样的内容

特斯拉创立开源人工智能系统OpenAI。其他工业巨头也纷纷斥巨资推动人工智能的发展,例如IBM的沃森系统、百度大脑计划、微软的同声翻译等等。

2016年的IBM正在率先推动全球人工智能的第一次商业化浪潮与核心业务转型。目前,深度学习的研究热点正在迅速转向基于深度卷积神经网络的物体检测与定位/分割能力,其突破将推动人工智能的实际应用与产业发展。目前研究热点是将深度卷积神经网络通过监督学习获得的表达,即所谓概念向量(thought vector)与推理、注意力、规划与记忆进行有机整合,涉及推理/规

划、注意力、短期/长期记忆、知识学习、知识蒸馏和知识迁移,小样本概念学习以及基于监督和再学习的大数据病历或棋谱的自动阅读与自主知识学习。

随着人工智能与大数据、云平台、机器人、移动互联网及物联网等的深度融合,人工智能技术与产业开始扮演着基础性、关键性和前沿性的核心角色。智能机器正逐步获得更多的感知与决策能力,变得更具自主性,环境适应能力更强;其应用范围也从制造业不断扩展到家庭、娱乐、教育、军事等专业服务领域。通过将大数

据转化为商业直觉、智能化业务流程和差异化产品/服务,人工智能开始逐步占据医疗、金融、保险、律师、新闻、数字个人助理等现代服务业的核心地位,并且不断渗入人们的日常生活。

7 展望

虽然人工智能取得了突破性进展,但是它还是在婴幼儿时期。联结主义的方法虽然摧枯拉朽、无坚不摧,但是依然没有坚实的理论基础。通过仿生学和经验积累得到的突破,依然无法透彻理解和预测。简单的神经网络学习机制加上机器蛮力,能否真正从量变到质变,这需要时间检验。如何通过小样本进行学习,特别是从周围环境自主学习(增强型学习),增加学习的泛化能力,这些都是人工智能研究的热点问题。

目前来看,人工智能在图像识别、语音识别、文本处理、游戏博弈、艺术美学、软件设计等诸多方面全面赶超人类。人工智能开始逐步占据医疗、金融、保险、律师、新闻、数字个人助理等现代服务业的核心地位,并且不断渗入人们的日常生活。

我们相信人工智能的发展将会为人类社会带来又一次技术革命,人工智能的浪潮正在汹涌澎湃!

(2016年5月10日收稿)■



参考文献

- [1] GAO X S, LI W, YUAN C M. Intersection Theory in differential algebraic geometry: generic intersections and the differential chow form [J]. Trans Amer Math Soc, 2013, 365(9): 4575-4632.
- [2] BREWER A A, LIU J J, WADE A R, et al. Visual field maps and stimulus selectivity in human ventral occipital cortex [J]. Nature Neuroscience, 2005, 8(8): 1102-1109.
- [3] SHARMAJ, ANGELUCCI A, SUR M. Induction of visual orientation modules in auditory cortex [J]. Nature, 2000, 404: 841-847.
- [4] VUILLERME N, CUISINIER R. Sensory supplementation through tongue electrotactile stimulation to preserve head stabilization in space in the absence of vision [J]. Investigative Ophthalmology & Visual Science, 2008, 50(1): 476-81.
- [5] MINSKY M, PAPERT S. Perceptrons: an introduction to computational geometry [M]. 1st ed. Cambridge: The MIT Press, 1969.
- [6] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back propagating errors [J]. Nature, 1986, 323(6088): 533-536.
- [7] YANN L C, BOSER B E, DENKER J, et al. Backpropagation applied to handwritten zip code recognition [J]. Neural Computation, 1989, 1(4): 541-551.
- [8] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets [J]. Neural Comput, 2006, 18(7): 1527-1554.
- [9] GLOTOT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks [J]. Journal of Machine Learning Research, 2011, 15: 315-323.
- [10] HINTON G E, SRIVASTAVA N, KRIZHEVSKY, et al. Improving neural networks by preventing co-adaptation of feature detectors [J]. Computer Science, 2012, 3(4): 212-223.
- [11] RAINA R, MADHAVAN A, NG A Y. Large-scale deep unsupervised learning using graphics processors [C]//Proceedings of 26th International Conference on Machine Learning, Montreal, 2009: 873-880.
- [12] DAN C C, MEIER U, GAMBARELLA L M, et al. Deep big simple neural nets excel on handwritten digit recognition [J]. Corr, 2010, 22(12): 3207-3220.
- [13] LE Q V, RANZATO M A, MONGA R, et al. Building high-level features using large scale unsupervised learning [C]//Proceedings of the 29th International Conference on Machine Learning, Edinburgh, Scotland, UK, 2012.
- [14] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database [M]//Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. Miami: IEEE, 2009: 248-255.
- [15] HINTON G, DENG L, YU D, et al. Deep neural networks for acoustic modelling in speech recognition: The shared views of four research groups [J]. IEEE Signal Processing Magazine, 29(6): 82-97.
- [16] GRAVES A, MOHAMED A R, HINTON G. Speech recognition with deep recurrent neural networks [J]. 2013. arXiv:1303.5778v1 [cs.NE].
- [17] AMODEI D, ANUBHAI R, BATTENBERG E, et al. Deep speech 2: end-to-end speech recognition in English and Mandarin [J]. Computer Science, 2015. arXiv:1512.02595v1 [cs.CL].
- [18] GATYS L A, ECKER A S, BETHGE M. Neural-style [EB/OL]. [2016-05-10]. <https://github.com/jcjohnson/neural-style>.

Historical review and current development of artificial intelligence

GU Xianfeng

Department of Computer Science, State University of New York at Stony Brook, NY 11794

Abstract This work gives a brief review of the history of artificial intelligence, and analyzes the current status of the field. The main principles and methodologies of the major branches in AI included symbolism and connectionism. Furthermore, the history, and booming reasons and major applications of deep learning are introduced as well.

Key words artificial intelligence, connectionism, symbolism, deep learning, image recognition, speech recognition, neuron network

(编辑: 段艳芳)