科学通报 2016 年 第 61 卷 第 33 期: 3548 ~ 3556
《中国科学》杂志社
SCIENCE CHINA PRESS
Science125个科学前沿问题系列解读(XIX)
What are the limits of learning by machines?

突破通过机器进行学习的极限

史忠植

中国科学院计算技术研究所, 北京 100190
E-mail: shizz@ics.ict.ac.cn

2016-06-27 收稿, 2016-09-01 修回, 2016-09-01 接受, 2016-09-20 网络版发表
国家重点基础研究发展计划(2013CB329502)资助

摘要 学习能力是人类智能的根本特征. 2016年3月, Google公司的AlphaGo把深度神经网络与蒙特卡罗树形搜索结合起来, 以4胜1负的成绩战胜了围棋世界冠军韩国的李世石. 这一结果标志人工智能取得了重大进展. 本文重点介绍AlphaGo采用的机器学习方法, 包括强化学习、深度学习、深度强化学习, 分析存在的问题和最新的研究进展. 为了突破通过计算机进行学习的极限, 提出认知机器学习, 列举可能的研究方向开展研究, 使机器智能不断进化, 逐步达到人类水平.

关键词 强化学习, 深度学习, 深度强化学习, 认知机器学习, 学习涌现, 学习进化

2005年7月1日, 在纪念*Science*创刊125周年之际, 科学家们总结出了125个问题, 其中第94个问题是“通过机器进行学习的局限是什么?”. 该问题的译文为“计算机已经可以击败世界上最好的国际象棋玩家, 他们在网络上可以抓取丰富的信息. 但抽象推理仍然超越任何机器”.

近期人工智能研究取得了很大的进展. 1997年, 人机大战兴起, IBM超级计算机“深蓝”击败了国际象棋大师卡斯帕罗夫. 2011年2月14日, IBM的“沃森”超级计算机正式登上美国最受欢迎的智力问答节目《危险边缘》(Jeopardy), 战胜了该节目的两名总冠军詹宁斯和鲁特尔. 2016年3月, Google公司的AlphaGo把深度神经网络与蒙特卡罗树形搜索结合起来, 以4胜1负的成绩战胜了围棋世界冠军韩国的李世石.

计算机击败了围棋世界冠军, 人工智能取得了重大进展, 但是在智能上与人类水平相比还相差很大. 必须把机器学习与脑认知结合起来, 开展认知机器学习, 使机器具有人类水平的智能.

1 学习是智能的必经之路

学习能力是人类智能的根本特征. 人从出生开始就不断地向客观环境和自身经历学习. 人的认识能力和智慧才能就是在毕生的学习中逐步形成、发展和完善[1].

100多年来, 心理学家和认知科学家在探讨人类学习理论的过程中, 由于各自的哲学基础、理论背景、研究手段的不同, 自然形成了各种不同的理论观点, 并形成了各种不同的理论派别, 主要包括行为学派、认知学派和人本主义学派. 有些心理学家用刺激与反应的关系, 把学习解释为习惯的形成, 认为通过练习使某一刺激与个体的某种反应建立一种前所未有的关系, 此种刺激反应间联结的过程, 就是学习. 因此, 此种理论被称为刺激反应论, 或称为行为学派. 行为学习理论强调可观察的行为, 认为行为的多次的愉快或痛苦的后果改变了个体的行为. 巴甫洛夫经典条件反射学说、华生的行为主义观点、桑代克的联结主义、斯金纳的操作条件反射学说以及班杜拉

引用格式: 史忠植. 突破通过机器进行学习的极限. 科学通报, 2016, 61: 3548–3556
Shi Z Z. Break through the limits of learning by machines (in Chinese). Chin Sci Bull, 2016, 61: 3548–3556, doi: 10.1360/N972016-00741

的社会学习理论可作为行为学派的代表学说.

有些心理学家不同意学习即习惯形成的看法, 他们特别强调理解在学习过程中的作用. 他们认为, 学习是个体在其环境中对事物间关系认知的过程, 这种理论被称为认知学派. 格式塔学派的学习理论、托尔曼的认知目的理论、皮亚杰的图式理论、维果斯基的内化论、布鲁纳的认知发现理论、奥苏伯尔的有意义学习理论、加涅的信息加工学习理论以及建构主义的学习理论均可作为认知学派的代表性学说. 认知主义学习理论的代表人物是皮亚杰、纽厄尔等.

人本主义心理学是20世纪50~60年代在美国兴起的一种心理学思潮, 其主要代表人物是马斯洛和罗杰斯(Rogers C R). 人本主义心理学家认为, 要理解人的行为, 就必须理解行为者所知觉的世界, 即要知道从行为者的角度来看待事物. 在了解人的行为时, 重要的不是外部事实, 而是事实对行为者的意义. 如果要改变一个人的行为, 首先必须改变他的信念和知觉. 当他看问题的方式不同时, 他的行为也就不同了. 换言之, 人本主义心理学家试图从行为者, 而不是从观察者的角度来解释和理解行为.

1956年, 人工智能正式创建以来, 机器学习是人工智能中发展最快的分支之一. 机器能否像人类一样具有学习能力呢? 1959年美国的Arthur Samuel设计了一个下棋程序, 这个程序具有学习能力, 它可以在不断的对弈中改善自己的棋艺. 4年后, 这个程序战胜了设计者本人. 又过了3年, 这个程序战胜了美国一个保持8年之久的常胜不败的冠军. 这个程序向人们展示了机器学习的能力, 提出了许多令人深思的社会问题与哲学问题.

一个学习系统总是由学习单元和环境两部分组成, 环境提供信息, 学习单元提供学习策略, 实现信息转换, 用能够理解的形式记忆下来, 并从中获取有用的信息. 20世纪80年代, 常见的机器学习方法有归纳学习、类比学习、分析学习、遗传学习、神经网络(连接)学习、等[2]. 1995年, Vladimir N. Vapnik 发表《统计学习理论的本质》(*The Nature of Statistical Learning Theory*)[3]. 1998年, Vladimir N. Vapnik 又发表《统计学习理论》(*Statistical Learning Theory*)[4], 把以支持向量机为代表的统计学习推向高潮, 贝叶斯概率等统计方法也得到广泛的应用[5].

2016年3月, Google围棋程序AlphaGo以4胜1负的成绩战胜了围棋世界冠军韩国的李世石, 标志人工智能研究, 特别在机器学习方面取得很大进展. AlphaGo在下棋过程中主要通过四步完成工作:

(1) 快速判断: 用于快速的观察围棋的盘面, 类似于人观察盘面获得的第一反应.

(2) 深度模仿: AlphaGo学习近万盘人类历史高手的棋局来进行模仿学习, 用得到的经验进行判断. 这个深度模仿能够根据盘面产生类似人类棋手的走法.

(3) 自学成长: AlphaGo不断与“自己”对战, 下了3000万盘棋局, 总结出经验作为棋局中的评估依据.

(4) 全局分析: 利用第三步学习结果对整个盘面的赢面判断, 实现从全局分析整个棋局.

AlphaGo把深度强化学习与蒙特卡罗树搜索算法结合起来[6], 取得了令人惊讶的成功.

2 强化学习

与人类一样, 智能体自己会学习获得成功的战略, 产生最大化的长期奖励. 这种通过试错、单纯地通过奖励或者惩罚完成的学习范式, 被称为强化学习(reinforcement learning, RL).

强化学习就是智能体从环境到行为映射的学习, 以使奖励信号(强化信号)函数值最大, 强化学习不同于连接主义学习中的监督学习, 主要表现在教师信号上, 强化学习中由环境提供的强化信号是对产生动作的好坏作一种评价(通常为标量信号), 而不是告诉强化学习系统如何去产生正确的动作. 由于外部环境提供的信息很少, 强化学习系统必须靠自身的经历进行学习. 通过这种方式, 强化学习系统在行动-评价的环境中获得知识, 改进行动方案以适应环境.

强化学习不是通过特殊的学习方法来定义的, 而是通过在环境中和响应外界环境的动作来定义的. 任何解决这种交互的学习方法都是一个可接受的强化学习方法. 强化学习也不是监督学习, 在有关机器学习的部分都可以看出来. 在监督学习中, “教师”用实例来直接指导或者训练学习程序. 在强化学习中, 学习主体自身通过训练, 误差和反馈, 学习在环境中完成目标的最佳策略.

强化学习技术是从控制理论、统计学、心理学等相关学科发展而来, 最早可以追溯到巴甫洛夫的条件反射实验. 但直到20世纪80年代末、90年代初强化学习技术才在人工智能、机器学习和自动控制等领域

中得到广泛研究和应用, 并被认为是设计智能系统的核心技术之一. 特别是随着强化学习的数学基础研究取得突破性进展后, 对强化学习的研究和应用日益开展起来, 成为目前机器学习领域的研究热点之一.

强化学习的模型如图1所示, 通过智能体与环境的交互进行学习[7]. 智能体与环境的交互接口包括动作(action)、奖励(reward)和状态(state). 交互过程可以表述为如下形式: 每一步, 智能体根据策略选择一个动作执行, 然后感知下一步的状态和即时奖励, 通过经验再修改自己的策略. 智能体的目标就是最大化长期奖励.

强化学习系统接受环境状态的输入s, 根据内部的推理机制, 系统输出相应的行为动作a. 环境在系统动作作用a下, 变迁到新的状态s'. 系统接受环境新状态的输入, 同时得到环境对于系统的瞬时奖惩反馈r. 对于强化学习系统来讲, 其目标是学习一个行为策略π: $S\rightarrow A$, 使系统选择的动作能够获得环境奖励的累计值最大. 在学习过程中, 强化学习技术的基本原理是: 如果系统某个动作导致环境正的奖励, 那么系统以后产生这个动作的趋势便会加强. 反之系统产生这个动作的趋势便减弱. 强化学习和生理学中的条件反射原理是相近的.

在实际应用中, 学习系统往往难以完全准确地观察到环境的真实状态, 而只能观察到真实状态的某一个或某几个方面. 这种状态观察上的不确定性为动作评估带来了更多的不确定性, 从而直接影响所选动作的好坏. 图2给出了部分感知状态马尔科夫模型. 在经典的马尔科夫模型上增加状态预测, 并对每个状态设置一个信度b, 用于表示该状态的可信度, 在决定动作时使用b作为依据, 同时根据观察值进行状态预测, 这样能解决一些非马尔科夫模型的问题.

由于部分感知问题的核心是观测状态的不确定性, 因此部分感知强化学习研究的核心是在学习过程中消除不确定性. 理论上, 这种不确定性可以通过概率来表示(信度), 然后构建基于信度的马氏决策过程. 但由于实际中信度是一个连续值, 问题转成大规模顺序决策任务, 需要用函数估计强化学习技术来解决, 效果并不理想.

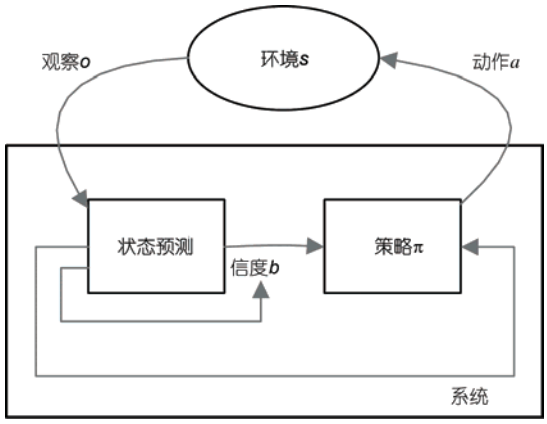

图 2 部分感知状态马尔科夫模型

Figure 2 Partially observable Markov decision process model

Soar强化学习的一个简单模型如图3所示[8], 反映智能体与环境的相互作用. 智能体试图采取动作, 期望获得未来最大的奖励. 奖励是一个信号, 它来自外部环境. 智能体的内部状态是由它的环境感知决定. 智能体维护一个值函数, 称为Q函数, 对于每个可以应用到状态的动作, 提供来自环境期望的奖励与状态的映射.

Soar对每个状态维护奖励结构, 提供外部和内部生成的奖励支持. 为了获得最大的灵活性, 该奖励值是通过智能体的知识(规则)设置. 该智能体可以基于从环境外部输入的奖励. 最简单的情况是通过复制特定的传感器奖励值. 该智能体可以自己判定, 创建内部奖励结构的适当值.

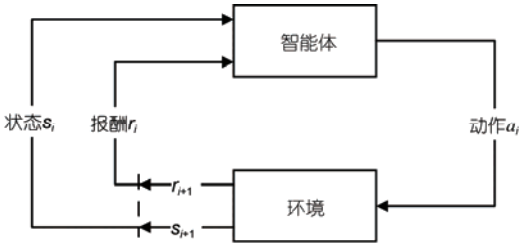

图 1 强化学习模型

Figure 1 Reinforcement learning model

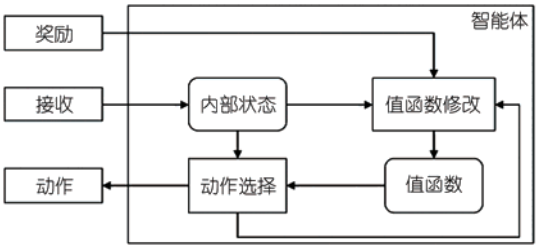

图 3 Soar的强化学习智能体

Figure 3 Soar reinforcement learning agent

3 深度学习

大脑善于处理非常高维的数据，并做出理解. 来自视觉神经的信息通常是百万量级的，并且几乎是即时的. 当我们过马路获得视觉输入是红灯时，我们就会停下来. 深度学习是机器学习研究中的一个新的领域，其核心思想在于模拟人脑的层级抽象结构，通过无监督的方式分析大规模数据，发掘大数据中蕴藏的有价值信息.

1981 年的诺贝尔医学奖颁发给了 David Hubel 和Torsten Wiesel，以及Roger Sperry. 前两位的主要贡献是“发现了视觉系统的信息处理”：可视皮质是分级的. 2002 年，多伦多大学的Geoffrey E. Hinton提出了一种名为 Contrastive Divergence (CD)的机器学习算法[9]，它可以高效地训练一些结构不太复杂的马尔可夫随机模型，其中就包括受限玻尔兹曼机(restricted Boltzmann machine, RBM). 这为后来深度学习的诞生奠定了基础. 2006年，Hinton等人[10]提出了一种深度信念网络. 一个深度神经网络模型可被视为由若干个RBM堆叠在一起，这样一来，在训练的时候，就可以通过由低到高逐层训练这些RBM来实现.

1989 年，LeCun 等人[11]提出卷积神经网络(convolutional neural network, CNN)，这是一种多阶段全局可训练的人工神经网络模型，它可以从经过少量预处理、甚至原始数据中学习到抽象的、本质的和高阶的特征. 1998年LeCun设计实现了LeNet-5，并用于手写数字识别. 目前在ImageNet视觉识别竞赛中，CNN击败了所有其他的算法. CNN的结构主要有稀疏连接和权值共享两个特点，具体操作包括：

(1) 特征提取：每一个神经元从上一层的局部接受域得到突触输入，因而迫使它提取局部特征.

(2) 特征映射：网络的每一个计算层都是由多个特征映射组成的，每个特征映射都是平面形式的. 平面中单独的神经元在约束下共享相同的突触权值集.

(3) 子抽样：每个卷积层后面跟着一个实现局部平均和子抽样的计算层，由此特征映射的分辨率降低. 这种操作具有使特征映射的输出对平移和其他形式变形的敏感度下降的作用.

2016年5月底，斯坦福大学召开了IEEE Computer Society 2016年认知计算会议，学者指出，ImageNet竞赛用于比较不同算法之间的性能可能有用，但比较不同算法与人类之间的区别则没多大用处. 鉴于网络架构，人类没有办法确保CNN是否会在全新的情境下造成毁灭性的错误. CNN训练好以后，无论是通过定性还是定量分析，几乎都没有办法预测网络会对新的输入产生怎样的结果. IBM 研究院的Welser[12]认为，关于未来的发展趋势并不是逐渐建立规模更大的 CNN，也不会只使用一种通用算法. Welser表示，系统需要很多不同的组件，每个组件都有特定的功能. 从宏观层面上讲，认知系统包括对真实世界模拟建模，以及对直接从现实世界采集的非结构化数据进行大数据分析. 认知系统将结合这两者并优化模型，得出一个所有可能情形的分布，并从中选择让能够自己接近目标的结论. 这种模型既可以用来形容自动驾驶汽车，同样说它是辩论机器人也没问题.

4 深度强化学习

深度强化学习是将深度学习与强化学习结合起来，实现从感知到动作的学习的算法. 简单地说，这和人类一样，输入感知信息比如视觉，然后通过深度神经网络，直接输出动作. 深度强化学习具备使机器人实现完全自主的学习一种甚至多种技能的潜力.

普通的强化学习虽然应用的比较成功，但是特征状态需要人工设定，对于复杂的场景，是个很困难的事情，特别容易造成维数灾难，同时表达的还不好. 2010年，Lange和Riedmiller[13]提出了采用Deep auto-encoder提取特征，用于基于视觉的相关控制. 2013年，DeepMind在NIPS (Neural Information Processing Systems)深度学习专题会上提出DQN (deep Q-network)方法[14]，采用卷积神经网络提取特征，再应用在强化学习上面. 他们不断改进，2015年在*Nature*上发表了改进版的DQN文章[15]，引起了广泛的关注.

DeepMind用一种由机器学习若干阶段组成的流程训练神经网络(图4)[6]. 开始阶段，直接使用人类高手的落子弈法训练监督学习(supervised learning, SL)型走棋策略网络p_σ. 此阶段提供快速、高效的带有即时反馈和高品质梯度的机器学习更新数据. 训练快速走棋策略p_π，能对走子时的弈法快速采样. 接下来的阶段，训练一种强化学习(reinforcement learning, RL)型的走棋策略网络p_ρ，通过优化那些自我博弈的最终结果，来提高前面的监督学习策略网络. 此阶段

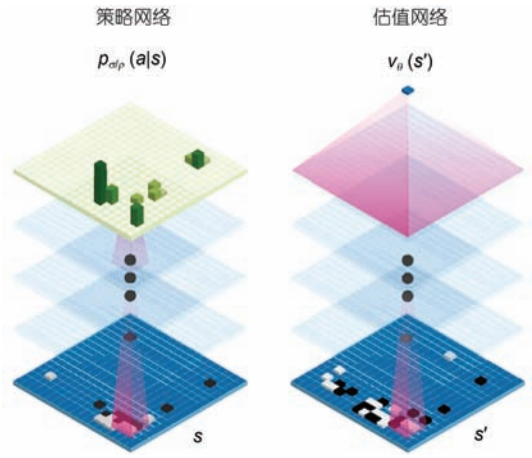

图 4 (网络版彩色)训练神经网络流程
Figure 4 (Color online) Neural network training pipeline

是将该策略调校到赢取比赛的正确目标上，而非最大程度的预测准确性. 最后阶段，训练估值网络V_θ来预测那些采用强化学习走棋策略网络自我博弈的赢家. 程序AlphaGo用蒙特卡洛树搜索(Monte Calro Tree Search, MCTS)[16]，有效结合了策略和估值网络.

近年来，在记忆强化神经网络领域的相关工作取得了一些很有趣的进展，例如2014年Alex Graves等人[17]提出的神经图灵机，是基于记忆强化神经网络的体系架构. 该架构包含2个基本组件：神经网络控制器和内存池. 这种架构通过加入可读写的外部存储器层，实现用极少量新观测数据就能有效对模型进行调整，从而快速获得识别未见过的目标类别的元学习能力，也就是可以利用极少量样本学习. 这种调整不是简单通过对新观测信息在存储器中查找匹配，而是基于强大的深度神经网络架构，结合长期观测得到的深度模型与根据新信息对存储内容灵活有效的读写更新.

2016年的机器学习会议上，Google DeepMind的Santoro等人[18]提出了新的存储读写更新策略，每次写操作只选择最少被用到的存储位置或者最近被用的存储位置. 这样的策略完全由内容决定，不依赖于存储的位置，而神经网络图灵机的更新策略则是由信息内容和存储位置共同决定的. Danihelka等人[19]的报道了“关联长短时记忆”在增加记忆但不增加网络参数数量的情况下，强化循环神经网络. 该系统具有基于复数向量的关联记忆，与全息约简表示(holographic reduction of rule, HRR)和长短时记忆(long short-term memory, LSTM)网络紧密相关.

Mnih等人[20]建立了一个大规模的分布式深度强化学习网络Gorila. 该算法使用并行actor-learner 更新一个共享模型，对于研究的3种基于估值的算法学习过程都有提升稳定性影响. 使用Google云平台，Gorila的训练速度提升了一个数量级.

Silver[21]指出，使用深度强化学习，让智能体在游戏上取得进展. 在DeepMind，把深度强化学习融合起来，创造了第一批能够在许多挑战性的领域实现人类级别表现的智能体. 我们的智能体必须持续地进行估值判断，以选择最佳行动. 这种知识的代表是一个Q-network，它能够评估一个智能体在采取了某一个具体的行动后可以获得的全部奖励. 许多技术都已经用到了现实生活中. 将来还有拓展到机器人的移动和控制以及医疗领域的可能. Google DeepMind还发展了深度增强学习博弈论，让程序成为了一个超人类的德州扑克手. 相对于围棋这种完美信息博弈(落子明确，无需猜测)，扑克是不完美信息博弈(需要猜测对手的牌)，也更接近现实生活情景. 程序在没有任何先验知识的前提下，使用可扩展的端到端学习近似纳什均衡的方法，结合深度强化学习技术和虚拟自我对局. 实验中，计算机通过自学成功掌握了德州扑克的技巧，其表现已经接近人类专家水平.

5 认知机器学习

陆汝钤院士在周志华著的《机器学习》序言中，引用了美国人工智能资深学者、俄亥俄州立大学Chandrasekaran教授的来信：“最近几年，人工智能在很大程度上集中于统计学和大数据. 我同意由于计算能力的大幅提高，这些技术曾经取得过某些令人印象深刻的成果. 但是我们完全有理由相信，虽然这些技术还会继续改进、提高，总有一天这个领域(指AI)会对它们说再见，并转向更加基本的认知科学研究. 尽管钟摆的摆回去还需要一段时间，我相信定有必要把统计技术和对认知结构的深刻理解结合起来”[22]. 这里，Chandrasekaran教授给出了明确的启示，机器学习必须与认知科学的研究结合起来.

在中国人工智能学会主持的“关于H. A. Simon学术思想研讨会”上，李衍达院士报告“沿着Simon开拓下去”[23]. 报告中他指出“如果计算机或者说机器脑跟人脑一样，不仅具有信息处理意义上的一致而且具有进化功能，那么我就相信人脑可能具有的一切

机器脑也会有.”

为了解决机器具有抽象推理的能力，计算机能够学习进化，必须开展认知机器学习研究. 认知机器学习是指把机器学习与脑认知机理结合起来.

图5给出了心智模型CAM[8]. 脑的高级认知功能包括学习、记忆、语言、思维、决策、情感等. 学习是通过神经系统不断接受刺激，获得新的行为、习惯和积累经验的过程，而记忆是指学习得到的行为和知识的保持和再现，是我们每个人每天都在进行着的一种智力活动. 语言和高级思维是人区别于其他动物的最主要因素. 决策是指通过分析、比较，在若干种可供选择的方案中选定最优方案的过程，也可能是对不确定条件下发生的偶发事件所做的处理决定. 情感是人对客观事物是否满足自己的需要而产生的态度体验. 长时记忆包括语义记忆、情景记忆、程序性记忆.

根据当前研究进展，下面列举几个认知机器学习可能感兴趣的问题.

5.1 学习涌现

毛泽东在实践论中指出: “认识的过程，第一步，是开始接触外界事情，属于感觉的阶段. 第二步，是综合感觉的材料加以整理和改造，属于概念、判断和推理的阶段. 只有感觉的材料十分丰富(不是零碎不全)和合于实际(不是错觉)，才能根据这样的材料造出正确的概念和论理来”[24]. 如何从感性认识上升为理性认识，即学习涌现.

Holland[25]的著作《涌现》(Emergence: From Chaos To Order)指出: “涌现的本质就是由小生大，由简入繁”. 涌现现象研究的方向和路标包括: 机制(积模块、生成器、主体)和永恒的新奇(大量不断生成的结构)，动态性和规律性(在生成结构中持续的重复发生的结构或模式)，具有层次的组织(在生成器结构的基础上，生成更高组织层次的生成器)，建模，尤其是计算机模型，可以提供许多涌现方面的例子. 具有自学习功能的西洋跳棋程序就是一个案例.

5.2 程序性记忆知识学习

2015年12月, *Science*刊登了纽约大学数据科学中心的Brenden Lake、多伦多大学计算机科学与统计学系的Ruslan Salakhutdinov和麻省理工学院大脑与认知科学系的Joshua Tenenbaum的文章，只需向学习系统展示一个来自陌生文字系统的字符，它就能很快学到精髓，像人一样写出来[26].

根据Brenden Lake等的介绍，他们在论文中分析了三个核心原则. 这些原则都很通用，既可以用在字符上，也可以用在其他的概念上: 组合性(compositionality)，表征是由更简单的基元构建而成; 因果性(causality)，模型表征了字符生成的抽象因果结构; 学会学习(learning to learn)，过去的概念知识有助于学习新的概念.

三位研究者采用的方法是“贝叶斯程序学习(Bayesian Program Learning: BPL)”，能让计算机系统对人类认知进行很好的模拟. 传统的机器学习方法

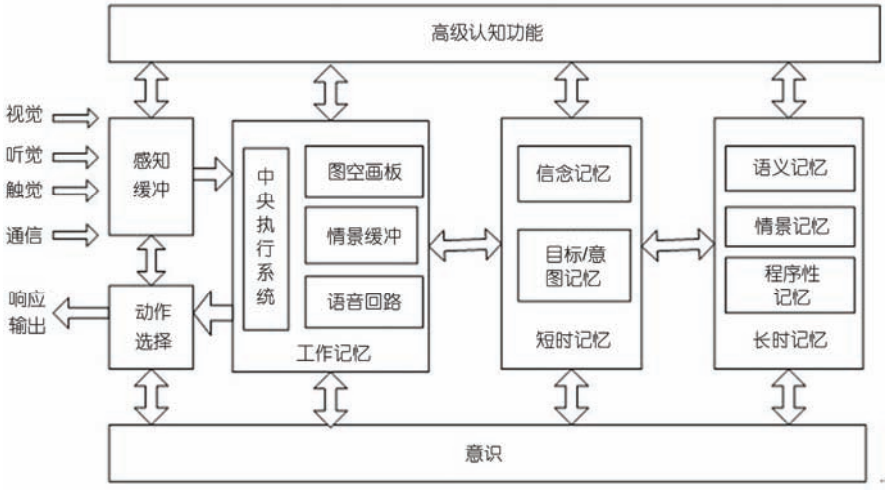

图 5 心智模型CAM
Figure 5 Mind model CAM

需要大量的数据来训练, 而这种方法只需要一个粗略的模型, 然后使用推理算法来分析案例, 补充模型的细节.

感知任务与操作任务相结合. 人类的感知工作的一个目的是为操作任务而服务, 而人类在执行操作任务的时候有时只需要对环境有大概认知即可. 以视觉感知为例, 感知任务样本之间差异性巨大, 因此需要大量标注样本, 相比而言, 操作任务真值的复杂度则小许多, 例如不管在什么环境中, 人类的行走方式是大致不变的. 若能直接从输入出发, 在获得一定程度的感知技能的基础上直接拟合操作任务的函数, 则有可能降低对于感知任务数据的需求.

5.3 学习进化

为适应外界而改变自身结构的进化, 是世界上最重要的机理之一. 进化、学习再经过高级的进化、学习的进化, 产生了目的性, 这个实际上是一个关键, 随机的无目的的机器能通过学习去探索其本身的目的. 19世纪中叶, 达尔文创立了生物进化学说. 生物通过遗传、变异和自然选择, 从低级到高级, 从简单到复杂, 种类由少到多地进化着、发展着.

对智力来说, 所谓的进化是指学习的学习, 这个学习的学习跟软件不同, 它是结构也跟着变化, 这是很重要的一条, 而且结构变化把学习的结果记录下来, 还改进了学习方法, 而且它的存储和运算是一体的, 这是目前计算机难以做到的. 这个地方研究计算机的学习进化模式恐怕就是一个新的课题, 是非常值得关注的课题.

对古人类头骨化石的研究揭示了人类大脑的发展, 在200万年的演变过程中, 人类大脑体积增加了3倍. 由于人类的食物、文化、技能、群体和基因等各种因素的共同作用, 最终导致了现代人类的大脑在20万年前进化成功. 现代人类脑容量约为1200~1400 mL. 在猿和类人猿阶段, 其智力发展是缓慢的, 到了能人、早期智人和晚期智人, 人类智力进化迅速提升, 许多人类特有的皮质中枢正是在这一时期产生, 如运动性语言中枢、书写中枢、听觉性语言中枢等. 同时, 大脑皮质还出现了欣赏音乐和绘画的中枢, 这些中枢都有明显的定位特点. 尤其是随着人类抽象思维发展, 人脑额叶迅速扩张. 由此可见, 现代人脑是不断进化而成的.

要使机器具有人类水平的智能, 突破通过计算机进行学习的局限, 必须让机器具有学习进化的功能. 通过学习, 不仅增长知识, 而且使机器的记忆结构发生变化.

6 结语

学习能力是人类智能的根本特征. Google公司的AlphaGo把深度神经网络与蒙特卡罗树形搜索结合起来, 通过学习, 以4胜1负的成绩战胜了围棋世界冠军韩国的李世石. 这一结果标志人工智能取得了重大进展. 本文重点介绍了AlphaGo采用的机器学习方法, 包括强化学习、深度学习、深度强化学习, 分析存在的问题和最新的研究进展. 为了突破通过计算机进行学习的极限, 本文提出了认知机器学习, 可以在学习涌现、程序性记忆知识学习、学习进化等方面开展研究, 突破机器学习的极限, 实现机器智能达到人类水平.

参考文献

1 Shi Z Z. Intelligence Science. Series on Intelligence Science-Vol.2. Singapore: World Scientific Publishing Co., 2012
2 Shi Z Z. Principles of Machine Learning. Beijing: International Academic Publishers, 1992
3 Vapnik V N. The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995
4 Vapnik V N. Statistical Learning Theory. New York: Wiley-Interscience Publication, John Wiley & Sons, Inc., 1998
5 Li H. Statistical Learning Methods (in Chinese). Beijing: Tsinghua University Press, 2012 [李航. 统计学习方法. 北京: 清华大学出版社, 2012]
6 Mnih V, Kavukcuoglu K, Silver D, et al. Mastering the game of Go with deep neural networks and tree search. Nature, 2016, 529: 484–489
7 Shi Z Z. Advanced Artificial Intelligence. Series on Intelligence Science-Vol.1. Singapore: World Scientific Publishing Co., 2011
8 Shi Z Z. Mind Computation (in Chinese). Beijing: Tsinghua University Press, 2015 [史忠植. 心智计算. 北京: 清华大学出版社, 2015]
9 Hinton G E. Training products of experts by minimizing contrastive divergence. Neural Comp, 2002, 14: 1771–1800
10 Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. Science, 2006, 313: 504–507

11 LeCun Y, Boser B, Denker J S, et a1. Handwritten digit recognition with a back-propagation network. In: Advances in Neural Information Processing Systems. Denver: Morgan Kaufmann, 1989. 396–404
12 Welser J J. Cognitive Computing: Augmenting Human Capability. In: IEEE New Frontiers in Computing 2016, Cognitive Computing: to the Singularity and beyond, 2016
13 Lange S, Riedmiller M. Deep auto-encoder neural networks in reinforcement learning. In: International Joint Conference on Neural Networks, 2010. 1–8
14 Mnih V, Kavukcuoglu K, Silver D, et al. Playing Atari with deep reinforcement learning. arXiv: 1312.5602, 2013
15 Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. Nature, 2015, 518: 529–533
16 Coulom R. Efficient selectivity and backup operators in Monte-Carlo tree search. In: 5th International Conference on Computers and Games, 2006, 72–83
17 Graves A, Wayne G, Danihelka I. Neural turing machines. arXiv: 1410.5401v2, 2014
18 Santoro A, Bartunov S, Botvinick M, et al. One-shot Learning with Memory-Augmented Neural Networks. International Conference on Machine Learning, New York, 2016. 1842–1850
19 Danihelka I, Wayne G, Uria B, et al. Associative Long Short-Term Memory. In: International Conference on Machine Learning, New York, 2016. 1986–1994
20 Mnih V, Badia A P, Mirza M, et al. Asynchronous Methods for Deep Reinforcement Learning. In: International Conference on Machine Learning, New York, 2016. 1928–1937
21 Silver D. Google's DeepMind creating artificial agents to achieve human-level performance. https://futuristech.info/posts/google-s-deepmind-creating-artificial-agents-to-achieve-human-level-performance, 06/22/2016
22 Zhou Z H. Machine Learning (in Chinese). Beijing: Tsinghua University Press, 2016 [周志华. 机器学习. 北京: 清华大学出版社, 2016]
23 Li Y D. Open up along Simon (in Chinese). 2016, http://www.caai.cn/index.php?s=/Home/Article/detail/id/167.html [李衍达. 沿着 Simon 开拓下去. 中国人工智能网站, 2016. http://www.caai.cn/index.php?s=/Home/Article/detail/id/167.html]
24 Mao Z D. Mao Zedong Anthology (Vol. 1) (in Chinese). Beijing: People's Publishing House, 1991. 282–298 [毛泽东. 毛泽东选集. 第一卷. 北京: 人民出版社, 1991. 282–298]
25 Holland J H. Emergence: From Chaos to Order. Cambridge: Perseus Publishing, 1999
26 Lake B M, Salakhutdinov R, Tenenbaum J B. Human-level concept learning through probabilistic program induction. Science, 2015, 350: 1332–1338

史忠植

博士生导师, 中国科学院计算技术研究所研究员, 中国计算机学会会士, 中国人工智能学会会士, IEEE 高级会员, AAAI 和 ACM 会员、IFIP 人工智能学会机器学习和数据挖掘工作组主席. 1964 年毕业于中国科技大学计算机专业, 1968 年毕业于中国科学院研究生院. 2013 年获得中国人工智能学会吴文俊人工智能科学技术成就奖. *Int J Intell Sci* 主编. 发表著作 16 本, 学术论文 500 多篇. 曾担任中国计算机学会秘书长, 中国人工智能学会副理事长.

Break through the limits of learning by machines

SHI ZhongZhi

Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

Learning ability is the basic characteristic of human intelligence. The July 1, 2005 issue of *Science* published a list of 125 important questions in science. Among them, the question 94 "What are the limits of learning by machines?". The annotation "Computers can already beat the world's best chess players, and they have a wealth of information on the Web to draw on. But abstract reasoning is still beyond any machine". In recent artificial intelligence has made great progresses. In 1997, the rise of the man-machine war, IBM Supercomputer Deep Blue defeated the chess master Garry Kasparov. On February 14, 2011, IBM's Watson supercomputer won a practice round against Jeopardy champions Ken Jennings and Brad Rutter. In March 2016, Google DeepMind's AlphaGo sealed a 4-1 victory over a South Korean Go grandmaster Lee Se-dol. This paper focuses on the machine learning methods of AlphaGo, including reinforcement learning, deep learning, deep reinforcement learning, analysis of the existing problems and the latest research progress. Deep reinforcement learning is the combination of deep learning and reinforcement learning, which can realize the learning algorithm from the perception to action. Simply said, this is the same as human behavior, input sensing information such as vision, and then, direct output action through the deep neural network. Deep reinforcement learning has the potential to learn a variety of skills for the robot to achieve full autonomy. Even though reinforcement learning is practiced successfully, but feature states need to manually set, for complex scene is a difficult thing, especially easy to cause the dimension disaster, and expression is not good. In 2010, Sascha Lange and Martin Riedmiller proposed deep auto-encoder neural networks in reinforcement learning to extract feature, which is used to control the visual correlation. In 2013, DeepMind proposed deep Q-network (DQN) in NIPS 2013, using convolution neural network to extract features, and then applied in reinforcement learning. They continue to improve and published an improved version of DQN on *Nature* in 2015, which has aroused widespread concern. In order to break through the limits of learning by machines, cognitive machine learning is proposed, which is the combination of machine learning and brain cognition, so that the machine intelligence is constantly evolving, and gradually reaches the human level of artificial intelligence. A cognitive model entitled Consciousness And Memory (CAM) is proposed by author, which consists of memory, consciousness, high-level cognitive functions, perception and motor. High-level cognitive functions of the brain include learning, language, thinking, decision making, emotion, and so on. Learning is a course to accept the stimulus through the nervous system and obtain new behavior, habits and accumulation experience. According to the current research progress of brain science and cognitive science, cognitive machine learning may be interested in learning emergence, procedural memory knowledge learning, learning evolution and so on. For intelligence, so-called evolution is refers to the learning of learning and the structure also follows the change. It is important to record the learning result by structure changing and improve the learning method.

reinforcement learning, deep learning, deep reinforcement learning, cognitive machine learning, learning emergence, learning evolution

doi: 10.1360/N972016-00741