

从图灵测试到深度学习:人工智能 60 年

万贇

美国休斯敦大学维多利亚分校,德克萨斯州维多利亚 77901

摘要 2015 年深度学习和 2016 年谷歌 AlphaGo 围棋在各国主流媒体上争取报道,也引发了大众对人工智能的兴趣。人工智能在 60 年前就已经出现并且几经起落,近两年的突破在很大程度上是由于云计算平台的发展。在过去的 20 年间,人工智能以智能软件代理、推荐系统、数据挖掘、知识管理、深度学习等各种应用层面的分支概念存在。追本溯源,本文概括介绍人工智能过去 60 年的发展历史。

关键词 人工智能;深度学习;云计算;电子商务;计算历史

1 人工智能诞生

2014 年英美合作的电影《模仿游戏》讲述了英国数学家艾伦·图灵 60 年前在二战期间帮助设计电子计算机破译纳粹德国军事密码的真实故事。不过影片的名字与图灵在电影中的事迹并没有直接关系,而是来自当时英国流行的游戏。游戏中一男一女分别躲在幕后,参加游戏的人向他们两个人不断提问题,他们通过无法识别笔迹的笔答方式来回答,提问题的人根据回答来判断这两个人的性别。1950 年图灵在《计算机与智能》^[1]一文中借用这个游戏做为判断计算机是否具有人类智能的标准,也就是把一个人和一台计算机放在幕后,让测试人员通过提问来判断哪一个是计算机,如果判断错误的话,就认为计算机通过了图灵测试,具有人的智能。后来人工智能学者将图灵这篇论文中描述的计算机称为图灵机,这一测试方式称为图灵测试(注:图灵本人曾经预测随着足够多内存的出现,50 年内(也就是到 2000 年)计算机能够达到图灵机的标准)。40 年后的 1990 年英国剑桥大学为了推动图灵机的实现设立了总额为 10 万美元的图灵测试比赛(The Leobner Prize),每年举办一次,

直到有人能够设计出达到标准的图灵机为止,结果 16 年后的今天比赛仍在继续。尽管如此,图灵测试为人工智能领域的发展竖立起一个目标。随着几代人工智能学者的不断研究,逐渐认识到人脑的高度复杂性和计算机的局限性。这些发现帮助我们不断把人工智能技术应用到生产和生活的诸多方面。

第二次世界大战期间,英国军方需要尽快破译纳粹德国军事密码,图灵参与并主导了英国电子计算机的研制。同时期美国科学家也投入到计算机的研制工作。战后电子计算机开始从纯军事领域转向民用,很多具有不同学术背景的专家聚拢到计算机的应用领域,其中不乏青年才俊。1956 年夏天,一批学者聚集在达特茅斯学院召开了一次关于人工智能的研讨会,该会议成为人工智能领域正式创立的标志。会议的召集人是达特茅斯学院的数学系助理教授麦卡锡(John McCarthy),参与人包括明斯基(Marvin Minsky),西蒙(Herbert Simon)以及西蒙的合作伙伴纽厄尔(Allen Newell)。由于不同的学术背景和对智能以及实现智能的不同看法,人工智能从一开始就形成两类不同的流派和方法。

第一类是以西蒙和纽厄尔代表的理性学派。这一学派认为人脑与计算机可以看成信息处理器。计算机的智能和人脑的智能主要表现在对抽象化问题的解决上。任何能够以一定的逻辑规则描述的问题都可以通过人工智能程序来计算解决,尤其是对人脑来说过于复杂的逻辑问题。西蒙有一段时间专门研究公司里人们的行为决策,他发现由于认知能力的限制,人在做决策时并不像经济学里的理性人描述地那样,能够总是去寻求最优解,大多数情况下是寻找能够满足最低要求的解决方案。他将这一现象称为“满意解(satisficing)”,用来区别对理性人所假设的“最优解(optimizing)”。他认为这一现象产生的原因是人理性的有限性或者说是有限理性的理性(bounded rationality)。在这一观念的影响下,他认为计算机带来的人工智能可以大大延伸人类理性。按照这一思路,西蒙和纽厄尔在人工智能领域取得了不少成果。1955 年他们设计的逻辑机程序成功证明了罗素和怀特海所著的《数学原理》一书提出的 52 个定理中的 38 个,其中不少证明比原书中的更加精彩。根据对逻辑机的研究,1957 年他们又设

收稿日期:2016-03-25

作者简介:万贇,副教授,研究方向为电子商务和互联网应用,电子信箱:WanY@uhv.edu

引用格式:万贇. 从图灵测试到深度学习:人工智能 60 年[J]. 科技导报, 2016, 34(7): 26-33; doi: 10.3981/j.issn.1000-7857.2016.07.003

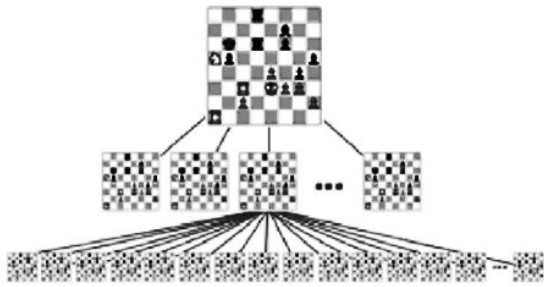


图1 无论是国际象棋还是围棋,目前都是通过对树结构的广度 b (每一步可能的下法)与深度 d (总回合)搜索相结合进行判断(bd),国际象棋的搜索维度(广度35;深度80)远远少于围棋(广度250;深度150)

计编写了一般问题解决器程序(general problem solver),希望以此来解决任何可以形式化的符号问题(定理证明、几何问题及国际象棋对抗等)。

西蒙所代表的理性流派虽然在机器定理证明和简单逻辑问题解决(比如汉诺塔问题)上取得了显著的成就,但一旦面对复杂的问题,有限的计算机内存空间很快就因为探索问题解决空间时考虑不同路径带来的组合爆炸而无法进行下去。由于同样的原因,很多人工智能专家认为计算机程序虽然可以击败人类国际象棋冠军,但可能永远无法击败人类围棋冠军,因为后者的探索空间太大(图1)。

与西蒙他们的理性学派在方法上形成对比的是感性学派。感性学派简单说就是通过对神经的模拟来获得人工智能。

人脑神经元一般由细胞体、树突和轴突组成。树突用来接收其他神经元传递过来的信息,轴突及其顶端的突触则用来传递信息给其他神经元。1949年加拿大神经心理学家赫布提出理论,

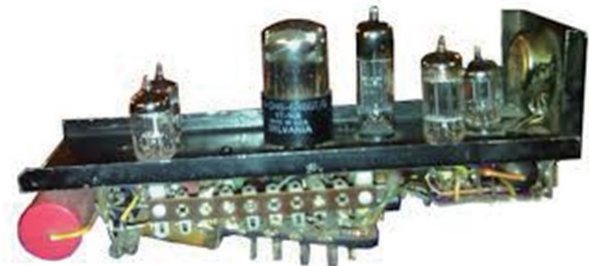


图2 SNARC使用的真空管是当时电子计算机的基本原件,每一个真空管通过开和关两种状态代表一个比特的信息

解释了人脑在学习过程中脑神经元发生的变化。赫布理论认为如果一个神经元持续激活另一个神经元,这种持续重复的刺激可以导致突触传递效能的增加。具体表现为前者的轴突将会生长出突触小体(如果已有,则会继续长大),并和后者

的胞体相连接,形成记忆痕迹。当时正在哈佛读本科的明斯基受到启发,产生了制作一个电子模拟神经网络实现人工智能的想法。1951年在美国心理学大师米勒的帮助下,明斯基和帕尔特(Seymour Papert)获得了美国海军经费资助设计出世界上第1台用来对迷宫求解的电子神经网络(SNARC),包含40个电子神经和若干内存。每一个电子神经元由6个真空管和一个马达组成,整个设备使用了300多个真空管(图2)。这一贡献使明斯基被认为是人工神经网络的先驱。

最早把神经网络原理成功应用到图像识别的是康奈尔大学的心理学教授罗森布拉特。他和明斯基都毕业于纽约布朗克斯科学高中,而且是后者的学长。1957年他利用神经网络原理成功制做了电子感知机(Perceptron,图3),该设备因为能够读入并识别简单的字母和图像而在当时引起轰动,使得很多专家预测在几年后计算机将具备思考功能。

明斯基和罗森布拉特设计的人工神经网络以及后来人工智能专家在计算机上虚拟生成的更复杂的人工神经网络,都是通过模拟人脑神经细胞的记忆结构来实现的。由于神经网络链接的权重分布需

要根据输入的信息不断调整,但是调整过程对外界来说是一个黑盒子,所以在设计不同的人工神经网络时,除了遵循一些基本原则外,更多需要通过经验和直觉来进行,据此有人称人工神经网络的设计为一门“艺术”而非“科学”,与西蒙等所倡导的理性学派形成了显著区别。

2 难以逾越的局限

20世纪50年代人工智能的诞生和流行除了与电子计算机的问世有直接联系外,也与当时美苏冷战有很大关联。1954年美苏冷战期间美国军方希望计算机产生的人工智能能够自动翻译俄语情报和资料,所以对在这一领域的突破赋予很高的期望,投入了大量的财力和人力。除了海军和空军,20世纪60年代初期互联网先驱立克里德在美国国防部高级计划署任职期间,每年资助麻省理工(明斯基)、卡梅(西蒙和纽厄尔)和斯坦福(麦卡锡)数百万美元的经费从事计算机以及人工智能方面的研究。

而人工智能的研发在经历了前面描述的这些突破后很快就变得停滞不

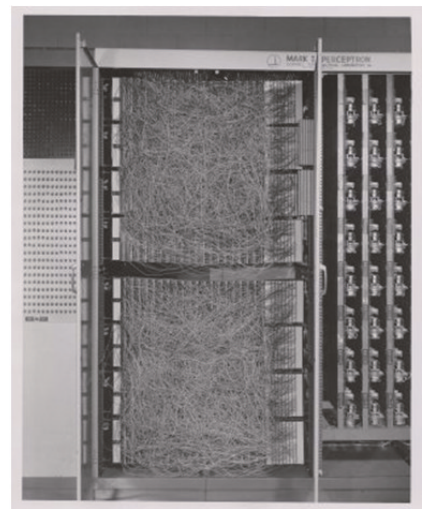


图3 1957年,罗森布拉特根据赫布原理用IBM 704完成了感知机的仿真。后来用马克1号制作了上面这台基于感知机的神经计算机,成功实现了一些英文字母的识别。1960年6月23日该神经计算机进行了公众展示,该计算机与一个能够摄取400个像素(20x20)的感光板相连

前,最主要的原因是计算机内存和运算速度的限制。比如 20 世纪 60 年代末 SDC 公司奎利恩(Ross Quillian)等设计的颇为成功的自然语言翻译程序所使用的计算机内存最多只能同时导入 20 个单词。1976 年,从事机器人研究的人工智能专家莫拉维克(Hans Moravec)估计仅仅让计算机具备人的视网膜识别物体的边界和探测移动的功能就需要至少 10 亿次/s 的运算能力,但当时最快的超级计算机的运算速度也仅有 8000 万~13000 万次/s。

除了计算能力的限制外,绝大多数人工智能应用,比如图像和声音的识别,需要通过大量的学习才能达到一定的准确率。20 世纪 60 年代数据库本身还处于发展阶段,根本没有积累如此多的数据,这与 40 年后深度学习出现时互联网上已经存在了海量信息所提供的学习环境是无法比拟的。于是从 1966 年美国政府组织的自动语言处理调查委员会给出 ALPAC 报告后^[2],人工智能开始遇冷。1973 年英国政府发布了剑桥大学教授莱特希尔的人工智能调查报告^[3],给第一轮人工智能的发展正式划上了句号。

3 专家与超算

第一代人工智能系统的兴起与衰落让人们人工智能系统有了更丰富的认识。20 世纪 80 年代出现的专家系统是人工智能第 2 阶段发展过程中最有代表性的例子。

西蒙和纽厄尔在设计人工智能系统时注重系统的特定形式和推理机制,但是这类系统往往无法解决纯逻辑推理外的具体生产领域问题,这是因为它们缺少这些领域的具体知识。于是 20 世纪 80 年代的人工智能领域开始把研究的重点放在为不同的系统提供它们所在领域的专业知识上,试图通过专业知识与推理机制相结合来达到专家水平。西蒙的学生,被称为专家系统之父的费根鲍姆将其总结为智能系统的强大来自于它们所掌握的知识而不是具体的规则和理论。

最早获得商业成功的专家系统是

卡梅大学为 DEC 开发的订单配置系统。20 世纪 70 年代诞生于麻省理工林肯实验室的 DEC 公司相当于 20 世纪 90 年代的戴尔,它是最早利用半导体和集成线路技术把大型机体积变小,价格变低,从而让学校和小公司也可以使用计算机的创新公司。因为在它之前的很多计算机公司已倒闭,为了不吓走风投资本,它的商业模式采取了提供计算机配件,让用户可以自由组装的方式,因此受到大学和研究机构用户的青睐。随着订单的不断增加,DEC 开始面临一个问题:销售人员经常搞错用户选购计算机时需要的各种配件(当时都是以人工方式生成),造成了运营成本的增加。于是卡梅大学帮助 DEC 开发了 XCON(订单专家)系统。1980 年,DEC 开始使用该系统后通过与销售人员互动积累了 2500 多条配置规则,运行 6 年共处理了 8 万多笔订单,准确率高达 95%~98%,为 DEC 公司节省了 2500 多万美元的成本。

20 世纪 80 年代,摩尔定律带来的内存容量和 CPU 运算速度的指数增长、关系数据库技术的成熟、个人计算机和局域网技术的普及等因素也是促成专家系统全面开花的重要前提。因为计算机成本的不断下降,企业不同部

门开始各自的局域网内搭建需要的专家系统,在客观上造成了对这些系统的需求。这些利好因素再加上像早期 XCON 这类系统的成功,使得人工智能领域通过专家系统进入第 2 次高速发展阶段。

专家系统的成功再次引发了学术泡沫。不少人认为专家系统可以很快应用到软件开发上,实现开发的自动化。面对这次学术大跃进,也有不少头脑冷静者。1986 年北卡罗来纳大学教堂山分校计算机系主任布鲁克斯发表了一篇后来成为计算机学术历史上经典的论文“没有银弹”^[4]。在这篇论文里布鲁克斯质疑了已经膨胀多年的关于自动编程可以很快实现的学术泡沫。后来他在《人月神话》一书中结合当年开发 IBM 360 操作系统的经历,进一步阐述了大型软件工程的复杂性(图 4)。

布鲁克斯在 20 世纪 60 年代曾经担任 IBM 360 大型机操作系统的开发负责人,对大型软件工程的复杂性有着切身体会。他以一个孕妇需要 10 个月才能生出一个孩子,但是 10 个孕妇加在一起也不可能在一个月生出一个孩子来比喻软件开发的不可避免的复杂性。他认为当时没有任何技术或者管理方式能够让编程效率在 10 年内增长 10 倍,事实也的确如此。布鲁克斯发表这篇论文直接或间接导致了人工智能的第 2 次低潮。因为这时专家系统没能再有大的突破。

专家系统在 20 世纪 80 年代的瓶颈除了计算机软硬件本身的限制外,还有其他因素。比如每个公司和研发团队研发的专家系统都是自成体系的封闭系统,没有开源软件和公开的数据标准来共享数据和策略,所以任何 2 个专家系统之间无法相互协作。另外尽管大家认识到专家系统中知识的重要性,但是知识体系的复杂性远远超过预期。直到 20 世纪 90 年代,学术界和产业界才通过系统地分析人类知识结构,发现其庞大性以及知识本身的显性和隐性之分^[5]。20 世纪 90 年代后期曾经有不少人工智能学者试图把人类的全部知

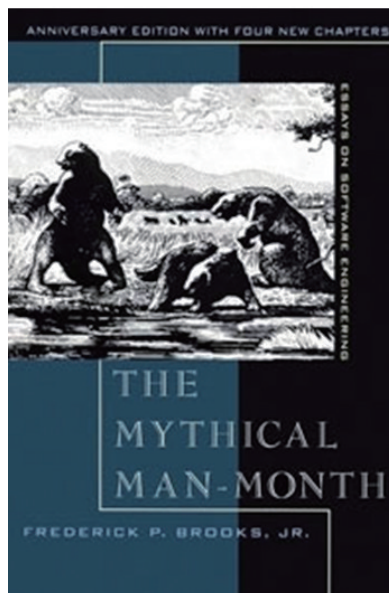


图 4 布鲁克斯的名著《人月神话》,封面用深陷沥青湖无法脱身的 3 只野兽来比喻大型软件开发过程的复杂性

识通过共享的方式构建到一个数据库里,但是就像受到诅咒的巴别塔一样,这一领域最有希望的两位学者先后英年早逝^[6]。

专家系统可以看做是理性流派的进一步发展。神经网络做为感性流派在这一时期虽然没有专家系统那样成功,也取得了一些进步。比如神经网络的架构从以前的输入层加输出层的单层结构转变为添加了中间的隐层。这一变化使得神经网络可以解决更加复杂的分类问题。

层数的增加为各个层的神经节点连接的权重选取带来新的困难。反向传播算法(backpropagation)的出现一定程度上解决了权重选取问题。另外计算机运算能力的提高使得多层神经网络的计算成为可能,在此基础上分支联结主义(connectionism)开始流行。通过分布式计算为多层神经网络提供架构支持成为新的发展方向。

与感性流派的神经网络相关的新人工智能(Nouvelle AI)思想也在这一时期出现。持新人工智能观点的学者认为,人工智能在过去30多年时间里把人类智慧狭义地理解为基于符号处理的智能模型是错误的,因为人类独有的很多智慧是无意识的技能和直觉,并不需要推理。该领域的主要学者莫拉维克(Hans Moravec)举例说:要让计算机如成人般地下棋是相对容易的,但是要让计算机有如一岁小孩般的感知和行动能力却是相当困难甚至是不可能的^[7]。这一观点不仅认为模拟人脑是研究人工智能的基础,还进一步认为人工智能的培育和发展需要通过不断感知外界环境,根据环境做出反应来循序渐进地进行。这一过程中需要的并不是大量的推理,而是对外界信息的大量获取和快速处理。

还需要提及的是专家系统及人工智能神经网络维度的深化推动了超级计算技术的发展。从这一领域衍生出的计算机集群技术成为20世纪90年代信息领域高科技公司的计算平台,而这些平台又成为新一代人工智能技术的硬件基础。当然人工智能的再度发展还需

要一个重要的因素,这一因素到了20世纪90年代才出现,这就是万维网。

4 网络人工智能

互联网的前身阿帕网起源于美苏冷战。20世纪80年代阿帕网已经通过大学和研究所等机构渗透到民间。1989年欧洲量子物理实验中心工作的伯纳斯李向实验中心正式提交了一份后来被称为“万维网蓝图”的报告^[8]。这份报告提出了万维网框架的运行机制和实施方案。1990年11月,他在NeXT工作站上制作了第一个万维网浏览器和第一个网络服务器,随后编写了执行万维网项目细节的网页,至此世界上第1个万维网站诞生。

1993年1月美国伊利诺大学为浏览万维网网站开发的UNIX版本马赛克浏览器被放到该大学计算中心的免费FTP服务器上,不到两个月的时间就被下载了上万次。1993年12月《纽约时报》商业版头版介绍了马赛克,称其将创造一个全新的产业。马赛克的流行使得覆盖互联网的万维网成为新的连接世界的平台,也引发了以硅谷为中心的电子商务革命。1993年1月马赛克刚出现时,全世界只有50个万维网服务器,10月份达到500个,1994年6月份增加到1500个,万维网开始以指数增长。

在万维网流行了3年后的1996年,斯坦福2个研究生发现用解 n 元一次方程组的办法,可以把万维网的所有网页按照重要性进行排名,从而解决了网络用户面对以指数增长的网页信息进行有效搜索的难题,他们后来成立了谷歌公司^[9]。差不多在相同的时间,刚上线

一年多的亚马逊与明尼苏达大学的几位计算机专家合作,开始分析每个用户在其网站上购买的商品,并与其他用户的购买商品进行比较和关联,用得到的结果来个性化对网站浏览用户的商品推荐^[10]。

谷歌的佩奇排名(PageRank)和亚马逊的协同过滤(collaborative filtering)推荐系统的共同特点是它们都通过使用简单的数学原理来处理海量数据达到目的。这种看似简单的指导思想解决了很多专家聚集在一起都很难解决的问题。这就是互联网时代超大规模并行计算所带来的网络人工智能的共同特点。

不同的是,网络人工智能不再使用昂贵的超级计算机,而是用大量联结在一起的廉价服务器甚至是个人计算机来取得相同甚至更好的效果。在知识来源上,网络人工智能往往依靠成千上万的大众点滴贡献(crowdsourcing)而不是专家智慧。比如谷歌在计算网页的排名时使用的是不同网页之间的超文本链接信息,而这些链接是用户在创建网页时提供的。亚马逊的数据来源则是每个用户购买的商品信息。这些在个人看来再简单不过的信息被整合到一起进行处理后发挥出巨大的潜力,也使得数据挖掘成为网络人工智能的流行用语和代名词。

网络人工智能给人类社会带来的不仅是日常生活的方便,它们在很大程度上开始从各个方面影响社会发展进程。谷歌的搜索结果可以决定一个人的言论被关注的程度,而亚马逊的推荐则可以把许多质量好但没能得到推广的商品推荐给大众,引出“长尾定律”^[11]。

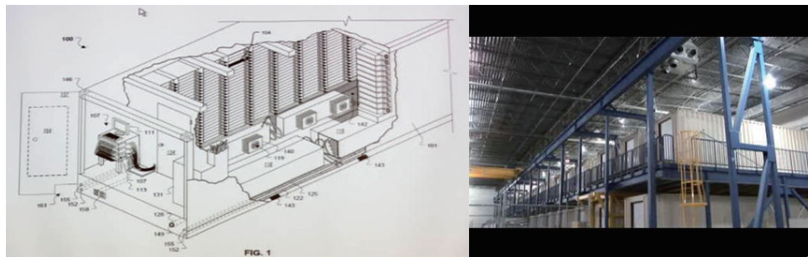


图5 谷歌最早采取了数据集装箱方式来搭建数据中心,几千台服务器在生产商处按照标准方式配置到专门设计的集装箱结构中,再直接运输到数据中心,这些集装箱在设计上采用了很多先进的散热方式,为谷歌节省了大量能耗

随着业务量和数据量的增加,这些网络平台不断扩张,具备越来越强大的存储和计算能力。谷歌从1999年购买二手服务器开始搭建自己的计算平台到2012年第一次对外展示谷歌数据中心,其服务器的数量从最初的几十台发展到100多万台。同一时期亚马逊的弹性云平台服务器总量大概在50万台左右。电商巨擎的计算平台不仅仅在单机数量上达到了以往未有的高度,在架构、软件控制,甚至能耗上与传统的数据中心相比也有诸多创新(谷歌在这一方面有最多的创新,比如谷歌的单机服务器自携备用电池,避免了为整个数据中心配备UPS所带来的额外成本;为了进一步节省空调能源消耗,部分数据中心采取循环水物理蒸发方式降温等(图5)。这些技术升级为2010年前后神经网络在深度学习方向的突破性发展提供了必要的平台。

5 深度学习

20世纪50年代感知机流行时,霍普金斯大学的住院医师休伯尔和年轻教授威塞尔结识并成为学术搭档,开始了后来长达20多年的使他们获得诺奖的视觉神经方面的研究。1959年他们通过观察猫脑部视觉中枢对视网膜进入图像的处理方式发现,有些神经元对不同光强度或者不同角度的斜线有反应,另一些神经元对光感和简单斜线没有反应,但是对物体移动具有敏感性,即便这些物体的边界发生形变也不影响其激发。他们两人将前一类称为简单细胞,后一类称为复杂细胞。休伯尔和威塞尔的这一发现和后来的研究给从事神经网络研究的计算机专家提供了重要的建模思路,比如视神经元对信息处理是分工分层进行的,不同神经元关注的对象特征不同,每一层神经元抽象出对象的部分特征进行处理,最后所有信息经过逐层激发,在中枢最高层激发对整个对象产生认知的“祖母神经元”。

20世纪60、70年代,通过回归分析来逐层学习和提炼最优的神经网络层

数和节点数的方式出现。乌克兰数学家伊万科夫(A. G. Ivakhnenko)及其同事在1971年发表过8层神经网络模型。1979年日本京都大学福岛邦彦(Kunihiko Fukushima)成功通过多层神经网络(neocognitro)学习进行手写字母的图像识别。1992年密歇根州立大学计算机系的翁巨扬提出了最大池化方式(max pooling)和卷积神经网络(convolution neural net)这一后来被广为传播和使用的深度学习算法。

深度学习能够在北美走向主流,与被戏称为“加拿大黑手党”的3个专注机器学习领域多年的计算机专家的坚持不懈有直接关系。他们是加拿大多伦多大学计算机系的杰弗里·辛顿(Geoffrey Hinton),曾经给辛顿做博士后的纽约大学计算机系教授燕存乐(Yann LeCun)和与燕存乐在贝尔实验室共同工作过的加拿大蒙特利尔大学教授约书亚·本吉奥(Yoshua Bengio)。

辛顿1970年从剑桥心理学专业本科毕业,1978年从爱丁堡大学获得人工智能专业博士学位。他毕业时恰好遇到人工智能的第2次高潮,于是投入到神经网络的研究中。他和燕存乐在1989年将反向传播算法应用到前馈多层神经网络学习过程,使得该算法能够识别出手写的邮政编码。不过随着深度加深,反向传播算法无法有效地调整神经连接之间的权重,导致过长的学习时间。20世纪80年代末人工智能研究进入低谷,很多同行开始研究别的领域,但是辛顿等却仍然在这一领域坚持

不懈。以至于有很长一段时间,多伦多大学的计算机系里私下流行的对新生的警言是不要去辛顿的实验室。

在欧洲学术界多层神经网络的研究在20世纪90年代也取得了一些进展。慕尼黑工业大学的人工智能专家于尔根·施米德休伯(Jürgen Schmidhuber)和他的学生团队在1992年提出的非监督学习时间递归神经网络(recurrent neural net)为语音识别和自然语言翻译提供了重要的模型。与前馈网络不同的是,递归神经网络允许各个神经节直接形成环状循环连接,甚至允许一个神经节点接受自己的输入,这种架构使得每个节点能够更好地利用历史权重,从而更适合处理音频和文字等序列数据。1997年他们提出了简化时间递归神经网络的长短期记忆人工时间递归神经网络(LSTM RNN),解决了时间递归网络在学习过程中的误差消失和放大的问题。

无论是北美还是欧洲,神经网络研究在20世纪90年代因为计算能力这一根本性限制均没能在应用领域取得突破。2004年机会出现,这一年辛顿从加拿大高等研究所获得每年50万美元的经费,在燕存乐和本吉奥的协助下启动了神经计算和适应感知项目。该项目将一批一流的计算机、生物学、电子工程、神经科学、物理学和心理学专家聚集在一起,共同探讨用神经网络深度学习的方法模拟人脑智能,该项目聚集了一批后来活跃在人工智能领域的专家,除了辛顿、燕存乐和本吉奥外,还有

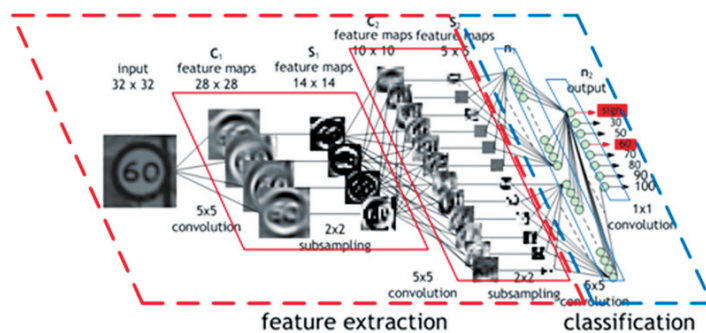


图6 卷积神经网络和最大池化是深度学习的核心设计,复杂的模式通过这种方式被逐层进行特征抽象,最终可以被有效地分析(<https://devblogs.nvidia.com/paralleforall/deep-learning-nutshell-core-concepts/>)

后来负责谷歌大脑计划的斯坦福计算机系教授吴恩达,以及指导奥巴马政府“人脑”计划的诺斯基(Terry Sejnowski)。诺斯基在20世纪80年代和辛顿一起发明了波兹曼机。

有了经费和智慧,辛顿他们很快利用集群计算技术搭配的大计算平台发现了更优化的算法,早期反向传播算法难以克服的困难陆续得到解决。用和时间递归网络类似的逐层培训方式,辛顿等在深度前馈网络中采取了预培训方式解决了反向传播算法的梯度消失问题,用逐层培训和抽象的方式来有效地降低观察对象的维度。简单地说就是将识别对象的特征(feature)从最底层(例如单个像素)开始通过池化方式逐层抽象收集,后一层把前一层的特征抽取结果做为新的识别对象,用同样的方法提取其新的特征。这种方式可以采用非监督的自动编码和解码的方式来矫正,不需要预先对识别对象进行标注(图6)。2006年前后这一算法在图像识别等分类型预测方面取得了突破性。同一时期施米德休伯等人的时间递归算法也在图像识别和语音识别方面取得突破。

2006年深度学习在欧美各地的突破与云计算技术的成熟有直接关系。用辛顿自己在2013年加拿大英属哥伦比亚大学的一次演讲中的话来总结,深度学习以前不成功是因为缺乏3个必要前提:足够多的数据、足够强大的计算能力和设定好初始化权重。事实也的确如此。2006年3月亚马逊正式向外界推出弹性云服务,同年8月谷歌总裁施密特开始使用云计算(cloud computing)一词来区别传统的“客户/服务器”模式。随后微软等其他公司纷纷步入其后尘,开始向用户提供类似的云计算服务。大数据和云计算为深度学习算法提供了海量数据和近乎无限的计算能力,打破了这两个限制人工智能发展的主要瓶颈。值得一提的是GPU的出现加速了深度学习的数据处理速度。与只拥有少量内核的CPU相比,GPU拥有上百个内核,可以同时处理上千个指令相同的线程。这意味着对神

经网络的权重计算这种高度一致的重复并行计算工作,GPU的处理效率有可能达到普通CPU的几十倍,从而可以高速有效地进行各种模式识别计算。

最早将深度学习引入语音识别的是微软。2009年圣诞前夕微软研究院在加拿大召开了一个小型学术会议,邀请辛顿介绍深度学习的研究成果。辛顿的算法并没有引起大多数与会专家的重视,但是微软的两个与会代表邓立和俞栋却在会后组织了人员和数据进行了测试,结果发现非监督的深度学习算法可以使得语音识别的准确率提升25%,远超5%的业界期望值。于是微软开始将这一算法通过不断优化应用到语音识别的各类产品和服务中。图7为2012年10月微软首席研究官拉希德(Rick Rashid)在天津21世纪的计算大会”上公开演示的后来一时热透中国互联网的全自动同声传译系统(错误率仅为7%)的视频(图7)。

微软在语音技术上的突破暗示出如果给予深度学习足够多的数据和计算资源,后者有可能从事更加复杂的图像识别工作。显然互联网上充足的数据资源为这一想法提供了重要的实验平台。2007年斯坦福大学的李飞飞和普林斯顿大学的李凯合作开发了ImageNet项目(图8)。

该项目团队从互联网上下载了10亿多图片,然后通过亚马逊机械土耳其人这一低成本网络众包的方式,雇佣了来自167个国家共5万多人对这些图片进行了分类

标注。截止2009年该项目成功产生了一个包含22000不同门类,共1500万图片的带标签的数据库。该数据库里的图片分类和标签质量超过以往任何数据库,其中仅猫这一门类就有62000张不同的图片,包含了所有种类的驯养和野生猫。建成这一数据库后,李飞飞及其团队利用深度学习方法,使得计算机通过监督学习方式识别包含各种物体的图像,而且能够用自然语言生成对每个图像中的物体关系的简单描述。这一成果进一步引发了学术界和科技界对深度学习的关注。

ImageNet项目主要为监督学习方式提供了数据标签支持,但是标签图片与整个互联网的所有图片数量相比,仍然是九牛一毛。绝大多数数据仍然是以无标签形式存在的。更重要的是人脑尤其是婴儿大脑的发育过程是以标签方式进行的。所以无监督学习和让



图7 2012年10月微软首席研究官拉希德(Rick Rashid)在天津召开的“21世纪的计算大会”上公开演示全自动同声传译系统(错误率仅为7%)(<https://www.youtube.com/watch?v=Nu-nlQqFCKg>)

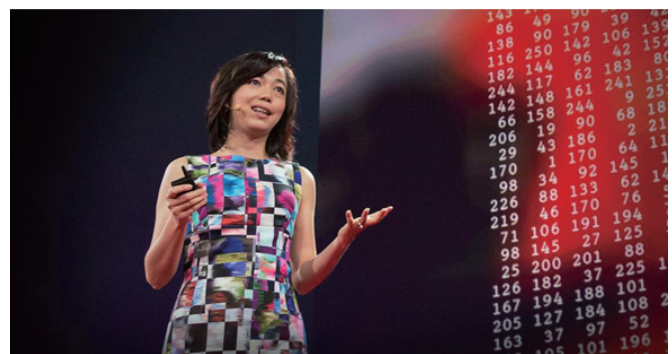


图8 李飞飞在TED上讲述ImageNet(How we're teaching computers to understand pictures),ImageNet目前以开源形式为各种深度学习算法的测试和比赛提供数据支持

神经网络自我学习成为另一个研究方向。

2011年谷歌的迪恩(Jeff Dean)和斯坦福计算机系的吴恩达在未来技术实验中心(Google X)联合发起了“谷歌大脑”研究计划。该计划通过模拟新生儿的大脑发育环境来研究人脑的物体识别和语言认知等功能。他们利用谷歌的云计算平台搭建了一个配备16000个CPU和10亿突触神经连接的谷歌大脑计算平台(1年后斯坦福大学和英伟达合作只用了16台配备GPU的服务器,每台配备2个8核CPU和4个英伟达GeForce GTX 680GPU,就达到了同样的计算效果)。为了收集有效的能够表征人类生活环境的数据,他们从谷歌YouTube随机选取了1000万个视屏,从每个视频里随即获取一个200×200像素的截屏,相当于模拟婴儿用眼睛不断观察到的周围环境。

数据收集完成后,他们用辛顿2006年提出的深度学习分层训练模型和自我编码解码校验方式去自动对这1000万张图片进行特征抽取和分析。这一项目的目标之一是查看该模型的分层抽象特征提取方式是否能够最终产生一批高度异化的“祖母神经元”。结果该实验不但发现了模拟状态的祖母神经元的存在,而且在抽象最高层形成物体判断的神经元中居然有一只是对应猫的面部图像(图9),也就是说通过深度学习,该人工大脑形成了对猫的



图9 这是谷歌大脑用深度学习方式分析了成千上万张猫的图片后形成的对猫的“记忆”

印象(古希腊哲学家柏拉图和其学生亚里士多德在人对客观世界的不同对象的概念形成中到底是先有绝对主观存在还是从客观形成有意见分歧。从谷歌的大脑实验来看,亚里士多德的客观形成论得到了验证)。

除了辛顿等在深度学习领域成果频出外,2010年以后在瑞士人工智能实验室(IDSIA)担任主任的施米德休伯和他的学生们也不断取得突破。他们利用长短记忆时间递归神经网络识别序列信息的优势在各种图像识别比赛中夺魁,其中包括德国的交通标志自动识别比赛、国际汉字、阿拉伯文等手写体识别比赛。2010年他的3个博士生创立了深思(DeepMind)公司,根据深度学习理论和对数据流的研究开发出了能够像人一样通过观察普通人玩街机游戏的画面来了解游戏规则,再通过学习的规则去玩同样的游戏,甚至击败人类对手。2014年该公司被谷歌以5亿美元收购。2016年3月,该团队设计的AlphaGo围棋对抗程序以4胜1负的成绩击败了人类围棋冠军李世石。

在目前这轮以深度学习为代表的人工智能新高潮里,企业界早已通过收购和人才猎取而投入了大量的资本。谷歌董事长施密特公开表示机器学习是谷歌目前最主要的工作。2013年谷歌通过收购辛顿的DNNresearch公司,将包括辛顿和他的几个学生在内的一批人才收揽在旗下。随后谷歌又在2015年收购了DeepMind公司,把欧洲的深度学习专家收拢到谷歌旗下。Facebook创始人扎克伯格在读完深度学习的论文后,从纽约大学高薪聘走了燕存乐。亚马逊资助西雅图的华盛顿大学计算机系从卡梅挖来了盖斯成(Carlos Guestrin)夫妇,后者与亚马逊合作开发机器学习项目。而在这之前,谷歌从该系挖走了7名计算机系教授。斯坦福大学的吴恩达则被百度从谷歌挖走,成为硅谷百度研究院的首席研究员。

这些投入会不会最终仍然以泡沫破裂收场?答案很有可能是肯定的。

6 智能与超越

2014年牛津大学的博斯特伦(Nick Bostrom)在《超级人工智能:路径,危险和策略》一书里提出,人工智能技术很可能在不久的将来孕育出在认知方面全面超越人类的超级智能(super intelligence)。他认为超级智能在给人类社会带来好处的同时也造成了人类本身的生存危机(existential risk)。2015年1月包括物理学家霍金、企业家马斯克等在内的主要来自欧洲的科学和科技界精英联名发表了一封公开信,敦促业界关注人工智能发展的稳健性和风险控制,通过合作和共识形成一个控制人工智能发展方向的框架,该公开信内容和所有签署人名单见http://futureoflife.org/AI/open_letter。

博斯特伦提出的超级智能可以追溯到冯诺依曼在1958年的奇点(Singularity)预测。冯诺依曼认为随着计算机技术的不断发展,终有一天计算机将能够自我设计,从而进入自我进化状态。一旦进入到这一奇点状态,人类智能将与计算机智能迅速拉开差距,前者将逐渐被后者取代。

尽管这些学术界和产业界的名人表示了忧虑,但人类对大脑的研究,尤其是什么是智能的认识却仍然处在初级阶段。从目前人工智能的发展水平来看,笔者认为这种对机器本身所带来的威胁的担忧是没有任何必要的,真正需要防范的是强势集团对人工智能技术的滥用。

首先,从计算机的计算能力来进行一下比较。根据目前对人脑的最新认识,人的大脑皮层共有大约860亿神经元,能够进行200 MB/s并行运算。新生儿大脑皮层每个神经元有大约2500个突触,2、3岁时增加到15000个,达到峰值。成年后每个神经元的突触数量大约为7500个。相比之下,2012年设计的谷歌大脑的总共16000个CPU神经元和10亿突触,远远低于人脑的复杂度。事实上就算把谷歌计算平台所拥有的全部服务器(总数量为几百万台)用于大脑认知模拟,也难以达

到一个普通人脑的神经元的数量和关联度。

其次,对人脑智能的机制理解仍然处于一个初级阶段。比如德莱弗斯(Hubert Dreyfus)提出人脑主要是通过图像来思考的,这一现象背后的机制是什么?当把一个小箱子放到一个大箱子上,然后把下面的大箱子取走,这时小箱子会怎样?普通人通过大脑的图像预演可以很简单地回答这一问题,而理性流派的人工智能往往需要通过大量的事实积累和逻辑推理才能做出这一推论。以感性流派为理念的深度学习神经网络虽然有可能通过模式识别实现这一推理,但目前来看还没有在这方面迈进一步。

奔迈公司的创始人霍金斯的HTM理论就这一现象给出了比较有意思的理论框架^[12]。该理论认为大脑皮层通过不断接收序列信息(sequence information)来学习建立外部世界的感

知-动作模型(sensory-motor model)的器官,通过接收—预测—反馈—调整再接收这一过程来调整记忆。对于大量涌入的序列数据(比如声音和动作),大脑皮层是通过稀疏分布的表达方式层层激发处理信息。信息激发到祖母神经元后根据以前记忆形成的预期开始向下激发相应的各种感知和动作神经元,这些预期与新的信息进行对比后,针对出现的偏差,相应的神经元会做出新的突触连接调整。遗憾的是在实际应用方面,根据该理论开发出的软件还未能取得突破性进展。

最后,涉及一个比较抽象的哲学问题,就是什么是人的智能。60年前图灵用图灵测试来避免回答这一问题。今天对智能的认识可能也无法给出满意的答案。目前流行的深度学习只是人认知能力的增强版,就像汽车和飞机虽然比人速度快,但不能认为汽车或者飞机具有意识或者它们的运动智能超

越了人类。

美国计算机和未来学家库兹韦尔(Ray Kurzweil)在2005年的《奇点已近》一书认为,人的智能其实就是人脑的神经连接状态^[13],他提出未来技术成熟后,可以通过释放足够多的纳米机器人到人体内,获得大脑神经元之间的每一个链接的各种生化指数,从而复制人的意识。这些纳米机器人可以不断监控脑神经的变化,从而在身体外部不断备份人的意识。像很多前人科学家一样,笔者认为库兹韦尔陷入了用人类可以理解的技术来解释人类智能和意识的陷阱。

最近10年很多科学研究发现自然界的各种生物现象与量子纠缠有密切关系,比如光合作用和候鸟迁徙^[14,15]。人的智能和意识是否也与此有关,只有等对这一现象有了更深刻的认识后才能知道。

参考文献(References)

- [1] Turing A M. Computing machinery and intelligence[J]. Mind, 1950, 59(236): 433-460.
- [2] Pierce J. Computers in translation and linguistics[R]. National Academy of Sciences/National Research Council, 1966.
- [3] Lighthill I. Artificial intelligence: A general survey[R]//Artificial Intelligence: A Paper Symposium. London: Science Research Council, 1973.
- [4] Brooks F P. No silver bullet essence and accidents of software engineering[J]. Computer, 1987, 20(4): 10-19.
- [5] Nonaka I, Takeuchi H. The knowledge-creation company: How Japanese companies create the dynamics of innovation[M]. New York: Oxford University Press, 1995.
- [6] Kushner D. Two AI Pioneers. Two Bizarre Suicides. What Really Happened?[J]. Wired Magazine, 2008-01-18.
- [7] Moravec H. Mind children: The future of robot and human intelligence[M]. Cambridge, MA, USA: Harvard University Press, 1988.
- [8] Berners-Lee T J. Information management: A proposal[J/OL]. [2016-03-15]. <http://faculty.georgetown.edu/irvinem/theory/Berners-Lee-HTTP-proposal.pdf>.
- [9] Page L. The PageRank citation ranking: bringing order to the web[J]. Stanford Infolab, 1998, 9(1): 1-14.
- [10] O'Connor M, Herlocker J. Clustering items for collaborative filtering[C/OL]. [2016-03-15]. http://www.csee.umbc.edu/~ian/sigir99-rec/papers/oconner_m.pdf in Proceedings of the ACM SIGIR workshop on recommender systems. 1999. UC Berkeley.
- [11] Anderson C. The long tail: Why the future of business is selling less of more[J]. Journal of Product Innovation Management, 2006, 24(3): 274-276.
- [12] Hawkins J, Blakeslee S. On intelligence[M]. Times Books, 2004.
- [13] Kurzweil R. The singularity is near: When humans transcend biology[J]. Cryonics, 2006, 85(1): 160-160.
- [14] Lee H, Cheng Y C, Fleming G R. Coherence dynamics in photosynthesis: Protein protection of excitonic coherence[J]. Science, 2007, 316(5830): 1462-1465.
- [15] Ritz T, Adem S, Schulten K. A model for photoreceptor-based magnetoreception in birds[J]. Biophysical Journal, 2000, 78(2): 707-718.

(责任编辑 刘志远)