

人工智能“大爆炸”时代刚刚开始



Tom Austin

高速并行的图形处理器、深度神经网络算法和来自于互联网的大数据，这三者撞击出了人工智能“大爆炸”。这次“大爆炸”将开启一个长达75年的人工智能时代。

“大爆炸”之前

1956年，约翰·麦卡锡(John McCarthy)将“人工智能”定义为“制造智能机器的科学工程”，尤其是“人类的智慧能精准地被一台机器模拟”这样的描述，反映出了20世纪中叶学者对“人工智能”的乐观预测，以及当时计算机科学界的宏大愿景。然而，这一愿望直到今天仍是遥不可及。正如《人工智能的冬天》一文所言，那些过于乐观的假设至少在20世纪就搁浅了9次。

Gartner在《2015年新兴技术成熟度周期报告》(见下图)中指出，“完善的人工智能”在目前，甚至未来10年内都难以实现，对于人工智能的理解才刚刚开始。设想一下，我们曾以为飞机的翅膀一定要上下震动起来才能飞，汽车必须要靠“腿”一样的零件才能迈开步子，而非滚动的“轮子”。

1997年，IBM超级计算机“深蓝”(Deep Blue)击败了国际象棋世界冠军卡斯帕罗夫(Gary Kasparov)。《新闻周刊》一定为成功预言了“机器将比发明它的人聪明”而感到高兴。但我们并不能认为“深蓝”的胜利就证明了机器比人类更聪明，我们甚至不认为“深蓝”能划归为人工智能，尽管它很强大，但“深蓝”的“智慧”都来自于其系统设计师和程序员所创造的算法——“深蓝”的行为亦步亦趋，它在代码加载不到的地方走不出任何一步棋。

既聪明又强大的DNNs

相比“深蓝”前辈，我们认为今天的DNNs(Deep neural networks, 深度神经网络)更聪明。DNNs擅长于搜索各种类型的数据，包括图像、声音、视频和文本。经过大量、反复地训练，DNNs学会了识别各类数据特征，无需程序员提前为其植入特定功能的探测装置，DNNs就能自如地运用自己强大的分析能力自力更生，甚至能概括出不同数据来源中的共性。

比如，2012年，Google的科学家们用1.6万块电脑处理器构建了全球最大的电子模拟神经网络，并通过向其展示自YouTube上随机选取的1000万段视频，考察其能够学到什么。结果显示，在无外界指令的自发条件下，该神经网络自主学会了识别猫的面孔。

在合适的条件下，DNNs表现出以下优势：

- 主动挖掘输入流(如单词、短语、句子等)中不同元素的分类规律，而非基于程序员编写的特定程序进行识别。
- 开发出一套独特的输入流分类规则体系。
- 能够容忍噪声、图像元素位置、声音频率等的变化。
- 多层次操作。

● 从开发人员未预测到的输入流中识别新模式。

“大爆炸”时代的到来

人工智能“大爆炸”的实现是三大要素汇合的结晶，即新型计算处理硬件的发展(规模运算系统与GPU超级计算机，GPU就是高度并行的图像处理单元)、DNNs算法和作用于DNNs上的大数据。这三个因素在过去几年中相互作用。截至2012年，这三者的结合达到了“群聚效应”。

“大爆炸”使得训练大规模DNNs模型成为可能，并能在短时间，甚至数小时之内见到结果。这使得在有限时间内进一步提升模型、训练数据、测试数据成为可能。反过来，其又能促进现有技术的发展，在缩减开发时间的同时，其开发速度和规模也在不断提升。

硬件：新型GPU芯片可以提供必需的算力。在固定的时间内，目前的芯片可以模拟2008GPU1000倍的DNNs神经元，2008GPU可以模拟出2007年X86处理器10倍的神元。

算法：DNNs算法在20世纪80年代得到进一步发展，学习(重新调整了“看待”数据的方式)取代了程式化的分类和推广。在训练中，这些模型是数据和集群计算。

数据：互联网上存在大量数据，提供了数以百万计的视频、语音录音、文本信息，以及其他训练材料。

DNNs训练、测试和运行

DNNs训练需要一个DNNs模型，需要将大量数据置于模型中，并消耗其计算能力。在此模型中，研究人员定义DNNs的结构和特点。他们设置DNNs节点(或神经元)的数

量、层数、层与层之间的关系，以及其他各种参数(如复发或回旋)。

然后，他们将数百万(甚至上亿)的图像、几万或几十万的数字音频录音或数量庞大的文字内容，以及这些输入内容的变体输入机器中运行数小时或数天(通常会加入噪声和其他信号来提高结果的稳定性)。

如今，DNNs训练所选用的硬件一般是大规模并行超级计算机。例如，百度在2014年研发出超级计算机“Minwa”，它含有72个互联Nvidia Titan GPUs，每个GPU拥有3072个图形处理核心处理器(总计221184个内核)，拥有高速共享内存864千兆字节。

“大爆炸”的应用成果

自20世纪80年代末以来，多伦多大学的杰夫·辛顿的团队、纽约大学的LeCun团队，以及其他研究者都在为提高DNNs算法能力而努力。由于GPU和大数据的出现，基于DNNs算法展示了人工智能在实时展现图像、视频、脸部、视觉场景、口语、语言翻译和书面文字方面惊人的能力。

以下是人工智能研究的一些研究成果展示，足以说明当今人工智能发展的最高水平：

● DNNs在图像识别测试(ImageNet)上的错误率低于人类在同等测试中的错误率。

● Google的面部识别能力已经超过人类区分两张脸的能力。

● 语音识别，特别是在嘈杂的环境中的语音识别，百度在2014年底发表了“Deep Speech”的带动下，比起以往有了显著改善，但仍然逊色于人类的表达。

● 自然语言处理——用DNNs来帮助构建包含大量既定事实的知识图也有了长足进步。

● 无监督的DNNs技术(如Google知识库服务)。

人工智能也要补短板

如今的人工智能存在着一些显著缺陷，大部分人工智能研究的进展停留在感知端。例如，人工智能在区分输入数据流(来自传感器、网络爬虫和其他来源)时能力惊人，在常规情境下的多个层面的运行表现良好。但是，在上文中，大部分人工智能的优异成绩都是在特定的测试条件下完成的，而人工智能的实时性能是受真实世界影响的。

我们乐观地预计，DNNs硬件的改进增强将进一步提高技术应用的广度。据估计，到2020年，DNNs的计算能力能达到2015年的100倍或1000倍。

然而，人工智能上一些被业界寄予厚望的高级功能还不成熟，比如自动驾驶。我们在智能视觉系统方面取得的巨大进步使得汽车控制器单元能够获得可视范围内所有信息的反馈。智能视觉系统从传感器(比如摄像头、雷达等)输入信息，在DNNs“看清楚”之后，按照概率、速度、物体类别和组件信息等标签分类，并将“加工”后的指令传递给某一对应的控制器进行确认并执行。

当车辆已经接收到了视觉信息信号，是汽车的控制单元而不是DNNs决定了汽车如何做出反应。这种控制器，由开发人员事先以C++语言进行编程，告诉汽车在特定情况下应该做A而不是做B。当然，现在的控制器已经可以根据车主的驾驶习惯、车速等参数来进行适配，但是请注意，这不是真正的人工智能。因为控制器不能通过反复试错来获得判断力的提升(天哪，也许这是好事)。

2015年新兴技术成熟度周期报告

