

# 为什么说数据是人工智能的粮食

机器智能来源有两类：一类是人类专家把自己掌握的知识教给机器，机器用这些知识进行推理，实现类似人的智能；另一类是把数据喂给机器，让他像人一样从中归纳总结出知识，再利用这些知识进行推理。早期的人工智能系统多采用第一种方案，现代人工智能系统多采用第二种方案。特别是深度学习发展起来后，数据的重要性越来越明显，人们将数据称为人工智能的粮食。

从数据中自动学习知识，这门学问称为“机器学习”。机器学习早在上世纪 60 年代就已经萌芽，但没有成为主流。后来人们渐渐发现，基于知识的人工智能系统受限严重，很难超过人类知识的上限，只有从数据中自我学习，才能实现更强大的智能。为了实现这一学习能力，研究者提出了包括概率模型、支持向量机、决策树、人工神经网络等各种学习结构，并取得了一系列成果。90 年代以后，机器学习已经成为人工智能的主流方法。然而，受数据规模所限，当时的智能系统很多还难以实用。

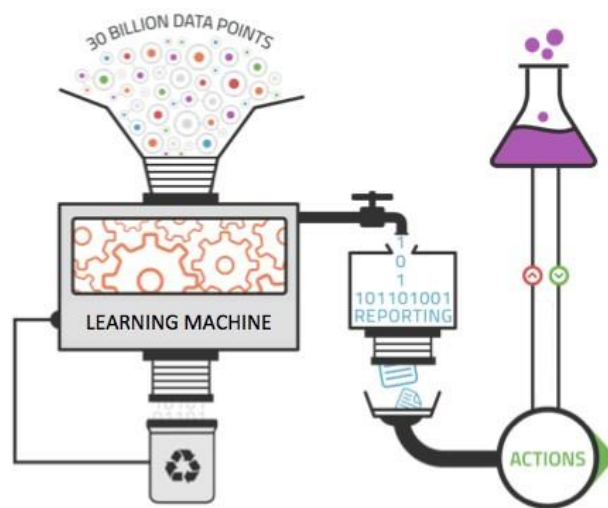


图 1：机器学习从大量数据中提取知识[1]

新世纪以后，数据积累快速增加，特别是互联网的发展让我们进入大数据时代。以语音信号处理为例，93 年由 MIT 等三家机构发布的 TIMIT 数据集只有 4.6 小时，而今天免费的公开数据集 Librispeech 已经达到 1000 小时，商业公司所拥有的数据可能达到数十万小时。数据量的大幅增长使得我们可以学习更多更可靠的知识，如果这些知识能覆盖实际应用中的绝大部分场景，那么我们的人工智能系统将变得非常强大。

数据量的增长带来另一个变化是学习方法的简化。数据量不大时，我们需要人为设计各种特征提取方法，并在模型结构上引入各种限制以适应目标任务的需要。在大数据时代，我们可以选用更灵活的模型，直接学习从原始输入到目标任务的决策函数(如分类目标、预测值等)，

即当前研究较多的基于深度神经网络的端到端系统。

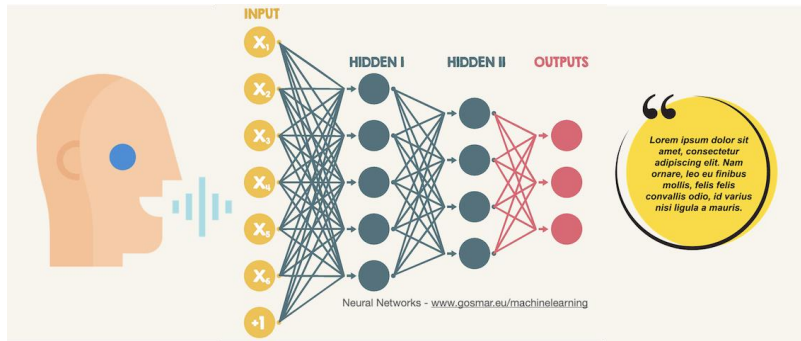


图 2：端到端语音识别系统[2]

大数据学习为机器智能突破人类智能创造了可能。人类的知识也是从数据中来的，只不过经过了人的理性加工，变得更有条理。机器无法做到像人那样理性地学习，但优势是学习的速度快，可以短时间处理海量数据，因而可以发现人类还没有发现的新知识，从而获得更强的智能。这一方面为设计更强大的智能机器提供了可能，另一方面也引发了人工智能潜在威胁的讨论。

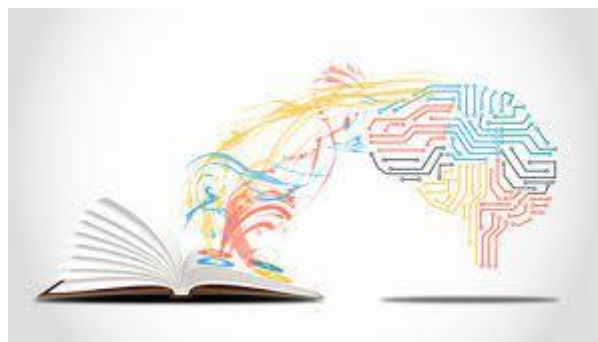


图 3：海量数据学习可能发现新的知识[3]

尽管大数据学习取得了令人瞩目的成就，这一方法也引起了一些学者的反思。一个重要的问题是数据驱动带来的知识偏差。当前的统计学习方法主要学习高频事件，这些高频事件确实是最主要的知识来源，但当这些知识学习饱和之后，更深刻的知识往往体现在那些与常识相违背的小概率事件中。人类的科学家们往往从这些小事件中发现世界的深层规律，而目前的机器学习方法很难从小数据中发现新知识。这一缺陷的背后是统计机器学习的天然缺陷，这一缺陷使得机器无法对知识像人那样进行整理，形成自洽的体系并进行自我检验。换句话说，目前机器已经可以学习，但还做不到像人一样学习。

可以想见，数据在可预见的未来依然是人工智能最重要的食粮，但如何对这些粮食进行更有效的消化利用，还需要深入探索。这就像一个小孩子一样，书本显然是最重要的，但要形成理性，需要的不只是学习，还需要对知识进行梳理、理解、实践。拥有这样能力的智能机器

或许才是人工智能的缔造者们所预想的样子。

[1] Let' s talk about Advanced Analytics: A brief look at Artificial Intelligence,  
[https://becominghuman.ai/lets-talk-about-advanced-analytics-a-brief-look-at-artificial-intelligenc  
e-bf1c7a7d3f96](https://becominghuman.ai/lets-talk-about-advanced-analytics-a-brief-look-at-artificial-intelligenc-e-bf1c7a7d3f96)

[2] Neural networks and speech recognition,  
<https://www.gosmar.eu/machinelearning/2020/05/25/neural-networks-and-speech-recognition/>

[3] Discover The Power of Big Data And Learning Analytics For Education,  
[https://www.smartdatacollective.com/discover-power-of-big-data-learning-analytics-for-educati  
on/](https://www.smartdatacollective.com/discover-power-of-big-data-learning-analytics-for-educati<br/>on/)