

机器学习算法概述及其应用场景

主讲人: 张聪聪

华润智慧能源有限公司

April 10, 2020

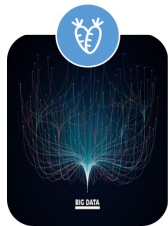
crp-logo.pdf

目录 I

crp-logo.pdf

crp-logo.pdf

大物移云智



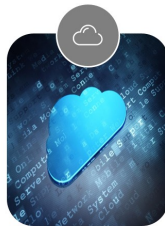
大数据



物联网



移动互联



云计算



人工智能

crp-logo.pdf

大数据定义

广义定义

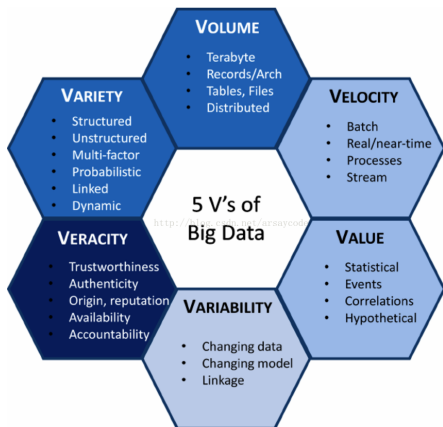
- 大数据是收集，组织，处理和收集大型数据集洞察所需的非传统策略和技术的总称。
- 大数据就是多，就是多。原来的设备存不下、算不动。
- 大数据，不是随机样本，而是所有数据；不是精确性，而是混杂性；不是因果关系，而是相关关系。
- 大数据，是指物理世界到数字世界的映射和提炼。通过发现其中的数据特征，从而做出提升效率的决策行为。

狭义定义

大数据，是通过获取、存储、分析，从大容量数据中挖掘价值的一种全新的技术架构。

crp-logo.pdf

大数据特点



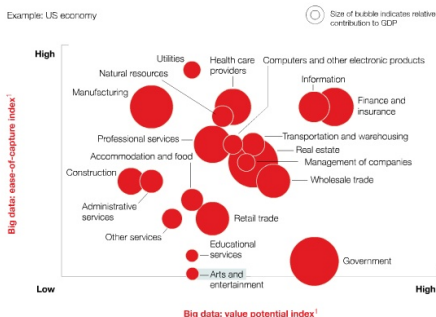
5V

- **Volume**: 数据量大, 包括采集、存储和计算的量都非常大。
- **Variety**: 种类和来源多样化。包括结构化、半结构化和非结构化数据, 具体表现为网络日志、音频、视频、图片、地理位置信息等等, 多类型的数据对数据的处理能力提出了更高的要求。
- **Value**: 数据价值密度相对较低, 或者说是浪里淘沙却又弥足珍贵。
- **Velocity**: 数据增长速度快, 处理速度也快, 时效性要求高。
- **Veracity**: 数据的准确性和可信赖度, 即数据的质量。

大数据的价值与应用场景

The ease of capturing big data's value, and the magnitude of its potential, vary across sectors.

Example: US economy



¹ For detailed explication of metrics, see appendix in McKinsey Global Institute full report *Big data: The next frontier for innovation, competition, and productivity*, available free of charge online at mckinsey.com/mgi.

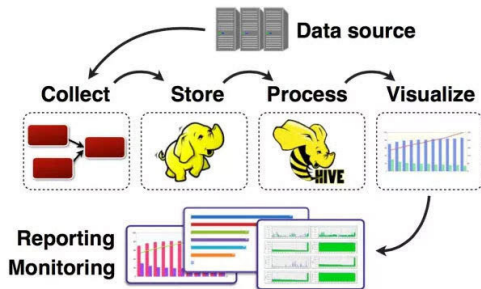
Source: US Bureau of Labor Statistics; McKinsey Global Institute analysis

四类价值方法

- 客户群体细分，然后为每个群体量定制特别的服务。
- 模拟现实环境，发掘新的需求同时提高投资的回报率。
- 加强部门联系，提高整条管理链条和产业链条的效率。
- 降低服务成本，发现隐藏线索进行产品和服务的创新。

crp-logo.pdf

数据怎么来



大数据的六个环节

- 数据提取
- 数据存储
- 数据清理
- 数据挖掘
- 数据分析
- 数据可视化



crp-logo.pdf

人工智能是什么



人工智能定义（Artificial Intelligence）

它是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学

crp-logo.pdf

人工智能的三个级别 I



弱人工智能

也称限制领域人工智能（Narrow AI）或应用型人工智能（Applied AI），指的是专注于且只能解决特定领域问题的人工智能。

crp-logo.pdf

人工智能的三个级别 II



强人工智能

又称通用人工智能或完全人工智能，指的是可以胜任人类所有工作的人工智能。

crp-logo.pdf

人工智能的三个级别 III

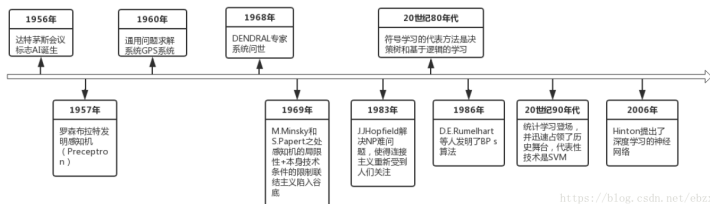


超人工智能

假设计算机程序通过不断发展，可以比世界上最聪明、最有天赋的人类还聪明，那么由此产生的人工智能系统就可以被称为超人工智能

crp-logo.pdf

人工智能发展史 I

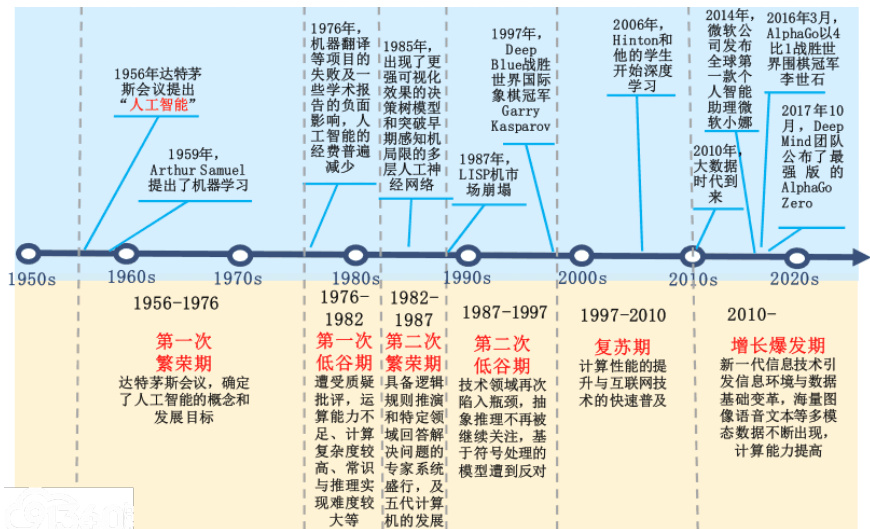


<https://blog.csdn.net/ebzxw>

六个时期

- 起步发展期：1956年—20世纪60年代初。
- 反思发展期：20世纪60年代—70年代初
- 应用发展期：20世纪70年代初—80年代中
- 低迷发展期：20世纪80年代中—90年代中
- 稳步发展期：20世纪90年代中—2010年
- 蓬勃发展期：2011年至今。

人工智能发展史 II



crp-logo.pdf

深度学习三巨头 I



(a)



(b)



(c)

Figure: a)Geoffrey Hinton; b)Yoshua Bengio; c)Yann LeCun

crp-logo.pdf

Geoffrey Hinton

- 1 **反向传播**: 证明了反向传播算法允许神经网络发现自己的数据内部表示, 这使得使用神经网络成为可能网络解决以前被认为超出其范围的问题。
- 2 **玻尔兹曼机**: 这是第一个能够学习不属于输入或输出的神经元内部表示的神经网络之一。
- 3 **卷积神经网络的改进**: 2012 年, Hinton 和他的学生 Alex Krizhevsky 以及 Ilya Sutskever 通过 Rectified Linear Neurons 和 Dropout Regularization 改进了卷积神经网络, 并在著名的 ImageNet 评测中将对象识别的错误率减半, 在计算机视觉领域掀起一场革命。

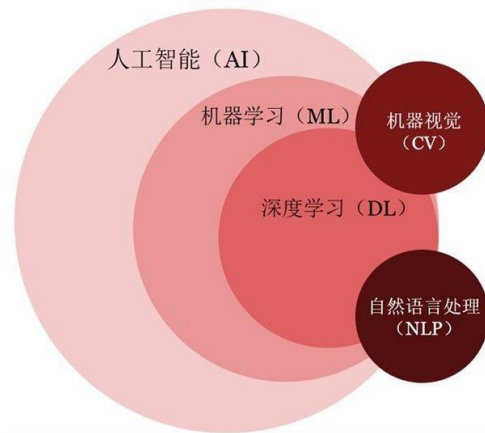
Yoshua Bengio

- 1 **序列概率模型**: 现代深度学习语音识别系统正在扩展这些概念。
- 2 **高维词嵌入和注意力模型**: 导致了机器翻译领域的突破, 并构成了深度学习序列建模的关键组成部分;
- 3 **生成对抗网络**: 生成对抗网络 (GANs), 在计算机视觉和计算机图形学领域引发了一场革命。

Yann LeCun

- 1 **卷积神经网络**: 提出卷积神经网络 (CNN), 卷积神经网络已经成为计算机视觉、语音识别、语音合成、图像合成和自然语言处理领域的行业标准.
- 2 **改进反向传播算法**: 加速了反向传播算法;
- 3 **拓宽神经网络**: 他将神经网络发展为一种计算模型, 用到一系列任务中, 他早期工作中的一些概念已成为 AI 发展的基石。

AI、机器学习、深度学习之间的关系



crp-logo.pdf

crp-logo.pdf

什么是算法



算法

算法是任何良定义的**计算过程**，该过程取某个值或值的集合作为**输入**并产生某个值或值的集合作为**输出**。

crp-logo.pdf

算法伪代码

crp-logo.pdf

算法伪代码

crp-logo.pdf

算法伪代码

crp-logo.pdf

算法伪代码

crp-logo.pdf

算法伪代码

crp-logo.pdf

crp-logo.pdf