人工智能

机器能思考吗?

毛昕渝 2021/11/22

人工智能（AI）是什么？

认知科学

人的思维过程是怎样的?

思维过程

逻辑主义

所有人都会死
苏格拉底是人

→苏格拉底会死

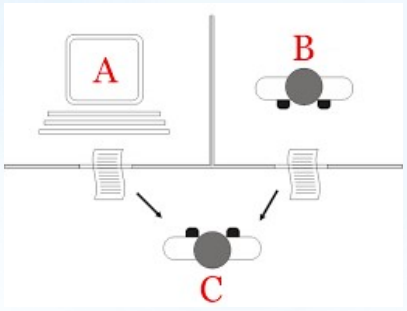
三段论

有像人一样的思维过程

能理性地思考

像人一样（现实状况）

理性（理想状况）



图灵测试

能像人一样行动

理性智能体（rational agent）

行为表现

理性人（Homo economicus）

经济学的一个经典工作假设：人是理性的、自利的。

发展历程

图灵的先见之明

- 在1950年的论文 *Computing Machinery And Intelligence* 中，图灵提到了：
  - 机器能思考吗?
    - 为了为这个问题提供一个判断准则，他提出了一种模仿游戏，今天被称为图灵测试。
  - 这种“能思考”的机器就是通用电子计算机
  - 回应了对机器能思考的各种反驳
  - 会学习的机器
    - *Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's?*
    - *We have thus divided our problem into two parts. The child-programme and the education process. These two remain very closely connected.*
    - *The use of punishments and rewards can at best be a part of the teaching process.*

诞生

- Marvin Minsky
- 1950年建造了第一台神经网络计算机SNARC
- 1969年图灵奖得主

一般被认为是人工智能领域的开端

- Herbert A. Simon
- 在1956年的达特茅斯人工智能会议上展示了推理程序“逻辑理论家（Logic Theorist）”
- 1975年图灵奖得主

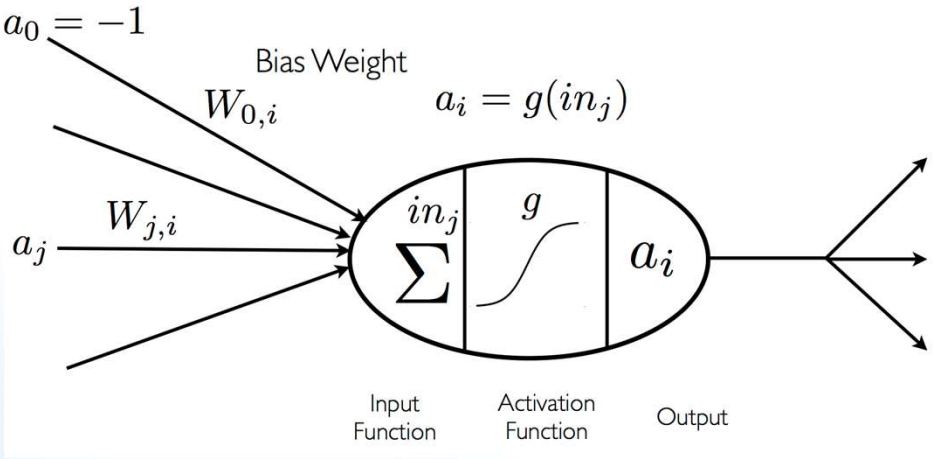
证明了《数学原理》第二章中的大部分定理

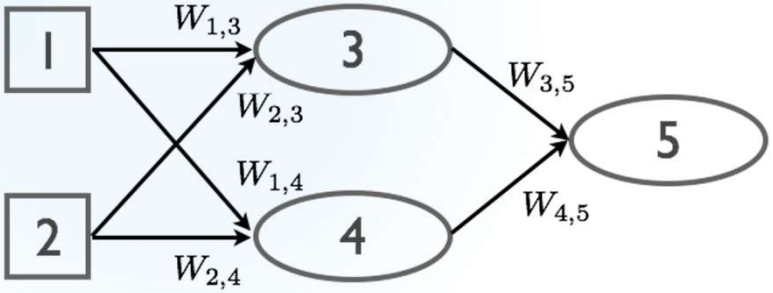
知识系统

- Edward Feigenbaum等人：专家系统

2011年，IBM-Watson 在 Jeopardy!问答比赛中夺冠。

连接主义：人工神经网络





- 第i个神经元的输入：$in_i = \sum_j W_{ji} a_j$.
- 第i个神经元的输出：$a_i = g(in_i) = g\left(\sum_j W_{ji} a_j\right)$.
- 激活函数 g 一般的是非线性的，这样的多层网络可以逼近很复杂的函数。
- “训练”神经网络：反向传播算法

Pictures from *Artificial Intelligence: A Modern Approach* by Stuart J. Russell and Peter Norvig

学习与进化

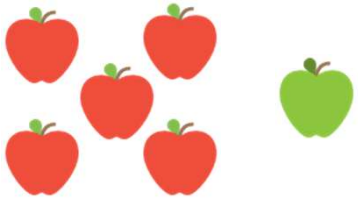
“君子曰：学不可以已。”——《荀子·劝学》

“学而时习之，不亦乐乎？”——《论语》

归纳的问题：学习何以可能?

休谟的问题

归纳推断为什么有道理?



David Hume
1711--1776

不变量假设

运用所学的环境和学习的环境不会有巨大的、根本性的差别。有一些规律是不变的。

规律可学习假设（来自计算机科学）

承上，有一些规律是不变的。并且这些规律：

1. 可以被某种计算过程（算法）检验；
2. 这种检验算法能够通过现实的资源和与环境的有限的交互来得到。

认同这些假设有利于实践，实践是一切怀疑论的解毒剂。

机器学习任务的分类

监督学习

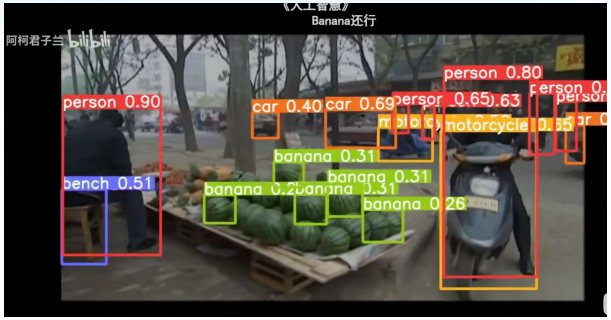
给定数据集

$$T = \{(x_i, f(x_i))\}_{i \in [|T|]},$$

试图学习函数f.

例子：

- 手写数字识别
- 图像标注



无监督学习

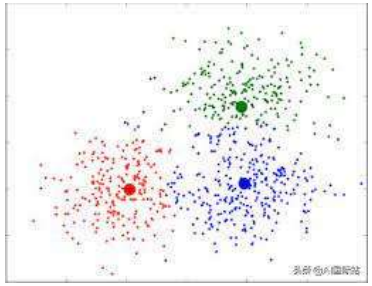
给定数据集

$$T = \{x_1, x_2, \dots, x_n\},$$

学习其中的隐含的模式和信息。

例子：

- 数据挖掘
- 聚类
- 网页排序（Pagerank）



强化学习

让机器在环境中行动，有时会得到反馈（奖励或惩罚）；我们希望机器根据反馈理性地行动（回报最大化）。

例子：

- AlphaGo
- 王者荣耀

理论的视角: PAC学习

- Probably Approximately Correct
- 基本设定
  - 要学习的目标函数f来自某个函数的集合$\mathcal{F} = \{f: X \to \{0,1\}\}$，称为概念类。
  - 学习得出的函数h 称为假设。

PAC可学习

一个概念类 $\mathcal{F} = \{f: X \to \{0,1\}\}$ 是PAC可学习的，如果存在满足以下条件的概率算法$\mathcal{A}$和函数 $T(\mathcal{F}, \epsilon, \delta)$:

- 任给X上任意的概率分布$\mathcal{D}$，函数$f^* \in \mathcal{F}$和整数$t \geq T$:

$$\Pr\left[\begin{array}{c}(x_1, \dots, x_t) \leftarrow \mathcal{D}^t \\ y_i := f^*(x_i) \qquad : \mathcal{D}(f \,\Delta\, h) \leq \epsilon \\ h \leftarrow \mathcal{A}(\{(x_i, y_i)\}_{i \in [t]}, \epsilon, \delta)\end{array}\right] \geq 1 - \delta.$$

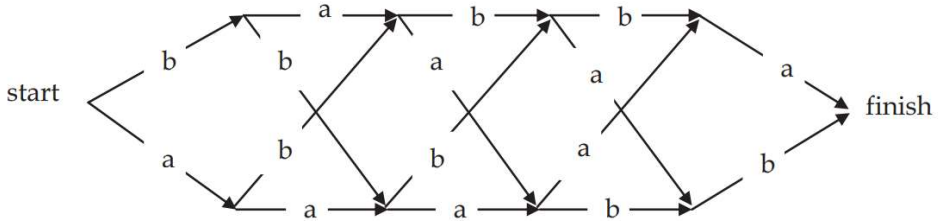
- $\mathcal{A}$在$\text{poly}\left(T, \frac{1}{\epsilon}, \frac{1}{\delta}\right)$时间内运行。

$$\mathcal{D}(f \,\Delta\, h) := \Pr_{x \leftarrow \mathcal{D}}[f(x) \neq h(x)]$$

PAC学习的能力与极限

- 可学习的例子：
  - $\mathcal{F}_{RECTANGLES}$ = 平面上与坐标轴平行的矩形
  - $\mathcal{F}_{BOOL}$ = n 个布尔变量的析取
  - $\mathcal{F}_{CNF}$ =长度不超过ℓ的CNF
- 不可学习的例子： $\mathcal{F}_{REGULAR}$ = 正则语言 （能被确定有限自动机识别的语言）
  - 正则语言的PAC学习算法可以用于破解RSA *

Leslie Valiant
2010年图灵奖得主



* M. Kearns and L. G. Valiant, "Cryptographic Limitations on Learning Boolean Formulae and Finite Automata," *Journal of the ACM* 41, no. 1 (1994): 67– 95.
** Picture from *Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World* By Leslie Valiant

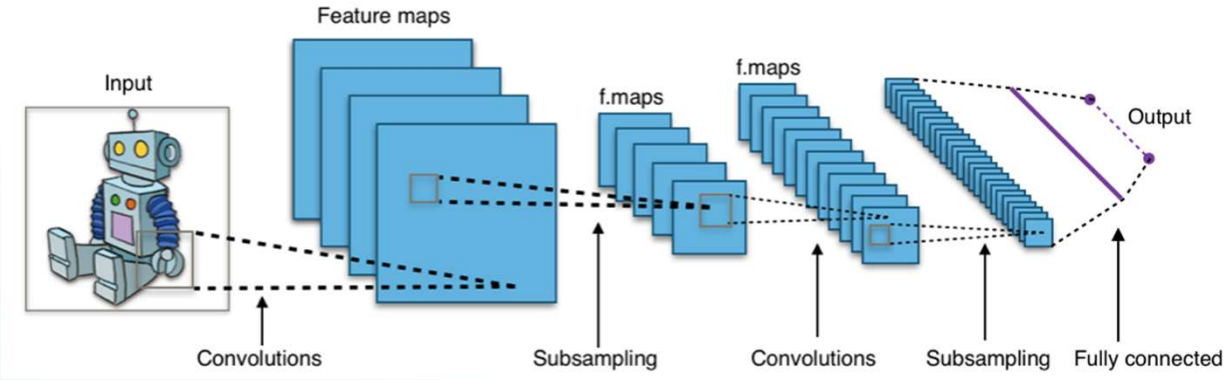
进化 vs. 学习：“人工”在哪里？

- 从学习的观点看生物进化：
  - 进化论的“量化版本”：生物界何以快速繁荣？
  - 学习目标：更好的适应自然环境
  - 例子：基因的表达
- 生物的进化与机器的学习有何不同？
  - 起始条件不同：机器学习可以人为地设置较好的初始假设
  - 反馈机制不同：自然的反馈机制有时更不直接，有时直接导致生物个体死亡。
- “人工”在哪里？
  - 硅基生物不同于碳基生物？采用了不同的算法？
  - 演化的环境和时间不同！

智能时代

这是最好的时代，这是最坏的时代。

深度学习：连接主义的复兴



2010至今：数据+算力

2018年，Yoshua Bengio, Geoffrey Hinton, Yann LeCun 因为深度神经网络的相关工作获得图灵奖。

哲学迷思：机器能思考吗?

- 中文房间
  - 思维不仅仅是程序和神经元的物理运作

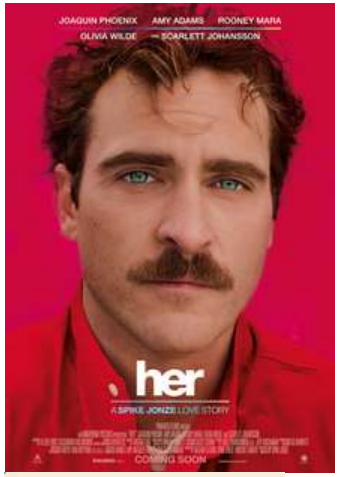
- “机器就是不能做某事”

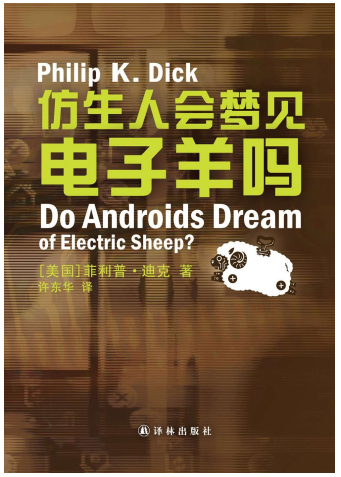
图灵在论文里举例：
“Be kind, resourceful, beautiful, friendly, have initiative, have a sense of humour, tell right from wrong, make mistakes, fall in love, enjoy strawberries and cream…”

他认为这只是人们对机器的固有印象。
人们需要“公平地对待机器”。

展望未来

- 智能化、精细化的社会
  - 交通信号灯的控制
  - 个性化的服务
- 隐私问题
  - 价格歧视
  - 保险
  - 老大哥在看着你?
- 多数人成为“无用之人”?





Thanks for listening ☺