

第一章 绪论

周世斌

中国矿业大学 计算机学院

May. 2022

1 引子

- AI 大起大落
- AI,ML,DM,PR
- 机器学习先修课程
- 工欲善其事必先利其器

2 人工智能概述

3 机器学习系统

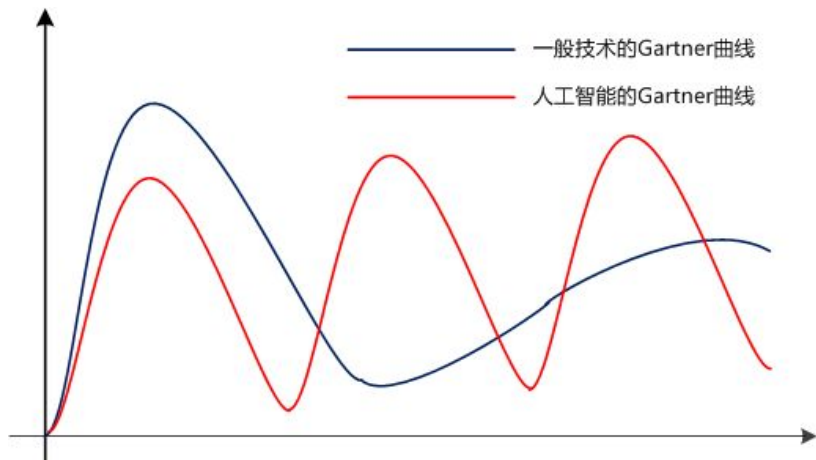
1 引子

- AI 大起大落
- AI,ML,DM,PR
- 机器学习先修课程
- 工欲善其事必先利其器

2 人工智能概述

3 机器学习系统

Gartner 曲线



六十多年前的达特茅斯会议，提出了“Artificial Intelligence”的课题，相关的研究也如火如荼地开展了起来。是为第一起。

当时，学者们解决人工智能问题的思路，是以人为师，通过专家编制规则的方法，教机器下棋、认字乃至语音识别。在今天看来，这样的方法是完全南辕北辙的——人类的视听器官虽然很发达，却并没有能力总结提炼其中的规律。于是，人工智能的美好憧憬中迎来了残酷的现实，学者们发现解决问题是如此遥远，围观群众也一度认为人工智能的学者都是骗子。是为第一落。

既然靠人指导不行，那就要祭出“实事求是”的法宝，从数据里统计规律。在这样数据 + 统计的方法论下，诸如人脸识别、手写识别等一些较为简单的问题取得了重大进展，到此时，我们感觉找到了解决人工智能问题的基本思路。是为第二起。

但是，拿语音识别为例，已经“基本可用”，但不“实用”，十几年过去，仍然无法实用，于是大家又转向悲观，觉得人工智能还只是个梦。是为第二落。

有一位老前辈 Jeffrey Hinton，和他的学生 Alex 一起，发现用 GPU 算神经网络，能大幅提高速度，于是这种模型居然可能实用了（在 ILSVRC2012 斩获大奖）。于是乎，深度模型的优势就发挥出来了，于是在语音识别、图像识别等领域带来了飞跃式的进展。是为第三起。各路大佬不论过去是做黑产、卖假货还搞劫持的，都摇身一变成了人工智能的忠实拥趸和业界先驱.....

1 引子

- AI 大起大落
- AI,ML,DM,PR
- 机器学习先修课程
- 工欲善其事必先利其器

2 人工智能概述

3 机器学习系统

Artificial intelligence

- In the field of computer science, artificial intelligence (AI), sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and other animals. Computer science defines AI research as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals. More specifically, Kaplan and Haenlein define AI as "a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation". Colloquially, the term "artificial intelligence" is applied when a machine mimics "cognitive" functions that humans associate with other human minds, such as "learning" and "problem solving".

Machine learning

- Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task **without using explicit instructions**, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model of sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are used in the applications of email filtering, detection of network intruders, and computer vision, where it is infeasible to develop an algorithm of specific instructions for performing the task. Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning.

Data mining

- Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. Data mining is an interdisciplinary subfield of computer science and statistics with an overall goal to extract information (with intelligent methods) from a data set and transform the information into a comprehensible structure for further use. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD .

Pattern recognition

- Pattern recognition is the automated recognition of patterns and regularities in data. Pattern recognition is closely related to artificial intelligence and machine learning, together with applications such as data mining and knowledge discovery in databases (KDD), and is often used interchangeably with these terms. However, these are distinguished: machine learning is one approach to pattern recognition, while other approaches include hand-crafted (not learned) rules or heuristics; and pattern recognition is one approach to artificial intelligence, while other approaches include symbolic artificial intelligence. A modern definition of pattern recognition is:
 - The field of pattern recognition is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions such as classifying the data into different categories.

1 引子

- AI 大起大落
- AI,ML,DM,PR
- 机器学习先修课程
- 工欲善其事必先利其器

2 人工智能概述

3 机器学习系统

- Artificial intelligence, Machine learning, Pattern recognition 与数学的关系
 - 线性代数、矩阵论
 - 最优化理论
 - 概率与数理统计
 - ...
- Artificial intelligence, Machine learning, Pattern recognition 和计算机类其他课程的关系
 - 计算机基础知识 ((数据结构、操作系统、算法设计、...))
 - 并行计算

1 引子

- AI 大起大落
- AI,ML,DM,PR
- 机器学习先修课程
- 工欲善其事必先利其器

2 人工智能概述

3 机器学习系统

人生苦短，我要学 Python

那 Python 是一种什么语言？

用任何编程语言来开发程序，都是为了让计算机干活，比如下载一个 MP3，编写一个文档等等，而计算机干活的 CPU 只认识机器指令，所以，尽管不同的编程语言差异极大，最后都得“翻译”成 CPU 可以执行的机器指令。而不同的编程语言，干同一个活，编写的代码量，差距也很大。

比如，完成同一个任务，C 语言要写 300 行代码，Java 只需要写 100 行，而 Python 可能只要 20 行。


Python 集成包, Anaconda

Anaconda 里面集成了很多关于 python 科学计算的第三方库, 主要是安装方便, 而 python 是一个编译器, 如果不使用 anaconda, 那么安装起来会比较痛苦, 各个库之间的依赖性就很难连接的很好

👉 下载并安装 anaconda

先到 <https://www.anaconda.com/products/individual> 下载 anaconda, 现在的版本 python3.8 版本, 下载好对应版本、对应系统的 Anaconda。

Anaconda Installers

Windows 

Python 3.8

64-Bit Graphical Installer (457 MB)

32-Bit Graphical Installer (403 MB)

MacOS 

Python 3.8

64-Bit Graphical Installer (435 MB)

64-Bit Command Line Installer (428 MB)


Linux 


Python 3.8


64-Bit (x86) Installer (529 MB)

64-Bit (Power8 and Power9) Installer (279 MB)

Python 机器学习: sklearn 库

 自 2007 年发布以来, scikit-learn 已经成为 Python 重要的机器学习库了, scikit-learn 简称 sklearn, 支持包括分类, 回归, 降维和聚类四大机器学习算法。还包括了特征提取, 数据处理和模型评估者三大模块。

 sklearn 是 Scipy 的扩展, 建立在 Numpy 和 matplotlib 库的基础上。利用这几大模块的优势, 可以大大的提高机器学习的效率。

 sklearn 拥有着完善的文档, 上手容易, 具有着丰富的 API, 在学术界颇受欢迎。sklearn 已经封装了大量的机器学习算法, 包括 LIBSVM 和 LIBLINEAR。同时 sklearn 内置了大量数据集, 节省了获取和整理数据集的时间。

Python 深度学习: PyTorch 库

- 👉 PyTorch 是一个开源的 Python 机器学习库。
- 👉 2017 年 1 月, 由 Facebook 人工智能研究院 (FAIR) 基于 Torch 推出了 PyTorch。它是一个基于 Python 的可续计算包, 提供两个高级功能: 1、具有强大的 GPU 加速的张量计算 (如 NumPy)。2、包含自动求导系统的深度神经网络。
- 👉 PyTorch 的前身是 Torch, 其底层和 Torch 框架一样, 但是使用 Python 重新写了很多内容, 不仅更加灵活, 支持动态图, 而且提供了 Python 接口。它是由 Torch7 团队开发, 是一个以 Python 优先的深度学习框架, 不仅能够实现强大的 GPU 加速, 同时还支持动态神经网络, 这是很多主流深度学习框架比如 Tensorflow 等都不支持的。PyTorch 既可以看作加入了 GPU 支持的 numpy, 同时也可以看成一个拥有自动求导功能的强大的深度神经网络

1 引子

- AI 大起大落
- AI,ML,DM,PR
- 机器学习先修课程
- 工欲善其事必先利其器

2 人工智能概述

3 机器学习系统

1 引子

- AI 大起大落
- AI,ML,DM,PR
- 机器学习先修课程
- 工欲善其事必先利其器

2 人工智能概述

3 机器学习系统