

超算与人工智能

南开大学 软件学院 张玉志

目录

- 一、超算简介
- 二、人工智能
- 三、超算与人工智能

一、超算简介

- 按照教科书，计算机历史发展分成了四代：
- 第一代(1946年～1958年)是电子管计算机。代表机型有:ENIAC、IBM650(小型机)、IBM709(大型机)等。
- 第二代(1959年～1964年)是晶体管计算机。代表机型有:IBM7090、IBM...
- 第三代(1965年～1970年)是集成电路计算机。代表机型有:IBM360系列、富士通F...
- 第四代(1971年至今)是大规模和超大规模集成电路计算机。
- 之后，尽管一直就作为第四代，其实发展很快

- 摩尔定律（不是物理定律，是经验定律）
 - 当价格不变时，集成电路上可容纳的元器件的数目，约每隔18-24个月便会增加一倍，性能也将提升一倍。
 - 换言之，每一美元所能买到的电脑性能，将每隔18-24个月翻一倍以上。
- 很神奇：从65年提出，近60年，一直遵循这个规律。计算机的性能提升了几万亿倍（五年十倍，十五年一千倍）

- 计算机类型另外一种分类：
- 1、巨型化（1946-1980的重点）
- 2、微型化（1980-2000的重点）
- 3、网络化（1990-现在 重点）
- 4、人工智能化（2015-现在 重点）
- 5、多样化
 - 嵌入式、手机/Pad、个人计算机、服务器（C/S,B/S）、云、超算、量子计算

- 定点运算和浮点运算
 - 定点运算就是整数的运算
 - 浮点运算就是小数的运算
 - 科学计数法, $0.xxx \times 10^n$
- 一般的计算是定点运算
- 超算主要是浮点运算能力很强, 普通计算机计算浮点运算很慢

CPU与GPU

- 中央处理器（英语：Central Processing Unit，缩写CPU），是计算机中负责读取指令，对指令译码并执行指令的核心部件。计算机系统中所有软件层的操作，最终都将通过指令集映射为CPU的操作。
- 图形处理器（英语：Graphics Processing Unit，缩写：GPU），专门做图像和图形相关运算工作的微处理器。GPU的构成相对简单，有数量众多的计算单元和超长的流水线，特别适合处理大量的类型统一的数据。但GPU无法单独工作，必须由CPU进行控制调用才能工作。CPU可单独作用，处理复杂的逻辑运算和不同的数据类型，但当需要大量的处理类型统一的数据时，则可调用GPU进行并行计算。














- 计算机中常说的几个单位
 - 1024^2 的十次方, 1K
 - 1024K就是1M (百万)
 - 1024M就是1G (十亿)
 - 1024G就是1T (万亿)
 - 1024T就是1P (千万亿)
 - 1024P就是1E (百亿亿, KP, MT, GG。就很少有人这么叫了, 就叫E, 或者1000P)
 - KMGTPPEZYBND

- 同构计算。同构计算是使用相同类型指令集和体系架构的计算单元组成系统的计算方式。同构超算只单纯使用一种处理器。
 - 日本超算“京”
 - 神威蓝光
 - IBM 的Mira和Sequoia
- 异构计算。使用不同类型指令集和体系架构的计算单元组成系统的计算方式。常见的计算单元类别包括CPU、GPU等协处理器、DSP、ASIC、FPGA等
 - 中国天河
 - 美国泰坦

- 超算的发展速度，超过了摩尔定律
- 大体上，每年翻番（总规模也在扩大）
- 超算十年速度大体上提高1000倍
- 价格18个月降一倍，十年大体上100倍
- 造价还是要提高十倍左右（量级）

- 天河一号（2010）2500万亿次，2.5P
- 天河二号（2014）3.39亿亿次，33.9P=>100P
- 太湖之光（2016）12.5亿亿次，125P
- 美国的顶点（2018年summit）201P
- 太湖之光（2019?）xxxP，几百P
- 天河三号（2021?）xxxxxP=? E级，就是过千P
 - 现在细节保密了

2021年6月 超级计算机TOP10

排行	品牌	地点	发布时间 ∨	峰值性能Tflops
		日本理化研究所	2020	549143.00
		美国能源部国家实验室	2018	200794.90
		美国能源部国家实验室	2018	125712.00
4		国家超级计算无锡中心	2015	125435.90
5		国家超级计算广州中心	2013	100678.70
6		美国能源部国家实验室	2021	89794.50
7		美国	2020	79215.00
8		德国于利希研究中心	2020	70980.00
9		沙特阿美	2020	55423.60
10		日本产业技术综合研究所	2021	54341.00



- 超算不是普通的计算机。普通的计算机主要是定点运算，超算主要强调浮点运算。
- 所以不能简单比较
- 非要比的话，普通的台式机可以算作十亿次量级，就是G量级。1P算力大体上折合上百万台台式机。天河三就是相当于几十亿台台式机
 - 相当于全国人民每人一台计算机
- 天河一号一年计算的题目，在天河三上跑不了一天
- 天河一号十年算过的所有题目，在天河三上一周内就可以计算完了

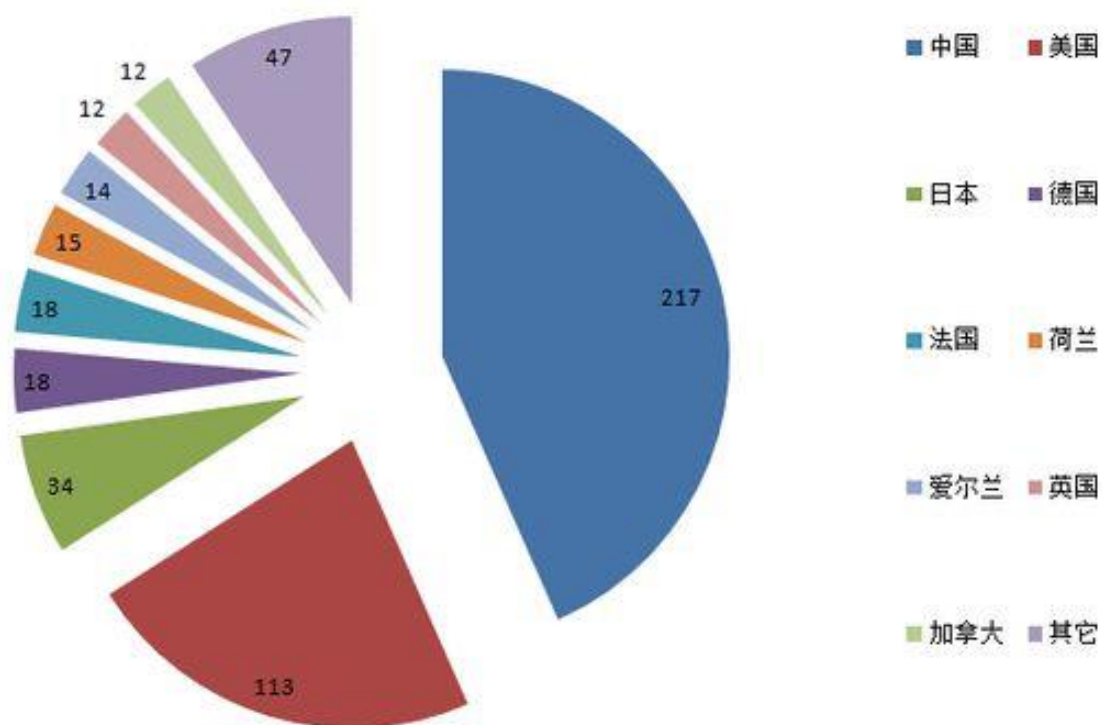
- 可以这样类比
- 普通计算机相当于小汽车，拉人为主可以运点货物
- 超算相当于轮船/火车/飞机/载重汽车，拉货为主偶尔拉人也可以
- 嵌入式或者手机，相当于电动自行车

- 超算用途是什么？
- 超算是国之重器
- 超级计算机拥有极快的运算速度，可以应用于多种尖端科技行业
- 弹道计算、核计算、石油勘探、天气预报、污染预报、海洋环境、数值风洞、碰撞仿真、动漫渲染、医药生物、材料科学、工业设计等等
- 人工智能计算(深度学习为代表)

- 自从天河一号，十年以来，中国的超算一直位居世界前列，自2011年起，9年中世界超算冠军都是中美轮流坐，其中中国凭借靠着天河二号、太湖之光蝉联了10届冠军，美国则是最近几年夺冠
- 天河二号之后，就开始阻止限制中国的超算，禁运了很多东西
- 这两年中国开始不公开说，很少参加评测了，应该是发展的很好

- 2021年6月最新公布的世界超算排名（参加评测的）
 - 第1名是日本的Fugaku（富岳），537P
 - 第2名是美国的Summit（顶点），200.79490P，148.60000P
 - 第3名是美国的Sierra（山脊），125.71200P，94.64000P
 - 第4名是中国的“神威太湖之光”，125.43590P，93.01460P
-
- 看上去，是几年以来我国的超级计算机首次掉出TOP500前三名
 - 我国上榜TOP500的超算数量总量仍据第一位，算力据第二位占据32.3%左右，美国的超算数量排第二位，算力排第一占据大概37.1%左右

2020年11月TOP500国家分布统计图（数量）



中国超算研究和天津的作用

- 中国第一台计算机诞生于1958年，比世界落后12年以上
- 银河一号，是中国第一台巨型机，1983年
- 前期的计算机基本上都是为了国防而研制的
- 当时禁运的很厉害
- 八十年代只有两个计算机院士：高庆狮、慈云贵

历史和故事

- 天河一号，落户在天津滨海新区
- 我是亲历者
- 这里的历史和故事，发生在2009年左右
 - 三家单位，准备做两台（P级），怎么办？
 - 用GPU做大规模的超算，行不行？
 - 天河是全世界第一家大规模采用的
 - 开始科技部验收不认可，只好增加规模
 - 美国人觉得很好，承认了
 - 逼出了世界第一
 - 于是掀起了业界一股风潮

- 中国超算主要是三家单位比较强
- 国防科技大学 “银河” “天河” 系列
- 总参56所 神威系列
- 中科曙光 曙光计算机
- 浪潮、联想、华为也有超算产品

- 天津因为安装了天河一号计算机
- 在中国乃至世界超算的发展进程中，发挥了重要的作用
- “天河一号”是具有重要作用和意义的超级计算机
 - 中国第一次拿到世界第一
 - 第一次采用GPU承担主要计算量的异构模式

二、人工智能与超算的关系

- 人工智能热点是深度学习
- 深度学习对超算的需求强劲
- 所以说，人工智能非常需要超算
- 超算对人工智能有重要的作用

- 最近红火的GPT-3，被大家传的很神，计算量非常大
 - 开始被推测需要1000万美元的计算费
 - 最新说法是460万美元
 - 最新型的GPU，约400个GPU年
 - 初步估计在天河三上要算2天
- GPT-3大体上是怎么回事？

- 人工智能要解决的一大类问题是：因为计算量太大，无法求“最优”解，只能寻求一个满意解
- 当计算能力上去了之后，就可以使用新方法

- 机器智能=算法+算力+（标注）数据
- 研究的对象是算法，数据和算力就非常重要了

- 新一代人工智能的实质：根据人给的大量例子，让机器模拟人的神经传导和计算方式，一次次调试优化到人的神经元的状态，从而实现像人一样的智能
- 人不需要去找出规律，机器自动的去找
- 传统的计算机是给出算法，机器去算。人工智能是让计算机去找出算法。难度不在一个层次上

- 目前的计算机，在给出算法的基础上，已经算的很快了，能够较好的满足人的需求，甚至超出了
- 但在人没有给出算法的条件下，需要机器找出方法，内部有大量的尝试，计算量就上去了
- 从另一个角度来看，未来发展最快的，也正正在此
- 因为代表着未来，所以重要和关键

- 超算也需要人工智能这样的“用武之地”发挥作用
- 实际上这里有一个卡脖子的重要事项
- 国产的加速部件，如何在人工智能计算中发挥出作用？
- 我们团队，一定意义上解决了这个问题

- CPU调用GPU（其他加速部件），怎么调用？
- 英伟达给出了一个标准，CUDA库
 - 英伟达的老板是黄仁勋（台湾出生），市值超过了Intel
- CUDA库做得不错，用起来简单，效率高
- 目前，基本上所有的重要的深度学习框架都是基于CUDA库的
- 但CUDA库只能用英伟达的GPU
- 而英伟达的GPU对中国有限制

- 其他的加速部件，非英伟达的GPU，对接的库是OPENCL
- 深度学习的框架（除了极个别很小的如Caffe）不支持OPENCL
- 怎么转到支持OPENCL？
- 清华和56所的团队没有攻下来
- 一个框架几百万行代码
- 工作量很大，我的团队已经转换完了TensorFlow，Mxnet

- 深度学习框架已经可以在天河3超算上运转了！

超算

人工智能

天津

能否成为
人工智能
聚集地？

- 天津应该利用好天河三号的机会，取得发展优势
- 围绕着天河三、人工智能，要建立一个人工智能的聚集地
- 如何形成产业链？
 - 各个部门拥有的数据有条件开放
 - 数据共享和交易机制
 - 标注数据形成一个行业
 - 人工智能应用成为一个行业
 - 改造传统制造业
 - 市场对外开放

补充：

- Science再次发布“全世界最前沿的125个科学问题”！
- 涉及：数学、化学、医学、生命、天文、物理、工程、材料、神经、生态、能源、信息科学、人工智能

人工智能

- 1. Will injectable, disease-fighting nanobots ever be a reality?
 - 可注射的抗病纳米机器人会成为现实吗？
- 2. Will it be possible to create sentient robots?
 - 是否有可能创建有感知力的机器人？
- 3. Is there a limit to human intelligence?
 - 人类智力是否有极限？
- 4. Will artificial intelligence replace humans?
 - 人工智能会取代人类吗？

- 5. How does group intelligence emerge?
 - 群体智能是如何出现的？
- 6. Can robots or AIs have human creativity?
 - 机器人或 AI 可以具有人类创造力吗？
- 7. Can quantum artificial intelligence imitate the human brain?
 - 量子人工智能可以模仿人脑吗？
- 8. Could we integrate with computers to form a human-machine hybrid species?
 - 我们可以和计算机结合以形成人机混合物种吗？

信息科学

- 1. Is there an upper limit to computer processing speed?
 - 计算机处理速度是否有上限？
- 2. Can AI replace a doctor?
 - AI可以代替医生吗？
- 3. Can topological quantum computing be realized?
 - 拓扑量子计算可以实现吗？
- 4. Can DNA act as an information storage medium?
 - DNA可以用作信息存储介质吗？

工程与材料科学

- 4. Is a future of only self-driving cars realistic?
- 纯无人驾驶汽车的未来是否现实？