



本课件仅用于教学使用。未经许可，任何单位、组织和个人不得将课件用于该课程教学之外的用途(包括但不限于盈利等)，也不得上传至可公开访问的网络环境

1

新媒体大数据分析

New Media Big Data Analysis

第一章 数据科学基础

黄振亚，朱孟潇，张凯

课程主页：

<http://staff.ustc.edu.cn/~huangzhy/Course/NM2025.html>

助教：齐畅 朱嘉骏

bigdata_2025@163.com

9/22/2025



数据科学基础

50

改变这个世界的四种力量

暴力



知识



大数据



世界著名未来学家托夫勒
《第三次浪潮》作者



金钱



数据科学基础

51

- 数据蕴含着巨大的价值——智慧医疗
 - 通过对患者建立AI电子病历
 - 整合患者的全时段、多模态的健康数据（病例文本、检查影像等）
 - 实现对患者的疾病诊断、病灶识别、药物推荐等

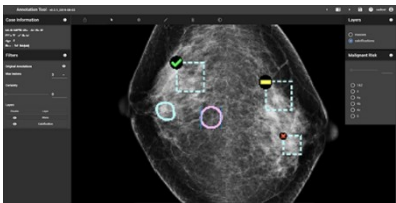


AI电子病历

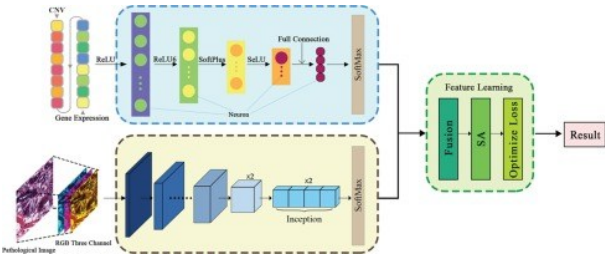


Input: EHR		Output: Diagnosis Results															
History of Present Illness: 3 days ago, Peter began to experience headache, cough, chest, non-productive cough, SOB, and chills. She says 3 people at work have been sick with headache, sore throat.... She was also hypertensive in 150's, so was given Dilzem.... Physical Exam: ...Response: decreased air movement through vs. Diffuse end-expiratory.... Major Surgical or Invasive Procedures: cardiac catheterization; intubation.... Discharge Diagnosis: Hospital-acquired Pneumonia; Pulmonary edema; Elevated cardiac enzymes; Hypertension; Arrhythmia.		 Clinical Diagnosis Model	<table><tr><th>Diagnosis Code</th><th>Diagnosis Description</th></tr><tr><td>486</td><td>Pneumonia Organism Unspecified</td></tr><tr><td>518.81</td><td>Acute Respiratory Failure</td></tr><tr><td>410.81</td><td>Unspecified Essential Hypertension</td></tr><tr><td>491.21</td><td>Obstructive Chronic Bronchitis with Acute Exacerbation</td></tr><tr><td>427.89</td><td>Other Specified Cardiac</td></tr><tr><td>...</td><td>...</td></tr></table>	Diagnosis Code	Diagnosis Description	486	Pneumonia Organism Unspecified	518.81	Acute Respiratory Failure	410.81	Unspecified Essential Hypertension	491.21	Obstructive Chronic Bronchitis with Acute Exacerbation	427.89	Other Specified Cardiac	...	...
Diagnosis Code	Diagnosis Description																
486	Pneumonia Organism Unspecified																
518.81	Acute Respiratory Failure																
410.81	Unspecified Essential Hypertension																
491.21	Obstructive Chronic Bronchitis with Acute Exacerbation																
427.89	Other Specified Cardiac																
...	...																

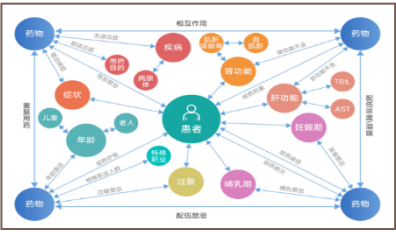
疾病诊断



病灶识别



多模态医疗数据挖掘模型



药物推荐



数据科学基础

52

- 数据蕴含着巨大的价值—安防领域
 - 公安监控智能分析



区间超速判定



“天眼”追凶

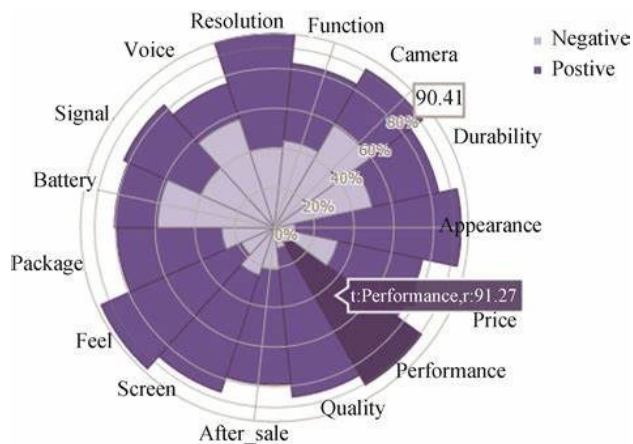


数据科学基础

53

数据蕴含着巨大的价值—安防领域

舆情监测



舆情情感分析



传播途径监测



数据科学基础

54

- 数据蕴含着巨大的价值——智慧教育
- 因材施教

学生的学习行为数据

1	[A]		[C]	[D]
2	[A]	[B]		[D]
3		[B]	[C]	[D]
4	[A]	[B]	[C]	
5	[A]	[B]		

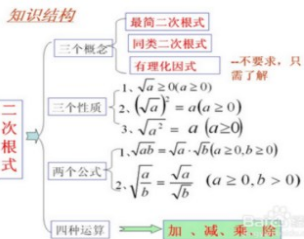


大数据分析

学生认知水平画像

试题难度等特征的预测

个性化学习推荐

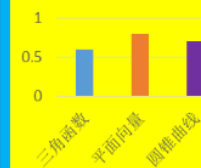


试题-知识点

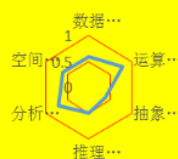


姓名	张三
学号	9527
平均正确率	85%
综合水平	90.562

考点掌握情况



能力分布图谱



$$9 - 3 \div \frac{1}{3} + 1 = ?$$

$$\frac{4}{7} \div 8 = ?$$

易



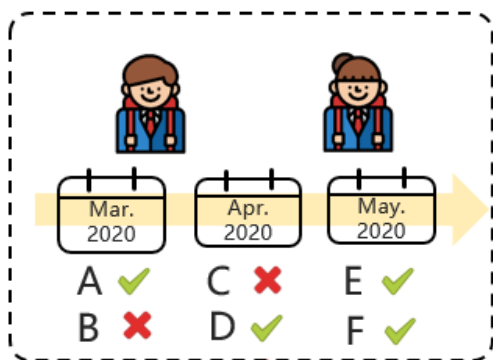


数据科学基础

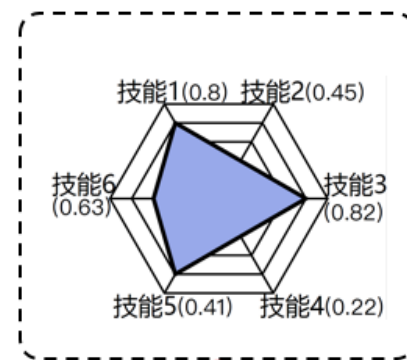
55

- 数据蕴含着巨大的价值——智慧教育
 - 学习者能力分析
 - 量表范式 到 数据驱动的模式

学习过程相关数据



知识能力及其发展变化



反馈调整教学策略



数据科学基础

56

- 数据蕴含着巨大的价值——智慧教育
 - 考试试题质量与公平性



浙江省教育考试院
ZHEJIANG EDUCATION EXAMINATIONS AUTHORITY

首页 | 组织机构 | 信息公开 | 政策法规 | 政策解读 | 2018年11月27日 星期二 11:11:39 | 请输入关键字 | 搜索

普通高考 | 学考选考 | 研究生考试 | 成人高考 | 自学考试 | 社会考试 | 教师资格考试 | 海外考试

关于英语科目考试成绩的说明

[发布时间:2018-11-27 阅读量:1570]

浙江省高考英语科目一年安排2次考试，考生可报考2次，选用其中较高1次的成绩。在2018年11月刚结束的英语科目考试中，根据答卷试评情况，发现部分试题与去年同期相比难度较大。为保证不同次考试之间的试题难度大体相当，浙江省招委组织专家研究论证，在制订评分细则时，决定面向所有考生，对难度较大的第二部分（阅读理解）、第三部分（语言运用）的部分试题进行难度系数调整，实施加权赋分。其他试题未作调整。

浙江省教育考试院

2018年11月27日



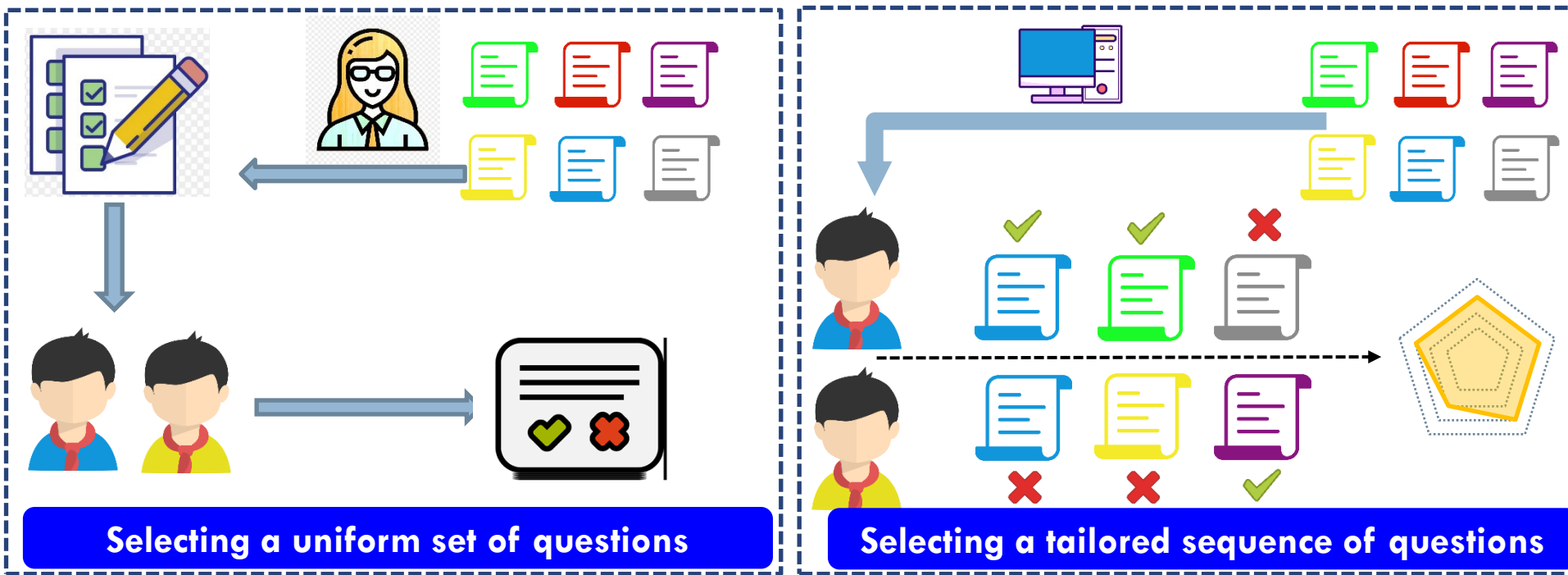
数据科学基础

57

- 数据蕴含着巨大的价值——智慧教育
 - 数据驱动的自适应测试（CAT）

GRE®

纸笔考试 ← 考试 → 计算机自适应测试





数据科学基础

58

- 数据蕴含着巨大的价值——智慧教育
 - 大模型驱动的智慧教育

学：智慧助教

教：数字教师

研：科研助手

管：以智助管

大模型学习助手

什么是布斯算法？已知二进制数 $X=0.1010$, $Y=-0.1101$ ，利用布斯算法求 $[X*Y]$ 的补码。

布斯算法是一种补码乘法算法，用于计算两个用补码表示的二进制数的乘积。在布斯算法中，被乘数和乘数均用补码表示，连同符号位一起参与运算。布斯算法的运算过程包括将两个数的补码相乘，并根据乘数相邻两位的情况进行特定操作，最终得到乘积的补码。

根据已知的二进制数 $X=0.1010$ ， $Y=-0.1101$ ，可以进行布斯算法求解 $[X*Y]$ 的补码：

1. 将 X 和 Y 表示为补码： $[X]_{补} = 00.1010$
 $[Y]_{补} = 11.0011$ $[-X]_{补} = 11.0110$

2. 根据布斯算法的运算过程，进行乘法运算，得到乘积的补码： $[X*Y]_{补} = 1.01111100$

因此，根据布斯算法， $[X*Y]$ 的补码为1.01111100。



星火科研助手

星火科研助手致力于创建高效、可靠的科技文献服务平台，为学术研究人员、企业专业人士和其它相关用户提供服务，助力科技发展创新。

成果调研 论文研读 辅助写作



关注订阅号
及时了解科研服务前沿信息



关注服务号
随时随地了解科研助手动态





数据科学基础

59

- 数据蕴含着巨大的价值——社会科学
 - ◆ 社交媒体 比 问卷调查 提供了更有代表性的结果
 - ◆ 智能引导社会成员的行为



15万名奥巴马支持者在Facebook安装了“奥巴马2012”应用，而通过这个程序，总统竞选团队可以间接得到这些支持者数百万的Facebook好友信息。



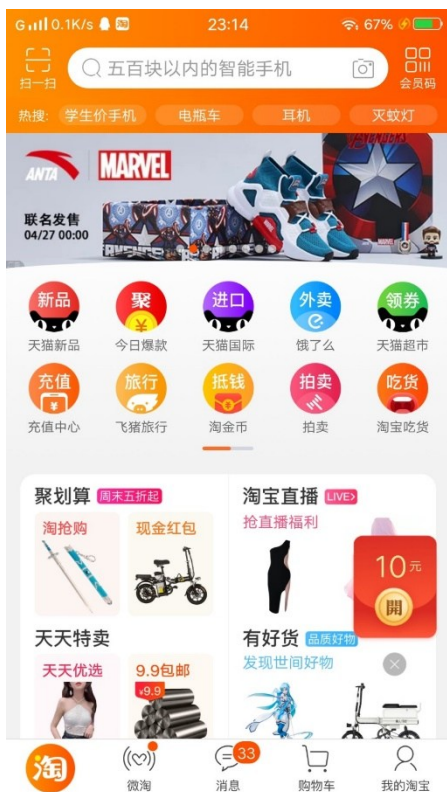
有一种说法称，特朗普的团队聘用数据分析公司，做了精准的广告投放，影响了那些徘徊不定的选民，拿下了决定性的关键州选举人票



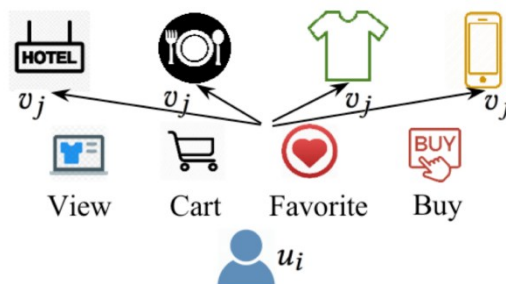


数据科学基础

- 数据蕴含着巨大的价值——电子商务
- 计算广告



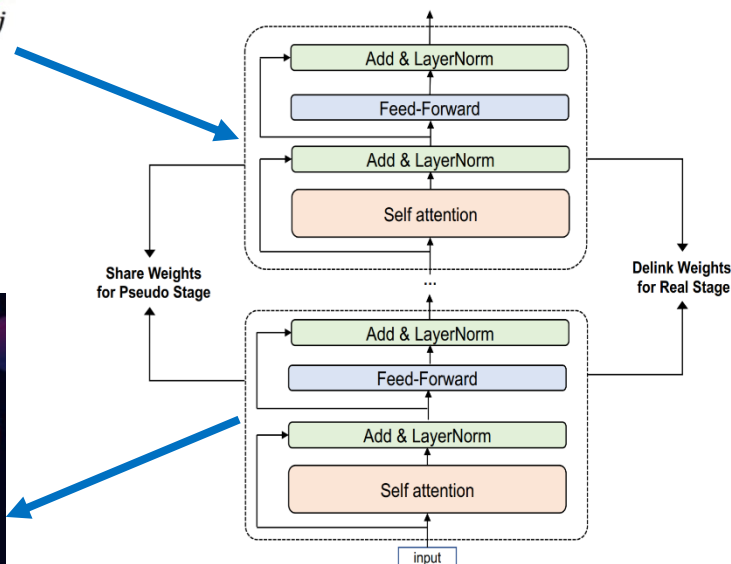
电商平台



海量用户多样交互行为



促进用户消费、提升平台收益



达摩院10万亿参数
M6-10T模型



数据科学基础

61

- 数据蕴含着巨大的价值——电子商务
 - 精准搜索、个性化消费推荐



天猫双11：破亿交易时间

年份	10亿	100亿	500亿
2014	3分	38分28秒	21时12分
2015	1分12秒	12分28秒	9时52分22秒
2016	52秒	6分58秒	2时30分20秒
2017	28秒	3分01秒	40分12秒
2018	21秒	2分05秒	26分03秒
2019	14秒	1分36秒	12分49秒

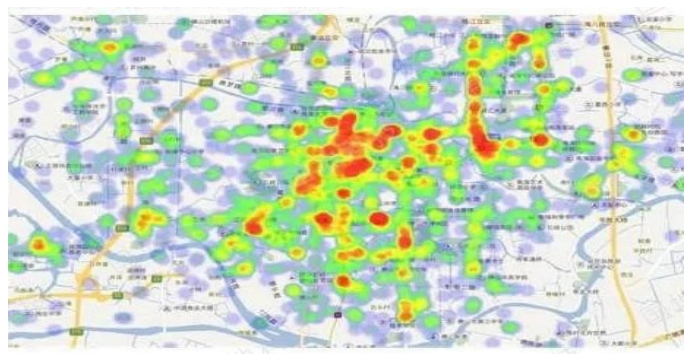


数据科学基础

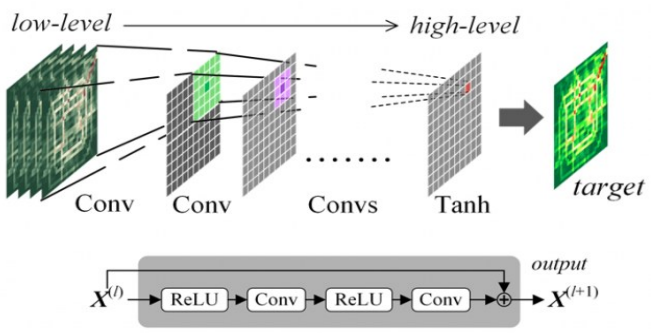
62

- 数据蕴含着巨大的价值 — 智慧城市
 - 基于运营商基站数据、交通路口和车辆移动轨迹数据等
 - 提升城市**交通管控**、**交通服务**和**规划水平**，实现**用户未来行程规划**、**交通路口信号灯调控**、**合理规划道路建设**等

车辆移动数据



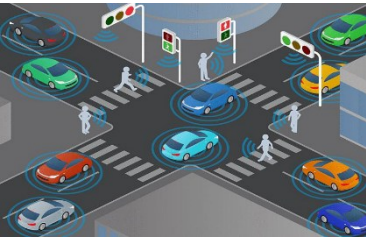
交通流预测模型



行程规划



信号灯调控



道路规划



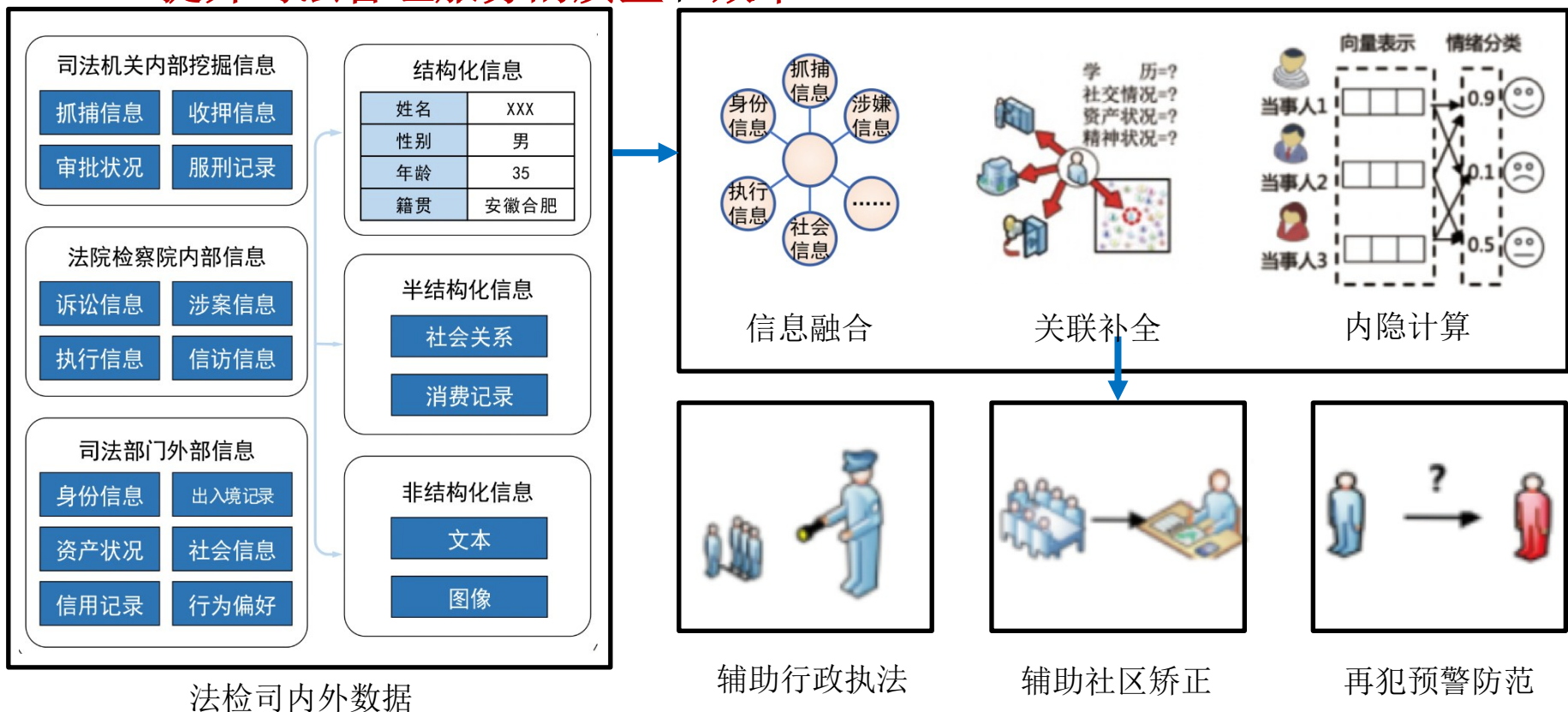


数据科学基础

63

数据蕴含着巨大的价值——智慧司法

- 基于法、检、司等部门关于涉案当事人的内部数据与外部数据等
- 构建涉案当事人画像，辅助行政执法、社区矫正、进行再犯预警防范等，**提升司法管理服务的质量和效率**





数据科学基础

64

- 数据蕴含着巨大的价值——文化娱乐
 - 纸牌屋效应：数据决定影视剧的内容



大卫·芬奇
凯文·史派西

老版《纸牌屋》

喜欢老版纸牌屋
及同类剧的用户

13集同时上线



数据科学基础

65

- 数据蕴含着巨大的价值——文化娱乐
 - 机器作诗



用户写作意图



江、船、秋风



春风、杨柳



松、竹、山、牧童



诗歌自动生成系统



对应诗词

江北江南万顷秋，
船头人去水悠悠。
一帆一棹秋风急，
又有离人万里愁。

杨柳千条拂地垂，
一川春水浸桃花。
游人不识湖中路，
游遍人间野水涯。

乔松古木两三间，
松竹阴中一径斜。
白鸟不知山路远，
牧童踏过野人家。



数据科学基础

66

□ 机器作诗 PK 古代诗人



哪首诗是人写的？

秋夕湖上

一夜秋凉雨湿衣，
西窗独坐对夕晖。
湖波荡漾千山色，
山鸟徘徊万籁微。

机器

秋夕湖上

荻花风里桂花浮，
恨竹生云翠欲流。
谁拂半湖新镜面，
飞来烟雨暮天愁。

宋代诗人葛绍体



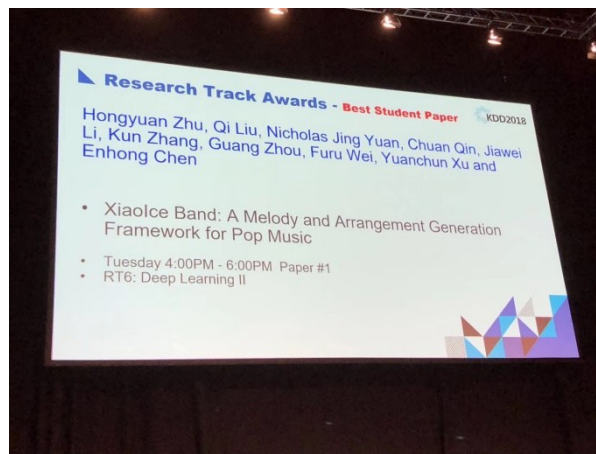
数据科学基础

67

数据蕴含着巨大的价值——文化娱乐

- 流行音乐的旋律与编曲生成
- 机智过人：

<http://tv.cctv.com/2017/11/24/VIDEo7JWp0u0oWRmPbM4uC Bt171124.shtml>



【KDD18最佳论文揭晓】中科大等斩获最佳学生论文，刘兵获创新奖，清华大学唐杰任副主席

● 首页 ● 新闻博览

我校获数据挖掘领域顶级国际会议KDD 2018最佳学生论文奖



数据科学基础

68

- 数据蕴含着巨大的价值—文化娱乐
 - 流行音乐的旋律与编曲生成

Sheet music for a musical score, measures 1-4. The tempo is marked $\text{♩} = 90$. The instruments listed are:

- 长笛 (Flute)
- 鼓组 (Drum Set)
- 无品电贝司 (Fretless Electric Bass)
- 乐队小提琴 (Violin)
- 古典吉他 (Classical Guitar)

The score shows the first four measures of the piece, with the Flute and Drum Set parts being the most active.

Sheet music for a musical score, measures 5-8. The tempo is marked $\text{♩} = 90$. The instruments listed are:

- Fl.
- D. Set
- Frd. El. B.
- Vlns.
- Guit.

The score shows the next four measures of the piece, continuing the melody and rhythm established in the first four measures.



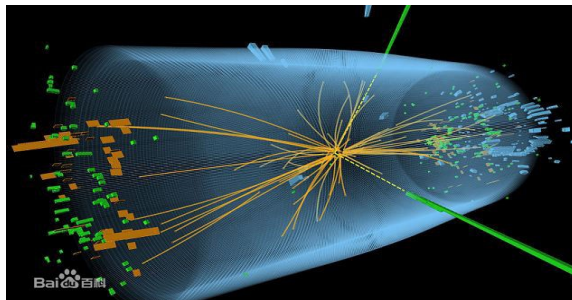
数据科学基础

69

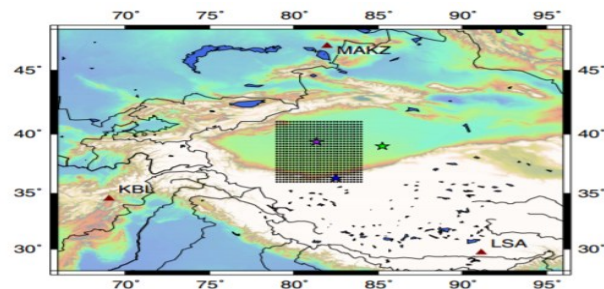
- 数据蕴含着巨大的价值—科学技术
 - ◆ 大数据推动科学新技术发现



天文大数据搜索新星



物理大数据预测分子属性



大数据地震速报、余震预测



生物大数据改良基因

科学
大数据



数据挖掘保护知识产权



数据科学基础

70

- 科技大数据来自于**物理世界**
 - 科学实验数据或传感数据
 - 技术描述型数据—专利、论文
- 集**多种特点**于一身
 - 采集的高代价性
 - 复杂性
 - 超高维度
 - 高度计算复杂性
 - 高度的不确定性
 - 学科知识壁垒
 - 信息与通信技术高度集成性

单一学科



数据驱动

多学科交叉



关系型数据库的鼻祖**Jim Gray**（右）





数据科学基础

71

2007年，Jim Gray总结出了四个科学范式





数据科学基础

72

□ 把握大数据带来的机遇

□ 零售业

- Winners: Amazon, Ebay
- Traditional: 传统书店、电子产品零售店

□ 旅游业

- Winners: Expedia, Ctrip
- Traditional: 旅行中介商

□ 金融服务业

- Winners: E*trade, TD Ameritrade
- Traditional: 股票中介商公司





数据科学基础



73

视频数据

- 把握大数据带来的机遇
- 影像租赁业
 - Winners: 视频流媒体公司(Netflix, Amazon, Hulu)
 - Traditional : DVD租赁公司
- 软件应用业
 - Winners: 软件数据服务公司(Salesforce.com)
 - Traditional : 软件产品公司
- 新闻报纸业
 - Winners: Google, Twitter, Facebook, Bloomberg
 - Traditional : 传统报纸业, Washington Post, WSJ
- 出租车行业
 - Winners: Uber, DiDi



数据科学基础

74

□ 新媒体

- 利用数字技术、网络技术和移动通信技术，通过互联网、宽带局域网、无线通信网和卫星等渠道，以电视、电脑和手机为主要输出终端，向用户提供视频、音频、语音数据服务、连线游戏、远程教育等集成信息和娱乐服务的所有新的传播手段或传播形式的总称，包括“新兴媒体”，也包括“新型媒体”

□ 新媒体大数据

- 新媒体服务场景中收集获取的数据
- 按数据的模态类型：语音，图片，视频，网络，文本



数据科学基础

75

多媒体大数据研究热点—语音大数据

语音识别

- 微软英语语音识别实现词错率5.9%的突破，第一次超越人类
- 科大讯飞语音识别词错率3%左右。中文领域突出





□ 多媒体大数据研究热点—图片大数据

□ 图像识别

- ImageNet图像数据库上，人工智能已达到2.99%的错误率（公安部三所），低于人类5.1%的错误率



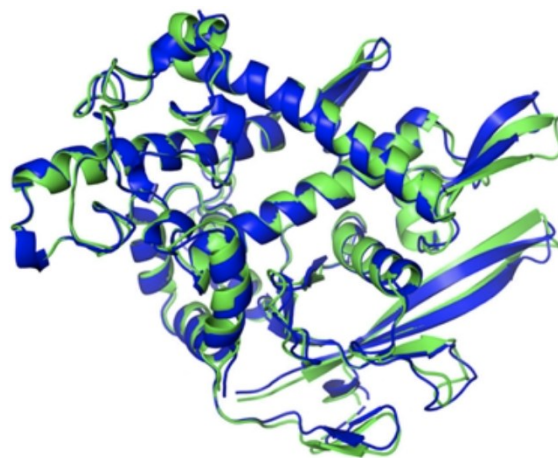
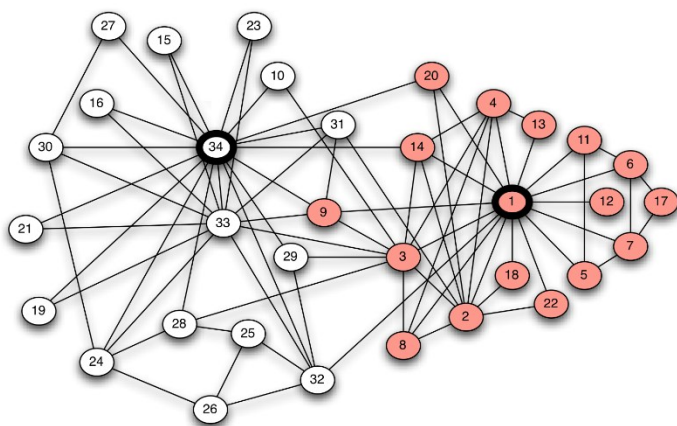
李飞飞
斯坦福大学、谷歌AI前任首席科学家



数据科学基础

77

- 多媒体大数据研究热点—网络大数据
 - 社会网络分析
 - 蛋白质结构图
 - 知识图谱

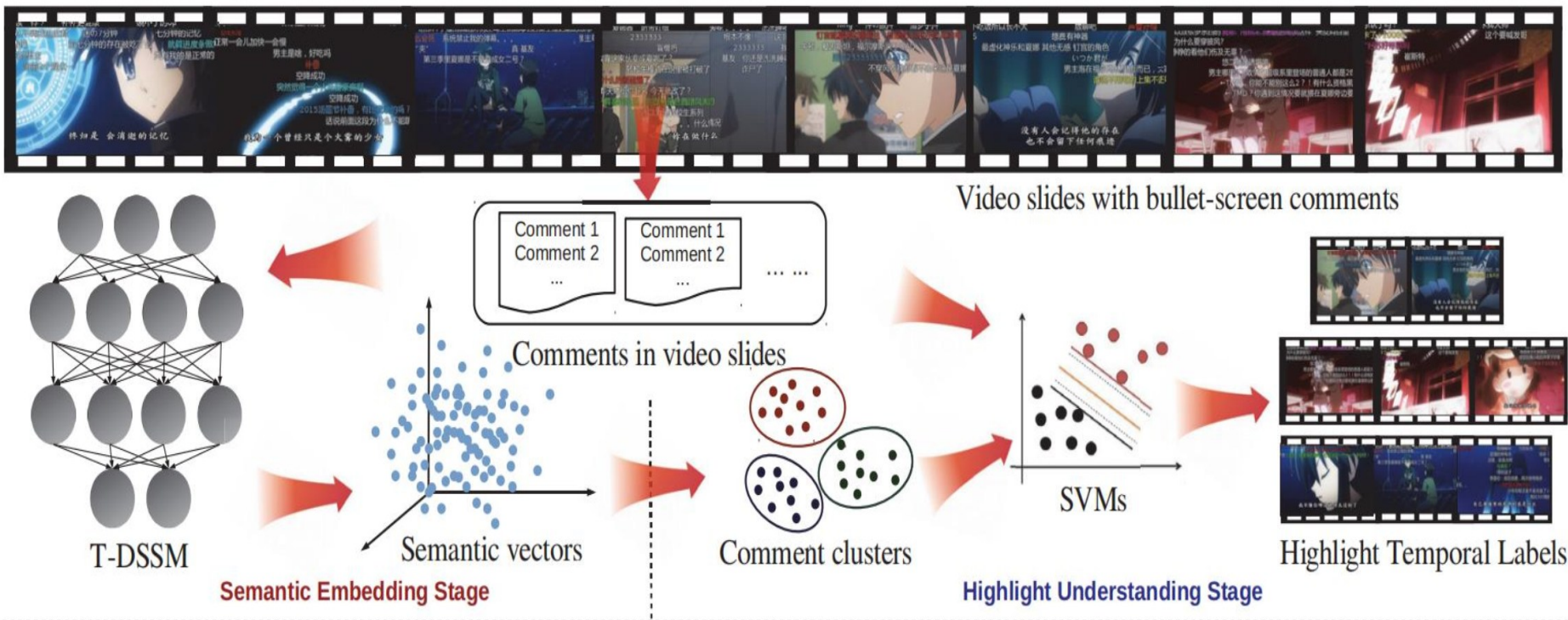




数据科学基础

78

- 多媒体大数据研究热点—视频大数据
 - 社会网络分析





数据科学基础

79

- 多媒体大数据研究热点—文本大数据
 - 自然语言处理
 - 机器阅读，机器翻译，文本推理，知识图谱等

通用语言理解评估（GLUE）基准

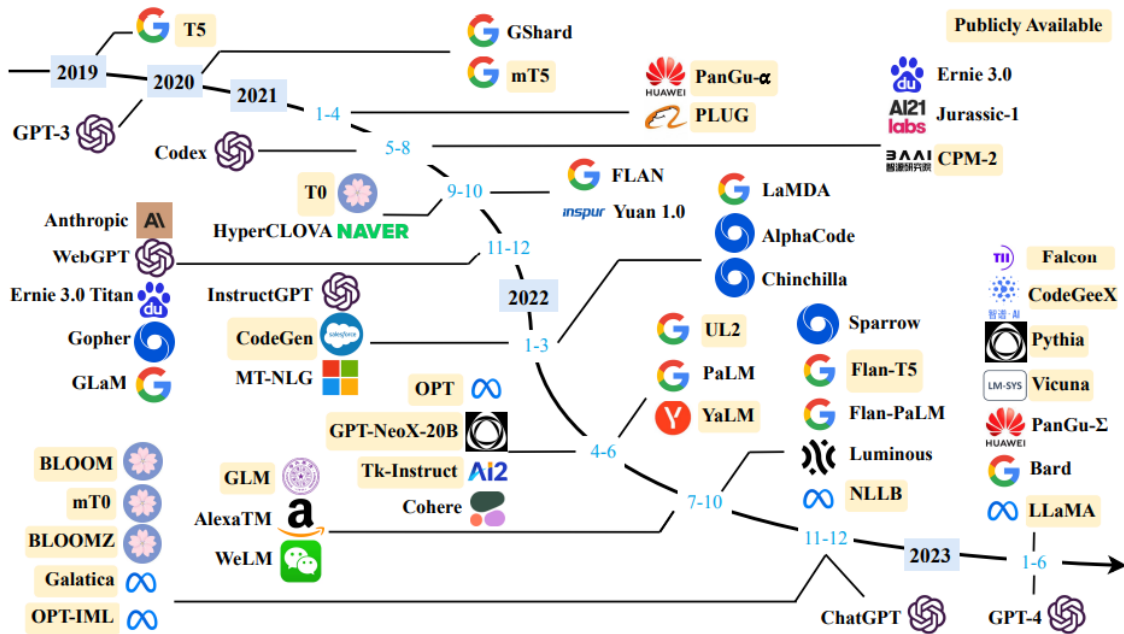
Rank Name		Model		URL	Score	CoLA	SST-2	MRPC	STS-B
+	1	Alibaba DAMO NLP	StructBERT		90.3	75.3	97.1	93.9/91.9	93.0/92.5
	2	T5 Team - Google	T5		90.3	71.6	97.5	92.8/90.4	93.1/92.8
	3	ERNIE Team - Baidu	ERNIE		90.1	72.8	97.5	93.2/91.0	92.9/92.5
	4	Microsoft D365 AI & MSR AI & GATECHMT-DNN-SMART			89.9	69.5	97.5	93.7/91.6	92.9/92.5
+	5	ELECTRA Team	ELECTRA-Large + Standard Tricks		89.4	71.7	97.1	93.1/90.7	92.9/92.5
+	6	Huawei Noah's Ark Lab	NEZHA-Large		88.7	67.4	97.2	93.2/91.0	92.2/91.6
+	7	Microsoft D365 AI & UMD	FreeLB-RoBERTa (ensemble)		88.4	68.0	96.8	93.1/90.8	92.3/92.1
	8	Junjie Yang	HIRE-RoBERTa		88.3	68.6	97.1	93.0/90.7	92.4/92.0
	9	Facebook AI	RoBERTa		88.1	67.8	96.7	92.3/89.8	92.2/91.9
+	10	Microsoft D365 AI & MSR AI	MT-DNN-ensemble		87.6	68.4	96.5	92.7/90.3	91.1/90.7
	11	GLUE Human Baselines	GLUE Human Baselines		87.1	66.4	97.8	86.3/80.8	92.7/92.6



数据科学基础

80

- 多媒体大数据研究热点—文本大数据
 - 大语言模型——多项技术融合
 - 如今大语言模型层出不穷，例如 ChatGPT, ChatGLM, LLaMa等，具备大量知识，能够以类人的方式和人类进行多轮对话，在多种自然语言理解相关任务上表现良好。





数据科学基础

81

- 多媒体大数据研究热点—文本大数据
- 大语言模型

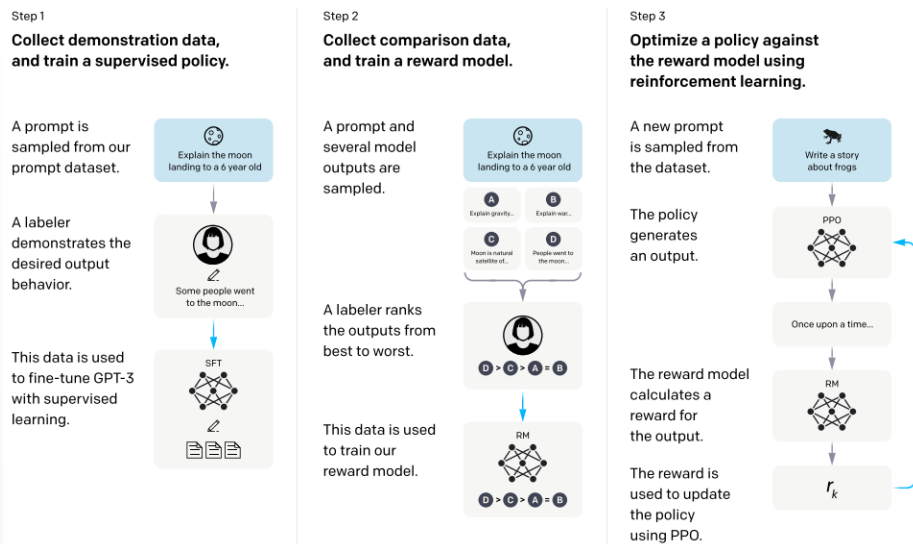
丰富大模型训练数据

➤ 数据类型

- 多轮对话^[2]、网页^[3]、源代码^[4]...

➤ 训练方式

- SFT
- 基于人类反馈的强化学习^[2] ...



例：基于人类反馈的强化学习

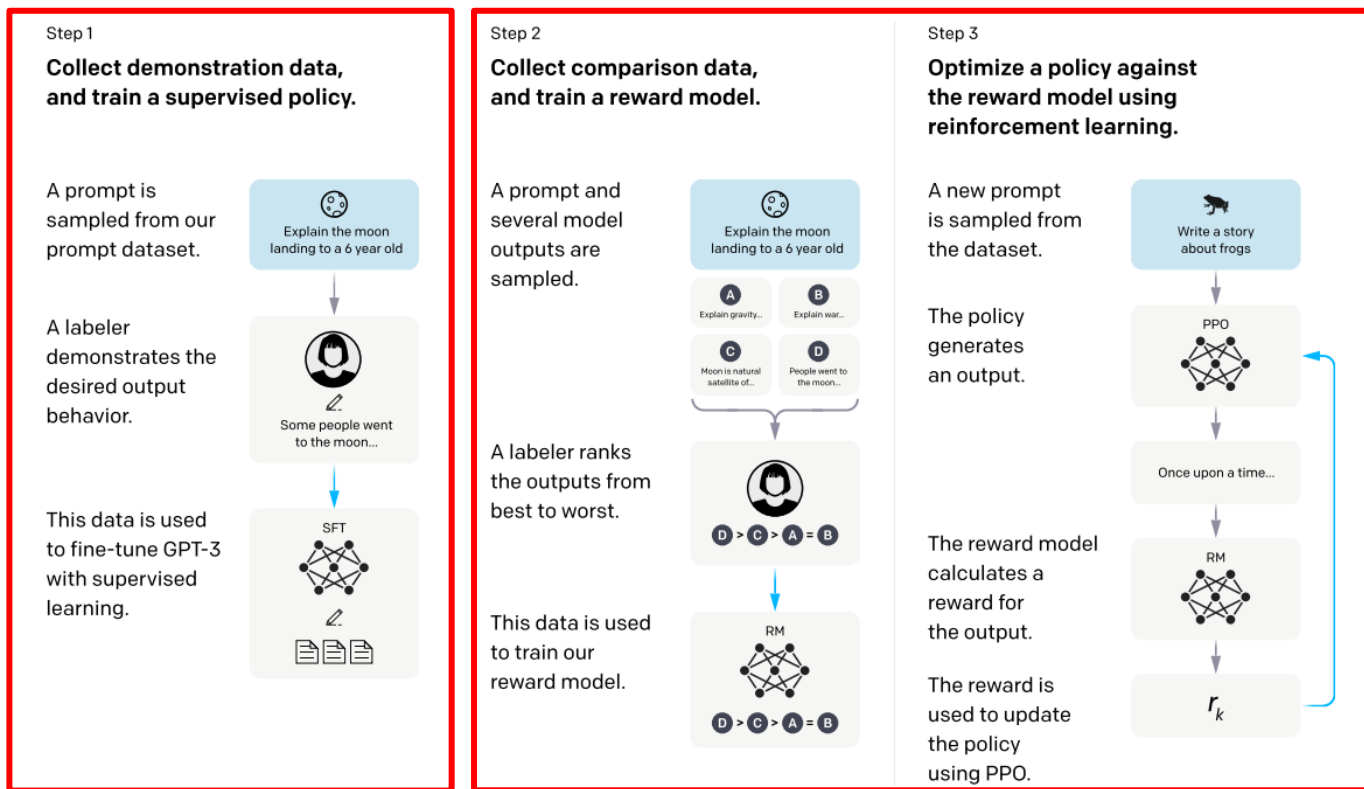


数据科学基础

82

- 多媒体大数据研究热点—文本大数据
 - 大语言模型

指令微调，
使大模型能够泛化到多种任务上



人工反馈的强化学习，使大模型与人类价值观对齐



数据科学基础

83

- 多媒体大数据研究热点—文本大数据
 - 大语言模型
 - 庞大的知识库
 - 通过指令实现在多种任务上的零样本/少样本学习

XB

什么是量子纠缠？



量子纠缠是一种奇特的量子力学现象，它发生在两个或多个粒子之间，当它们之间存在一种特殊的量子状态时，这些粒子之间的属性会彼此紧密关联，无论它们之间有多远的距离，这些关联性都会保持下去。这些属性可以包括粒子的自旋、位置、动量等。

ChatGPT作为知识库回答问题

XB

将文本中的语句分类成积极和消极两种情感

文本1: 这张照片拍的不错。
文本2: 今天雨下的很大，我的鞋湿了。



文本1: 这张照片拍得不错。 - 积极情感
文本2: 今天雨下得很大，我的鞋湿了。 - 消极情感

ChatGPT仅通过任务描述完成文本情感分类



数据科学基础

84

- 多媒体大数据研究热点—文本大数据
 - 大语言模型
 - 通过思维链（CoT）解决复杂问题

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ❌

Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✅

向大语言模型展示少量样本并解释推理过程，模型在回答时也会显示推理过程从而引出正确回答

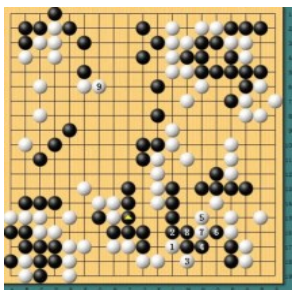


数据科学基础

□ 大数据的研究—数据与知识融合，让人工智能更“聪明”

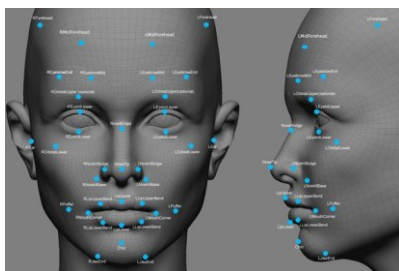
计算智能

- 规则明确、特定领域



感知智能

- 语音、图像、视频



认知智能

- 理解、推理、解释



能穿多少穿多少



能穿多少穿多少



数据科学基础

□ 大数据的研究—从数据中的相关性到世界的因果推断

逻辑关系

- 归纳法、数理逻辑、布尔代数系统

$$(a \vee b) \vee c = a \vee (b \vee c) \quad \text{重推理}$$
$$(a \wedge b) \wedge c = a \wedge (b \wedge c)$$

相关关系

- 贝叶斯网络、机器学习、深度学习



重分析
(学习)

因果关系

- 因果关系是有方向的、存在时序先后性

万有引力



数据分析+
逻辑推理



数据科学基础

➤ L1: 基础对话者

- 对话与交互能力: 当前(如GPT-4),能流畅的对话
- 特点: 以对话为核心, 提供信息、解答问题、辅助创作等

➤ L2: 推理者

- 推理与问题解决能力: OpenAI-O1接近L2。PhD-Level。
- 特点: 像人类一样分析问题、制定方案, 甚至在某些领域超越人类

➤ L3: 智能行动者

- 自主决策能力: AI能独立思考并根据复杂情境采取行动(接近人类)
- 特点: 不仅能够思考, 还能在真实世界中执行决策, 实现人机协同

➤ L4: 创新者

- 创造能力: AI能进行创造性思维和协助人类进行发明和创造, 推动科技进步
- 特点: 具备创新思维和创造力, 能够提出新的想法和解决方案

➤ L5: 组织者

- AGI: AI能执行和组织人类所有工作, 标志着真正的人工通用智能的实现
- 特点: 具备人类智能和组织能力, 能管理和优化整个社会的运行。

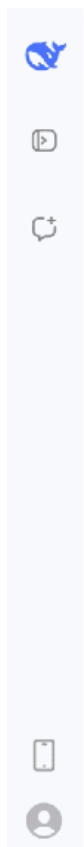
OpenAI's 5 Step to AGI

- | | |
|---------|---|
| Level 1 | Chatbots, AI with conversational language |
| Level 2 | Reasoners, human-level problem solving |
| Level 3 | Agents, systems that can take actions |
| Level 4 | Innovators, AI that can aid in invention |
| Level 5 | Organizations, AI that can do the work of an organization |



数据科学基础

□ 大数据的研究—推理大模型



我是 DeepSeek, 很高兴见到你!

我可以帮你写代码、读文件、写作各种创意内容, 请把你的任务交给我吧~

给 DeepSeek 发送消息



DeepSeek



联网搜索

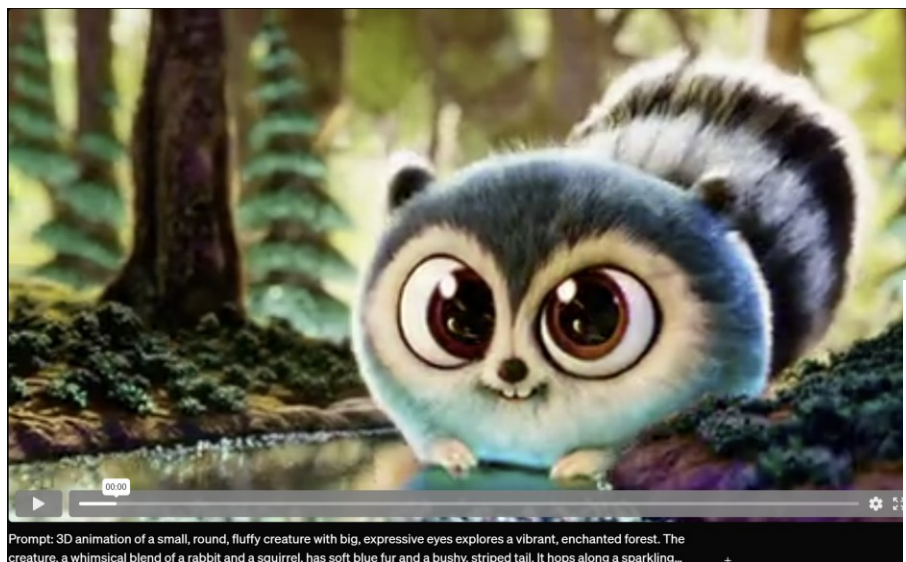




数据科学基础

□ 大数据的研究—多模态大模型

- SORA: <https://openai.com/index/sora/>
- 智象未来: <https://hidreamai.com/home>





数据科学基础

90

□ 大数据与人工智能

□ ABC当前AI的技术体系

Big data

数据

大数据是人工智能发展的**基石**，人工智能的核心在于数据支持。

机器学习算法是人工智能的**核心**，是今天引领人工智能发展潮流的一大类算法

算法

Algorithm

AI

算力

Computation

人工智能算法的实现需要强大的计算能力**支撑**，特别是深度学习算法的大规模使用，对计算能力提出了更高的要求。

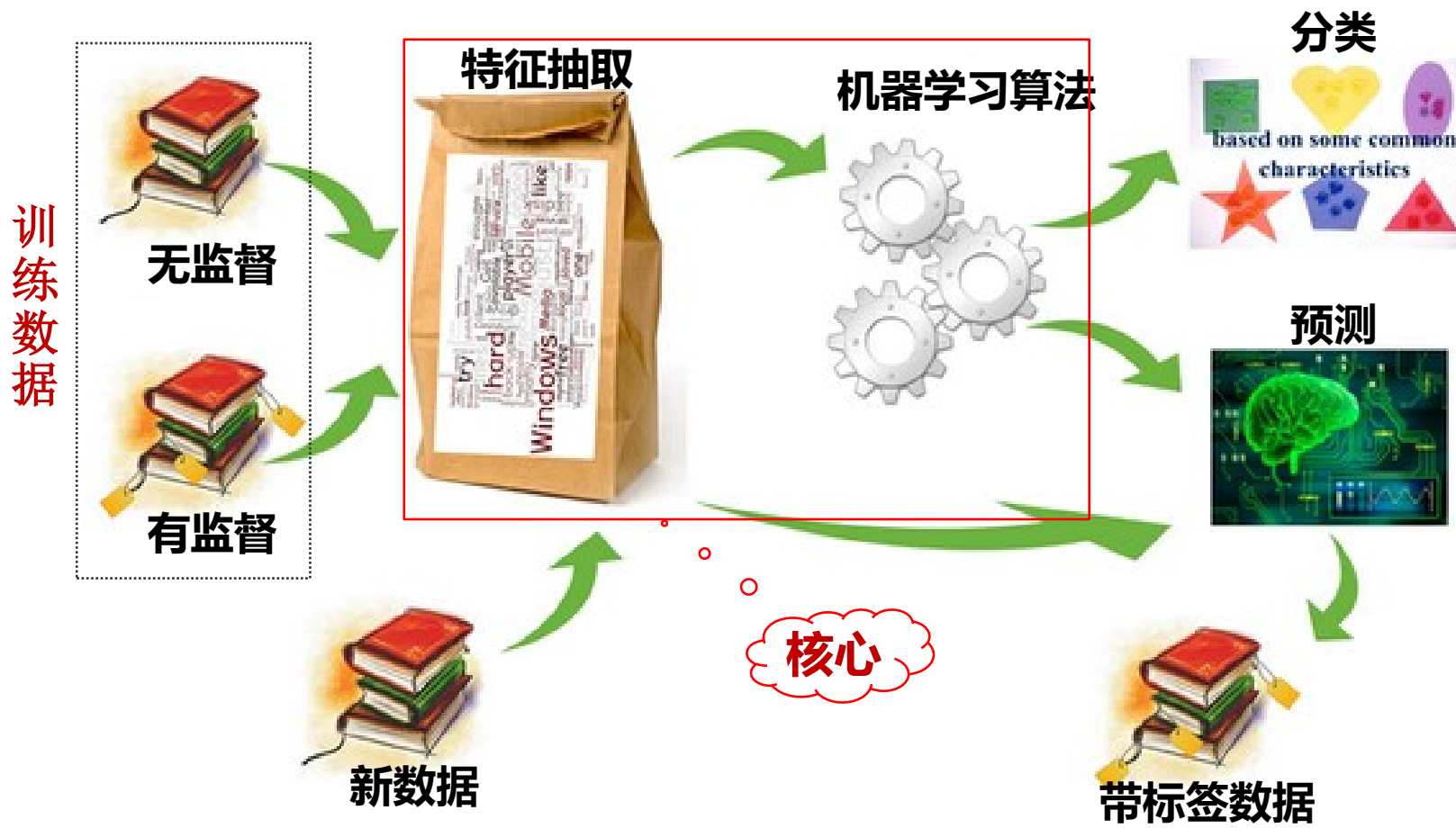


数据科学基础

91

大数据与人工智能

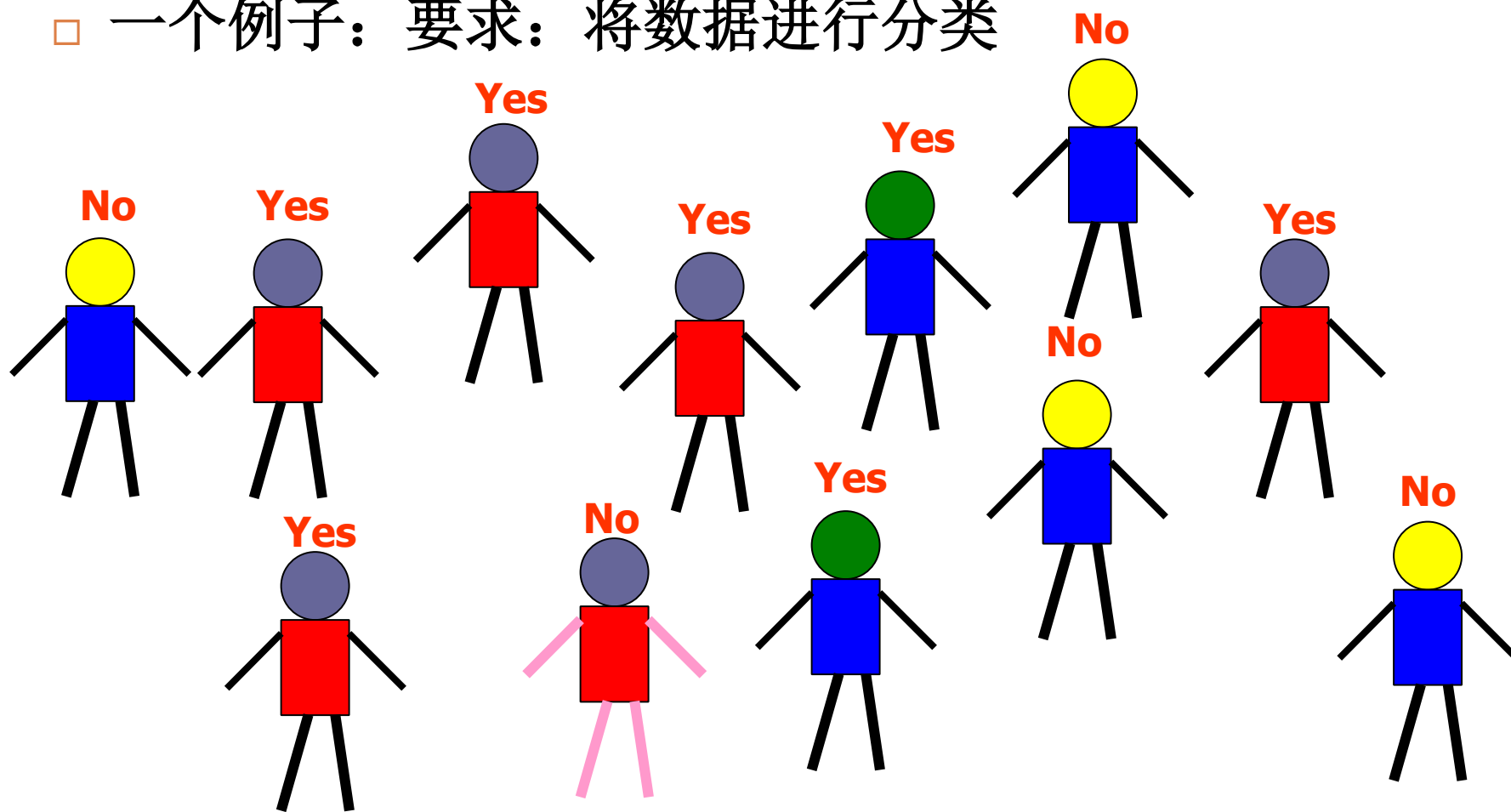
现阶段，人工智能的核心是对大数据进行的**特征抽取**与**机器学习算法**





数据+分类学习的方法

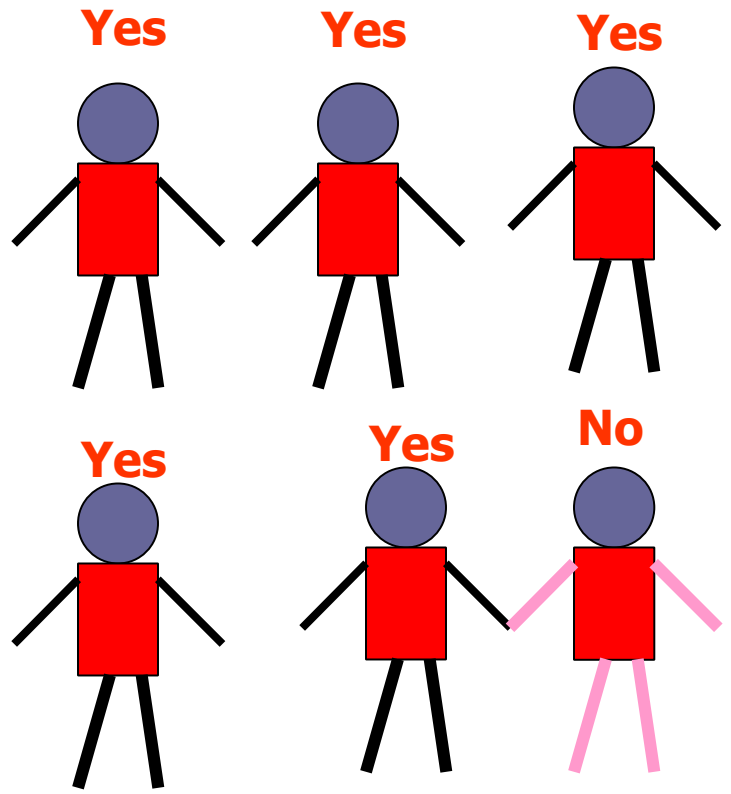
□ 一个例子：要求：将数据进行分类



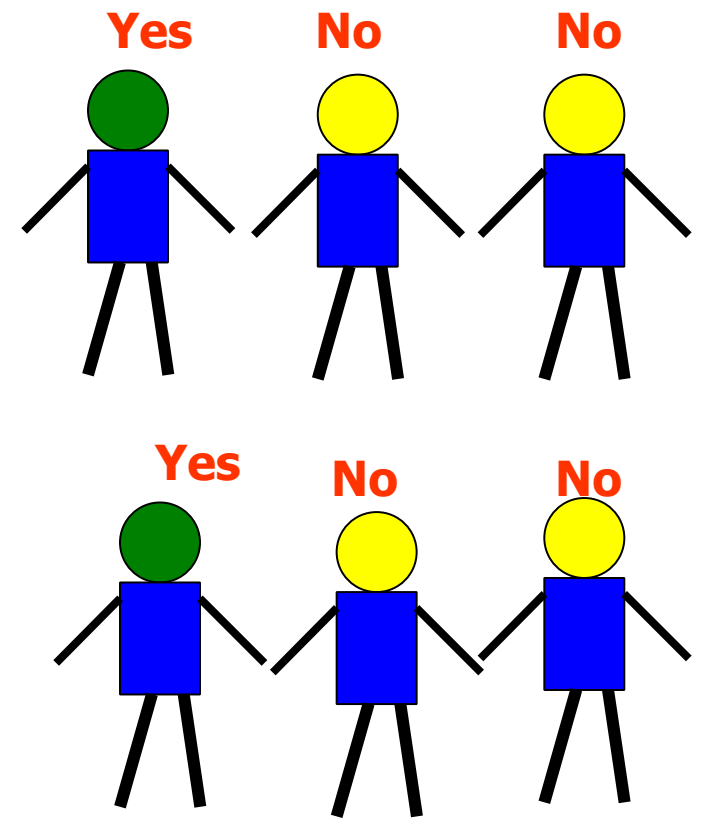


数据+分类学习的方法

躯干：红色



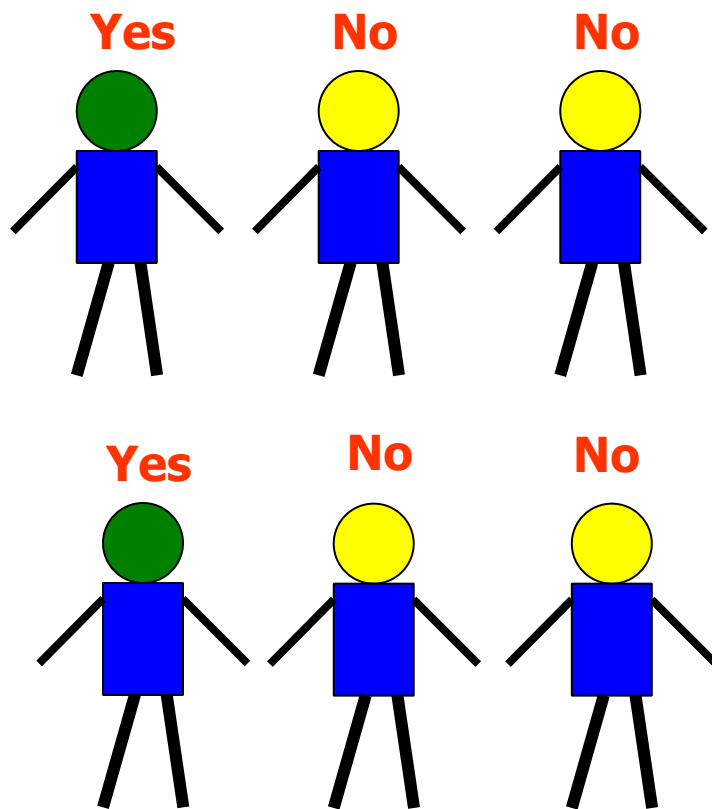
躯干：蓝色





数据+分类学习的方法

躯干：蓝色



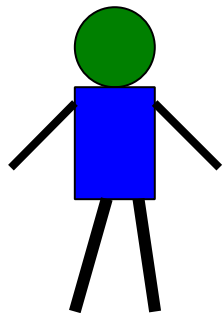


数据+分类学习的方法

躯干：蓝色

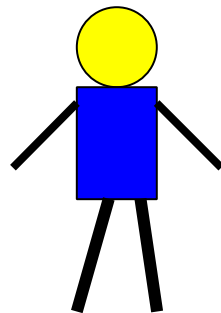
头：绿色

Yes

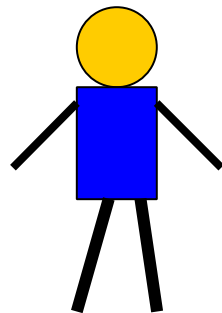


头：黄色

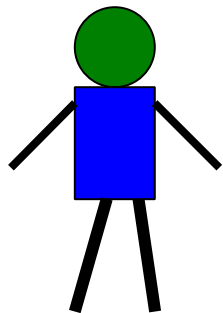
No



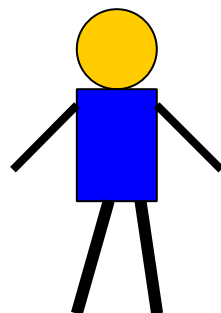
No



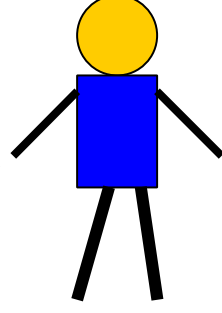
Yes



No



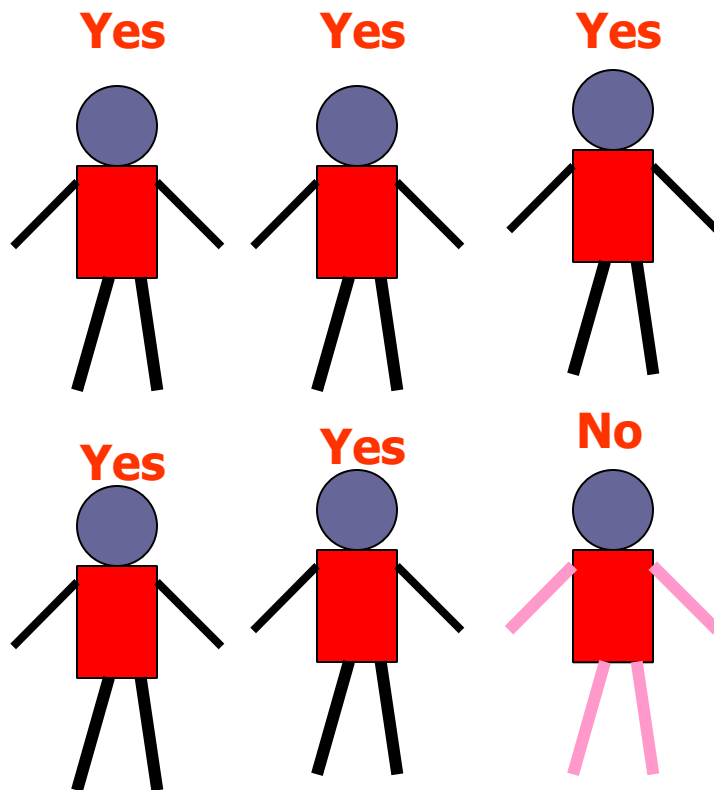
No





数据+分类学习的方法

躯干：红色



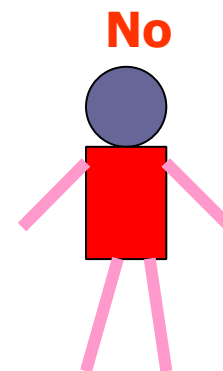
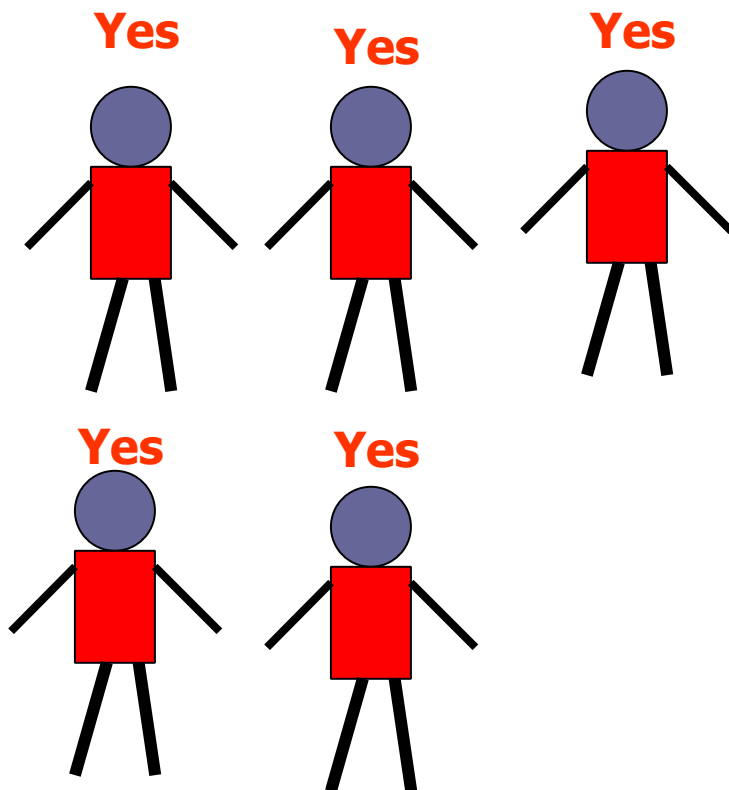


数据+分类学习的方法

躯干：红色

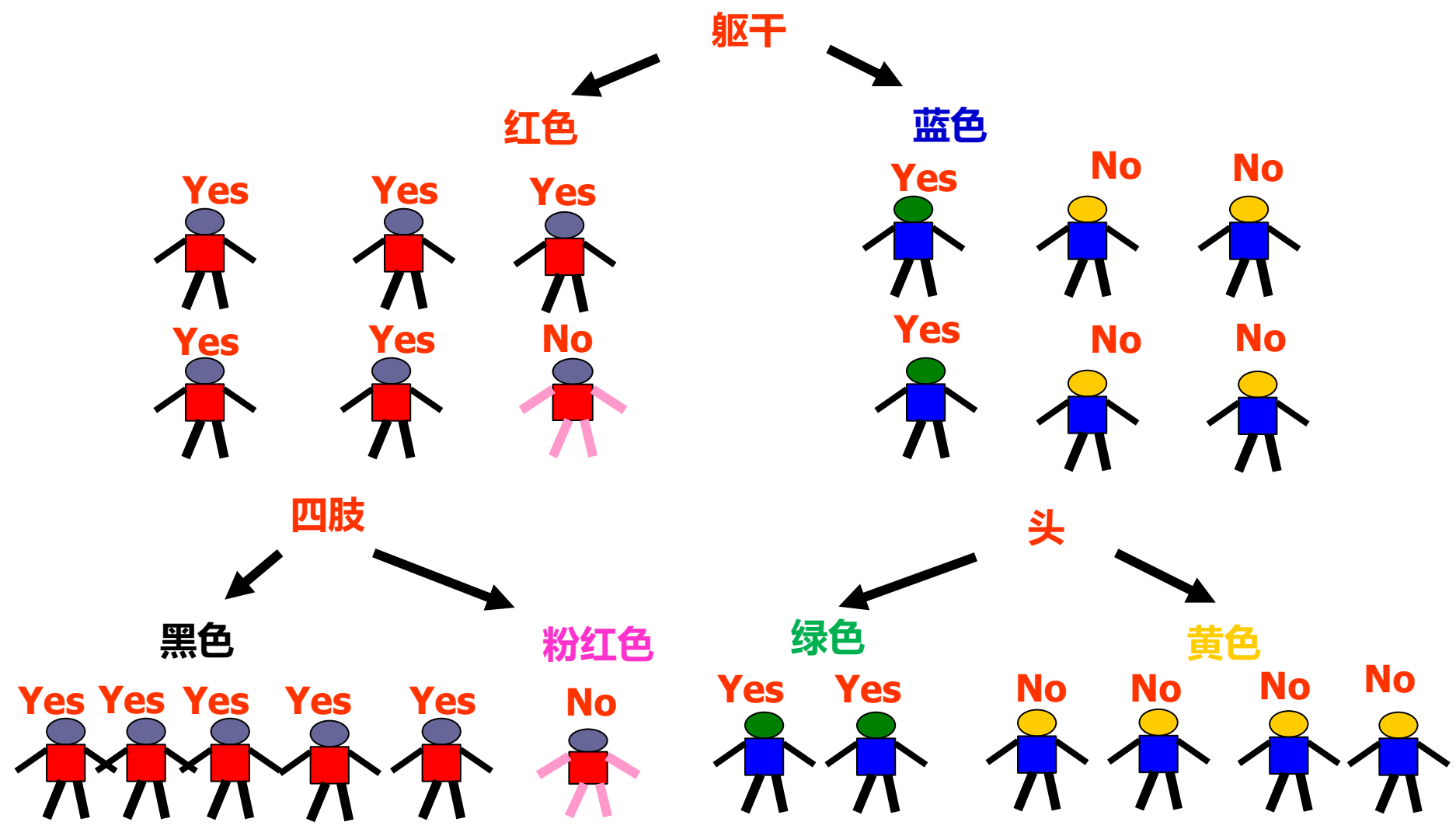
四肢：黑色

四肢：粉红色





数据+分类学习的方法



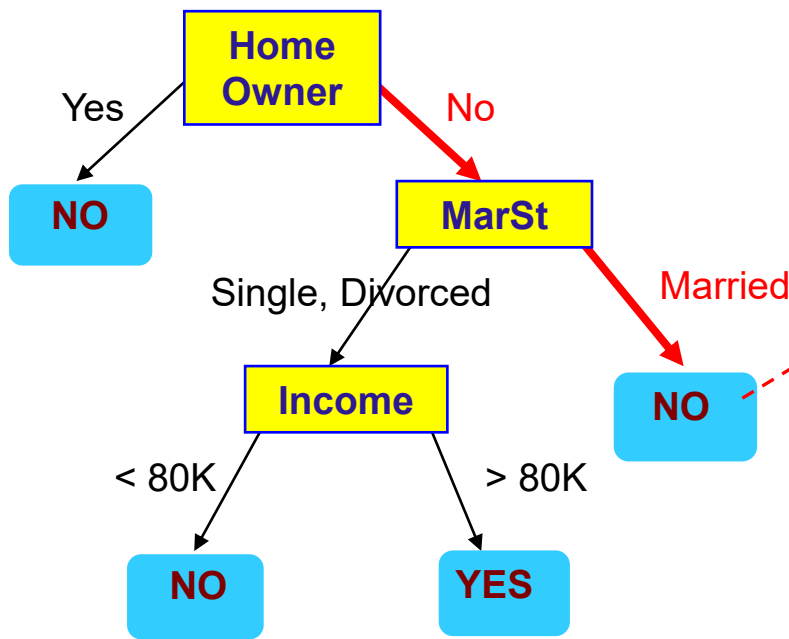


数据+分类学习的方法

□ 决策树——使用模型对测试数据分类

Test Data

Home Owner	Marital Status	Annual Income	Defaulted Borrower
No	Married	80K	?



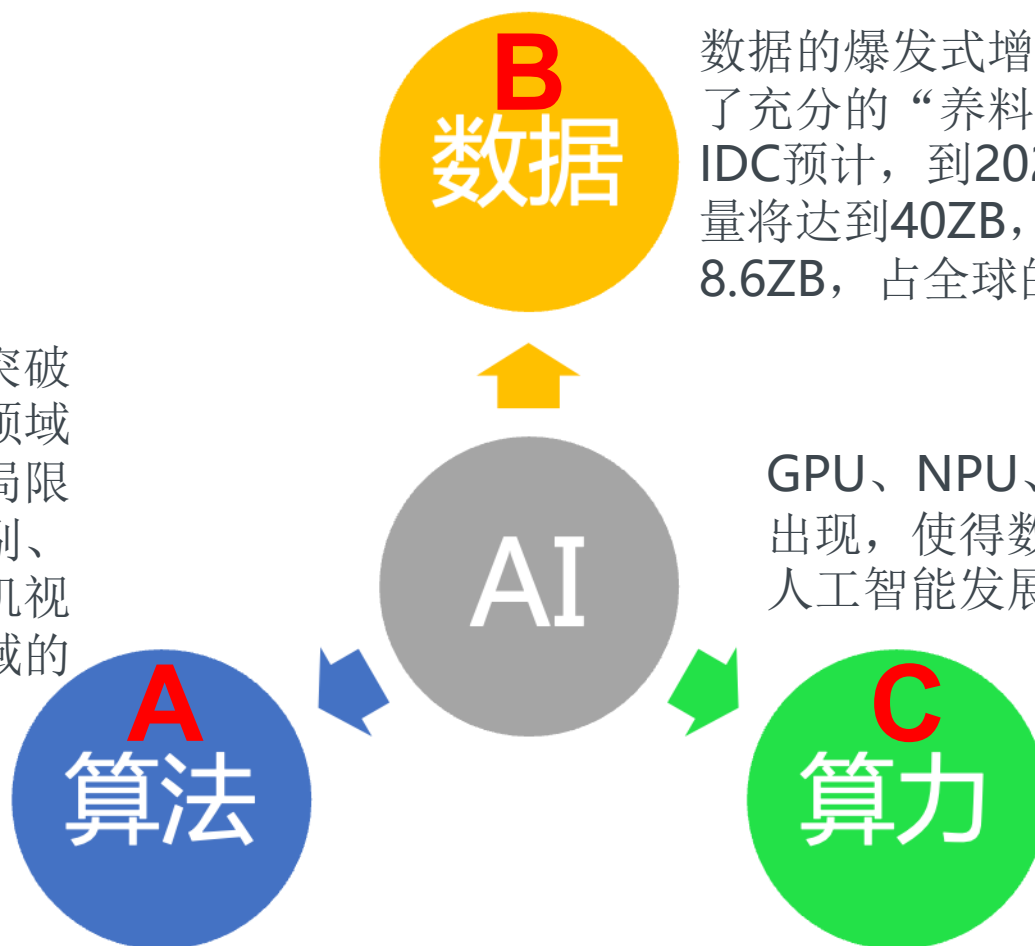
Assign Defaulted to
"No"



数据科学基础

□ 大数据的研究—数据驱动人工智能成熟与商业化

深度学习的出现突破了过去机器学习领域浅层学习算法的局限，颠覆了语音识别、语义理解、计算机视觉等基础应用领域的算法设计思路



数据的爆发式增长为人工智能提供了充分的“养料”，市场调研机构IDC预计，到2020年，全球数据总量将达到40ZB，我国数据量将达到8.6ZB，占全球的21%左右。

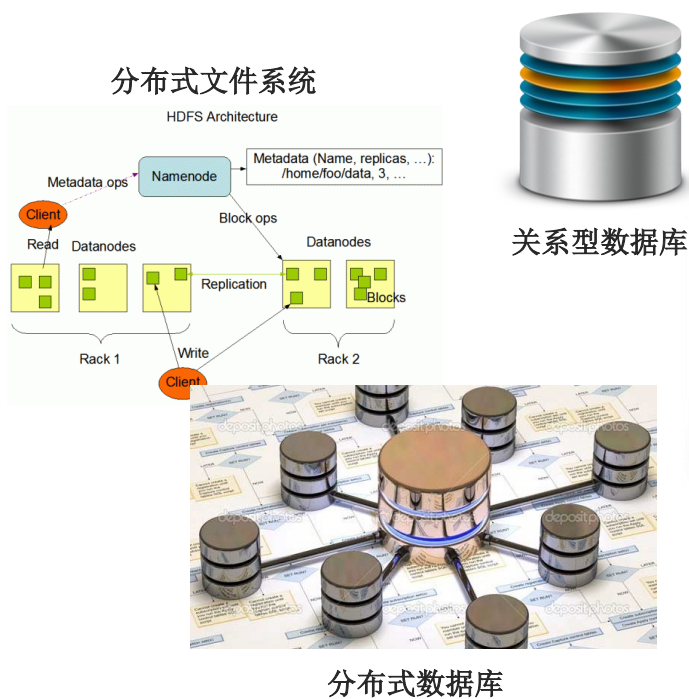
GPU、NPU、FPGA等专用芯片的出现，使得数据处理速度不再成为人工智能发展的瓶颈



数据科学基础

101

- 包括高效的CPU/GPU、云计算、AI芯片、多机集群并行化处理等技术手段
 - 数据处理和智能计算任务的多元化促使相关软件的多样化



9/22/2025



数据科学基础

102

- 包括高效的CPU/GPU、云计算、AI芯片、多机集群并行化处理等技术手段



- 云计算: EPYC (霄龙) 处理器; Project 47服务器



- 新一代处理器架构**VOLTA**: 新一代NVIDIA NVLink高速互联技术
- 云计算: Tesla V100 GPU 加速器; DGX-2 全球最大GPU; GPU云平台
- 机器人: Jetson Xavier 机器人专用AI芯片
- GPU工作站: 基于Volta架构的GV100
- 自动驾驶: Drive Xavier 首个自动驾驶处理器



- 云计算: 至强可扩展处理Purley; Xeon+FPGA (云端/设备端低功耗性能计算); Xeon Phi+Nervana (云端高性能计算)
- 无人驾驶: EyeQ 4/EyeQ 5 SoC
- 边缘计算: Myriad X VPU. Movidius Myriad X 视觉处理单元 (VPU), 是全球首个配备专用神经网络计算引擎的SoC
- 自学习神经元芯片: Loihi



- 云计算: 可重配置加速堆栈 (FPGA-Accelerator Stack)
- 设备端: reVISION 加速堆栈



- CPU架构: Cortex-A76
- GPU架构: Mali G76



- 云计算: TPU 3.0; Cloud TPU
- 移动端: Pixel VisualCore



- 移动端: 骁龙855处理器
- 智能驾驶: C-V2X芯片组



- 移动端: A12 芯片



- 移动端: 麒麟980芯片



- 跨界处理器: i.MX RT1060



数据科学基础

103

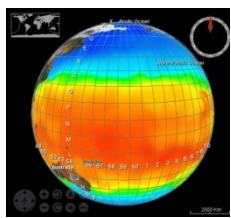


存储（如硬盘、数据库）

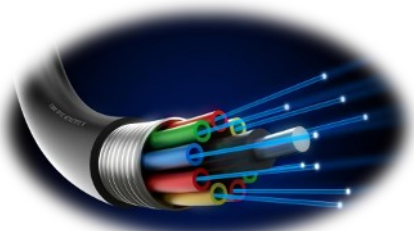


$$\min_X f(X) + \lambda \cdot \text{rank}(X)$$

分析、挖掘和学习



可视化



收集、传输



数据安全与个人隐私



生产、记录



基本程序与算法



计算（平台与架构等）

.....