



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](mailto:bigdata_2025@163.com)

9/22/2025



# 数据科学基础

3

- 国内外优秀高校信息与传播学院开展“数据分析”等专业课程，以培养科教融合、协同育人人才培养模式的新模式

地区	学校	院系	课程	学分	类别	年级
中国	中国人民大学	新闻学院	统计数据分析	2	通识教育	大一、大二
中国	复旦大学	新闻学院	数据分析与信息可视化	3	专业基础课	大三
中国	山东大学	新闻传播学院	大数据推断基础（双语）	2	专业选修课	大一
中国	华中科技大学	新闻与信息传播学院	人工智能与数据科学	2	学科基础课	大二
中国	中国科学技术大学	计算机科学与技术学院	数据科学导论	2	核心通识	大二及以上
美国	University of Illinois at Urbana, Champaign	Journalism	Data Science Discovery	4	专业课	undergraduate
美国	Stanford University	Humanities & Sciences	Data Science 101	5	专业课	undergraduate
美国	The University of Chicago	Humanities	Data Analysis I: Introduction to Statistics	100-units	专业课	undergraduate
美国	Carnegie Mellon University	Humanities and Arts	Methods for Statistical & Data Science	9units	专业课	undergraduate
美国	University of Florida	Journalism and Communications	Advanced Data Journalism	3	专业课	undergraduate



# 课程基本情况

4

- 科技传播系（25系）新增开设“网络与新媒体”专业
  - 以原有的传播学专业为基础，借助理工科培养的优势，初步形成了“文理管交叉、产学研一体”的复合型人才培养模式
  - 致力于培养具有深厚的理工知识背景、扎实的网络传播和新媒体应用技能、卓越的新媒体策划与运营理念、出色的中英文表达能力及国际化视野的高级专门人才
- 开展《新媒体大数据分析》课程建设，更好地适应该专业人才培养目标
  - 面向科技传播系高年级本科学生开放，将通过理论教学、案例分析与课题实践相结合的方式，全景介绍大数据科学分析的基础概念、基本方法、应用实践与前沿概况
  - 使学生树立运用大数据分析和相关算法工具解决实际问题的基本思路，并初步掌握新媒体大数据分析中解决实际问题应用问题的能力。



# 课程基本情况

5

通过开展《新媒体大数据分析》课程建设，希望达到以下本科教学培养目标：

## □ 知识目标

- 了解数据科学的基本概念、思路
- 初步掌握数据处理、数据分析、结果可视化等入门级知识
- 熟悉新媒体大数据分析前沿概况和经典应用场景

## □ 能力目标

- 树立运用数据和相关算法工具解决实际问题的基本思路
- 初步掌握用于数据分析手段解决实际应用问题的能力，即如何使用科学的方法研究和应用数据（包括用数据的方法研究科学）

## □ 素质目标

- 培养正确的“**数据价值观**”，增强从数据科学的角度认识社会、理解社会、服务社会的意识



# 课程基本情况

6

《新媒体大数据分析》专业基础课程共**40学时理论+20学时实验**：

## □ 第一部分：数据科学基础

- 重点介绍**课程定位**，概述数据科学与大数据分析的发展历程，和重点案例

## □ 第二部分：大数据分析基础

- 介绍大数据分析的生命周期流程，培养学生认识数据的特点，**掌握数据分析基本原理和方法**，掌握数据特征的提取和评价方法等。内容包括：
  1. 数据采集与存储
  2. 数据预处理与特征工程
  3. 数据统计分析



# 建设内容

7

## □ 第三部分：大数据分析算法

- 介绍**大数据分析的基本算法**。培养学生掌握分类、聚类、关联规则挖掘等基本分析算法：
  1. 大数据分析 with 分类
  2. 大数据分析 with 聚类
  3. 大数据分析 with 关联规则挖掘

## □ 第四部分：新媒体大数据分析应用

- 围绕新媒体数据，介绍大数据分析成功案例。培养学生掌握对不同类型的**新媒体数据****进行实际分析和应用的能力**，内容拟包括：
  1. 文本大数据分析案例
  2. 图片与视频大数据分析案例
  3. 网络大数据分析案例

## □ 实验介绍：

- **Python**基础、分类与聚类实验等



# 课程几点说明

9

- 2025年课程（NNM2012）
  - 主要面向科技传播系“网络与新媒体专业”高年级本科生
  - 以及感兴趣的工科方向学生
  
- 同类型课程：**学分不重复认定**
  - 大数据学院：《数据分析与实践》DS3001
  - 计算机学院：《数据科学导论》CS4023
  
- 关于其他专业的学分认定请咨询本专业教秘



# 课程几点说明

10

- **2025年课程（NNM2012）**
  - 面向科技传播系“网络与新媒体专业”高年级本科生
  
- **时间: (1-18周)周一下午6,7节: 14:00-15:35**
- **地点: 东区2303**



# 课程要求与考核方式

11

- 课程目标：用科学的方法研究和应用数据
- 课程要求
  - 1. 课堂出勤，日常作业
  - 2. 文献调研报告（独立完成）
  - 3. 实验任务（独立完成，需要编程），之后布置给大家
  - 4. 课程交流与课程汇报
- 考核方式
  - 课堂+作业（30%）+调研报告（30%）+实验报告（40%）



# 联系方式

12

- 课程主页：  
<http://staff.ustc.edu.cn/~huangzhy/Course/NM2025.html>
- 授课教师：黄振亚，朱孟潇，张凯
- 助教：
  - 齐畅 朱嘉骏
  - 课程邮箱：[bigdata\\_2025@163.com](mailto:bigdata_2025@163.com)
- QQ群：1029150801



25届新媒体大数...

群号: 1029150801





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

13

# 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

[bigdata\\_2025@163.com](mailto:bigdata_2025@163.com)

9/22/2025



# 课程目标

14

- 全面了解数据科学的基础知识
  - 包括数据分析的常用技术、发展前沿和应用案例
  - 了解数据的“能”与“不能”
- 树立数据科学的基本思路
- 初步掌握使用数据分析手段解决实际应用问题的能力

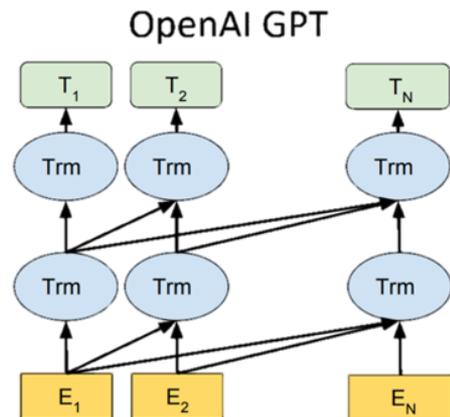
## 用科学的方法研究和应用数据

选修新媒体大数据分析课程的同学将来可能从事不同行业的科学研究、技术开发、产品管理等，希望这门课程带给你们的是终身受用的数据思维和创新力。



# 数据科学基础

- ChatGPT: 大数据催生人工智能新浪潮
  - 参数量从1.17亿增加到1750亿
  - 数据量从5GB增加到45TB
    - 96%以上是英文，其它20个语种不到4%



## GPT

无监督预训练，有监督微调

**5G**文本数据 | **1.17亿**模型参数

在**9/12**任务上最优，包括问答、语义相似度、文本分类

2018

## GPT-2

多任务、零样本学习 (**zero-shot**)

**40G**文本数据 | **15亿**模型参数

在**7/8**任务上最优，包括阅读理解、翻译、问答

2019

## GPT-3

小样本学习 (**few-shot**)

**45T**文本数据 | **1750亿**模型参数

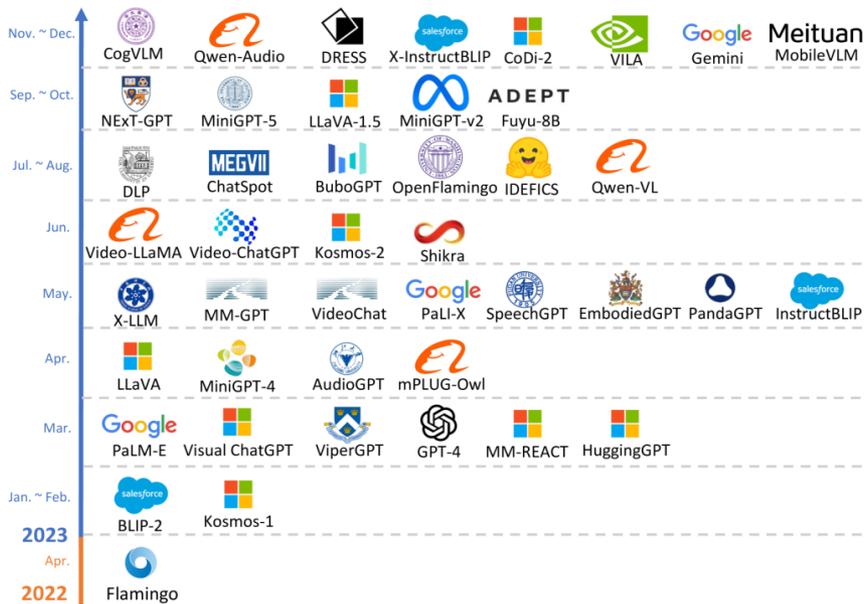
在阅读理解任务上超越当时所有**zero-shot**模型

2020



# 数据科学基础

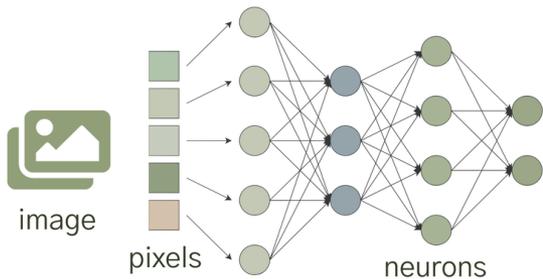
- 大数据催生人工智能新浪潮- 多模态模型（GPT-5、Sora 等等）-2023至今
  - 任务：多模态对话、多模态内容生成
  - 数据量：GPT-4：45TB文本数据增加到1PB多模态数据
  - 参数量：GPT-4：1750亿增加到1.76万亿参数



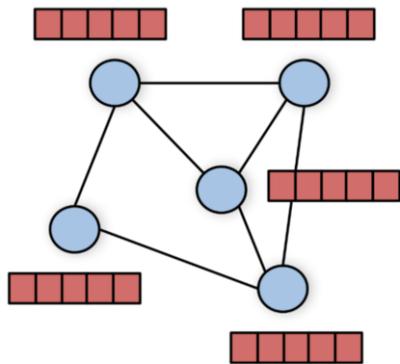


# 数据科学基础

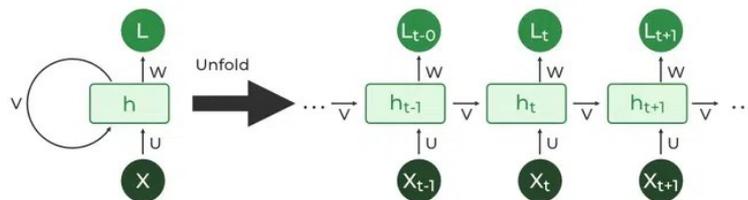
- 以模型为中心(model-centric)的数据科学技术
  - 围绕目标任务特性，设计不同模型结构



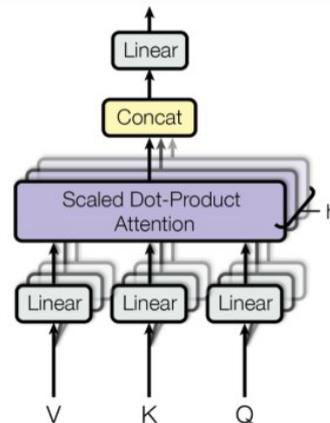
卷积神经网络CNN



图神经网络GNN



循环神经网络RNN

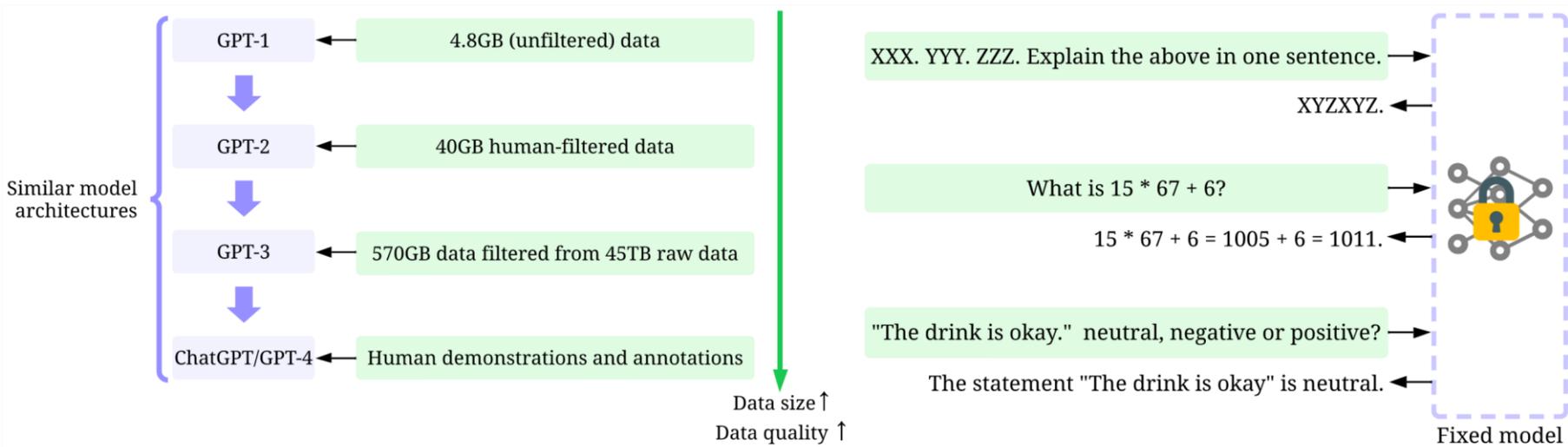


自注意力神经网络Transformer



# 数据科学基础

- 人工智能逐渐从以模型为中心过渡到以数据为中心
  - GPT成功的数据基石**: GPT进化中，模型结构保持相似，训练数据的规模、质量得到极大提升
  - 数据导向的模型应用**: 当模型足够强大，仅仅需要修改推理数据（提示工程）便可完成目标任务

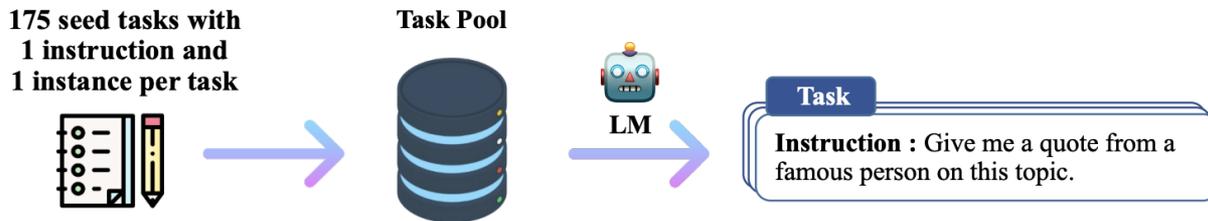




# 数据科学基础

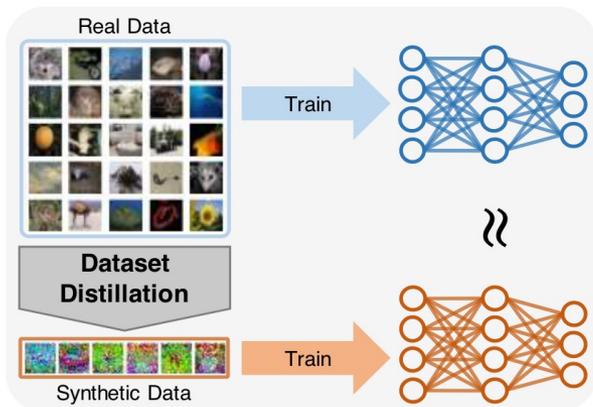
## 以数据为中心(data-centric)的数据科学技术

### 增加数据数量



数据生成

### 改善数据质量



数据蒸馏



数据选择



# 数据科学基础

## □ 数据

- 从计算机科学的角度，所有能够输入到计算机并被计算机程序处理的符号的总称
- 新时代的生产要素（十四五）
- “人-机-物”三元融合，世界已经成为数据化的世界



当文字成为数据



当方位成为数据

一切事物的数据化



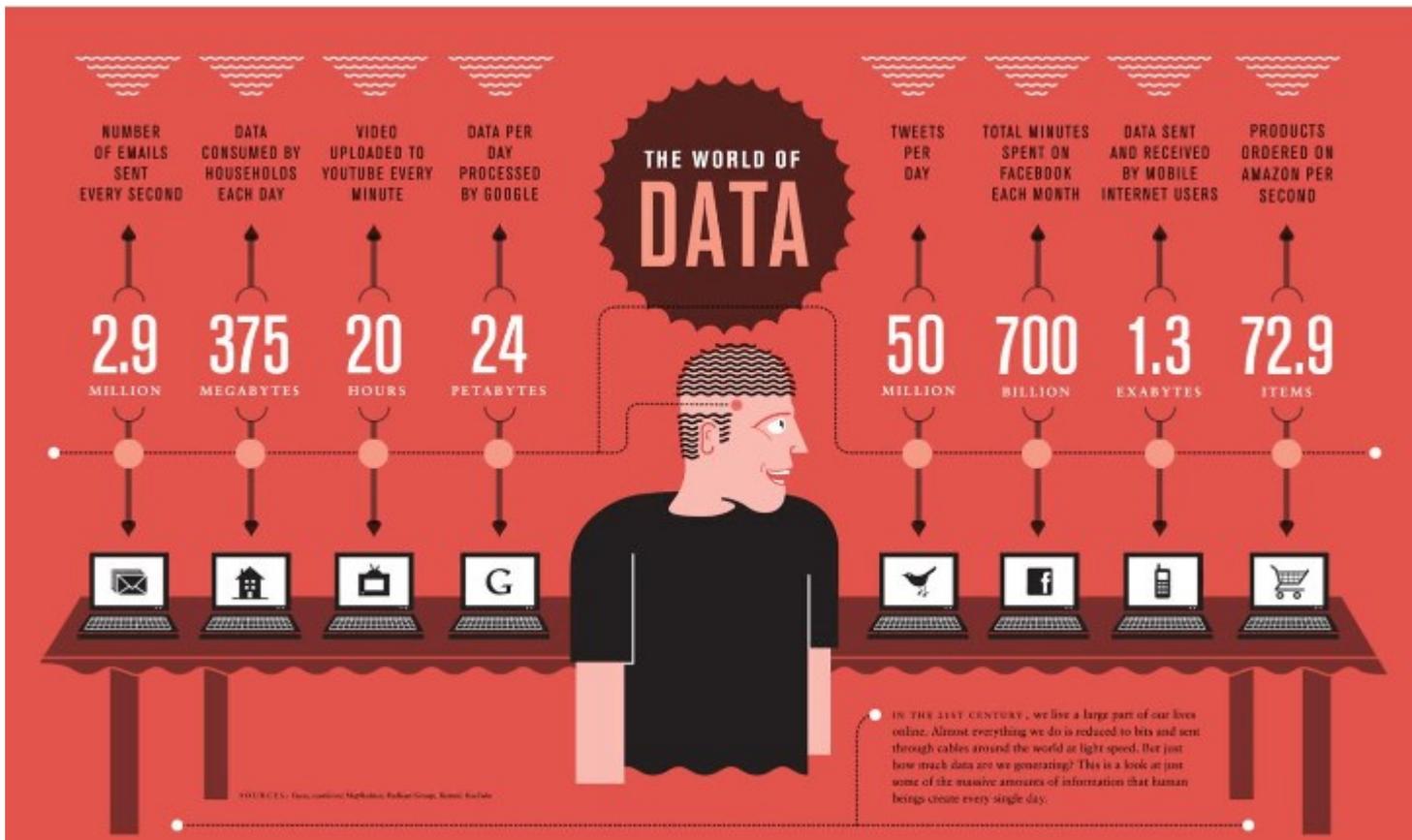
当沟通成为数据



# 数据科学基础

21

- 我们生活在数据中，所有人都在制造和分享数据

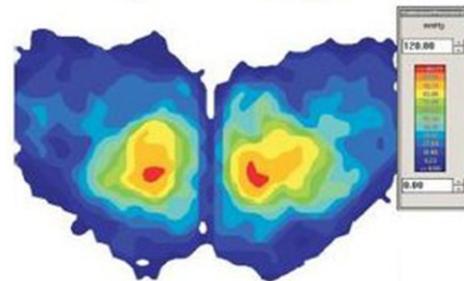
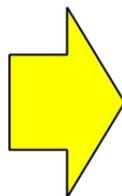




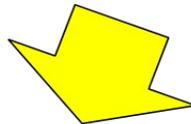
# 数据科学基础

- 例子：最不可能的地方获得数据
  - 当一个人坐着的时候，他的身形、姿势和重量分布都可以量化和数据化。

在汽车座椅下部安装  
360个压力传感器



测量人对椅子施加的压力，  
用0~256的数值量化



- 把人体屁股特征转化成了数据，产生独属于每个乘坐者的精确数据资料。
- 汽车可以准确的识别乘坐者的身份：**汽车防盗系统**



# 数据科学基础

23

## □ 大数据的提出



从2008年9月,《Nature》杂志首次出版一期大数据专刊,科学家们提出“大数据真正重要的是新用途和新见解,而非数据本身”



# 数据科学基础

## □ 铺天盖地的“大数据”

**大数据新闻1亿篇**

Baidu 百度 大数据

Q 网页 图片 视频 资讯 贴吧 文库 知道 地图 采购 更多

百度为您找到相关结果约100,000,000个

搜索智能精选

大数据\_百度百科

“大数据(big data), 或称巨量资料, 指的是所涉及的数据量规模巨大到无法透过主流软件工具, 在合理时间内达到获取、管理、处理, 并整理成为帮助企业经营决策更积极目的的资讯。在维克托·迈尔-舍恩伯格及肯尼

大数据 - 官方网站 官方

大数据 双月刊 ISSN: 2096-0271



大数据“完全占领”了互联网和IT领域之后，开始进入各行各业，形成了政府大数据、教育大数据、医疗大数据、交通大数据、金融大数据、保险大数据、公安大数据、法院大数据、旅游大数据、.....



# 数据科学基础

## 大数据有多大？

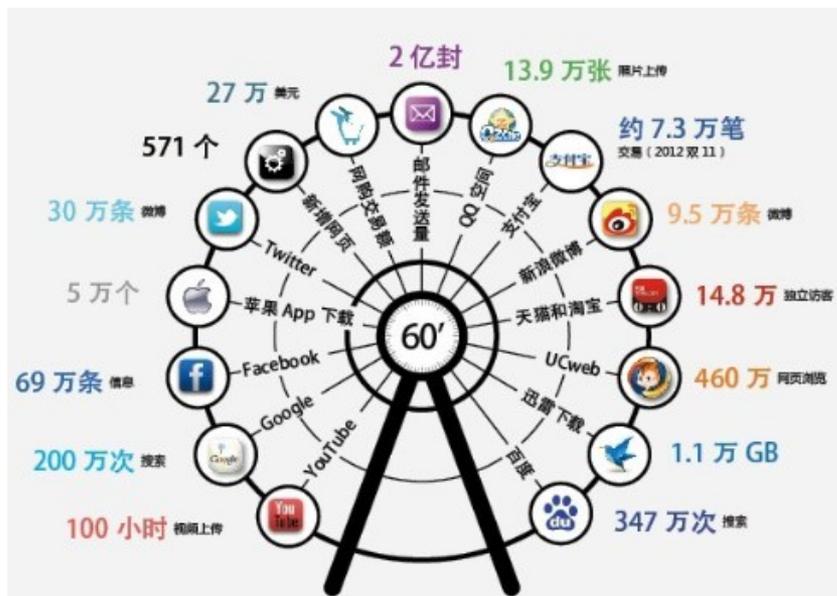
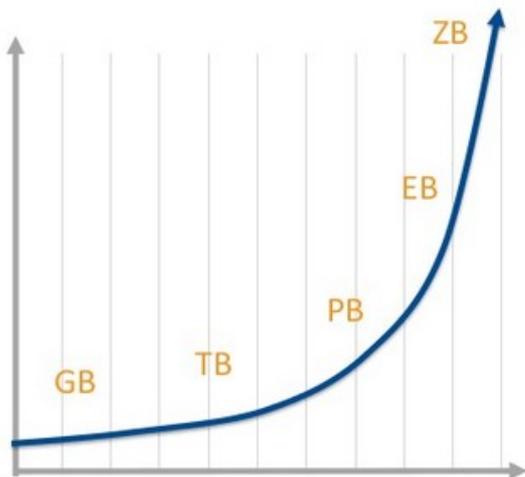
PB是大数据层次的临界点

◆ 数据量已到**ZB**等级

KB->MB->GB->TB->**PB**->EB->ZB->YB->NB->DB

PB以上级别的数据，最有效的传输方式是空运，而不是网络

◆ 大数据不仅仅只是量大！

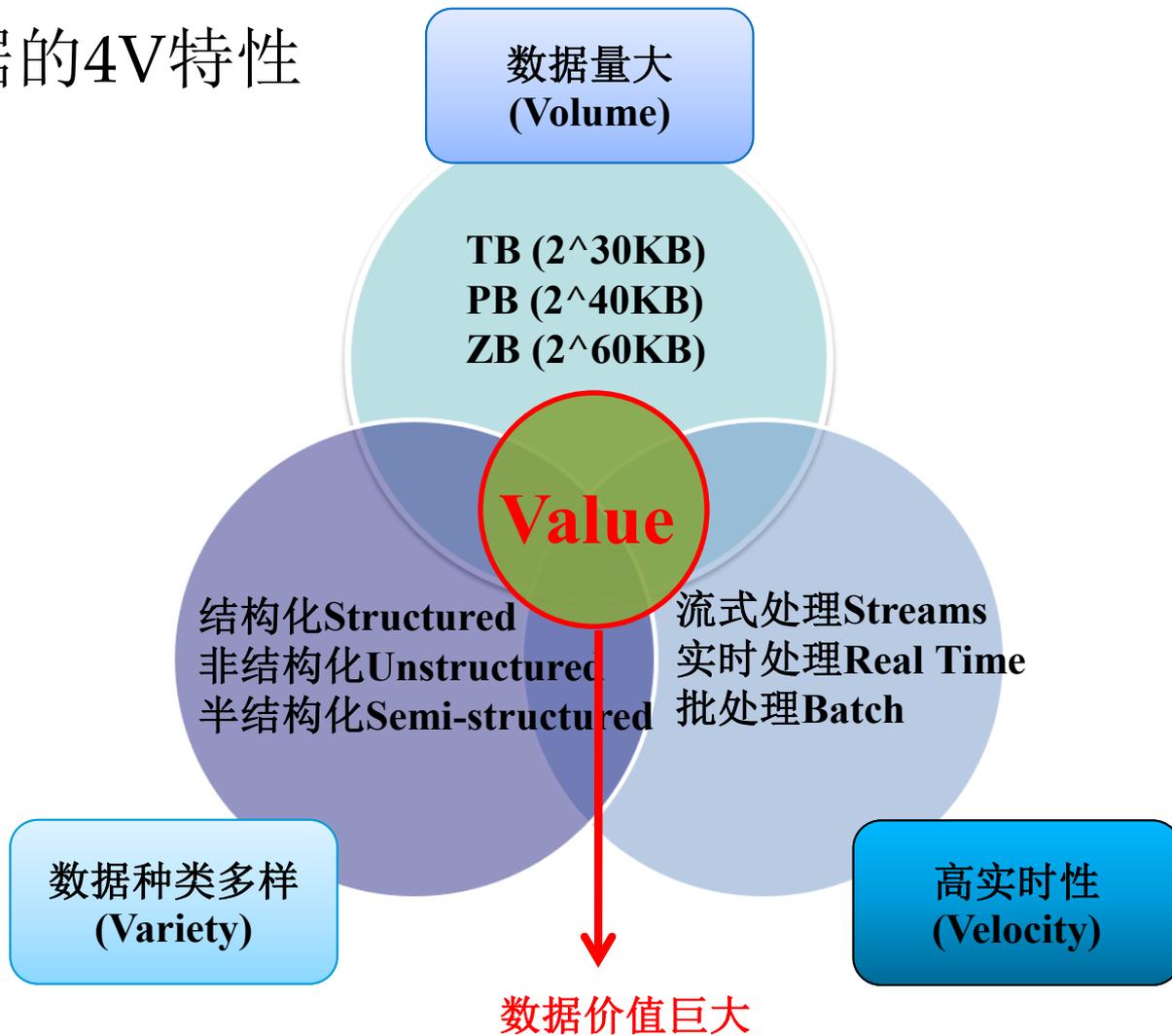


60秒，我们能产生多少数据？



# 数据科学基础

## 大数据的4V特性





# 数据科学基础

29

## □ 大数据---Volume(数据量巨大)

阿里所保有的、经过清洗的历史数据已超过**100PB**。

——阿里数据仓库负责人七公（汪海）

百度现在的数据规模已经到了**EB级**，每天处理的数据量到了上百PB。

——百度大数据部总监薛正华

全球数据总量在2020年达到**60ZB**，预计2025年达到175ZB，我国数据占18%。

——IDC互联网数据中心

$$1 \text{ ZB} = 2^{10} \text{ EB} = 2^{20} \text{ PB} = 2^{30} \text{ TB} = 2^{40} \text{ GB}$$

- 1 ZB = 地球上沙粒的总量，1 EB = 4000个美国国会图书馆的藏书



# 数据科学基础

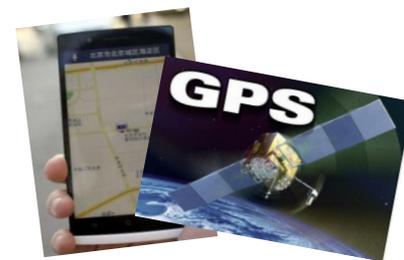
## 大数据--- Variety(数据类型多)

### 数据形式的多样:

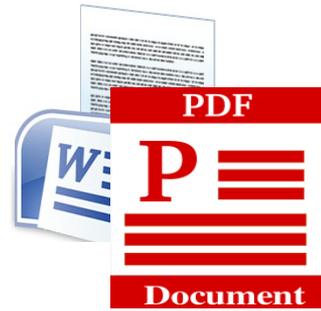
- 结构化数据, 半结构化数据, 非结构化数据
- 关系数据库数据、xml/JASON文档、音视频数据

### 数据来源的多样性:

- 不同的IT应用系统
- 各种设备 (手机、手环)
- 互联网、物联网
- 其它



时空数据



文本数据



图像数据



事务数据



视频数据



音频数据



# 数据科学基础

## 大数据--- Velocity(高实时性)

**1秒定律**: 对于大数据应用而言, 必须要在1秒钟内形成答案, 否则这些结果可能就是过时的、没有意义的

在百度输入关键字:  
“汽车维修”、“挖掘机 学习”

某在线电影网站



某IT业界资讯网站



例如用户在合肥某台PC上, 打开百度输入关键字片刻之后, 再打开其它网站, 就会看到相关的广告, 并且所推荐的是地理位置信息相关的 (合肥、安徽)



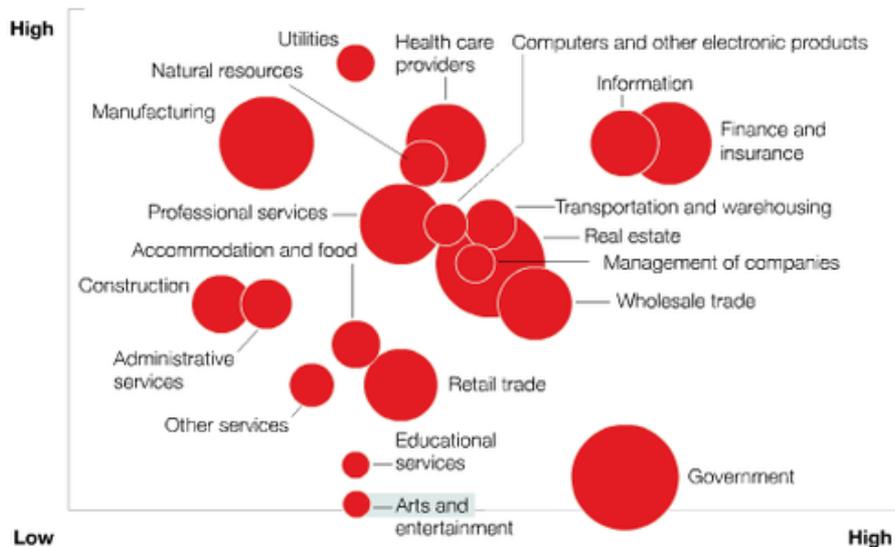
# 数据科学基础

## 大数据--- Value(价值巨大但价值密度低)

挖掘大数据中的价值类似沙里淘金，需要从海量数据中挖掘稀疏但珍贵的信息

所有产业都可以应用大数据产生价值

价值获取难度



潜在价值高低



● 各产业GDP占比 (以美国经济为例)

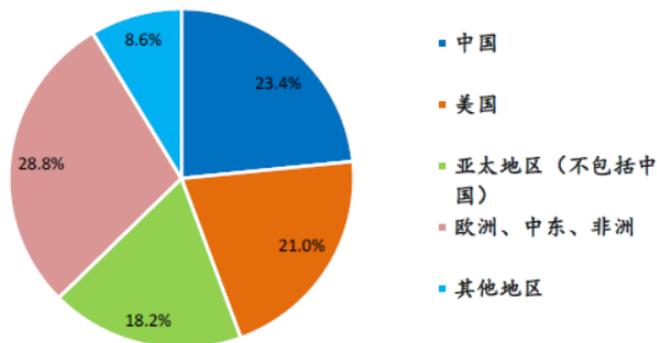
图：麦肯锡对各个行业从大数据中获得价值难易程度的分析



# 数据科学基础

33

- 我国是数据产生和应用最大的国家
  - 大数据是推动**数字经济**发展的关键生产要素
    - 2020年我国数字经济规模占GDP比重达到**38.6%**(39.2万亿元)
  - 大数据是重塑国家**竞争优势**的重大发展机遇
    - 2018年，中国数据圈占全球**23.4%**(7.6ZB)；预计2025年成为**最大数据圈**
  - 大数据是实现**治理能力现代化**的重要创新工具
    - 2021年我国数字政府行业市场规模有望达到**5000亿元**
  - 大数据是建设**数字中国**的关键创新动力
    - 2021年全国工业互联网产业增加规模预计突破**4万亿元**





# 数据科学基础

34

## □ 我国的大数据战略

- 2015年，十八届五中全会首次提出“**国家大数据战略**”，标志着大数据战略正式上升为国家战略。
- 2017年，《大数据产业发展规划(2016-2020年)》实施
- 2018年，在**十三届全国人大一次会议**中，国务院总理李克强在作政府工作报告时，三次提到大数据
- 2020年，《关于构建更加完善的要素市场化配置体制机制的意见》将大数据正式**列为新型生产要素**
- 2021年，《**“十四五”大数据产业发展规划**》：明确了大数据发展的四大任务
- 2024年，十七部门关于印发《**“数据要素×”三年行动计划（2024—2026年）**》的通知：构建**以数据为关键要素**的数字经济。



# 数据科学基础

35

## □ 十四五规划中的“大数据”

中共中央关于制定国民经济和社会发展第十四个五年规划的建议  
(2020年10月29日中国共产党第十九届中央委员会第五次全体会议)



- 加快构建全国一体化**大数据中心体系**，强化算力统筹智能调度，建设若干国家枢纽节点和**大数据中心集群**，建设**E级**和**10E级**超级计算中心
- 培育壮大人工智能、**大数据**、区块链、云计算、网络安全等新兴数字产业
- 推动**5G**、**大数据中心**等新兴领域能效提升
- 完善适用于**大数据**环境下的数据分类分级保护制度（大数据安全）



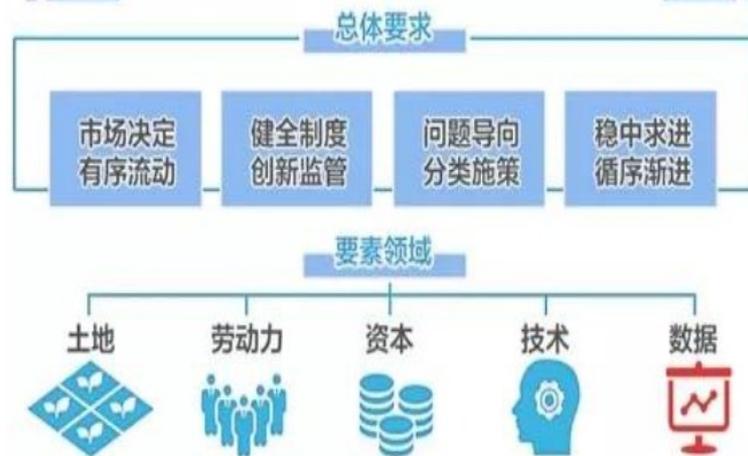
# 数据科学基础

## □ 十四五规划中的“大数据”

系统布局新型基础设施，加快第五代移动通信、工业互联网、**大数据中心**等建设。



推进土地、劳动力、资本、技术、**数据**等要素市场化改革。



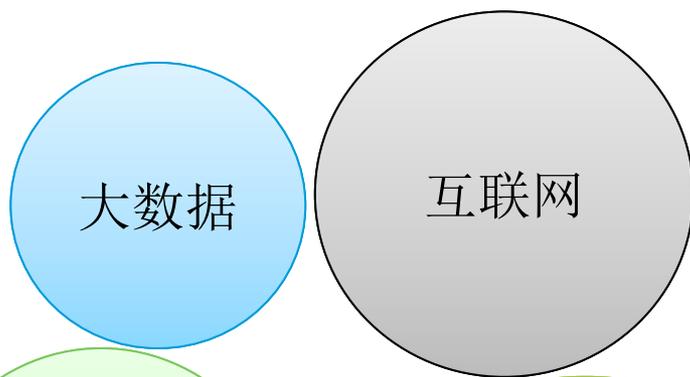


# 数据科学基础

## □ 十四五规划中的“大数据”

### 关键词

- 新时代生产要素
- 国家基础战略资源
- 大数据技术



- 互联网+服务
- 互联网+监管



- 新一代人工智能
- 量子信息
- 集成电路
- 脑科学
- 基因生物技术
- 深空深地深海和极地探测

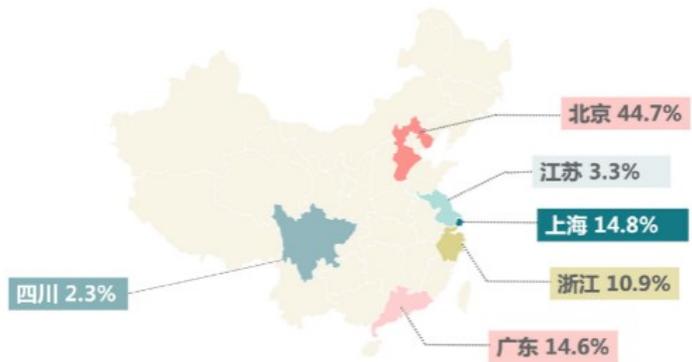
- 数字中国
- 数字经济
- 数字社会
- 数字政府
- 数字乡村
- 贸易数字化
- 产业数字化转型
- 数字化服务普惠应用



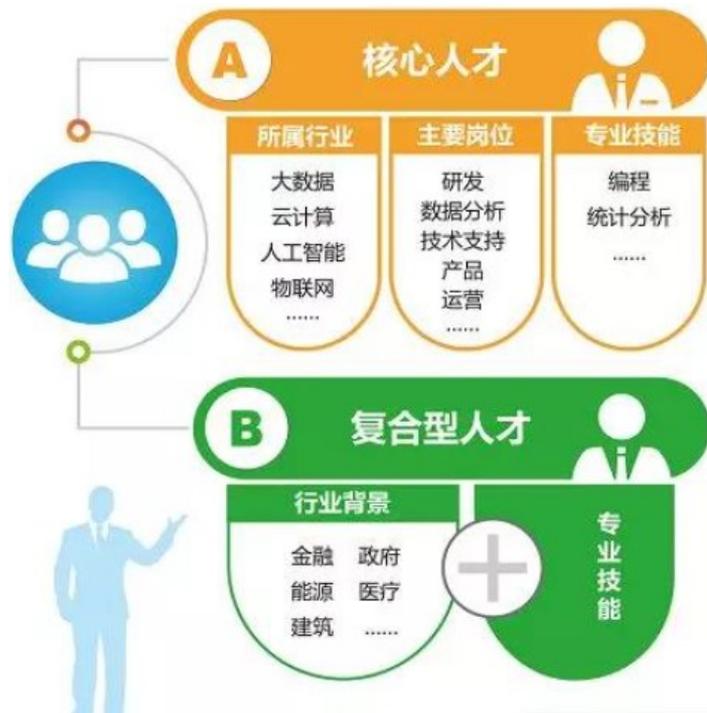
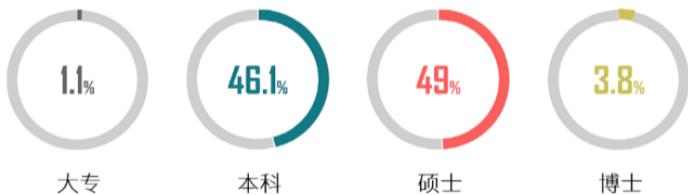
# 数据科学基础

## 大数据人才缺口—市场需求

- 市场对大数据人才的需求日益增加，供求关系不成正比，2025年人才缺口可达到230万
- 行业发展对大数据人才提出更高要求



公司对人才学历要求高，半数要求硕士及以上





# 数据科学基础

39

## □ 大数据人才缺口—专业需求

- **传统的工科人才**培养模式难以满足新兴产业的发展要求，亟需探索大数据工科人才培养的新途径和新模式
- **大数据工科人才**培养面临的挑战

校内外都存在针对大数据技术的广泛需求，但**专业划分偏窄、偏细**，使得现有大数据相关的技术体系和积累较为零散，给跨学科交叉创新与技术应用带了诸多限制

新工科时代涌现出大量的新兴经济行业（如共享经济、数字经济），使得**传统的专业设置略显陈旧**，培养的人才知识体系不能有效应对新兴行业的大数据人才需求

很多工科学生的理论知识扎实，但实验实训、人文基础不够，创新性受限，**向专业数据科学研究人才的转化率较低**



# 数据科学基础

- 国内外优秀高校信息与传播学院开展“数据分析”等专业课程，以培养科教融合、协同育人人才培养模式的新模式

地区	学校	院系	课程	学分	类别	年级
中国	中国人民大学	新闻学院	统计数据分析	2	通识教育	大一、大二
中国	复旦大学	新闻学院	数据分析与信息可视化	3	专业基础课	大三
中国	山东大学	新闻传播学院	大数据推断基础（双语）	2	专业选修课	大一
中国	华中科技大学	新闻与信息传播学院	人工智能与数据科学	2	学科基础课	大二
中国	中国科学技术大学	计算机科学与技术学院	数据科学导论	2	核心通识	大二及以上
美国	University of Illinois at Urbana, Champaign	Journalism	Data Science Discovery	4	专业课	undergraduate
美国	Stanford University	Humanities & Sciences	Data Science 101	5	专业课	undergraduate
美国	The University of Chicago	Humanities	Data Analysis I: Introduction to Statistics	100-units	专业课	undergraduate
美国	Carnegie Mellon University	Humanities and Arts	Methods for Statistical & Data Science	9units	专业课	undergraduate
美国	University of Florida	Journalism and Communications	Advanced Data Journalism	3	专业课	undergraduate



# 数据科学基础

41

- 科技传播系致力于培养的“科技媒体和科技传播”英才应具备以下素质



专业基础扎实，有卓越的数理基础、创意设计、管理能力



实践能力强，具有处理多媒体大数据的能力

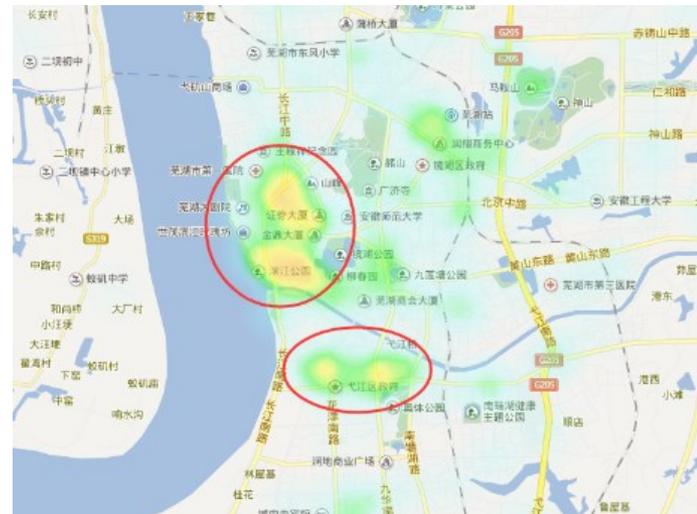
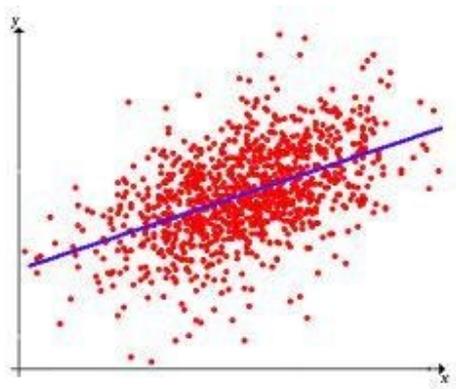
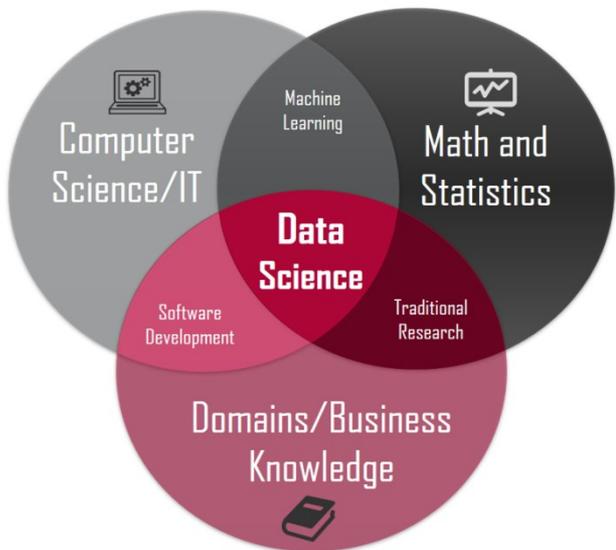


跨界能力强，能够解决新媒体行业的大数据应用问题



# 数据科学基础

- 大数据新工科人才需要具备以下素质
  - 学习理论知识：数学（基础）+ 计算机科学 + 交叉学科知识
  - 锻炼实践能力：编程、数据分析、数据可视化等
  - 培养跨界能力：应用场景、领域知识





# 数据科学基础

43

- 1. 理论基础扎实，能理解运用数据科学中的理论模型
- 数学是学习数据科学的基础
  - 数学与优化：数学分析的应用
    - 梯度下降
    - 搜索方向：负梯度方向、牛顿方向
    - 算法收敛性
  - 数学与聚类：线性代数的应用
    - 社交网络聚类的问题形式化
    - 线性代数知识求解
  - 数学与图卷积网络：傅里叶变换的应用
    - 图表征学习
    - 图上的傅里叶变换与卷积
  - 。 。 。



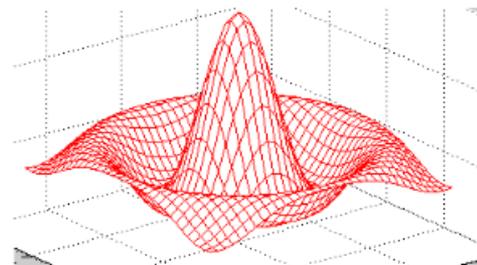
# 数据科学基础

- 1. 理论基础扎实，能理解运用数据科学中的理论模型
  - 数学是学习数据科学的基础
    - 数学与优化：数学分析的应用



模型学习 (机器学习)

找到合适的  $w$  , 使  $f(w, x)$  最接近  $D$



例如，线性回归损失函数

$$L(w) = \sum_{d_i \in D} f(w, x_i) - y_i$$

$$w = \operatorname{argmin}_w L(w)$$

优化方法



常见问题：  $\min_{x \in R^n} f(x)$

- 梯度下降
- 牛顿法/拟牛顿法
- . . .



# 数据科学基础

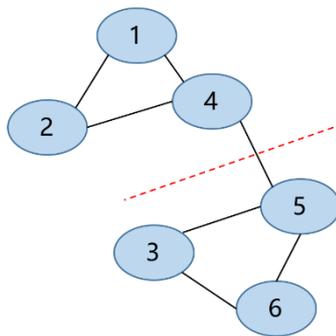
- 1. 理论基础扎实，能理解运用数据科学中的理论模型
  - 数学是学习数据科学的基础
    - 数学与聚类：线性代数的应用

社交网络划分：物以类聚，人以群分

无向图分割问题



问题转化



$$W = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

$$D = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

常用知识:

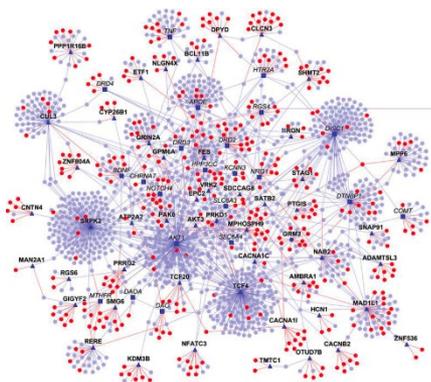
- 特征值分解
- 奇异值分解
- QR分解
- 矩阵求逆相关定理

- ✓ 将全校学生划分为不同班级？
- ✓ 将员工划分为不同公司？
- ✓ 将用户划分为不同追星圈？

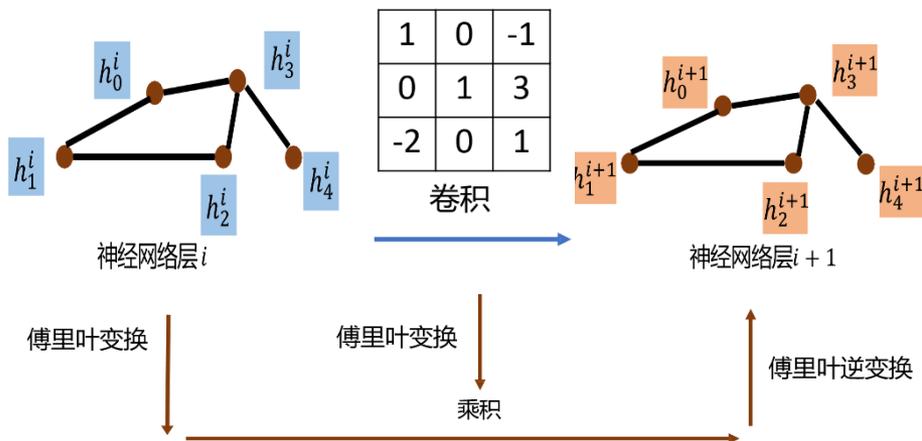
关键点：利用不同个体之间的联系

- 1. 理论基础扎实，能理解运用数据科学中的理论模型
  - 数学是学习数据科学的基础
    - 数学与图卷积网络：傅里叶变换的应用

图数据：分子图、社交网络等



图卷积网络：一类典型方法



✓ 典型任务

- ✓ 节点分类，关系（边）预测等
- ✓ 图分类，图属性预测，图生成

- ✓ Idea: 卷积定理：函数卷积的傅里叶变换是函数傅立叶变换的乘积
- ✓ 一般傅里叶变换 至 图上傅里叶变换



# 数据科学基础

47

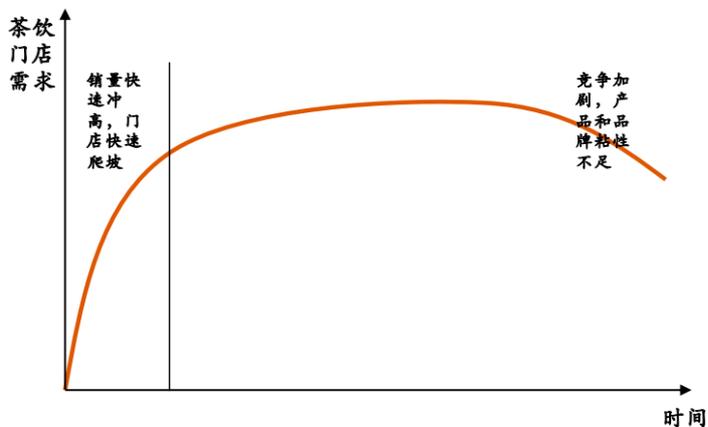
## 1. 创意设计与管理能力

运用数据分析结论帮助产品设计与营销管理

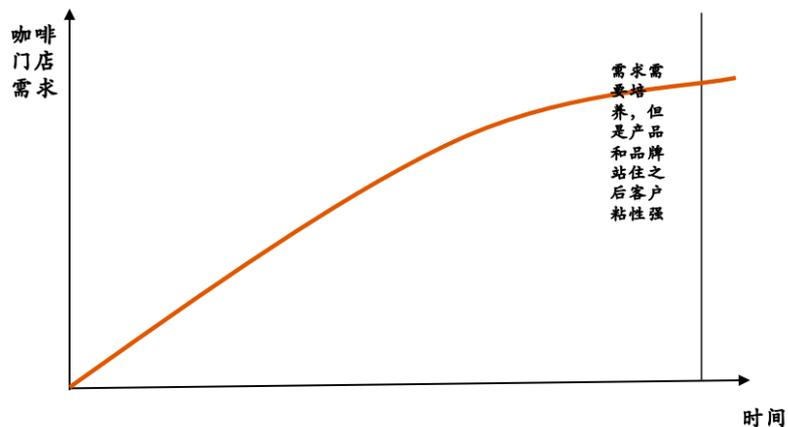
例：瑞幸咖啡

- 竞品分析，价格分析，用户分析，地理分析等
- 新媒体营销等

茶饮需求相对成熟，门店能够快速形成赚钱效应



咖啡需求需要培养，单店爬坡周期更长





# 数据科学基础

- 2. 实践能力强，具有处理大数据的能力
  - Python等编程技术，Web技术、数据库技术、可视化技术等
  - 常用工具使用：**如大模型**，可视化工具

```
1 def SumOfSubArray(larr, n, k): SumOfSubArray(arr, n, k):
2
3 Sum = 0Sum = 0
4 S = deque()= deque()
5 G = deque()= deque()
6 for i in range(k):for i in range(k):
7 while (len(S) > 0 and arr[S[-1]] >= arr[i]):while (len(S) > 0 and arr[S[-1]] >= arr[i]):
8 S.pop().pop()
9 while (len(G) > 0 and arr[G[-1]] <= arr[i]):while (len(G) > 0 and arr[G[-1]] <= arr[i]):
10 G.pop().pop()
11 G.append(i).append(i)
12 S.append(i).append(i)
13 for i in range(k, n):for i in range(k, n):
14 Sum += arr[S[0]] - arr[G[0]]Sum += arr[S[0]] + arr[G[0]]
15 while (len(S) > 0 and S[0] < i - k):while (len(S) > 0 and S[0] < i - k):
16 S.popleft().popleft()
17 while (len(G) > 0 and G[0] < i - k):while (len(G) > 0 and G[0] < i - k):
18 G.popleft().popleft()
19 while (len(S) > 0 and arr[S[-1]] >= arr[i]):while (len(S) > 0 and arr[S[-1]] >= arr[i]):
20 S.pop().pop()
21 while (len(G) > 0 and arr[G[-1]] <= arr[i]):while (len(G) > 0 and arr[G[-1]] <= arr[i]):
22 G.pop().pop()
23 G.append(i).append(i)
24 S.append(i).append(i)
25 Sum += arr[S[0]] - arr[G[0]]Sum += arr[S[0]] + arr[G[0]]
26 return Sumreturn Sum
27
```





# 数据科学基础

## 3. 跨界能力强，能够解决特定行业的大数据应用问题

