



# 数据科学基础 I (Matlab)

— 东北大学 —





## 人类文明形态



农业文明



工业文明



信息文明





## 信息文明



### 什么是信息？

- .....
- 美国数学家诺伯特·维纳（Norbert Wiener, 1894~1964）（控制论的奠基人）：‘信息就是信息，既不是物质也不是能量’
- 美国电信工程师香农（Claude Elwood Shannon, 1916~2001）（信息论的奠基人）：信息是有别于物质与能量的第三种东西，是对事物运动状态或存在方式的不确定性的描述。
- .....



## DIKW金字塔模型

对未来的预测、解释、发现的能力

智慧Wisdom

知识

Knowledge

信息

Information

数据Data

加工处理后的有逻辑的数据，反映了现实世界的现象

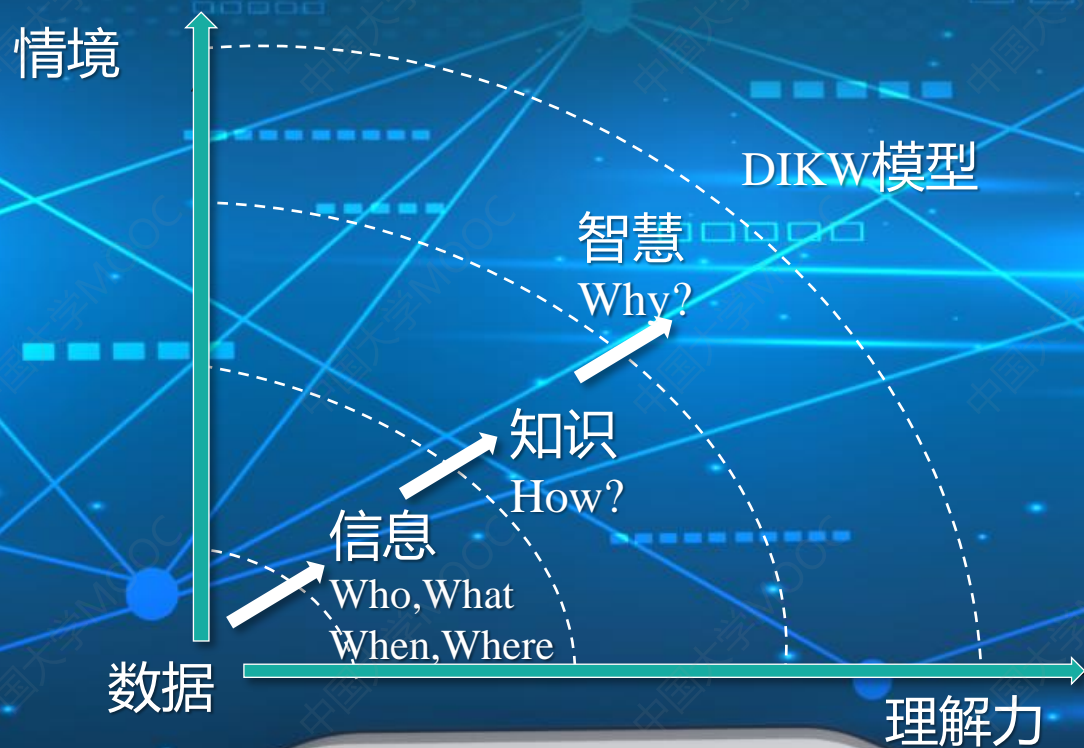
提炼信息之间的联系，发现的共性规律、模式、模型、理论、方法等

现实世界的记录，是原始素材





## 换个视角





## 数据的形式



数字



文本



音频



图像



视频



.....





# 什么是数据科学？



## 数据科学

是一门通过系统性研究来获取与数据相关的知识体系的科学。

通过对数据的研究，为自然科学和社会科学的研究提供一种新的方法，其目的在于揭示自然界和人类行为的现象和规律。



## 数据科学的提出



**1974年**

丹麦计算机科学家Peter Naur (图灵奖得主)提出了‘数据学’概念，建议计算机学界不仅要关注科学计算，也要关注数据处理



**1996年**

数据科学(Data Science)这一名称（主要是从统计学领域）开始出现



**2002年**

数据科学杂志创办：主要以科学数据为对象





## 数据科学的提出

○ 2007年

Jim Gray提出数据密集型科学发现（科学研究的第四范式）

○ 2013年

有些媒体称为中国的大数据元年



# 大数据时代

**Baidu 百度** 大数据  **百度一下**

[Q 网页](#) [图 资讯](#) [图 视频](#) [图 图片](#) [? 知道](#) [图 文库](#) [图 贴吧](#) [图 采购](#) [图 地图](#) [更多](#)

百度为您找到相关结果约100,000,000个 [搜索工具](#)

**Google** 大数据   

[全部](#) [图 图片](#) [图 新闻](#) [图 视频](#) [图 图书](#) [更多](#) [设置](#) [工具](#)

找到约1,010,000,000条结果 (用时 0.39 秒)

**Google** big data   

[全部](#) [图 图片](#) [图 新闻](#) [图 视频](#) [图 图书](#) [更多](#) [设置](#) [工具](#)

找到约8,000,000,000条结果 (用时 0.58 秒)





## 数据大到什么程度

**2018年**，微信报告每天有450亿次信息发送出，有4.1亿次音视频呼叫成功

**2019年**，全球每天收发2936亿封电子邮件  
一辆联网汽车每天将产生4TB的数据

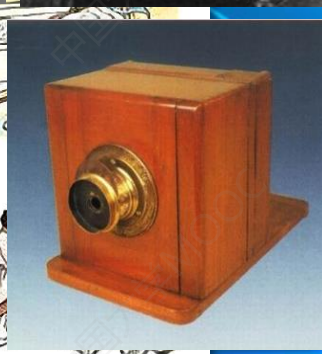
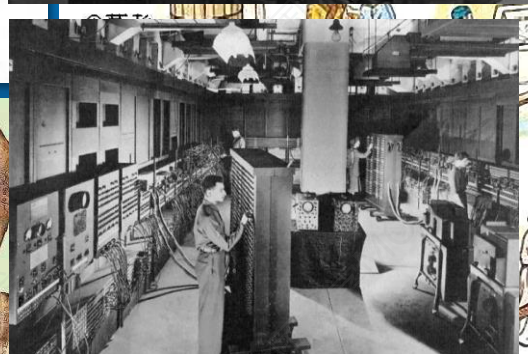
**2020年**，全球可穿戴设备将产生28PB的数据  
IDC预测，2025年全球每天产生的数据量将达到491EB



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



## 人类和数据



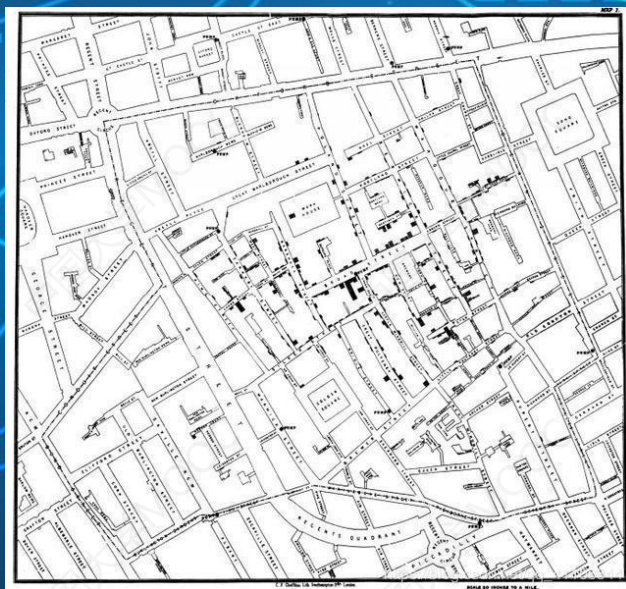




## 数据驱动的成功案例



### 数据为疾病提供解决方案







## 数据驱动的成功案例



### 数据阻止一场战争

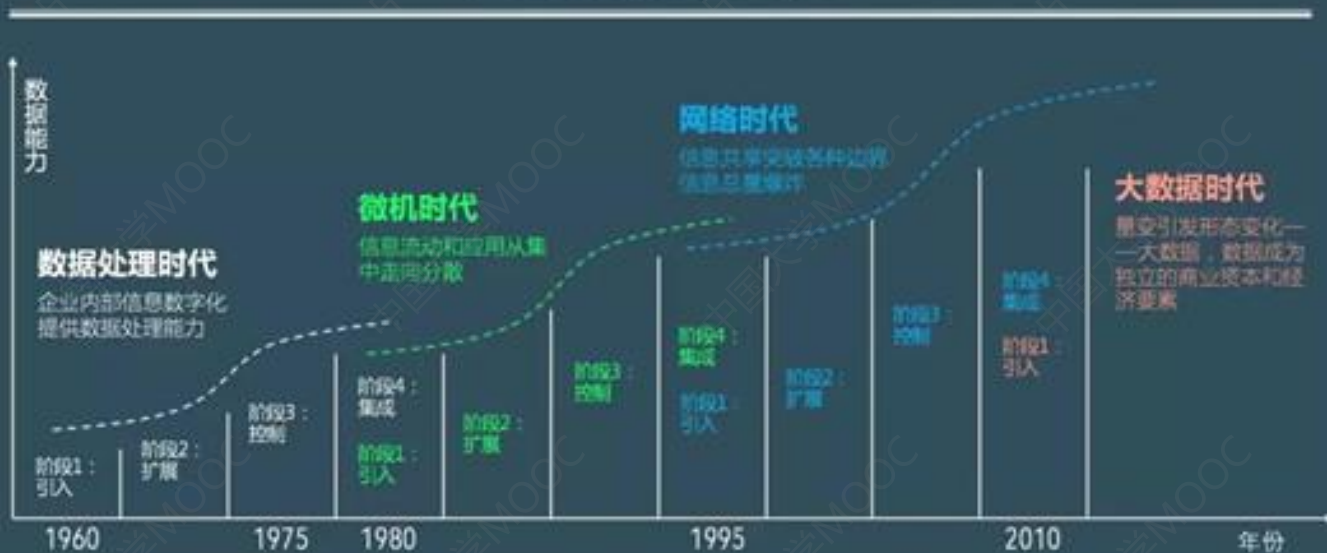






## 越来越大的数据

### 大数据发展态势



大数据持续升温，国内 2020 年预计达到 1 万亿  
计算形态发生变化，进入第四范式：密集型科学计算



# 大数据的特征







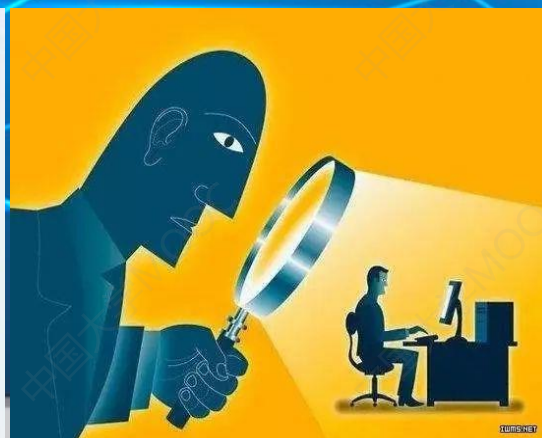
真的越大越好吗?

点外卖

打车

导航

购物





## 真的越大越准吗？

