

Data Mining: Method and Application

——Data Visualization——

徐华

清华大学 计算机系 智能技术与系统国家重点实验室 2021年5月31日

1

目录



- 相关背景概述
- 几种典型数据的可视化
- 大数据可视化的常见形式



相关背景概述



◎ 大数据(Big Data)是继云计算和物联网之后又一个具有革命性的信息技术

- ◆ 2008年《Nature》出版专刊,专题性讨论海量数据对于互联网、经济、环境以及生物等各方面的影响与挑战
- ◆ 2011年《Science》出版专刊,专题性讨论如何应对数据洪流(Data Deluge)

● 大数据分析是大数据研究领域的核心内容之一

- ◆ "数据正在变得无处不在、触手可及;而数据创造的真正价值,在于我们能否提供进一步的稀缺的附加服务.这种增值服务就是数据分析"(Google 首席经济学家、 UC Berkeley 大学 Hal Varian 教授)
- ◆ 美国《时代》杂志于 2012 年 11 月指出,奥巴马的成功连任背后所依托的关键即是两年来对大数据的分析与挖掘(http://swampland.time.com),例如,通过对海量选民微博的分析得出选民对总统候选人的喜好

◉ 数据的分析过程往往离不开机器和人的相互协作与优势互补

- ◆ 从机器或计算机的角度出发,强调机器的计算能力和人工智能,以各种高性能处理算法、智能搜索与挖掘 算法等为主要研究内容
- ◆ 从人作为分析主体和需求主体的角度出发,强调基于人机交互的、符合人的认知规律的分析方法,意图将人所具备的、机器并不擅长的认知能力融入分析过程中,这一研究分支以大数据可视分析(Visual Analytics of BigData)[6,7]为主要代表。

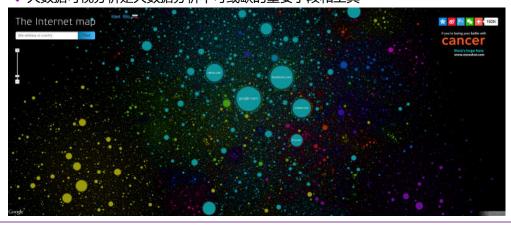
3

相关背景概述



● 一幅图胜过千言万语

- ◆ 人类从外界获得的信息约有 80%以上来自于视觉系统
- ◆ 大数据可视分析是大数据分析不可或缺的重要手段和工具



信息可视化、人机交互、可视分析的基本概念



- 信息可视化(Information Visualization):对抽象数据使用计算机支持的、交互的、可视化的表示形式以增强认知能力
 - ◆ 通过可视化图形呈现数据中隐含的信息和规律
 - ◆ 研究的创新性可视化表征旨在建立符合人的认知规律的心理映像(Mental Image)
- 人机交互(Human-Computer Interaction):人与系统之间通过某种对话语言,在一定的交互方式和技术支持下的信息交换过程
- 可视分析(Visual Analytics)一种通过交互式可视化界面来辅助用户对大规模复杂数据集进行分析推理的科学与技术
 - ◆ 可视分析概念提出时拟定的目标之一即是面向大规模、动态、模糊,或者常常不一致的数据 集来进行分析

5



信息可视化的参考模型 • 信息可视化参考模型 Data Visual Form task Raw Data Visual Data Tables Structures Data Visual View Transformations Mappings Transformations **Human Interaction**

面向大数据主流应用的信息可视化技术



- 大数据可视化技术涉及传统的科学可视化和信息可视化,从大数据分析将掘取信息和洞悉知识作为目标的角度出发,信息可视化技术将在大数据可视化中扮演更为重要的角色。
- 根据信息的特征把信息可视化技术分为一维信息(1-Dimensional)、二维信息(2-Dimensional)、 三维信息(3-Dimensional)、多维信息(Multi-Dimensional)、层次信息(Tree)、网络信息 (Network)、时序信息(Temporal)可视化

7

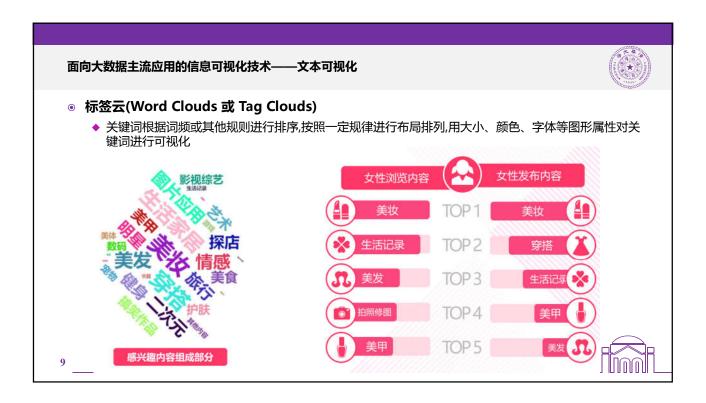


目录



- 相关背景概述
- 几种典型数据的可视化
- 大数据可视化的常见形式





面向大数据主流应用的信息可视化技术——文本可视化

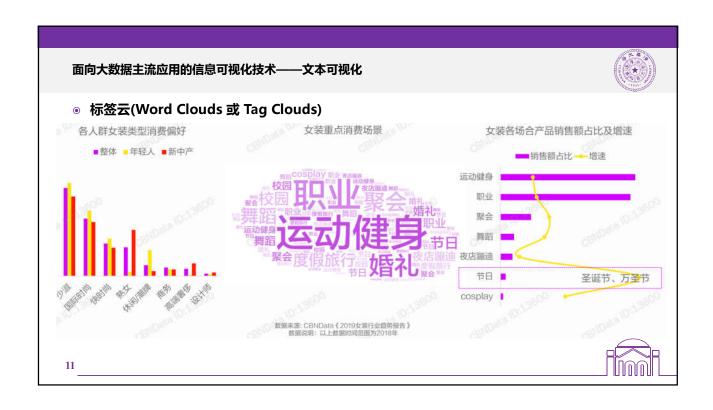


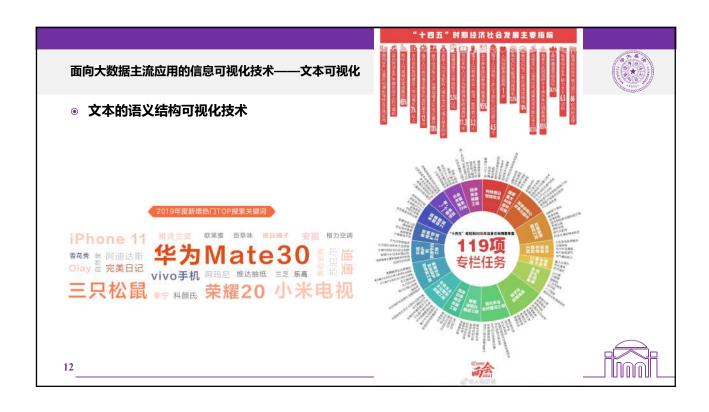
- 标签云(Word Clouds 或 Tag Clouds)
 - ◆ 关键词根据词频或其他规则进行排序,按照一定规律进行布局排列,用大小、颜色、字体等图形属性对关键词进行可视化

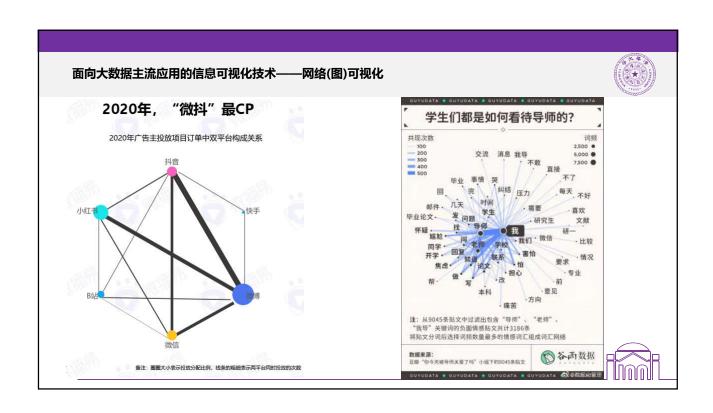


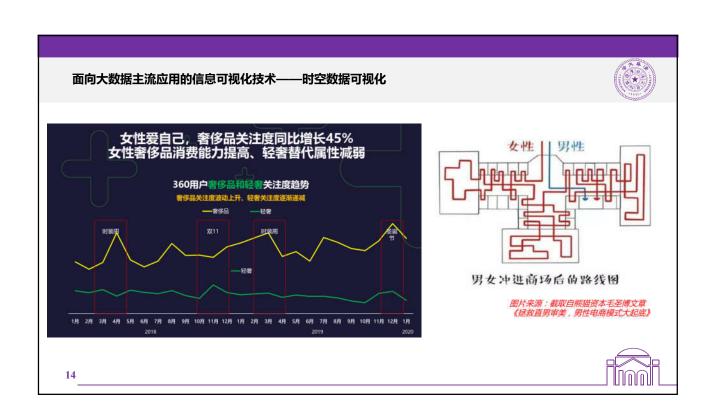


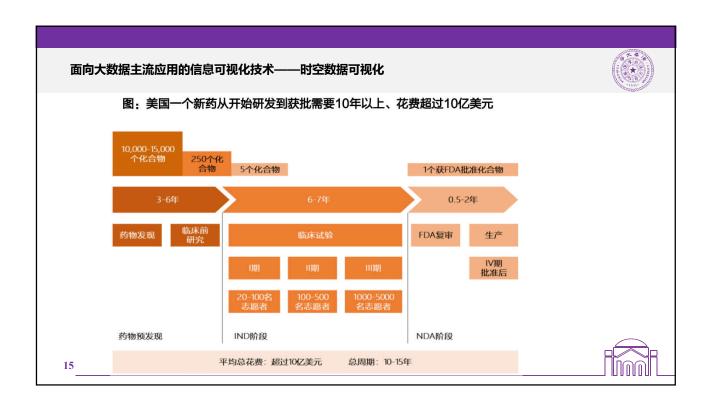


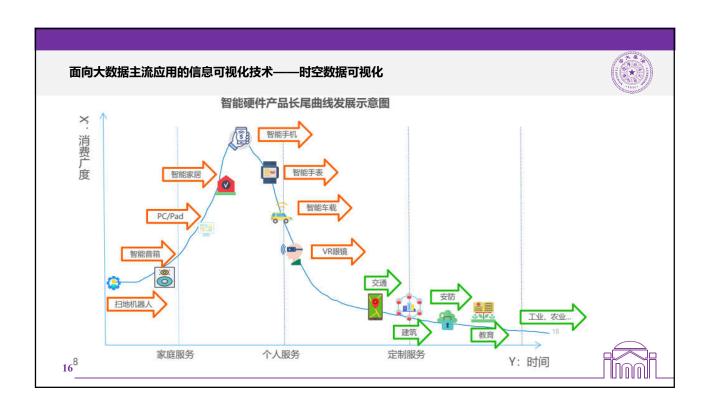












目录



- 相关背景概述
- 几种典型数据的可视化
- 大数据可视化的常见形式

17



大数据分析结果的可视化 ② 数据分析的可视化实例-基本统计数字的呈现(1) ◆ 单一的趋势曲线也能够抓人眼球 2018-2024E全球线上零售市场规模 2018-2024E全球线上零售市场规模 2018 2019 2020 2021E 2022E 2023E 2024E ○ 线上零售市场规模(万亿美元) - 线上零售市场规模(万亿美元) - 线上零售占社消总额比例 ◆ 年幣长率

