



## 第4章 大数据应用

(PPT版本号: 2022--V1)





## CONTENTS

4.1 大数据在互联网领域的应用

4.2 大数据在生物医学领域的应用

4.3 大数据在物流领域的应用

4.4 大数据在城市管理领域的应用

4.5 大数据在金融领域的应用

4.6 大数据在汽车领域的应用

4.7 大数据在零售领域的应用

4.8 大数据在餐饮领域的应用

4.9 大数据在电信领域的应用

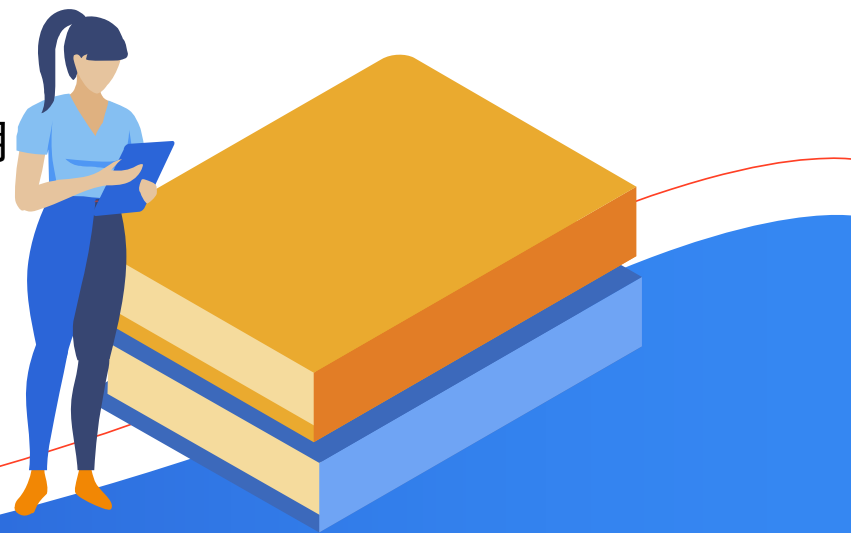
4.10 大数据在能源领域的应用

4.11 大数据在体育和娱乐领域的应用

4.12 大数据在安全领域的应用

4.13 大数据在政府领域的应用

4.14 大数据在日常生活中的应用





# 4.1

## 大数据在互联网领域的应用

4.1.1 什么是推荐系统

4.1.2 长尾理论

4.1.3 推荐方法

4.1.4 推荐系统模型

4.1.5 推荐系统的应用

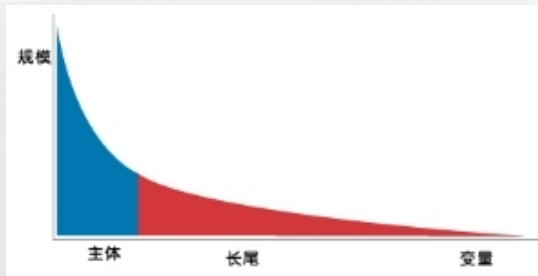


## 4.1.1 什么是推荐系统

- 互联网的飞速发展使我们进入了信息过载的时代，搜索引擎可以帮助我们查找内容，但只能解决明确的需求
- 为了让用户从海量信息中高效地获得自己所需的信息，推荐系统应运而生。推荐系统是大数据在互联网领域的典型应用，它可以通过分析用户的历史记录来了解用户的喜好，从而主动为用户推荐其感兴趣的信息，满足用户的个性化推荐需求
- 推荐系统是自动联系用户和物品的一种工具，和搜索引擎相比，推荐系统通过研究用户的兴趣偏好，进行个性化计算。推荐系统可发现用户的兴趣点，帮助用户从海量信息中去发掘自己潜在的需求

## 4.1.2 长尾理论

- 推荐系统可以创造全新的商业和经济模式，帮助实现长尾商品的销售
- “长尾”概念于2004年提出，用来描述以亚马逊为代表的电子商务网站的商业和经济模式
- 电子商务网站销售种类繁多，虽然绝大多数商品都不热门，但这些不热门的商品总数量极其庞大，所累计的总销售额将是一个可观的数字，也许会超过热门商品所带来的销售额
- 因此，可以通过发掘长尾商品并推荐给感兴趣的用户来提高销售额。这需要通过个性化推荐来实现





## 4.1.2 长尾理论

- 热门推荐是常用的推荐方式，广泛应用于各类网站中，如热门排行榜。但热门推荐的主要缺陷在于推荐的范围有限，所推荐的内容在一定时期内也相对固定。无法实现长尾商品的推荐
- 个性化推荐可通过推荐系统来实现。推荐系统通过发掘用户的行为记录，找到用户的个性化需求，发现用户潜在的消费倾向，从而将长尾商品准确地推荐给需要它的用户，进而提升销量，实现用户与商家的双赢。

## 4.1.3 推荐方法

•推荐系统的本质是建立用户与物品的联系，根据推荐算法的不同，推荐方法包括如下几类：

专家推荐：人工推荐，由资深的专业人士来进行物品的筛选和推荐，需要较多的人力成本

基于统计的推荐：基于统计信息的推荐（如热门推荐），易于实现，但对用户个性化偏好的描述能力较弱

基于内容的推荐：通过机器学习的方法去描述内容的特征，并基于内容的特征来发现与之相似的内容

协同过滤推荐：应用最早和最为成功的推荐方法之一，利用与目标用户相似的用户已有的商品评价信息，来预测目标用户对特定商品的喜好程度

混合推荐：结合多种推荐算法来提升推荐效果

## 4.1.4 推荐系统模型

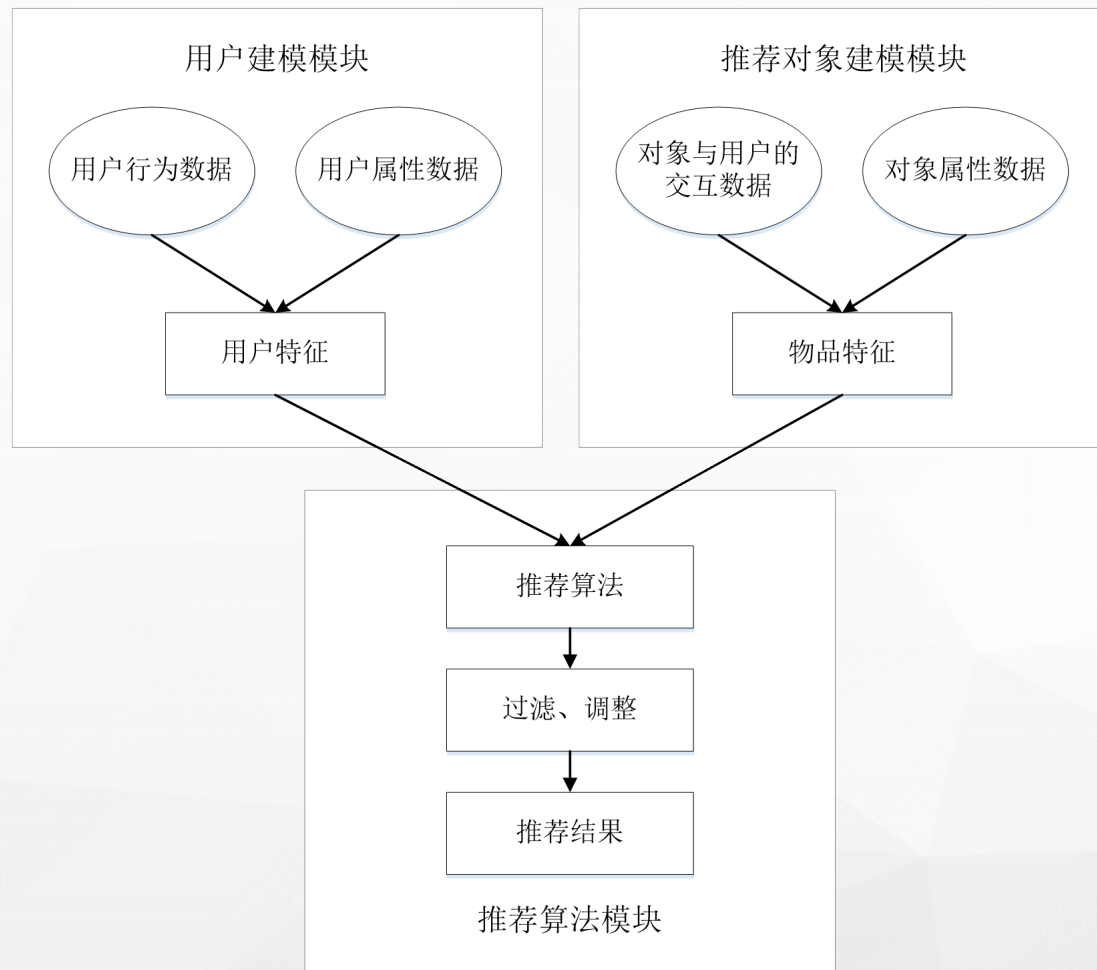


图 推荐系统基本架构



## 4.1.5 推荐系统的应用

- 目前推荐系统已广泛应用于电子商务、在线视频、在线音乐、社交网络等各类网站和应用中
- 如亚马逊网站利用用户的浏览历史记录来为用户推荐商品，推荐的主要是用户未浏览过，但可能感兴趣、有潜在购买可能性的商品

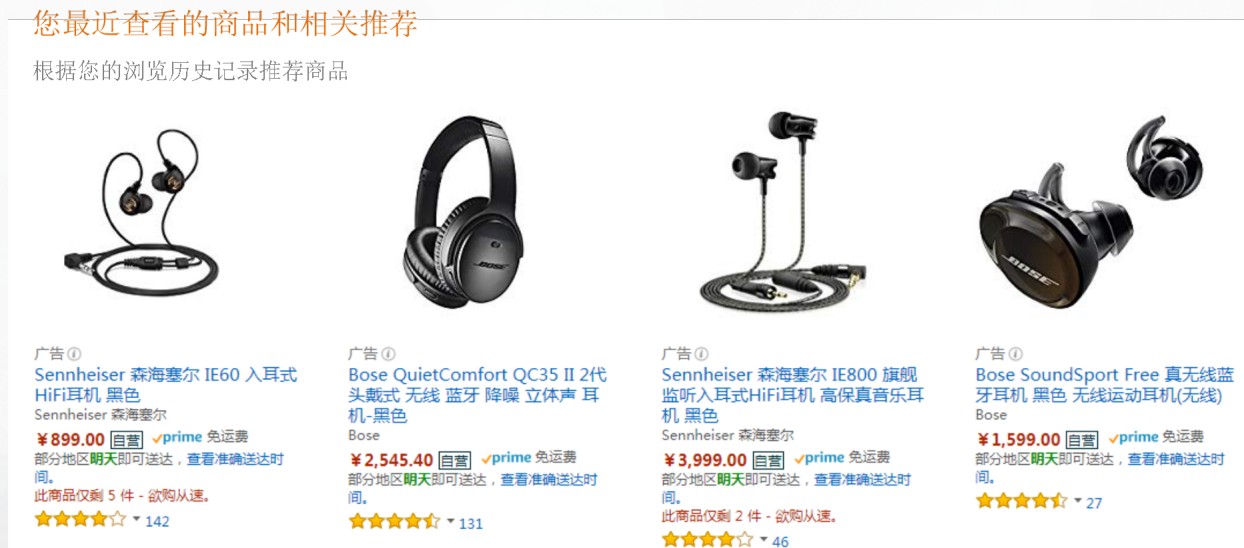


图 亚马逊网站根据用户的浏览记录来推荐商品



# 4.2

## 大数据在生物医学领域的应用

4.2.1 流行病预测

4.2.2 智慧医疗

4.2.3 生物信息学

4.2.4 案例：基于大数据的综合健康服务平台

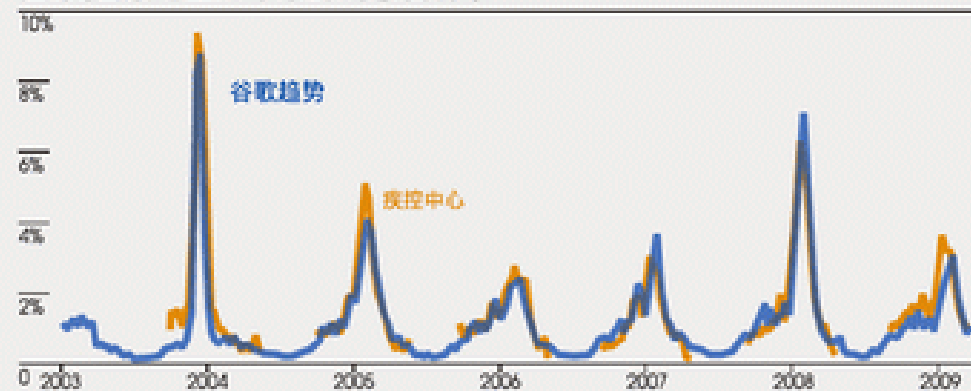


## 4.2.1 流行病预测



- 从谷歌流感趋势看大数据的应用价值
- “谷歌流感趋势”，通过跟踪搜索词相关数据来判断全美地区的流感情况

图:美国某地区历年来的流感发病率



数据来源: 谷歌趋势, 美国各地疾病预防控制中心



## 4.2.1 流行病预测

• 百度疾病预测（<http://trends.baidu.com/disease>）就是具有代表性的互联网疾病预测服务，其基本原理是：流行病的发生和传播有一定的规律性，与气温变化、环境指数、人口流动等因素密切相关，每天网民在百度搜索大量流行病相关信息，汇聚起来就有了统计规律，经过一段时间的积累，可以形成一个个预测模型，预测未来疾病的活跃指数。

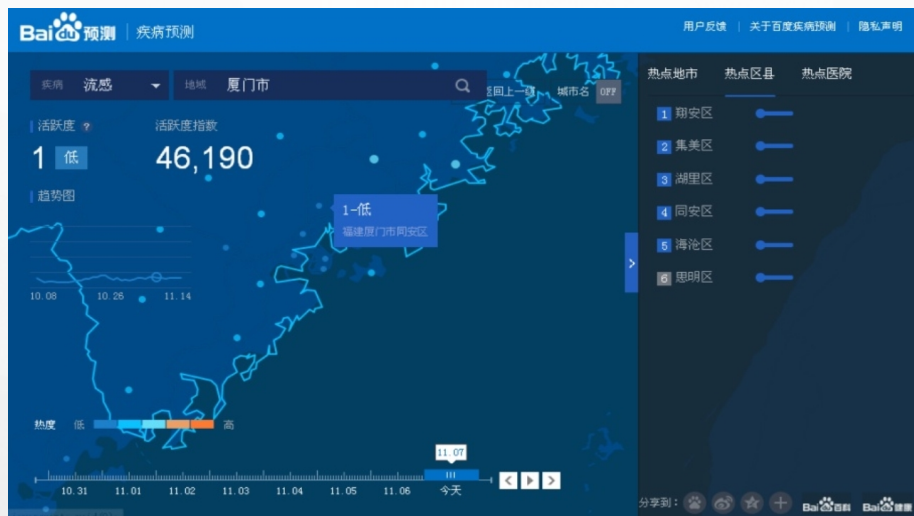


图 百度疾病预测

## 4.2.2 智慧医疗



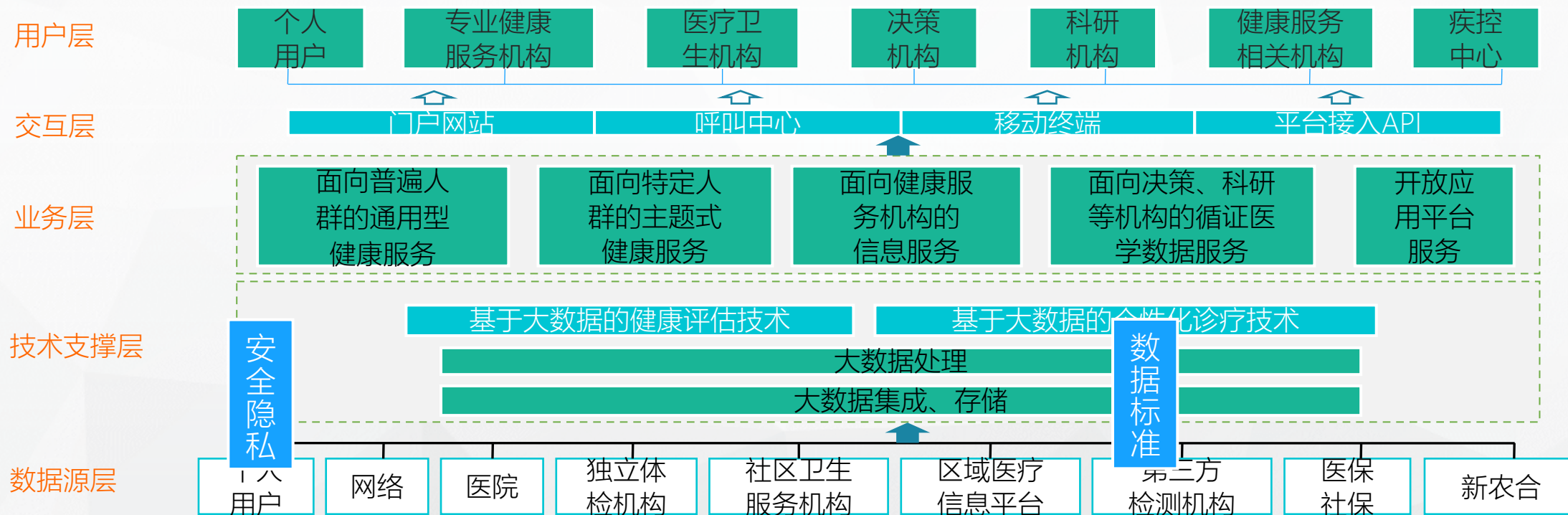


## 4.2.3 生物信息学

- 生物信息学是研究生物信息的采集、处理、存储、传播、分析和解释等方面的学科，也是随着生命科学和计算机科学的迅猛发展、生命科学和计算机科学相结合形成的一门新学科，它通过综合利用生物学、计算机科学和信息技术，揭示大量而复杂的生物数据所蕴含的生物学奥秘
- 和互联网数据相比，生物信息学领域的数据更是典型的大数据
- 生物大数据使得我们可以利用先进的数据科学知识，更加深入地了解生物学过程、作物表型、疾病致病基因等
- 用大数据分析技术，可以从个人健康档案中有效预测个人健康趋势，并为其提供疾病预防建议，达到“治未病”的目的

## 4.2.4 案例：基于大数据的综合健康服务平台

- 目标：构建覆盖全生命周期、内涵丰富、结构合理的以人为本全面连续的综合健康服务体系，利用大数据技术和智能设备技术，提供线上线下相结合的公众健康服务，实现“未病先防、已病早治、既病防变、愈后防复”，满足社会公众多层次、多方位的健康服务需求，提升人民群众的身心健康水平。





# 4.3

## 大数据在物流领域的应用

- 4.3.1 智能物流的概念
- 4.3.2 智能物流的作用
- 4.3.3 智能物流的应用
- 4.3.4 大数据是智能物流的关键
- 4.3.5 中国智能物流骨干网——菜鸟



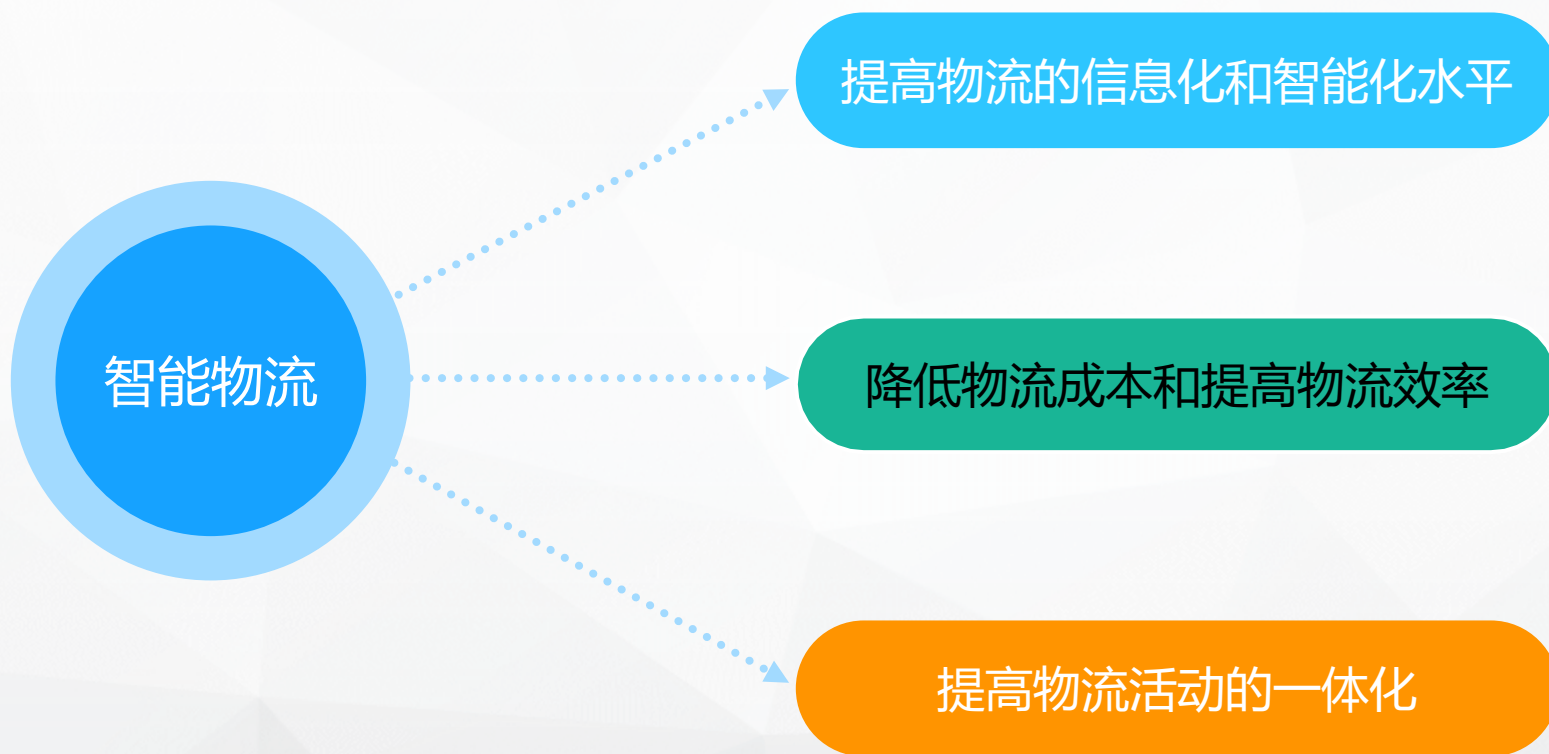
## 4.3.1 智能物流的概念

- 智能物流, 又称智慧物流, 是利用集成智能化技术, 使物流系统能模仿人的智能, 具有思维、感知、学习、推理判断和自行解决物流中某些问题的能力, 从而实现物流资源优化调度和有效配置、物流系统效率提升的现代化物流管理模式。





## 4.3.2 智能物流的作用





### 4.3.3 智能物流的应用

- 国内许多城市都在围绕智慧港口、多式联运、冷链物流、城市配送等方面，着力推进物联网在大型物流企业、大型物流园区的系统级应用
- 还可以将射频标签识别技术、定位技术、自动化技术以及相关的软件信息技术，集成到生产及物流信息系统领域，探索利用物联网技术实现物流环节的全流程管理模式，开发面向物流行业的公共信息服务平台，优化物流系统的配送中心网络布局，集成分布式仓储管理及流通渠道建设，最大限度地减少物流环节、简化物流过程，提高物流系统的快速反应能力
- 此外，还可以进行跨领域信息资源整合，建设基于卫星定位、视频监控、数据分析等技术的大型综合性公共物流服务平台，发展供应链物流管理

## 4.3.4 大数据是智能物流的关键

- 在物流领域有两个著名的理论——“黑大陆说”和“物流冰山说”
- 这两个理论都旨在说明物流活动的模糊性和巨大潜力
- 发现隐藏在海量数据背后的有价值的信息，是大数据的重要商业价值。大数据是打开物流领域这块神秘的“黑大陆”的一把金钥匙
- 大数据将推动物流行业从粗放式服务到个性化服务的转变，颠覆整个物流行业的商业模式

## 4.3.5 中国智能物流骨干网——菜鸟

智能物流案例：阿里巴巴的中国智能物流骨干网（地网）



菜鸟网络到底是什么？

- 中国智能物流骨干网，又名“菜鸟”
- 菜鸟网络计划在5到8年内，打造一个全国性的超级物流网。
- 这个网络能在24小时内将货物运抵国内任何地区，能支撑日均300亿元(年度约10万亿元)的巨量网络零售额。

1000亿元投资物流基础设施 强强联手共建智能骨干网络  
物流信息系统向所有的制造商、网商、快递公司、第三方物流公司完全开放

### 阿里物流体系

天网

天猫牵头负责与各大物流快递公司对接的数据平台

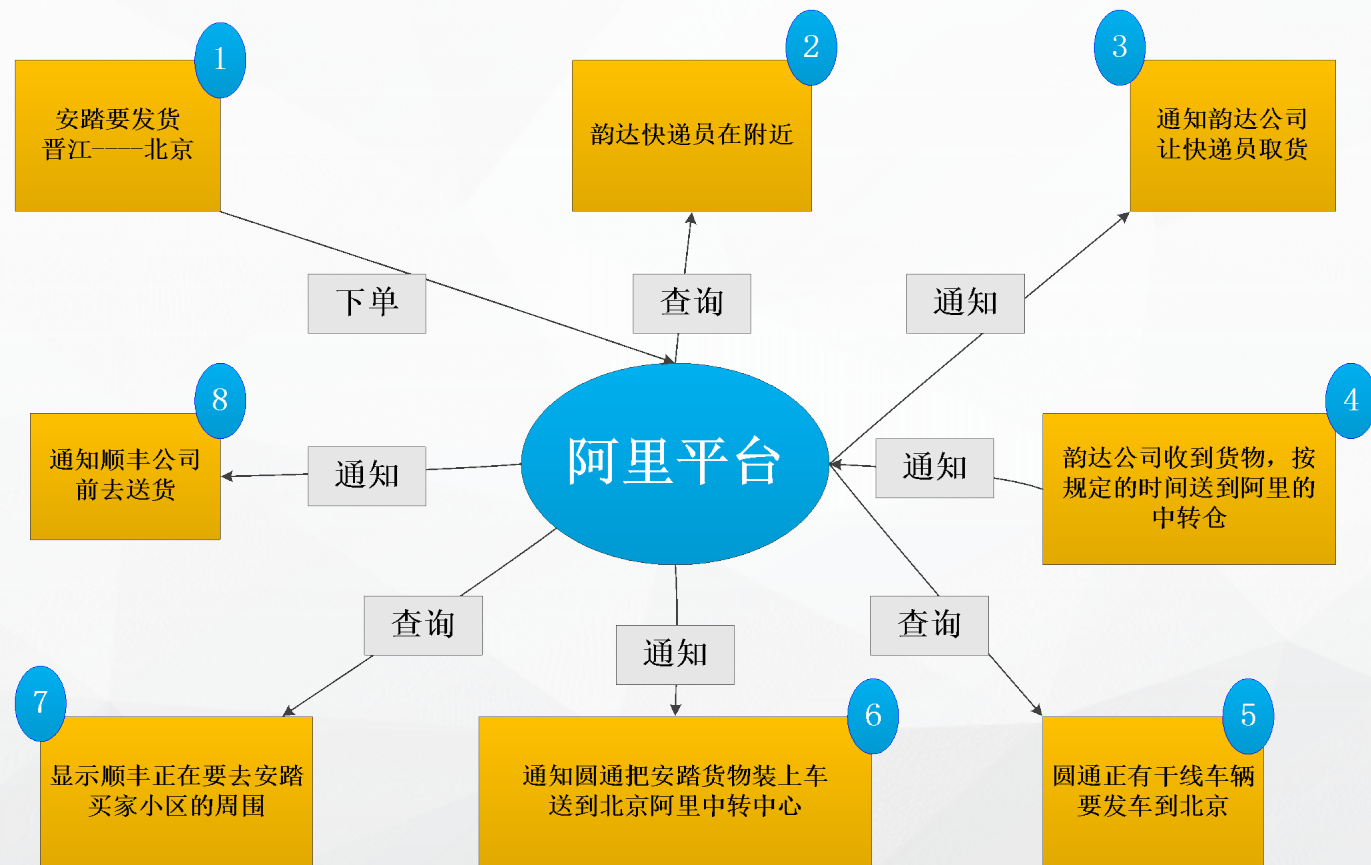
地网

即“菜鸟”，又称“中国智能物流骨干网（CSN）”



## 4.3.5 中国智能物流骨干网——菜鸟

### 菜鸟网络未来畅想





# 4.4

## 大数据在城市管理领域的应用

4.4.1 智能交通

4.4.2 环保监测

4.4.3 城市规划

4.4.4 安防领域





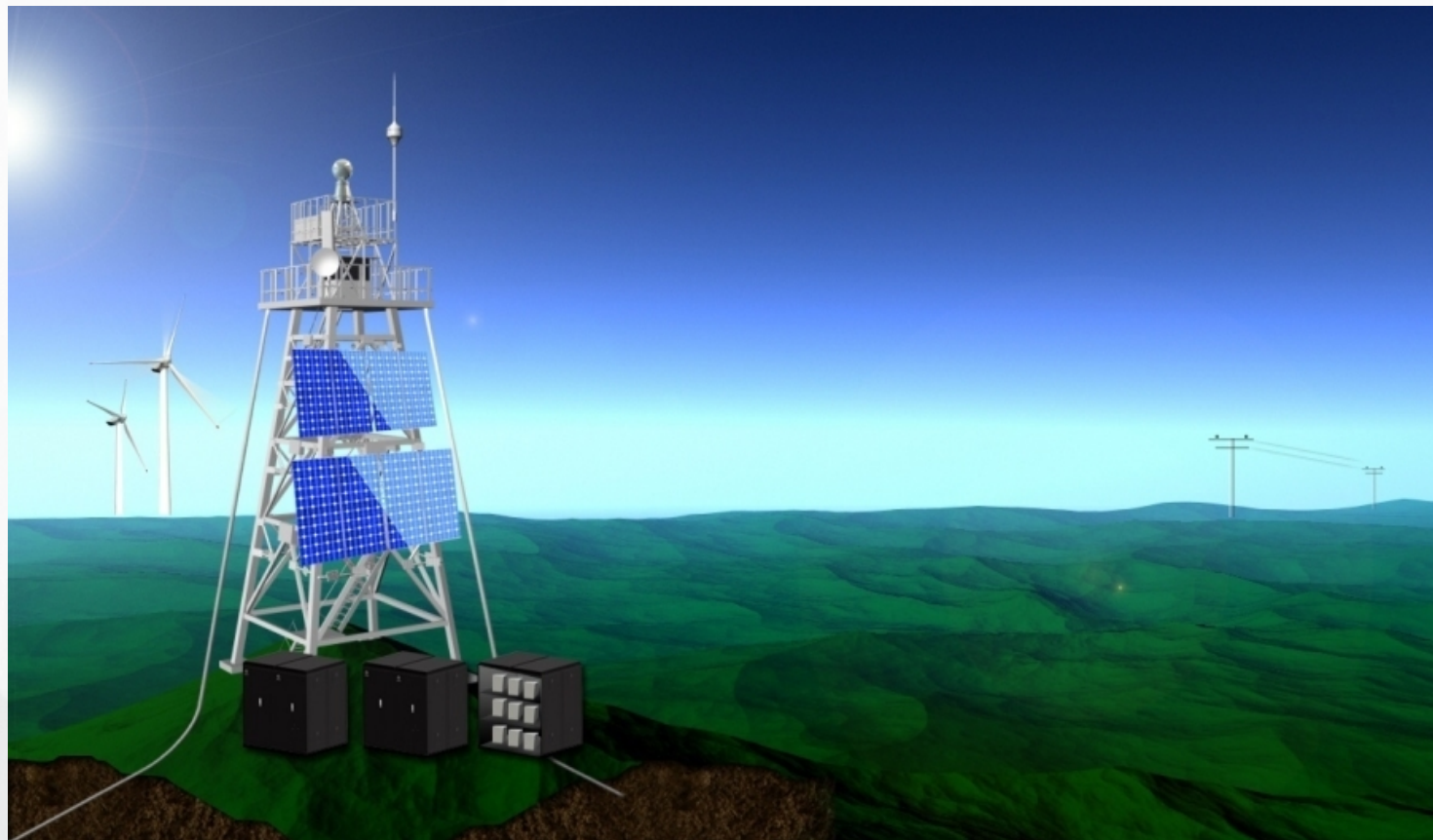
## 4.4.1 智能交通

- 智能交通将先进的信息技术、数据通信传输技术、电子传感技术、控制技术以及计算机技术等，有效集成并运用于整个地面交通管理，同时可以利用城市实时交通信息、社交网络和天气数据来优化最新的交通情况。



## 4.4.2 环保监测

### 1. 森林监视



## 4.4.2 环保监测

### 2. 环境保护

- 大数据已经被广泛应用于污染监测领域，借助大数据技术，采集各项环境质量指标信息，集成整合到数据中心进行数据分析，并把分析结果用于指导下一步环境治理方案的制定，可以有效提升环境整治的效果
- 中国水污染地图
- 中国空气污染地图
- 中国固废污染地图
- 汽车尾气污染治理



### 4.4.3 城市规划

- 利用地理数据可以研究全国城市扩张模拟、城市建成区识别
- 利用公交IC卡数据，可以开展城市居民通勤分析、职住分析、人的行为分析、人的识别、重大事件影响分析、规划项目实施评估分析等
- 利用移动手机通话数据，可以研究城市联系、居民属性、活动关系及其对城市交通的影响
- 利用社交网络数据，可以研究城市功能分区、城市网络活动与等级、城市社会网络体系等。利用出租车定位数据，可以开展城市交通研究
- 利用搜房网的住房销售和出租数据，同时结合网络爬虫获取的居民住房地理位置和周边设施条件数据，就可以评价一个城区的住房分布和质量情况

## 4.4.4 安防领域

- 中国的很多城市都在开展平安城市建设，在城市的各个角落密布成千上万个摄像头，7×24小时不间断采集各个位置的视频监控数据，数据量之大，超乎想象
- 除了视频监控数据，安防领域还包含大量其他类型的数据，包括结构化、半结构化和非结构化数据







# 4.5

## 大数据在金融领域的应用

- 4.5.1 高频交易
- 4.5.2 市场情绪分析
- 4.5.3 信贷风险分析
- 4.5.4 大数据征信



## 4.5.1 高频交易

- 高频交易（High-Frequency Trading, HFT）是指从那些人们无法利用的极为短暂的市场变化中寻求获利的计算机化交易，比如，某种证券买入价和卖出价差价的微小变化，或者某只股票在不同交易所之间的微小价差
- 为了从高频交易中获得更高的利润，一些金融机构开始引入大数据技术来决定交易。



## 4.5.2 市场情绪分析

- 市场情绪分析是交易者在日常交易工作中不可或缺的一环，根据市场情绪分析、技术分析和基本面分析，可以帮助交易者做出更好的决策。大数据技术在市场情绪分析中大有用武之地。





### 4.5.3 信贷风险分析

- 大数据分析技术已经能够为企业信贷风险分析助一臂之力。通过收集和分析大量中小微企业用户日常交易行为的数据，判断其业务范畴、经营状况、信用状况、用户定位、资金需求和行业发展趋势，解决由于其财务制度的不健全而无法真正了解其真实经营状况的难题，让金融机构放贷有信心、管理有保障





## 4.5.4 大数据征信

大数据征信就是利用信息技术优势，将不同信贷机构、消费场景、支离破碎的海量数据整合起来，经过数据清洗、模型分析、校验等一系列流程后，加工融合成真正有用的信息。





# 4.6

## 大数据在汽车领域的应用



## 4.6 大数据在汽车领域的应用

- 为了实现无人驾驶的功能，谷歌无人驾驶汽车上配备了大量传感器，包括雷达、车道保持系统、激光测距系统、红外摄像头、立体视觉、GPS导航系统、车轮角度编码器等，这些传感器每秒产生1GB数据，每年产生的数据量将达到约2PB
- 大数据分析技术将帮助无人驾驶系统做出更加智能的驾驶动作决策，比人类驾车更加安全、舒适、节能、环保



图 谷歌无人驾驶汽车



# 4.7

## 大数据在零售领域的应用

4.7.1 发现关联购买行为

4.7.2 客户群体细分

4.7.3 供应链管理





## 4.7.1 发现关联购买行为

### 啤酒与尿布的故事



## 4.7.2 客户群体细分

- 美国Target超市比孩子父亲还早发现他女儿已经怀孕



### 4.7.3 供应链管理

- 亚马逊、联合包裹快递（UPS）、沃尔玛等先行者已经开始享受大数据带来的成果，大数据可以帮助它们更好地掌控供应链
- 美国最大的医药贸易商McKesson公司，对大数据的应用也已经远远领先于大多数企业。该公司运用先进的运营系统，可以对每天200万个订单进行全程跟踪分析，并且监督超过80亿美元的存货





# 4.8

## 大数据在餐饮领域的应用

4.8.1 餐饮行业拥抱大数据

4.8.2 餐饮O2O



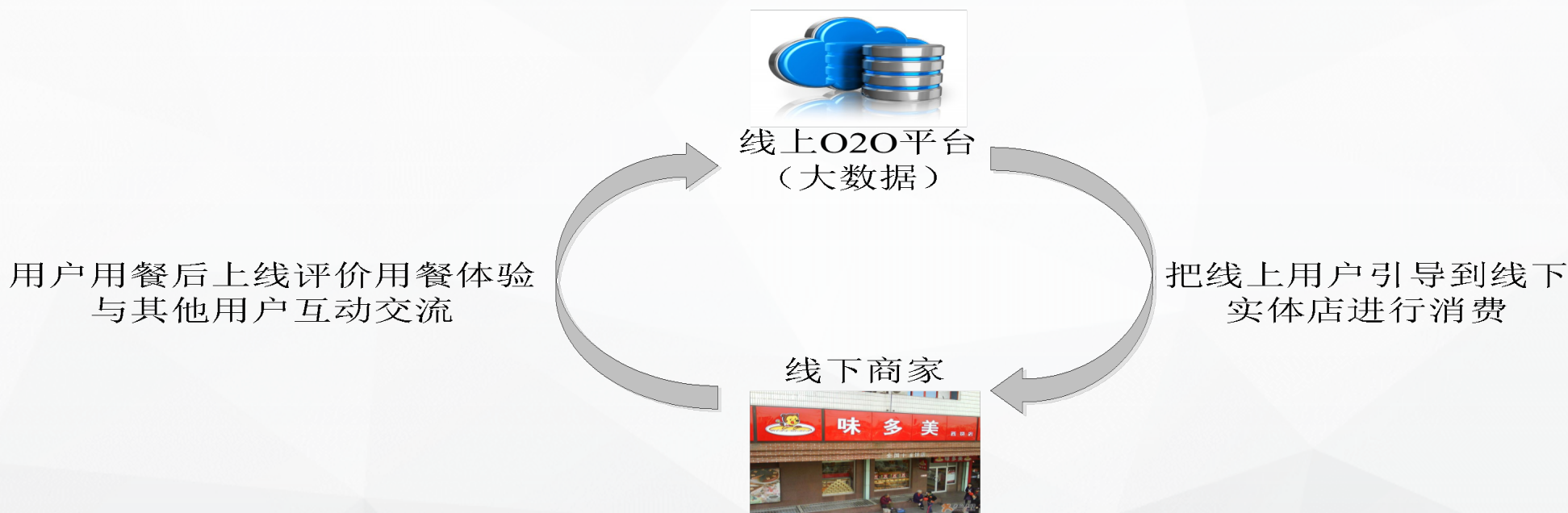


## 4.8.1 餐饮行业拥抱大数据

- Food Genius是一家总部位于美国芝加哥的公司，聚合了来自美国全国各地餐馆的菜单数据，对超过350 000家餐馆的菜单项目进行跟踪，以帮助餐馆更好地确定价格、食品和营销的趋势。这些数据可以帮助餐馆获得商机，并判断哪些菜可能获得成功，从而减少菜单变化所带来的不确定性。
- Avero餐饮软件公司则通过对餐饮企业内部运营数据进行分析，帮助企业提高运营效率，如制定什么样的战略可以提高销量、在哪个时间段开展促销活动效果最好等。

## 4.8.2 餐饮O2O

• 餐饮O2O（Online To Offline）模式是指无缝整合线上线下资源，形成以数据驱动的O2O闭环运营模式



- 1. 大数据驱动的团购模式
- 2. 利用大数据为用户推荐消费内容

- 3. 利用大数据调整线下门店布局
- 4. 利用大数据控制店内人流量



# 4.9

## 大数据在电信领域的应用



## 4.9 大数据在电信领域的应用

- 预测客户行为，发现行为趋势，并找出公司服务过程中存在缺陷的环节，从而帮助公司及时采取措施保留客户

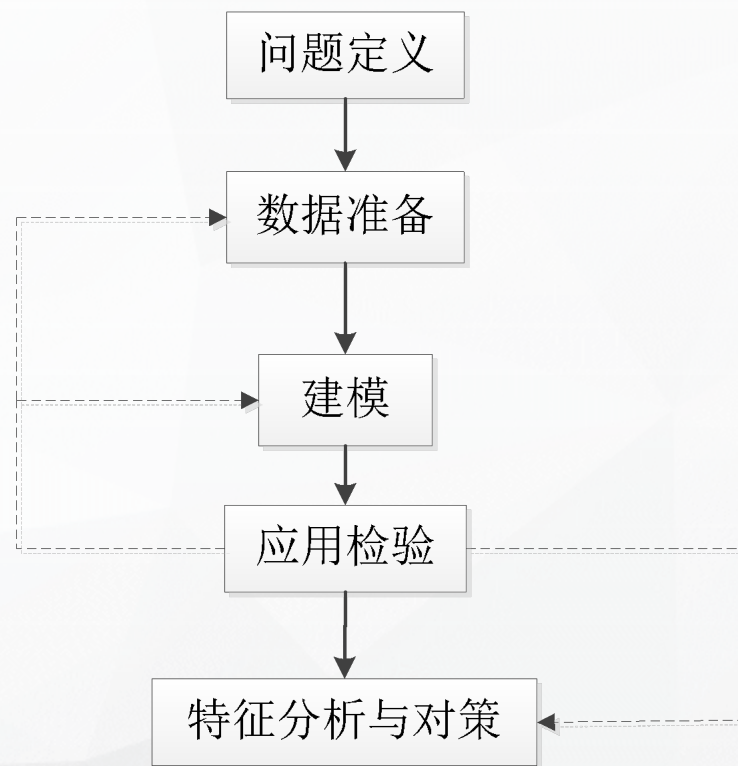


图 电信客户离网分析





# 4.10

## 大数据在能源领域的应用



## 4.10 大数据在能源领域的应用

- 智能电网的发展，离不开大数据技术的发展和运用，大数据技术是组成整个智能电网的技术基石





# 4.11

## 大数据在体育和娱乐领域的应用

4.11.1 训练球队

4.11.2 投拍影视作品

4.11.3 预测比赛结果



## 4.11.1 训练球队

- 大数据可以有效帮助一支球队进一步提升整体实力和水平



图 2014年巴西世界杯决赛中德国队队长拉姆的技术数据



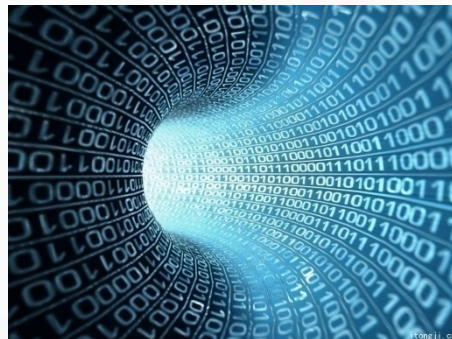
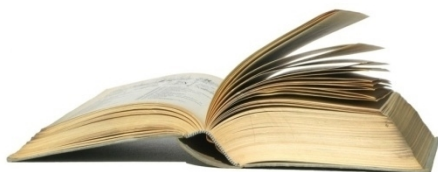
## 4.11.2 投拍影视作品



Kevin Spacey



David Fincher



大数据分析



### 4.11.3 预测比赛结果

- 大数据可以预测比赛结果却是具有一定的科学根据的，它用数据来说话，通过对海量相关数据进行综合分析，得出一个预测判断
- 本质上而言，大数据预测就是基于大数据和预测模型去预测未来某件事情的概率
- 利用大数据预测比赛结果，将对人们生活产生深刻的影响



# 4.12

## 大数据在安全领域的应用

4.12.1 大数据与国家安全

4.12.2 应用大数据技术防御网络攻击

4.12.3 警察应用大数据工具预防犯罪

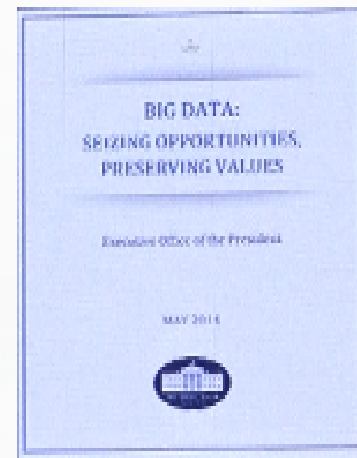




## 4.12.1 大数据与国家安全



美国前国防部长拉姆斯菲尔德多次强调：保证能力一枚导弹没有一条情报能更有效地应对恐怖活动



•美国政府2014年5月发布的大数据报告：大数据可以极大增强





## 4.12.3 警察应用大数据工具预防犯罪

根据大数据预测犯罪

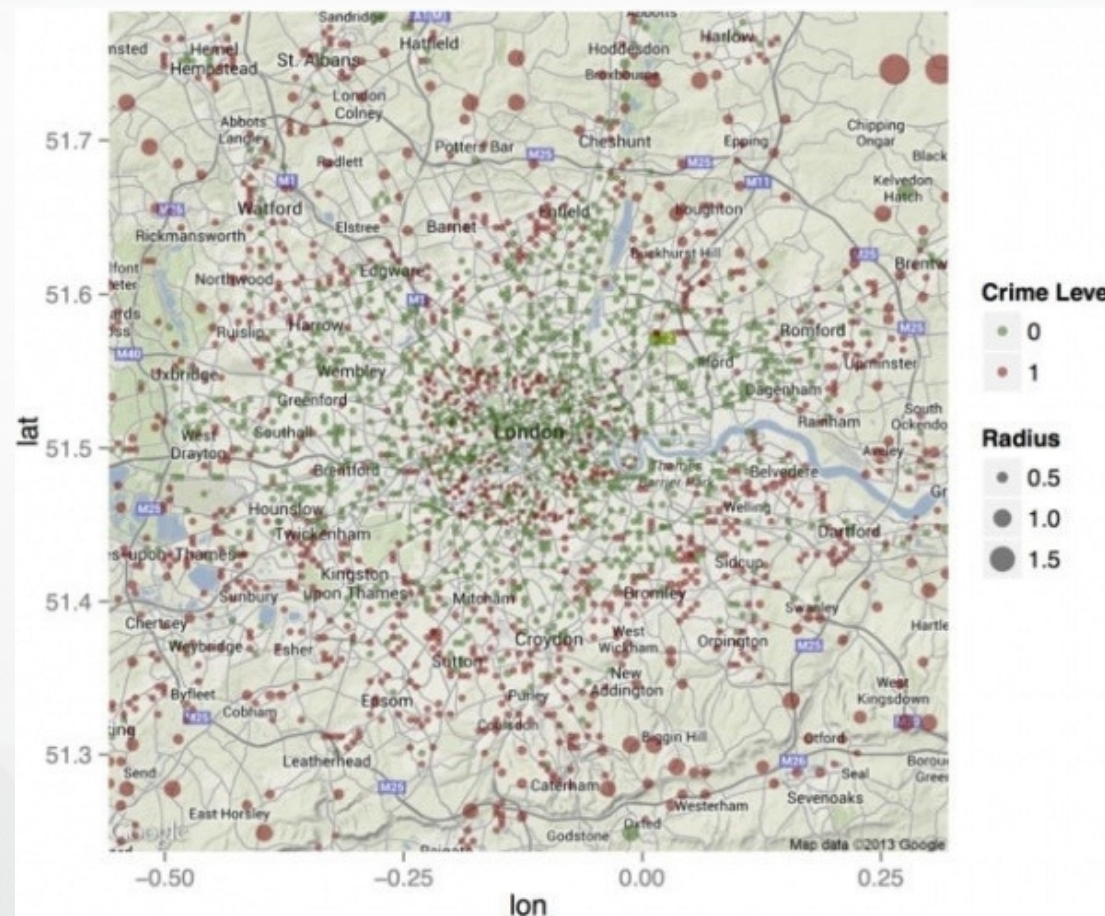
- 提供破案线索
- 实时犯罪预警（摄像头、突发新闻、聊天记录）



- 洛杉矶警察局是第一个跃进大数据时代、采取大数据公安警务模式的公安机构

## 4.12.3 警察应用大数据工具预防犯罪

根据大数据预测犯罪



伦敦犯罪事件预测地图



# 4.13

## 大数据在政府领域的应用





## 4.13 大数据在政府领域的应用



大数据分别帮助奥巴马和特朗普获得美国总统大选胜利





# 4.14

## 大数据在日常生活中的应用



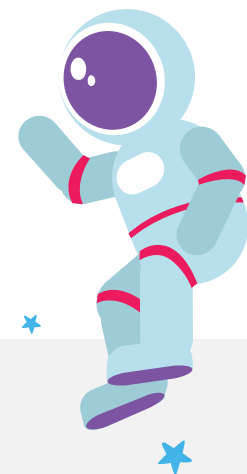
## 4.14 大数据在日常生活中的应用

- 在信息化社会，我们每个人的言行都会留下以数据形式存在的轨迹，这些分散在各个角落的数据，记录了我们的通话、聊天、邮件、购物、出行、住宿以及生理指标等各种信息，构成了与每个人相关联的“个人大数据”
- 分析个人大数据就可以深刻了解与之关联的自然人，了解他的各种生活行为习惯
- 了解了个人的生活行为模式，一些公司就可以为个人提供更加周到的服务



大数据与统计系

大数据系列课程



# 大数据



大数据与统计系数字教师网

<http://hssj.wxcgi.com/>