



《大数据导论（通识课版）》

教材官网: <http://dbllab.xmu.edu.cn/post/bigdataintroduction/>

温馨提示: 编辑幻灯片母版, 可以修改每页PPT的厦大校徽和底部文字

第2章 大数据与云计算、物联网、人工智能

(PPT版本号: 2019年秋季学期)



扫一扫访问教材官网

林子雨

厦门大学计算机科学系

E-mail: ziyulin@xmu.edu.cn ▶▶

主页: <http://www.cs.xmu.edu.cn/linziyu>





课程教材

- 林子雨 编著 《大数据导论——数据思维、数据能力和数据伦理（通识课版）》
- 高等教育出版社，2019年11月



提纲

- 2.1 云计算
- 2.2 物联网
- 2.3 大数据与云计算、物联网的关系
- 2.4 人工智能
- 2.5 大数据与人工智能的关系



高校大数据课程

公共服务平台

百度搜索厦门大学数据库实验室网站访问平台





2.1 云计算

2.1.1 云计算概念

2.1.2 云计算服务模式和类型

2.1.3 云计算数据中心

2.1.4 云计算的应用

2.1.5 云计算产业



2.1.1 云计算概念



云计算示意图

云计算概念

•通过网络、以服务的方式，为千家万户提供非常廉价的IT资源

云计算特点

■超大规模计算、虚拟化、高可靠性和安全性、通用性、动态扩展性、按需服务、降低成本



2.1.1 云计算概念

水资源获取方式





2.1.1 云计算概念

水资源获取方式——挖井取水



- (1) 初期成本高，周期长
- (2) 后期需要自己维护
- (3) 供水量有限



2.1.1 云计算概念

水资源获取方式——自来水



- (1) 初期零成本，瞬时可获得
- (2) 后期免维护，使用成本低
- (3) 在供水量方面“予取予求”



2.1.1 云计算概念

传统IT资源获取方式：自建机房



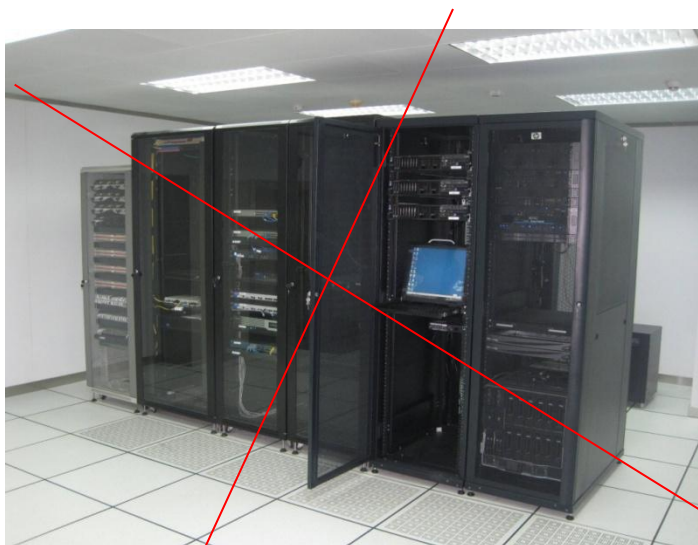
- (1) 初期成本高，周期长
- (2) 后期需要自己维护，使用成本高
- (3) IT资源供应量有限



2.1.1 云计算概念

云计算

企业不需要自建IT基础设施，可以租用云端资源



无需自建

- (1) 初期零成本，瞬时可获得
- (2) 后期免维护，使用成本低
- (3) 在供应IT资源量方面“予取予求”

企业用户

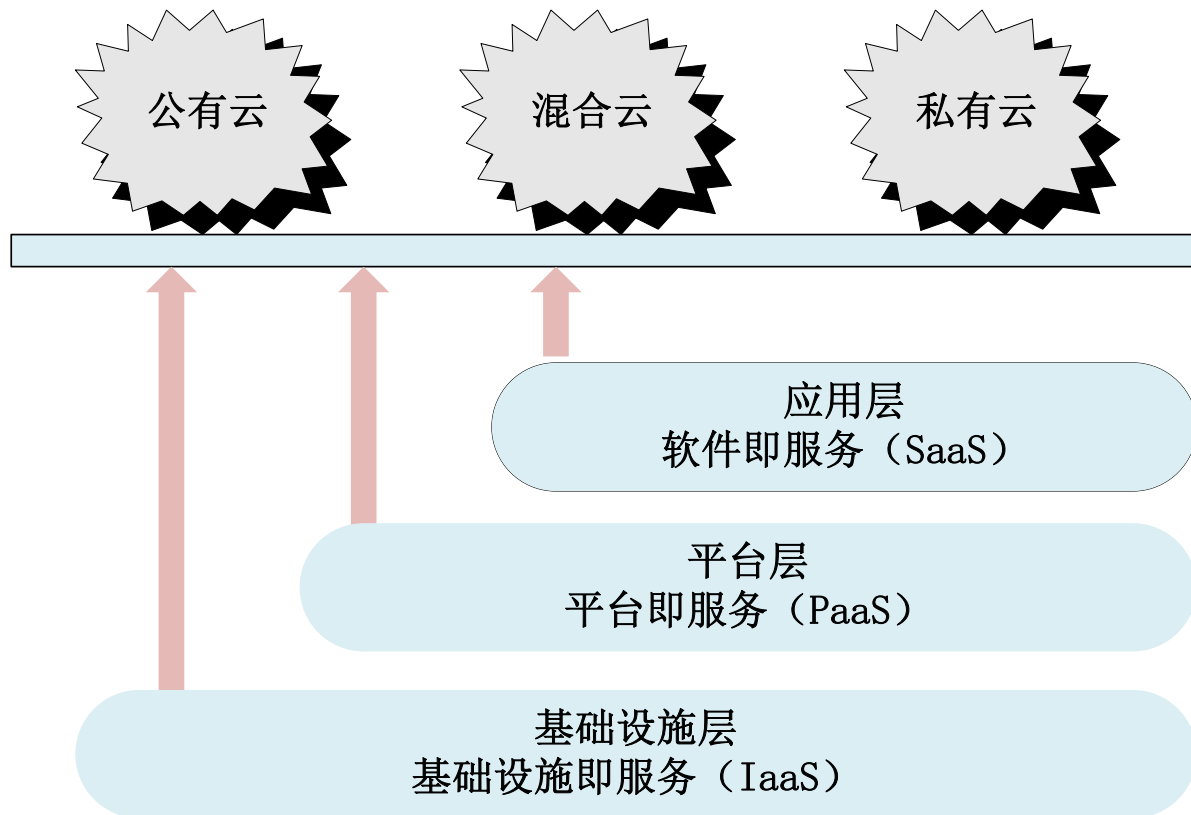


租用云端资源





2.1.2 云计算服务模式和类型





2.1.2 云计算服务模式 and 类型

SaaS

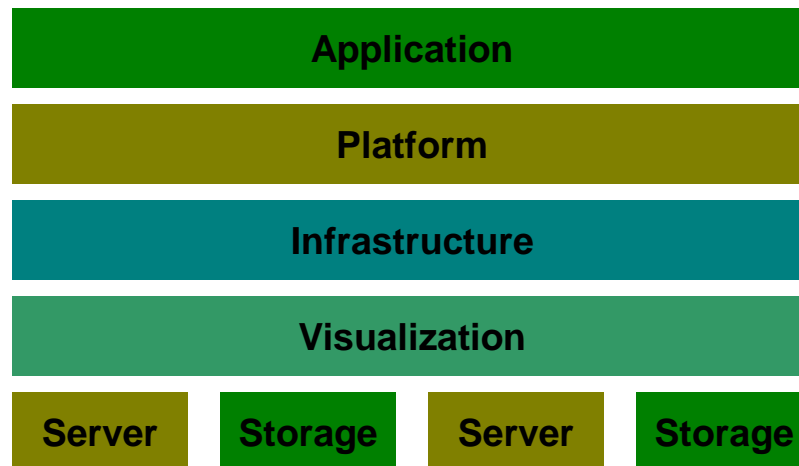
从一个集中的系统部署软件，使之在一台本地计算机上(或从云中远程地)运行的一个模型。由于是计量服务，SaaS 允许出租一个应用程序，并计时收费

PaaS

类似于 IaaS，但是它包括操作系统和围绕特定应用的必需的服务

IaaS

将基础设施(计算资源和存储)作为服务出租



SaaS

Software as a Service

Google Apps, Microsoft “Software+Services”

PaaS

Platform as a Service

IBM IT factory, Google App Engine, Force.com

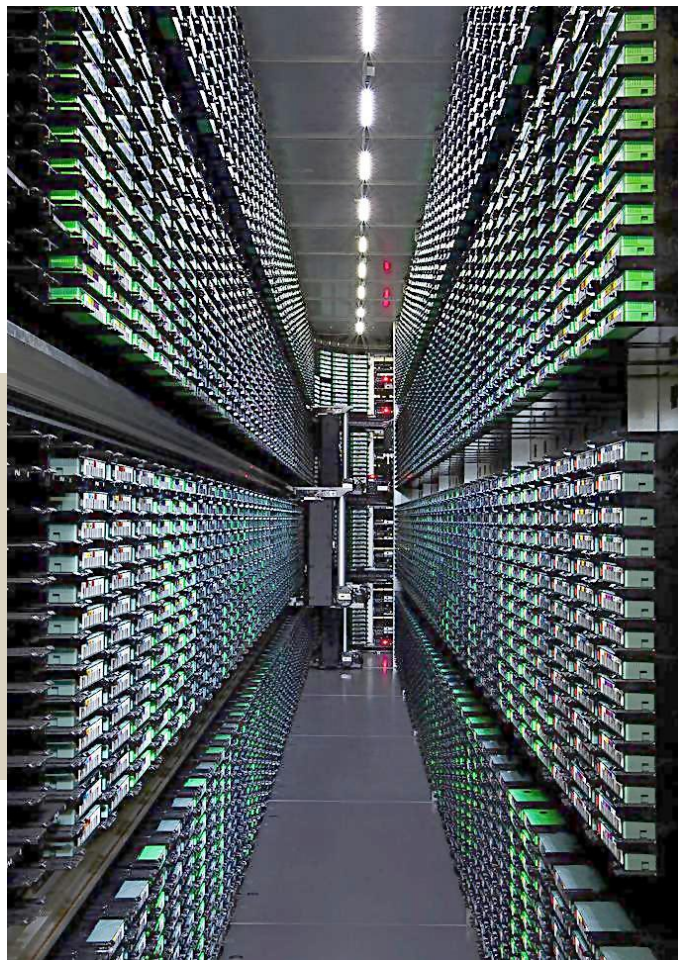
IaaS

Infrastructure as a Service

Amazon EC2, IBM Blue Cloud, Sun Grid



2.1.3 云计算数据中心



- 数据中心是云计算的温床
- 云计算推动数据中心向虚拟化和云架构的转型，不断提高IT基础架构的灵活性，以降低IT、能源和空间成本，从而让客户能够快速提高业务敏捷性



“那些正在兴建最大规模数据中心的公司一般都在云计算方面拥有巨大野心”

-- 《数据中心知识》杂志主编 Rich Miller



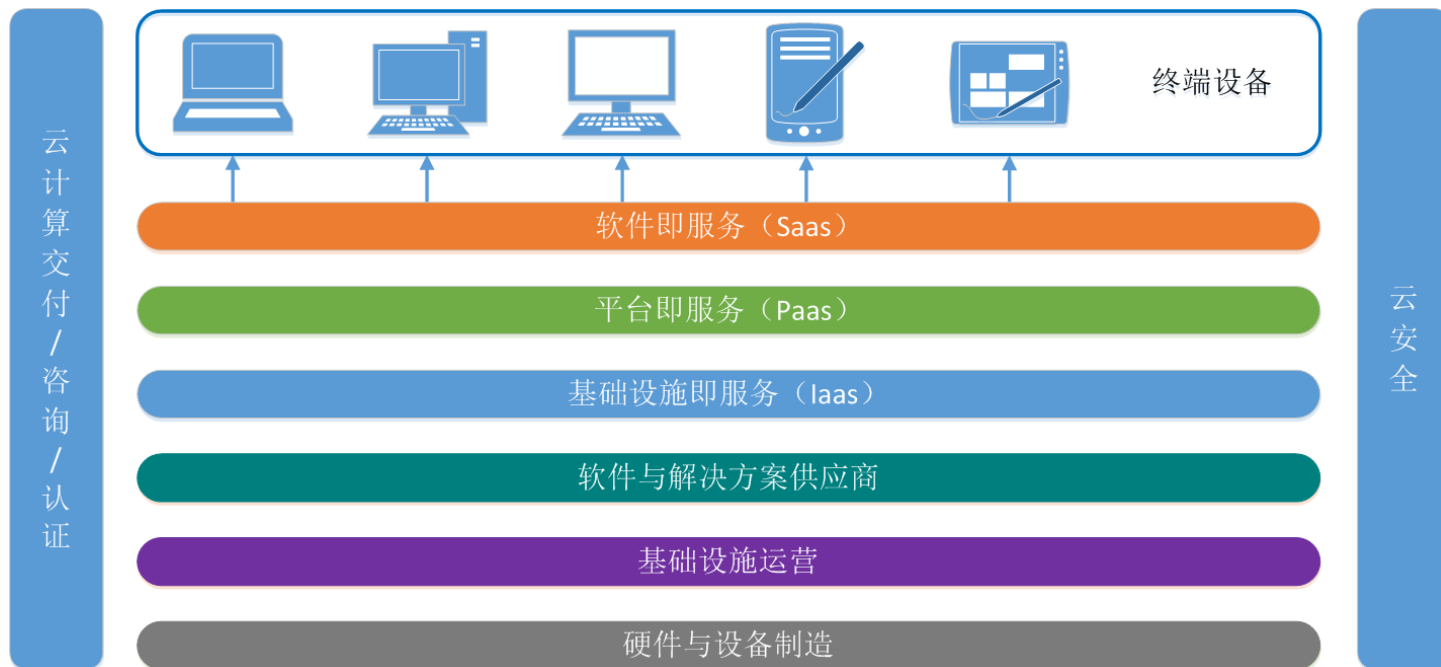
2.1.4 云计算的应用

- 政务云上可以部署公共安全、应急管理、智能交通、社会保障等应用，通过集约化建设、管理、运行，可以实现信息资源整合和政务资源共享，推动政务管理创新，加快向服务型政府转型
- 教育云可以有效整合幼儿教育、中小学教育、高等教育以及继续教育等优质教育资源，逐步实现教育信息共享、教育资源共享及教育资源深度挖掘等目标
- 中小企业云能够让企业以低廉的成本建立财务、供应链、客户关系等管理应用系统，大大降低企业信息化门槛，迅速提升企业信息化水平，增强企业市场竞争力
- 医疗云可以推动医院与医院、医院与社区、医院与急救中心、医院与家庭之间的服务共享，并形成一套全新的医疗健康服务系统，从而有效地提高医疗保健的质量



2.1.5 云计算产业

- 云计算产业作为战略性新兴产业，近些年得到了迅速发展，形成了成熟的产业链结构，产业涵盖硬件与设备制造、基础设施运营、软件与解决方案供应商、基础设施即服务（IaaS）、平台即服务（PaaS）、软件即服务（SaaS）、终端设备、云安全、云计算交付/咨询/认证等环节





2.2 物联网

2.2.1 物联网的概念

2.2.2 物联网关键技术

2.2.3 物联网的应用

2.2.4 物联网产业



2.2.1 物联网的概念

- 物联网是物物相连的互联网，是互联网的延伸，它利用局部网络或互联网等通信技术把传感器、控制器、机器、人员和物等通过新的方式联在一起，形成人与物、物与物相联，实现信息化和远程管理控制

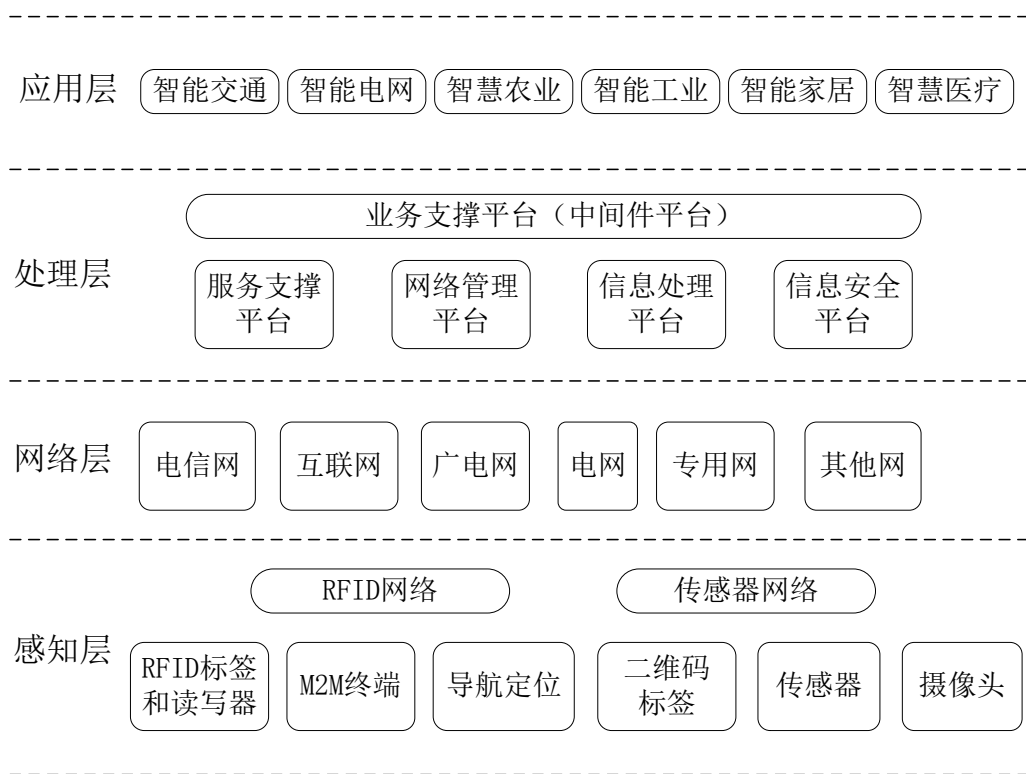


图 物联网体系架构



2.2.2 物联网关键技术

- 物联网中的关键技术包括识别和感知技术（二维码、RFID、传感器等）、网络与通信技术、数据挖掘与融合技术等

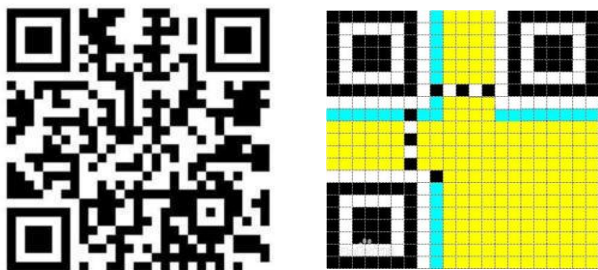


图 矩阵式二维码

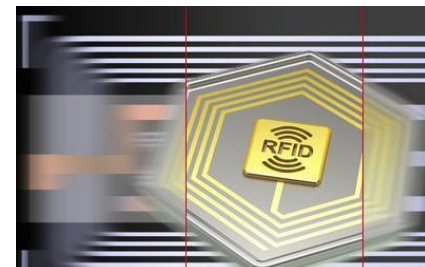


图 采用RFID芯片的公交卡



(a)温湿度传感器



(b)压力传感器



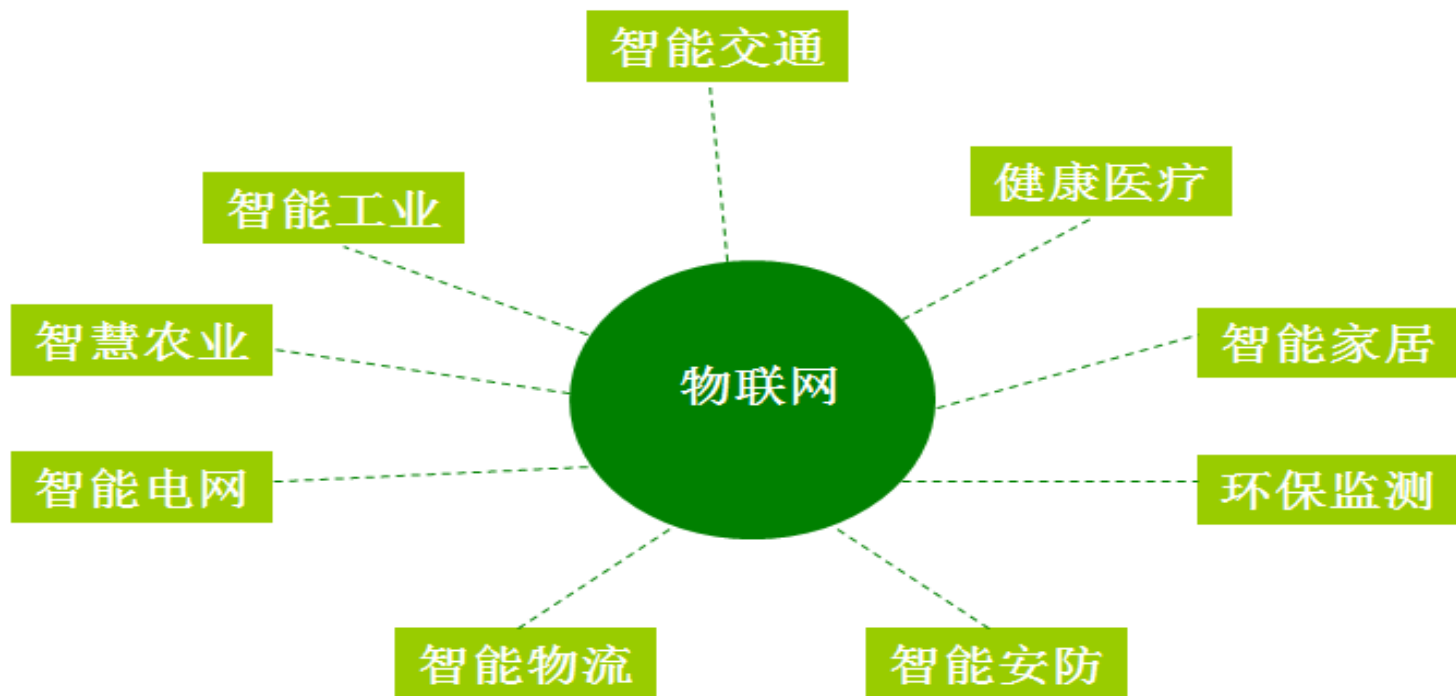
(c)烟雾传感器

图 不同类型的传感器



2.2.3 物联网的应用

- 物联网已经广泛应用于智能交通、智慧医疗、智能家居、环保监测、智能安防、智能物流、智能电网、智慧农业、智能工业等领域，对国民经济与社会发展起到了重要的推动作用





2.2.4 物联网产业

- 完整的物联网产业链主要包括核心感应器件提供商、感知层末端设备提供商、网络提供商、软件与行业解决方案提供商、系统集成商、运营及服务提供商等六大环节

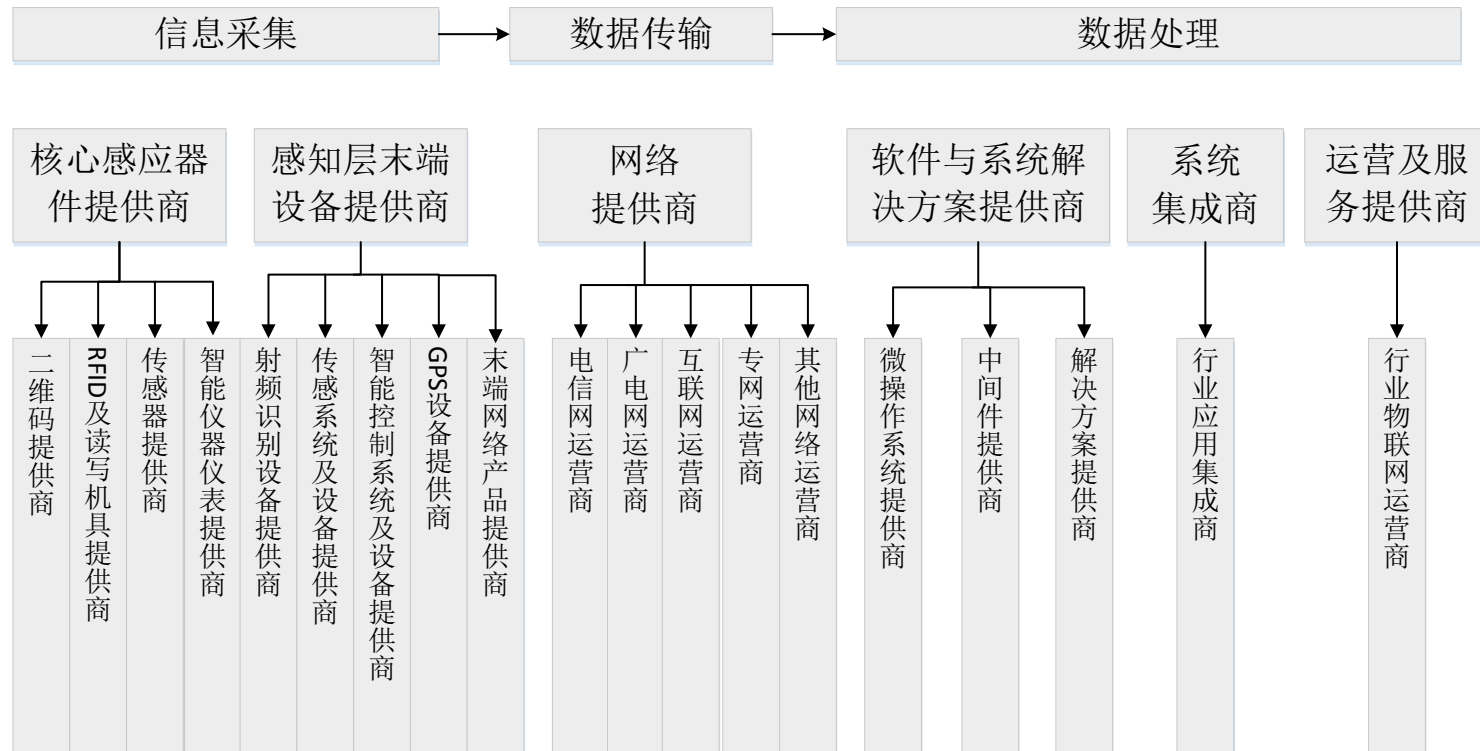


图 物联网产业链



2.3 大数据与云计算、物联网的关系

- 云计算、大数据和物联网代表了IT领域最新的技术发展趋势，三者既有区别又有联系

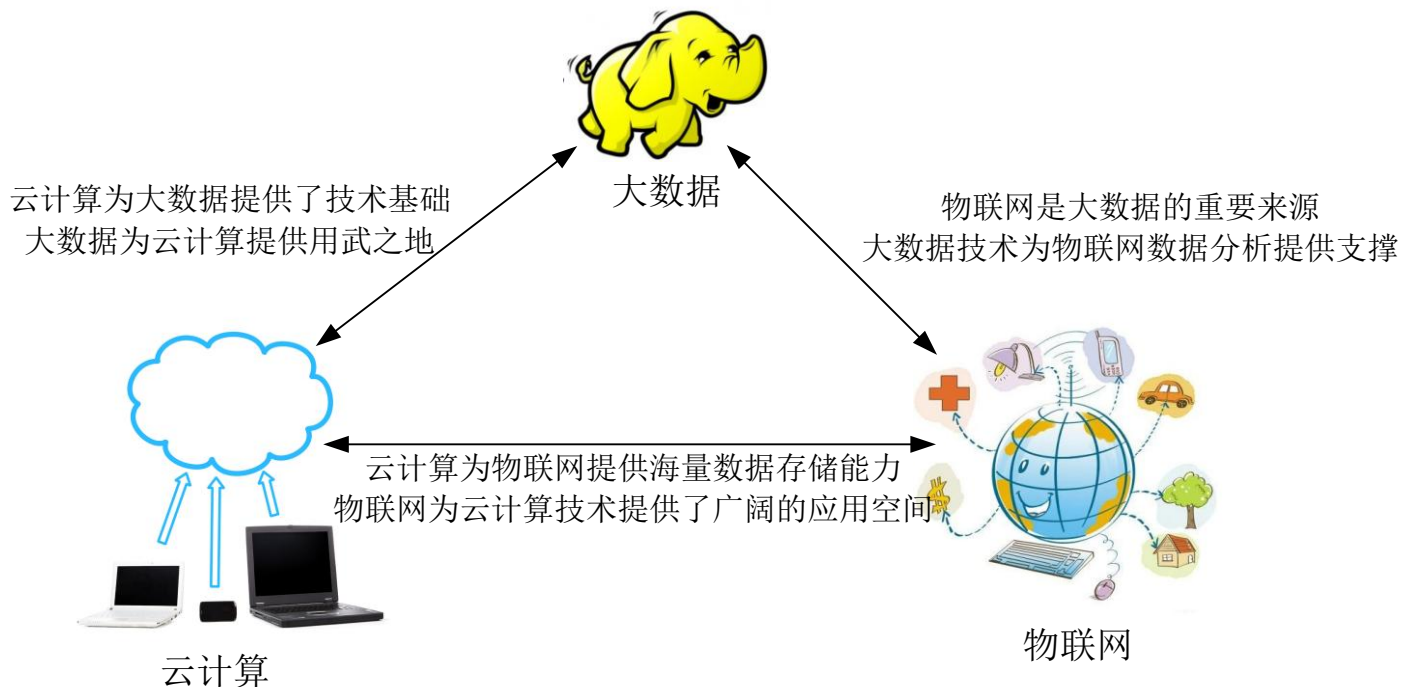


图 大数据、云计算和物联网之间的关系



2.4 人工智能

2.4.1 人工智能概念

2.4.2 人工智能关键技术

2.4.3 人工智能的应用

2.4.4 人工智能产业

2.5 大数据与人工智能的关系

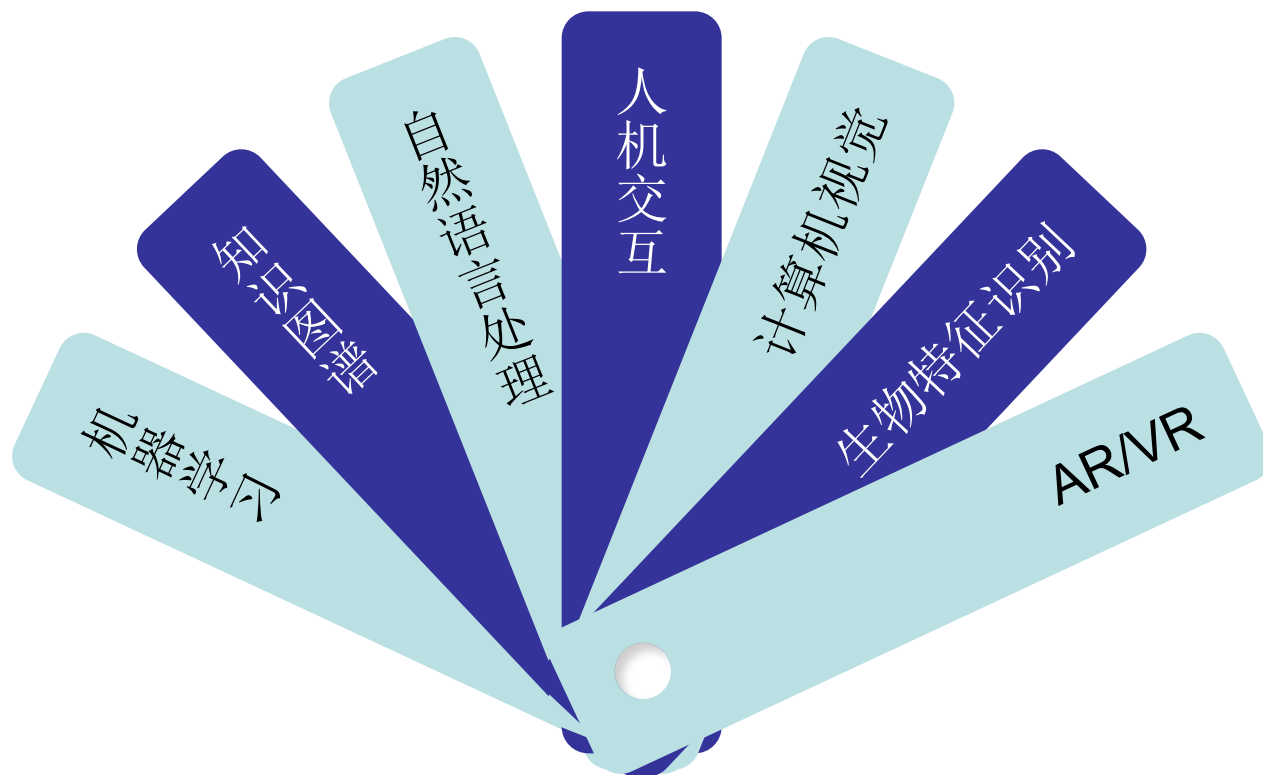


2.4.1 人工智能概念

- 人工智能（**Artificial Intelligence**），英文缩写为**AI**，是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。
- 人工智能是计算机科学的一个分支，它企图了解智能的实质，并生产出一种新的能以与人类智能相似的方式做出反应的智能机器，该领域的研究包括机器人、语言识别、图像识别、自然语言处理和专家系统等。



2.4.2 人工智能关键技术





2.4.2 人工智能关键技术

1. 机器学习

强调三个关键词：算法、经验、性能，其处理过程如图所示。在数据的基础上，通过算法构建出模型并对模型进行评估。评估的性能如果达到要求，就用该模型来测试其他的数据；如果达不到要求，就要调整算法来重新建立模型，再次进行评估。如此循环往复，最终获得满意的模型来处理其他数据。

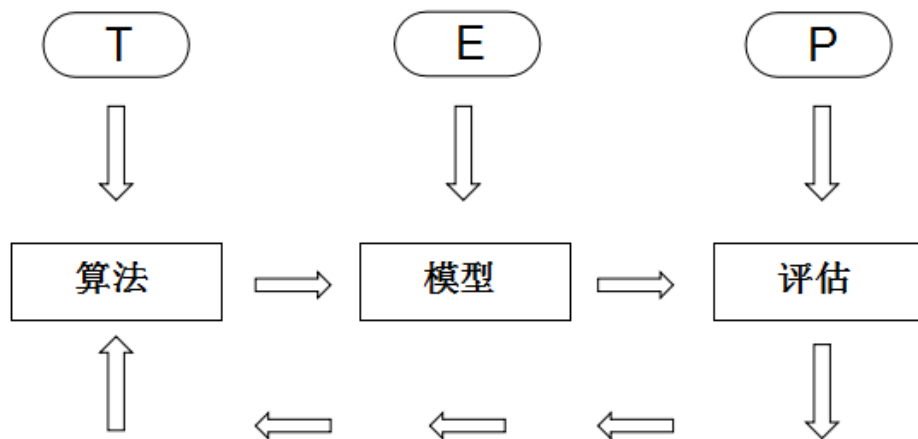


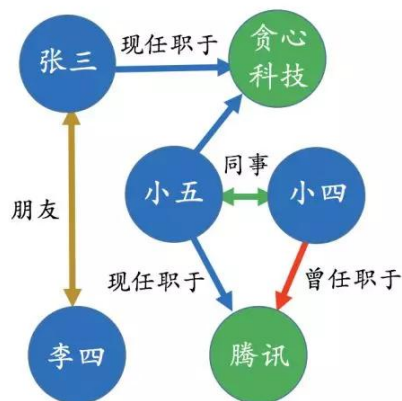
图2-13 机器学习处理过程



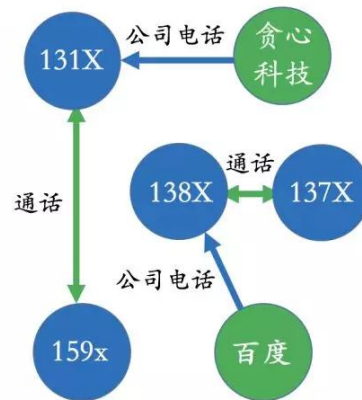
2.4.2 人工智能关键技术

2.知识图谱

知识图谱（Knowledge Graph）又称为科学知识图谱，在图书情报界称为知识域可视化或知识领域映射地图，是显示知识发展进程与结构关系的一系列各种不同的图形，用可视化技术描述知识资源及其载体，挖掘、分析、构建、绘制和显示知识及它们之间的相互联系。



案例：社交网络



案例：风控知识图谱



2.4.2 人工智能关键技术

3. 自然语言处理

- 自然语言处理是计算机科学领域与人工智能领域中的一个重要方向。它研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法。
- 自然语言处理的应用包罗万象，例如：机器翻译、手写体和印刷体字符识别、语音识别、信息检索、信息抽取与过滤、文本分类与聚类、舆情分析和观点挖掘等



2.4.2 人工智能关键技术

4. 人机交互

- 人机交互是一门研究系统与用户之间的交互关系的学科。系统可以是各种各样的机器，也可以是计算机化的系统和软件。
- 人机交互界面通常是指用户可见的部分。用户通过人机交互界面与系统交流，并进行操作。
- 人机交互是与认知心理学、人机工程学、多媒体技术、虚拟现实技术等密切相关的综合学科。

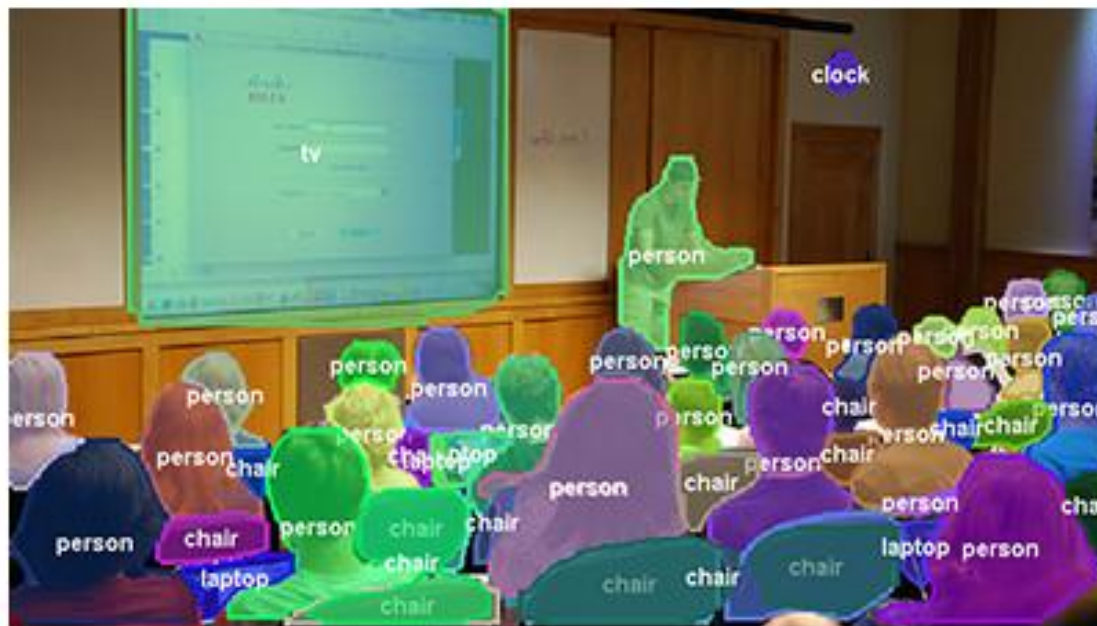




2.4.2 人工智能关键技术

5. 计算机视觉

计算机视觉是一门研究如何使机器“看”的科学，更进一步地说，是指用摄影机和电脑代替人眼对目标进行识别、跟踪和测量的机器视觉，并进一步做图形处理，成为更适合人眼观察或传送给仪器检测的图像

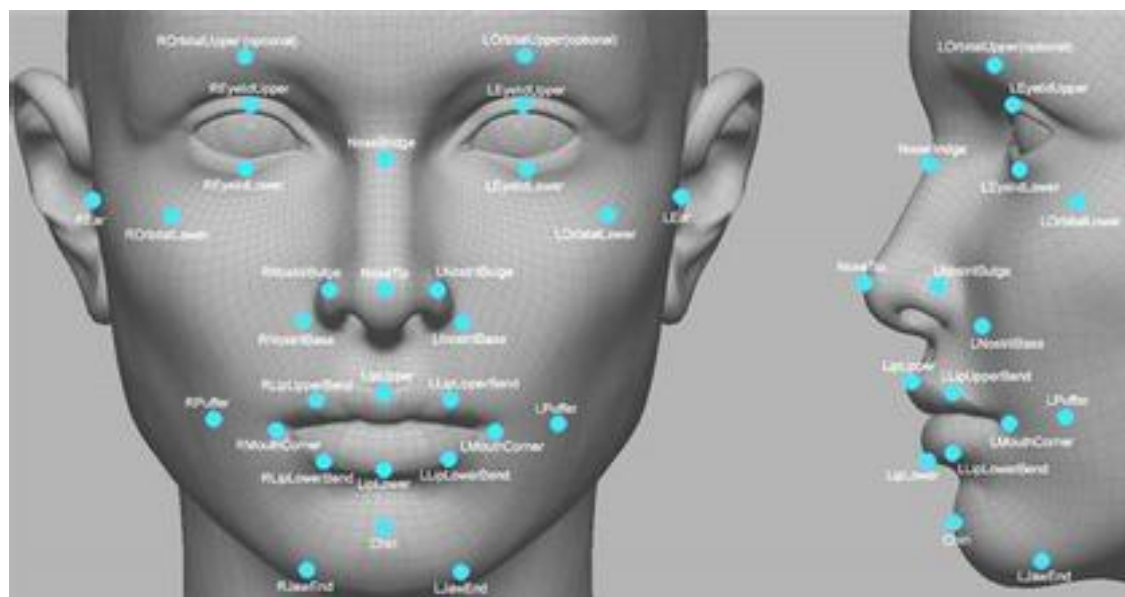




2.4.2 人工智能关键技术

6.生物特征识别

生物特征识别技术涉及的内容十分广泛，包括指纹、掌纹、人脸、虹膜、指静脉、声纹、步态等多种生物特征，其识别过程涉及到图像处理、计算机视觉、语音识别、机器学习等多项技术





2.4.2 人工智能关键技术

7.VR/AR

虚拟现实（VR）/增强现实（AR）是以计算机为核心的新型视听技术。结合相关科学技术，在一定范围内生成与真实环境在视觉、听觉、触感等方面高度近似的数字化环境。



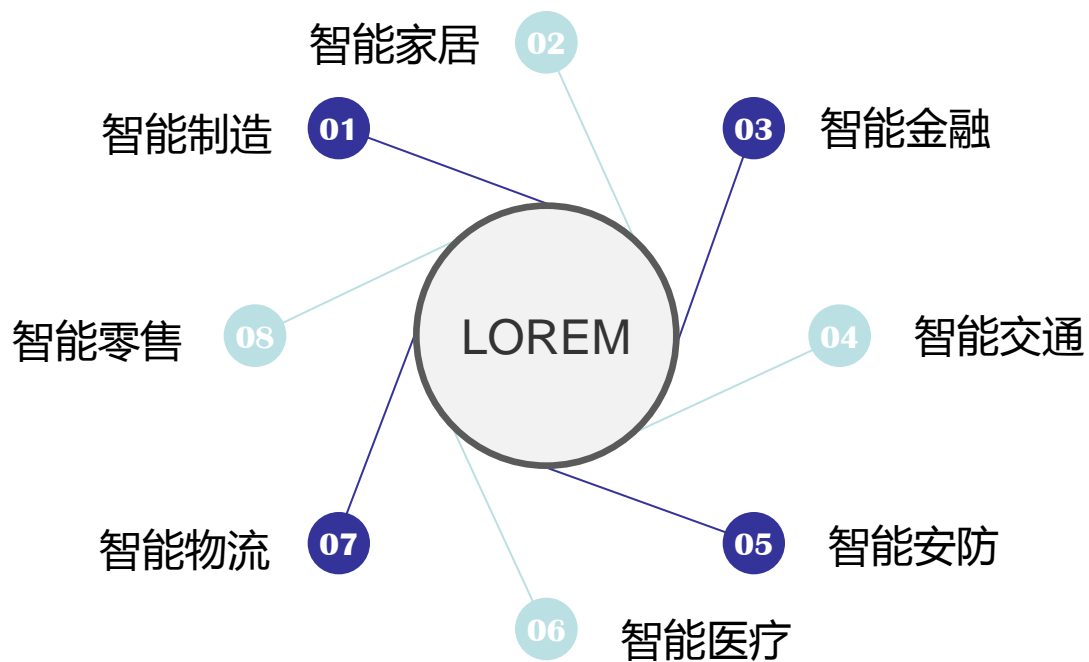
采用虚拟现实技术的虚拟弓箭



利用Tilt Brush在虚拟三维空间里绘画



2.4.3 人工智能的应用





2.4.3 人工智能的应用

1. 智能制造

智能制造（Intelligent Manufacturing, IM）是一种由智能机器和人类专家共同组成的人机一体化智能系统，它在制造过程中能进行智能活动，诸如分析、推理、判断、构思和决策等





2.4.3 人工智能的应用

2.智能家居

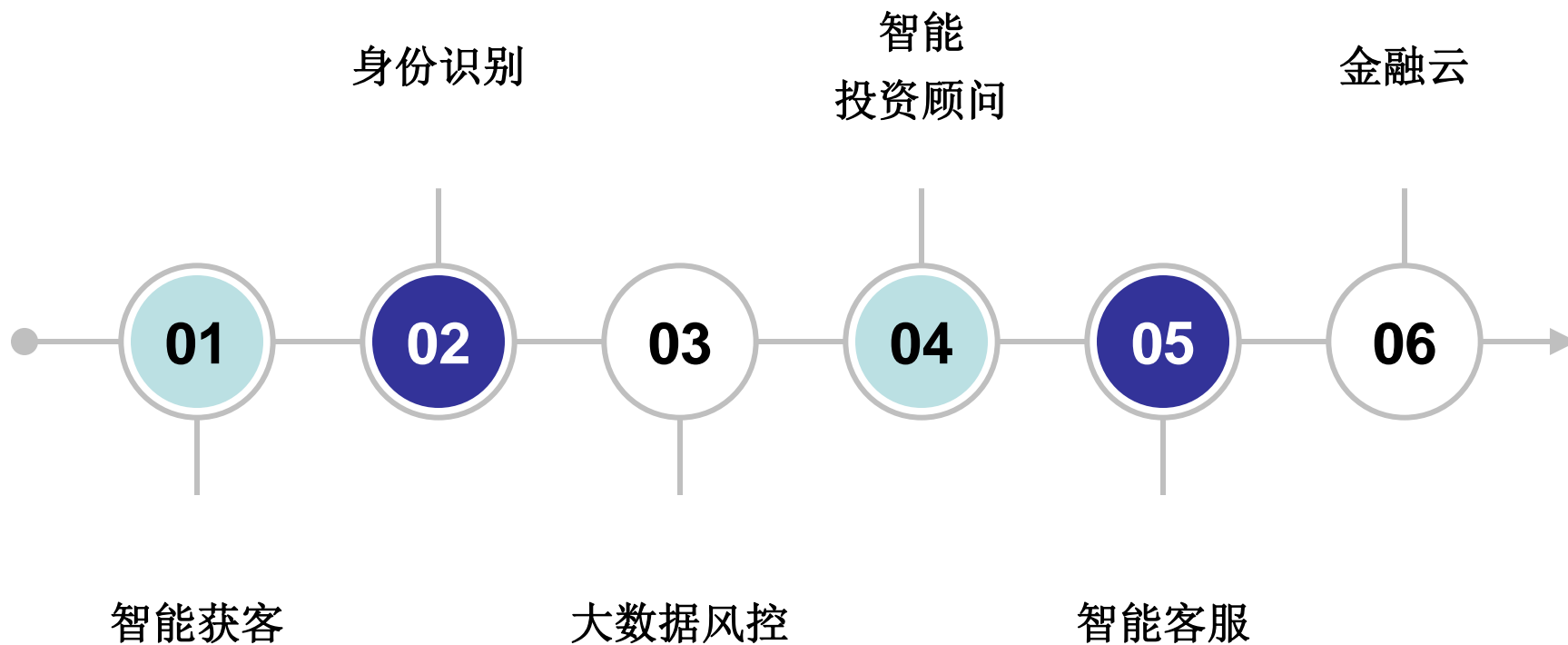
智能家居通过物联网技术将家中的各种设备（如音视频设备、照明系统、窗帘控制、空调控制、安防系统、数字影院系统、影音服务器、影柜系统、网络家电等）连接到一起，提供家电控制、照明控制、电话远程控制、室内外遥控、防盗报警、环境监测、暖通控制、红外转发以及可编程定时控制等多种功能和手段





2.4.3 人工智能的应用

3. 智能金融

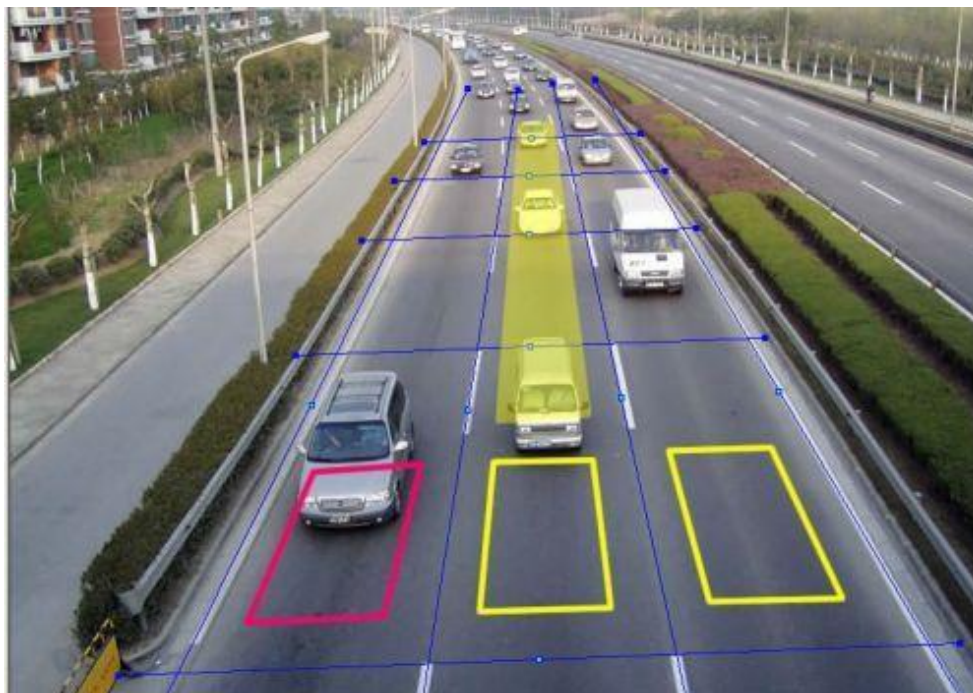




2.4.3 人工智能的应用

4. 智能交通

智能交通是未来交通系统的发展方向，它是将先进的信息技术、数据通讯传输技术、电子传感技术、控制技术及计算机技术等有效地集成运用于整个地面交通管理系统而建立的一种在大范围内、全方位发挥作用的，实时、准确、高效的综合交通运输管理系统。





2.4.3 人工智能的应用

5. 智能安防

- 智能安防是兼顾了整体城市管理系统、环保监测系统、交通管理系统、应急指挥系统等应用的综合体系
- 特别是车联网的兴起，在公共交通安全管理上、车辆事故处理上、车辆偷盗防范上可以更加快捷准确的跟踪定位处理
- 还可以随时随地的通过车辆获取更加精准的灾难事故信息、道路流量信息、车辆位置信息、公共设施安全信息、气象信息等等信息来源



2.4.3 人工智能的应用

6. 智能医疗

借助于物联网/云计算技术、人工智能的专家系统、嵌入式系统的智能化设备，可以构建起完善的物联网医疗体系，使全民平等地享受顶级的医疗服务，解决或减少由于医疗资源缺乏，导致看病难、医患关系紧张、事故频发等现象





2.4.3 人工智能的应用

7. 智能物流

使用智能搜索、推理规划、计算机视觉以及智能机器人等技术，实现货物运输过程的自动化运作和高效率优化管理，提高物流效率



图 京东智能分拣



2.4.3 人工智能的应用

8. 智能零售

人工智能在零售领域的应用已经十分广泛，无人便利店、智慧供应链、客流统计等等都是的热门方向





2.4.4 人工智能产业





2.4.4 人工智能产业

1. 智能基础设施建设





2.4.4 人工智能产业

2. 智能信息及数据

在人工智能数据采集、存储、处理和分析方面的企业主要有两种：

- 一种是数据集提供商
- 另一种是数据采集、存储、处理和分析综合性厂商



2.4.4 人工智能产业

3. 智能技术服务

- 提供人工智能的技术平台和算法模型
- 提供人工智能的整体解决方案
- 提供人工智能在线服务



2.4.4 人工智能产业

4. 智能产品

- 智能机器人
- 智能运载工具
- 智能终端
- 自然语言处理
- 计算机视觉
- 生物特征识别
- VR/AR
- 人机交互



2.5 大数据与人工智能的关系

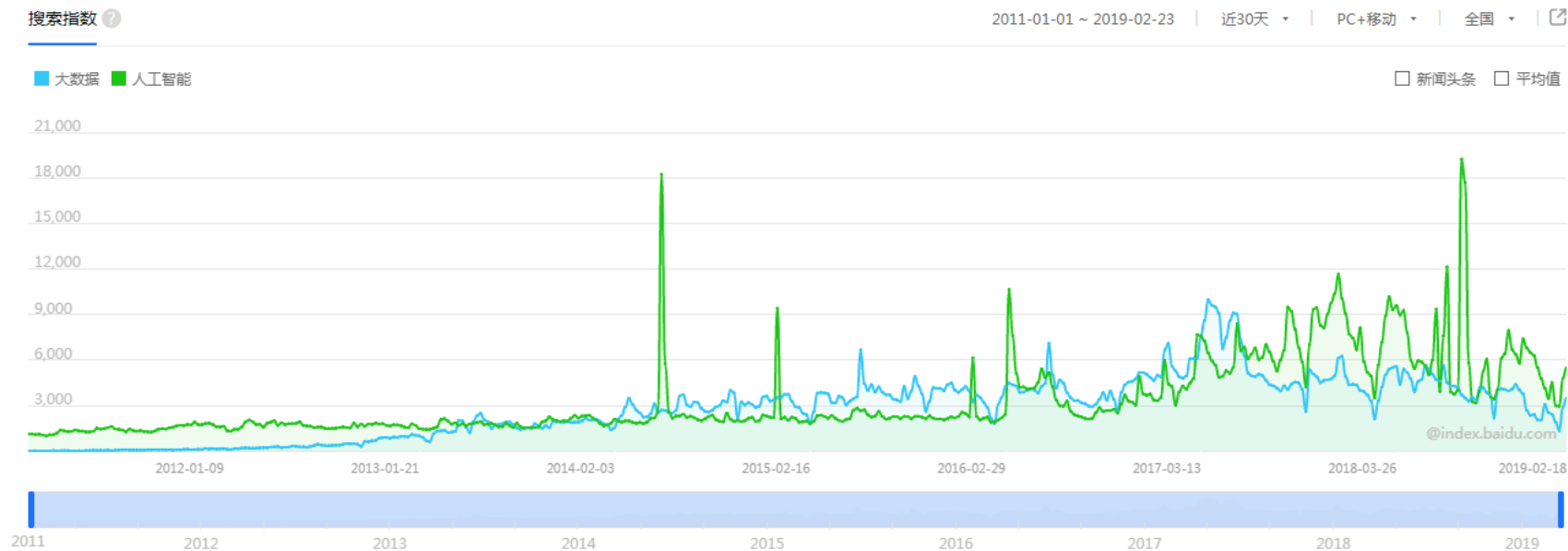


图 人工智能和大数据的百度指数



2.5 大数据与人工智能的关系

1. 人工智能与大数据的联系

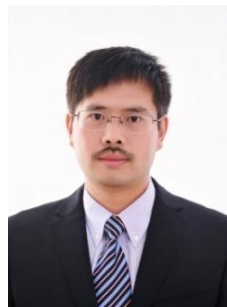
- 人工智能需要数据来建立其智能，特别是机器学习
- 大数据技术为人工智能提供了强大的存储能力和计算能力

2. 人工智能与大数据的区别

- 人工智能与大数据也存在着明显的区别，人工智能是一种计算形式，而大数据是一种传统计算，它不会根据结果采取行动，只是寻找结果
- 二者要达成的目标和实现目标的手段不同



附录A：主讲教师林子雨简介



主讲教师：林子雨

单位：厦门大学计算机科学系

E-mail: ziyulin@xmu.edu.cn

个人网页: <http://dblab.xmu.edu.cn/post/linziyu>

数据库实验室网站: <http://dblab.xmu.edu.cn>



扫一扫访问个人主页

林子雨，男，1978年出生，博士（毕业于北京大学），现为厦门大学计算机科学系助理教授（讲师），曾任厦门大学信息科学与技术学院院长助理、晋江市发展和改革委员会副局长。中国计算机学会数据库专业委员会委员，中国计算机学会信息系统专业委员会委员。国内高校首个“数字教师”提出者和建设者，厦门大学数据库实验室负责人，厦门大学云计算与大数据研究中心主要建设者和骨干成员，2013年度和2017年度厦门大学教学类奖教金获得者，荣获2017年福建省精品在线开放课程、2018年厦门大学高等教育成果特等奖、2018年福建省高等教育教学成果二等奖、2018年国家精品在线开放课程。主要研究方向为数据库、数据仓库、数据挖掘、大数据、云计算和物联网，并以第一作者身份在《软件学报》《计算机学报》和《计算机研究与发展》等国家重点期刊以及国际学术会议上发表多篇学术论文。作为项目负责人主持的科研项目包括1项国家自然科学基金青年基金项目(No.61303004)、1项福建省自然科学基金项目(No.2013J05099)和1项中央高校基本科研业务费项目(No.2011121049)，主持的教改课题包括1项2016年福建省教改课题和1项2016年教育部产学协作育人项目，同时，作为课题负责人完成了国家发改委城市信息化重大课题、国家物联网重大应用示范工程区域试点泉州市工作方案、2015泉州市互联网经济调研等课题。中国高校首个“数字教师”提出者和建设者，2009年至今，“数字教师”大平台累计向网络免费发布超过500万字高价值的研究和教学资料，累计网络访问量超过500万次。打造了中国高校大数据教学知名品牌，编著出版了中国高校第一本系统介绍大数据知识的专业教材《大数据技术原理与应用》，并成为京东、当当网等网店畅销书籍；建设了国内高校首个大数据课程公共服务平台，为教师教学和学生学习大数据课程提供全方位、一站式服务，年访问量超过100万次。



附录B：大数据学习路线图



大数据学习路线图访问地址：<http://dblab.xmu.edu.cn/post/10164/>



附录C： 《大数据技术原理与应用》 教材

《大数据技术原理与应用——概念、存储、处理、分析与应用（第2版）》，由厦门大学计算机科学系林子雨博士编著，是国内高校第一本系统介绍大数据知识的专业教材。人民邮电出版社 ISBN:978-7-115-44330-4 定价：49.80元



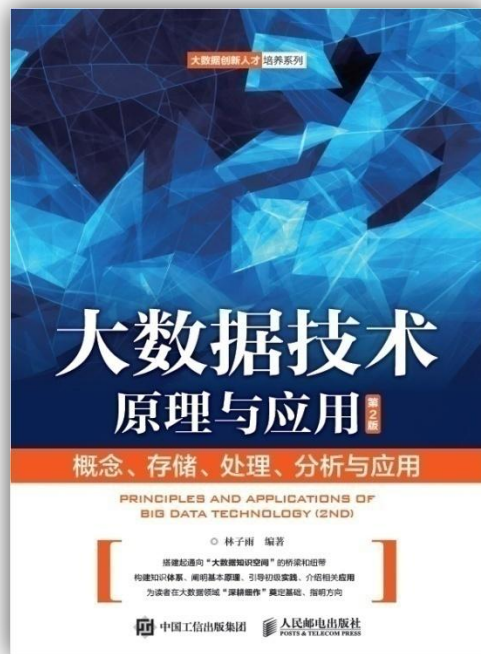
扫一扫访问教材官网

全书共有15章，系统地论述了大数据的基本概念、大数据处理架构Hadoop、分布式文件系统HDFS、分布式数据库HBase、NoSQL数据库、云数据库、分布式并行编程模型MapReduce、Spark、流计算、图计算、数据可视化以及大数据在互联网、生物学和物流等各个领域的应用。在Hadoop、HDFS、HBase和MapReduce等重要章节，安排了入门级的实践操作，让读者更好地学习和掌握大数据关键技术。

本书可以作为高等院校计算机专业、信息管理等相关专业的大数据课程教材，也可供相关技术人员参考、学习、培训之用。

欢迎访问《大数据技术原理与应用——概念、存储、处理、分析与应用》教材官方网站：

<http://dbl原因.xmu.edu.cn/post/bigdata>





附录D：《大数据基础编程、实验和案例教程》

本书是与《大数据技术原理与应用（第2版）》教材配套的唯一指定实验指导书

大数据教材



1+1黄金组合
厦门大学林子雨编著

配套实验指导书



- 步步引导，循序渐进，详尽的安装指南为顺利搭建大数据实验环境铺平道路
- 深入浅出，去粗取精，丰富的代码实例帮助快速掌握大数据基础编程方法
- 精心设计，巧妙融合，五套大数据实验题目促进理论与编程知识的消化和吸收
- 结合理论，联系实际，大数据课程综合实验案例精彩呈现大数据分析全流程

清华大学出版社 ISBN:978-7-302-47209-4 定价：59元



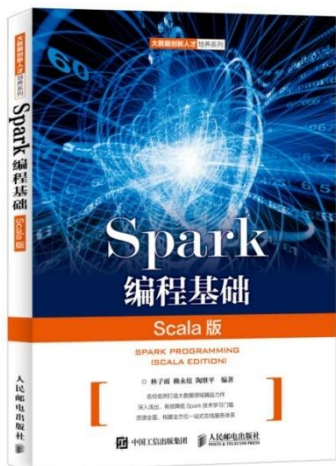
附录E：《Spark编程基础（Scala版）》

《Spark编程基础（Scala版）》

厦门大学 林子雨，赖永炫，陶继平 编著

披荆斩棘，在大数据丛林中开辟学习捷径
填沟削坎，为快速学习Spark技术铺平道路
深入浅出，有效降低Spark技术学习门槛
资源全面，构建全方位一站式在线服务体系

人民邮电出版社出版发行，ISBN:978-7-115-48816-9
教材官网：<http://dmlab.xmu.edu.cn/post/spark/>



本书以Scala作为开发Spark应用程序的编程语言，系统介绍了Spark编程的基础知识。全书共8章，内容包括大数据技术概述、Scala语言基础、Spark的设计与运行原理、Spark环境搭建和使用方法、RDD编程、Spark SQL、Spark Streaming、Spark MLlib等。本书每个章节都安排了入门级的编程实践操作，以便读者更好地学习和掌握Spark编程方法。本书官网免费提供了全套的在线教学资源，包括讲义PPT、习题、源代码、软件、数据集、授课视频、上机实验指南等。



附录F：高校大数据课程公共服务平台



高校大数据课程

公 共 服 务 平 台

<http://dblab.xmu.edu.cn/post/bigdata-teaching-platform/>



扫一扫访问平台主页



扫一扫观看3分钟FLASH动画宣传片



附录G：高校大数据实训课程系列案例教材

为了更好地满足高校开设大数据实训课程的教材需求，厦门大学数据库实验室林子雨老师团队联合企业共同开发了《高校大数据实训课程系列案例》，目前已经完成开发的系列案例包括：

《基于协同过滤算法的电影推荐》

《电信用户行为分析》

《实时日志流处理分析》

《微博用户情感分析》

《互联网广告预测分析》

《网站日志处理分析》

系列案例教材将于2019年陆续出版发行，教材相关信息，敬请关注网页后续更新！

<http://dbllab.xmu.edu.cn/post/shixunkecheng/>



扫一扫访问大数据实训课程系列案例教材主页

The background of the slide features a blue gradient with several faint, light-blue silhouettes of people. At the top, there are two groups of people standing and holding hands. On the right side, a person is shown in profile, looking towards the center. On the left side, two people are shown in profile, one appearing to be speaking or gesturing towards the other. The overall theme is one of community and collaboration.

Thank You!

Department of Computer Science, Xiamen University, 2019