



Introduction to AI and Big Data

賈維嘉講座教授
智慧城市物聯網國家重點實驗室副主任
數據科學中心主任



提纲

- 1 人工智能
- 2 大数据应用
- 3 智慧城市物联网国重@UM
- 4 小结

提纲

- 1 人工智能
- 2 大数据应用
- 3 智慧城市物联网国重@UM
- 4 小结

人工智能概述

- 1.1 人工智能 (AI) 定义
- 1.2 AI简史
- 1.3 AI研究
- 1.4 AI构成

作为智能体的人类

- 智能体(Agent)
- 我们为什么能够思考?
- 怎么感知、理解、预测和应对庞大和复杂的世界?
- AI不仅试图理解智能体，而且想建造智能体!?

1.1 什麼是AI?

- AI两个目标：
- 1、近期(理性思维/行动)：研究人脑功能
 - 判断、理解和推理、形成概念、适当的反应和适应环境的总体能力，用计算机去模拟它，使计算机更聪明、更象人。
- 2、远期 (类人思考/行动)：研究人脑的微观结构和宏观功能
 - 制造出和人脑的结构一致的智能机，能完成人脑的宏观功能。

1.1 什麼是AI?

- 两大学派：
- (1) 数学：求解智能全靠算法 (Algorithm)，求解一切智能问题都建立数学模型并利用算法来求解问题。
- (2) 心理学：依靠心理学模型，把人在解决各种问题时所使用的经验、方法、策略、窍门等都编成启发式程序来求解智能问题。
- 应该使用以上两个智能算法和启发式程序共同工作。

探索初期的学科

- AI是新兴和激动人心的学科. Russell声称：不同于物理学，这里还可能出现几个爱因斯坦
- Why? 研究主观世界的成果远少于研究客观世界的成果

AI -- 4种定义

像人一样	理性地思考的系统—机器学习
要使计算机能思考……有头脑的机器(Haugeland, 1985) [使之自动化]与人类的思维相关的活动,诸如决策、问题求解、学习等活动(Bellman, 1978)	通过对计算模型的使用来进行心智能力的研究 对使得知觉、推理和行动成为可能的计算的研究
像人一样行动	理性地行动的系统
创造机器来执行人需要智能才能完成的功能(Kurzweil, 1990) 研究如何让计算机能够做到那些目前人比计算机做得更好的事情(Rich & Knight, 1991)	计算智能是对设计智能化智能体的研究(Poole et al., 1998) AI关心的是人工制品中的智能行为(Nilsson, 1998)

4种不同定义

- 像人一样行动的系统 — 类人行为
- 像人一样思考的系统 — 类人思考
- 理性地思考的系统 – 机器学习
- 理性地行动的系统

4种不同定义 (1)

- 类人行为：图灵测试(1950)（理性地思考）
- 图灵建议：不是问“机器能否思考”，而是问“机器能否通过关于行为的智能测试”
- 测试过程：让机器与人进行5分钟对话 / 猜测交谈对象是程序还是人？如果在30%测试中程序成功地欺骗了询问人，则通过了测试
- 图灵期待最迟2000年出现这样的程序，今天面对训练有素的鉴定人，没有一个程序接近30%的标准

4种不同定义 (2)

- 要想通过图灵测试，还需：
 - 自然语言处理 (NLP)，机器可以用人类语言交流
 - 知识表示，存储机器获得的各种信息（例如知识图谱）
 - 自动推理，运用知识来回答问题和提取新结论
 - 机器学习，适应新环境并检测和推断新模式
 - 计算机视觉，机器感知物体
 - 机器人技术，操纵和移动物体
 - 软、硬件自我保护、修复技术

4种不同定义(3)

- AI研究者并未止步：研究智能的根本原则远比复制样本重要.
- 如同飞机的产生



4种不同定义 (4)

- **类人思考:**
- **认知模型方法:** 人类是如何思考的? 通过自省—捕捉人类思维过程和通过心理测试
 - 关心对程序的推理步骤轨迹与人类个体进行比较
- **认知科学:** AI模型与心理学相结合, 创立一种精确而且可检验的人类思维工作方式的理论
- **关心程序实现了什么功能, 不关心AI技术和人类认知的异同**



4种不同定义 (4)

- 理性地思考：“思维法则”方法
- 19世纪, 发现了描述世界上一切事物及其彼此关系的命题符号；1965年, 程序可以求解任何用逻辑符号描述的可解问题(消解法)
- AI逻辑主义希望通过编制上述程序来创造智能系统
- **难点:** 非形式化的知识难以用逻辑符号形式化 / “原则上”可解决问题≠实际解决问题



4种不同定义 (5)

- **理性智能体：**
 - 有别于“简单的”程序：具有自控操作、感知环境、适应变化...
 - 通过自己的行动获得最佳期望结果
 - 正确地推理和行动是理性智能体的部分功能，而不是理性的全部内容

概念理解

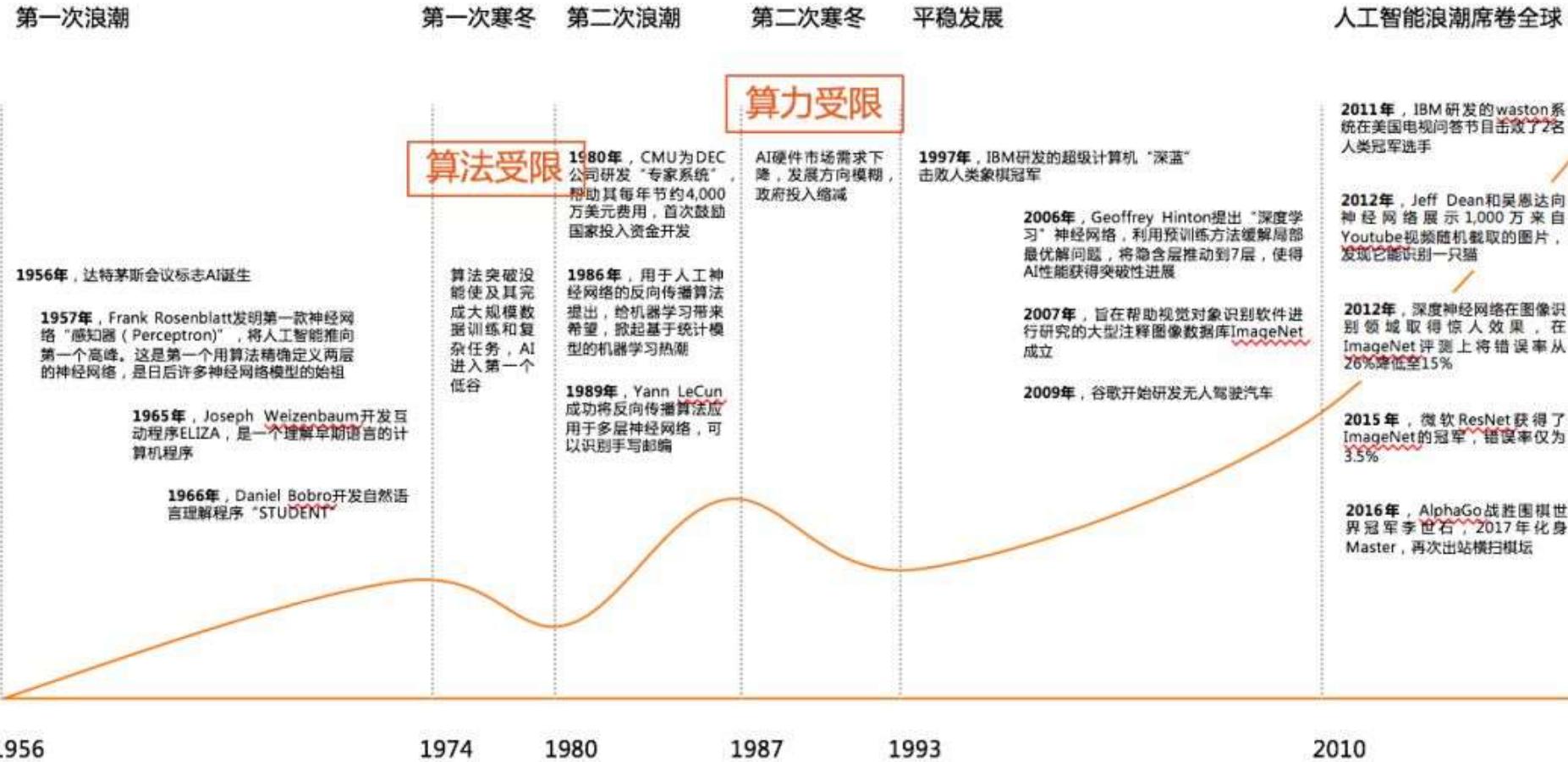
- 上述定义见仁见智
- 重要的是学习AI方法、在实践中逐步深入领会AI这个词的含义
- AI就是一种运行在我们自己机器中的程序，它的智能都是我们给的！



AI简史

7个时期：
孕育期/诞生/早期的成功与期望
困难期/知识系统崛起
AI工业/AI科学

AI六十多岁了！ -- 还很年轻



时期	时间	事件
人工智能“推理”时代	1942年	美国科幻巨匠阿西莫夫提出“机器人三定律”
	1956年	达特茅斯会议标志人工智能诞生
	1963年	自动定理证明系统完成数学原理第二章的证明
	1968年	DENDRAL世界上第一例专家系统诞生
人工智能“知识工程”时代	1969年	早期的系统适用于更宽的问题选择和更难的问题时效果均不理想， 美国、英国相继缩减经费支持
	1983年	莱斯利提出概率近似正确模型
	1986年	DE Rumelhart等人发明了BPs算法
	1992年	第5代计算机（人工智能计算机）由于技术线路背离计算机工业发展方向， 项目宣告失败
人工智能“数据挖掘”时代	2006年	Hinton提出了深度学习的神经网络
	2012年	CNN超第二名10个百分点夺冠ImageNet
	2014年	港中文实验室DeepID算法首次超过人眼识别人脸率
	2016年3月	谷歌Alpha Go以4:1战胜人类顶尖围棋选手李世石
	2016年9月	美国五大科技巨头（谷歌、微软、Facebook、亚马逊、IBM）成立AI联盟。斯坦福大学发布人工智能百年研究计划
	2017年3月	“人工智能”首次写入政府工作报告
	2017年5月	AlphaGo 2.0以3:0完胜围棋现排名第一的柯洁九段
	2017年7月	国务院出台《新一代人工智能发展规划》， 中国人工智能产业创新联盟成立
	2017年11月	科技部公布人工智能四大平台（百度、阿里、腾讯、科大讯飞）
	2017年12月	AlphaGo的又一版本，AlphaZero闪亮登场

8个时期

- AI～六十年：
 - 孕育期(1943~1955)
 - 诞生(1956)
 - 热情与期望(1952~1969)
 - 现实的困难(1966~1973)
 - 基于知识的系统: 钥匙? (1969~1979)
 - 成为工业(1980~现在)
 - 成为科学(1987~现在~未来) / 神经网络的回归 (1986~现在) / 智能体出现(1995~现在)
 - (深度) 机器学习 (成果或者又一个泡沫?)

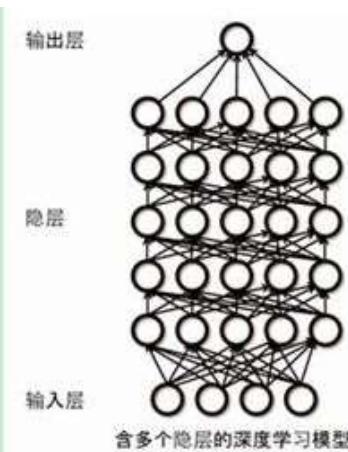
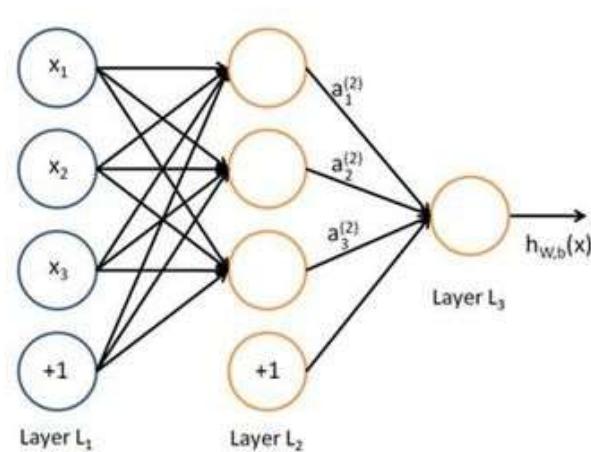
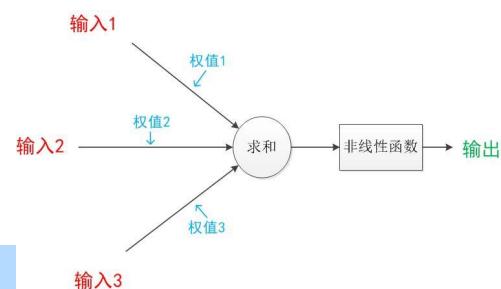
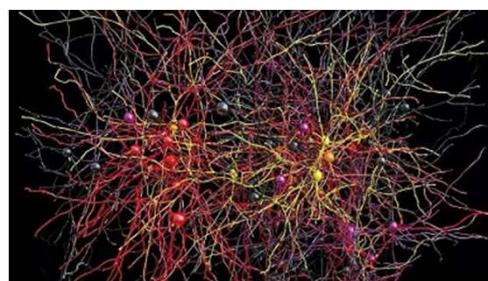
1.2 AI简史

- 1. 孕育期(1956年以前)
 - (1) 公元前4世纪的希腊哲学家Aristotle的形式逻辑：三段论：
 - 既析取： $\neg P \wedge (P \vee Q) \Rightarrow Q$
 - 假言三段论： $(P \rightarrow Q) \wedge (Q \rightarrow R) \Rightarrow P \rightarrow R$ ，至今仍然是演绎推理的基本依据。
 - (2) 英国哲学家培根 (F. Bacon) 提出了归纳法，喊出“知识就是力量”，
 - 对70年代AI转向以知识为中心的研究都产生了很重要的影响。

1.2 孕育期(1943~1955)

- 神经网络

- 1943年Warren McCulloch和Walter Pitts人工神经元模型 ==> 可计算的函数 ==> 神经元连接成的网络 ==> 网络能够学习
- 1951年，普林斯顿大学数学系研究生Marvin Minsky(明斯基)和Dean Edmonds建造了第一台神经元网络计算机



图灵的论文

- (3)德国数学家莱不尼茨 (G.Leibniz)
 - 万能符号和推理计算 ==》 数理逻辑的产生和发展。
- (4)英国逻辑学家布尔 (G.Boole)
 - 创立了布尔代数。
- (5)英国数学家图灵
 - 1936年提出了一种理想计算机模型—图灵机。为后来数字计算机的问世奠定了理论基础。图灵1950年的论文第一个清晰地描绘出AI的完整图像
- (6)美国生理学家麦克洛奇 (W.McCulloch)与匹兹 (W.Pitts)
 - 1943年建成了第一个神经网络模型 (M-P模型) 为后来人工神经网络的研究奠定了基础。
- (7) 美国数学家莫克利 (J.W.Mauchly)和埃柯特 (J.P.Eckert)
 - 1946年研制出了世界上第一台电子数字计算机ENIAC为AI的研究奠定了物质基础

AI诞生(1956)(1)

56年夏, AI正式诞生于达特茅斯大学

- John McCarthy(麦卡锡)自普林斯顿大学毕业以后去了达特茅斯大学,说服了另外2个人帮助召开了为期2个月的研讨会
- 会议组织者4人: 麦卡锡、 Minsky(明斯基)、 Claude Shannon(香侬)、 IBM 的Nathaniel Rochester(罗切斯特), 参加者共10人
- 其他6位是: 普林斯顿大学摩尔、 IBM的Arthur Samuel(塞缪尔)、 MIT的Ray Solomonoff和Oliver Selfridge、 CMU的纽厄尔和西蒙



2006 年, 会议 50 年后, 当事人重聚达特茅斯 (左起: 摩尔、麦卡锡、明斯基、塞弗里奇、所罗门诺夫)

AI诞生(1956)(2)

- 历史意义的会议，AI学科正式诞生。主要成果：

- (1) 机器学习：塞谬尔1956年研制出了一个跳棋程序。这个程序能从棋谱中学习。1959年它击败了塞谬尔，1962年它又击败了一个州的冠军。（遗憾的是从此以后它没再取胜过）。
- (2) 定理证明：美籍华人王浩在1960年在机器上实现了命题逻辑的证明问题。被命名为王浩算法。



AI诞生(1956)(3)

- 纽厄尔和西蒙最为活跃,介绍了他们的推理程序:逻辑理论家
- 聚集了AI的主要人物,也成为了美国AI研究的3大基地:
 - MIT—明斯基
 - Stanford—麦卡锡(先在MIT后去了Stanford)
 - CMU—纽厄尔和西蒙
 - IBM

AI诞生(1956)[4]

- 最为长久的贡献就是麦卡锡起的名字：“人工智能” (Artificial Intelligence--AI)
- 为什么AI成为一个新领域?
 - 目标不同：一开始就承载着复制人的创造性、自我修养、语言功能等思想，没有任何一个其他领域涉及这些问题
 - 方法论不同：是唯一一个明确属于计算机科学的分支，因非数学或者控制论或其他学科的分支
 - AI试图建造在复杂和变化的环境中自动发挥功能的机器

早期的期望 (1952~1969)(1)

- CMU: 纽厄尔和西蒙通过心理学实验总结出了通用问题求解程序GPS (General Problem Solving),
 - 成功运用物理符号系统假设, 用来求解11种不同类型的问题。
- 例如: 中国的人工智能学家吴文俊先生研究的初等几何定理证明问题, 可解决初等几何中所有定理证明问题, 还可推广到欧式几何中去。

早期的期望 (1952~1969)(2)

- IBM: 1959建造了几何定理证明机;
- 1952年起, 塞缪尔的西洋跳棋程序, 通过学习可达业余高手的级别
- MIT: 1958年麦卡锡作出了三项重要贡献
 - 1: 定义LISP表结构语言, 曾经是建造智能系统的重要工具。
 - 2: 与MIT其他人发明了分时技术
 - 3: 发表《Program with Common Sense》的论文--实现了知识表示和推理的中心原则: 具备明确的知识表示, 并能通过演绎过程处理这些表示

早期的期望(1952~1969)(3)

- Stanford: 1963年麦卡锡启动了斯坦福的AI实验室==>逻辑推理通用方法以及机器人研究
- MIT: 1958年明斯基发展出反逻辑的观点
 - 贡献: 积木世界==> 视觉项目、自然语言理解项目(Terry Winograd)、规划器等
- 模式识别, 1959年塞尔夫里奇推出了一个模式识别程序, 1965年罗伯特编制出了一个可分辨积木构造的程序



(1952~1969)(4)专家系统：

美国斯坦福大学的费根鲍姆1968年研制成功了专家系统DENDRAL。

- 一个实用的智能系统，
- 探索知识的表示、存储、获取、推理及利用等技术，对AI的发展产生了深刻的影响，远远超出了系统本身的价值。
- （专家系统使AI走出低迷时期面向向实用阶段…）

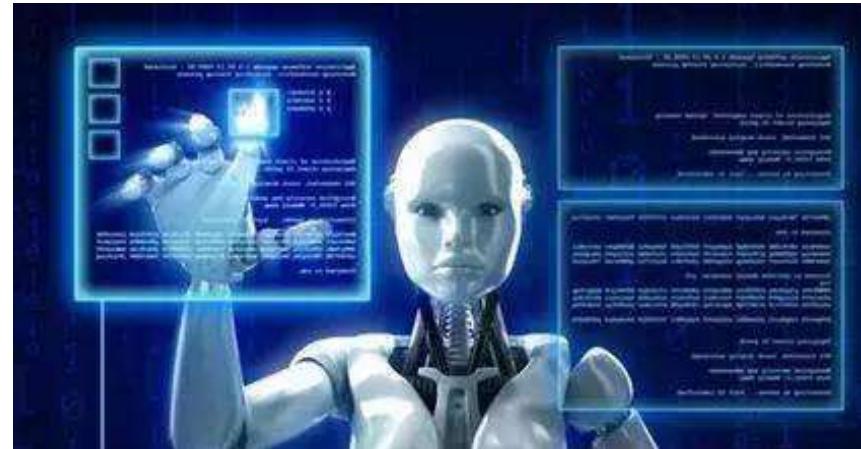
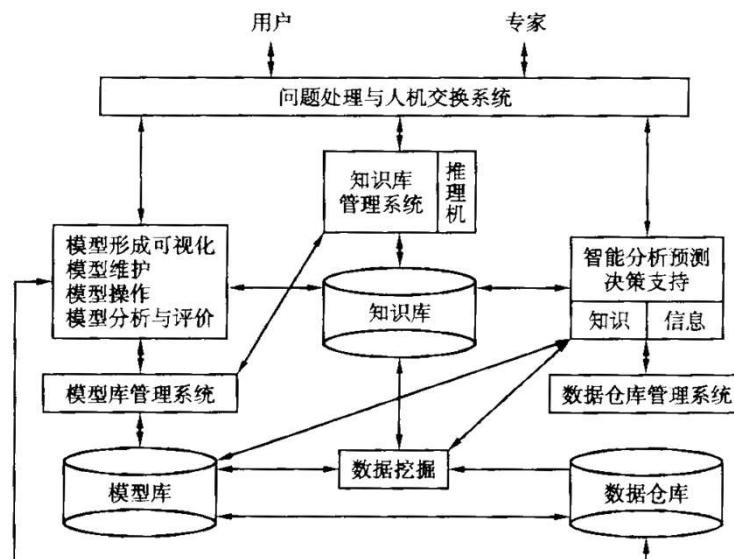


图 1 维修智能决策系统结构

早期的期望(1952~1969)(5)

- 1969国际人工智能联合成立会 (International Joint Conferences on Artificial Intelligence,简称IJCAI),
- 每两年召开一次国际学术交流会
- 2019 年夏在澳门召开

AI寒冬(1966~1973)(1)

- AI研究者过于乐观
 - AI系统在试图解决更宽范围和更难的问题时,都悲惨地失败了。
- 原因何在?
 - 缺少主题知识(通用而非专门化)
 - 典型例子: 机器翻译(MT) -- 最早对AI研究的发难始于机器翻译
 - 时至今日, MT==》 不完善但是被广泛期待, 作为一种辅助文档处理工具 (e.g. Google 翻译)

AI寒冬(1966~1973)(2)

- 第二类困难：AI试图解决的很多问题是不可操作的(NP类)
- 计算复杂性理论，“问题放大”(从玩具到现实)的认识局限于速度和存储容量
- 例子：包含超过几十条事实的定理证明--早期遗传算法实验(1958~59)
- 无限计算能力的幻觉：原则上能解≠实际上有找到解的机制
- 1973年英国政府终止了除2所大学以外所有的AI研究资助

AI寒冬(1966~1973)(3)

- 第三类困难：基本结构存在限制
 - 例子：1969年Minsky和Papert证明了简单的神经网络能表示的东西很少(单层感知器无法解决XOR函数)
 - 神经网络研究由此沉寂了20年，直到80年代后期多层网络的反向传播算法出现引起了神经网络的复兴

发展期 (1969~1979)(1)

- 早期的通用搜索机制称为弱方法，通用但不能扩展到大规模/困难问题
- 需要更强有力的、领域相关的知识
 - DENDRAL
 - 根据质谱仪信息推断分子结构
 - 知识和推理分开
 - 80年代专家系统的典型结构

发展期 (1969~1979) (2)

- 由DENDRAL起，
- MYCIN—检测血液感染的专家系统
 - MYCIN知识库的特点: 直接来自经验 / 反映出知识的不确定性
- 自然语言理解领域的专家系统:
 - 耶鲁大学Roger Schank和其学生们开发的一系列程序 (1977~1983)

发展期（1970年以后）

- 斯坦福大学的肖特里菲于1972年研制成功用于治疗和诊断感染性疾病的专家系统MYCIN。
- 1970年国际人工智能杂志（Artificial Intelligence）创刊，推动了AI发展和交流。
- 1972年法国马赛大学研制逻辑程序设计语言Prolog(Programming in logic的缩写)。
- 日本的第五代机使用Prolog。

发展期（1970年以后）

- AI又迎来了蓬勃发展的新时期，即以知识为中心的时期 - 系统如雨后春笋！
- 例如，地矿勘探专家系统PROSPECTOR，拥有15种矿藏知识，能根据岩石标本及地质勘探数据对矿藏资源进行估计和预测，对矿床、储藏量、品位、开采价值等进行推断，制定合理的开采方案，成功地找到了超亿美元的钼矿。
- 专家系统MYCIN则能识别51种病菌，正确使用23种抗菌素，协助医生诊断、治疗细菌感染性血液病。
- 内科诊断专家系统CADUCEUS正确地诊断出了许多疑难病症。还有XCON计算机配置专家系统，及信用卡认证决策专家系统等。

发展期（1970年以后）

- 1977年费根鲍姆在第五届IJCAI上提出了“知识工程”。
- 中国1978年把“智能模拟”作为国家科学技术发展规划。
- 1981年成立了中国人工智能学会（CAAI）
- 2016年后美国、欧盟、英国、日本、德国均将人工智能上升为国家战略。
- 2017年国务院印发《新一代人工智能发展规划》，人工智能正式上升为国家战略；
- 2018十三届全国人大一次会议发布《政府工作报告》，人工智能再次被写入。

发展期（1970年以后）

- 时代在发展，人类在进步。网络Internet的出现对人工智能也提出了新的需求，如电子商务：电子签证；网上购物、售票、网上银行、求医、求学等等。
- 神经网络技术，中间件技术（Agent技术），
- 人脸，语音识别等等都是人工智能发展的产物。

AI成为工业(1980~现在)(1)

- 1982年, DEC第一个商用专家系统R1开始运转, 到1986年每年为公司节省4千万美元
 - AI工业在1980年只是几百万美元, 1988年涨到数十亿美元
 - 以知识为中心: 知识表示、利用、获取等的研究取得了较大的进展,
 - 不确定性知识的表示与推理: 主观Bayes理论、确定性理论、证据理论、可能性理论等 (模式识别, 自然语言理解等)。

AI成为工业(1980~现在)(2)

- 在八十年代的AI研究热潮中，1981年日本提出五代机计划，目的是建造运行Prolog程序的智能机
- 美国则对应成立了MCC研究集团
- 其中的AI部分从未实现其野心勃勃的目标
- 实际上，“AI成为工业”目前在一些家电中可以找到影子(智能洗衣机等)

AI成为工业(1980~)(3)

很快又进入了“AI的冬天”：

- 塞谬尔的下棋程序自从胜了一个州的冠军后，再没赢过；
- 1996年国际象棋程序“深蓝”赢了它的设计人，但1997年深蓝失败；机器翻译也出现了一些问题。
- 美国和日本等国家对人工智能研究的第一个十年计划遭到了重挫。其中日本投资540亿日元，想要研制出具有人脑结构的智能机宣告失败。
- 费根鲍姆关于以知识为中心开展人工智能研究的观点被大多数人接受。

神经网络回归(1986~现在)

- 神经网络: Frank Rosenblatt 1962年提出感知器, 证明了感知器收敛定理, 1969年以后沉寂
- 反向传播算法引起了神经网络研究的复兴
- **连接主义方法**崛起, 被认为是 Newell 和 Simon 提出的符号模型和 McCarthy 主张的逻辑方法的直接竞争者
- 当前的观点是: 连接主义和符号主义方法是互补的

AI成为科学(1987~现在)(1)

- AI研究在内容和方法论方面的特点：
 - 在已有的理论基础上进行研究
 - 理论建立在严格定理或者确凿实验证据基础上而不是靠直觉
 - 与现实世界应用的相关性而不是与玩具样例的相关性

AI成为科学(1987~现在)(2)

- 从对控制论和统计学的某种叛逆到开始接受这些领域的理论和方法
- 通过互连网进行测试数据和程序代码的共享
- 典型：语音识别中HMM模型应用 / 贝叶斯网络，深度学习等

弱AI和强AI

- 弱人工智能(Weak AI)断言：“机器能够智能地行动”
- 强人工智能(Strong AI)断言：“机器确实在思考”
- 大多数AI研究者认为弱AI假设是当然的，目前AI的成就可以证明
- 关于强AI，还是哲学上的争论

人工智能的研究领域

- 专家系统 自然语言处理
 - 机器学习 定理证明
 - 分布式人工智能 机器人
 - 模式识别 智能决策支持
 - 博弈 计算机视觉
 - 商务智能 智能体技术
 - 城市计算 智能物联网
 - 大数据处理 深度学习

1.4 对AI有贡献的学科

- 哲学(BC428~现在)
- 数学(800~现在)
- 经济学(1776~现在)
- 神经科学(1861~现在)
- 心理学(1879~现在)
- 计算机工程(1940~现在)
- 控制论(1948~现在)
- 语言学(1957~现在)

提纲

- 1 智慧城市基础：人工智能概述
- 2 大数据应用
- 3 智慧城市物联网国重@UM
- 4 小结

2.1

什么是大数据

信息技术革命的周期

- 架构化 --1950-1970
- 数字化--1970-1990
- 网络化-- 1990-2010+
- 今天： 互联网 +
- 5G移动互联网
- 云计算
- 物联网
- 大数据 + AI

大数据从哪里来？

人类社会、信息空间、物理世界都是大数据的实体空间

人类社会
Social



人类社会：
对事物看法、经验、认知

例如：
交通拥堵、环境污染、诊断、舆情等

信息空间
Cyber



信息空间：
互联网应用产生的数据

例如：
网页、邮件、即时通信、
微博、微信等

物理世界
Physical

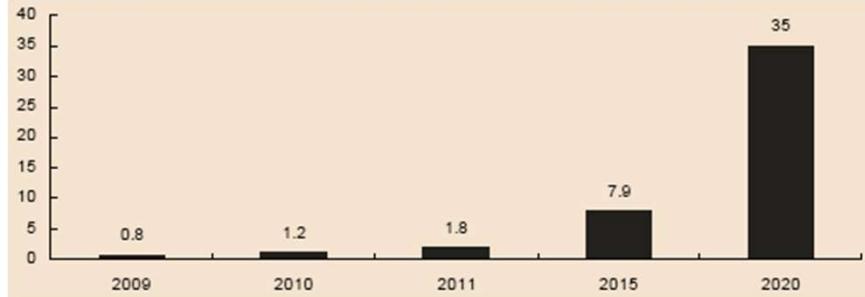


物理世界：
感知设施产生的数据

例如：
人车位置、环境PM2.5、
生产销售、资源消耗等

为什么叫“大”数据

数据量增加

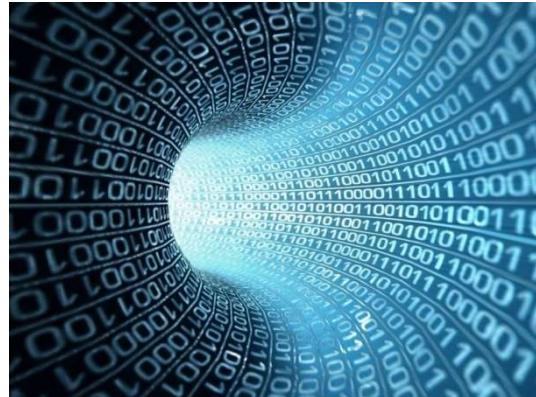


根据IDC 监测，人类产生的数据量
正在呈指数级增长，**每两年翻一番。**

TB → PB → EB → ZB

数据结构日趋复杂

非结构化,半结构化数据爆发式增长



- 人类产生的数据早已经远超了目前人力所能处理的范畴，例如：
 - 美国国会图书馆所有资料为 462 TB；
 - 2012年末全世界所有资料为 2.72 ZB。

大数据的4V

大数据是指无法在一定时间内用传统数据库软件工具对其内容进行抓取、管理和处理的数据集合

1. Volume

数据量巨大

全球10年正式进入ZB时代，
IDC预计到2020年，全球将总共拥有35ZB的数据量。

2. Variety

结构化、半结构化和非结构化数据

数据类型早已不是单一的文本形式，
订单、日志、音频，能力提出了更高的要求

3. Value

沙里淘金，价值密度低

以视频为例，一部一小时的视频，在连续不间断监控过程中，可能有用的数据仅仅只有一两秒。

4. Velocity

实时获取需要的信息

如今已是ZB时代，处理数据的效率就是企业的生命



2.2

什么是大数据：数据表示

数据表示：计算机存储单位

- 布尔: True/False (Boolean: 一个字节)
- 字符: a, b, c, ..., A, B, C, (char: 一个字节)
- (基本) 整型: -32768~32767 (int: 两个字节)
- 短整型: -32768~32767 (short[int]: 两个字节)
- 长整型: $-2^{31} \sim 2^{31}-1$ (long[int]: 4个字节)
- 浮点;
- 双精度浮点
- 枚举; 结构; 字符串
- 固定长度数组;
- 可变长度数组
- 联合

数据表示：计算机存储单位

- bit、Byte、KB、GB、TB等等都意味着多少数据呢？
- bit(比特)是binary digit的英文缩写，量度信息的单位 (0, 1)。
- 8个bit组成一个Byte(字节)，
 - 容纳一个英文字符，一个汉字需要两个字节的存储空间
- 普通英文单词则需要十个左右的字节。

数据表示：计算机存储单位换算

- 8 bit = 1 Byte 一字节
- 1024 B = 1 KB (KiloByte) 千字节
- 1024 KB = 1 MB (MegaByte) 兆字节
- 1024 MB = 1 GB (GigaByte) 吉字节
- 1024 GB = 1 TB (TeraByte) 太字节
- 1024 TB = 1 PB (PetaByte) 拍字节
- 1024 PB = 1 EB (ExaByte) 艾字节
- 1024 EB = 1 ZB (ZetaByte) 皆字节
- 1024 ZB = 1 YB (YottaByte) 佑字节
- 1024 YB = 1 NB (NonaByte) 诺字节
- 1024 NB = 1 DB (DoggaByte) 刀字节

数据表示：形象数据量

- 一张电报：100字节
- 一个笑话：1KB
- 一页书籍：10KB
- 一张照片：100KB
- 一部小说：1MB
- 一次胸透：10MB
- 一部百科全书：100MB
- 一卷磁带：200MB
- 一张光盘：500MB

数据表示

- 一部电影： 1GB
- 一卷大型数字磁带： 100GB
- 十万棵树制成的纸： 1TB
- 一套大型存储系统： 50TB
- (美国国会图书馆总量： ~500 TB)
- NASA对地观测系统三年数据： 1PB
- (人类功能记忆的容量预计在1.25个PB)
- 所有印刷材料： 200PB
- 全人类说过的所有的话： 5EB
- 人类所有存储资料： 500 EB
- 全世界海滩上的沙子数量总和： 1ZB

数据表示

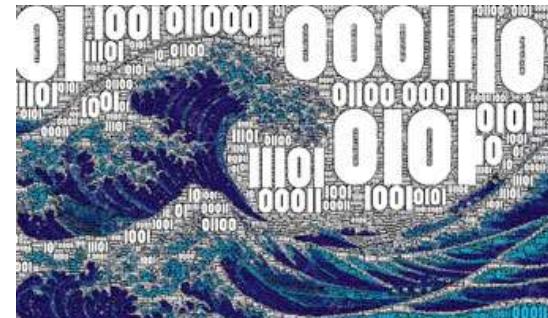
- 5% 数据是结构化且能适用于传统数据库。
- 95% 非结构化数据都无法被利用, 如网页和视频资源。
- 大数据: 接受不精确性, 走入从未涉足的世界

大数据定义

大数据泛指在一定时间内无法用常规软件工具对内容进行抓取、管理和处理的数据集合。

有4V特征-Volume、Velocity、Variety、Value。

- 1、Volume指容量大；2、Velocity指速度快；
- 3、Variety指数据类型丰富；4、Value指价值。



数据本身：

数据分析

大数据已成为一种新的资产和财富类型，蕴含只有通过大数据才能显现的价值；

经验直觉

基础架构：

通过大数据技术构建的数据存储、分析系统具有成本低、扩展性强的显著优点

大数据及大数据的处理技术

具有多重价值。

处理容量

大数据技术能够处理的数据规模远超过传统数据处理技术，在更短的时间内能够挖掘更多的数据；

处理能力

大数据擅长处理非结构化的数据，在挖掘的维度、深度上均有明显优势，如各种分类、聚类、推荐算法。

2.3

什么是大数据 - 应用例子

大数据分析应用

- 源于搜索排名：搜索引擎都试图对搜索请求排出网页的相关性次序。
- 广告追踪：记录每次用户会话中每个页面事件的海量数；迅速完成一次广告位置、颜色、大小、用词和其他特征的试验以表明更改是否促成了更好的点击行为。
- 位置和邻近追踪：安全分析、导航和社交媒体中加入 GPS 位置追踪、更新。
- 精确定位：为 GPS 测定点附近其他位置的海量相关数据带来销售或服务的机会。

恐怖袭击—奥兰多，脉动！

- 美国911后最大恐怖袭击：2016年6月12日奥兰多—脉动同性恋酒吧阿富汗裔奥马尔·马丁枪杀49人伤53人。
- 这和大数据什么关系？
- 1966-2012：美国发生90次 ≥ 4 人死亡的枪击事件，同期全世界发生292起。美国人口占全球5%而枪击却占31%！美国民间有大约3亿只枪！
- 防范恐怖分子：FBI 两次调查奥马尔·马丁，但认定他与恐怖活动关联弱，放了他一马！导致。。。。
- 如果FBI 用大数据分析潜在的恐怖分子，也许情况不一样！

巴黎、布鲁塞尔恐袭



2015.11.13-伊斯蘭國
(IS) 極端組織聲稱的巴
黎恐怖襲擊事件中，**129**
人喪生。攻擊持續了3個
小時！

2016年3月22日，比利时
首都布鲁塞尔市郊的扎芬
特姆机场和市内欧盟总部
附近地铁站先后发生爆炸，
造成**30多人死亡**，**200多**
人受伤。

案例3：CIA、FBI、金融诈骗

CIA、FBI等情报机构掌握着成千上万个数据库：

- 财务、DNA样本、语音、录像片段以及世界各地的地图！
- 挑战：对一些数据实时感知和这些数据之间建立关联，从浩如星海的数据中快速找出有价值的线索，提前掌握恐怖份子可能发动袭击的消息！

美Palantir（帕兰提尔）公司最为显著“战绩”：

- 帮银行追回前纳斯达克主席Bernard Madoff所隐藏起来的数十亿美元巨款。
- 追捕本·拉登的行动发挥的情报分析作用。
- 2010年，摩根大通用Palantir的技术查找盗取客户账号的欺诈者。

谷歌2009年“自然”論文：預測冬季流感傳播

- 用搜索记录来完成这个预测至全美范围到特定的地区和州
- 历史搜索记录 + 30亿搜索指令 / 天
- 比较5千万条最频繁检索词条和美国疾控中心在2003-2008年间季节性流感传播时期的数
- 预测07-08年，再对比美国疾控中心相应记录
- 相关性 = 97%!

破译人体基因密码—机器炒股

- 辛苦工作了十年，03年第一次破译人体基因密码（三十亿对基因排序）
- 今天，基因仪只要10分钟！
- 美国股市每天的成交量>70亿股 (~1千亿美元)
 - 2/3交易由程序自动完成
 - 用大数据来预测利益和降低风险。

西雅图交通数据处理公司Inrix

- Inrix汇集美欧近亿辆汽车交通数据：私家车，出租和货车，私家车主的移动电话。
- 建立手机APP：提供免费交通信息同时得到同步数据。
- 预测交通状况：数据与历史交通数据进行比对+天气和当地时事信息。
- 分析结果：同步到汽车卫星导航系统中，政府和商用车队都使用。
- 2011年，美国经济复苏开始放缓，政客们否定。但被交通状况分析给披露：
 - Inrix分析发现，上下班高峰期交通状况变好=>失业率增加=>经济状况变差。
- Inrix把数据卖给了一个投资基金，该基金把交通情况视作一个大型零售商场销量的代表，一旦附近车辆很多，就说明商场的销量会增加。

大数据帮助企业：促进业务

- 零售业实时精准营销
- 银行多渠道交叉销售、增值销售、客户关怀
- 金融投资风控
- 贷款风险分析和保单承保—评价风险：
 - 偿付历史、信用报告、就业数据, 财务资产披露内容等; 贷款抵押物或被保项目...。
- 公司竞争性：
 - 客户流失分析—流失预测性因素, 客户行为、经济、年龄段和其他人口统计学信息。

大数据分析：

- 基因组分析：种子基因组测序。
- 棉花序列识别基因和基因族：
 - 更高产、更高纤维品质、适应环境压力、抗虫害和病害。(SeedQuest,2010 年 9 月 22 日)。

大数据分析：客户特定群体

- 企业使用客户队列群体来识别一般的人口统计学趋势和行为历史。
- 羊群效应：买相同样一本的其他客户也买类似这些书。
 - 如果能将产品或服务出售给群体一名成员，剩下的成员都有希望。

大数据分析：检测和预防金融欺诈

- “空头支票”：钱在两个独立账户之间来回快速转账
- 短时间内跨越两个单独机构的交易
- 多个帐户的行为模式
- 合谋经纪人不断抬高价格售出证券
- 经纪人获利快速退出
- 黑客入侵检测和干预：不寻常的系统进入模式。

大数据分析：在线游戏手势跟踪

- 记录玩家每次点击和移动 “遥测数据”
- 对屡战屡败 (灰心) 玩家干预、 向将要离开的玩家提供额外剧情或游戏目标、 新游戏和新剧情的体验！
- 电视收看： DVR / 机顶盒捕获遥控器按键行为、 录像事件、 播放事件, 画中画观看以及收视指南的背景情况。
- 所有玩家动作/事件发回服务提供商。

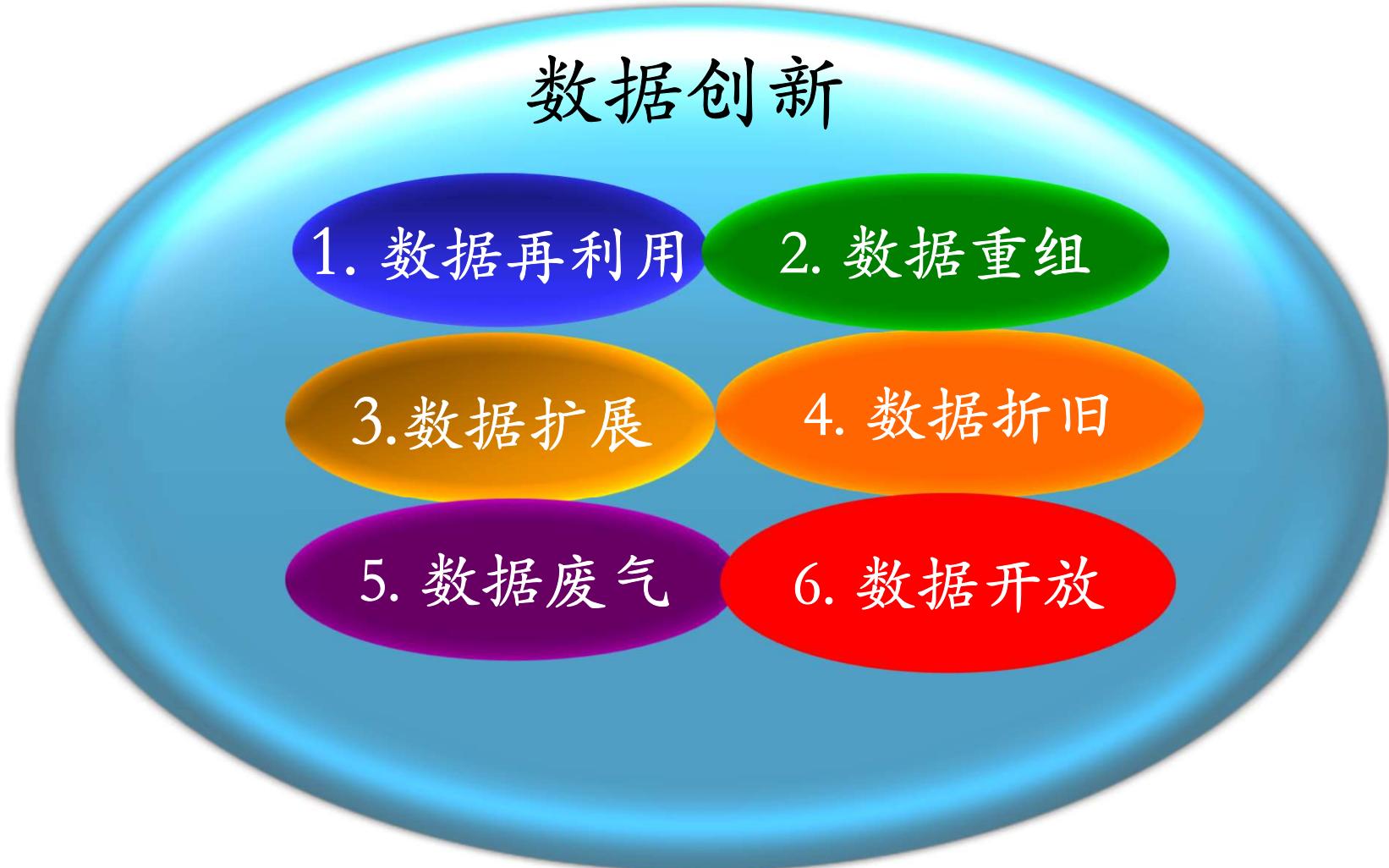
大数据分析：在航飞机状态

- 传感器技术：发动机、燃油系统、液压和电力系统数以百计的变量组成了在航状态,不到几微秒就被测量和发送一次。
- 分析工程遥测数据,促进实时自适应控制、燃油使用、零件故障预测和飞行员通报。

2.4

大数据创新应用

六大创新应用



数据再利用

故事1：运营商收集用户位置信息：

- 数据具有技术用途。
- 发布个性化位置广告服务和促销。
- 数据价值从汇总中体现出来。
- 例如：人们周五晚上聚集在哪里？汽车在哪个地段行驶多慢等？
- 信息集合可用来确定房地产价值或广告牌价格。

数据再利用

故事2：谷歌在07年在搜索列表中加入语音识别服务：

- 没有自己的语音识别技术，急需购买许可。
- 谷歌与Nuance公司达成合作协议。
- 合同中没有规定由谁来保存语音翻译记录！
- 谷歌保存了数据！
- 数据使得谷歌从头创建了一个新的语音识别系统。
- Nuance忽视了数据，认识到后，它立即与移动运营商和手机制造商达成其语音识别服务从而进行数据收集。
- 2016.3.24谷歌向第三方开发者开放其语音识别免费API，与Nuance及其它的语音识别公司竞争。

创新1：数据再利用

故事3：亚马逊与AOL（美国在线--曾是最大Internet服务提供商之一）：

- 亚马逊早期与AOL达成了一项协议，提供后台技术服务。
- 看似普通的外包协议，亚马逊真正的用意在于掌握用户的数据：他们在看什么、买什么。“这些数据可以帮助亚马逊提高它的推荐引擎性能。”
- AOL没有意识到这一点，只看到了销售这个基本的利益；亚马逊却从二次数据利用中获利。
- 后继故事：2015.5.11，Verizon宣布44亿美元收购AOL。
- 现在亚马逊市值>10,000亿美元

创新2：数据重组

- 重组：新的数据价值可通过与另一截然不同的数据集结合释放出来。
- 数据的总和比部分更有价值。重组总和本身的价值也比单个总和更大。
- 用户都熟悉基本的混搭式应用，即将两个或多个数据源以一种新的方法结合起来。
- =》 房地产网站可将房地产信息和价格添加在社区地图上，如社区近期交易和物业规格=》预测区域内每套住宅的价值。

数据重组

- 故事4：手机是否增加致癌率？
- 丹麦拥有85年以来所有手机用户的数据库。分析了1990—2007个人用户，共358403人。记录所有癌症患者的信息（共有10729名中枢神经系统肿瘤患者）。
- 结合两个数据集寻找两者关系：手机用户是否比非手机用户显示出较高的癌症发病率？使用手机时间较长的用户是否比较短的用户更容易患上癌症？
- 研究的规模很大“样本=总体”的准则，包括所有癌症患者和移动用户。
- 没有发现使用手机和癌症风险增加之间存在任何关系！
 - 2011年10月研究结果在《英国医学杂志》上发布时，并未在媒体中引起任何轰动。

创新3：数据扩展

- 收集多个数据流=》每个数据流中更多数据点的额外成本往往较低。
- 二次用途并使其具有扩展性。
- 数据潜在价值：“一份钱两份货”单一数据集有多种不同的用途，具有双重功能。
- 二次用途是一定的！

可扩展数据

- 案例5: 谷歌街景与GPS采集
- 谷歌收集数据扩展性: 街景汽车不仅拍摄了房屋和道路的照片, 同时采集GPS数据, 检查地图信息; 无线网络名称。
- 一辆谷歌街景汽车积累大量离散数据流。谷歌进行了大量二次使用。
- 对谷歌自动驾驶汽车运作功不可没。



可扩展数据

故事6: 2011年， 西雅图Decide.com公司为顾客预测商品价格。

- 所有电子产品价格=> “大数据” => “大文本” => 数据分析 =>产品是否下架 =>新产品发布=>产品价格影响。
- 2012: Decide.com分析400万产品>250亿条价格信息。发现怪现象! 新产品发布时=>旧的产品价格短暂上浮。大部分人都认为旧产品更便宜，会选择买旧产品，付出的金钱比购买新产品还要多。
- Decide.com 发现不正常、不合理价格高峰， 告知用户何时才是购买电子产品的最佳时机。
- 准确率=77%， 平均顾客／产品节省100美元。

数据折旧

- 企业有经济动机来保存数据并再使用。但是效果有限。
- 故事7：亚马逊利用客户购买的产品、浏览的页面和评论来推荐新的产品，重复使用数据。
- 公司应该永远保存这些数字记录？
- 时间的推移，数据都会失去一部分基本用途。继续依赖于旧的数据不仅不能增加价值，实际上还会破坏新数据的价值。
- 十年前在亚马逊买了一本书，可能对它完全不感兴趣。继续用这个数据来向你推荐其他书籍是否合理？
- 旧数据的存在可能会破坏新数据的价值。

数据折旧

- 故事8: 亚马逊决定只使用产生价值的数据=>更新数据库、淘汰无用信息。挑战: 如何得知哪些数据不再有价值? 仅据时间来判断显然不够;
- 建立复杂的模型来分离有用和无用的数据
- 客户浏览或购买了一本基于以往购买记录而推荐的书, 这项旧的购买记录仍然代表着客户的喜好->评价旧数据的有用性—“折旧率”更具体。
- 收集尽可能多的使用数据并保存尽可能长的时间。
- 与第三方分享数据, 要保留“延展性”权利。
- 由数据再利用而产生的任何商业价值, 原始数据拥有者都能分一杯羹。

数据折旧

- 故事9:
- 并非所有数据会贬值 ->谷歌拒绝将IP地址从旧的搜索查询中删除。
- 谷歌要每年的同比数据，如假日购物搜索等。了解搜索者的位置，改善搜索结果的相关性。

创新5：数据废气

- 人们在网上留下的数字轨迹：
- 网络公司捕捉用户在其网站上做的所有事情；将每个离散交互作为“信号”，为网站个性化、提高服务或创建新数字化产品的反馈。
- => 用户在线交互的副产品
 - 浏览了哪些页面、停留了多久、鼠标光标停留的位置、输入了什么信息等。
- 数据废气循环利用，改善现有或开发新服务
 - 谷歌是这方面的领导者，不断地“从数据中学习”
 - 用户每一个动作都被认为是一个“信号”，谷歌对其进行分析并反馈给系统。
 - 数据废气=>服务背后的机制=>语音识别、垃圾邮件过滤、翻译等。用户指出语音识别程序误解了他们的意思时，实际上在“训练”系统，让它变得更好。

数据废气

故事10：微软与谷歌的拼写检查

- 微软为其Word软件开发出了一个强大的拼写检查程序；
- 频繁更新字典正确拼写、比较用户键入的字符流进行判断。
- 不断编译和更新字典，每年要花费数百万美元创建和维护费用。
- 谷歌拥有世界上最完整的拼写检查器，涵盖了世界上的每一种语言。这个系统一直在不断地完善和增加新的词汇—人们使用搜索引擎的附加结果。
- “免费”获得拼写检查，30亿/天的查询中的错误拼写，巧妙的反馈循环可以将用户实际想输入的内容告知系统。
- 直接在用户访问的页面上显示正确拼写的结果=》正确的拼写高度相关。
- 那些“不合标准”、“不正确”或“有缺陷”的数据非常有用。

数据废气

- 故事11：2000年，雅虎也看到了从用户输错的查询中创建拼写检查系统的可能性，只想没做。旧的搜索查询数据被当成了垃圾。
- Infoseek和Alta Vista这两个早期流行的搜索引擎，都拥有那个年代世界上最全面的错别字数据库，未懂得欣赏其中的价值。在用户搜索过程中，系统将错别字作为“相关词”进行了处理，而不是用户交互的总和应用。
 - 1998年中，迪斯尼成为Infoseek的控股公司，将Infoseek转型为门户网站。
 - AltaVista于95年由Digital Equipment Corporation创立。03年被Overture以1.4亿美元现金+股票购得同年转手雅虎。雅虎宣布13年7月8日关闭AltaVista服务。
 - 2016年7月25日消息：威瑞森确信以48.3亿美元购买雅虎互联网核心业务并整合于美国在线（AOL）。

数据废气

故事12：早期Facebook的数据科学家们研究了数据废气的丰富信息，发现人们会采取某种行动（如回帖、点击图标等）

- 重要的预测指标就看到了周围的朋友也这么做。
- Facebook重新设计了它的系统，收集用户反馈使每个用户的活动变得可见并广播出去，成为网站的良性循环。

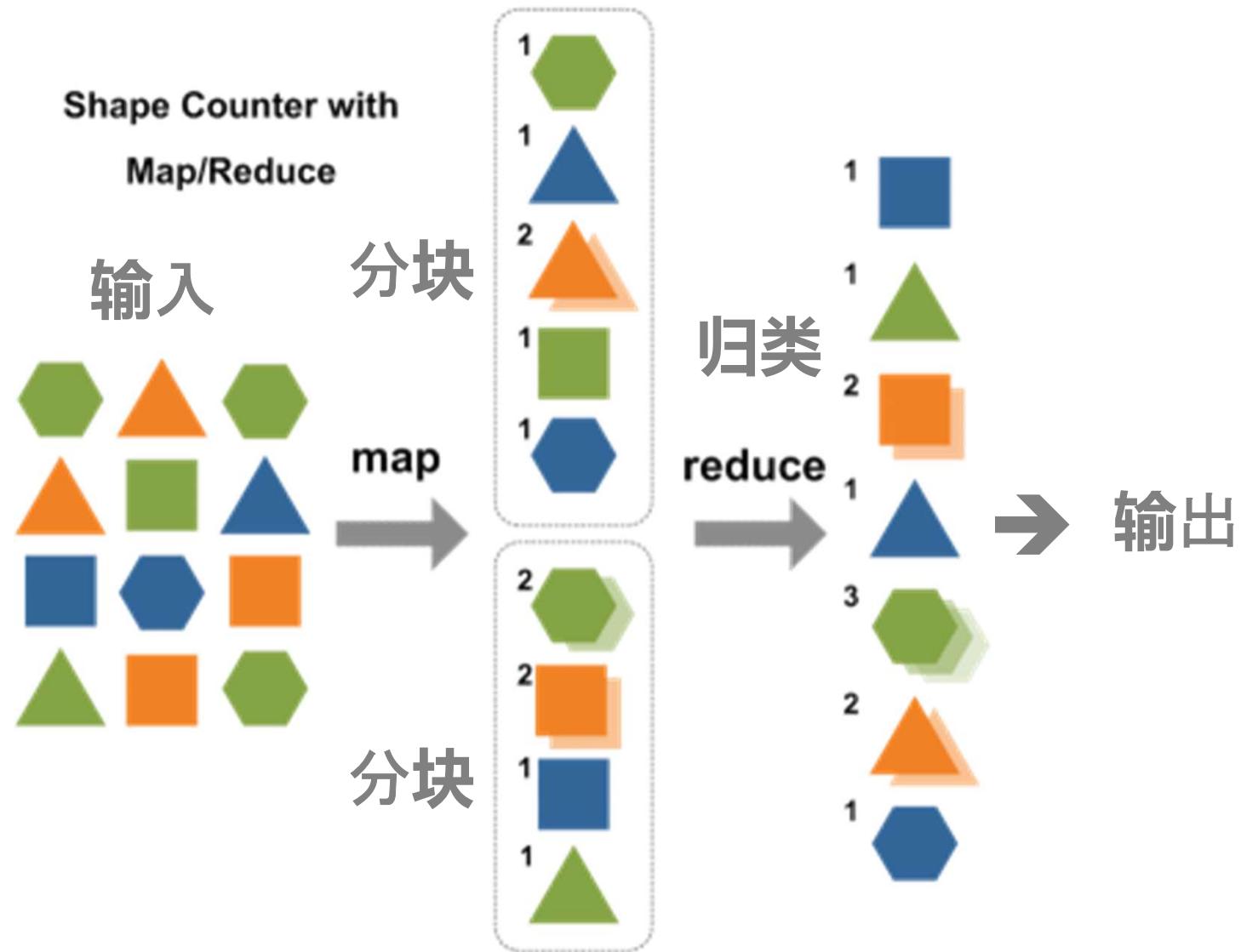
创新6：数据开放

- 政府=>大规模信息的原始采集者，还在与私营企业竞争他们所控制的大量数据。政府与私营企业数据持有人之间的主要区别就是，政府可以强迫人们为他们提供信息，而不必加以说服或支付报酬。因此，政府将继续收集和积累大量的数据。
- 提取政府数据价值=>允许私营部门和社会大众访问。国家收集数据时代表的是其公民，因此它也理应提供一个让公民查看的入口(危害到国家安全或他人隐私权的情况除外)。

开放数据

- 故事13: 2008年1月21日，奥巴马在就职的第一天发表了一份总统备忘录，命令美国联邦机构的负责人公布尽可能多的数据，使开放政府数据的想法取得了进展。“面对怀疑，公开优先。”指令促成了data.gov网站的建立，网站从2009年的47个数据集迅速发展起来，到2012年7月三周年时，数据集已达45万个左右，涵盖了172个机构。
- 英国政府已经颁布相关规定鼓励信息公开，并支持创建由万维网的发明者蒂姆·伯纳斯（Tim Berners-Lee）参与指导的开放式数据中心。
- 欧盟宣布开放数据。
- 澳大利亚、巴西、智利等也相继出台并实施了开放数据策略。同时，世界各地越来越多的城市和地区也已经加入开放数据的热潮，
- 世界银行公开了数百个之前被限制的关于经济和社会指标方面的数据集。
- Web开发人员和有远见的思想家组成了数据团队来最大化开放数据价值，如阳光基金会（美国）和开放知识基金会（英国）。

大数据计算：MapReduce/Hadoop 系统



大数据价值链的3大构成

- 三种大数据公司：数据本身、技能与思维。
- 1. 数据公司：公司拥有大量数据；不一定有从数据中提取价值或者用数据催生创新思想的技能。Twitter拥有海量数据，它的数据都通过两个独立的公司授权给别人使用。
- 2. 技能公司：咨询公司、技术供应商或者分析公司。掌握了专业技能并不一定拥有数据或提出数据创新性用途的才能。沃尔玛和Pop-Tarts这两个零售商就是借助天睿公司（Teradata一家大数据分析公司）的分析来获得营销点子。
- 3. 思维公司：想法获得价值。例如通过用户分享到网上的旅行照片来为人们推荐下次旅行的目的地。
- 数据和技能并不是成功的关键。创新思维 => 挖掘数据的新价值的独特想法。

体会

- 大数据的研究和应用是一片蓝海；
 - 水还是很深！
- 数据从何而来？
 - 要：1. 向政府要； 2. 向企业要；
 - 自己采： 网络空间人-机-物。
 - 得有工具😊
 - 人工智能是方向--» 机器学习

提纲

- 1 智慧城市基础：人工智能概述
- 2 大数据应用
- 3 智慧城市物联网国重@UM
- 4 小结

The University of Macau



澳门大学智慧城市物联网国家重点实验室

7月获国家科技部批准，10月8号揭牌仪式

City Development
城市发展
为主线

SKL of IoT-SC@UM
智慧城市物联网
国家重点实验室

Fundamental
Research & Key
Technology
基础研究/关键技术
为支撑

Talent Training
人才培养
为平台

Applications
应用支持
为特征



背景和意义

第一部分：勇于挑战、把握机遇

新建智慧城市物联网SKL **立足澳大，贡献澳门，融入湾区，面向世界**

第二部分：立足技术、持续创新

智慧城市物联网SKL的**基础研究领域及关键技术**

第三部分：支撑智慧城市建设、助力产业发展

智慧城市物联网SKL的**四大示范应用**

澳門智慧城市的挑戰與機遇



萬物互聯、動態資料；資源優化、聚合多贏！

澳门智慧城市的挑战与机遇

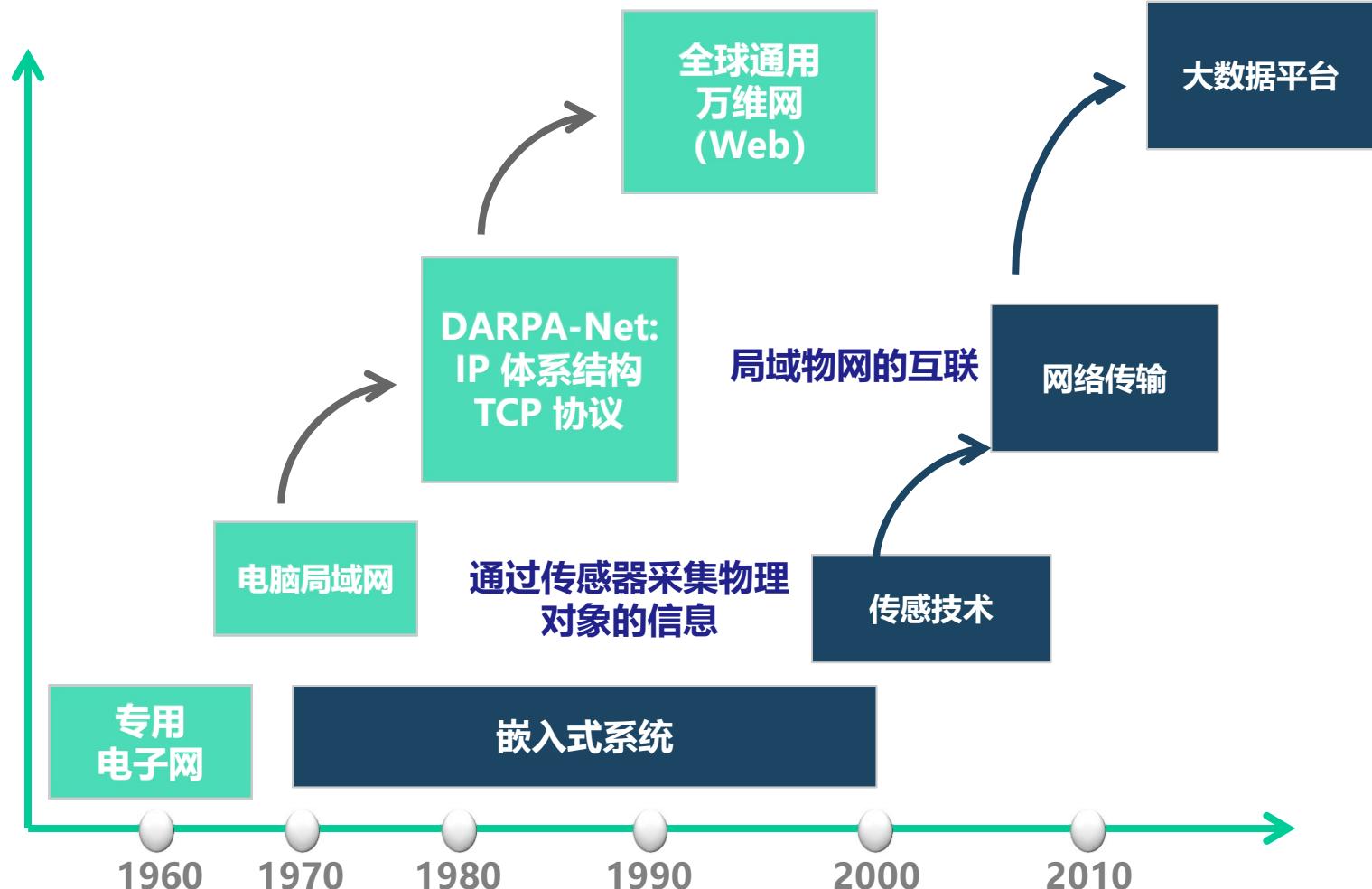


万物互联、动态数据；资源优化、聚合多赢！

物联网的发展：



数据的大规模共享和分析



新建智慧城市物联网国家重点实验室



智慧城市物联网必须有所为！



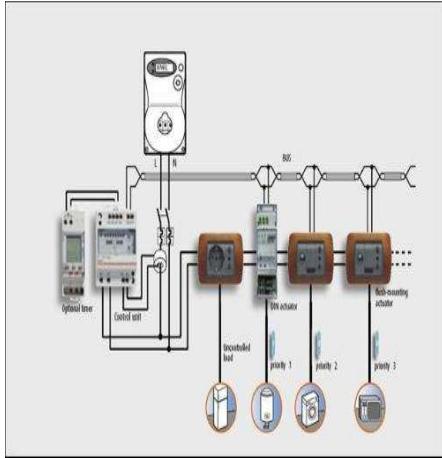


新建智慧城市物联网国家重点实验室的 必要性和可行性





新建智慧城市物联网国家重点实验室的 必要性和可行性



智慧城市物联网实验室将透过科技和创新，在能源管理、智能交通、公共安全和灾害防治、城市大数据等领域开发新技术，优化城市承载力，为市民创造更美好的工作生活环境。作为“互联网+”的新风口，物联网技术将持续带动传统行业的发展，为推动澳门经济适度多元和可持续发展出力。

突破澳门城市发展的瓶颈！



新建智慧城市物联网国家重点实验室的 必要性和可行性



充分利用“一个平台”的优势，实现物联网模块深度融合，智慧应用百花齐放，为建设粤港澳大湾区发挥独特作用。粤港澳大湾区的互联互通、深入融合，需要湾区城市群形成物联网网络，促进经济要素流动，和谐共赢。澳门透过本地物联网建设范本，辐射湾区、闪耀海上丝绸之路，实现资源共享，力争建成“大湾区+海丝”物联网合作示范区。

坚守澳门城市发展的定位！

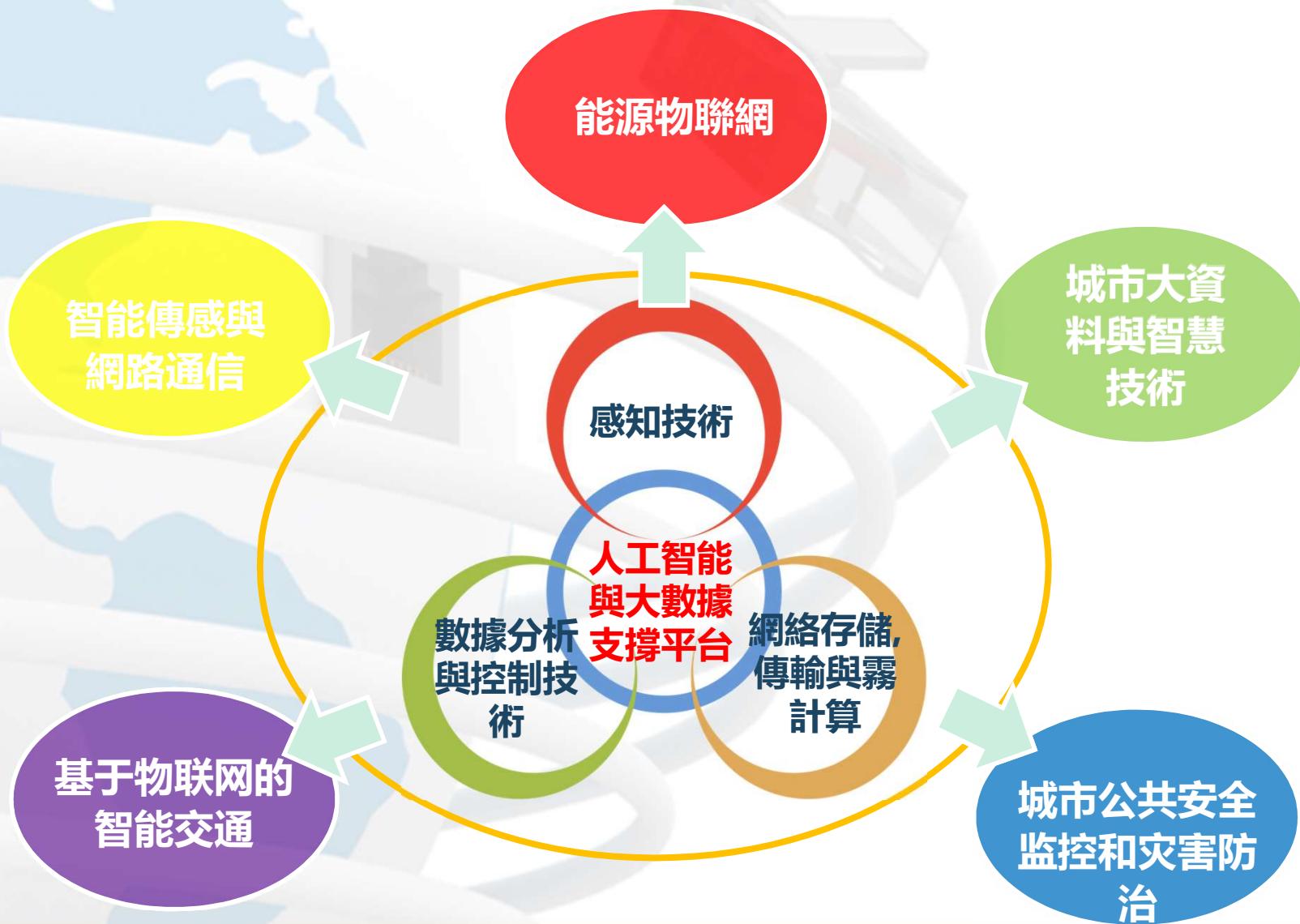


智慧城市物联网国家重点实验室

- 多方共建，利于发挥澳门的地缘和制度优势，形成与北京航空航天大学的软件开发环境国家重点实验室的优势互补、强强联合。
- 立足新的高度和战略布局，澳门大学为澳门物联网技术的持续发展及多学科的交叉融合奠定良好基础。澳门大学校长宋永华教授长期引领能源物联网领域，做出开创性的重大贡献，在智能交通物联网等领域成果丰硕，在海内外享有盛誉，是广受尊重的行业领导人。在物联网其他关键技术领域，如传感、标签、识别、定位、追踪、数据分析等，澳门大学均有优异的本土研究团队。
- 协助市民上升流动，培养和积累物联网技术和人才。在澳门本地拟建立一个高质量的技术及人力资源平台，包括软件设计和测试设备的硬件平台，多方位培养建设澳门、服务国家的高素质人才，为支撑粤港澳大湾区的经济增长提供有力的人才储备。

构建澳门城市发展的物联网生态系统！

以人為本的智能澳門 “物聯網+”



关键技术1 - 感知技术



感知天地、智创未来：传感器和相关的器件是支撑物联网的关键产业，方兴未艾，来日方长



- 1) 物联网振动、音频、视听感知技术：**传感器向微小型、低能耗、高可靠和多功能发展
- 2) 智能终端技术和协同传感：**通过传感器从城市中采集实时信息，融合监测、定位、健康管理等系统

赢者不能通吃，澳门和大湾区有发展传感器产业的潜力和优势



焦点问题

1. 高灵敏度实用型功能性材料
2. 新型传感器及集成技术
3. 传感信号处理、智能监测与控制
4. 智能终端技术和协同传感

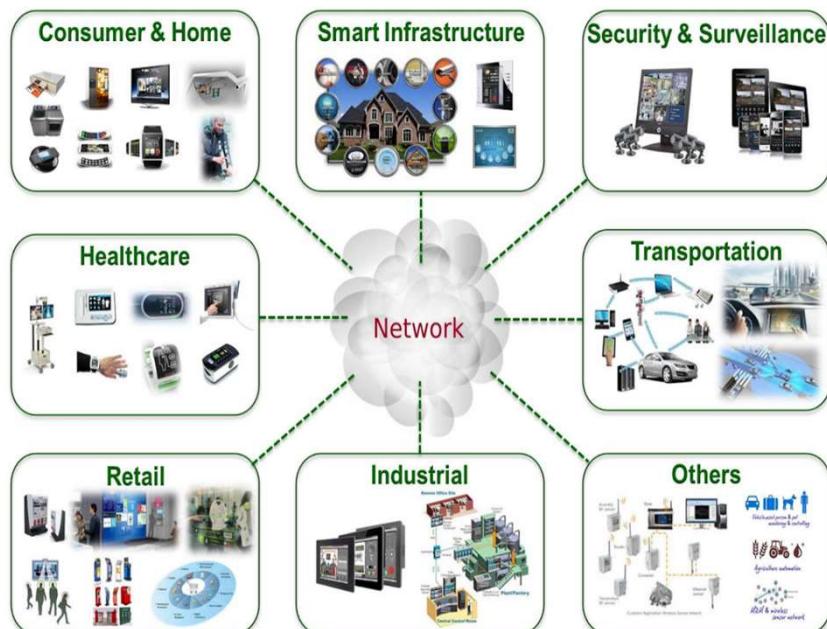


关键技术2 - 网络存储与传输技术

挑战: 如何把众多封闭的、不互联的物网 (NoT)
互联成统一的物联网 (IoT) 并有效管理数
据?

基础研究领域:

- 高效可靠低能耗通信技术
- 开放、安全、实时无线传输技术
- 物联网数据预处理 (数据辨析、抽取、清洗)
- 物联网大数据存储 (异构数据融合技术, 数据组织)
- 物网、互联网互联融合、远程控制





关键技术2 - 网络存储与传输技术

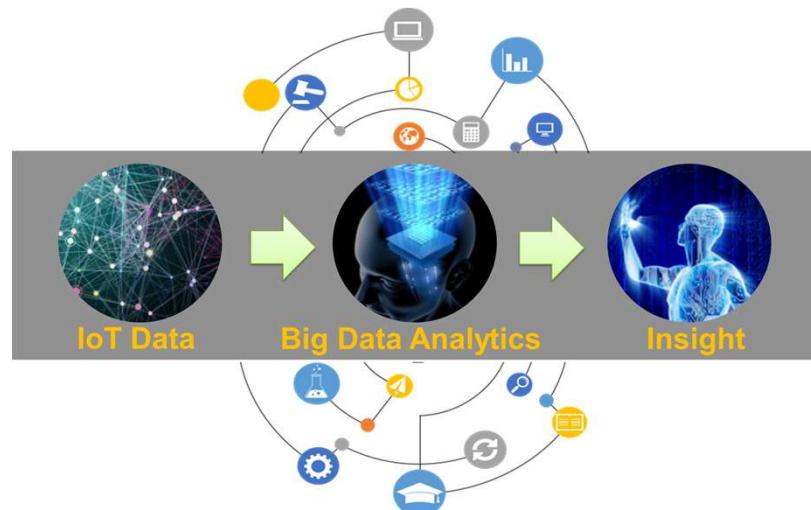
焦点问题

1. 通过对物联网大数据的动态传算，确定物理对象的实时状态。
2. 物联网大数据的高效预处理、存储及管理
3. 合理地分配网络资源以有效地实施分布式实时计算
4. 响应李克强总理号召，促进中国与七个葡语系国家更紧密合作中，通过葡语系国家的物联网平台互联，使澳门发挥“无形之桥”的作用。

关键技术3 - 数据分析与控制技术

数据分析与控制技术

挑战: 智慧城市物联网大数据结构复杂、种类繁多，具有分散性、多样性和复杂性等特征，这些特征给大数据处理带来极大的挑战。



基础研究领域:

- **云计算技术**
- **大数据分析及挖掘技术** (人工智能、知识发现与机器学习、模式识别与智能系统、神经网络、智能计算、图像处理与计算机视觉、预测模型、关联规则、系统仿真、优化、机器翻译、序列模式发现、异常和趋势预测等)
- **大数据展现**(大数据检索、可视化等)
- **智能物联网全数据共享**



关键技术3 - 数据分析与控制技术

焦点问题

1. 从底层实现基础设施状态实时分析、反馈和控制 - 实现智能化；
2. 设施群内部、群之间的状态共享与监控；
3. 大数据挖掘技术：数据分析服务，弹性计算技术，交互式数据可视化技术，数据深度和广度学习。
4. 依托国家物联网标识管理公共服务平台（NIOT），协助澳门与大湾区物联网产品国际化



智慧城市物联网国家重点实验室

示范应用



焦点问题

- 重点研发具有良好商业化前景的、配合澳门智慧城市定位的关键应用
- 注重研究核心算法和协议
- 打造集成平台，支持原型快速构建，积极参与国家和国际标准的制定 (譬如 工业4.0和中国制造2025在能源、交通、安全中的标准)



示范应用1 – 能源物联网

基于广域低功耗网络的能源物联网 - 开创分布式能源新世界

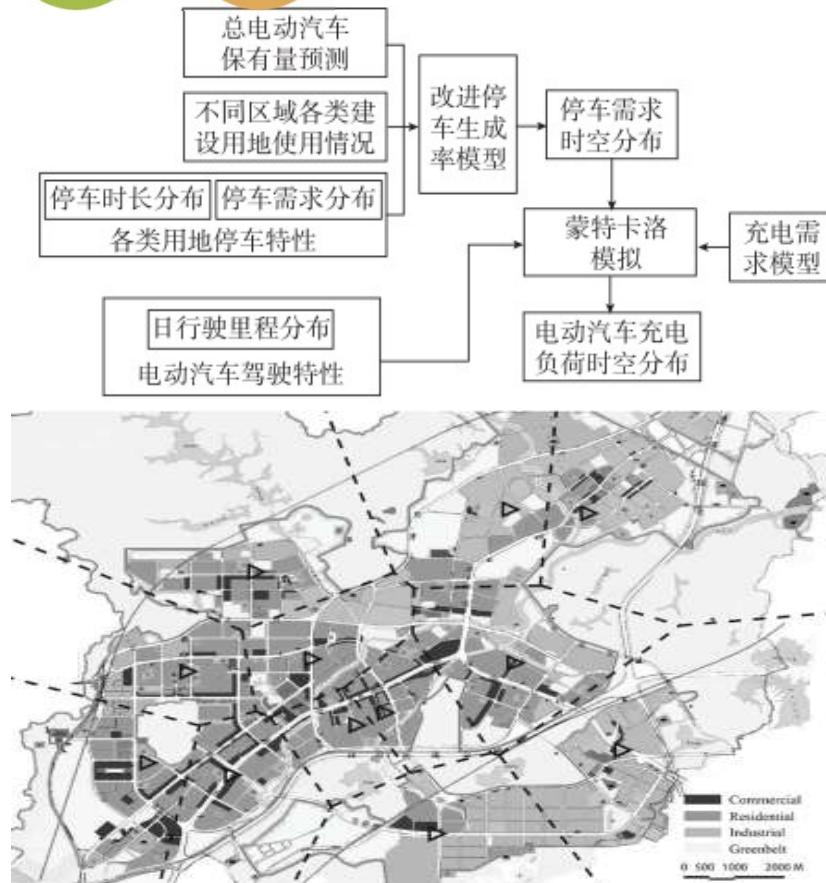
目标：推进能源互联网这一新型能源系统：通过支撑海量能源终端间的接入与互动操作，实现能源互联网内电、气、热等多种能源类型的高效集成，和能源从生产到消费的全生命周期管理。

内容：建立能源物联网的物理架构、运行模式与市场机制；开发基于环境和设备工况智能预知的智能节电控制系统与模型；应用于新能源微网、电动汽车、分布式储能以及主动配电网等前沿领域。

关键技术：广域低功耗网络技术、传感网、云操作系统、多能物联网关、能源互联网需求侧的典型应用。

示范 应用

示范应用1-能源物联网



考虑时空分布的城市中不同类型插电式电动车充电设备的集成规划框架和充电负荷预测

关键技术：

- 1) 用于插电式电动车的不同充电设备（公共充电桩和路边快速充电站）的优化布置；
- 2) 基于蒙特卡洛模拟方法，考虑不同时间、不同空间的停放、驾驶以及充电行为,预测电动汽车充电负荷的时空分布特性；
- 3) 居民区配电网规划。

示范应用2—基于物联网的智能交通



背景和意义：国家“十二五”规划中，澳门要建设成世界旅游休闲中心，如何提升交通服务效率和质量，令交通更加绿色化，科技化和文明化是实现智慧城市的重要科学问题



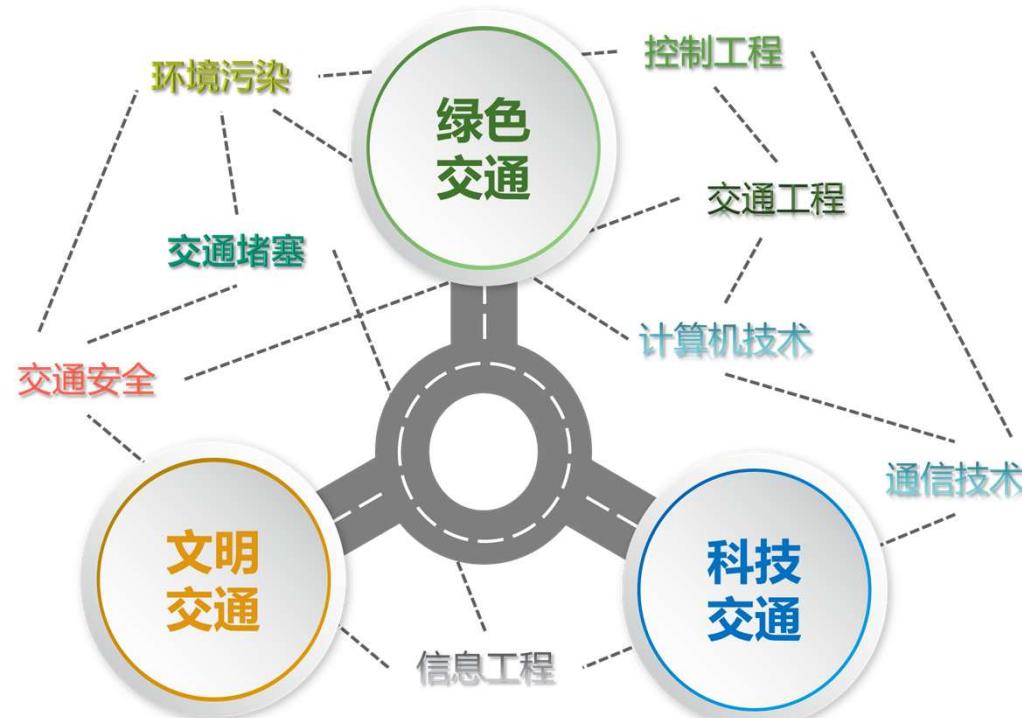
挑战：

- 1) 澳门地少人多，32.8平方公里的陆地面积长住人口达到65万
- 2) 游客数量庞大，2017年澳门全年入境旅客逾3200万人次
- 3) 国内、外旅客量屡创历史新高，年增长率达6~8%
- 4) 交通堵塞、人潮攒动、道路安全成为澳门交通系统的新挑战



示范应用2—基于物联网的智能交通

实现澳门智慧城市的核心方案



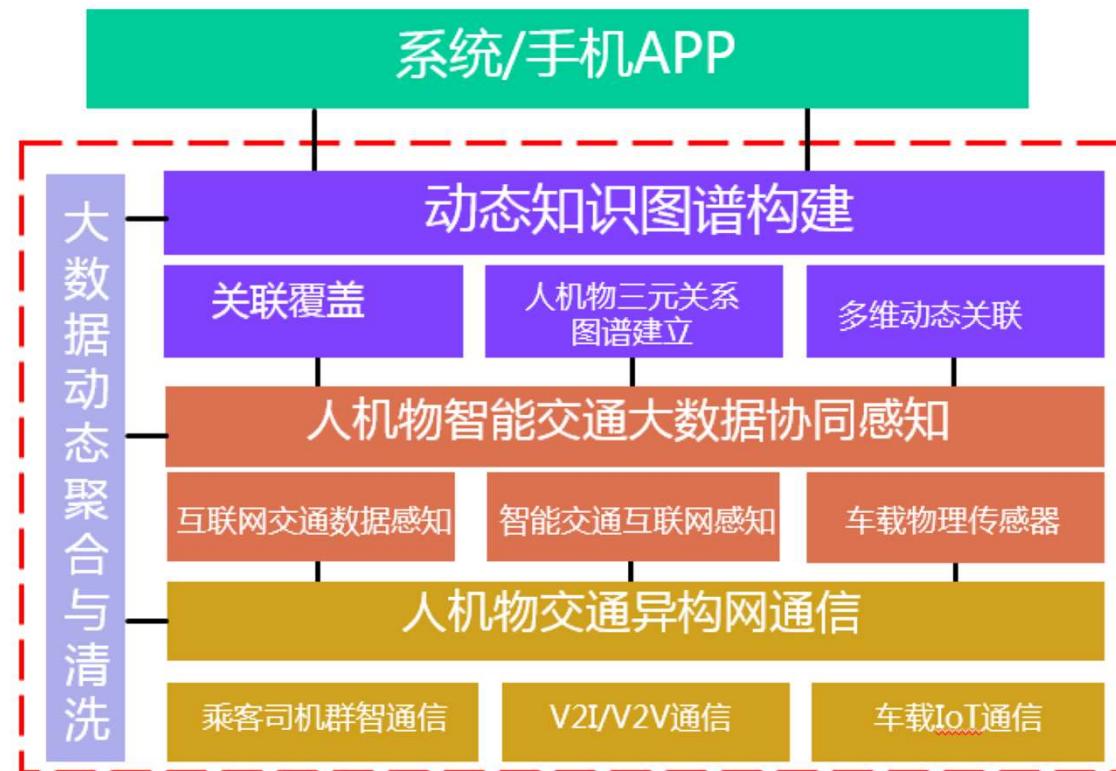
前期成果：

1. 电动汽车充电设施规划优化
2. 基于“车载终端、无线数据链路和监控中心软件”的软、硬件综合监管系统
3. 香港屯门高速公路智能检测系统
4. 在线众包地图匹配平台
5. 准确辨识违规驾驶行为
6. 自主驾驶汽车传感器数据融合，基于雷达信号的道路建模，以及道路交通标记识别

示范应用3—基于物联网的智能交通

澳门智能交通的未来发展

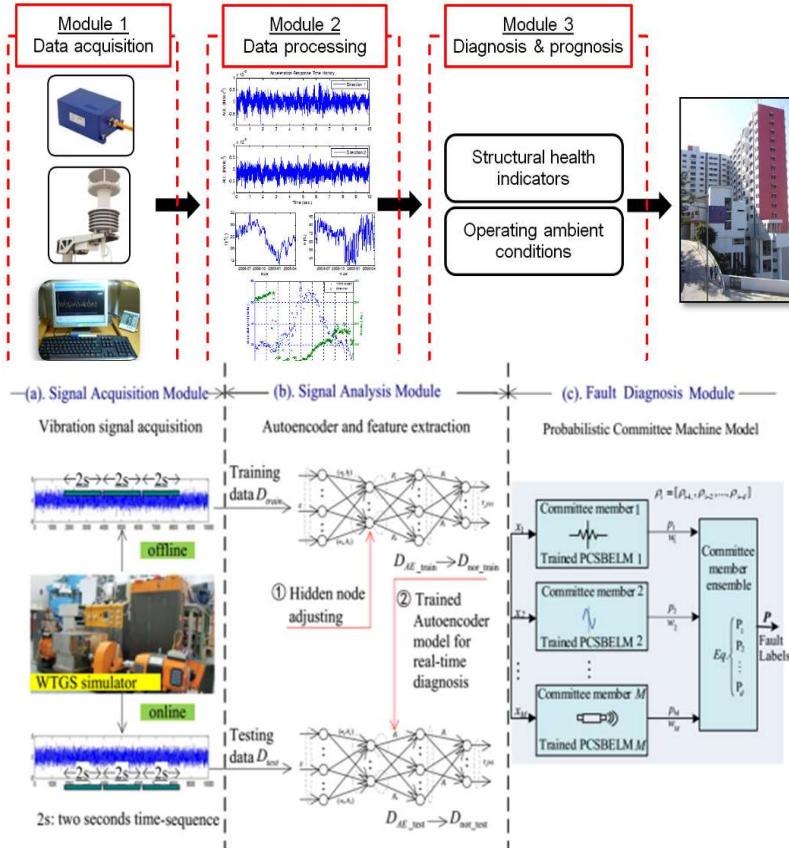
建立交通大数据知识库平台，实现全方位基于物联网的澳门智能交通建设。





示范应用3-城市公共安全监控和灾害防治

基于物联网的城市基础设施（建筑结构和机电设备）健康监测

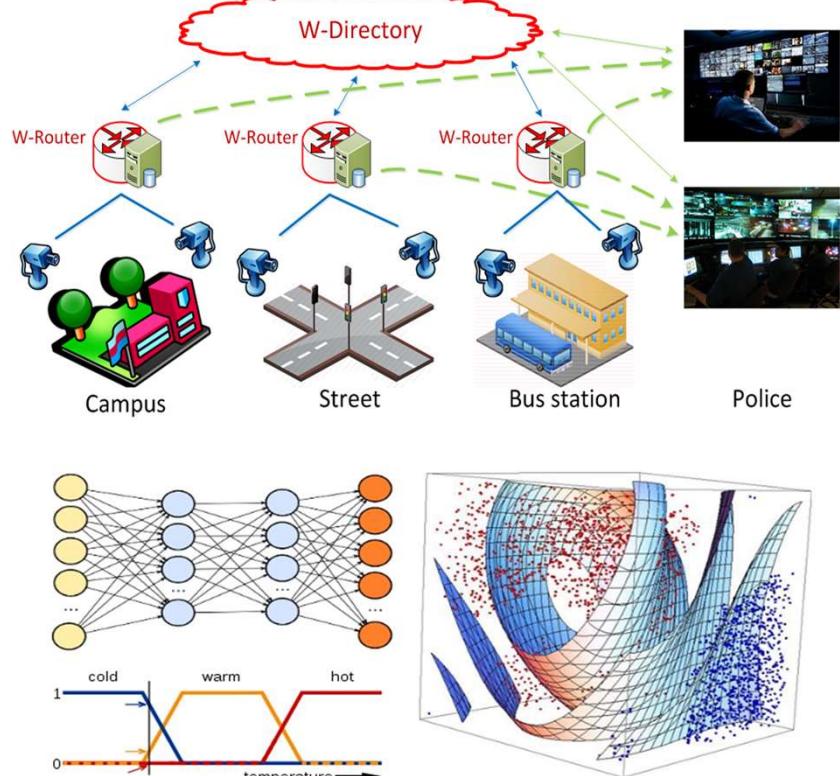


关键技术：

- 开发有效的实时模态和刚度参数识别方法，用于可靠的结构完整性评估并诊断结构中的可能损害
- 开发新的子结构识别方法
- 提出新的贝叶斯实时结构健康监测框架
 - 更新结构的有限元模型
 - 对异常数作出实时识别
 - 对卡曼滤波器参数作实时识别
 - 进行实时模型群选择，并提出自适应模型群
- 实时监测城市重要机电设备（供电、供水、医疗设备），进行智能健康管理并故障预测
- 灾害防控系统（海洋、楼宇、设备等）



示范应用3-城市公共安全监控和灾害防治



**基于物联网的智能社区、城市的
安全和边界监控，以及灾害防治
关键技术：**

- 智能集成公共安全智能视频监控服务平台
- 基于移动机器人的三维物体识别和操控
- 群体飞行器分布式控制系统感知突发事件的发生
- 提高食品、生产、环境、全过程监控（工业4.0）
- 物联网在减灾防灾、疾病瘟疫中的对抗能力



实验室目标

引领智慧城市物联网研发，推动本地，
国家，和全球的物联网产业。

实验室策略

引领产业，为产业化服务

注重基础，瞄准国际前沿

全面覆盖，主攻应用难点

优势互补，整合内外资源

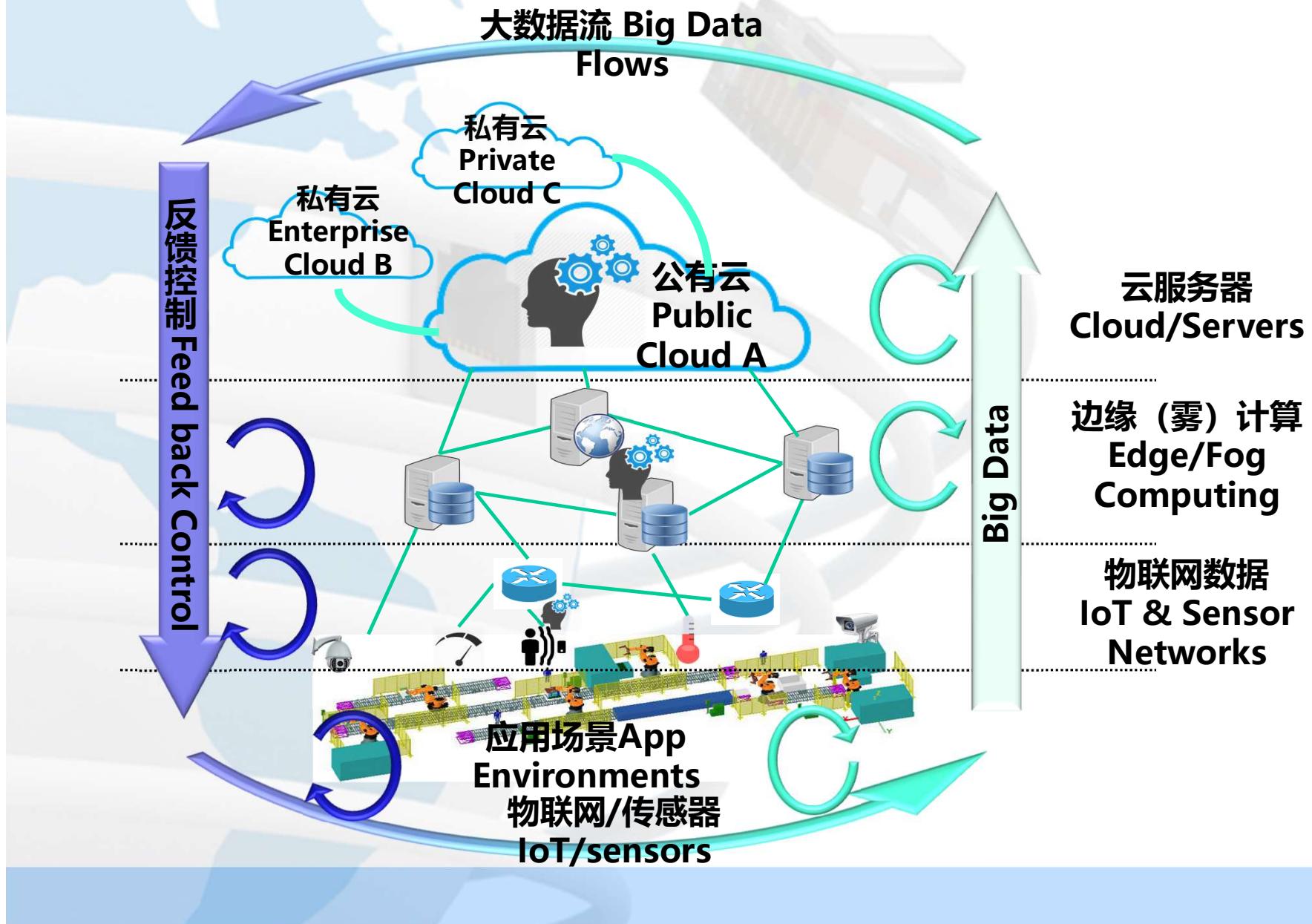
THE MAP OF THE MACAO
SPECIAL ADMINISTRATIVE REGION
OF THE PEOPLE'S REPUBLIC OF CHINA



澳門特別行政區水域示意圖



研究方案



Applications: Smart Estate Management System

“智慧物业” 安全管理和运营解决方案



Smart-home 智能安防产品

- 采用欧洲标准的超高频加密射频技术，超低功耗，室内有效距离远，穿墙能力强，外设数量多，防串码和密码，双向通信，安全可靠
- 视频实时联动，推送照片和报警录像到用户手机，随时随地查看现场
- 微信全面操作，微信推送报警和故障消息，微信个人服务查询服务报告，查询系统状态和操控安防智件



系列化高清云摄像机

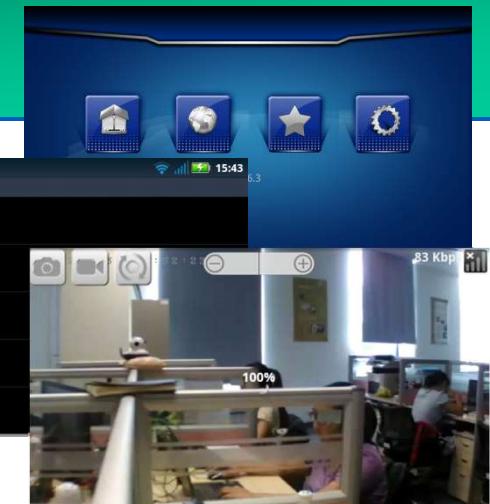
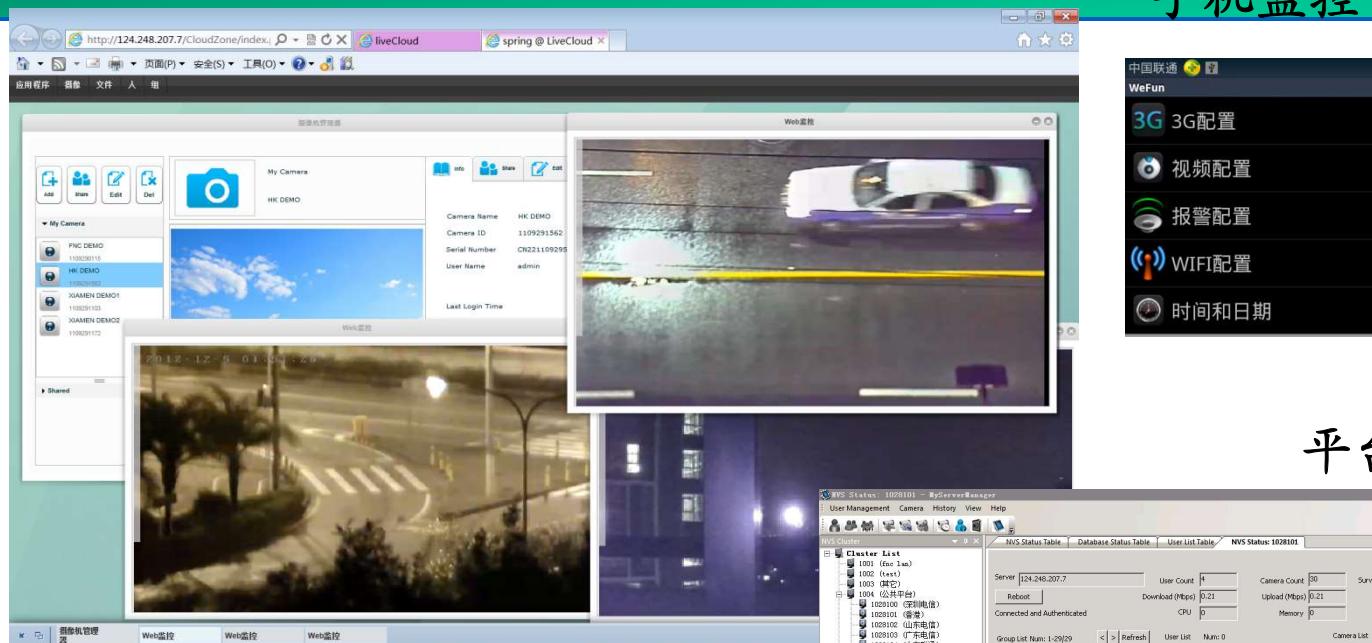
- 云眼：云网络摄像机
- 4G眼：4G LTE摄像机
- 狗狗看家：云报警摄像机
- 户外型网络摄像机



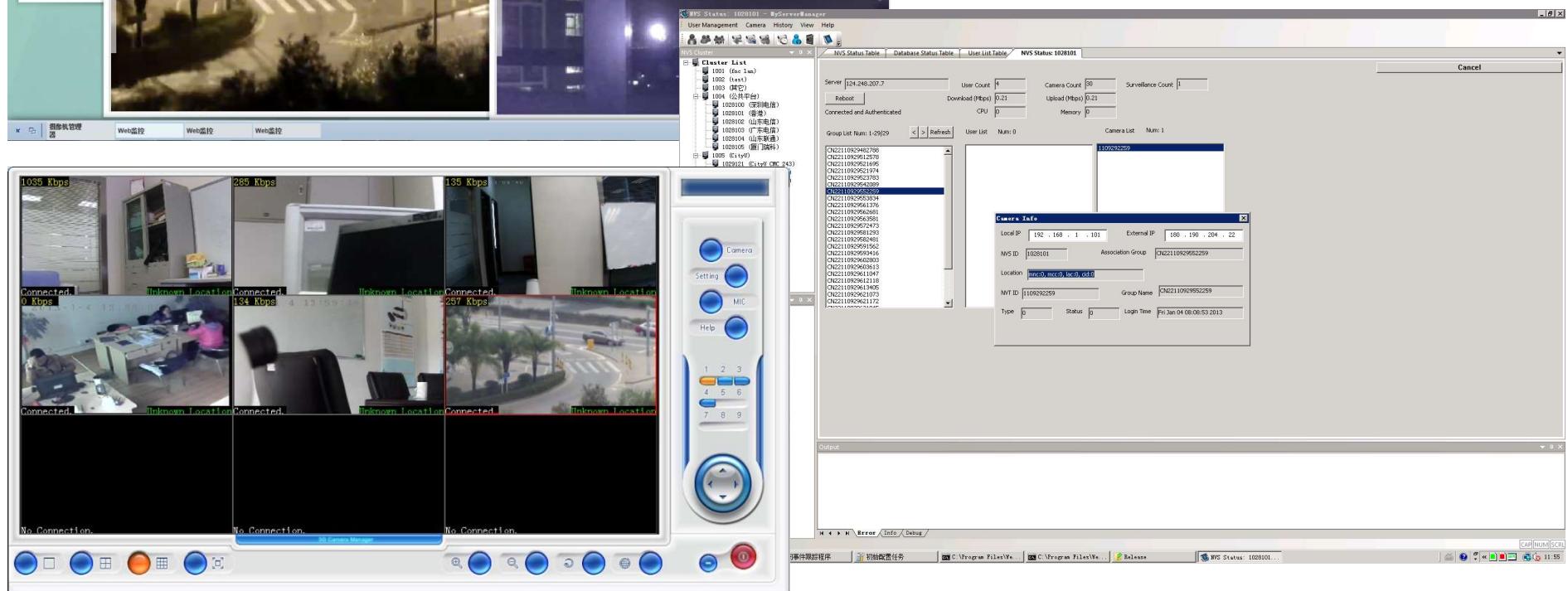
序号	产品图片	货物名称、型号及规格	4G/3G	WIFI	FSK	云台
1		4G/3G 高清网络摄像机 (B2X720-LTE/TD/W/E)	√	√	√	√
2		4G/3G 高清网络摄像机 (HS-G10D-LTE/TD/W/E)	√	√	可选	√
3		4G/3G 高清防水红外枪机 (GUN720-LTE/TD/W/E)	√	√	√	支持无线云台
4		高清网络摄像机 (H8-187V)	不支持	√	可选	√
5		云报警网络摄像机 (VD25)	不支持	√	√	√
6		高清防水红外枪机 (GUN720)	不支持	√	不支持	支持RS485云台

云空间Web监控

手机监控



平台控制中心



电脑集控软件

135

Fog Node (Gateway) 安全管理服务器

产品	安全管理服务器（俗称：安防系统“黑匣子”）
产品型号	ALB17
机型	2U标准上架式整机
主板	AM9275主板
处理器	Intel i7 3770 3.4GHZ四核处理器
芯片组	Intel B75芯片组
内存	8GB DDR3 内存
硬盘	1TB 3.5英寸SATA硬盘
功能	<ol style="list-style-type: none">独立安防服务器系统，支持可视化安全管理中心软件功能最大支持300路ONVIF标准摄像机，通过协议转换接入云服务平台，提供大屏视频墙和手机远程监控兼容海康、大华等主流网络摄像机和NVR视频监控设备，监测设备运行数据安全物联网报警主机管理，以及各网格化防区管理巡检视频监控和报警设备，自动记录、报告和统计设备故障情况记录报警日志，存储报警抓拍录像记录机房环境数据，数据异常时发出报警
前面板界面	前置2个USB、HDD LED、Power LED、电源开关按钮
后面板界面	2*千兆网口、4个USB2.0、RS232、HDMI
电源	工业电源，额定功率300W，115V/60Hz、230V/50Hz自我调整
尺寸	2U标准，450mm深
环境	最高50°C(122°F)，非冷凝湿度90%



Public Safety 公共设施安全管理需求



- 公共设施存在的安全管理需求：
 - 电梯井基坑，渗漏水造成积水，导致电梯故障
 - 弱电间或电气竖井，线路老化存在自燃隐患，现有消防系统不易在火灾早期发现和预警
 - 常闭式防火门未处于关闭状态，存在消防隐患
 - 楼顶、地下室等无人值守出入口或通道的非法闯入
 - 小区周界防范报警存在误报，无法及时核警

System Interface

服务控制台

- 联网报警服务
- 实时报警
- 历史记录
- 任务分派
- 系统通知
- 用户设备管理
- 部门人员管理
- 服务公司管理
- 物业安防管理

地图 | 平面图 | 防区 | 视频点
地图模式

Baidu 地图
© 2017 Baidu - GS(2016)2089号 - Data © 长地万方

全部	异常	操作	报警	业务
2017-09-13 18:58:47	设备报警		设备:1212253146[ef308b]	
2017-09-13 18:50:49	设备报警		设备:1212253146[ef308b]	
2017-09-13 18:28:02	设防		设备:1212253146[ef308b]	
2017-09-13 18:27:52	设备报警		设备:1212253146[ef308b]	
2017-09-13 18:22:57	设防		设备:1212253146[ef308b]	

实时报警		系统通知				
任务ID	用户	防区	报警时间	状态	警情类型	用户操作
33480	李鹏(SY1-D2-103)	防区:2 人体红外探测器(ef308b)	2017-09-14 02:58:34	未分派	未知	无
33479	李鹏(SY1-D2-103)	防区:2 人体红外探测器(ef308b)	2017-09-14 02:50:30	未分派	未知	无
33478	李鹏(SY1-D2-103)	防区:2 人体红外探测器(ef308b)	2017-09-14 02:26:57	超时结束	未知	无
33477	李鹏(SY1-D2-103)	防区:2 人体红外探测器(ef308b)	2017-09-14 02:02:02	超时结束	未知	无
33475	李鹏(SY1-D2-103)	防区:2 人体红外探测器(ef308b)	2017-09-14 01:54:39	超时结束	未知	无
33474	李鹏(SY1-D2-103)	防区:2 人体红外探测器(ef308b)	2017-09-14 01:49:59	超时结束	未知	无
33473	李鹏(SY1-D2-103)	防区:2 人体红外探测器(ef308b)	2017-09-14 01:45:38	超时结束	未知	无
33472	李鹏(SY1-D2-103)	防区:2 人体红外探测器(ef308b)	2017-09-14 01:39:52	超时结束	未知	无
33471	李鹏(SY1-D2-103)	防区:2 人体红外探测器(ef308b)	2017-09-14 01:36:20	超时结束	未知	无
33446	苏昊(20170630SH)	防区:1 人体红外探测器(72eb)	2017-09-13 07:30:11	超时结束	未知	无
33445	苏昊(20170630SH)	防区:1 人体红外探测器(72eb)	2017-09-13 07:29:48	超时结束	未知	无
33444	苏昊(20170630SH)	防区:1 人体红外探测器(72eb)	2017-09-13 07:28:45	超时结束	未知	无
33443	苏昊(20170630SH)	防区:1 人体红外探测器(72eb)	2017-09-13 07:28:27	超时结束	未知	无
33442	苏昊(20170630SH)	防区:1 人体红外探测器(72eb)	2017-09-13 07:28:08	超时结束	未知	无
33441	苏昊(20170630SH)	防区:1 人体红外探测器(72eb)	2017-09-13 07:27:44	超时结束	未知	无
33440	苏昊(20170630SH)	防区:1 人体红外探测器(72eb)	2017-09-13 07:27:12	超时结束	未知	无

未分派 2
超时结束 14
已处理 0
已分派 0
已接受 0

其他
录像
暂存
提交

任务ID: 33479

时间: 2017-09-14 02:50:30

用户: 李鹏(177.3331.7965) NO.SY1-D2-103

地址: 一期D2-103, 立鹏药业

设备: (1212253146)

防区: 防区:2 (ef308b)

接警口令:

劫持口令:

接警: 2017-09-13 19:19:40

人员: 沈平(8615999557536)

警情: 请选择

亲友: 已通知 未通知

警察: 已通知 未通知

火警: 已通知 未通知

急救: 已通知 未通知

备注:

138

Looking inside of App Fog

服务控制台

欢迎, 沈平 上海某公安分局(ID:10021) 插件下载(1.0.1) 简体中文

地图 平面图 防区 视频点

实时报警

历史记录 任务分派 系统通知

用户设备管理 部门人员管理 服务公司管理 物业安防管理

任务ID: 33482
时间: 2017-09-13 20:24:08
用户: 沈平(15999557536) NO.20170609SP
地址: 奎州中药城测试 (深圳设备)
设备: 门廊(1401287073)
防区: 防区:0 大门(12722)
接警口令: 123
劫持口令: 456
接警: 2017-09-13 22:26:29
人员: 沈平(8615999557536)
警情: 请选择
亲友: 已通知 未通知
警察: 已通知 未通知
火警: 已通知 未通知
急救: 已通知 未通知
备注:

实时报警 系统通知

未分派 0 超时结束 11 已处理 3 已分派 0 已接受 0

智慧小区

服务控制台

地图 平面图 防区 视频点

实时报警

历史记录 任务分派 系统通知

用户设备管理 部门人员管理 服务公司管理 物业安防管理

实时报警 系统通知

未分派 2 超时结束 8 已处理 4 已分派 0 已接受 0

欢迎, 沈平 上海某公安分局(ID:10021) 插件已启动 Test 简体中文

接警模式 其他 录像 停存 提交 任务ID: 33459 日期: 2017-09-13 14:24:10 20170609SP (深圳设备) 1401287073(1401287073) 大门(12722) 123 456 13 14:37:11 5999557536 季 通知 未通知 通知 未通知 通知 未通知 急救: 已通知 未通知 备注:

提纲

- 1 智慧城市基础：人工智能概述
- 2 大数据应用
- 3 智慧城市物联网国重@UM
- 4 小结

小结

再回过头来讲讲AI
AI成功的例子
目标
未来

例子(1)

- 博弈：IBM公司的“深蓝”成为第一个在国际象棋比赛中战胜世界冠军的计算机程序
- 1997年，一次公开赛中3.5/2.5比分战胜卡斯帕罗夫，他说从棋盘对面感到了“一种新智能”（但是，连“深蓝”的设计者也不认为用了什么人工智能技术）
- Deepmind的AlphaGo打败围棋高手李世石，一次又一次的加速了AI前进的速度。

例子(2)

- 自主控制：CMU研制的ALVINN计算机视觉系统安置在NAVLAB计算机控制微型汽车中，用于汽车导航行驶在高速公路上
- 全程2850英里(约4586.5公里)，其中98%时间由这个系统掌握方向盘，2%时间由人驾驶，几乎都在高速公路出入口处

例子(3)

- 后勤规划：1991年海湾战争中美国军队配备了动态分析和重规划工具DART，用于自动后勤规划与运输调度
 - 该系统同时涉及50000个车辆、货物和人，而且要考虑起点、目的地、路径，解决所有参数之间的冲突。使用AI技术使规划在几小时内完成，而传统方法需要几个星期
 - DARPA称就此一项投资足以补偿DARPA在AI方面30年的投资

例子(4)

- 对抗垃圾信息：每天，学习算法都会识别出上百万条垃圾信息
- 机器人：现在机器人行业也在崛起
- 机器翻译：Google翻译

AI目标(1)

- 完美理性：从环境中获得的信息，智能体每时每刻都以使其效用最大化方式行动
- 计算理性：最终返回理性的选择，考虑和环境交互的时机(错误时刻的正确答案没有价值)
- 有限度理性：思考/训练足够长的时间，得到一个“足够好”的答案只能用于固定环境—缺乏泛化与应用价值（环境改变一点点则错误百出）

AI目标(2)

- 有界最优化(Bounded Optimization):
已知计算资源, 一个有界最优化智能程序的期望 (平均) 效用至少会与同一台机器上运行的其他任何智能体程序的期望效用一样高
 - 至少总存在一个最佳程序

未来

- AI在中级水平上的成功将影响所有人的日常生活
- 目前AI还没有像互连网和手机一样对社会造成普遍深入的影响
- “我们只能向前看到很短的距离，但是我们能够看到仍然有很多事情要做”（图灵）
- 今天各个公司拼命想要AI落地！

社会伦理问题

AI+大数据的应用可能会导致：

- 失业
- 太多/太少时间
- 失去个人隐私
- 责任感丧失
- 失去作为人的独一无二的感觉
- 机器越来越像人，。。。人越来越像机器！
- 这...? 人类种族的终结 ?!

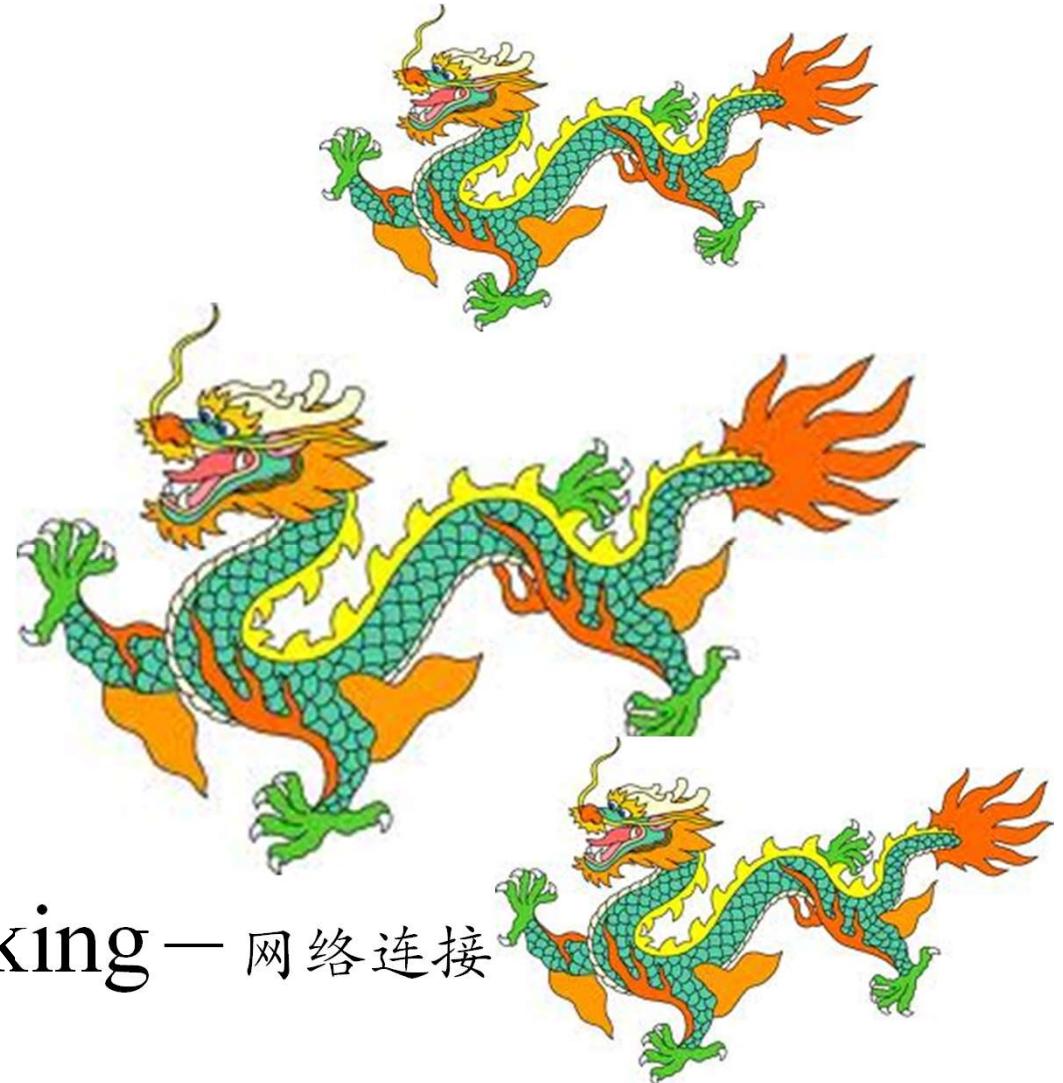
为什么关注社会伦理？

一个只有专业型、科技型的知识分子，而没有公共性、批判性的知识分子的社会是没有人文精神的；而没有人文精神的国度，将是一个什么样的国度？放眼世界，便一目了然。

– 沙叶新（1939-2018.7.26）《往事如雷》
http://blog.sina.com.cn/s/blog_9f1550fb010139ff.html

众龙腾飞 (DRAGONS) !

- Data, data, data
- Relations — 关联
- A.I. — 人工智能
- Games — 游戏
- Openness — 开放性
- (Neural) - Networking — 网络连接
- Smart City — 智慧城市



感谢各位
Q&A