大数据

对智慧城市的意义

清华大学大数据研讨会 2012年11月17日



大数据之形成

大数据

共同作用

1965 摩尔定律

1988 普适计算 1989 数据挖掘 2004 社交媒体 **2006** 云计算

2010: 大数据 (Big Data)

- 1998年,大数据开始明确指代那些大小超出传统软件工具能够捕捉、 存贮、管理和分析能力的数据
- 2000年, Prof. Francis X. Diebold @ University of Pennsylvania 第一次为"大"给出定义: 200太(TB, 2⁴⁰)
- 2012年3月成为和互联网、超级计算同等重要的美国国家战略:投入2 亿多美元启动大数据研发
- "大"是虚指,大数据之大是指人类有能力发现数据当中蕴藏的大价值

从数据到智慧



从信息时代、知识时代向智能时代迈进

- 信息时代:可以迅速找到自己所需要的信息
- 知识时代:人类各个领域都会出现一批优秀的 知识讲解视频,知识无处不在
- 智能时代: 计算机和网络更加智能, 人与人之间的合作, 任务之间的对接会更精确, 整个国家和社会的运行成本会越低
 - 2011年,IBM设计的机器人"沃森"参加综艺节目危险边缘战胜了人类的回答问题的冠军
 - 图书正在数字化、借书的数量不受限制,图书 谷歌在问世、跨语言搜索跳转将实现
 - 谷歌无人驾驭汽车计划5年后上市
 - 机器学习(Machine Learning)成为新的引爆点



IBM设计的机器人"沃森"



谷歌无人驾驭汽车在美国多个州获得上路行驶权

"智慧城市"的起源

- 2008年11月,IBM提出Smart Planet
 - 理念:全面感知、整合协作
 - 其后衍生出Smart City, 2010年12月IBM为巴西里约热内卢完成最大的Smart City项目,整合了30多个城市管理部门的系统和数据, 耗资\$14M
 - Cisco主推Smart Connected Communities
 - 中国"智慧城市"的提法和翻译带有文化含义和更强的战略色彩

理解"智慧城市"

- 智慧城市之定义
 - 利用信息技术对城市的状态进行实时管理、数据分析和预测,为城市管理的方方面面,例 如教育、医疗、公共安全、交通、公用事业,提供更智能、更互动互连、更高效的服务。
 - 不同的维度:智慧校园、医院、银行、智慧社区、智慧电网,智慧路灯、智慧大厦
- 为什么需要智慧城市?
 - 越来越多的城市人口,给城市管理和服务带来巨大的挑战
 - 能源的压力和环境危机
 - 预测: 更有效调配资源、解决问题
 - 智慧城市的核心目标:高效、智能、可持续发展
 - 不仅有经济效益、还有社会效益、环境效益,但后两者难以量化

下一波软、硬件发展的重要助推器



大数据:智慧城市的大脑

- 物联网、云计算、大数据被认为是三大核心技术
- 物联网是城市的数据触角、云计算是躯干
- 数据整合、分析、展现(数据可视化)才是智慧城市的核心问题
- 围绕数据的收集和使用,出现了城市分析学、城市信息学、数据驱动城市等等新的提法
- 智慧城市首先是一个开放的城市: 信息自由, 开放数据
- 不仅仅是技术,社会创新也是智慧城市的一个重要成分

物联网:城市数据的触角

- 2011年10月,美国国家气象局 (NWS) 在2000辆客运大巴上装了传感器,沿途收集温度、湿度、露水、光照度
- 美国国家邮政局 (U.S. Postal Service) 规划在邮车上安装传感器,在投递邮件时采集社区的空气质量、污染指数和噪声数据

- 垃圾管理正在成为中国城市管理者的头痛难题
 - 2009年,MIT开始利用感觉器跟踪垃圾的处理过程
 - 电子垃圾、含有危险成分的垃圾要在系统中平均"旅行" 1500公里,才能最终被处理
 - 智能垃圾桶:发送垃圾桶是否装满的数据,提高垃圾处理的效率

城市的数据开始多元化



精确数据

各类信息系统的数据 ,每一条都有具体的 含义和价值

模糊数据

社交媒体的单条数据 可能没有明确含义、 确定价值,由大众产 生的数据

准增值数据

包括公共数据、例如 人口、天气等等,以 及其它公司的数据

多元数据的使用

私营领域:

迪斯尼

- 100多个景点、100多条队伍
- 利用十多年的历史数据、结合天气数据、旅游数据预测每一条队伍每一天、
- 每一小时所需要的排队等候时间,为游客计算最佳的园内景点游览次序
- 同时实时收集推特数据,处理突发情况,更新每一条队伍的排队等待时间
- 使用这些数据的游客平均每人节省4小时

公共领域:

宾西法尼州政府

- 分析全州感冒药品的销量、对比系统保存的历史数据,以确定可能发生的 大面积流感
- 分析儿童的就症率,对比历史数据,以确定可能发生的大面积流感
- 对推特进行文本分析,实时监控各个地区的流感爆发、传播、分布情况

开放数据和数据权

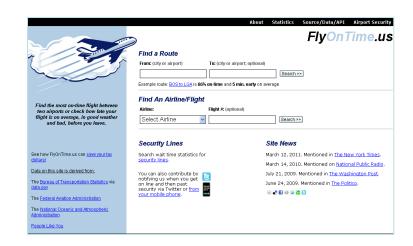
- 公共数据有很大一部分是一个社会的基础性数据,缺少它,数据无法有效整合:天气、人口、地理、经济发展指标、社会福利等等
- 公共数据是用纳税人的钱收集的
- 商业基础数据也应开放,例如公司注册数据、 顾客行为记录
- 创新的力量正在向大众流动: 大众创新

新的"数据权" 最令人激动。这将确保人民有权向政府索取各式各样的数据,用于社会或者商业创新。通过这些措施,我们可以创建一个最开放、最负责和最透明的政府。使用这些数据、开发这些数据,让我们负起责来,一起努力,创建一个现代民主的典范。

——戴维·卡梅伦,英国首相,2010年11月19日

Data.Gov激发的创新

- 由政府主导、向全社会开放公共数据的 Data.Gov,为更多的社会创新、大众创新提供了一个平台
- 2011年12月,Data.Gov网站上开放了 近40万项数据,汇集了1140个应用程序 和软件工具、85个手机应用插件,其中 300多个是由民间的程序员、公益组织 等社会力量自发开发的
- 例子: 航班延误分析系统



	Average (median)	Be Prepared For (85 th percentile)	Cancelled or diverted
In Good Weather	5 min. early	19 min. late	1%
In Bad Weather	3 min. late	53 min. late	7%
In Fog (7%)	12 min. late	87 min. late	14%
In Rain (36%)	4 min. late	54 min. late	7%
In Snow (10%)	7 min. late	53 min. late	12%
In Thunder (5%)	2 min. late	53 min. late	4%
"Be Prepared For"	gives the lo	ngest delav vou can	n reasonably

"Be Prepared For" gives the longest delay you can reasonably expect to occur. Only the unluckiest 15 percent of flights experience longer delays.



Current weather is 25.0 F (-3.9 C) and A Few Clouds, 39% humidity, and 10.00 miles visibility at BOSTON LOGAN INT'L ARPT, US, MA as of Sun, 18 Dec 2011 16:54:00 -0500.

旧金山市政府: HeyGov!

