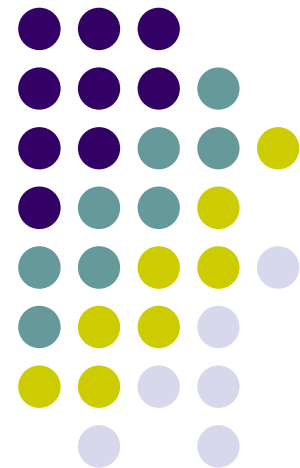




大数据时代的教学变革

王明文
江西师范大学





提纲



- 引言
- 大数据带来的新应用
- 教学挑战



引言



1970年



引言



未来的教育

小班化、多师同堂、在家上学趋势、在线和多媒体教育、回到社区；

培养学生适应临时组织的能力；

培养能作出重大判断的人、在新环境迂回前行的人、敏捷地在变化的现实中发现新关系的人和在未来反复、或然性和长期的设想下的通用技能。



引言



今天，

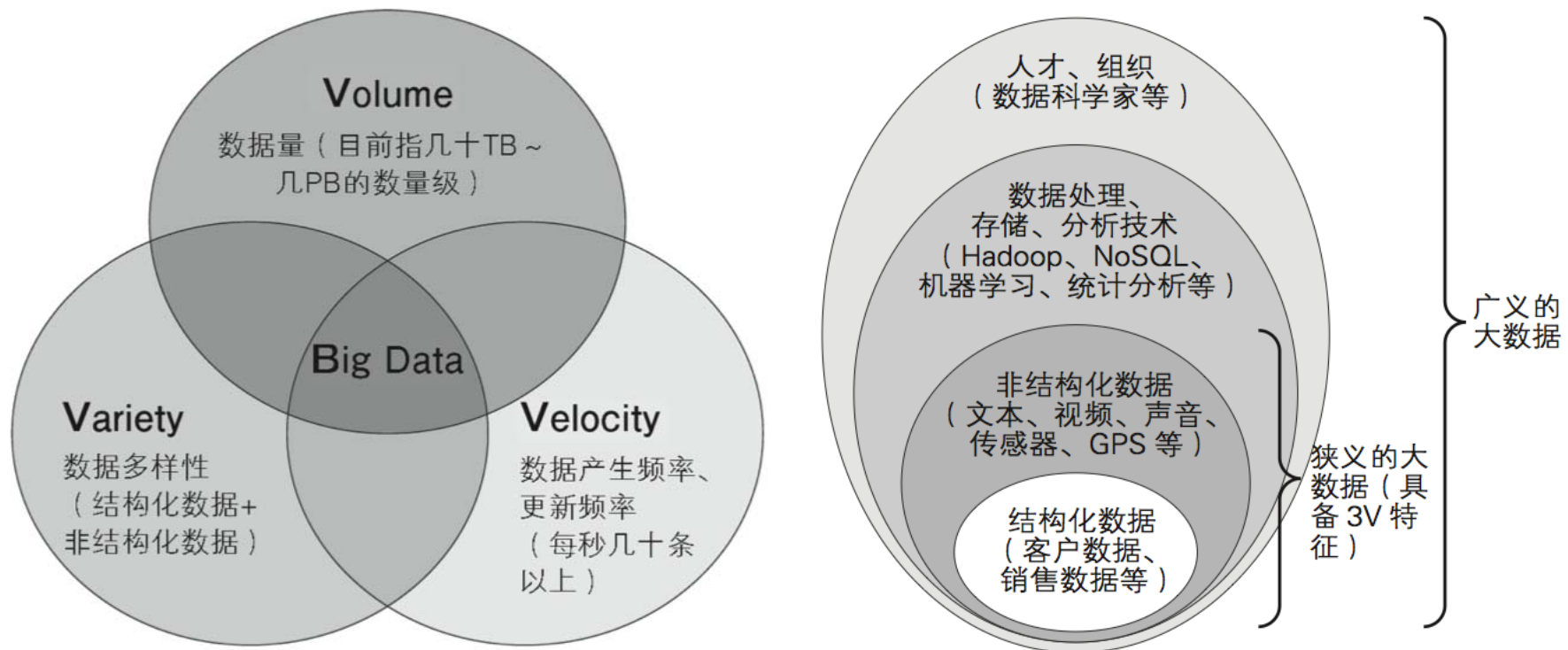
云计算、物联网、数据科学、社交网络

信息不仅仅是一种视觉和感官的东西，更是可捕捉、可量化、可传递的数字存在。

教育正悄悄地发生着一场革命，而今天，我们已经明确知道带来这场革命的真正原因：那就是大数据。



什么是大数据



所谓大数据，就是用现有的一般技术难以管理的大量数据的集合



大数据时代数据理念上的转变



- 抽样 = 全体
- 要效率不要绝对精确
- 要相关不要因果
- 万事万物数据化、数据交叉复用
- 产业生态环境、数据安全隐私、信息公正公开



大数据时代的思维变革——更多

- “更多”——不是随机样本，而是全体数据
 - 当数据处理技术已经发生翻天覆地的变化时，在大数据时代进行抽样分析就像在汽车时代骑马一样。一切都改变了，我们需要的是所有的数据，“样本 = 总体”

- ★ 让数据 “发声”
- ★ 小数据时代的随机采样，最少的数据获得最多的信息
- ★ 全数据模式，样本 = 总体



大数据时代的思维变革——更杂

- “更杂”——不是精确性，而是混杂性
 - 执迷于精确性是信息缺乏时代和模拟时代的产物，只有5%的数据有框架且适用于传统数据结构。如果不能接受混乱，剩下95%的非框架数据都无法被利用，只有接受不精确性，我们才能打开一扇未涉足的世界的窗户

- ★ 允许不精确
- ★ 大数据的简单算法比小数据的复杂算法更有效
- ★ 纷繁的数据越多越好
- ★ 混杂性，不是竭力避免，而是标准途径
- ★ 新的数据结构的诞生



大数据时代的思维变革——更好

- “更好”——不是因果关系，而是相关关系
 - 知道“是什么”就够了，没必要知道“为什么”。在大数据时代，我们不必非得知道现象背后的原因，而是要让数据自己“发声”

- ★ 关联物，预测的关键
- ★ “是什么”，而不是“为什么”
- ★ 改变，从操作方式开始
- ★ 大数据，改变人类探索世界的方法



数据化——一切皆可“量化”

- “数据化”——一切皆可“量化”
 - 大数据发展的核心动力来源于人类测量、记录和分析世界的渴望。信息技术变革随处可见，但是传统信息技术变革的重点在“T”（技术）上，而不是在“I”（信息）上，现在，我们应该把去光灯打向“I”，开始关注信息本身了

- ★ 数据，从最不可能的地方提取出来
- ★ 数据化，不是数字化
- ★ 量化一切，数据化的核心
- ★ 当文字、方位、沟通变成数据
- ★ 一切事物的数据化



大数据带来的新应用



1、PRADA的试衣间



PRADA（普拉达，一个意大利的奢侈品牌）在纽约的旗舰店中每件衣服上都有RFID（射频识别）码。每当一个顾客拿起一件PRADA衣服进试衣间，RFID会被自动识别。同时，数据会传至PRADA总部。每一件衣服在哪个城市哪个旗舰店什么时间被拿进试衣间，停留多长时间，数据都被存储起来加以分析。如果有一件衣服销量很低，以往的做法是直接干掉。但如果RFID传回的数据显示这件衣服虽然销量低，但进试衣间的次数多，也许这件衣服的下场就会截然不同，也许对某个细节作微小改变就会重新创造出一件非常流行的产品。



2、中国的粮食统计



采用遥感卫星，通过图像识别，把中国所有的耕地标示、计算出来，然后把中国的耕地网格化，对每个网格的耕地抽样进行跟踪、调查和统计，然后按照统计学的原理，计算(或者说估算)出中国整体的粮食数据。这种做法是典型采用大数据建模的方法，打破传统流程和组织，直接获得最终的结果。



3、公安部门的“犯罪地图”



北京市怀柔区警方通过运用大数据、云计算和科学分析模型，整合历年案件信息，建立了犯罪数据分析和趋势预测系统，能够预测犯罪趋势，指导警力投入。这套系统共收录了怀柔区近9年来1.6万余件犯罪案件数据，通过标准化分类后导入系统数据库，同时采用地图标注，将怀柔分成16个警务辖区，抓取4700余个犯罪空间坐标，实施空间网格编号。



3、电子商务



Amazon基于大量购买历史记录和点击流数据做出了“购买了本商品的顾客还购买了……”的商品推荐功能，这种做法现在已经随处可见了，但像这样为客户推荐合适的商品，过去只有经验丰富的销售人员和熟悉客户的店员才能做到，是“具有人类属性”的行为，现在却能够由计算机来完成，这一点是具有划时代意义的



4、Facebook的社交网络圈子



Facebook十分重视“您可能认识.....”的功能，并对用户找到好友需要的时间进行了监控

- 通过运用精确的用户追踪技术和分析技术，Facebook掌握了一个规律，即如果一个用户能够在一定时间内找到一定数量以上的好友，则该用户就很可能长期使用Facebook
- 事实上，Facebook为了能够让新用户尽早找到一定数量以上的好友，在服务的设计上倾注了大量的心血



QQ中的朋友推荐



The screenshot displays the QQ desktop application interface. The main window is titled "圈子 - 在这里,在一起" (Circle - Here, Together). The left sidebar shows the user's profile (566) and a list of friend groups: "研究生同学 (226)", "分组 (10)", and "其他 (29)". The main area shows a grid of recommended friends with their avatars and names. A notification banner at the top of the grid says "来, 分享你的头像历程 立即查看 x".

Avatar	Name	Text
	李凯	突然有个疯狂的想法!
	吕宏伟	闻道有先后, 术业有专攻!
	陈亚慧	天怎么这么暗啊
	吴哲林	哥对盲僧的理解已经突破了天际-0-
	吴润秀	It is through isolation and dedication
	郭茂林	
	刘凯	知行合一, 无欲则刚
	许生模	
	李智彪	
	梁艳招	
	化志章	学而不思则罔, 思而不学则殆。
	易云	小宝健康快乐, 茁壮成长。
	盛俊	
	乐丹	2013年已过半了! 主动沟通, 提
	杨乐	跟着时代的脉搏走!
	周世科	
	王乐	南昌待留4天, 重温南大翻墙复习考
	李云洪	命里有时终须有, 命里无时莫
	许璟	压力, 如影随行
	段润英	追求心灵的富足
	钟旭	大数据时代, 我能做什么
	肖小方	诡异天气, 郁闷心情
	邹成武	讨厌自己.... [会话] [加好友]
	熊能	正式启用新号 18107910411,

At the bottom of the window, it says "研究生同学, 共226人, 其中QQ好友51人, 当前筛选条件为 '全部'". The status bar at the very bottom shows "v2.5.5" and "应用".



微博



请订阅
我的微信



- 添加微信号
李开复/kaifu
- 扫描二维码



李开复 (设置备注) <http://weibo.com/kaifulee>

创新工场CEO, 媒体联系: press@chuangxin.com

射手座 | 北京 东城区 | 公司创新工场 | 标签

514 48846565 12566
关注 粉丝 微博

已关注 | [取消](#)

新浪认证

创新工场董事长兼首席执行官
[申请认证](#)

他的主页 个人资料 关注/粉丝 相册

全部 微博

搜索他说的话

全部 原创 图片 视频 音乐 标签 心情

什么是动态 动态隐私设置

1) 曾太太、大女儿在牢里, 是否见到了曾成杰最后一面? 2) 邮寄给女婿, 收到时已处决, 不算通知, 3) → → [@曾成杰之女](#):我爸爸知道我们的电话号码, 再说法院肯定有爸爸的律师的联系方式! 法院的这种做法, 直接剥夺了我父亲的临终告别权和我们的临终会面的权利, 对这种不人道的行为我们表示极大的愤慨!

[@人民日报](#)

【湖南高院新闻发言人详解曾成杰集资诈骗案】湖南省高级人民法院新闻发言人曾鼎新: 据我们了解, 长沙市中级法院于执行当日将曾成杰的死刑执行布告贴于法院公告栏内。因为案卷材料中没有曾成杰的二女儿、儿子的联系方式, 故当日执行后法院将执行通知书邮寄送达曾成杰的女婿(大女儿丈夫)。详细 [↓](#)



微关系

我们之间的共同关系

共同关注 (4) [更多 >](#)



吴军博士



王小川



张宏江



张亚勤

我关注的人也关注他(58) [更多 >](#)



蔡-东风



丁国栋_1...



叶浩



刘群MT-t...

他的粉丝还关注了 [更多 >](#)



薛蛮子



汪华



潘石屹



蔡学镛

[+ 关注](#)

[+ 关注](#)

[+ 关注](#)

[+ 关注](#)

INNOVATION WORKS
[返回顶部](#)



微博推荐



为你推荐

关注人太少？那怎么能知天下呢？



JDZ宝石村  

简介：退一步，海阔天空！
你们可能是朋友

[+ 关注](#) 健康



白色流星1022

你们可能是朋友

[+ 关注](#) 游戏动漫



Z_箬

简介：...人在荆棘中，不动不刺；心在...
你们是同学

[+ 关注](#) 腐 看书



JJL小璇_时线TimeLine 

简介：我怎么走怎么还原 怎么飞回那...
你们是同学

[+ 关注](#) 各种宅 小清新



宋佳Sylvia

简介：一切的一切都可以是真的，只要...
Ta可能在你附近

[+ 关注](#) 森林少女 隐隐约约的



罗永浩  

锤子科技创始人，老罗
我关注的人中：[王海勋...](#) 等11人关注了他
彪悍的人生不需要解释

[+ 关注](#)

[我不想被别人找到，去隐私设置](#) [关注他们](#)



5、Google的服务



- Google利用“搜索历史记录”这一用户看来毫无用处的“数据垃圾”，接二连三地推出有价值的新服务，如智能关键字修正、手写输入、Google翻译、语音搜索等
 - 大数量的数据往往要胜于优秀的算法。这句话的意思是，相比用复杂的算法来识别每一条新输入的数据来说，对大量存储的正数据进行分析，在统计学上往往能够得出最合适的结果
 - Google的智能关键字修正（您要搜索的是.....），是对每月900亿次的搜索记录进行分析，找出用户在搜索时可能打错的，或者是输入法转换错的关键字，以及之后又重新输入的，或者是用户点击的正确关键字，通过机器学习的方式来进行分析处理



6、客户行为分析



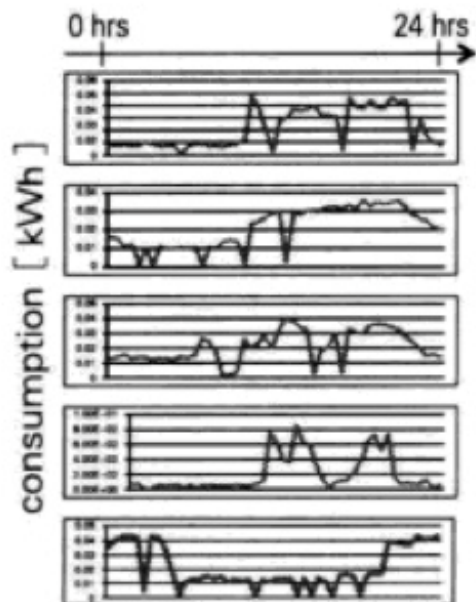
- 在超市中，可将由植入购物车中的IC标签收集到的顾客行动路线数据，和POS等销售数据相结合，从而分析出顾客买或不买某种商品的理由
- 此外，还可以通过分析监控摄像机的视频资料，来分析店内顾客的行为
- 以前也有对店内的购买行为进行分析的方法，不过，那种分析大多由调查员肉眼观察并记录，这种记录是非数字化的，成本很高，而且收集到的数据也比较有限



7、能源



通过分析客户的能源消耗模式，在客户分类、费用菜单开发、将来消费趋势预测等方面进行运用



每天超过数千万条的大量数据

① 在如此大量的关于能源消耗模式的数据中，蕴含着让客户更好地进行能源管理的有价值信息。但是，仅仅将来自电表的数据进行可视化还不够。

② 模式识别算法对内存中大量的采样数据进行高速处理，从而归纳出客户的使用模式。

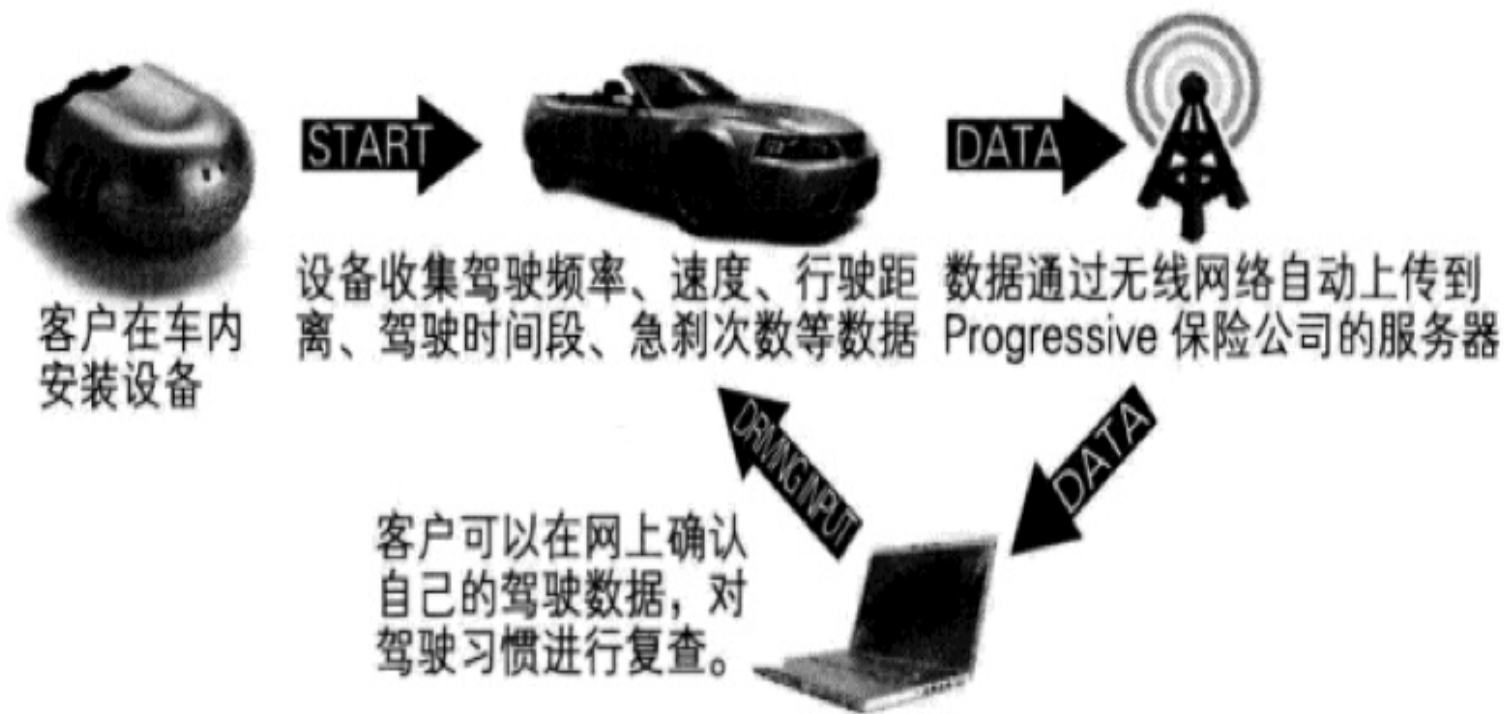
③ 不必对上百万签约用户的数据逐个进行确认，只要看几个典型的使用模式，就足够把握客户的行为特征了。

- 根据使用数据进行用户分类
- 费用菜单的开发
- 对电表采集数据的准确性进行检查
- 对未来消费趋势的预测

对智能电表的大量数据进行模式识别的实例



8、保险业



美国 Progressive 保险公司提供的根据客户驾驶习惯对保费给予相应折扣的 Pay as You 计划



9、教育



在美国宾州，有一个叫做EDLINE的网站，将学生的每次作业、每次考试记录在网上，完成学生的日常GPA积累，这个网站的技术并不难，然而能够坚持下来的数据积累，对于学生、家长和教育管理非常重要，大家都知道，美国的大学入学GPA非常重要。依靠这个GPA 再加上学生的SAT和ACT所提供的分析报告以及志愿者活动资料，就决定了学生的大学去向。



10、足球训练



SAP Match Insights足球解决方案

球员的鞋内、护胫中被放置了传感器，偌大的训练场地内也到处布满传感器。通过这些装置，SAP的系统得以捕捉球员的各种细节动作与位置变化。他们的跑动及传球路线被实时传回到SAP HANA平台上。

通过该系统对球员跑动、传球等数据的捕捉和分析，教练能够评估每场比赛的主要状况和每个球员的特点，并以“数字和事实”来优化备战方案，提升球队的成绩。



教学挑战



Big Data for Education:

Data Mining, Data Analytics,
and Web Dashboards

Darrell M. West

Brookings Institution
美国布鲁金斯学会



未来的学习场景



12岁的苏珊正在进行一门旨在提高她的阅读技能的课程。她一直在阅读一些短篇小说，每隔一周，老师都会给她和她的同学进行纸笔测试，测量他们词汇和阅读理解能力。几天后，苏珊的老师批改完试卷，并返回成绩。测验表明，她的词汇掌握得不错，但在关键概念上还需加强。



未来的学习场景



她的弟弟理查德是通过电脑软件程序学习阅读。

他每阅读一个故事，计算机都会收集他学习过程的相关数据。每个任务后，会弹出一个小小测验在他的屏幕上，是有关词汇和阅读理解的问题。

理查德每回答一个问题，他都将得到即时反馈，显示他的答案是否正确。

对于困难的题目，电脑会推送给他更详细的解释词语和概念的网站链接。

他的综合表现情况会可视化地显示在一个仪表盘上，包括作业和测试的正确率，已经掌握的概念列表，以及学习表现与同学乃至全国的学生的比较情况等。



未来的学习场景



在一个学习环节结束时，他的老师会收到一封自动邮件，其中显示了理查德和班上其他学生的个人表现以及全班情况的汇总，包括阅读时间，词汇知识，阅读理解，补充电子资源的使用情况等。

教师根据汇总信息，能很快发现需要额外帮助的学生，学习时间不够的学生，以及全班大部分学生都有困难的内容。



未来的学习场景



教师可能会通过集中讲授、个别辅导，人为干预学习系统以适合学生的学习步调，或为没掌握的技能增加额外学习材料。



未来的学习场景



学校管理者，通过数据分析可视化仪表盘能查看年级、全校乃至整个学区的学生的学习情况。如果发现低效的课堂和学习表现不佳的学生群体（根据性别，收入情况等划分），管理者会依据学习分析结果决定是否给予特定的干预。对于更大范围的异常表现，管理者会依据学习分析数据调整管理策略，以适应教师更好的教和学生更好的学。



挑战一：大数据如何实现自适应学习



挑战二：大数据如何实现个性化学习



挑战三：大数据如何关注学生运动健康



挑战四：大数据如何识别高危学生



挑战五： 如何获取数据



挑战六：各种教学数据格式的规范



挑战七：学校如何使用数据



谢谢！