



大数据热点问题和 2013年发展趋势预测

CCF大数据专家委员会专家调研结果发布
2012年11月30日





调研的目标、方法和过程



- **70位委员**
 - 学术界46位、产业界14位、海外10位
- **思考热点问题**
- **给出2013年大数据发展趋势预测**

CCF大数据专家委员会



热点问题候选项

一、科学问题

- C.1 数据的科学问题 (55)
- C.2 大数据的基本内涵 (43)
- C.3 计算模式 (61)

二、技术问题

- C.4 数据的多样性与数据态 (53)
- C.5 大数据的空间维问题 (44)
(分布、感知与传输)
- C.6 大数据的时间维问题 (63)
(流式化、时效性、在线处理)

- C.7 如何将大数据变小 (42)
- C.8 数据的价值提炼 (49)

三、应用实践问题

- C.9 大数据应用领域 (49)
- C.10 大数据对于系统的要求(49)
- C.11 数据质量问题 (41)
- C.12 大数据安全和隐私 (59)

四、大数据生态问题

- C.13 数据资源化和共享管理 (33)
- C.14 大数据的生态环境 (32)



【1】数据科学与大数据的学科边界

55

数据的科学问题

- 数据界与物理界、人类社会之间的关联与差异?
- 是否存在独立于领域的数据科学?
- 数据科学的分类谱系
- 大数据的复杂性主要来自个体之间的联系
- 学习理论和认知理论等应当是数据科学的重要组成部分

大数据的基本问题

(定义、结构等)

43

- 需要对“大数据”给出科学定义，清晰说明其内涵与外延
- 大数据区别于其他数据的关键特性是什么?
 - 3V
 - 高价值总量、低价值密度
- 大数据意味着全数据?
- 需要为动态、高维、复杂大数据建立形式化、结构化描述，并在此基础上发展大数据处理技术



【2】 数据计算的基本模式与范式

61

- 数据密集型计算的基本范式?
- 数据计算的效率评估与数据计算复杂性理论?
- 从中心化的/top-down模式转为去中心化的/自组织的计算模式?
- 基于数据的智能：会有越来越多靠“数据的体量+简单的逻辑”的方法去解决复杂问题



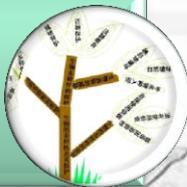
【3】 大数据特性与数据态

53

- 多来源多模态数据：图像、视频、音频、数据流、文本、网页...
- 关联关系异质、结构模式复杂
- 互为因果，动态变化

关系维

复杂性



44

- 三元空间大数据的产生、状态感知与采集
- 柔性粒度数据传输、移动、存储与计算
- 数据空间范围和密度的非均

空间维

复杂性



63

- 数据的生命周期
- 数据的时间维状态与特征
- 流化分析、增量学习、在线推荐
- 离线与在线时效性要求

时间维

复杂性





【4】大数据的作用力和变换反应

物理



将大数据变小（物理作用）

42

- 在尽量不损失价值的条件下，减小数据规模
- 不改变数据基本属性情况下的数据清洗
- 抽样、去重、过滤、筛选、压缩、索引、提取元数据等等方法，可以直接将大数据变小，这种作用类似于所谓的物理式的变小

化学



价值提炼（化学反应）

49

- 大数据探索式考察与可视化将发挥作用，人机的交互分析可以将人的智慧作用融入
- 通过群体智慧、社会计算、认知计算对数据价值的发酵和提炼
- 从数据分析到数据制造



【5】 大数据安全和隐私问题

59

- 安全和隐私，永远的问题
- 随着数据的增多，组织面临的重大风险跨越了一个复杂的威胁面，要遵守更多合规规定，传统的数据保护方法常常无法满足

• 挑战

大数据规模的密码学
分布式编程框架中的安全计算
非关系型数据存储
安全的数据存储和事务日志
终端输入的确认/过滤
实时安全/合规监测
可扩展的、可组合的、脱敏(无隐私)的数据挖掘和分析
强制的访问控制和安全通信
粒度访问控制
数据来源和数据通道



【6】 大数据对IT技术架构的挑战

49

- 大数据对于系统提出了很多非常极限的要求。不管是存储、传输还是计算
- 大数据平台将是技术高峰。这个平台可能是计算平台、传输平台、存储平台、交互平台等等
- 现有数据中心技术难以满足大数据需求，整个IT架构的革命性重构势在必行。
- 存储能力的增长远远赶不上数据的增长，设计最合理的分层存储架构已成为信息系统的关键。分布式存储架构不仅需要scale up式的可扩展性，也需要scale out式的可扩展性
- 大数据也许会颠覆整个系统结构



【7】大数据的应用及产业链

49

- 大数据一定要与领域知识相结合。在不同的领域环境和不同的应用需求，大数据的获取、分析、反馈的方式都会不同
- 针对不同行业与领域业务需求，展开数据特征与业务特征的研究，进行大数据应用分类与技术需求分析，构建从需求分析→业务模型→数据模型→数据采集→数据分析→总结反馈→再到数据分析的全生命周期应用模型
- 不同的应用环境和应用目标代表了不同的价值导向，对于大数据的价值密度有很大影响



【8】大数据的生态环境问题

资源和共享管理

33

- 数据所有权，既是技术问题，也有法理问题
- 数据权益的认定与保护
- 如何保护多方利益的前提下解决数据共享问题
- 共享可能遇到的障碍：
 - 法律或信誉顾虑
 - 保护竞争力的需要
 - 数据存储的位置和方式不利于数据的访问和传输等

其他生态问题

32

- 涉及与政治、经济、社会、法律、科学等等的交叉影响问题
- 大数据将对国家治理模式、企业的决策、组织和业务流程、个人生活方式都将产生巨大的影响，影响模式值得研究



发展趋势预测的候选项

☀ 整体态势和发展方面

- T1. 更大的数据
- T2. 数据资源化
- T3. 数据价值凸显
- T4. 数据的私有化
- T5. 数据共享联盟

☀ 大数据与学术

- T6. 最令人瞩目的学科是： ...
- T7. 数据科学兴起
- T8. 数学学科发生变革

- T9. 自组织计算
- T10. 基于海量知识的智能
- T11. 大数据分析的革命性方法
- T12. 学术活动盛行

☀ 大数据与人

- T13. 更奇妙的人机互连
- T14. 分析平民化
- T15. 数据化生存

☀ 大数据的安全和隐私

- T16. 大数据隐私问题
- T17. 大数据安全
- T18. 数据保护是国家安全

☀ 大数据应用

- T19. 最令人瞩目的应用是： ...
- T20. 决策支持
- T21. 大企业大数据

- T22. 精确个性化推荐系统
- T23. 数据清洗
- T24. 政府大数据
- T25. 大数据加强军队
- T26. 犯罪预测

☀ 大数据系统和处理

- T27. 处理能力难以满足需要
- T28. 处理模式多样化
- T29. 网络压力突出

☀ 大数据对产业的影响

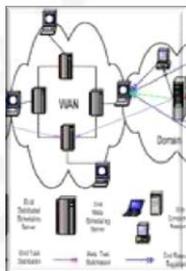
- T30. 资本高度关注
- T31. 产业政策
- T32. 非结构化数据处理
- T33. 数据产品丰富
- T34. 数据外包
- T35. 产业垂直整合
- T36. 大数据新职业
- T37. 大数据与云计算等深度融合



最令人瞩目的学科



大数据分析 & 预测 (25/40)



分布式计算 (25/40票)



社会计算 (20/40票)



最令人瞩目的应用





候补项 (排名11/12, 22票)

大企业大数据

- 传统大企业继续数据化，并开始采用大数据技术。
- 舆情监控将从政府应用过渡到企业应用

资本高度关注

- 资本高度关注大数据领域，相关的融资、并购、IPO纷纷出现
- 大数据行业内，大企业并购一些有特色的中小企业将成为热潮
- 2013年将真正开始投资。



【10】更大的数据（24票）

- 现在的大数据，将来都不够大
- 2013年，大数据的获得更多的关注、研究、开发和应用。而所带来的现象就是，体现大数据特征的体量大、快速、多样性、价值密度低等的几个V特性将更加极致。
- 大数据的价值密度会越来越低——数据不断地增长，如何去除大数据中的噪点，从中挖掘和提取出有价值信息的难度也随之增大。



【9】 大数据新职业 (25票)

- 大数据产生一批新的就业岗位，如：**数据分析师、数据科学家、数据工程师**
- 具有丰富经验的数据分析人才成为稀缺资源，数据驱动型工作机会成爆炸性增长
- 大数据领域最优秀的科学家们纷纷转行**股票、期货，甚至赌博**（能比别人多看远一秒钟，就是效益）。



【8】 数据共享联盟 (28票)

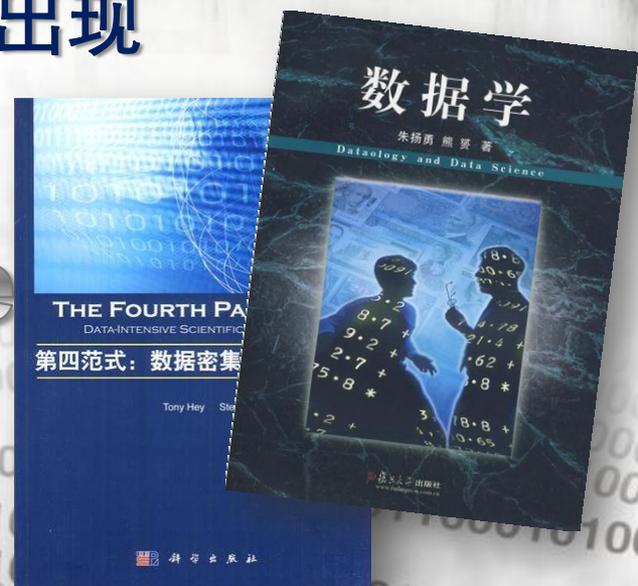
- 数据共享联盟将逐渐壮大成为产业的核心一环
- 数据是基础，之前在科技部的支持下，已建立了多个领域的数据共享平台，包括气象、地震、林业、农业、海洋、人口与健康、地球系统科学数据共享平台。之后、数据共享将扩展到企业层面。



【7】 数据科学兴起 (29票)

- 数据科学作为一个与大数据相关的新兴学科出现
- 将有专门针对数据科学的专业形成，有博士、硕士甚至本科生出现。
- 有大量数据科学的专著出现

Dataology
Data Science



CCF大数据专家委员会



【6】 大数据安全 (29票)

- 大数据的安全令人担忧
- 大数据的保护越来越重要---大数据的不断
增加，对数据存储的物理安全性要求会越来越高，从而对数据的多副本与容灾机制提出更高的要求。
- 网络和数字化生活使得犯罪分子更容易获得关于人的信息，也有了更多不易被追踪和防范的犯罪手段，可能会出现更高明的骗局。大数据已经把你出卖。



【5】大数据分析的革命性方法 (31票)

- 在大数据分析上，出现革命性的新方法
- 就像计算机和互联网一样，大数据可能是新一波技术革命。基于大数据的数据挖掘/机器学习和人工智能可能会改变小数据/小世界里的很多算法和基础理论,这个方面很可能会产生理论级别的突破





【4】基于海量知识的智能 (32票)

- 会有更多基于海量知识的智能成果出现。
 - 人工大脑的可能性，至少Chinese Room彻底解决——能想到的问题，问之前就都已经被人回答过了。在没有思考和逻辑的情况下，可以利用前人的经验，同样可以起到脑的功能。甚至也是可能通过大数据进行推理的。



【3】 大数据与云计算等深度融合 (40/70票)

- 大数据处理离不开云计算技术，云计算为大数据提供弹性可扩展的基础设施支撑环境以及数据服务的高效模式，大数据则为云计算提供了新的商业价值，大数据技术与云计算技术必有更完美的结合
- 云计算、物联网、移动互联网等新兴计算形态，既是产生大数据的地方，也是需要大数据分析方法的领域。



【2】 大数据隐私问题 (44/70票)

- 大数据对于隐私是一个重大挑战
- 2013年隐私相关的标准和条例颁布
- 现有的隐私保护法规和技术手段难于适应大数据环境
- 有偿隐私服务可能出现
- 个人隐私越来越难以保护
- “面罩” 流行



【1】数据资源化 (51/70)

- 大数据在国家企业和社会层面成为重要的战略资源
- 数据成为新的战略制高点，是大家抢夺的新焦点。
- 大数据将不断成为机构的资产，成为提升机构和公司竞争力的有力武器





大数据 大价值

八大热点问题

- 1、数据科学与大数据的学科边界
- 2、数据计算的基本模式与范式
- 3、大数据的作用力和变换反应
- 4、大数据特性与数据态
- 5、大数据安全和隐私问题
- 6、大数据对IT技术架构的挑战
- 7、大数据的生态环境问题
- 8、大数据的应用及产业链

2013年十项发展趋势预测

1. 数据的资源化
2. 大数据的隐私问题突出
3. 大数据与云计算等深度融合
4. 基于大数据的智能的出现
5. 大数据分析的革命性方法
6. 大数据安全
7. 数据科学兴起
8. 数据共享联盟
9. 大数据新职业
10. 更大的数据



大数据共享联盟（Big Data Alliance）

- 隶属于中国计算机学会大数据专家委员会，由大数据相关产学研机构和个人自愿组成的学术性、公益性的社会组织。
- 宗旨是建立并持续积累和维护运行一个大数据收集、展示和共享平台，促进各种基于大数据的研究与开发应用。

大数据联盟
共享平台

大数据收集

大数据展示

大数据共享

促进大数据研发

CCF大数据专家委员会



大数据共享联盟 (Big Data Alliance)

更多联盟信息，敬请关注报告：

12月1日 (周六)
16:30-17:00

大数据共享倡
议及实践

齐红威
(数据堂)

地点

主题论坛五：大数据的技术挑战与发展趋势 (昆明厅ABC)

Track 5: Challenges and Trends in Bigdata (Kunming ABC)

CCF大数据专家委员会