学科动态专题报道

2015 年第 1 期

大数据专题

主办者：图书馆学科服务部

2015.1

为传播科学知识，促进业界交流，特编辑《学科动态专题报道》，仅供个人学习、研究使用。

前 言

大数据正在以不可阻拦的磅礴气势，与当代同样具有革命意义的最新科技进步 (如纳米技术、生物工程、全球化等)一起，揭开人类新世纪的序幕。可以简单地说，以往人类社会基本处于蒙昧状态中的不发展阶段，即自然发展阶段。现在，这一不发展阶段随着 2012 年的所谓“世界末日”之说而永远成为了过去。大数据宣告了 21 世纪是人类自主发展的时代，是不以所谓“上帝”的意志为转移的时代，是“上帝”失业的时代。

对于地球上每一个普通居民而言，大数据有什么应用价值呢?只要看看周围正在变化的一切，你就可以知道，大数据对每个人的重要性不亚于人类初期对火的使用。大数据让人类对一切事物的认识回归本源；大数据通过影响经济生活、政治博弈、社会管理、文化教育科研、医疗保健休闲等等行业，与每个人产生密切的联系。

本期学科动态报道选取“大数据”专题，主要分如下几个模块进行跟踪：

《海外资讯》专栏选取国外“大数据”的相关资讯进行翻译，信息主要来源于国外专业的大数据网站及大数据相关会议网站，内容涵盖国外大数据研究现状、会议议题内容及国外专家学者对大数据相关观点。

《SSCI 高被引文献推介》专栏内容来自 Web of Science 的社会科学引文索引（SSCI），SSCI 收录了世界上不同国家、地区的社会科学期刊和论文，并进行了一定的统计分析，划分了不同的因子区间，是当今社会科学领域重要的期刊检索与论文参考渠道。本专栏将 SSCI 中关于大数据的一些高被引文章进行汇总、介绍，推介给大家。

《国内资讯》专栏信息主要来自中国信息产业网、光明网以及各种大数据相关网站，将国内关于大数据的最新报道呈现给大家，以供交流参考。

《国内文献计量分析》的工具主要是 CNKI，通过该工具分析“大数据”学科的学术关注度和用户关注度、研究热点和趋势等内容，为研究人员提供参考、研究材料。

目 录

【海外资讯】.....1
大数据与金融工作会议 1
在线跟踪和隐私：埃尔文德·纳拉亚南演讲摘要 4
为何深度分析变得如此重要 6
大数据量需要更少的车道 7
穿越数据湖泊：数据中心 9
大数据有潜力改革医疗保健 11
【SSCI 高被引文献推介】.....1
【国内资讯】.....8
盘点 2014:大数据现状与国人思维误区 8
2015 令人心动的新兴行业大数据行业上榜 10
大数据将加速形成新的技术经济范式 17
有数据就是这么任性，2014 年谁在玩转大数据？ 19
不仅仅是机遇 细数大数据领域待解决问题 21
Connection Analytics 引领下一代分析技术.....23
2015 贵阳国际大数据产业博览会将于五月开幕 25
当当举办"中国童书年会" 大数据描绘中国童书全景图.....28
大数据开放可提升政府公信力 30
【国内文献计量分析】 33
“大数据”学术关注度 33
“大数据”用户关注度 34

“大数据”热门被引文章.....35
“大数据”热门下载文章.....36
“大数据_数据处理”研究热点.....37
“大数据_数据挖掘”研究热点.....37
“大数据”2013 年立项课题.....46

【海外资讯】

大数据与金融工作会议

张春玲 编译 郝晓雪 校对

麻省理工学院的特别会议聚集了来自行业和政府的利益相关者去分享想法和探索金融方面的大数据应用。会议强调了目前正在进行的研究项目例子，这些项目可能对未来的金融和经济数据是如何收集、管理和使用产生重大影响。该会议也提供了一个论坛用于讨论不同的优先权和行业面临的挑战以及在麻省理工学院由创立一个大数据的价值+金融领域组的价值。

动机

伴随过去二十年技术的进步金融市场经历了快速转型，银行、资产管理公司、管理数字支付公司、电信、IT 企业、政府和监管机构这些金融部分机构都产生、存储和管理了大量数据。我们的目标旨在更好的理解大数据在金融和经济学的角色，并展示其社会和商业价值以及识别潜在的缺陷和挑战。

当涉及到有关共享和使用包括个人和消费等不同类型的数据，麻省理工学院处于一个独特的位置去汇集整个价值链的利益相关者和计算机科学、统计、经济、金融和管理学科这些跨学科的研究人员；推进这些领域中大数据的使用；告知下一代大数据使用系统的发展；推进数据分析方法的提升；解决隐私和监管问题的挑战。

研究进展

这次会议期间，由 John Guttag 教授和麻省理工学院的杰出代表 Andrew Lo Guttag 指导，由斯隆商学院提供了一些正在进行大数据与金融有关研究的例子:

斯隆管理学院的应用经济学教授 Roberto Rigobon 讨论了十亿价格项目(BPP)，BPP 是一个正在进行的利用收集每天来自世界各地成百上千的在线零售商的商品价格进行经济研究项目。BPP 的目标是构建全球经济的新措施和报告协议，尤其是通货膨胀率方面，继而令中央银行的工作更有针对性。BPP 对数以百万计的不同种类的产品和服务的价格进行收集(例如在 26000 个城市乘坐出租车的价格和在每一个国家一件 H&M 衬衫的价格)，然后用这些数据来计算实际通货膨胀率—辅助时间。今天，统计办公室所能做都受到严格的监管(确保没有操纵)，他们开发新方法的能力也因此受到限制。在麻省理工学院，我们可以创新和实验去学

习新的和有效的收集数据和计算重要经济数据的方法，以及与统计办公室和其他能够把这些知识付诸实践机构一起合作。 （网址：http://bpp.mit.edu/）

CSAIL 的副教授 Nickolai Zeldovich 讨论了 VerSum，它是一个允许客户将昂贵的计算规模庞大且经常变化的数据结构外包，同时还确保了客户能够验证结果的系统。例如，激励的应用程序是分析手机比特币交易，它难以有效执行给定的比特币率的大小。VerSum 表明只要其中一个服务器仍然是“诚实”，可通过多个服务器管理和保证由大数据组成的不同计算。VerSum 是辅助效果自由和允许重用计算而有效地验证每个计算步骤。(网址：http://people.csail.mit.edu/nickolai/papers/vandenhooff-‐versum.pdf）

EECS 职业发展助理教授 Vinod Vaikuntanathan、Steven 和 Renée Finn 讨论了研究密码学领域的最新进展的全同态加密方案 (FHE)，他允许通过不同的媒介和与密钥时加密但不失其计算功能。保持数据安全免遭内部攻击或数据泄露是机构面临的挑战，此该项目研究是一个重大进展。例如，新创建的一般功能加密(GFE)意味着金融服务公司将加密的数据存储在云端，同时计算系统测量和风险管理的总曝光。就未来而言，Vinod 说，“加密将允许你有蛋糕(隐私)并且吃掉它(计算)!”更多信息详见：“同态加密可以实现吗？”

（网址：https://eprint.iacr.org/2011/405.pdf; http://www.cs.toronto.edu/~vinodv/FHE-‐focs--‐survey.pdf）

斯隆管理学院金融学教授 Antoinette Schoar 讨论如何利用行为经济学和消费金融心理学从金融交易数据了解消费者行为。该研究表明，金融产品微小变化可以显著影响消费者行为和金融决策。Antoinette 认为大数据为更好地理解人类行为提供契机，以便其提供更好的用户定制的金融商品和服务，并能提高客户的体验和降低成本。更多信息详见：“记住支付吗？提醒还是提醒和金融激励贷款？” （网址：http://www.nber.org/papers/w17020）

斯隆管理学院的金融教授兼 CSAIL 首席研究员 的 Andrew Lo 提供了对几个项目的进行了简短概述，包括系统性风险的隐私保护措施研究展示了使用安全的多方计算的价值，其允许金融服务公司分享数据，继而监管机构可以使用该数据计算系统性风险管理的总曝光。

https://secure.brightworkinc.net/~andrewlo/documents/Abbe Preserving_Methods.pdf

通过机器学习进行消费信贷风险管理项目，该项目使用大型银行和信用咨询公司的数据检测模式，其能预测消费者违约和拖欠行为。通过使用大量的银行和 FICO 数据能准确的预测消费者何时将会发生违约以及预测发生的时间。(网址：

http://www.argentumlux.org/documents/CRisk_final.pdf)

众包系统性风险的措施研究是使用收集想法的竞赛和聚合方法来决定更好的系统性风险的措施。

人工迟钝和投资者心理研究是它模拟常见的投资者行为，即消费投资行为是什么模式和为什么消费者使用该消费行为方式，并确定基于这些行为改进的产品。（网址：https://secure.brightworkinc.net/~andrewlo/documents/Learning_Connections_in_Financial_Time_Series.pdf）

法律代码研究是以使用美国合法的代码为例，法律和软件代码之间的结构相似、影响法律的复杂性、代码协议应用到金融监管的可能性(例如多德-弗兰克辅助紧密耦合)。这一发现可能有助于更有效地法律起草，并且比当前的实践更有效且有更少的破坏性。

总结：

该会议讨论且探索了听众感兴趣以及大数据面临的挑战，主要主题包括：

- 财务数据在收集、验证、保护和使用中面临的挑战；
- 使用大数据改变组织文化来提高公司的生产力、个人和集体消费行为；
- 使用大数据来预测消费者信贷违约；
- 预测长期资本市场、评估市场不同阶段的走势；
- 监管机构使用收集来的数据对相关机构进行压力测试。

麻省理工学院教师表达了他们有兴趣与企业继续谈话，以便更好地理解财经大数据面临的挑战和机遇。建议的进一步讨论主题包括创建数据协作和分析的应用、探索研究人员存储、管理和共享数据的新方法。

这个会议开启了麻省理工学院、行业和政策制定者之间的对话，旨在探索创建一个伙伴关系关注共享数据、协同工作和支持研究的伙伴关系，研究有关解决大数据与金融问题以及最终努力创建和维护一个更好的、稳定、高效的金融体系。

编译自：

http://bigdata.csail.mit.edu/

在线跟踪和隐私：埃尔文德·纳拉亚南演讲摘要

郝晓雪 编译 王凯艳 校对

在线隐私跟踪反向工程，透明制和问责制

——埃尔文德·纳拉亚南（普林斯顿大学副教授）演讲摘要

当我们浏览网页时，我们产生的数据将会被跟踪、收集、创造性使用。但是由于透明度不够使得网络跟踪存在很强的争议性。在麻省理工大学最近的一个大数据演讲中，纳拉亚南阐述了他研究的几个方面，并且强调了正在进行的普林斯顿网络透明制和问责制项目（WEBTAP）。WEBTAP 的研究目标是纠正网络隐私的市场失灵，加快对网络隐私、安全和道德问题的管制，以实现一次大范围的公共辩论。

“当谈及大数据时，”他说，“其实是在谈论大数据系统——在线跟踪系统（此系统不由用户建立，不受用户控制）和关于输入输出设计方法的研究。”他还说，这是一个新兴领域，发展迅速，需要研究的问题比这方面的专家还多，所以，现在是研究大数据的最好时机。

第三方在线跟踪

当用户访问网页时，如纽约时报（NYT）主页，网站主要内容属于 NYT，但是图片可能来自其他媒体，第三方广告由多方提供。谁可以查看用户浏览记录编译文件的权限并不透明。纳拉亚南说：研究表明，排名前 50 的网站有几十个跟踪机制。

在线跟踪的目的是多样的，从广告服务到为用户定制新闻。其可能产生的危害也是多样的。拿广告来说，网站可能会根据用户信息改变产品价格，我们不清楚这会不会导致不公平的差别定价。在新闻网站中，过度调整的利益是否会导致过滤气泡，基于其之前的浏览记录，个体只会接收到越来越狭小的信息流。更极端点说，这可能会使人们的世界观变小——““自我强化气泡”——基于个体现有的信念和偏好来提供持续积极的反馈和定制的内容。

“不仅用户没有注意到这点” 纳拉亚南说，“就是网站运营商自身也没有意识到第三方跟踪系统产生的后果。”“圆形监狱”的观念可以作为一个不祥的类比：人们只知道自己被监视，但是他们不知道具体的监视细则或目的，这种权利是不对称的。同样，对于在线跟踪来说，人们不知道自己的哪些信息被收集、被使用了。

为了了解这些网站的运行机制，纳拉亚南和他的团队开始研究他们发现的环境和编译数据。“我们的中心论点是单个模块化平台可以进行隐私冲击反向工程的各种试验。”在试验中，自动化的模拟用户浏览网站，我们监测并分析用户数据流，测试网站如何为这些机器人进行

信息定制。

纳拉亚南说:“虽然隐私问题很严重，但进行隐私保护的一个简单方法就是测量它。它可以解决基本的信息不对称问题，产生更多的公共辩论，促进现有法律的规范和执行。测量并公布调查结果将大大促进这些问题的解决。”

帆布指纹和其他的隐藏识别机制

关于浏览器指纹的一个新技术叫“帆布指纹”，这样第三方不使用浏览器 cookie 就可以唯一识别和跟踪用户。在对 10 万个网站的研究中，纳拉亚南团队和在 KU 的另一团队发现超过 5,500 的网站使用了帆布指纹技术。“这项研究结果很重要，”纳拉亚南说，“研究结果公布之后，网站做法遭到了强烈攻击，一些企业也决定停止这样做。”

高级在线跟踪机制的第二个研究领域是 cookie 的再生和同步，这些都是针对第三方试图在用户设备中安装永久 cookies 的做法。纳拉亚南的项目还包括对 ID cookie 探测以及联合登录影响的研究。

在线跟踪不仅能提供有针对性的广告，也可以通过用户的支付意愿改变产品和服务的价格（差别定价）。研究发现虽然少数网站确实存在差别定价问题，但并不会大规模发生；通过对在线机票价格的研究，结果表明并没有明确证据可以证明系统差别定价的存在。

挑战及未来任务

纳拉亚南及其团队面临的第一个挑战是网站自身特性——他们发现网站抵制自动化，这就导致了试验的频繁失败。团队已经在其系统上建立了几个抽象层，以实现良好的错误恢复率和并行。第二个问题是统计严格保护，因为被评估网站可能创建了个人设备及其浏览器的配置文件，因此，在进行某些交互测试时要使用不同类型的设备。另外，网站也可能进行自身的用户调查，这会使研究结果产生偏差，造成双向测试场景。

理想状态是，团队能模拟现实世界的交互，但是目前还没有一种完全自动的方法来完成这项试验。纳拉亚南说:“这是一个寻求理解的地方，因为大数据技术的存在——有 10 亿美元的投资用于第一方和第三方跟踪系统基础结构建设。”公司推销个性化产品，对产品使用后你会发现他们并没有真正实现个性定制。

未来将更多的侧重于机器学习使用的研究，以进一步深化对帆布指纹和其他现象的研究。对测量驱动隐私工具的研究也在增多，这种隐私工具可以实现对行为而非行为者的拦截。同时，项目也开始使大家达成对网站隐私的共识，并探寻第一方问责制方法。

纳拉亚南总结说：“应该举行一次公共辩论，以就用户数据使用问题达成共识。同时，随着系统的飞速发展，这将是技术人员加入并创建独立监督制和透明制的最佳时机。如果能实现这些，这些系统将会朝社会可接受的方向发展，同时也能实现系统建造者的商业愿景。”

编译自：

http://bigdata.csail.mit.edu/node/217

为何深度分析变得如此重要

郝晓雪 编译 王凯艳 校对

11 月，在奥斯汀举行的第四届戴尔全球大会上，数据分析成为大会重点。无论是迈克尔·戴尔的主题报告还是关于数据经济的承诺或是对大数据及数据分析的讨论，都证明了将数据变成洞察力的重要性。

“大数据的一个问题是大量噪音的存在，所以我们需要一定的工具来过滤这些噪音。”戴尔著名工程师马克·戴维斯说。

这并非偶然，每年的戴尔全球大会上，我们关注的内容都与客户紧密相关，企业在数据经济时代的生存发展能力将不可避免的与他们通过深度分析获得商业优势的能力联系在一起。虽然深度分析平台本身并不是新的，但是现代技术可以帮助我们连接并整合数据，使数据分析变得更为容易。结合分析平台自身的发展，这些新功能构成了深度分析利用的典型案例。

深度预测分析，不再仅仅是降低成本的机制，而成为商业盈利的主要工具。这就意味着，深度分析能力正成为企业占领中端市场的必备能力，并且能够为每个供应商提供可信的点对点数据管理方案。这就是戴尔年初就采用史丹索特的原因。史丹索特数据统计分析软件的深度分析平台成为现代信息管理的重要组成部分。

利用史丹索特数据统计分析软件，我们可以与用户进行更多的讨论。以下是深度分析帮助发现数据价值，并将其转化为收益的五种途径：

1. 提高对客户的理解。当你了解并预测到了客户的行为模式及购买模式时，你将更容易在需要的时间、需要的地点为他们提供需要的产品和服务。收益就是这样产生的。这也是客户支持的重要体现。深入分析能帮助企业更好地进行客户需求预测，并主动提供客户支持，客户支持进而产生收益。
2. 加快决策速度。时间就是金钱，像史丹索特数据统计分析软件这样的平台可以将分析直

接嵌入到运营决策过程中。

3. 提供广泛的过程改进。通过将分析嵌入到日常工作活动中，可以提高团队的工作效率，实现过程改进。
4. 保持与在线竞争对手和零售商的竞争力。与传统企业相比，在线企业对客户需求拥有更为敏捷和快速的反应，因此在线企业拥有固有优势。如果你的企业属于传统的大企业，那么就需要通过深入分析进行更为深入的客户观察，提高客户支持，加快决策制定，优化业务流程，这对保持与小企业和在线企业在速度和敏捷性上的竞争力有着重要作用。
5. 开发垂直-具体模型。企业数据分析需要正变得更专业、更具体、更垂直化。垂直领域（如制造业、医疗保健业、制药业、银行业）的客户及市场需求分析变得日益具体化。现代深度分析平台能帮助企业获得垂直-具体的洞察力，这对降低企业成本、保持健康的收入来源具有重要作用。史丹索特数据统计分析软件，就是在这一方面取得竞争优势的。

当然，深度分析对客户的价值远非我们所列的这几点。相信，不就将来，深度分析将为企业创造出更大的价值。因此，这只是个开始，我们将持续与客户保持紧密联系。未来几个月，我们还将加大对深度分析能力的投入，同时我们也希望客户能这样做。

编译自：

http://en.community.dell.com/dell-blogs/direct2dell/b/direct2dell/archive/2014/12/03/why-advanced-analytics-are-more-valuable-than-ever

大数据量需要更少的车道

王凯艳 编译　　张春玲 校对

一个组织能否在不增加数据管理环境的复杂性基础上而承担大数据的挑战？有一种方法。

正如我已经不止一次地注意到，在过去的几个月中，传统数据库的固有局限性已成为各种组织充分利用他们的数据资产的阻碍。这成为试图不断加快业务步伐的企业的明显问题。大数据带来的挑战将使这些限制成为焦点。不断增长的数据量，行动时间的缩短，并在加入多种类型的非结构化数据等都以不同的方式暴露了传统的数据库。该大型厂商推出了解决这些问题的修复以使数据管理环境更加的错落有致。

一个有趣的比喻是“修复”这些厂商（在所有被发现的地方）在努力减少交通拥堵。当在交通繁忙的道路上驾驶时，我们从不动的车道出来变换到能动的车道。意图减少交通拥堵

的建筑师往往兴建更多道路，或者扩大原有道路。

不幸的是，这些“最优策略”被证明是错的。研究表明，变道会使交通变慢。这就是在时间就是金钱的今天，为什么英国会贴出不要变道的标示的原因。此外，研究表明，虽然违反直觉，但是仍然是这样的：超过一定点，增加道路拓宽车道无助于减少交通拥堵。

IT 人员用数据做同样的事情。就像增加更多的车道，使交通拥堵更糟，在您的联机事务处理系统，电信运营支撑系统，数据仓库，数据集市，高速缓存，立方体，柱状数据库，无共享架构等等添加更多的数据管理解决方案是不会改善您的运行数据量和运行速度的。事实上，可能会使他更糟糕。另外，这些数据解决方案都需要您将大量数据“变道”，这种做法仍旧会使数据运行速度缓慢。

当然，近期的 Hadoop 和 NoSQL 能够数据管理环境提供巨大的可能。设有 CIO 职位并鼓励其找寻能够容纳更多数据的管理解决方案的公司不告诉任何新的补充。结果已成定局：更多的膨胀，更混乱，更多移动的部件。允许大数据环境的巨大复杂性的解决方案添加了更多的复杂性。但是，还有另外一种方法。在 SAP，我们认为正确的答案，而不是只针对大数据的所有企业计算，始于一个核心原则：简化。通过反思旧观念和发明新的方法少开始。

我最近描述了在 SAP HANA 在这些方面的发展思路：让我们简化问题。让我们将整个数据库嵌入到内存中，切割磁盘和所有磁盘访问问题和延误出来的图片。让我们来运行整个企业 —所有的应用程序，所有的交易，所有的分析，简单的一点，以令人眼花缭乱的快速始终保持最新的数据库。当我们重新想象数据管理，我们知道，对于企业来说，增加复杂性永远无法得到答案。SAP HANA 提供了一种数据交互和自身处理的创新，这种创新能够管理数据量。

这就是为什么 SAP HANA 删除在应用、事务处理和分析之间的所有障碍：将所有资源放在在一个单一的平台上。这就是为什么 HANA 从企业环境中删除不需要的数据的副本，实现一切只有一个副本就可以。那就是 HANA 无需在数据库层和应用层之间移动数据的理由；数据在哪里，处理就在哪里。这就是为什么 HANA 消除交易和分析之间的所有延迟：没有更多的批处理作业，没有更多的 ETL 等。最重要的是，这就是为什么 HANA 摒弃了需要使用数据（企业用户）的人和管理数据的人之间的隔膜（数据架构师。）

HANA 为企业数据环境的根本的彻底的简化提供创新需要。因此，它是真正的大数据解决方案的完美起点。当所有的杂波不见了，一个企业才能真正开始认识到技术，如 Hadoop

的或 NoSQL 的价值。能够管理任何结构的数据、任何规模的数据对象，以及管理无限期大范围的数据量的能力—— 这是一个完美的补充 HANA 革命性的实时方法。

现在面临的大数据时代的挑战背景下的公司可以找出其特定要求的最佳解决方案。对于核心数据，有 SAP HANA，它支持交易和关键任务的分析。对于关联数据 - 其中包括社会渠道的数据，遥测传感器和其他设备的数据，外部数据源的数据，以及任何和决策者想获取的决策相关数据——有 Hadoop 和 NoSQL。

这些技术结合起来，提供一个完整的解决方案，这能够更加积极有效的满足大数据需求。这是第一次，他们带来的这些结果不是通过使环境更加困难和复杂，而是通过使其更简单，更从容和更有准备。

编译自：

https://blogs.saphana.com/2014/09/24/big-data-volumes-need-fewer-lanes/

穿越数据湖泊：数据中心

王凯艳 编译　　张春玲 校对

组织争相利用大数据进行业务分析。他们已经开发了一些创造性的新数据管理理念。数据湖泊就是其中之一。

廉价商品服务器、集群、云、Hadoop 和分布式数据存储方法的基础就是数据湖泊。在创建一个数据湖过程中，你需要把所有的数据结构化、半结构化、非结构化到中央池。数据池中是保留着源格式的数据。所有用户都可以访问数据池，并且每个都可以根据具体需求分析计算出如何使用它。我们的想法是，没有最典型的管理开销将数据存储在关系数据库和其他传统系统的话，你可以保存所有的数据以后使用。

数据湖泊的挑战

数据湖泊具有吸引力也有挑战。

数据湖泊的终极目标是数据即时，并能无处不在地自助分析。但在实现这一目标上，只是池数据并不能够解决这个困难的问题。有几个问题和相关的技术差距：

- 如何找到你所需要的数据集并且快速了解它们所包含的使用什么样的可视化和抽样方法？
- 如何获得允许人们去操作的数据集格式?如何将相关数据集联系并整合在一起？
- 随着用户的增加、时间的推移，如何保持数据集之间的关系并跟踪他们的进化、合

并、扩展吗?某种形式的版本控制系统是必要的。

- 你怎么在组织内外部与他人共享数据，并且如何才能在用户不断添加或删除记录后保持数据的完整？

这些问题影响着商户、研究者和科学家等。

数据中心：一个大数据管理的分布式的版本控制系统

今天科学家们希望收集数据,分析数据集和协作得到即时的见解,提炼知识和发现创造。协作在进行用户最佳数据访问和可视化工具实验的基础上经常发生、迭代和试错。受到像 Git 和 github 上的软件版本控制系统的启发，马里兰大学和伊利诺伊大学厄巴纳 - 香槟分校跨越 MIT CSAIL，提出了一个解决方案：数据中心。该解决方案提供数据的湖泊（数据和分析解放），同时克服其挑战（糟糕的用户界面，缺乏治理，潜在的“坏”/失同步数据）的优势。该解决方案由两个紧密集成的系统：1）数据集版本控制系统（DVCS），它使用户能够创建，分支机构，合并，差和搜索的数据集大，不同的集合的能力，2)一个 DataHub 平台，使用户能够在良好的版本控制下执行建立的协同数据分析的能力。该数据集的版本控制系统提供了多版本数据集中管理。它的目标是提供一个共同的衬底，使数据科学家捕捉到他们的修改，最大限度地减少存储成本，使用声明的语言推理版本，找出版本之间的差异，并与其他科学家共享数据集。该平台是并使数据意义上的组织,管理，共享，协作，托管平台。它提供了一个高效的平台和易于使用的工具/接口：

- 发布自己的数据（托管，共享，协作）
- 使用他人数据（查询，链接）
- 检测数据（分析，可视化）

该平台提供了通过可伸缩的、平行的、基于 SQL 的分析数据处理为对大型数据集极低延迟操作优化引擎对数据的访问。这克服了前几代产品的数据库管理系统在处理大数据，如索引，主内存数据库，列式数据库的限制.该系统目前正在不同的数据集大湖被测试 - 其中包括麻省理工学院的大数据生活实验室，数据在麻省理工学院的研究人员中共享的平台，并支持一组动态分析。

跃过数据湖，而不是投入其中

DataHub 克服数据湖泊的问题，将数据解放出来，使数据科学家和研究人员可以在大规模协作数据分析中有效参与。

编译自：

http://bigdat a.csail.mit.edu/node/205

大数据有潜力改革医疗保健

张春玲 编译 郝晓雪 校对

大数据有可能适宜的改变一切。在我们的指尖越来越多的信息被用于分析和使用，从预测犯罪到加速企业发展。现在，创新型公司使用越来越多的信息会使医疗保健做的更好。

越来越多的医疗数据为预测流感疫情、防止疾病、降低医疗成本、提高整体病人护理提供线索。鉴于这些情况，毫无疑问，医疗数据备受瞩目，故此它吸引创业公司和投资者寻求由数据驱动的医疗保健。根据旧金山的种子资金 Rock Health 公司调查，今年医疗技术风险投资公司增加了 176%投资，其中大部分的资金分配给数据分析公司。

随着电子健康记录和健康保险索赔的形式的发展，该领域聚集了越来越多的信息，一些机构使用它进行个性化定制和加强护理。同样，消费者转向使用个人健康跟踪装置掌握自己的健康。但机构如何真正利用生产和聚集的信息来发挥它的潜力以改善医疗保健和捕捉商业机会?

现状

毫无疑问，最近麻省理工学院有关数据驱动的医疗保健技术的评论报告称，最大的挑战是打破信息孤岛以聚达到集聚所有的数据。目前，医学数据被禁锢于多方来源，这种情况令企业无法获得正在发生什么事情的发展的、全面的信息。大多数的由数据驱动的医疗保健应提供一个便利的解决方式：仅从一个来源而不是所有来源获取数据。

所有的数据来自哪里？根据麻省理工学院技术评论报，今天的医疗保健存在于多个数据包，这些数据包括来自传感器、公共卫生记录、电子医疗记录、保险索赔、基因组数据、家庭健康史和移动健康应用的数据。通过分析这些层的数据才能获得最有价值的，患者、医生、研究人员和商业用户都可以理解的见解。

这些新旧数据的来源一起提供详细的信息，它可能对疾病预防和治疗产生新的见解。为保证成功，数据分析解决方案必须能够分析所有这些来源——甚至是新的数据来源。

增长的个人健康追踪设备是一种新兴数据源的来源。目前，用于测量卫生指标的移动健康应用程序已经超过 100000 台，它能对病人健康提供持续关注。Manhattan Research Cybercitizen Health 研究估计，目前有 9500 万美国人使用至少一个移动医疗技术，并且他们

生成的大多数数据存储在移动设备上。根据相同的报告显示，未来五年内将超过 170 万智能手机用户会下载一个移动健康应用。伴随着所有这些增长的信息来源，最有效的数据分析技术必须能够整合每一个来源的数据。

治疗

解决呈分散生态系统的医疗数据的有效方法是创新技术，该技术可以将所有可用的无论其体积、速度或品种的病人数据合并起来。对多来源的数据而言，一个灵活的数据分析解决方案是至关重要的。

所以当寻找解决所有由数据驱动的医疗保健带来的弊端，找到能连接每一个单一的数据源并对结构化、非结构化和实时数据进行分析的解决方案是很重要的。当采用能完成所有这三个任务的技术时，机构可确保包含所有的重要数据以及包括临床叙述、医生笔记、床边监测仪获得的实时数据这些最新数据。若没有能力包含这些信息，见解只是没有价值的。

更重要的是分析所有的数据后，必须对可行的见解进行描述，该描述对病人和医生而言是可视的和可理解的。毕竟，如果这些数据不能由使用他们的人掌控，又怎么会是好的医疗见解？

由数据驱动医疗保健公司的最大的挑战是分析所有这些有价值的见解，而当代的技术正在展开。使用正确的技术配方，机构可以找到治疗改善医疗保健的方法。

编译自：

http://www.datameer.com/ceoblog/big-data-potential-transform-health-care/

（该板块部分文献摘自 2014 年召开的大数据年会，更详细内容见 http://bigdata.csail.mit.edu/）

【SSCI 高被引文献推介】

1、Title: Big data: The future of biocuration

Author(s): Howe, D; Costanzo, M; Fey, P; Gojobori, T; Hannick, L; Hide, W; Hill, DP; Kania, R; Schaeffer, M; St Pierre, S; Twigger, S; White, O; Rhee, SY

Abstract: The exponential growth in the amount of biological data means that revolutionary measures are needed for data management, analysis and accessibility. Online databases have become important avenues for publishing biological data. Biocuration, the activity of organizing, representing and making biological information accessible to both humans and computers, has become an essential part of biological discovery and biomedical research. But curation increasingly lags behind data generation in funding, develop ment and recognition.

Full Text：01

2、Title： Big data and the future of ecology

Author(s): Hampton, SE; Strasser, CA; Tewksbury, JJ; Gram, WK; Budden, AE; Batcheller, AL; Duke, CS; Porter, JH

Abstract: The need for sound ecological science has escalated alongside the rise of the information age and "big data" across all sectors of society. Big data generally refer to massive volumes of data not readily handled by the usual data tools and practices and present unprecedented opportunities for advancing science and informing resource management through data-intensive approaches. The era of big data need not be propelled only by "big science" - the term used to describe large-scale efforts that have had mixed success in the individual-driven culture of ecology. Collectively, ecologists already have big data to bolster the scientific effort - a large volume of distributed, high-value information - but many simply fail to contribute. We encourage ecologists to join the larger scientific community in global initiatives to address major scientific and societal problems by bringing their distributed data to the table and harnessing its collective power. The scientists who contribute such information will be at the forefront of socially relevant science - but will they be ecologists? Front Ecol Environ 2013;11(3):156-162, doi: 10.1890/120103 (published online 12 Mar 2013)

Full Text：02

3、Title： The Parable of Google Flu: Traps in Big Data Analysis

Author(s): Lazer, D; Kennedy, R; King, G; Vespignani, A

Abstract: In February 2013, Google Flu Trends (GFT) made headlines but not for a reason that Google executives or the creators of the flu tracking system would have hoped. Nature reported that GFT was predicting more than double the proportion of doctor visits for influenza-like illness (ILI) than the Centers for Disease Control and Prevention (CDC), which bases its estimates on surveillance reports from laboratories across the United States (1, 2). This happened despite the fact that GFT was built to predict CDC reports. Given that GFT is often held up as an exemplary use of big data (3, 4), what lessons can we draw from this error?

Full Text：03

4、Title： Large-scale electrophysiology: Acquisition, compression, encryption, and storage of big data

Author(s): Brinkmann, BH; Bower, MR; Stengel, KA; Worrell, GA; Stead, M

Abstract: The use of large-scale electrophysiology to obtain high spatiotemporal resolution brain recordings (>100 channels) capable of probing the range of neural activity from local field potential oscillations to single-neuron action potentials presents new challenges for data acquisition, storage, and analysis. Our group is currently performing continuous, long-term electrophysiological recordings in human subjects undergoing evaluation for epilepsy surgery using hybrid intracranial electrodes composed of up to 320 micro- and clinical macroelectrode arrays. DC-capable amplifiers, sampling at 32 kHz per channel with 18-bits of A/D resolution are capable of resolving extracellular voltages spanning single-neuron action potentials, high frequency oscillations, and high amplitude ultra-slow activity, but this approach generates 3 terabytes of data per day (at 4 bytes per sample) using current data formats. Data compression can provide several practical benefits, but only if data can be compressed and appended to files in real-time in a format that allows random access to data segments of varying size. Here we describe a state-of-the-art, scalable, electrophysiology platform designed for acquisition, compression, encryption, and storage of large-scale data. Data are stored in a file format that incorporates lossless data compression using range-encoded differences, a 32-bit cyclically redundant checksum to ensure data integrity, and 128-bit encryption for protection of patient information. (C) 2009 Elsevier B.V. All rights reserved,

Full Text：04

5、Title: Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate

Author(s): Ansolabehere, S; Hersh, E

Abstract: Social scientists rely on surveys to explain political behavior. From consistent overreporting of voter turnout, it is evident that responses on survey items may be unreliable and lead scholars to incorrectly estimate the correlates of participation. Leveraging developments in technology and improvements in public records, we conduct the first-ever fifty-state vote validation. We parse overreporting due to response bias from overreporting due to inaccurate respondents. We find that nonvoters who are politically engaged and equipped with politically relevant resources consistently misreport that they voted. This finding cannot be explained by faulty registration records, which we measure with new indicators of election administration quality. Respondents are found to misreport only on survey items associated with socially desirable outcomes, which we find by validating items beyond voting, like race and party. We show that studies of representation and participation based on survey reports dramatically misestimate the differences between voters and nonvoters.

Full Text：05

6、Title: Bioinformatics clouds for big data manipulation

Author(s): Dai, L; Gao, X; Guo, Y; Xiao, JF; Zhang, Z

Abstract: As advances in life sciences and information technology bring profound influences on bioinformatics due to its interdisciplinary nature, bioinformatics is experiencing a new leap-forward from in-house computing infrastructure into utility-supplied cloud computing delivered over the Internet, in order to handle the vast quantities of biological data generated by high-throughput experimental technologies. Albeit relatively new, cloud computing promises to

address big data storage and analysis issues in the bioinformatics field. Here we review extant cloud-based services in bioinformatics, classify them into Data as a Service (DaaS), Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS), and present our perspectives on the adoption of cloud computing in bioinformatics.
Full Text：06

7、Title: Functional Interactions as Big Data in the Human Brain
Author(s): Turk-Browne, NB
Abstract: Noninvasive studies of human brain function hold great potential to unlock mysteries of the human mind. The complexity of data generated by such studies, however, has prompted various simplifying assumptions during analysis. Although this has enabled considerable progress, our current understanding is partly contingent upon these assumptions. An emerging approach embraces the complexity, accounting for the fact that neural representations are widely distributed, neural processes involve interactions between regions, interactions vary by cognitive state, and the space of interactions is massive. Because what you see depends on how you look, such unbiased approaches provide the greatest flexibility for discovery.
Full Text: 07

8、Title: Towards big data science in the decade ahead from ten years of InCoB and the 1st ISCB-Asia Joint Conference INTRODUCTION
Author(s): Ranganathan, S; Schonbach, C; Kelso, J; Rost, B; Nathan, S; Tan, TW
Abstract: The 2011 International Conference on Bioinformatics (InCoB) conference, which is the annual scientific conference of the Asia-Pacific Bioinformatics Network (APBioNet), is hosted by Kuala Lumpur, Malaysia, is co-organized with the first ISCB-Asia conference of the International Society for Computational Biology (ISCB). InCoB and the sequencing of the human genome are both celebrating their tenth anniversaries and InCoB's goalposts for the next decade, implementing standards in bioinformatics and globally distributed computational networks, will be discussed and adopted at this conference. Of the 49 manuscripts (selected from 104 submissions) accepted to BMC Genomics and BMC Bioinformatics conference supplements, 24 are featured in this issue, covering software tools, genome/proteome analysis, systems biology (networks, pathways, bioimaging) and drug discovery and design.
Full Text:0 8

9、Title: Trends in big data analytics
Author(s): Kambatla, K; Kollias, G; Kumar, V; Grama, A
Abstract: One of the major applications of future generation parallel and distributed systems is in big-data analytics. Data repositories for such applications currently exceed exabytes and are rapidly increasing in size. Beyond their sheer magnitude, these datasets and associated applications' considerations pose significant challenges for method and software development. Datasets are often distributed and their size and privacy considerations warrant distributed techniques. Data often resides on platforms with widely varying computational and network capabilities. Considerations of fault-tolerance, security, and access control are critical in many applications (Dean and Ghemawat, 2004; Apache hadoop). Analysis tasks often have hard deadlines, and data quality is a major concern in yet other applications. For most emerging

applications, data-driven models and methods, capable of operating at scale, are as-yet unknown. Even when known methods can be scaled, validation of results is a major issue. Characteristics of hardware platforms and the software stack fundamentally impact data analytics. In this article, we provide an overview of the state-of-the-art and focus on emerging trends to highlight the hardware, software, and application landscape of big-data analytics. (C) 2014 Elsevier Inc. All rights reserved.

Full Text:0 9

10、Title: Early Prediction of Movie Box Office Success Based on Wikipedia Activity Big Data

Author(s): Mestyan, M; Yasseri, T; Kertesz, J

Abstract: Use of socially generated "big data" to access information about collective states of the minds in human societies has become a new paradigm in the emerging field of computational social science. A natural application of this would be the prediction of the society's reaction to a new product in the sense of popularity and adoption rate. However, bridging the gap between "real time monitoring" and "early predicting" remains a big challenge. Here we report on an endeavor to build a minimalistic predictive model for the financial success of movies based on collective activity data of online users. We show that the popularity of a movie can be predicted much before its release by measuring and analyzing the activity level of editors and viewers of the corresponding entry to the movie in Wikipedia, the well-known online encyclopedia.

Full Text: 10

11、Title: Content Analysis in an Era of Big Data: A Hybrid Approach to Computational and Manual Methods

Author(s): Lewis, SC; Zamith, R; Hermida, A

Abstract: Massive datasets of communication are challenging traditional, human-driven approaches to content analysis. Computational methods present enticing solutions to these problems but in many cases are insufficient on their own. We argue that an approach blending computational and manual methods throughout the content analysis process may yield more fruitful results, and draw on a case study of news sourcing on Twitter to illustrate this hybrid approach in action. Careful combinations of computational and manual techniques can preserve the strengths of traditional content analysis, with its systematic rigor and contextual sensitivity, while also maximizing the large-scale capacity of Big Data and the algorithmic accuracy of computational methods.

Full Text: 11

12、Title: Big Data: New Opportunities and New Challenges

Author(s): Michael, K; Miller, KW

Abstract: We can live with many of the uncertainties of big data for now, with the hope that its benefits will outweigh its harms, but we shouldn't blind ourselves to the possible irreversibility of changes-whether good or bad-to society.

Full Text: 12

13、Title: Optical storage arrays: a perspective for future big data storage

Author(s): Gu, M; Li, XP; Cao, YY
Abstract: The advance of nanophotonics has provided a variety of avenues for light-matter interaction at the nanometer scale through the enriched mechanisms for physical and chemical reactions induced by nanometer-confined optical probes in nanocomposite materials. These emerging nanophotonic devices and materials have enabled researchers to develop disruptive methods of tremendously increasing the storage capacity of current optical memory. In this paper, we present a review of the recent advancements in nanophotonics-enabled optical storage techniques. Particularly, we offer our perspective of using them as optical storage arrays for next-generation exabyte data centers.
Full Text: 13

14、Title: Big Data: A Survey
Author(s): Chen, M; Mao, SW; Liu, YH
Abstract: In this paper, we review the background and state-of-the-art of big data. We first introduce the general background of big data and review related technologies, such as could computing, Internet of Things, data centers, and Hadoop. We then focus on the four phases of the value chain of big data, i.e., data generation, data acquisition, data storage, and data analysis. For each phase, we introduce the general background, discuss the technical challenges, and review the latest advances. We finally examine the several representative applications of big data, including enterprise management, Internet of Things, online social networks, medial applications, collective intelligence, and smart grid. These discussions aim to provide a comprehensive overview and big-picture to readers of this exciting area. This survey is concluded with a discussion of open problems and future directions.
Full Text: 14

15、Title: Social-Network-Sourced Big Data Analytics
Author(s): Tan, W; Blake, MB; Saleh, I; Dustdar, S
Abstract: Very large datasets, also known as big data, originate from many domains. Deriving knowledge is more difficult than ever when we must do it by intricately processing this big data. Leveraging the social network paradigm could enable a level of collaboration to help solve big data processing challenges. Here, the authors explore using personal ad hoc clouds comprising individuals in social networks to address such challenges.
Full Text: 15

16、Title: 'Big data', Hadoop and cloud computing in genomics
Author(s): O'Driscoll, A; Daugelaite, J; Sleator, RD
Abstract: Since the completion of the Human Genome project at the turn of the Century, there has been an unprecedented proliferation of genomic sequence data. A consequence of this is that the medical discoveries of the future will largely depend on our ability to process and analyse large genomic data sets, which continue to expand as the cost of sequencing decreases. Herein, we provide an overview of cloud computing and big data technologies, and discuss how such expertise can be used to deal with biology's big data sets. In particular, big data technologies such as the Apache Hadoop project, which provides distributed and parallelised data processing and analysis of petabyte (PB) scale data sets will be discussed, together with an overview of the

current usage of Hadoop within the bioinformatics community. (C) 2013 Elsevier Inc. All rights reserved.

17、Title: Data Mining with Big Data
Author(s): Wu, XD; Zhu, XQ; Wu, GQ; Ding, W
Abstract: Big Data concern large-volume, complex, growing data sets with multiple, autonomous sources. With the fast development of networking, data storage, and the data collection capacity, Big Data are now rapidly expanding in all science and engineering domains, including physical, biological and biomedical sciences. This paper presents a HACE theorem that characterizes the features of the Big Data revolution, and proposes a Big Data processing model, from the data mining perspective. This data-driven model involves demand-driven aggregation of information sources, mining and analysis, user interest modeling, and security and privacy considerations. We analyze the challenging issues in the data-driven model and also in the Big Data revolution.

18、Title: Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management
Author(s): Waller, MA; Fawcett, SE
Abstract: We illuminate the myriad of opportunities for research where supply chain management (SCM) intersects with data science, predictive analytics, and big data, collectively referred to as DPB. We show that these terms are not only becoming popular but are also relevant to supply chain research and education. Data science requires both domain knowledge and a broad set of quantitative skills, but there is a dearth of literature on the topic and many questions. We call for research on skills that are needed by SCM data scientists and discuss how such skills and domain knowledge affect the effectiveness of an SCM data scientist. Such knowledge is crucial to develop future supply chain leaders. We propose definitions of data science and predictive analytics as applied to SCM. We examine possible applications of DPB in practice and provide examples of research questions from these applications, as well as examples of research questions employing DPB that stem from management theories. Finally, we propose specific steps interested researchers can take to respond to our call for research on the intersection of SCM and DPB.

19、Title: Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb
Author(s): Crampton, JW; Graham, M; Poorthuis, A; Shelton, T; Stephens, M; Wilson, MW; Zook, M
Abstract: This article presents an overview and initial results of a geoweb analysis designed to provide the foundation for a continued discussion of the potential impacts of 'big data' for the practice of critical human geography. While Haklay's (2012) observation that social media content is generated by a small number of 'outliers' is correct, we explore alternative methods and conceptual frameworks that might allow for one to overcome the limitations of previous analyses of user-generated geographic information. Though more illustrative than explanatory, the results of our analysis suggest a cautious approach toward the use of the geoweb and big data that are as mindful of their shortcomings as their potential. More specifically, we propose five extensions to the typical practice of mapping georeferenced data that we call going 'beyond the geotag': (1) going beyond social media that is explicitly geographic; (2) going beyond spatialities of the 'here and now'; (3) going beyond the proximate; (4) going beyond the human to data produced by bots

and automated systems, and (5) going beyond the geoweb itself, by leveraging these sources against ancillary data, such as news reports and census data. We see these extensions of existing methodologies as providing the potential for overcoming existing limitations on the analysis of the geoweb. The principal case study focuses on the widely reported riots following the University of Kentucky men's basketball team's victory in the 2012 NCAA championship and its manifestation within the geoweb. Drawing upon a database of archived Twitter activity - including all geotagged tweets since December 2011-we analyze the geography of tweets that used a specific hashtag (#LexingtonPoliceScanner) in order to demonstrate the potential application of our methodological and conceptual program. By tracking the social, spatial, and temporal diffusion of this hashtag, we show how large databases of such spatially referenced internet content can be used in a more systematic way for critical social and spatial analysis.

20、Title: Predicting mobile hotel reservation adoption: Insight from a perceived value standpoint
Author(s): Wang, Hsiu-Yuan; Wang, Shwu-Huey
Abstract: With the attempt of further supporting the overwhelming demand for reservation, a few hospitality corporations have launched mobile hotel reservation (MHR) services. For the acceptance of MHR by individuals is indispensable to the successful implementation of MHR, it is critical for practitioners and academics to understand the factors influencing the adoption of MHR. This study examines the adoption of MHR from the value perspective by proposing and examining a new research model that can capture both gain and loss elements influencing individual value perceptions on behavioral intention to adopt MHR. Data from 235 usable questionnaires, collected in Taiwan, were tested against the research model using the structural equation modeling approach. The results indicated that perceived value was a predictor in explaining the customer's adoption of MHR. From the benefits point of view, perceptions of information quality and system quality were the two critical components significantly influencing perceived value of MHR. On the sacrifice side, the effects of technological effort and perceived fee on perceived value were significant. This study will be helpful to researchers in developing and testing MHR related theories, as well as to hospitality firms in understanding individual value perceptions of utilizing MHR and implementing successful MHR system to attract more customers. Theoretical and managerial implications of our results are discussed.

温馨提示：
以上文献来自来自 Web of Science 的社会科学引文索引（SSCI），其中有具有“**Full Text**”标识的是提供原文下载的，如有需求可自行下载附件。

【国内资讯】

盘点 2014:大数据现状与国人思维误区

近两年，“大数据”已成为业界和学术界舌尖上的热词，从央视的春运迁徙图到美国奥巴马政府宣布投资 2 亿美元启动“大数据研究与开发计划；从两会期间的两会大数据到预报旅游热点，“大数据”被人们推到了一个前所未有的高度。但是，在欢呼和激动了数年后，我们更需要认真思考如何利用大数据、如何正确挖掘出大数据的价值。2014 年底，记者与 Informatica 中国区的几位资深技术专家，就大数据的现状、思维、技术和发展等问题进行了深入探讨与剖析。

大数据现状：思路已有，离成功尚远

大数据真正开始做始于去年，通过两年的尝试、积累，思路已有，但离成功还很远。一些国外的大数据案例、大数据故事无非是商务智能(BI)、数据仓库(BW)的改头换面，新瓶装旧酒而已。就如数据仓库一样，建设了近 20 年才让每个企业真正承认其价值，大数据也不能期望很快就获得成功，需要一个沉淀时间。在 Informatica 技术专家看来，如果要给个期限，那这个时间至少需要 10 年。

大数据发展可以用一个波浪式的图来形容，现在还处于第一个峰顶，必须经过低谷再升起，几轮反复。这期间，大家可能会看到许多大数据真实的案例，不管是成功的还是失败的都会给我们启示。只要尝试了就不一定完全失败，就如数据仓库建设，几年前很多报告都显示 80%的项目失败，但仔细分析后发现，只是在发展过程当中没有达到预期价值而已。前人淌过的路，后边的人可以少走一些雷区。

大数据应用的必要前提：数据治理

越来越多的行业和企业开始关注数据这一企业核心资产，但对于数据如何治理，如何管控却没有合适的方法体系的产品支撑，大数据就必须以数据治理为基础，没有数据治理谈不上大数据，数据家家都有，但不治理根本用不上，而这些恰恰是 Informatica 公司的核心竞争力所在。

在纷繁杂乱的大数据面前，没有良好的数据质量，没有更加良好的数据管理策略，用于业务应用的投资将随着应用组合在企业内的增长和扩展而日渐缩水。做大数据，90%的企业走的路子都不可能实现放烟花式的很炫效果，他们首先还是要踏踏实实地解决数据整合、数据质量和主数据管理等问题。Informatica

技术专家建议道。

大数据市场：安全先行

在生活中我们常会有这样的经历，浏览新闻网页时跳出的淘宝推荐商品竟然是你想买的东西，在家里休息时会突然接到各种保险推销电话。对于这种司空见惯的信息数据泄露人们似乎习以为常。而当更加隐私、敏感的 12306 数据的泄露事件，还是让不少人感到十分后怕。 进入大数据时代后，数据将更加透明，数据信息安全的挑战变得越来越严峻。

近两年，国家政府着重强调信息安全，企业都非常关注数据安全问题。敏感的个人、财务和健康信息受到多种不同行业和政府数据隐私法规的管制，如果企业无法保持数据隐私，他们就会面临严重的财务和法律惩罚，同时还会在客户与市场信心方面蒙受可观损失。

IT168 记者了解到，2014 年，Informatica 数据安全方案因满足市场热点需求而成为业务增长较快的单元。大数据的发展还有许多亟待解决的难题，但无疑解决大家最担心的数据安全问题应当是重中之重。

大数据思维：允许数据的不精确性

以前，由于可获得的数据量比较小，为此我们必须尽量准确的记录下所获得的所有数据，做出个 KPI 供领导参考，采样过程的精确度被放在重要的地位。显然，这种对精确性的执着是信息缺乏时代的产物。大数据时代，数据的收集问题不再成为困扰，采集全量的数据成为现实，但海量数据的涌现一定会增加数据的混乱性且造成结果的不准确性，如果仍执迷精确性，那么将无法应对这个新的时代。

大数据通常都用概率说话，且大数据处理之前是可以对之进行清洗从而减少部分的错误数据。所以，与致力于避免错误相比，对错误的包容将会带给我们更多信息。其实，允许数据的混杂性和容许结果的不精确性才是我们拥抱大数据的正确态度，只要做到 10%准确结果，能够达成业务数十倍的增长即可，这是真正的大数据思维，未来我们应当习惯这种思维。

大数据思维：大数据不是单纯的技术问题

大数据不是一个纯技术问题，会包含很多管理、业务方面的内容。并不是说，购买了一套数据挖掘工具，组建了一个 Hadoop 环境，就能称为做了大数据。除

了设备、技术上的投资，企业还需要从组织结构、人员意识、管理方式、企业文化等方面都有一个转变。大数据的前期准备工作很多，这是一种思维上的全面变革。大家都是摸着石头过河，走一步想一想，然后再走一步再想，直到最后成功上岸。

在这样的一个过程当中，人们的思想还要跟随大数据技术的发展不断更新，同时也要对一些过去的想法进行纠正和改变。当然，这个时间不会像以前数据仓库那样花费 20 年，大数据可能会缩短一半时间。因为数据仓库时代是从无到有，而大数据时代是从有到更好，人们已经从建设数据仓库中积累了很多的经验、技术、教训，甚至有效的管理方法，可以很好地借鉴。

大数据思维：大数据技术解决的不仅仅是非结构化数据问题

新兴的大数据技术提供了非常有效的手段，让人们可以花很低的代价去分析、处理非结构化的数据，但是这些非结构化数据有一个特点，就是密度还很低，它远不如结构化数据有非常高的价值密度，可能 100G 的非结构化数据，最终有效的才 1G。这表明，非结构化数据是对数据完整度的很大补充，但是并不能说大数据就是做非结构化数据，其实最终的目的还是要发掘数据价值。另外一方面，传统的数据仓库已经能够完成现有结构化数据 90%的利用程度，在这种背景下，人们才会把大数据的焦点放在对非结构化的处理上。

当前，非结构化数据大量产生，如机器日志、传感器的数据、社交媒体的数据，都是以非结构化形式存在，而传统的方式对这些数据的处理能力比较欠缺。如果用木桶效应来比喻，首先要把这个短板补上，与结构化数据处理的效率和能力齐平之后，更多的就是围绕数据如何使用来进行更深一步的研究。还要认识到一点，大数据技术能够处理半结构化、非结构化的数据，不过，这些数据总是要转换成结构化的数据才能分析，算法可能输入的是非结构化的，如视频信息，但是刚进来不到 10 秒就变成结构化，最后显示出来的还是表格式结构化的结果。

链接地址：

http://www.cnii.com.cn/Bigdata/2015-01/05/content_1508624_2.htm

2015 令人心动的新兴行业大数据行业上榜

2014 的脚步已远去，2015 年的钟声接着敲响。在全球的动荡与不安中，各行各业在痛苦或幸福中度过了 2014 年，有的企业遭受了淘汰，有的还处在挣扎

中，当然，也有的企业在寒冬中脱颖而出，成为新的佼佼者。同样，2015 年会是一个新的开始，动荡和不安也许还会继续，淘汰和挣扎依然会如期上演，但更令人期待的，还有那些新黑马出现，它们注定将在未来不断闪人眼球

可穿戴设备新潮流

几年前，没人知道可穿戴设备为何物。但是在 2014 年，智能手表和运动腕带如雨后春笋一般冒出来。现在，健身追踪类可穿戴设备的技术含量越来越高，价格却更低了。

苹果已经宣布它将在并不遥远的将来发布 Watch，这似乎点燃了可穿戴设备市场的星星之火。可穿戴设备正在日新月异地向前发展，变得越来越美观。

与此同时，智能珠宝和智能服装越来越受关注。虽然很多可穿戴产品仍将专注于健身功能，但是高端时尚公司也将加入进来。

不管你如何看待它，可穿戴设备已经渐成气候。2015 年，它们将变得像智能手机一样普及和实用。

智能眼镜热潮

尽管谷歌眼镜遇到了不少的挫折，但是这并没有让其他的厂商却步。索尼、三星和东芝等很多厂商已经推出或正在开发智能眼镜类可穿戴设备，并且在不同的方向上发力。

虽然这类设备大多不太精美，但是智能眼镜已经开始让人看着越来越顺眼了。谷歌眼镜配备了有型的镜框，其中有些出自著名的时尚设计师之手。2015 年，谷歌将推出新款智能眼镜，并联手英特尔公司，在“医院网络，制造商，并将开发新的工作场所使用此设备”。

与谷歌相比，索尼的智能眼镜解决方案更为“灵活”。2015 年，索尼将推出一款智能眼镜配件，能让普通眼镜变成智能眼镜。这个装置可以附加在普通的眼镜上，然后通过配件上的显示屏阅读手机上的通知、信息甚至是你的运动信息等。

“只需要将它简单地附加到消费者的近视镜、太阳镜或者其他类型的眼睛上，就可以获得有用的信息，为生活带来极大的便利。”索尼在官方博客中表示。

日本眼镜制造商 Jins 公司推出的 Jins Meme 智能眼镜看起来跟普通眼镜一样，但包含了不少的科技技术。Meme 甚至可以监控你的警觉性，以免你在开车时睡着。

智能珠宝讨好女性

到目前为止，可穿戴设备的目标用户仍是男性消费者。大多数可穿戴设备公司只是顺带着提到女性，声称自己的智能手表“男女皆宜”。然而，这种情况将在 2015 年发生变化。

进入 2015 年之后，不仅很多专注于时尚的众筹可穿戴设备将上网销售，而且苹果的首款智能手表也将在春季发布。自从苹果 9 月宣布 Watch 以来，它已经在巴黎时尚周、时尚中国版和其他时尚杂志上亮相。

目前，Indiegogo 和 Kickstarter 平台上冒出了不少关于智能珠宝的众筹项目。很多项目比如 Ringly、Mota SmartRing 和 Cuff 已经完成了融资目标，这说明女性消费者或许对笨重的智能手表不感冒，但她们非常喜欢智能珠宝的概念。

最近，英特尔也凭借其 MICA 智能手镯进入了可穿戴市场。它的这款可穿戴设备配备了不少的功能和蛇皮材料的外皮和宝石装饰。

英特尔进入可穿戴市场之后，科技领域的一些大品牌也开始开发智能珠宝。即便他们没有这样做，类似 Ringly、Mota 和 Cuff 这样的小厂商也将推动智能珠宝市场在 2015 年进一步发展壮大。

智能服装大突破

智能手表并不是可穿戴科技的终点。科技分析公司高德纳的研究表明，智能服装将是未来几年内可穿戴技术发展的重大突破。该公司预计 2015 年智能服装的销售额会超过 1000 万美元，2016 年达到 2600 万美元。

“智能衬衫可以让更多传感器接近我们的皮肤，”高德纳公司的研究主管向《卫报》说道，“他们可以收集更多的信息，产生更好的数据，例如心跳的全波脉冲而不仅仅是脉搏。”

代号为 D-Shirt 的 Cityzen Sciences 项目可以作为完美阐述下一代可穿戴技术的例子。这款高科技衬衫附有传感器检测运动、心率、速度和呼吸模式。2015 年，还将推出衣物纤维内附有 GPS 定位的上衣和自行车短裤。

3D 打印的服装已经问世了，很多著名的时尚品牌比如拉尔夫劳伦和维多利亚的秘密已经在高调展示它们的智能服装。包括内置 LED、能够随着用户情绪变色的裙子、光纤制成的服装以及能够在下雨前检测空气湿度的衣服。

大数据前景大好

仅仅几年时间里，大数据技术就从之前的炒作阶段逐渐发展成为新数字时代中的核心技术之一。2014 年，企业内部的大数据计划慢慢地从测试阶段走向研发和生产。2015 年，大数据行业的发展重点将不再是以低成本收集和存储数据，而是转移为一旦得到这些数据如何利用的问题。

有了这些数据就能够回答那些高层提出的一个简单的问题：“为什么销售达不到上个月的预期?” 他们可以通过图表、图形、详细情况及其理由等数据来回复这个问题。

MapR 联合创始人兼首席执行官约翰·施罗德(John Schroeder)认为，2015 年企业和相关的组织机构将进行大数据的部署工作，并推进到实际的应用中。“这个行业里的领导者已经将新的大数据平台同他们的"运行"数据分析进行了整合，以便对其业务产生一定的影响”。

数据更为灵活

随着传统数据库(database)和数据仓库(datawarehouse)的运行越来越缓慢，并很难满足企业业务的发展需要，数据的灵活性就成为了推动大数据技术发展的一个重要推动力。

约翰·施罗德指出，2015 年，随着企业逐渐从简单地收集和管理数据过渡到真正使用这些数据，数据灵活性将越来越重要。

所谓的“分析”常常是一个模糊的范畴，但这需要有所改变。为了成为市场关注的焦点，经历其他领域所期待的增长水平，一些公司正成为这一领域潜在的领导者。

在金融科技领域，Money.Net 是一家致力于重建传统分析终端市场的公司，它通过利用一个从零建起的新平台达成目标。Reorg Research 研究公司利用内容与技术的结合挑战固定收益分析领域的现状。Madrone 凭借独特的方法和对冲基金数据库，试图重新定义性能分析。MarketProphit 则是第一批定义并标记出社交媒体分析的平台之一。

入侵传统产业

不过，大数据不会仅仅止步于硅谷或科技公司。Emergence Capital 的凯文·斯潘(Kevin Spain)分析指出：“新型传感器和无人飞机的发展会使农业及土地测量等产业的数据搜集方法实现革命性的改进。”

而 Bessemer 公司的布莱恩·范斯坦(Brian Feinstein)和肯特·班尼特(Kent Bennett)则指出将有更多数据搜集软件被开发出来，所有产业都会受影响，包括石油和天然气等传统产业。

在 2014 年年末的一场风险投资人聚会上，种子阶段投资公司精简创投(Streamlined Ventures)的创始人乌拉斯·奈克(Ullas Naik)就表示，自己投资了一家名为 Tachyus 的初创企业，这家公司利用传感器和其他技术来帮助石油和天然气生产商提高产量。

而另一位投资人红点创投(Redpoint Ventures)合伙人斯科特·拉尼(Scott Raney)则表示对软件即服务(Software As A Service，简称 SAAS)和数据驱动型软件的结合感兴趣。他认为，SAAS 公司意识到了他们正坐在能够加以利用的数据宝库上。

自助服务成为主流

约翰·施罗德指出，随着大数据工具和服务的发展，2015 年，IT 行业将逐渐缓解发展瓶颈的局面，许多商业用户和数据科学家将会借助相关工具和服务访问大量数据。

在此之前，IT 行业要求建立一种集中的数据结构，但是这非常消耗时间和成本一些先进的组织机构将会通过数据绑定的运行模式而非集中的结构来满足持续的需求。这种自助服务模式将促进企业更好地利用新的数据资源，同时又能够抓住新的市场机遇，应对问题和挑战。

在过去几年里，文本分析已经变得越来越复杂，这一趋势还将会继续发展下去。计算机将能够更为熟练地“阅读”一篇文章，并能够理解文章的主题和情感。这意味着这些文章能够像结构型数据那样被分类和分析。

2015 年，自助服务大数据将成为 IT 行业的一种趋势，它允许商业用户可以通过自助服务接触大数据。自助服务还可以帮助开发者、数据科学家和数据分析师直接进行数据探索和处理工作。

金融科技潜力无限

2014 年是金融科技行业快速发展的一年。所谓“金融科技”(fintech)是指“金融”(finance)和“科技”(technology)相结合的行业。金融科技公司，在简化支付过程、降低诈骗损失、提升理财规划等方面都有所作为，并悄悄地推动着传统金

融行业跨出自我革新的一步。

2014 年 10 月，在互联网创新最为活跃的美国，金融科技行业的筹资金额创下了逾 10 亿美元的历史纪录。包括谷歌、苹果在内的科技巨无霸企业，凭借着自身庞大的机构和大量的现金流，正在广泛地进入这一令人激动和快速成长的领域。

2014 年，金融科技终于走出了 2008 年信贷危机所带来的阴影。这样的增长势头看起来似乎潜力无限。展望 2015 年，又有哪些因素会继续推动该领域的发展?

数字财富管理崛起

如所有行业一般，金融科技也在经历自身的技术成熟度曲线。美国金融数据服务公司 Xignite 创始人斯蒂芬·杜布瓦(Stephane Dubois)认为，无论是 P2P 借贷、数字钱包、加密货币或是众筹，都已经度过了期望膨胀期，一切不切实际的幻想正在走向幻灭。而数字财富管理作为一种金融科技门类，当下仍然处于上升期。

事实上，受公共云规模扩大，应用开发日趋成熟，以及应用程序接口(API)日益普及的影响，当前金融科技创新繁荣发展早在多年前就开始了。诸如 Wealthfront 和 Betterment 这样的大型资产管理公司就已经开始找寻快速的集资组合服务与技术。

Wealthfront 是一家主要面向科技领域新财富拥有者的理财咨询管理平台，创立于 2011 年。Wealthfront 的最大特点是“量身定制”以及“透明”。

Wealthfront 会评估用户能接受的风险水平，然后根据评估结果为用户量身定制一个投资计划，并将一些经过精心挑选的交易所交易基金(ETFs)推荐给用户。同时平台会随时监控投资动态，并定期更新投资组合计划，以维持用户所需要的风险水平。在任何时候，用户都可以清楚地查看、跟踪自己投资的最新动态。

与 Wealthfront 类似的竞争者还包括 Personal Capital、Betterman 等在线理财平台。斯蒂芬·杜布瓦认为，随着越来越多的科技创业公司上市或被收购，在线理财领域的需求会越来越大。目前已有 Facebook、Skype、LinkedIn 等公司的众多早期员工准备使用这个平台。而在未来，很有可能出现中国的 Wealthfront，德国的 Personal Capital，还有澳大利亚的 Betterment。

B2B 模式转型

在斯蒂芬·杜布瓦看来，金融科技领域的第二个趋势是商业模式转型，从零售(B2C)模式转向机构(B2B)模式。“投资型初创公司的第一波浪潮集中于传统金融的非中介化也叫金融脱媒，这一点也不奇怪。这也是红极一时的硅谷风投向这些公司进行投资的目的：市场越大影响越大。”斯蒂芬·杜布瓦说。

但是，紧跟而来的不是企业寻求非中介化，而是帮助传统金融服务适应当今的变化的新潮流。例如，Algomi 利用社会观念来简化债券交易，Tradier 提供全方位的经纪服务，例如应用程序界面(APIs)

Algomi 创立于 2012 年，从诞生之初其目标就是为银行提供服务。近年来，金融行业面临越来越严格的监管。而 Algomi 则能够帮助银行在严格的监管条件下，满足资产负债表的要求，优化固定收益部门的收益。如今，Algomi 已经有了 9 家银行客户，并计划在今年对买方投资机构开放。

“2015 年以后，我们期望看到许多以机构为侧重点的高科技公司在这一领域能取得进展并成功引发公众注意。”斯蒂芬·杜布瓦说。

支付平台创新

2014 年，金融科技领域涌现了大量电子支付平台，这些初创公司致力于简化资金流动的手续，为用户提供更快速便捷的服务。

Zipmark 公司创立于 2010 年，总部位于纽约。Zipmark 公司致力于为用户提供“省钱的电子支票”。创始人杰伊·巴塔查里亚(Jay Bhattacharya)说：“我们的目标就是要告诉小企业，"你们不再受限于 PayPal 或信用卡，也不用支付这些费用"。”

Zipmark 的运作模式如下：企业可以告诉客户在它的网站上选择 Zipmark 支付方式，或在发票上扫描他们的二维码，这个二维码可以链接到 Zipmark 移动应用的下载地址，待下载完成后，客户可以在 App 上提交他们的支付交易。

在国内，一些传统金融行业也开始在金融科技领域发力，其中就包括为移动互联网用户搭建更为便捷的支付平台，包括上海证券“闪电通”在内的移动端 APP 被广泛使用。

“我们身处的行业正在发生变化。”杰伊·巴塔查里亚说道，“以前创业者的心态是，只要你有一个黑底色，就能变出各种颜色。而现在，每位客户的需求都不一样，需要因地制宜，为他们提供多种方式的支付选择。”

链接地址：

http://cio.it168.com/a2015/0105/1695/000001695683_all.shtml

大数据将加速形成新的技术经济范式

“关键生产要素”需要具备三个基本条件，一是成本较低并且相对成本迅速下降，二是在长期内几乎无限的供应能力，三是在整个经济系统中具有广泛的应用前景。围绕大数据技术而发生的移动终端、云计算、物联网等技术的集成应用，完全符合“关键生产要素”的基本特征。

技术经济范式是在一定社会发展阶段，由主导技术推动宏观和微观经济结构和运行模式发展的过程，并由此决定经济生产的范围、规模和水平。在新的技术经济范式形成过程中，占主导地位的科学技术将以革命性的方式迅速实现产业化、市场化，并不断对整个经济结构进行呈几何级数的渗透扩散，并逐渐改变原有的生产方式、管理方式、营销模式以及整个经济增长形态。当前，我国正处在实施创新驱动发展战略的关键时期，以新一代信息技术、生物技术、新能源、新材料为代表新兴技术群正在形成新的技术经济范式。

新技术经济范式形成的关键，在于是否基于主导技术形成了新的“关键生产要素”。这种“关键生产要素”需要具备三个基本条件：一是成本较低并且相对成本迅速下降；二是在长期内几乎无限的供应能力；三是在整个经济系统中具有广泛的应用前景。其中，围绕大数据技术而发生的移动终端、云计算、物联网等技术的集成应用，完全符合“关键生产要素”的基本特征，已经迅速地应用于经济社会发展的各方面，对技术开发、生产加工、商业模式等方面产生了深刻的影响，在新的技术经济范式形成过程中将成为决定性因素之一。这种影响，主要体现在数据资源、研发组织、技术融合和创新链衔接等方面。

大数据的集成分析将大幅度提高创新资源的使用效率。大数据的本质是面向海量数据的数据挖掘，发现隐藏的知识和规律，这为基础优化创新资源配置开辟了新的空间。根据美国麦肯锡公司 2013 年的报告，充分利用大数据技术能使零售商提高利润率 60%以上，使美国医疗保健行业降低成本 8%。经过多年的积累，我国形成了大量的科技文献、监测数据等大量科技基础信息。同时，也积累了大量面向市场的科技数据资源，例如技术成果、技术交易数据、高新技术企业、研发机构、大学科技园、科技企业孵化器等数据。这些数据往往形成相对独立、难

以探索的数据孤岛，而大数据的信息关联、智能决策等功能，能够对这些分割、离散的数据信息进行集成，并提供智能化、商业化的增值服务。

促进研发活动的去组织化和再组织

化。一方面，与传统以课题组、科研机构为基本单元的研发组织载体相比，社会化的研发组织将更为普遍，伴随移动互联网、社交网络的发展，研发活动的参与者越来越能够以个体的身份脱离学科领域、学术地位、空间等因素的限制，围绕特定主题参与到研究的策划和实施。另一方面，大数据技术将促使研发活动由精细化的单向组织管理走向趋势化的复合组织管理，对全局性预测的准确性和实时性要求更高，特别是对研发数据的在线收集和即时分析，为大规模研发活动的组织和协调提供支持。

促进跨领域的技术和产品研发。以生物医药产业和信息技术的融合为例：在研发环节，很多发达国家正尝试运用信息技术建立“虚拟人”，将药品临床试验的某些阶段虚拟化；针对电子健康档案海量、即时数据的挖掘和分析将有助于招募特定基因型的患者开展临床试验，研发基因导向型的个体化药物，这将大大加快药品研发效率，降低研发费用。在生产流通环节，无线射频识别标签、智能尘埃（超微型传感器）、温度传感器将在药品流通中广泛应用，提高药品流通行业集中度和流通效率。在医疗服务环节，电子病历、智能终端、物联网、网络社交软件等将使有限医疗资源让更多人共享，形成新的医患关系，并推动个体化的医疗服务。这些活动正在促使生物医药、信息技术两类传统意义上边界清晰的领域开始融合，而融合所必需的对海量即时数据的分析处理，都要以大数据、云计算等技术系统为前提。

缩短基础研究、应用开发到创新的进程。大数据带来的管理、检测等流程的优化将大大缩短研发周期。在基础研究方面，对海量数据的预测建模能帮助识别那些具有更高可能性的方案，这在药物分子筛选方面尤为明显。另一个案例来源于英特尔，其采用大数据技术开发的预测分析解决方案，能够收集生产过程中的历史数据，由此带来更快速的芯片研发，并将芯片的测试时间缩短 25%。

大数据在促进技术经济范式形成的过程中，需要相应的制度规范和保障。例如，在数据应用方面，既要鼓励科技数据，特别是财政投入形成的数据，实现更大范围、更及时的开放共享，也要通过立法和有效执法加强知识产权保护，注重

数据资产的价值，防止数据被滥用，明确界定数据挖掘、利用的权限和范围。在研发组织方面，虽然大数据在构建创新网络上具有明显优势，但也存在一定的局限性。欧盟最近的一项调查认为，在创新网络形成过程中，面对面的交流仍是不可或缺的因素。因此，大数据技术作为一项高效便捷的组织工具，其收集、分析和研判得出的关联机制，需要与学术研讨会、创新创业大赛、创业公开课等常规的、更加具象化的交流沟通方式紧密结合。在促进跨领域、跨环节的融合方面，需要各主管部门依托各类创新示范区、高新区、经济开发区，面向产品、服务、技术标准、合格评定程序等方面，集成各类创新资源开展大数据的试点示范，为大数据产业快速发展提供更加清晰的市场信号。

链接地址：

http://health.gmw.cn/newspaper/2015-01/05/content_103461642.htm

有数据就是这么任性，2014 年谁在玩转大数据？

借助大数据的力量进行巫术般地精准营销，年初爆红的美剧《纸牌屋》将大数据引入了普通人的视野。大数据无疑是当下除移动互联网外 IT 领域最热的讨论，简言之，从各种各样类型的数据中，快速获得有价值信息的能力，就是大数据技术。年底将至，今年互联网圈子里都是谁在接棒大数据，又玩出了什么新花样呢？

360 手机卫士十亿号码“提纯”10KB 专治 iPhone 骚扰电话

日前，困扰 iPhone 手机用户 7 年的骚扰电话问题终于得到解决，360 手机卫士 iOS 版发布更新，向非越狱的 iPhone 手机用户提供骚扰电话识别功能。据了解，360 手机卫士通过五年来形成的十亿级骚扰号码数据库，综合使用几十种聚类算法、十余种身份识别以及地域识别算法，通过 200 多个标签信息对手机用户进行分类细化。

此外，每天更新的骚扰号码库数据，会依据标记趋势调整骚扰号码库中各类数据比例，也就是说，每一位 360 手机卫士用户手机中的 1000 个骚扰号码都是动态的，随地域、身份以及骚扰趋势的变化而变化。

支付宝发布十年对账单 剁手党表示“我想静一静”

支付宝对外发布十年对账单，为用户梳理自支付宝诞生至今的购物、理财、生活缴费等全套数据。在一串串光鲜的数字背后，也成就了一段段剁手族的败家

史。不过不少网友在收到自己的账单后，惊呼“不忍直视”、“不知不觉可以买房了”、“原来自己也是高富帅”、“我的首付给了支付宝”等。

自 2004 年支付宝成立以来，全国人民十年网络总支出笔数为 423 亿笔 2014 年的移动支付占整体支付比例已经稳超 50%。各省移动支付占比的排名中，西藏、陕西、宁夏、内蒙垄断前四名。

双十一京东趣闻大数据　京东网友性福指数羞答答出炉

11.11 这一天，京东商城卖出了 80 万块香皂，重量约 115 吨，相当于 23 头大象；基情无限的同时，手纸卖出 900 万卷，8 亿多抽手纸，按一秒钟扯一抽的话，至少要扯 3 年，按一卷纸 30 米算，900 万卷至少可绕地球 7 圈。

在京东的这份数据里，性福指数分析占据了很大篇幅，北京城区性福指数对比结果是，昌平区性福指数最高，朝阳区居然是一片灰白色，性福指数是最低的。按道理讲朝阳区北京夜生活最丰富的地区了，曾经的天上人间，灯红酒绿的三里屯都在这里，双十一朝阳区的青年们都跑哪里去了？

情侣间的小心机　大数据带你解读“37 次想你”现象

一部小成本制作的爱情片，《37 次想你》连续数周停留在电影排行榜 Top10，最高纪录更是直逼 Top1 稳坐 Top2 。他们用 30 天的坚持，创造了 3799 次转发的“不断回忆过去，是一种以痛补痛的方式”；又用 3122 次转发，将一个简单的“寄明信片”活动变成了全民狂欢，甚至引发了情侣间的小风波。

8 个问题，不足 15 个页面，上线 5 天，参与人数突破 14 万。平均每一分钟就有将近 50 人参与此测试，后台每一小时会收到将近 3000 人为心中的 Ta 填写完的地址以及联系方式。其实，这个互动 h5 页面做为电影《37 次想你》营销链条中一个重要的结点，以宣传电影为主，却出乎意料的收获了如此强烈的反映，当然，其实更让人出其不意的是后台数据平台上所显示的一些参与数据。这些数据也惊现了全国范围内情侣间互不为知的“小心机”。

IBM 社交大数据技术：“上天台” 请留步

四年一度的顶级足球赛事已经告一段落，171 个进球让全世界为之振奋，球迷们在这一个月中过足了足球瘾。中国队虽然与世界杯无缘，但是中国人是此次世界杯不可缺少的人群。根据 IBM 数据分析，此次世界杯每天有有超过 1.2 亿人在社交媒体发声，微博上相关讨论超过十亿次。其中关于足彩的讨论脱颖而出，

以 355678 次的频率成为网友讨论最多的话题。

在本次世界杯期间，IBM 便通过社交大数据技术对舆情数据进行了分析，在半决赛开始前已经通过大家的支持率准确判断出了四强名次。德国、巴西、阿根廷、荷兰这四强中，德国队以 16%得到最高的支持率，而东道主巴西队因为当家球星内马尔的受伤，不被广大球迷看好，仅以 8%位列四强之尾。阿根廷与荷兰分别以 15%和 10%位列二、三位。如果大家都看了 IBM 的分析数据再买彩票，估计都不用“上天台”了。

链接地址：

http://www.adquan.com/post-1-29399.html

不仅仅是机遇 细数大数据领域待解决问题

大数据，现在已经不仅仅是人们日常工作和生活当中的必需品了，很多国家已经开始将大数据技术和应用上升到国家的战略层面，在 2012 年 3 月，美国政府就宣布将大数据以及相关产业上升为国家战略，很多行业包括军事、能源等都被列入到了大数据应用领域。

其实从上述内容我们不难看出，大数据的诞生和发展带给我们的不仅仅是机遇，同时在技术和应用层面用户也面临着很多挑战和困难，放眼国内的大数据领域市场，有很多行业压力摆在我们面前，本期我们就来说说国内目前的大数据仍然面临的几大问题。

数据来源良莠不齐

我们都知道，我国国内的人口众多，大数据给我们带来的机遇和压力都不小，作为一个新兴领域，尽管大数据意味着大机遇，拥有巨大的应用价值，但同时也遭遇工程技术、管理政策、人才培养、资金投入等诸多领域的大挑战。只有解决这些基础性的挑战问题，才能充分利用这个大机遇，让大数据为企业为社会充分发挥的最大价值与贡献。

丰富的数据源是大数据产业发展的前提。而我国数字化的数据资源总量远远低于美欧，每年新增数据量仅为美国的 7%，欧洲的 12%，其中政府和制造业的数据资源积累远远落后于国外。

现在很多企业时时刻刻都在产生着大量数据，但这些数据如何归集、提炼始终是一个困扰。而大数据技术的意义确实不在于掌握规模庞大的数据信息，而在

于对这些数据进行智能处理，从中分析和挖掘出有价值的信息，但前提是如何获取大量有价值的数据。

大数据时代，我们需要更加全面的数据来提高分析预测的准确度，因此我们就需要更多便捷、廉价、自动的数据生产工具。除了我们在网上使用的浏览器有意或者无意记载着个人的信息数据之外，手机、智能手表、智能手环等各种可穿戴设备也在无时无刻地产生着数据。

云计算平台和大数据之间的相辅相成关系是现在 IT 业界所共识的，机等各种网络入口以及无处不在的传感器等，都会对个人数据进行采集、存储、使用、分享，而这一切大都是在人们并不知晓的情况下发生。

数据分析模型建设困难

现在越来越多的用户开始试图用大数据分析技术来去解决很多问题，但是大数据的大，一般人认为指的是它数据规模的海量。随着人类在数据记录、获取及传输方面的技术革命，造成了数据获得的便捷与低成本。

大数据的真正价值不在于它的大，而在于它的全面：空间维度上的多角度、多层次信息的交叉复现；时间维度上的与人或社会有机体的活动相关联的信息的持续呈现。

要以低成本和可扩展的方式处理大数据，这就需要对整个 IT 架构进行重构，开发先进的软件平台和算法。这方面，国外又一次走在我们前面。特别是近年来以开源模式发展起来的 Hadoop 等大数据处理软件平台，及其相关产业已经在美国初步形成。

用户使用权和隐私的平衡

很多人现在一说到大数据就“谈虎色变”，究其很重要的原因之一就是大数据挖掘和分析技术带来的用户隐私的泄露。有专业人士指出，中国人口居世界首位，但 2010 年中国新存储的数据为 250PB，仅为日本的 60%和北美的 7%。2012 年中国的数据存储量达到 64EB，其中 55%的数据需要一定程度的保护，然而目前只有不到一半的数据得到保护。

笔者在以前的文章当中曾经写过，大数据技术其实是一把双刃剑，我们如何在推动数据全面开放、应用和共享的同时有效地保护公民、企业隐私，逐步加强隐私立法，将是大数据时代的一个重大挑战。

数据增值的关键在于整合，但自由整合的前提是数据的开放。在大数据的时代，开放数据的意义，不仅仅是满足公民的知情权，更在于让大数据时代最重要的生产资料、生活数据自由地流动起来。

数据的管理难度

海量数据通过挖掘、收集、存储、分析、最后被应用在不同行业当中，这当中的众多步骤在管理方面都是需要仔细计划的。因为显而易见，大数据的用户体验效果很有可能直接影响到企业以及个人用户的一些决策。

大数据能够真正发挥作用，深层次看，还要改善我们的管理模式，需要管理方式和架构的与大数据技术工具相适配。大数据应用领域仍窄小，应用费用过高，制约大数据应用。国内能利用大数据背后产业价值的行业主要集中在金融、电信、能源、证券、烟草等超大型行业。

链接地址：

http://www.bigdatas.cn/article-1448-1.html

Connection Analytics 引领下一代分析技术

2014 年时曾提到，业内大数据分析面临几个重大的转折点。我们需要新技术和新工具，帮助更多用户更合理地利用数据，而且迫切需要更广泛的数据分析功能，从不同来源的所有数据中发现它们之间的关系，并获得洞察力。为此，Teradata 的创新型技术 Connection Analytics 技术将工具为大数据分析行业开辟了新的格局。

今年一月，我在展望 2014 年时曾提到，业内大数据分析面临几个重大的转折点。随着数字 化时代数据规模和复杂度呈指数级增长，我们需要新技术和新工具，帮助更多用户更合理地利用数据。我还提到，我们迫切需要更广泛的数据分析功能，从不同来源 的所有数据中发现它们之间的关系，并获得洞察力。

仅关注客户或网络等特定分析实体的内容已不足以满足企业需求，我们还需要了解这些实体之 间的关系情境，通过跟踪用户、产品及过程之间对结果产生影响的关系变化，获得洞察力并创造价值。但这不仅仅是数据科学家的专利，我们还需要通过各种途径帮 助普通商业用户轻松、直观地获得并运用这些洞察力。目前，Teradata 天睿公司已推出全新分析功能，以满足这些要求，对此我倍感骄傲。

我们在上周举办的 2014 年 Teradata 合作伙伴大会上发布 Connection Analytics，这是一套高级情境分析功能，能够以较低成本大规模应用于大型多结构数据集。Connection Analytics 基于 Teradata Aster 强大的 MapReduce 及 Graph 引擎，可运用 100 多种预置算法帮助数据科学家乃至普通商业用户理清复杂的关系，并从中梳理出获得全新业 务洞察力并创造价值的成功模式。Connection Analytics 将作为 Teradata 统一数据架构下 Teradata Aster 探索平台的重要组件供用户即时使用。

据我们发布的 Connection Analytics 新闻稿，Connection Analytics 能够在用户可访问的环境下实现上述功能，并与现有基于 SQL 的可视化能力及商业智能应用无缝整合，在业内率先将高级情境分析能力与易用 性完美结合。这将为更多商业用户提供多种洞察力，帮助他们梳理各种关系，用于预测业务欺诈行为或客户流失，开展精密策划的病毒式营销活动，提升公共网络健 康度与安全性及优化推荐引擎。

到目前为止，情境分析仍存在高难度、高成本等挑战，因为它需要专用系统及难以企及的独特 技能组合，并结合多种算法，才能发现这些错综复杂的关系。现有基于内容的决策模型侧重用户、产品或过程的个体特性分析，而 Connection Analytics 拥有基于情境的决策模型，可分析这些实体之间的相互关系。部署 Connection Analytics 后，数据科学家乃至商业用户将能够运用熟悉及易用的工具增强现有决策模型，实现大数据分析最前沿技术的普及应用。

但所有这些讨论都仿佛是在纸上谈兵。现在，我将介绍一些即将发布的价值驱动型用例。例如 （怎样减少）客户流失：通过部署 Connection Analytics，用户能够将传统统计方法、机器学习及情感分析与影响因素分析相结合，调查客户满意度，并在客户群中准确找出最具影响力的群体。这将帮 助企业减少客户流失，并在客户流失时尽量避免连锁反应。Connection Analytics 还能够找出对购买产品构成最直接及间接影响的因素，为病毒式营销活动有针对性地提供信息。

Connection Analytics 还帮助企业监测 IP、网络、服务器和通信日志不断生成的各种数据流，实现网络威胁的近实时监测。Connection Analytics 可跟踪用户、产品、过程及其它“实体”之间关系，这对于破解组织严密的诈骗团伙至关重要。当诈骗人员创建新的身份，或改变其诈骗手段 时，如仅使用基于内容的决策模

型，用户将轻易上当。但通过使用 Connection Analytics，将帮助用户运用基于情境的决策模型，增强传统上较为肤浅的分析视图，获得暴露可疑活动并识别诈骗集团的算法模式。

在当前数字化时代中，万物皆有联系。因此，企业和公共部门机构需通过关系建模分析，了解 不同数据集之间的关系。Teradata 天睿公司推出 Connection Analytics，为情境式决策专门开发出可供用户访问的高性能分析平台，率先为广泛的用户群体提供企业级分析能力，为大数据分析行业开辟了新的格局。 在大数据分析技术处于重要转折点时，Teradata 天睿公司将通过技术创新，推动行业不断发展，并帮助客户取得成功。

链接地址：

http://www.bigdatas.cn/article-1430-1.html

2015 贵阳国际大数据产业博览会将于五月开幕

中新网 1 月 8 日电 今天上午，2015 贵阳国际大数据产业博览会暨全球大数据时代贵阳峰会(以下简称数博会)新闻发布会在京举行。贵州省经济和信息化委主任李保芳，贵阳市委副书记、市长刘文新，遵义市委常委、副市长郑欣，贵安新区管委会副主任欧阳武，北京市贸促会副会长林彬，中国互联网协会副理事长高卢麟，宽带资本董事长田溯宁出席新闻发布会，并就数博会相关内容回答了媒体提问。贵阳市委常委、宣传部部长兰义彤出席会议；贵阳市副市长高卫东主持会议。

据了解，此次数博会将于 2015 年 5 月 26 日至 29 日，在贵阳国际会议展览中心举行，主办单位为贵阳市人民政府、遵义市人民政府、贵安新区管委会、贵州省经济和信息化委、北京市贸促会、中国互联网协会，协办单位为工信部国际经济技术合作中心、中国信息协会大数据分会、中国呼叫中心与 BPO 产业联盟、IDC 国际数据公司、中关村大数据产业联盟、中关村大数据交易联盟，以“专业展会、国际平台、促进合作、共谋未来”为原则，以“大数据时代的变革、机遇和挑战”为主题，届时将举行展览展示、峰会论坛和创新大赛等活动，综合呈现大数据技术、应用和发展趋势。

此次数博会将呈现出四个特点：一是展示前沿技术。将设国际精英馆、大数据应用馆、大数据设备馆、大数据软件和服务馆四个展馆，面积约 4 万平方米。

其中，国际精英馆是本次展览的主题馆，将汇聚世界顶尖企业以特装方式展示新成果、新产品、新技术。大数据应用馆集中展示以大数据为核心支撑的热门行业，重点是智慧城市、大数据金融、大数据营销、移动互联网、车联网、大数据健康等。大数据设备馆将吸引大数据产业硬件设备及其制造商和解决方案提供商，包括存储及服务器板块、网络通信设备板块、大数据信息安全板块、机房设备板块、可穿戴设备板块等。大数据软件和服务馆定位在大数据软件和数据处理技术、数据交易平台及关联服务等方面的展览展示。

二是探讨发展趋势。将举办 1 个峰会和若干分论坛。邀请国内外大数据领域知名企业家、专家学者，在峰会上发表主旨演讲并展开高峰互动对话交流，探索大数据产业发展趋势及前沿理论。将围绕“大数据的交易和互换”、“大数据时代下政府的‘智’与‘治’”、“大数据驱动金融创新”、“民生与健康大数据”、“大数据技术发展趋势和产业变革”、“大数据的战略与方向”等专题开展分论坛。

三是催生新兴业态。将以“云上贵州·数聚贵阳”为主题，紧紧围绕政务数据开放和数据交易，举办大数据创新应用大赛，并揭晓贵阳正在策划开展的“大数据推动政府改革”、“大数据改善民生”应用创意征集活动首批成果。同时，率先在世界发布数据确权、数据定价、数据保险、数据货币，以及数据的登记、交割等一系列大数据交易及相关标准，将促进大数据应用由“条数据”向“块数据”突破，打破传统的信息不对称和物理区域、行业领域对信息流动的限制，培育一批基于大数据的信息消费、金融服务、先进制造等新兴业态。

四是云集业界精英。举办以大数据为主题的博览会和峰会在全球尚属首次，阿里巴巴董事长马云、富士康科技集团董事长郭台铭等世界著名企业家已明确表示将出席本次活动，惠普公司、趋势科技、神州数码等也已接受了组委会的邀请，目前组委会正在向 IBM、微软、苹果等全球领先的大数据企业发出诚挚邀请，将云集更多业界精英，共谋创新发展。

贵阳市政府是举办此次数博会的最先发起方，贵阳市发展大数据产业具有五大优势：

一是生态环境优势。“爽爽的贵阳”自然条件与有印度“硅谷”之称的班加罗尔相似，气候凉爽，清新的空气稍微过滤就可以直接进入机房，符合精密制造业研究发展的要求和创新创业者的宜居选择。贵阳的地质构造稳定，地震、台风

等灾害罕见，信息网络设备的“安全系数”很高，对大数据产业企业和高智商、高知识、高投资、高收入群体的吸引力很强。

二是资源禀赋优势。贵阳是天然的“大宝库”，磷、铝、煤等矿产资源储量丰富，特别是作为“西电东送”的起源，电力水火互济，稳定可靠，电力价格具有相对的优势，而发展大数据产业需要电力作为保障，在这方面贵阳具有独特的优势支撑。

三是产业基础优势。贵阳的电子产业起步较早，上世纪五六十年代就以 011、061、083 军工基地为核心，布局形成了航空、航天、电子三大产业体系。当前，在做大产业存量的同时，积极做优新的产业增量，以创新驱动、转型升级为载体，以中关村贵阳科技园统筹全市产业和园区发展，形成了先进科技资源持续引入的有效机制，着力构建大数据产业发展生态环境。去年全市新注册大数据及关联企业 227 家，云上企业超过 2900 家，上线产品达 1.4 万个，阿里巴巴、惠普、富士康、亿赞普等一大批国际国内知名企业相继入驻，越来越多的人才、技术、服务等要素汇集贵阳，为发展大数据产业提供了后续的有力支撑。

四是市场优势。近来，中国电信、移动、联通三大运营商数据中心落户贵安新区，惠普、中电乐触、高新翼云、翔明科技等企业和北京市供销合作总社正在贵阳建设 50 万台服务器规模的数据中心，数十家银行、保险等金融机构也将数据中心搬到贵阳，使贵阳及周边区域成为国内乃至全球最大的数据聚集地之一。从现在全国的数据中心布局看，贵阳是长江以南重要的大数据节点城市，而且是南方数据灾备中心。这些都让贵阳从信息产业的末梢变成了中枢节点，成为大数据市场的中心地段。

五是政策叠加优势。从 2012 年“国发二号文”支持贵州培育发展战略性新兴产业等政策，到贵州省政府出台加快大数据产业发展应用若干政策，再到本市制定实施相应的支持政策，贵阳的相关政策环境进一步优化，目前工信部已确定在贵阳市、贵安新区创建全国首个大数据产业发展集聚区，也将有相应的支持措施，这一系列利于大数据发展的支持政策，强力推动着贵阳大数据产业发展。五大优势，有贵阳自身的先天优势，也有经过努力后形成的后发优势，现在都聚在一起集中释放能量，2014 年 3 月贵阳被中国数据中心产业联盟授予“最适合投资数据中心的城市”称号，贵阳正在大步迈入大数据时代。

在此次新闻发布会上，北京市贸促会副会长林彬表示：“我们希望将本届贵阳数博会打造成大数据领域专业的博览会，全面展示和共同分享大数据领域前沿科技、创新产品和解决方案，为国内外大数据领域企业和机构提供一个专业化，国际化的平台。”

链接地址：

http://news.sina.com.cn/c/2015-01-09/135731380635.shtml

当当举办"中国童书年会" 大数据描绘中国童书全景图

中新网 1 月 9 日电 1 月 7 日，主题为“我们的世界”第三届中国童书编辑与营销年会在京举行，300 多位全国优秀的童书编辑齐聚一堂，共话中国童书未来。

中国少读工委主任、中国少年儿童出版总社社长李学谦为本次大会致辞，著名儿童文学作家殷健灵女士应主办方邀请客串圆桌论坛主持，与著名儿童文学作家、诗人金波先生、著名儿童插画家朱成梁、陈泽新先生与江苏凤凰少年儿童出版社青少年读物事业部主任陈文瑛女士一起探讨作家、画家、编辑如何凝聚力量创作出好作品，是本次年会的亮点之一。

与往届相比，本届年会在内容和环节上都进行了很大的丰富，除重量级嘉宾致辞、演讲等环节外，还与大家分享了儿童品牌与童书文化跨界联合的成功案例等诸多精彩内容。作为会议的主办方，当当也在年会上发布了 2014 年童书数据，并就当当童书在 2015 年的发展战略进行了披露。

中国童书新“命题”：我们的世界

与前两届的编辑年会主题相比，今年当当的主题是“我们的世界”。当当认为，“我们生活的这个世界是丰富而复杂的，是温暖而艰辛的，是美丽而残酷的，更是宽广而深厚的……当当童书希望孩子们通过阅读去弥补体验的不足；通过阅读和孩子们一起，用眼、用心、用脑，探寻我们的别样世界，”基于此，当当希望在即将到来的 2015 年，和小读者们一起，探寻“我们的世界”。

在“我们的世界”这一主题下，当当在年会上与全国优秀编辑就童书的各种话题进行了共同研讨。一直以来，当当以“让中国孩子的阅读与世界同步，把更多的好书给更多的孩子”作为当当童书的使命。某种程度上说，“我们的世界”不仅是当当本届童书会的主题，也是当当童书以及全国各大出版机构共同面对的课

题，如何去理解、探索、满足孩子们的阅读需求和心理世界，如何让阅读成为孩子们成长的助力器，是包括当当童书在内的相关从业者共同的责任。

大数据描绘“全景图”中国童书市场前景广阔

在年会上，当当童书发布了 2014 年童书数据，为与会人士和业界分享了真实的童书“市”界。数据显示，当当童书 2014 年销售了 1.1 亿册童书，销售码洋 23.2 亿。占据中国线上线下童书零售市场份额的 1/3 以上，高端的图画书、科学书、玩具书更是占有 60-70%的市场份额。这些亮眼的数据与顾客特征、地域特征、年龄分布、品类结构等一道构成了 2014 年中国童书的“全景图”，让我们看到了过去一年中国儿童的阅读消费情况，也为未来童书市场的发展提供了宝贵参考。

当当在年会现场发布的未来童书品类发展态势表明，科普书、原创文学市场需求大，占比一直逐年增加；0-2 岁婴儿读物销售增加，更多的妈妈关注婴幼儿早期阅读。与中国 3.7 亿中国儿童数量相比，当当童书过去一年 1.1 亿册童书的销售，意味着中国童书将会有更为广阔的市场空间。作为中国最大的童书平台，当当童书更是首当其冲，肩负着中国儿童普及阅读的使命，将更多的孩子从游戏机前拉回到阅读中。

加码图百混合模式，当当婴童全渠道模式再升级

值得一提的是，在本届年会上，当当童书还发布了 2015 年的发展模式：继续深度合作图百混合，深化婴童全渠道发展战略。作为中国图书市场老大，当当在近年来不断加大服装日百等非图书业务的发展力度，凭借图百混合的发展战略，也确实推动当当在 2013-2014 年连续四个季度实现盈利。

有道是趁热打铁，从本届年会上透露的信息看，当当正是要趁着图书和非图书业务相得益彰、互促互进的火候，发力婴童全渠道战略，从线上到线下，深入跨界合作，升级其图百混合模式。本届大会协办方小星辰品牌集团在会议上介绍了目前与当当的深入合作，未来当当将与小星辰在 2015 年开展战略合作，通过当当文化消费大数据，打通品牌商、出版商及作家之间的产业合作，升级婴童文化产业链。同时，当当童书还已与小星辰陆续在线下各大城市高端商场实体店联合打造童书阅读体验区，并定期举办亲子阅读活动，为提升亲子阅读体验创作环境，同时为家长一站购物提供可能。

作为中国最大的童书编辑与营销盛会的主办方，当当致力于为优秀的童书作家、出版人和编辑们建立一个良好的沟通和交流的平台，迄今已连续举办三届。在每年年会上，与会人士都共同见证中国优秀童书出版人在过去取得的骄人成绩，分享成功经验，探讨童书的发展趋势，感悟优秀编辑的价值，提升营销策略。也正因为此，当当童书倾力打造的“中国童书编辑与营销年会”吸引了众多行业精英，已成为一场中国童书界的思想盛会和传播主流价值观的行业盛典。当当童书立志与所有关爱儿童阅读的人一道，打造中国儿童阅读的风向标。

链接地址：

http://finance.ifeng.com/a/20150109/13418709_0.shtml

大数据开放可提升政府公信力

1 月 5 日，美国哥伦比亚新闻评论网发表的文章《21 世纪的新闻审查》（21st-Century Censorship）提出，互联网技术的发展产生了越来越多的媒体平台和自媒体。有观点认为，在大数据时代下，政府对信息的监管和审查会变得日益乏力。然而，学者们则认为，事实可能恰恰相反。

政府与媒体关系发生变化

美国杜克大学斯坦福公共政策学院教授菲利普·班尼特（Phillip Bennett）表示，对新闻业来讲，首先，互联网对新闻媒体的影响是破坏性和颠覆性的，这是不可阻挡的趋势；其次，它催生了互联网通信工具的创新。同样，政府公共部门与媒体间的关系也发生转变。政府不再像以前一样“单纯”地面对并处理与几

家主流媒体间的关系，而是面临着由公共舆论所引发的更加复杂多样的挑战。

据美国互联网数据中心数据显示，互联网数据每年约增长 50%，每两年便翻一番，而目前全球 90%以上的数据是近几年才产生的。有人曾预言，随着互联网和大数据平台的日趋开放，面对海量数据的“无从监管”，政府对新闻信息的监管和审查将随之消失。但事实上，大数据概念从诞生开始，政府监管在其中扮演的角色从未被低估。正如大数据技术的战略意义不在于掌握庞大的数据信息，关键在于对这些有意义的数据进行专业化处理的能力。比如，对新闻信息的处理，就不仅仅是个技术问题。

班尼特表示，在“自由的”网络世界和各国政府对海量新闻信息源的处理过

程中，我们看到了这个时代的一种悖谬的“传播风格”，一方面，人们更加善于表达自己，言论途径更加开放；另一方面，政府监管随着大数据的发展在某些方面“愈加严格”

“当前，一个事实是：监管和审查随着信息时代的发展而不断完善。”班尼特说，理论上，新技术的发展应该使新闻信息的审查更加困难，并最终成为“不可能的审查”，政府难以控制信息的流通。然而，现实中越来越多的政府开始加大监管、审查新闻及各路媒体信息的力度，不仅是发展中国家，也包括发达国家，尤其是在互联网安全意识的提高下，对新媒体的审查越来越“用力”。

英国伦敦城市大学学者格兰达·库珀（Glenda Cooper）日前在澳大利亚对话网站刊文表示，全民新闻时代的来临，打破了传统的派记者去现场、报道、发布新闻的流程，现在很多记者都是在办公室里从社交媒体中搜索有价值的新闻。但这种新的方式，在新闻的隐私性、验证真伪、品位方面都有值得质疑的地方。自媒体新闻往往是倾向性明显的新闻，常伴随伪造、暴力的图片，给社会造成不良影响。

据了解，目前，许多国家通过立法加强对新闻媒体行为的规范，一方面是制定专门的新闻法规，如瑞士、法国、意大利、丹麦等国；另一方面，即使没有整体立法，但相关法律条文“散见”各处，如美国、英国、日本等国，尤其是针对互联网信息的传播。

杜克大学教授莫伊塞斯·纳伊姆（Moises Naim）表示，各国政府正在从数字革命的“观众”转变为数字时代先进技术的最早采用者，政府越来越有能力掌握新闻记者和自媒体的情况，并在一定程度上控制信息的流通过程和方式。政府的监管能力发生了变化，与媒介的关系也发生了变化。

监管方式的模仿与创新

纳伊姆说，今天，一方面许多国家的政府正在使信息更加开放；另一方面政府也需要在监管和审查方式上进行创新。当前，在一些国家，如匈牙利、厄瓜多尔、土耳其和肯尼亚，政府正在模仿俄罗斯等国家，对重要新闻进行更大程度的管控，并力图建立国家媒体品牌的影响力。

“虽然互联网的发展是一个全球化的趋势，但对于信息的监管和审查是每个国家或地区的自主行为，比如，在匈牙利，政府规定权威媒体有权收集记者的详

细信息、广告内容和社论内容。”纳伊姆说，我们无法单纯以“新闻自由”为幌子全盘否定政府对媒体和信息的审查和监管行为，尤其是在今天的海量数据面前，每个国家都要考虑互联网时代的信息安全。纳伊姆反问道，当网民在网络平台上遭受他人肆意而名目张胆的人身攻击时，是否还能看到“言论自由”的积极面？

政府监管将更加透明化

许多学者认为，不论新闻及其他信息的监管和审查面临多少偏见和争议，在大数据平台迅速发展和广泛应用背景下，政府监管方式都将更加透明。

大数据的开放能够增加政府的透明度，提高政府的公信力。纽约大学法学院教授贝丝·诺维克（Beth Novick）表示，数据的开放可以让政府公职人员和民众一起参与进来，解决政府无法完成的、棘手的问题，更广泛地发挥群众力量，借助大数据平台进行更好的社会管理。

当然，大数据的存储和处理模式，不可避免地会带来信息安全、隐私保护等问题。一方面，大数据时代的国家信息安全需要引起政府高度重视，应大力推动国家的网络信息安全建设；另一方面，从全球范围来看，目前已有 50 多个国家依靠法律形式规范个人信息数据的管理与使用。显然，大数据时代，人们的隐私应该受到更好的法律保护。

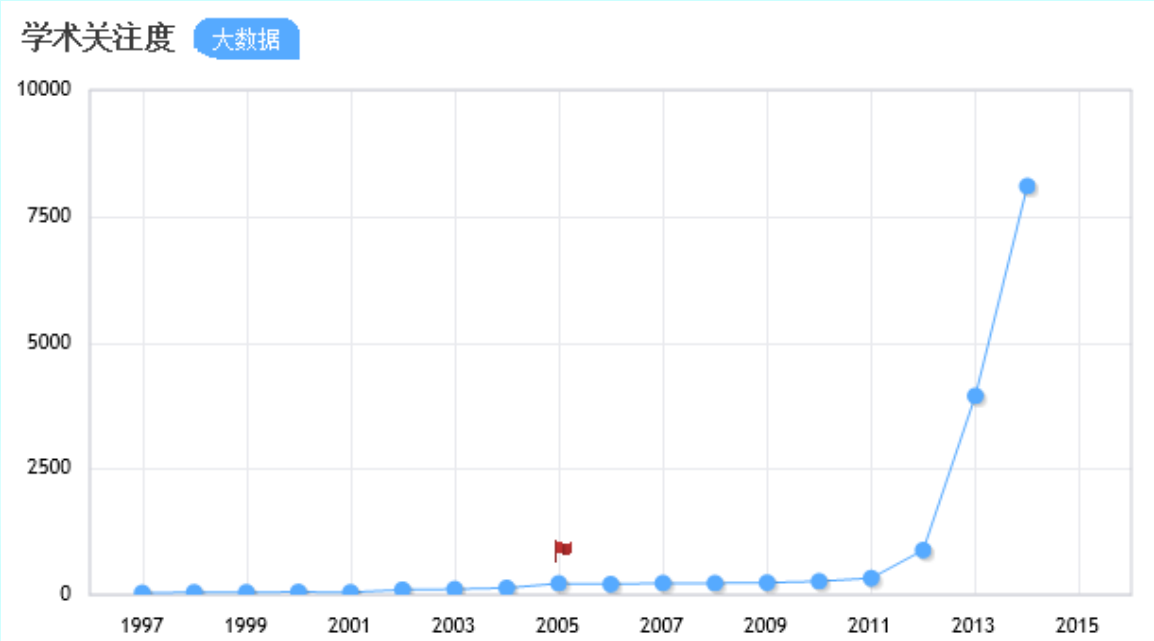
链接地址：

http://www.gmw.cn/xueshu/2015-01/09/content_14459032.htm

【国内文献计量分析】

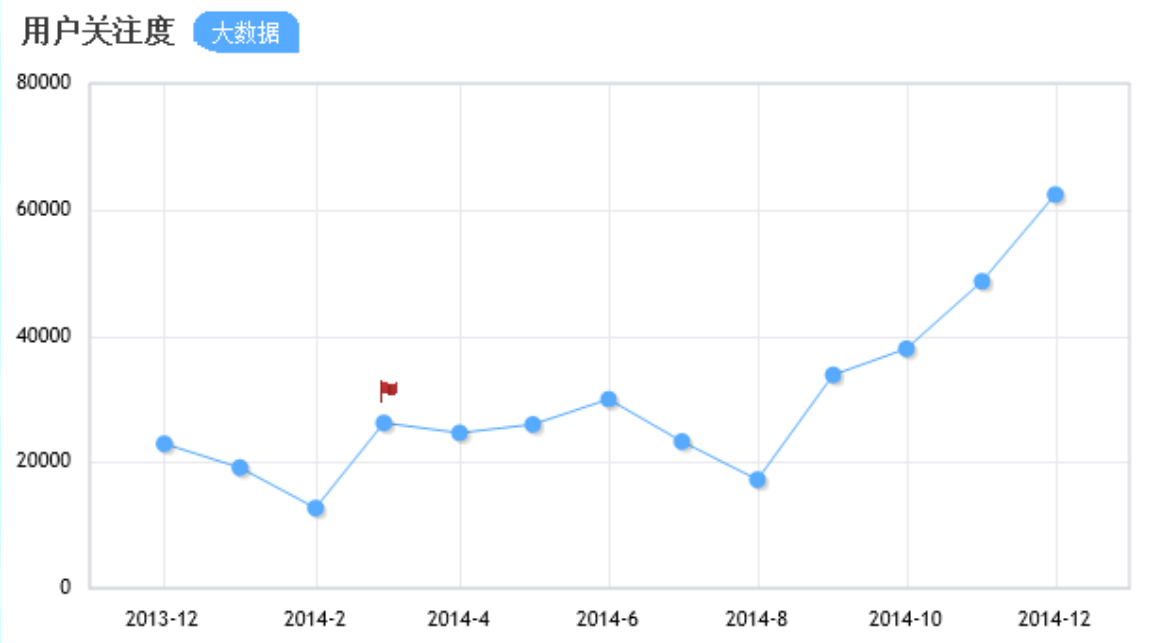
“大数据”学术关注度

本文献计量分析以“CNKI 学术趋势”为分析工具，该工具依托于中国知识资源总库和千万用户的使用情况，提供学术发展趋势分析。该分析以“大数据”为检索点，关注本研究领域的学术热点，展示其学术发展历程，发现经典文献。



本趋势图表明“大数据”1997—2001 年还未引起学术届广泛关注，自 2002 年开始有相对多的研究学者发表该领域的论文，但随后几年关注量增幅也不明显。到 2005 年标识点数高于前后两个点，随后几年随略有增长，但从趋势图上看并不明显，自 2011 年，相关论文收录量开始呈直线式激增，到 2014 年收录量达到顶峰，数量达 8097 篇。

“大数据”用户关注度



该图表是关于 2013 年 12 月至 2014 年 12 月有关“大数据”文章的用户下载情况，这一年之中，用户下载量波动不是很大大，2013 年 12 月后下载量有小幅度递减后，至 2014 年 2 月下载开始逐年递增，至 2014 年 3 月下载量达到一个顶峰值，标识点数值高于前后两点，数量为 26223 篇，随后下载量又开始小幅度增长，并出现波动。2014 年 8 月开始，下载量出现直线式增招趋势，研究者可通过该标识点进行调研，进一步发现研究思路。

“大数据”热门被引文章

序号	文献名称	作者	文献来源	发表时间	被引频次
1	多媒体传感器网络及其研究进展	马华东;陶丹	软件学报	2006-09-30	415
2	三维 GIS 的基本问题探讨	肖乐斌,钟耳顺,刘纪远,宋关福	中国图象图形学报	2001-09-25	382
3	粗糙集理论及其应用进展	胡可云,陆玉昌,石纯一	清华大学学报(自然科学版)	2001-01-30	351
4	大数据管理:概念、技术与挑战	孟小峰;慈祥	计算机研究与发展	2013-01-10 07:44	347
5	架构大数据:挑战、现状与展望	王珊;王会举;覃雄派;周烜	计算机学报	2011-10-15	250
6	大数据研究:未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考	李国杰;程学旗	中国科学院院刊	2012-11-15	214
7	基于近邻传播算法的半监督聚类	肖宇;于剑	软件学报	2008-11-15	169
8	大数据分析——RDBMS 与 MapReduce 的竞争与共生	覃雄派;王会举杜小勇;王珊	软件学报	2011-09-09	167
9	Internet 上的文本数据挖掘	王伟强;高文;段立娟	计算机科学	2000-04-15	151
10	基于数据挖掘的 SVM 短期负荷预测方法研究	牛东晓;谷志红;邢棉;王会青	中国电机工程学报	2006-06-25	137

“大数据”热门下载文章

序号	文献名称	作者	文献来源	发表时间	下载频次
1	大数据管理:概念、技术与挑战	孟小峰;慈祥	计算机研究与发展	2013-01-10	25290
2	大数据研究:未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考	李国杰;程学旗	中国科学院院刊	2012-11-15	13861
3	网络大数据:现状与展望	王元卓;靳小龙;程学旗	计算机学报	2013-06-15	13450
4	架构大数据:挑战、现状与展望	王珊;王会举;覃雄派;周烜	计算机学报	2011-10-15	10147
5	大数据分析——RDBMS 与 MapReduce 的竞争与共生	覃雄派;王会举;杜小勇;王珊	软件学报	2011-09-09	6326
6	大数据研究综述	陶雪娇;胡晓峰;刘洋	系统仿真学报	2013-07-25	6241
7	大数据研究	严霄凤;张德馨	计算机技术与发展	2013-04-10	4875
8	大数据环境下技术创新管理方法研究	朱东华;张嶷;汪雪锋;李兵;黄颖;马晶;许幸荣;杨超;朱福进	科学学与科学技术管理	2013-04-10	4378
9	大数据与大数据经济学	俞立平	中国软科学	2013-07-28	4032
10	大数据的概念、特征及其应用	马建光;姜巍	国防科技	2013-04-20	3947

温馨提示:

以上文章可在本期专题报道的附件中获得!

“大数据_数据处理”研究热点

序号	热点主题	主要知识点	主题所属学科名称	热度值↓	主要文献数	相关国家课题数	主要研究人员数
1	matlab;数据处理;化工计算;	matlab;仿真;化工计算;化学吸收;信号与系统;可视化;图像处理;数据处理;优化设计;simulink;符号工具箱;神经网络;excel;图形用户界面;化学计量学方法;数学建模;机械优化设计;漆水河;机械控制工程基础;曲线拟合;	无机化工;计算机软件及计算机应用;	★★★★★	188	13	356
2	matlab;应用;数据处理;	matlab;数据处理;应用;图像处理;excel;仿真;曲线拟合;图像压缩;小波变换;测量;可视化;图形用户界面;信号与系统;优化设计;excel link;工具箱;边缘检测;数值计算;神经网络;matlab 语言;	计算机软件及计算机应用;	★★★★★	298	2	542

“大数据_数据挖掘”研究热点

序号	热点主题	主要知识点	主题所属学科名称	热度值↓	主要文献数	相关国家课题数	主要研究人员数	主要研究机构数
1	数据挖掘;数据挖掘技术;关联规则;	数据挖掘;数据挖掘技术;关联规则;关联规则挖掘;数据仓库;apriori算法;频繁项集;决策树;电力系统;知识发现;关联规则算法;算法;应用;频繁项目集;遗传算法;挖掘算法;最小支持度;故障诊断;数据库;频集;	计算机软件及计算机应用;	★★★★★	551	68	981	341
2	网络营	网络营销;中小企业;数	贸易	★★★★★	248	95	220	146

	销;数据挖掘;企业;	据挖掘;电子商务;营销策略;企业;搜索引擎;消费者;互联网;传统营销;博客;网络广告;对策;博客营销;营销模式;策略;产品;中小企业网络;企业网站;数据挖掘技术;	经济;					
3	数据挖掘技术;电子商务;数据挖掘;	数据挖掘技术;电子商务;数据挖掘;web 挖掘;数据仓库;关联规则;应用;客户关系管理;电子商务网站;子商务;知识发现;聚类分析;关联规则挖掘;web 数据挖掘;方法;数据库;电子商务发展;会计信息系统;xml;客户关系管理系统;	计算机软件及计算机应用;	★★★★★	259	48	367	204
4	数据挖掘;数据挖掘技术;web 挖掘;	数据挖掘;数据挖掘技术;web 挖掘;关联规则;数据挖掘算法;数据仓库;知识发现;关联规则挖掘;入侵检测系统;聚类算法;文本挖掘;web;聚类分析;xml;电子商务;web 服务;web 数据挖掘;聚类方法;入侵检测;web 日志挖掘;	互联网技术;计算机软件及计算机应用;	★★★★★	413	65	723	274
5	数据挖掘;数据挖掘技术;客户关系管理;	数据挖掘;数据挖掘技术;客户关系管理;crm;数据仓库;知识管理;应用;关联规则;客户关系管理系统;电子商务;客户关系管理(crm);商业智能;数据库;知识发现;客户细分;交叉销售;客户流失;呼叫中心;知识管理系统;聚类分析;	计算机软件及计算机应用;	★★★★★	354	44	542	248
6	数据挖掘;关联规	数据挖掘;关联规则挖掘;关联规则;数据仓	计算机软件	★★★★★	250	40	430	215

	则;关联规则挖掘;	库;apriori 算法;数据挖掘技术;频繁项集;知识发现;中医药;遗传算法;应用;关联规则算法;数据库;算法;决策树;聚类算法;聚类分析;挖掘算法;人工智能;最小支持度;	及计算机应用;					
7	数据仓库;数据挖掘;决策支持系统;	数据仓库;数据挖掘;数据挖掘技术;联机分析处理;决策支持系统;olap;决策支持;元数据;数据库;电力系统;关联规则挖掘;关联规则;商业智能;客户关系管理;联机分析;数据集市;电力营销;资本市场;数据仓;继电保护;	计算机软件及计算机应用;	★★★★★	257	49	435	201
8	数据仓库;数据挖掘;联机分析处理;	数据仓库;数据挖掘;数据挖掘技术;联机分析处理;决策支持系统;olap;商业智能;客户关系管理;数据库;关联规则;数据质量;商务智能;关联规则挖掘;数字图书馆;决策支持;crm;数据分析;his;商业银行;数据模型;	计算机软件及计算机应用;	★★★★★	182	50	273	154
9	粗糙集理论;属性约简;数据挖掘;	粗糙集理论;属性约简;粗糙集;决策表;数据挖掘;不完备信息系统;差别矩阵;约简算法;决策规则;属性重要性;知识约简;约简;属性重要度;条件属性;信息系统;罗杰斯;遗传算法;草鸟;可拓学;特征提取;	自动化技术;计算机软件及计算机应用;	★★★★★	123	17	250	108
10	概念格;数据挖掘;形式背景;	概念格;形式背景;形式概念分析;数据挖掘;关联规则;属性约简;关联	计算机软件及计算	★★★★★	156	19	249	86

		规则挖掘;知识发现;粗糙集理论;算法;内涵缩减;决策规则格;规则提取;形式概念;渐进式算法;粗糙集;赤霉素;协调集;规则;多媒体;	机应用;					
11	数据挖掘;客户关系管理;数据仓库;	数据挖掘;客户关系管理;数据仓库;数据挖掘技术;关联规则;关联规则挖掘;crm;决策树;知识发现;应用;大客户;数据挖掘工具;聚类算法;分类;apriori 算法;支持向量机;算法;客户流失;聚类分析;数据库;	计算机软件及计算机应用;	★★★★★	232	46	365	185
12	数据挖掘;关联规则;数据仓库;	数据挖掘;关联规则;数据挖掘技术;关联规则挖掘;智能检索;数据仓库;网络信息挖掘;知识发现;搜索引擎;数字图书馆;遗传算法;web 挖掘;电子政务;高校图书馆;知识管理;个性化服务;web;web 数据挖掘;apriori 算法;人工智能;	计算机软件及计算机应用;	★★★★★	180	33	299	152
13	商务智能;数据仓库;数据挖掘;	商务智能;数据仓库;数据挖掘;商业智能;元数据;联机分析处理;olap;数据挖掘技术;智能系统;决策支持;erp;数据仓库技术;决策支持系统;客户关系管理;电子证据;电子商务;知识发现;政府管理方式;智能技术;解决方案;	计算机软件及计算机应用;	★★★★★	208	27	213	109
14	供应链风险;数据挖掘;风险管理;	供应链风险;数据挖掘;风险管理;供应链管理;供应链;模糊聚类分析;牛鞭效应;客户关系管	企业经济;	★★★★★	90	41	150	74

		理;危机;风险识别;中国画;科技评估;需求链管理;风险分析;防范措施;风险评估;农产品物流;供应链系统;层流冷却;模糊综合评判;						
15	聚类分析;数据挖掘;聚类算法;	聚类分析;数据挖掘;聚类算法;遗传算法;数据挖掘技术;蚁群算法;信息公开;短期负荷预测;干酪乳杆菌;耐冷性;聚类方法;对应分析;粒子群优化算法;计算机应用;客户细分;行政机构改革;vlsi;开放教育试点工作;网上教学;关联规则;	计算机软件及计算机应用;	★★★★★	70	38	114	58
16	关联规则;数据挖掘;电力市场营销;	数据挖掘;关联规则;关联规则挖掘;apriori 算法;频繁项集;遗传算法;电力市场营销;数据挖掘技术;火电厂;olap;最小支持度;数据仓库;知识发现;电网故障诊断;数据库;算法;频繁项目集;序列模式;频繁模式树;并行算法;	计算机软件及计算机应用;电力工业;	★★★★★	77	37	144	77
17	web 数据挖掘;数据挖掘;电子商务网站;	web 数据挖掘;xml;电子商务网站;数据挖掘;电子商务;web 使用模式挖掘;数据挖掘技术;聚类;搜索引擎;相关技术;应用;数据预处理;客户关系管理;uddi;结构挖掘;孟山都公司;关系数据库;元搜索引擎;使用挖掘;互联网;	计算机软件及计算机应用;互联网技术;	★★★★★	99	18	150	85
18	数据挖掘;电子政务;数据仓	数据挖掘;电子政务;数据仓库;数据挖掘技术;子商务;竞争情报系统;	计算机软件及计算	★★★★	116	82	176	104

	库;	关联规则;决策支持系统;知识管理;人工智能;决策支持;隐私权;电子政务系统;电子商务发展;经营创新;数据库;数据仓库技术;模式识别;应用;crm 整合;	机应用;					
19	数据挖掘;群决策;知识发现;	数据挖掘;群决策;数据挖掘技术;数据仓库;关联规则;知识发现;不完全信息;数据;应用;决策树;kdd;crm;模式识别;人工智能;语言判断矩阵;web 使用模式;联机分析处理;一致性指标;三角模糊数;快速汽门;	计算机软件及计算机应用;	★★★★	130	41	225	96
20	数据挖掘;故障诊断;关联规则;	数据挖掘;关联规则挖掘;关联规则;故障诊断;数据挖掘技术;决策树;气门;apriori 算法;专家系统;人工智能;遗传算法;分类属性;频繁项集;模式识别;凝汽器;电网故障诊断;旋转机械;数据仓库;短期负荷预测;状态监测;	计算机软件及计算机应用;自动化技术;	★★★★	63	33	92	52
21	知识发现;语义网格;数据挖掘;	知识发现;数字图书馆;语义网格;数据挖掘;知识网格;网格服务;数据仓库;面向服务;模糊概念格;网格计算;高校数字图书馆;知识管理;web 服务;语义 web 服务;电子政务;非相关文献;普适计算;网格技术;决策树算法;地理信息系统;	计算机软件及计算机应用;	★★★★	62	28	116	66
22	webgis;arcims;数据挖掘;	webgis;arcims;数据挖掘;森林资源信息;mapxtreme;客户端;森林资源;数据库技术;	计算机软件及计算机应	★★★★	59	4	131	61

		空间数据库;信息发布系统;管理信息系统;地球空间;信息系统;森林资源信息管理;gis;共享;服务器端;商品劳务税;webservices;网络地理信息系统;	用;					
23	遗传算法;自适应光学系统;数据挖掘;	遗传算法;自适应光学系统;数据挖掘;大气湍流;波前传感器;数据挖掘技术;分类算法;关联规则;可变形反射镜;背包问题;林隙;波前校正;旅行商问题;自适应光学技术;复制;变形镜;适应度函数;区间水平;证券投资组合;算子;	自动化技术;物理学;	★★★	78	15	151	54
24	决策树;数据挖掘;决策树算法;	决策树;数据挖掘;神经网络;id3 算法;信息增益;id3;c4;粗糙集;协方差;数据分类;决策树算法;数据挖掘技术;分类器;信息熵;支持向量机;粗糙集理论;属性;熵;归纳学习;旅游开发扶贫;	自动化技术;计算机软件及计算机应用;	★★★	166	39	344	142
25	模糊聚类;聚类分析;数据挖掘;	模糊聚类;聚类分析;数据挖掘;层次聚类算法;聚类算法;加权欧氏距离;类属特征;划拨土地使用权;分类指标;动态直接聚类法;评估;短期负荷预测;干酪乳杆菌;数据对象;模糊模型辨识;ahp;模糊关联规则;模糊聚类方法;遗传算法;综合排序;	计算机软件及计算机应用;	★★★	77	3	152	73
26	竞争情报系统;数据挖掘;竞争情报;	数据挖掘;竞争情报系统;竞争情报;知识管理;数据挖掘技术;数据仓库;竞争情报研究;企业	企业经济;计算机软件及	★★★	148	31	215	121

		信息化;竞争情报工作;企业;电子政务;城市林业;关联规则;企业竞争情报系统;竞争情报活动;企业知识管理;反竞争情报;应用;犯罪信息;知识发现;	计算机应用;					
27	web 日志挖掘;数据预处理;数据挖掘;	web 日志挖掘;数据预处理;数据挖掘;关联规则;用户识别;会话识别;频繁访问路径;序列模式识别;多哈回合;数据预处理方法;模式识别;apriori 算法;用户聚类;频繁访问模式;算法;个性化服务;web 日志;web 挖掘;用户访问;频繁访问页组;	互联网技术;计算机软件及计算机应用;	★★★	71	2	136	57
28	数据库营销;数据挖掘;营销数据库;	数据库营销;网络营销;商函;营销数据库;营销策略;数据挖掘技术;数据库;旅行社;客户关系管理;问题;数据仓库;数据挖掘;消费者;客户数据库;对策;顾客数据库;竞争优势;直复营销;土地权利;策略;	企业经济;	★★★	130	0	102	93
29	时间序列;数据挖掘;相似性搜索;	时间序列;数据挖掘;相似性搜索;动态时间弯曲;复种指数;相似性;小波变换;聚类;ar 模型;算法;多边形边界约简;预测;支持向量机;序列距离;反分析;时间序列数据;形态距离;云发生器;复杂非线性系统;贷款户;	计算机软件及计算机应用;	★★★	76	37	127	62
30	粗糙集;数据挖掘;约简;	粗糙集;粗糙集理论;不完备信息系统;决策表;数据挖掘;约简;属性约	自动化技术;计	★★★	60	13	137	64

		简;知识获取;故障诊断;湿害;决策规则;信息系统;模糊集;信息熵;知识;数据模型;遗传算法;包含度;模糊辨识;模糊聚类;	算机软件及计算机应用;					
31	大客户;客户流失;数据挖掘;	客户流失;大客户;数据挖掘;中国电信;电信企业;电信运营商;数据挖掘技术;客户流失分析;服务业对外开放;运营商;决策树;客户关系;大客户管理;预测模型;客户服务;流失;id3 算法;支持中心;电信业;精准营销;	信息经济与邮政经济;	★★	87	5	86	63

“大数据”2013 年立项课题

序号	项目名称	项目来源	承担单位/负责人	立项年份
1	[在研中] 大数据在政府统计中的应用研究	2013 年国家社科基金年度项目	国家统计局/鲜祖德	2013
2	[在研中] 基于大数据技术的微博问政话题挖掘研究	2013 年国家社科基金年度项目	华南理工大学/王和勇	2013
3	[在研中] 大数据时代全球信息传播格局可视化统计研究	2013 年国家社科基金年度项目	浙江大学/韦路	2013
4	[在研中] 基于大数据的互联网阅读行为模型研究	2013 年国家社科基金年度项目	中国新闻出版研究院/林晓芳	2013
5	[在研中] 大数据时代网络媒介生态环境下个人信息保护体系的构建研究	2013 年国家社科基金年度项目	中央财经大学/章宁	2013
6	[在研中] 基于大数据的个人信用评分建模及违约风险管理研究	2013 年国家社科基金青年项目	清华大学/韩璐	2013
7	[在研中] 大数据的高维变量选择方法及其应用研究	2013 年国家社科基金青年项目	厦门大学/马双鸽	2013
8	[在研中] 基于大数据的产业竞争态势动态预警机制研究	2013 年国家社科基金青年项目	武汉纺织大学/吴金红	2013
9	[在研中] 我国政府部门基于大数据的决策模式研究	2013 年国家社科基金青年项目	燕山大学/迪莉娅	2013
10	[在研中] 推进上海大数据产业发展对策研究	2013 年软件和集成电路发展专项资金课题研究项目	上海软件产业促进中心	2013
11	[在研中] 大数据技术：数据驱动下的警务模式与控制犯罪问题研究	2013 年国家社科基金年度项目	浙江警官职业学院/付艳茹	2013
12	[在研中] 面向海上移动作业的海洋大数据挖掘与仿真重构	2013 年度国家自然科学基金委员会与法国国家科研署绿色信息通信技术领	中国海洋大学;Institut Mines T é l é com – T é l é com Bretagne/陈戈;Ronan Fablet	2013

序号	项目名称	项目来源	承担单位/负责人	立项年份
		域合作研究项目		
13	[在研中] 公安机关大数据管理平台及智能搜索引擎的研发	安徽省 2013 年度科技计划项目	安徽大学	2013
14	[在研中] 基于公安业务的视频大数据挖掘技术研究	安徽省 2013 年度科技计划项目	安徽云端信息技术有限公司	2013
15	[在研中] 基于公安大数据的道路交通安全决策支持关键技术研究与系统开发	安徽省 2013 年度科技计划项目	安徽科力信息产业有限责任公司	2013

主编：刘雁　周莉

编辑：赵冉　杨幸然　陈辰　张春玲　王凯艳　郝晓雪