大数据在市场研究中的国际前沿追踪

国家信息中心 路明

摘要：近些年来，IBM、Google、阿里巴巴、Facebook等国际知名公司都在运用大数据进行市场研究、开拓新技术来发展自己的业务。利用网络平台进行市场调研，利用Hadoop平台分析，更精准的定位客户、预测市场，已成为当今市场研究的前沿技术。

关键词：大数据 市场研究 精准定位客户

1. 大数据在当今市场研究中的重大意义

大数据，是由数量巨大、结构复杂、类型众多的数据构成的数据集合，可通过云计算的处理技术与应用模式，在合理时间内达到选取、管理、处理、并整理成为可以帮助企业经营决策的资讯。大数据具有数量大、实时性、多样性、真实性等特点。

从2009年开始，“大数据”逐渐成为互联网信息技术的热门词汇。美国互联网数据中心指出，互联网上的数据平均每年增长50%，每两年便翻一番，所以目前世界上90%以上的数据是最近几年才产生的。此外，数据不光指人们在互联网上发布的信息，全世界的工业设备、汽车、电表上有着无数的数码传感器，随时测量和传递着有关位置、运动、震动、温度、湿度乃至空气中化学物质的变化，也产生了海量的数据信息。

数据已经渗透到了全世界市场中的各个领域，并逐渐成为重要的组成因素，对海量数据的运用将提升人们对市场以及未来走向的预知和把控。在互联网的时代，数据本身就是资产，而大数据则意味着这些资产正在变得庞大无比，在现实生活中，善于运用大数据的企业已经从数据中获得了丰厚的回报。

例如，华尔街的德温特资本市场公司分析3.4亿个微博账户的留言，判断民众情绪，依据人们在高兴时买股票、焦虑时抛售股票的规律，决定公司股票的买入或卖出。阿里巴巴根据天猫网上中小企业的交易状况筛选出财务健康和讲究诚信的企业，对他们发放无需担保的贷款。目前已放贷300多亿元，坏账率仅0.3%，低于商业银行的平均水平。

在宏观经济预测方面，2008年，阿里巴巴根据对交易数据的分析结果提前预测了金融危机的爆发，帮助一些企业躲过了金融危机的冲击。印第安纳大学利用谷歌公司推出的心情分析工具，从近千万条网民留言中归纳出六种心情，进而对道琼斯工业指数的变化进行预测，准确率高达87%。

在公共事业领域，大数据也在发挥着不可小觑的重要作用。欧洲多个城市通过分析实时采集的交通流量数据，指导司机选择最佳出行路线，从而改善城市交通状况。联合国也推出了名为“全球脉动”的项目，希望利用“大数据”来促进全球经济发展。

“大数据是未来的新石油”，这已经成为社会的共识。

1. 国际知名企业在大数据上的前瞻研究

大数据在市场研究中的精髓是价值，即数据本身可以给企业带来的商业价值，在未来4-5年中，越来越多的公司会像IBM、Google、阿里巴巴、Facebook等技术领先的公司一样，学会如何在大数据集群中运用技术来发展自己的业务。下面我们看几个国际知名企业在大数据在市场研究中的前沿走向。

IBM — 大数据可视化

　IBM近些年来，一直致力于大数据的前瞻研究，有自己的数据管理系统、数据仓库、Hadoop System 、Stream Computing流计算、信息整合管理平台这一整套的大数据技术，其中数据可视化功能是IBM 的一个重要大突破，即分析工具生成的信息以可视化形式呈现在用户面前，给用户一个非常强烈的直觉判断。可视化分析方法，主要包括可视化查询、链接分析、路径分析、群集分析、社会网络分析等分析算法与分析工具。

未来IBM在大数据可视化商业化研发的三大工作方向：1. 从结构化、半结构化、非结构化的音频和视频当中抽取特殊数据，即语义分析和语境分析，来帮助企业进行决策和分析；2. 将分析得出的结果以可视化的方式让业务用户能够理解；3. 可视化与地理位置信息相结合，例如在港口监控所有船只时，一旦出现异常情况，可视化软件可以让工作人员更准确的判断现实情况并得到更快的解决方案。

Facebook — 通过大数据精准定位客户群

Facebook一直是大数据技术最积极的应用者和开拓者，因为它拥有的数据量极其巨大，

目前，Facebook在全球有9亿用户，其中日常活跃用户达5.26亿，每天会采集到500+TB的数据。面对这么大量的数据，Facebook正在分类、提炼，发挥数据真正的价值，根据海量用户的使用习惯做数据挖掘，然后对用户进行“画像”，更精准地把握用户需求和广告主的需求，Facebook大数据技术被广泛应用在广告、新闻源、消息/聊天、搜索、站点安全、特定分析、报告等各个领域。

海量数据处理的第一步就是归类，将用户发表的评论、上传的图片、音乐、视频这些碎片化、非结构化的数据进行分析，使其集结、归类成结构化的数据信息。第二步，是要将这些结构化的数据进行解读，深入到数据背后的潜在意义。每当用户登录Facebook，Cookie会一直驻留在用户的浏览器中，从此用户的浏览行为、浏览页面的关键字会被记录，通过对关键字和上传信息的持续分析，Facebook很容易得出用户的长期爱好和近期需求。再加上对你的朋友圈的分析，可以获得你的教育、工作、收入、地理位置等等诸多方面，这种挖掘和解读往往比个人主动填写的信息还要全面、真实。

为了更便捷、更真实的获得用户的资料，Facebook发布了一款大数据新产品——“时间线”Timeline，它是一个用户可以自我编辑的个人时间轴，在这条时间线的页面记录个人生活故事的应用。Timeline通过帮用户创建个人的时间线和电子传记这一形象化的工具，进行用户数据捕获、存储，将Facebook的数据收集工作带入历史领域。而一旦拥有了这些历史数据，Facebook就如同一个和你从小一起长大的人，对你的档案了如指掌。用户留下的数据越多，Facebook就越了解用户，投放的广告就会更加精准。

Google — 大数据作预测

Google就是大数据时代的开拓者，Google的大数据技术架构一直都是全球互联网企业争相学习和研究的重点，在市场研究中Google所提供的大数据分析主要是客户情绪分析、交易风险分析、产品推荐、消息路由、诊断、客户流失预测、法律文案分类、电子邮件内容过滤、政治倾向预测、物种鉴定等多个方面。据称，大数据已经给Google每天带来2300万美元的收入。具体应用如下：

(1)基于Map Reduce（映射化简模式），Google提供了包括数据存储、数据分析、日志分析、搜索质量以及其他数据分析应用。

(2)基于Dremel系统(在线可视化系统)， Google推出其强大的数据分析软件BigQuery，它也是Google自主开发的一个云数据分析引擎。BigQuery引擎可以快速扫描高达70TB未经压缩处理的数据，并且可马上得到分析结果。这个服务，能帮助企业用户在数秒内完成万亿字节的扫描。

(3)Google的趋势图应用。通过用户对搜索词的关注度，很快的了解社会上的热点是什么。对广告主来说，它的商业价值就是很快的知道现在用户在关心什么，他们应该在什么地方投入一个广告。如“Brand Lift in Adwords”、“Active GRP”等工具，可以帮助广告客户分析和评估其广告活动的效率，再利用Google Analytics，可全面掌控营销投资回报率。

Google的大数据平台架构仍在演进中，追求的目标是更大数据集、更快、更准确的分析和计算，这将进一步引领大数据技术发展的方向。

阿里巴巴 — 大数据下的新C2B模式

阿里巴巴作为中国最大的电商企业，已经通过所掌握的数据以及分析成果，去指导这些生产线的研发、设计、生产、定价。

用户的搜索浏览、驻留时间、商品对比、购物车、下单、评价数据被全程记录，同时用户的个人资料，例如性别、地域、年龄、职业、消费水平、偏好、星座等也已进行画像。这时候阿里巴巴可以对用户进行交叉分析、定点分析、抽样分析、群体分析，基于这些分析结果得出市场需求趋势，再通过地域和时间分析指导生产线不同季节不同物品的产量和不同地域不同产品的库存，适时调整生产、销售、推销策略。

阿里巴巴已经启动了数据共享计划，将它们沉淀的行业数据分享给厂商，从价格分布、关键属性、流量、成交量、消费者评价等维度建模，挖掘出功能卖点、主流价格段分布、消费者需求、增值卖点等来指导厂家的研发、设计、生产。并可以将这种模式复制到更多厂家，让他们去承包生产线，引入更多厂商。

这是一种用户不知不觉参与的C2B模式，可以总结为“大数据定制”。它既帮助厂家更好地满足用户的需求，也有助于帮助厂家减少库存、提升销量。这种C2B模式的C是全网用户，所以就不再需要兴师动众地组织团购，组织投票，组织调研。

未来这种基于大数据的C2B模式将会从小家电扩展到服装、家居以及一些日用品。除承包生产线之外，阿里巴巴还会尝试其他的一些大数据C2B定制模式，例如有偿提供大数据成果或定制服务给一些厂家、其他电商卖家或普通互联网。

3. 大数据时代市场研究的新方式

传统的市场研究虽然以严谨的抽样理论为基础，但由于受制于主持人的访问技巧、街头拦访的不确定性等缺点，不能完全真实反映总体的客观情况。而大数据的调研方法为市场研究人员提供了以“隐形人”身份观察消费者的可能性，超大样本量的统计分析使得研究成果更接近市场的真实状态，同时具有丰富性、实时化、低投入等特点。总结一下，在大数据时代下新的市场研究方法主要有以下两点：

1. 搭建网络平台进行市场调研

无论是IBM、Google、Facebook、还是中国的阿里巴巴，都是基于强大的网络平台来进行大数据的采集、分析、研究、盈利。通过网络平台获取海量的非结构化数据，从大量的音频、视频、搜索记录等数据中抽取特殊数据用来市场的决策和分析，包括影像背后语义分析、语境分析、消费者态度和心理分析，得出的结果通过可视化的方让用户能够理解、分析、判断。这些数据都属于用户主动披露的，与传统访谈形式的被动挖掘相比信息的真实性更高。总之，网络调研具有传统调研无可比拟的便捷性和经济性。

1. 基于云计算的数学分析模型

传统市场调研的采集数据的结构化较好，一般的统计软件SPSS甚至Excel就能满足数据处理过程，而大数据的数据是海量的、非结构化的，需要如Hadoop Map Reduce、Info Sphere Streams等云计算的数据处理技术来实现数据分析，即通过单遍扫描实现对海量数据集的高效聚类，是对AP聚类算法的扩展，通过稀疏化各子集，然后融合各子集稀疏化后的数据再次AP聚类，从而实现对大规模数据的高速聚类。基于云计算的数学分析模型可以将碎片化信息还原为完整的消费过程信息链条，更好地帮助营销人员研究消费行为及消费心理。这些碎片化的信息包括消费者在不同时间、不同地点、不同网络应用上发布的消费信息、购买信息、商品评论信息等，并且通过连续追踪可以形成一个专属的时间链，更加精准的对想得到的结果作趋势性判断。

未来大数据的Hadoop平台将普及，云计算智能化分析的成本低、效率高、收益大的特点促使着IBM、Google、Facebook等这些技术领先的公司不断开拓新的大数据分析软件和模型。

大数据时代新的市场研究方法使“无干扰”真实还原消费过程成为可能，智能化的信息处理技术使低成本、大样本的定量调研成为现实，这将推动消费行为及消费心理研究达到一个新的高度，帮助企业更为精准地捕捉商机。

参考文献：

1. 王云蔚 ，大数据背景下的消费市场研究， 第22卷第1期， 2013.4，北京印刷学院学报
2. 赵刚， 2013-2014年中国互联网行业大数据应用年度研究报告， 2013 10 25，北京赛智时代信息技术咨询有限公司
3. 沈浩，如何实现数据的地理信息呈现，2013 10
4. Joel Gurin，Open Data Now，2014.1
5. <http://www.bigdatas.cn/>
6. <http://www.bigdatas.cn/article-1164-1.html>
7. <http://www.bigdatas.cn/article-906-1.html>
8. <http://www.bigdatas.cn/article-979-1.html>
9. <http://www.bigdatas.cn/article-426-1.html>
10. <http://www.meijiezazhi.com/zt/yw/2012-11-05/11340.html>
11. <http://www.enet.com.cn/article/2013/1127/A20131127334913.shtml>
12. <http://www.ibm.com/developerworks/cn/industry/ind-sc-ibmi2introduction/index.html>
13. <http://www-03.ibm.com/software/products/zh/category/bigdata>

注：本文发表于《汽车工业研究》2015年第8期（总第255期）研究版块。