



数据分析

CHINA DATA ANALYSIS 用数据说话·做理性决策

++ 中国商业联合会数据分析专业委员会 主办 ++



《中国数据分析》会员特刊

2017年第03期 总第31期 (季刊)

咨询热线: 010-59000991 / 59000339

<http://www.chinacpda.org/>

投稿邮箱: xiehui@chinacpda.org



China Data Analysis

/ 数据驱动价值，创新引领未来 /

2017年，大家对“大数据”这个词听过太多遍，似乎没什么新意。但是，如果还只是停留在大数据的基本概念，你就out了，多数的行业对大数据的认知变得理性和客观，更多的企业关注的是大数据的思维、大数据的商业模式、以及大数据的应用，其实这是一种成熟的转变，我们如果就此认为大数据时代已经进入风平浪静的时期，那么就有可能错过一场颠覆三观的革命。

近期苹果召开了2017发布会，一下推出了三款苹果新手机，相信大家的朋友圈围绕Face ID也被各种“槽”点刷了屏，但千万不要只停留在“刷脸”的层面看人脸识别。背后的视频分析、视频人工智能、视频三维识别技术，将重新定义办公、社交、娱乐、教育、医疗、支付六大需求，视频互联网和视频物联网时代有望到来，这样必将产生出更大的数据量，构建有商业价值回报的大数据应用场景。

然而，不管数据的规模如何爆发式的增长，类型多么的复杂多变，如果还只是把数据简单的进行统计分析，那它就不算是有价值的、真正的大数据应用。界定大数据是不是有价值的、真正的大数据，还是有指标可循的。首先，数据的来源是否从多个途径、多个结构的综合；其次，数据一定要脱离简单的统计，要对数据进行真正的预测分析；再则，数据分析一定要融入到每一个业务流程里，在每一个业务流程中进行再应用、再清洗。

大数据的分析应用，要深扎根、细耕耘，深扎根是数据一定要采集全面并对数据足够管控，经过反复的清洗和解析将数据统一基因，充分的挖掘数据内在的核心价值；细耕耘是抛开工具论，充分从数据的分析着手，即要有全局的视角，也要有突出的重点，抓住关键，有的放矢，如果大数据的分析应用还是停留在数据采集、数据分析、决策支持的传统逻辑，就不能发挥大数据分析的价值。从爆发式市场动向，到大数据的价值创造，继而实现“数据货币化”，必由之路是大数据业务应用思路的转换和大数据分析的创新。

只有大数据应用分析的创新，才能引领未来！

中国商业联合会数据分析专业委员会

邵东生

本期目录 CONTENTS

卷首语

- 01 数据驱动价值，创新引领未来

峰会指南

- 03 “分析引领未来”
——第五届中国数据分析行业峰会即将盛大开幕！
- 07 关于CPDA学员参加
“2017中国第五届数据分析行业峰会”邀请函

协会动态

- 08 中国大数据人才培养体系标准即将发布！
- 09 2017年数据分析行业个人会员执业教育工作圆满结束
- 09 第八届全国优秀数据分析师事务所评选开始！

会客厅

- 11 零售行业大数据的挖掘与应用
- 13 手把手教你：如何用数据驱动业务

人才培养

- 15 数据分析方法论在业务实战中的应用心得分享
- 19 “十三五”期间全国快递行业发展建模与预测

数业专攻

- 25 关于流式大数据实时处理技术、平台及应用

运数有道

- 29 大数据告诉你，近十年来苹果中国消费者变迁
- 31 大数据分析：《战狼2》火爆荧屏背后的必然

事务所风采

- 34 贵州中科项目数据分析师事务所



主办单位

中国商业联合会数据分析专业委员会

编委成员

周燕、李缘

出版时间

2017年第3期 总第31期 10月出版

美工设计

崔峻珩

联系我们

中国商业联合会数据分析专业委员会

地址：北京市朝阳区朝外soho C座9层

电话：010-59000991 / 010-59000339

传真：010-59000991转 607

投稿

欢迎广大读者踊跃投稿，内容包括学术观点、教学体验、教学活动、学习感悟、实战经验、随笔文章等。稿件附图格式为JPG或TIFF格式，大于1M，分辨率在300dpi以上。

感谢您对《中国数据分析》的支持！

投稿邮箱：xiehui@chinacpda.org



/ “分析引领未来” ——

第五届中国数据分析行业峰会即将盛大开幕！ /

文 / 协会市场处 张楠 编辑 / 协会会员处 周燕 日期 / 2017-09

随着大数据全球战略布局的全面升级，大数据发展从概念推广进入到全面落地的转折期。2017年，对大数据的讨论不再停留于宽泛的概念，技术创新的热忱也在削减，随着大数据技术寒冬的来临，数据应用的春天指日可待。如何让大数据变成真正的生产力，让“数据驱动价值”的理念变成现实，成为各行业急需思考和探讨的核心。

被业界誉为“知识和实战盛宴”的中国数据分析行业峰会，自2010年开始，至今已成功举办四届，得到了社会各界的聚焦和广泛关注。本届峰会在传承往届宗旨的同时，又提出了新的行业发展方向——“分析引领未来”。

届时，来自零售、金融、互联网、消费、教育、医疗等诸

多行业的一线大数据实战派行家将齐聚一堂，共同探讨大数据产业发展及大数据人才储备等的前沿话题，分享企业真实“落地”案例，还原大数据应用场景，探索各行业之间面临的数据分析应用问题的解决途径。

一、峰会概况

举办时间： 2017年11月18日-19日（9:00-17:00）
会议地点： 广东亚洲国际大酒店
（广州市越秀区环市东路326号之一）
参会人数： 预计超过1000人
主办单位： 中国商业联合会
中国商业联合会数据分析专业委员会

协办单位： 北京中盈达科技有限责任公司、
北京犀数科技有限公司、
广州数据场科技有限公司
支持媒体： CCTV、广东电视台、环球网、新华网、
腾讯网、搜狐网、千龙网、和讯、广州日报、
羊城晚报、数据圈

二、峰会亮点

1、主题演讲全部以实战案例为主

汇集来自零售、金融、互联网、消费、教育、医疗等诸多行业一线大数据实战家，分享所在行业的一手实践案例，真实展示大数据在业务流程中的应用，让更多参会者亲眼见证同行企业应用大数据变现的效果。真正使得本届行业峰会成为一场“知识与实战”的顶级盛宴。

2、演讲嘉宾全部来自企业一线实战派领军人士

演讲嘉宾来源于企业大数据应用的一线人员，有数据部门的负责人、数据运营官、首席数据分析师、企业高管…同行之

间既可现场进行“技艺”切磋，又可结识更多业内精英，形成大数据行业的专业交流圈。

3、峰会主题鲜明：分析才是数据产生价值的核心

当大数据概念炒作回归常态的时候，数据驱动价值的认知成为各企业关注的新热点。IDC预测，企业基于大数据分析的商机高达5000亿美元以上。由此可见，大数据的核心价值在于分析。因此，如何实现对数据的充分利用，成为本届峰会的重要议题。通过一线大数据实战派的案例分享、热点话题的巅峰探讨，共同寻找大数据产业的未来发展之路。

三、特色活动

1、百家企业战略合作公益项目广州行

“35.1%以上的企业已经开始应用大数据实现业绩增长”，这个过程中，企业面临着数据平台搭建费用高、算法不精准、数据与经营决策衔接不到位，以及大数据分析人才紧缺等问题。针对上述问题，协会以扶持行业发展的公益角度出发，在对企业数据化程度进行评估的前提下，从数据平台搭建、数据部门构建、数据分析人才培养等方面，为亟待转型或战略升级的中国企业提供大数据应用解决方案。

2、论“数”有道巅峰对话

大数据未来的发展方向是什么？数据分析如何引领企业实现商业变现？不同行业的专家，对社会上大数据的热点话题又会碰撞出什么样的火花？这一切都将在主会场及分论坛的巅峰对话环节得到答案，相信这会是一次汇聚各家思想精华的顶级“对决”。

3、“聚焦应用”行业专题论坛

从战略高度到具体操作环节、从前端销售、市场到后端

生产、管理，数据分析存在于企业的各个部门以及所有运营环节，在描述、决策、预测方面发挥着重要作用。

通过全方面、多角度展示企业成功应用案例，为与会者提供行业榜样。一场峰会，领略各个行业的精彩应用，真实感受数据分析带来的价值。

4、大数据人才群英汇

2016-2017年间，35所高校获批“数据科学与大数据专业”，这预示着中国大数据人才进入了一个新的发展阶段，大数据人才要从理论、实战、应用层面提升专业能力。而只有对大数据人才的重视，才是助推企业数据化进程的关键。本届峰会将针对大数据人才培养进行深入探讨，更有由协会牵头，联合全国百余位大数据专家、数据分析师事务所、国内外知名企业参与评审的首个《中国大数据人才培养体系标准》现场发布。真正为大数据从业者、企业、人才服务中介、高校搭建线上线下交流平台，成为中国数据分析师的行业盛会。

四、精彩议题

主题方向	议题范围
峰会主会场 （ 11 月 18 日 9：00-12：00 ）	
战略发展	分析引领未来——大数据行业 2017 年发展新趋势
热点议题	大数据如何解决行业挑战？ 如何规避大数据应用中的误区？ 万物互联后的下一个目标
实战案例	一个行业案例，带你看透数据分析真相 大数据如何驱动商业裂变？ 大数据在传统企业产业链中的应用 电商平台数据分析场景和方法探研
高峰对话	论“数”有道：数据分析如何引领企业实现商业变现？
峰会分论坛 （ 11月18日 13:30-17:00 ）	
论坛 1 大数据在行业中的应用场景	金融、电商、医疗、教育、商业等领域的应用案例展示 主题对话：大数据如何在企业不同发展阶段发挥最大作用？
论坛 2 传统企业战略转型	传统企业转型的最佳切入点 大数据在服务行业中的融合应用 零售业在转型过程中的难点及解决方案 主题对话： 不同阶段传统企业如何引入大数据
论坛 3 数据分析师专场	没有基础，如何成为数据分析师？ 大数据未来，人人都是数据分析师 大数据时代催热大数据专业 大数据专业毕业后的一公里 主题对话：优秀的数据分析师应该具备哪些标准？
论坛4 大数据与区块链	大数据与区块链的关系和区别 区块链助力破解大数据难题 主题对话：区域链如何在大数据领域发挥作用？
论坛 5 大数据与信息安全	大数据安全任重而道远 数据开放与隐私保护 主题对话：共享经济，如何保证信息安全？
中国数据分析行业内部会议 （ 11月19日 09:30-17:00 ）	
行业发展	中国数据分析行业未来发展趋势 大数据人才的道与术 数据分析师未来发展空间有多大？
最新技能	做个与时俱进的数据分析师——大数据人才必备新技能 年薪百万的数据分析师是怎样养成的 数据分析师的四大决胜能力
案例分享	优秀授权中心展示 优秀数据分析师事务所展示
交流研讨	授权中心、数据分析师事务所、行业主管协会共话发展！

（备注：会议内容以现场公布为准，请登录www.chinacpda.org或随时来电了解会议动态最新日程。）

五、部分拟邀嘉宾（排名不分先后）

姓 名	职 位
邹东生	中国商业联合会数据分析专业委员会 会长
田日辉	联想云数据服务业务集团 副总裁
傅志华	360 公司数据中心 副总经理
沈志勇	百度云计算事业部 首席数据科学家
戎鹏	甲骨文股份有限公司 北京区经理、数据科学家
单艺	猎聘网 首席数据官
王兴海	深圳雷柏科技 副总裁、数据科学家
许建真	中原集团数据中心 总监、首席分析官
蒋杰	腾讯数据中心 总经理、首席分析师
冯一村	海云数据 创始人兼 CEO
鲍忠铁	北京腾云天下科技有限公司 首席金融分析官
段林峰	QUEST MOBILE 副总裁
陈毓洁	硕文拓管理信息科技(上海)有限公司 总经理
李宁	携程机票 资深数据&产品经理

（备注：会议内容以现场公布为准，请登录www.chinacpda.org或随时来电了解会议动态最新日程。）



2017中国大数据与分析创新大会
暨第五届中国数据分析行业峰会，
一场知识和实战的盛宴，大门即将开启！期待您的参与！

/ 关于CPDA学员参加 “2017中国第五届数据分析行业峰会”邀请函 /

文 / 协会市场处 冯雪 编辑 / 协会会员处 李缘 日期 / 2017-09



全国CPDA学员：

本次峰会由中国商业联合会、中国商业联合会数据分析专业委员会主办,预期观众1000多人次,多家专业大众媒体将会对此次峰会进行全方位多层次的报道。CPDA学员作为行业发展的奠基者,无数学员为行业发展作出了重要贡献,本届行业峰会我们邀请全国的数据分析师参加,共享行业盛宴。

对于参加峰会的学员,将享有以下安排：

针对大数据人才培养方面,从主会场到分论坛均设置精彩演讲！

主题涵盖数据分析师职业发展方向、数据分析人才应该具备的能力以及知识框架、数据分析师在企业中的价值,以及优秀数据分析师的修炼等几方面内容。

实战派嘉宾手把手教你如何成为优秀数据分析师！

演讲嘉宾均来自企业大数据应用的一线人员,有数据部门

的负责人、数据运营官、首席数据分析师、企业高管,既有同行之间现场“技艺”切磋,又可结识更多业内精英。

专为你设立的数据分析师专场！

身为数据分析师,一定想要了解行业的从业方向、职业发展前景和趋势。会议现场为大家解读关于人才的走向、企业的招聘需求,获得职场转型或者提升的机会；

CPDA学员参加行业峰会的,可获得一次免费复训机会,有效期自峰会之日起一年！

CPDA学员证书未年检的,参加本次峰会,即可获得免费证书激活一次！

CPDA学员免费获得精美礼品！

FIN.



/ 中国大数据人才培养体系标准即将发布！ /

编辑 / 协会会员处 周燕 日期 / 2017-09

职业社交平台LinkedIn发布的《2016年中国互联网最热职位人才库报告》,其中作为四大核心发现之一的人才需求与供给方面,指出数据分析人才最为紧缺。伴随着大数据在互联网行业更多的应用,尤其是数据分析人才愈加重要。站在互联网风口上的大数据,直接催热了大学里的大数据专业。

在国外,美国麻省理工学院、哥伦比亚大学、加州大学圣地亚哥分校、普渡大学等十几所名校都围绕大数据开设了分析类专业,在师资建设和人才培养方面行动更早、学科建设也更完善。与国外高校相比,虽然从2016年至今,我国共有35所高校成功申请设立了大数据专业,但开设还显不足,需要加快建设步伐,迎头赶上。

中国商业联合会数据分析专业委员会为此进行了预测,到2018年,大数据或数据工作者的岗位需求将激增,大数据专业人才的缺口将达到180万,巨大的潜在需求,迫切的需要人才的填补。不管是从时间还是经验看,目前国内培养大数据人才的院校都尚处于起步阶段,学校教育 with 大数据市场需求脱节严重,也正因为此,社会培训机构对大数据人才的培养所起的作用不容忽视。

中国商业联合会数据分析专业委员会作为主管行业协会,通过分析目前大数据培训课程设置,发现培训机构对大数据人才的培养存在严重的方向偏移。今年8月份开始,由中国商业联合会数据分析专业委员会牵头,联合全国百余位大数据专家、数据分析师事务所、国内外知名企业参与评审的首个《中国大数据人才培养体系标准》将在今年11月18日第五届中国数据分析行业峰会上公布！

《中国大数据人才培养体系标准》的确立,在行业方面,弥补了大数据人才市场标准缺失的空白；在个人方面,帮助大数据人才找到行业位置、发展方向；在企业方面,助力企业培养大数据人才,搭建大数据部门；对于猎头,提升人才和企业需求匹配度。

《中国大数据人才培养体系标准》从四个方面进行了确立,概要如下：

一、大数据及大数据人才介绍

二、大数据发展与大数据人才培养现状

（一）大数据发展进程

（二）大数据人才高校培养现状

（三）大数据人才社会培训现状

三、大数据人才知识体系架构

（一）从实践应用需求中构建出的大数据人才知识体系

（二）案例：旅游大数据人才知识体系架构的应用落地

四、大数据人才培养战略

（一）指导思想

（二）基本原则

（三）未来五年战略目标

相信随着《中国大数据人才培养体系标准》发布,未来大数据人才的培养将逐渐符合大数据市场的需求！

FIN.

/ 2017年数据分析行业个人会员执业教育工作圆满结束 /

编辑 / 协会会员处 李缘 日期 / 2017-07

根据《会员管理办法》和《会员执业教育管理办法》的相关规定，鉴于个人会员的从业需求和行业发展的实际情况，协会于2017年7月1日面向全国所有个人会员举办了“2017年数据分析行业会员执业教育”。

2017年数据分析行业个人会员执业教育安排如下：

开展时间：2017年7月1日——7月31日

参加对象：全国范围内数据分析行业个人会员

举办形式：网络远程课程（扫描二维码进入）

教育内容：（分两期进行）

1、第一期

课程名：传媒大数据技术应用

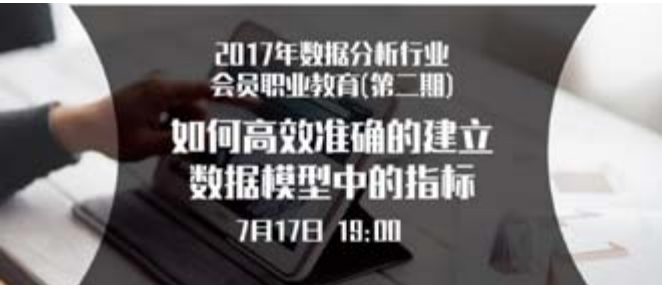
主讲人：王鑫



2、第二期

课程名：如何高效准确的建立数据模型中的指标

主讲人：李军



针对每期讲解的内容设置了相应的习题，供会员练习。

总体，个人会员对此次执业教育活动有一定的重视，在繁忙的工作同时认真学习执业教育的课程，并通过课后习题的测试。至此，2017年中国数据分析行业个人会员执业教育工作圆满结束！ **FIN.**

/ 第八届全国优秀数据分析师事务所评选开始！ /

编辑 / 协会会员处 周燕 日期 / 2017-09

“优秀数据分析师事务所评选活动”是由中国商业联合会数据分析专业委员会主办，每年举行一次，是协会重点活动之一。在通过了解和掌握各地事务所的市场环境和经营现状，同时有针对性的进行行业规范和指导，意促进事务所的发展，加强协会与事务所，事务所之间的互动交流。

本届活动为协会第八次举办。回顾历届情况，既有众望所

归的事务所，也有异军突起的事务所，每年的活动都会有亮点和黑马出现。

随着大数据环境的越发火热，事务所在整体运营，业务类型及报告质量等方面也在不断改变中，出现了经营决策业务比例升高，企业合作更加紧密，涉及领域越来越广等现象，事务所在引领大数据行业发展方面起着不可或缺的作用。



一、参评对象

此次评选针对在我会在册备案全国范围内的所有会员单位

二、评选指标

- 1、事务所运营状态
- 2、行业宣传推广、活动参与度
- 3、遵守行业准则，尊重且执行从业规范
- 4、人才培养及团队建设情况

三、评选流程

评选时间	评选进程
09.18-09.22	事务所提交参选文件、评优活动答疑
09.27-10.10	微信平台公开投票
10.11-10.17	评审小组根据投票结果进行综合评审
10.18	公布评审结果

四、参评方式

- 1、通过协会官网www.chinacpda.org下载参评文件：
附件一：《2017年事务所经营情况调研问卷》
附件二：《2017年度事务所项目承接表》
- 2、事务所相关电子版照片若干（格式：jpg/png、Logo大小：300像素*300像素，容量<=1M）
- 3、联系方式：
邮箱：xiehui@chinacpda.org
电话：010-59000339转652，李老师



“数据分析师事务所”作为由数据分析师组成的从事数据分析服务的专业中介机构，接受中国商业联合会数据分析专业委员会的监管，具有统一的服务标准规范，是中国数据分析行业走向规范和自律的中流砥柱和促进行业健康发展的中坚力量！

“数据分析师事务所”服务范围：

- 企业数据化服务
- 量化咨询服务
- 算法优化和改进服务
- 常规性数据服务
- 数据定制化服务

期待更多有能力有梦想的数据分析师踏上创业之路，共同为行业的发展做出更大的贡献！

FIN.



/ 零售行业大数据的挖掘与应用 /

文 / CPDA学员 吴向华 编辑 / 协会会员处 李缘 日期 / 2017-07

过去流传过一句戏言：网店、网店、网店，这是零售行业取得成功所需的三个最重要的条件。如今，变成了“数据、数据、数据”不难看出，未来的零售行业一定是数据驱动型企业。零售行业产生的海量数据，通过挖掘、整合、分析，必定会显现强大的商业价值。大数据如何为传统零售行业赋能？协会特别邀请CPDA学员吴向华先生，和大家一起探讨大数据如何找到落地业务场景。将从数据获取、数据预处理、相关性处理、数据验证、聚类分析、租金预测等六大部分，和大家分享门店租金预测实际场景案例分解，以及简易的资金管理系统搭建思路等。



主讲人：吴向华
(CPDA学员，北京航空航天大学软件工程硕士)
项目一级管理师，熟悉各主流数据库，精通EXCEL VBA。

从主数据管理、数据支持到数据分析，管理数据的全生命周期。现专注于数据分析思维在传统企业落地，寻找数据分析应用业务场景。

很高兴跟大家探讨关于寻找数据分析应用场景这个主题。现在，无论是普通

中小型企业还是大型互联网企业，甚至是国家政府部门以及事业单位也经常用到数据分析，证明数据分析发展的趋势非常好。

应用歧途	● 唯理论化 ● 唯工具化
悟道	● 舍得 ● 不唯因果
准备	● 培养数据驱动的意识 ● 将线下业务变为线上业务 ● 尝试新的数据采集手段
案例	● 门店绩效因子分析 ● 简易资金管理系统构建

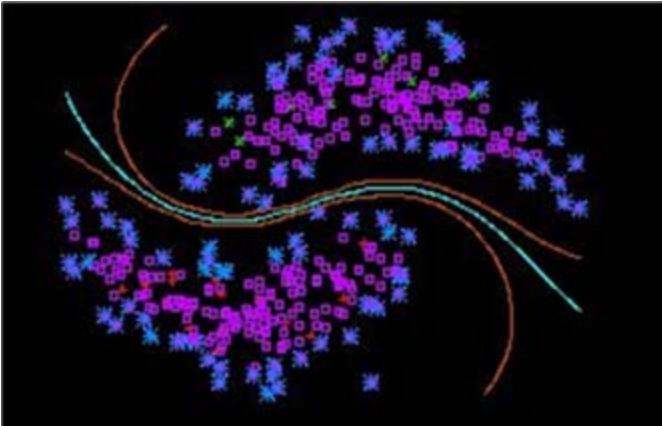
我从事的零售行业数据分析相关的工作，就以传统企业数据分析部门为出发点，讲讲我们是如何把数据分析相关的知识运用到我们实际的工作场景中。不缺大数据、不缺算法，我们缺少的就是如何将数据分析落地，在各种业务场景中灵活应用。

在实际应用之前我也找了很多的资料，很多代码等等，从统计学理解，从各种算法理解，总觉得离我们实际应用角度“相差甚远”。

一、应用误区

现在市面关于大数据应用，理论化内容比较多，偏“高冷”。通常那些课程、书籍只告诉你算法是怎么来的，解决关联度分析（像啤酒+尿布），但是这些离企业很远，如果我在企业中遇到实际问题，我应该怎么用？这个挺难的。我个人认为，术业有专攻，算法研究还是交给数据科学家来处理，但是我们要会用，理解。

做预算管理工作的時候，我們的系統會有很多報表、明細的支持，對於預算管理，各個業務部門可以看到明細，但對於費用報銷、各項支出審批的時候，沒有整體直觀的概念。所以，我們做了一個跨平台的数据报表自动生成工具，方便各部门负责人。



二、误道

数据分析是一种思维方式，要树立正确的决策观。现在企业真正的核心就是怎样做精细化管理，以前管理比较定性、粗放；现在任何一个新生的企业都要有一套系统，有很多的数据支持。像互联网企业如果没有一定的数据支持，是无法生存的。

传统企业如果做精细化市场划分，精准营销加管理，没有精确的数据分析，效果不见得有多好，所以，归根结底任何一家企业都要进行数据化精确管理，数据驱动意识。

过去，数据是精简主义，只保留当时认为最有用的信息；现在，每个数据都有意义，无法取舍。以前，因为……所以……，科学界普遍追求因果关系检验；现在，数据量庞大，因素并不单一，在复杂的情况下，如何做分析？以相关性分析为基础，真实的数据往往都是不完美的，一堆概率事件交织作

用后形成可能的结果。数据分析的应用场景需要去寻找对原有工作颠覆性创新，知行合一。

三、准备

将线下业务变为线上业务，大数据分析的基础是系统化。尝试新的数据采集手段，跨平台数据库工具；抓取的第三方数据。

四、案例

门店租金预算编制预测

根据现有门店及2018年新开店计划和销售目标，合理测算租金支出

1）数据获取：租金条款手工录入

固定租金

保底租金

抽成租金

根据商品品类适用不同的扣率

2）数据预处理：现有租金条款系统化

延续2018年底，缺失值处理，按每年增长幅度推算

3）验证2017年租金决算数据

找到差异修正

4）相关性处理

门店信息包含：店面属性、品牌、经营模式、店铺形式、所属大区，销售城市、城市等级、开业时间、闭店时间、门店面积、门店级别、门店用途等。

5）门店聚类分析

K-means

交叉验证

6）新开店面租金预测

FIN.

/ 手把手教你：如何用数据驱动业务 /

文 / CPDA学员 鉴真 编辑 / 协会会员处 李缘 日期 / 2017-08



呈现真实工作场景，你手里的大数据不该是数据本身，而是如何利用数据驱动业务价值。22年IT从业经验，4年数据分析行业经验，一定会带给你不一样的视角……

数据驱动是什么？大数据驱动又是什么？是管理、业务、运营、模式、产品、市场、营销……有的教材还提到变革、优化、革命话题就更大……今天主要跟大家聊一聊管理层面的数据驱动。

跟数据驱动相对应的话题就是流程驱动，相信在IT工作经验在10-20年以上的朋友会对MRP这个单词很熟悉，它实际是ERP的前身，做ERP部分的时候强调最多的是流程驱动，而流程驱动和数据驱动是相互补充。

我们在做流程驱动的时候，要做产供销，然后是人、财、物的管理，从产品的设计开始到工艺设计制造，再到销售，整个链条从头做到底。有人会问，我们讲的数据驱动是强调数据，而流程驱动是不是不强调数据？这是不对的，流程驱动更强调数据，只不过，我们的数据是来自产品设计模块的输出，

是产品工艺设计模块的输入，产品工艺设计模块的输出又是产品制造模块的输入。

所以，最后要形成数据的时候，是从头到尾得到的数据报告。这样，我们在改任何数据的时候，除非是从头到尾，再形成结果。

数据驱动则不同，数据在整个流程中，是形成报表，分析，再进行报表，再进行行动，再测价值，再到数据。

所以，总结一般数据驱动和流程驱动的相同点：都有流程和数据、都强调标准化/数据量化、都要求数据正确；不同点：侧重点不同、数据采集点不同、数据影响面不同、数据结果显示不同、管理者要求不同、采取的行动点不同、灵活度不同。

数据量化，不管在做流程驱动还是数据驱动，我们在数据量化时候都会有非常大的观念，我们要做数据的准确性。做数

据准确性需要说明一点，在流程驱动时候，我们大部分服务的对象，实际上对于我们的设备和软件，而在数据驱动的时候，我们的服务对象大部分都是以人为主。

所以，我们在说数据量化模型的时候，都是以锥形模型来阐述数据量化的可能性。把数据服务的人分成三个部分，高层管理者、中层管理者、基层管理者。

对于高层管理者看重的是结果，中层管理者看重的是过程，而基层管理者看中的是操作。而我们每一个层面都要对上一个层面的管理者负责。

形成锥形模型的结果就是因为高层管理者看重市占率、利润率，中层管理者看重转化率、人才流失率、利用率、效率……，而基层管理者看中的是具体量化数据……

数据驱动的基础，我认为是管理者最重要的一环。首先我们要改变数据考核和激励机制，如果这两方面不改，数据驱动几乎很难进行。数据驱动当中的考核机制大部分公司做的是OKR，对应型是KPI。这两个考核机制有很大的区别，在OKR考核的时候注重阶段性目标，因为这两个机制单词中最重要的是O，他对目标很明确。OKR在考核过程中非常注重沟通关系，而KPI可能就是一种强制性的机制。

我们现在也在批评KPI没有人性化的管理，OKR就比较偏人性化。把这两种机制做到我的数据驱动和流程驱动的时候，个人感受，KPI更适合流程驱动的公司（这里并不是说流程驱动不好，各有各的好处），KPI从制定到考核也是比较成功的，只不过随着互联网发展，我们可能数据更驱动OKR这种模式。这就是我们把这两种机制进行比较。

我们在描述数据的时候，一定在脑子里要有维度的概念。

我们数据驱动，驱动什么？

- 1) 了解过去发什么了什么？
- 2) 当前正在发生什么？
- 3) 明确各个时间/问题发生的原因
- 4) 预测未来可能发生什么事——调整
- 5) 惊醒全局优化。

这就是数据驱动我们简单要做的工作。

数据驱动和数据分析的关系是什么？

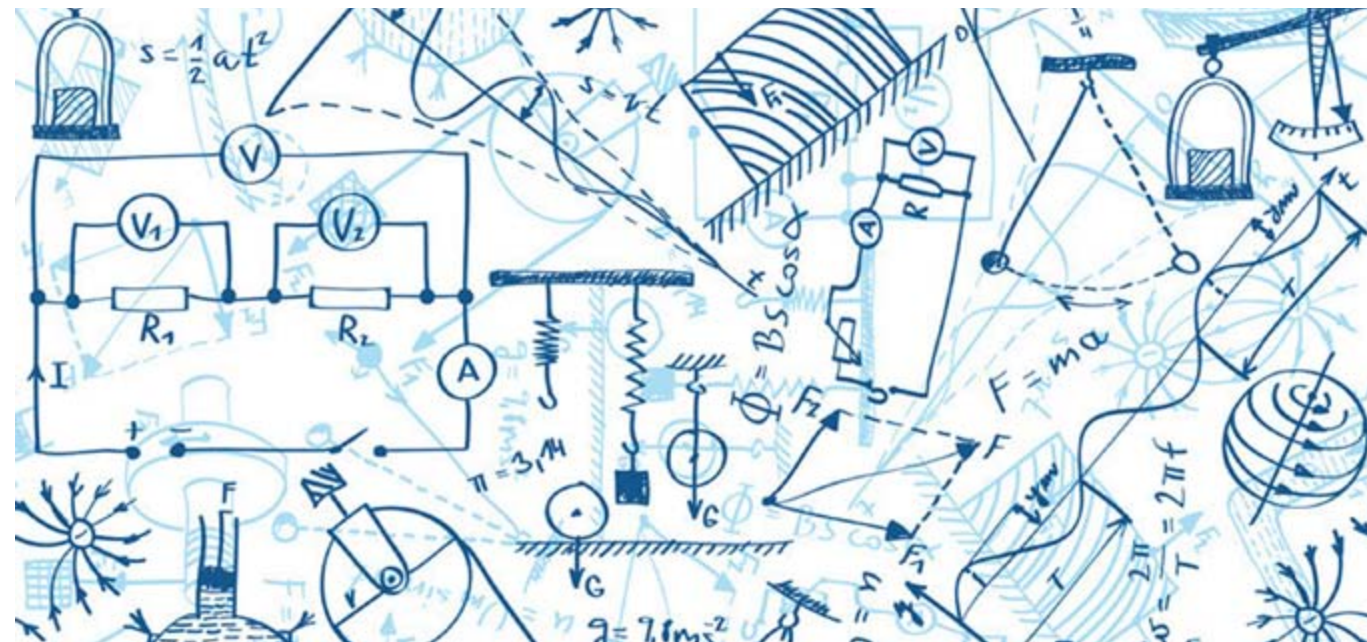
其实，数据驱动依赖数据分析，所以数据驱动的帮手是数据分析。

之后，许老师结合自身公司实例，分别讲解了数据分析的数据域如何划分；主数据和元数据的管理；主数据、元数据的关系；元数据分类；业务元数据；客户主数据划分；KANO模型；行为分析；数据可视化……并进行了非交互性&交互性数据展示、个性化数据展示、数据的可用性分析、利用可视化工具进行个性化分析……

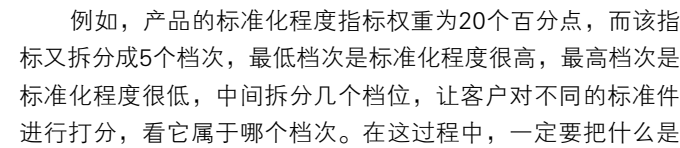
总结这次的分享：

1. 数据驱动和流程驱动的不同
2. 考核绩效是数据驱动的指南
3. 数据整理是数据驱动基础
4. 数据可视化是数据驱动的助手

FIN.



文 / CPDA学员 刘程浩 编辑 / 协会会员处 李缘 日期 / 2017-06



4) 一致性检验。检验比较矩阵的



- 18 -

/ “十三五” 期间全国快递行业发展建模与预测 /

文 / CPDA学员 刘嘉 编辑 / 协会会员处 李缘 日期 / 2017-06



“桃花潭水深千尺，不及汪伦送好评”、“洛阳亲友如相问，就说我在等快递”、“一骑红尘妃子笑，无人知是快递来”、“劝君更尽一杯酒、西出阳光不包邮”……古诗完全快递范儿了，快递业第一大国岂是浪得虚名。飞速发展的快递业将何去何从？

协会特邀北京邮政科学研究院及设计院软科学研究中心刘嘉先生，带你一起预见2020年的快递行业。



主讲人：刘嘉

（CPDA学员，北京工业大学计算机软件学院工程硕士）
从事邮政企业的规划，标准（规范）编制，可研和数据分
析，建模倡导“业务是核心，数据是驱动”。

我们已经学了许多CPDA的知识、技术，自己企业中也有些数据和工具，那么以上这些怎么更好的为业务驱动服务？又怎样预测行业的发展？

一、选题与综述

2013年，全国快递业务量首次超过函件业务量。2014年，中国经济进入新常态，经济增长从高速增长转为中高速增长，全国快递量增幅也有所下滑，但仍超过140亿件，首次超过美国，跃居世界第一，成为中国经济的一匹黑马。2015年，全国快递业务量首次突破200亿件大关，继续稳居世界第一，中国快递业在加速流通、扩大内需、调整结构、吸纳就业、普惠民生等方面的基础性作用日益显现。

国家高度关注快递行业发展。2016年，全国快递业务量突破300亿件，全球占比超过40%，日均服务超过2.5亿人次，支撑网络零售额超过4万亿元，占社会消费品零售总额比重达到12.5%，新增就业20万人以上。2017-2020年，中国快递业，将越来越与中国经济相互依存、相互影响、共同发展，经济结构进一步调整与优化，作为‘新经济’的代表，快递业未来的发展将达到什么水平？-业务量、收入的预测问题。

1.1 快递行业规划中的发展目标

规划的核心：规划，是个人或组织指定的比较全面长远的发展计划，是对未来整体性、长期性、基本性问题的思考和考量，设计未来整套行动的方案。主要包括：指导思想、基本思路、发展目标（业务量、业务收入）；其次是：主要任务和保障措施。

2012年，有人利用平稳时间序列（ARIMA），根据全球快递四大家族在10年间（2001年~2010年）各季度业务量，预测2011年~2013年全球各季度快递业务量。2014年，有人利用微分方程GM(1,1)——一阶一个变量的灰微分方程模型，分析2008-2013年Q2各季度全国快递收入数据，分别建立季度收入模型，预测2013年~2017年全国各季度快递收入。2016年，有人利用季节性差分自回归滑动平均模型（SARIMA），分析2003年~2014年各月全国快递业务量数据，预测2015年1月~4月全国快递业务量。

以上模型，仅考虑快递业务本身的历史数据，在快递业务量、业务收入的周期性（月、季）发展中，短期预测效果好，而对两年以外的预测，则误差越来越大。

对于快递业务中长期的发展（战略管理），不仅要考虑快递业务本身的历史数据，还应考虑其发展的诱因及这些诱因的发展变化趋势，更要考虑国家宏观政策对快递业发展的引导、国民经济发展的趋势，以及社会环境、技术环境的发展变化所带来的各种影响。

二、思路与方法

问题分析？——用已知的知识、方法、技术，解决未知的问题。

“快递业作为服务业的重要组成部分，是推动流通方式转型、促进消费升级的现代化先导产业。”——《国务院关于促进快递业发展的若干意见》。截止2017年6月，中国网民7.51亿，占全球网民的1/5，手机网民占96.3%。截止2017年6月，中国网购用户5.14亿，手机网购用户4.80亿。互联网+消费（网购），推动快递业的发展。

网购-零售：“5294互联网零售，指不包括在网络销售中，仅提供网络支付的活动，以及仅建立或提供网络交易平台和接入的活动”，属于“52零售业”。

快递包裹-网购：2009年，超过50%的快递包裹来自网购；2013年，超过60%；2015年，已接近70%。网络零售逐步成为推动我国快递发展的重要动力。

快递服务-第三产业：“6020快递服务，指在承诺的时限内快速完成的寄递服务”，属于“60邮政业”，是第三产业。

综上，快递行业是服务业，属于第三产业，其行业发展与第三产业的发展必然存在一定的关系；同时，快递行业也与网购、零售业的发展存在密切的联系。因此，研究思路确定如下：

1）采集数据：从国家统计局、国家邮政局及专业资讯网站统计宏观经济和快递行业的相关数据；2）分析数据；对这些

数据，进行相关分析，确定建模的最终数据；3）建模、检验及评估；对选定的数据，进行回归分析，建立模型并检验。经检验通过后，并比对实际收入，根据误差、选取最佳模型；4）预测未来；根据PEST分析，提出合理假设，利用选定的模型，预测2017-2020年全国快递行业业务量、业务收入的区间。这里用到了CPDA知识中的市场需求分析、相关分析、回归分析、外部环境PEST分析。

三、数据采集与分析

采集什么数据？去哪采集？怎么分析？

3.1 数据采集

年度	第三产业增加值 (万亿元)	社会零售 品总额 (万亿元)	网购交易 规模 (万亿元)	快递业务量 (亿件)	快递业务收入 (亿元)
2005 年	7.49	6.84	0.02	8.66	239.70
2006 年	8.86	7.91	0.03	10.60	299.70
2007 年	11.14	9.36	0.05	12.02	342.60
2008 年	13.13	11.48	0.12	15.13	408.40
2009 年	14.80	13.30	0.25	18.58	479.00
2010 年	17.36	15.80	0.46	23.40	574.60
2011 年	20.52	18.72	0.78	36.70	758.00
2012 年	23.19	21.44	1.19	56.90	1055.30
2013 年	26.22	24.28	1.89	91.90	1441.70
2014 年	30.67	27.19	2.79	139.60	2045.40
2015 年	34.16	30.10	3.78	206.70	2769.60

（数据来源：国家统计局、国家邮政局、艾瑞咨询）

3.2 数据分析

2005-2015年，第三产业增加值年平均增长16.4%；

2005-2015年，社会零售品总额年平均增长16.0%；

2005-2015年，网购交易规模年平均增长68.9%；

2005-2015年，，快递业务量年均增长37.3%，快递收入年均增长27.7%。

第三产业增加值、社会零售品总额、网络购物市场交易规模、全国快递业务量、全国快递业务收入，均不服从正态分布。在10年间持续增长，趋势不变。对不服从正态分布的数据，应使用秩相关来描述其关联性。

因此，对快递行业的业务量、业务收入的影响因素分析，均采用秩相关分析，常用的有Kendall（肯德尔）分析和Spearman（斯皮尔曼）分析。在数据量较小时，斯皮尔曼的秩相关系数最为可靠。

综上，自变量之间两两显著相关，但从业务角度出发考虑快递行业与网购市场的天然联系，远远超过第三产业增加值、

社会零售品总额。因此，最终确定两个模型的自变量，均为网购交易规模。

相关性		Y1：全国快递业务量（亿件）	X1：第三产业增加值（亿元）	Y2：社会零售品总额（万亿元）	X2：全国网购交易规模（万亿元）
前建立模型	Y1：全国快递业务量（亿件）	相关系数 显著性（双尾）	1.000 11	1.000 11	1.000 11
	X1：第三产业增加值（亿元）	相关系数 显著性（双尾）	1.000 11	1.000 11	1.000 11
	Y2：社会零售品总额（万亿元）	相关系数 显著性（双尾）	1.000 11	1.000 11	1.000 11
	X2：全国网购交易规模（万亿元）	相关系数 显著性（双尾）	1.000 11	1.000 11	1.000 11
前建立模型	Y1：全国快递业务量（亿件）	相关系数 显著性（双尾）	1.000 11	1.000 11	1.000 11
	X1：第三产业增加值（亿元）	相关系数 显著性（双尾）	1.000 11	1.000 11	1.000 11
	Y2：社会零售品总额（万亿元）	相关系数 显著性（双尾）	1.000 11	1.000 11	1.000 11
	X2：全国网购交易规模（万亿元）	相关系数 显著性（双尾）	1.000 11	1.000 11	1.000 11

相关性		Y2：全国快递业务收入（亿元）	X1：第三产业增加值（亿元）	Y2：社会零售品总额（万亿元）	X2：全国网购交易规模（万亿元）
前建立模型	Y2：全国快递业务收入（亿元）	相关系数 显著性（双尾）	1.000 11	1.000 11	1.000 11
	X1：第三产业增加值（亿元）	相关系数 显著性（双尾）	1.000 11	1.000 11	1.000 11
	Y2：社会零售品总额（万亿元）	相关系数 显著性（双尾）	1.000 11	1.000 11	1.000 11
	X2：全国网购交易规模（万亿元）	相关系数 显著性（双尾）	1.000 11	1.000 11	1.000 11
前建立模型	Y2：全国快递业务收入（亿元）	相关系数 显著性（双尾）	1.000 11	1.000 11	1.000 11
	X1：第三产业增加值（亿元）	相关系数 显著性（双尾）	1.000 11	1.000 11	1.000 11
	Y2：社会零售品总额（万亿元）	相关系数 显著性（双尾）	1.000 11	1.000 11	1.000 11
	X2：全国网购交易规模（万亿元）	相关系数 显著性（双尾）	1.000 11	1.000 11	1.000 11

四、模型建立与评估

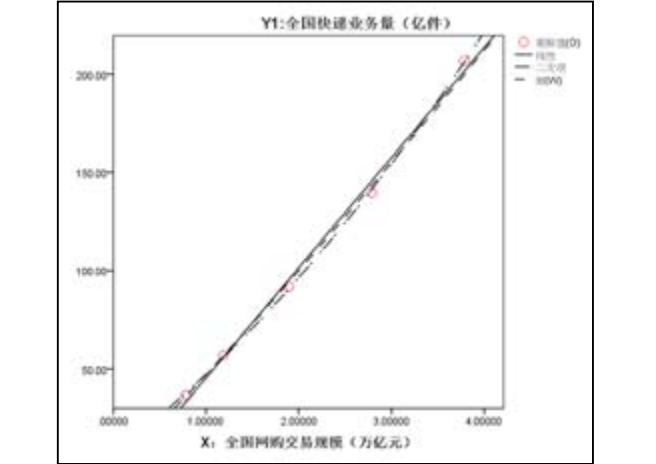
快递业务量、收入，与网购交易规模存在相关关系，如何用方程表示？

4.1 选择数据



根据快递行业影响因素分析结果，参考“十一五”、“十二五”期间网购交易规模的增幅变化趋势，结合2016年快递业发展的最新数据，本研究预测“十三五”期间快递业的发展，无论从时间上看，还是从业务量、收入的增幅上看，距离“十二五”期间的数据差距均较小。因此，选择2011~2015年的“快递业务量、收入”和“网购交易规模”五年的数据，进行建模。

4.2 业务量建模



在计量经济学领域，经常假定增长不是线性的，因此需要在线性回归之外，增加曲线拟合进行比较。

具体根据图像观察结果呈现出可能存在线性、二次项或幂函数，利用SPSS软件进行曲线拟合，探索建立比较三种回归拟合模型，得到拟合图像及检验结果。

F临界值=FINV（0.01，1，3）=34.116

t临界值=TINV（0.01，3）=5.841

序号	曲线类型	R平方	F 检验	t 检验	P 检验	备注
1	线性	0.995	625.831	25.017	0.000	检验通过
2	幂函数	0.999	2124.125	46.088	0.000	检验通过
3	二次项	1	2433.141	8.099, 4.607	0.015, 0.044	t检验不通过 P检验不通过

Y’ 1 = 56.051X – 10.645

Y’ 2 = 47.106X1.085

X：网购交易规模；Y：快递业务量。

4.3 业务量模型评估

年度	X网购交易规模（万亿元）	Y快递业务量（亿件）	Y ₁ 预测	Y ₂ 预测	Y ₁ -Y：绝对误差	Y ₂ -Y：绝对误差	(Y ₁ -Y)/Y：相对误差	(Y ₂ -Y)/Y：相对误差
2011年	0.785	36.70	33.33	36.20	-3.37	-0.50	-9.19%	-1.36%
2012年	1.187	56.90	55.89	56.74	-1.01	-0.16	-1.77%	-0.28%
2013年	1.892	91.90	95.43	94.11	3.53	2.21	3.84%	2.41%
2014年	2.789	139.60	145.68	143.34	6.08	3.74	4.35%	2.68%
2015年	3.784	206.70	201.47	199.61	-5.23	-7.09	-2.53%	-3.43%

数据带入方程，对比实际数据，得到相对误差，其中：Y’ 2（幂函数回归）方程的预测效果更好，预测误差均在[-5%，5%]区间。

另外，从函数的形态上看，Y’ 2 = 47.106X1.085 的增长速度，早期（2011、2012年）要快于Y’ 1 = 56.051X – 10.645，后期（2012年之后）则放缓。这个情况，与2012年中国GDP增长率首次跌破8%，网购交易总额增长开始下降到60%以下的宏观经济发展情况，密切相关。同时，2013年中国快递业务量发展的体量，已经超过90亿件。快递业务量的持续高增长，已经难以维系。这也是符合市场发展规律的，业务解释上也同样合理。



根据预测效果的评估、宏观经济发展趋势和快递业务发展解释，选择Y’ 2 = 47.106X1.085作为全国快递业务量建模的结果。

五、发展分析与预测

快递业务量、收入的回归方程确定了，网购交易未来几年的数据如何获得？（通过对外部环境的PEST分析，合理假设）

5.1 PEST分析——政策环境

1) 快递业发展纳入国家战略

2015年10月，颁布《国务院关于促进快递业发展的若干意见》，明确快递业发展的四大内容、五项任务、六大举措；2016年3月，国家发改委发布《中华人民共和国国民经济和社会发展第十三个五年规划纲要》，三次提及快递。

2）快递业管理纳入法制建设

《快递条例》立法工作得到国务院高度重视，连续三年列入立法计划。

2015年5月，国家邮政局、国务院法制办联合开展快递条例立法调研。11月16日，《快递条例（征求意见稿）》公开征求意见。2016年3月3日，俞正声在全国政协十二届四次会议开

幕式上，在工作报告中首次提及“快递业”。2017年3月3日，俞正声在全国政协十二届五次会议开幕式的工作报告中，回顾2016年工作再次指出：聚焦法治建设和社会社会治理重要问题，对《快递条例》的制定等调研议政。2017年3月20日，国务院办公厅印发的《国务院2017年立法工作计划》指出，快递条例已经列入今年的立法计划。

3）快递业发展进入新蓝海



2016年1月27日正式发布中央一号文件——《中共中央国务院关于落实发展新理念加快农业现代化实现全面小康目标的若干意见》，将实施“快递下乡”工程写入其中。2017年2月——《中共中央国务院关于深入推进农业供给侧结构性改革加快培育农业农村发展新动能的若干意见》正式发布，实施“快递下乡”工程被纳入其中。2017年3月5日，国务院总理李克强在政府工作报告中明确提出要“促进电商、快递进社区进农村，推动实体店销售和网购融合发展”。

5.2 PEST分析——经济环境/GDP



发现1：第三产业增长率（除2010年外），均不低于同年国民生产总值的增长率；2013年及以后，第三产业增长率均高于第二产业增长率。

发现2：第三产业，在国民生产总值中的占比不断上升；到2012年，与第二产业占比基本持平；到2015年，第三产业占比超过50%。

发现3：第三产业平均增长率，分别高出同期国民生产总

值平均增长率1.6和3.6个百分点。

指标	“十一五” 期间	“十二五” 期间
第三产业增加值平均增长率(%)	18.7	12.5
国内生产总值平均增长率(%)	17.1	8.9
第三产业增加值平均增长率(%)÷国内生产总值平均增长率(%)	1.6	3.6

5.3 PEST分析——经济环境/通货膨胀

在经济学上，通货膨胀意指整体物价水平持续性上升。一般性通货膨胀为货币贬值或购买力下降。通货膨胀的严重程度是用通货膨胀率来反映的，它说明了一定时期内商品价格持续上升的幅度。通货膨胀率一般以消费者物价指数来表示。

消费者物价指数，是世界各国普遍编制的一种指数，它可以用于分析市场价格的基本动态，是政府制定物价政策和工资政策的重要依据，同时反映与居民生活有关的产品及劳务价格统计出来的物价变动指标，通常作为观察通货膨胀水平的重要指标，我国称之为居民消费价格指数。

通货膨胀率

=

报告期消费者物价指数

—

基期消费者物价指数

基期消费者物价指数

x 100%

指标	居民消费价格指数 (上年=100)	CPI 计算通货膨胀率 (%)	通货膨胀率 平均值 (%)
1996 年	108.3	8.3	1.9
1997 年	102.8	2.8	
1998 年	99.2	-0.8	
1999 年	98.6	-1.4	
2000 年	100.4	0.4	1.4
2001 年	100.7	0.7	
2002 年	99.2	-0.8	
2003 年	101.2	1.2	
2004 年	103.9	3.9	3.0
2005 年	101.8	1.8	
2006 年	101.5	1.5	
2007 年	104.8	4.8	
2008 年	105.9	5.9	2.8
2009 年	99.3	-0.7	
2010 年	103.3	3.3	
2011 年	105.4	5.4	
2012 年	102.6	2.6	2.8
2013 年	102.6	2.6	
2014 年	102.0	2.0	
2015 年	101.4	1.4	

通过观察，自“九五”期间，通货膨胀率平均值为1.9%，到“十二五”期间，通货膨胀率平均值为2.8%，通货膨胀率有上升的趋势。

据国家发改委经济研究所课题组研究成果显示：“十三五”期间中国经济平均增速为7%左右，物价总水平维持在2.4%~4.6%之间。也就相当于通货膨胀率将被控制在2.4%~4.6%之间。



另据中国农业银行战略规划部课题组的研究成果，预计“十三五”期间，我国通货膨胀率维持在同比2%~3%。

5.4 PEST分析——经济环境/未来经济发展

摩根士丹利报告：详尽地给出看好中国经济前景的原因：一是中国的消费格局正在发生改变，消费成为引领经济增长主力，中国已成为全球最大的消费市场；二是服务业将是增长最快的领域；三是中国正向高附加值制造强国迈进；四是企业继续转型升级提质增效；五是过剩产能问题正在逐步解决。

国家统计局：2016年全年，最终消费支出对经济增长贡献率高达64.6%，较2015年进一步提高4.9个百分点；

渣打银行：发布《2017年全球市场展望》称，2017年中国经济很可能保持6.5%左右的增速，继续作为亚洲各国和地区及发展中国家的“稳定锚”，为全球贸易和商品价格提供支撑；

《华尔街日报》：中国经济已经进入稳定且持续增长的新常态，主要推动力是家庭支出和服务业；

普华永道：2017年2月的研究报告《长远前景：2050年全球经济排名将会如何演变？》指出：以购买力评价（Purchasing Power Parity，简称PPP）计算，中国的GDP已超越美国，成为全球最大经济体；以市场汇率计算，中国将在2030年之前超越美国成为全球最大经济体。

国际货币基金组织：最新《世界经济展望》报告中，将2016年中国经济增长从6.6%上调至6.7%，超过印度重回全球经济增长率第一，并且将对中国2017年经济增速预期从2016年10月份的6.2%上调0.3个百分点为6.5%。

通过国内外著名的投资公司、咨询公司、经济组织预测：消费将推动中国经济中高速发展（6.5%），潜力巨大，服务业将是增长最快的领域。

5.5 PEST分析——社会环境

我国电信网、互联网、移动互联网、智能设备，以及电子支付、移动支付、电子商务、网络购物的普及，为快递业的发展插上了翅膀。

从全国人均快递量的统计结果上看，在“十一五”期间，全国人均快递量年平均增幅为18.82%，处于低速发展期；而“十二五”期间，全国人均快递量年平均增幅达到53.81%，处于高速发展期；2016年，全国人均快递量依然超过50%，是2006年的26.8倍，如下图所示。但相比发达国家美国30~32件、日本30件，还存在一定的差距，未来提升潜力充足。

从区域上看，电商发展好的江浙沪地区，年人均快递量在30件至50件；而在中西部和农村地区，远低于全国平均水平，具有巨大升值空间。近年来，全国人大、国务院及相关部门出台多项推动快递业发展的政策文件。2016年，“快递下乡”被纳入中央一号文件，成为国家战略，为农村网购市场和快递业的快速发展注入新的活力。



5.6 行业发展分析——快递平均单价

从总体上看，2007~2016年全国快递平均单价呈现下滑趋势，但降幅趋于平稳。



从业务种类上看，同城业务，2007~2012年，在7.6~9元/件之间波动；2013~2016年，在7.2~7.6元/件之间波动，幅度明显减少，且2016年同城业务的价格已接近2008年的水平。而异地业务，2007~2016年一直呈现下降趋势，2016年异地业务的价格不足2007年价格的0.39，价格略高于同城，降幅空间不大。国际及港澳台，2007~2016年价格降多增少，2016年国际及港澳台业务的价格不足2007年价格的50%。

5.7 行业发展分析——快递收入

据国家邮政局发布的“十一五”、“十二五”期间快递业收入数据，结合同期第三产业增加值数据，发现：快递业收入平均增长率，分别是同期国民生产总值平均增长率的0.9倍和3.1倍。

指标	“十一五” 期间	“十二五” 期间
第三产业增加值平均增长率(%)	18.7	12.5
快递业务收入平均增长率(%)	17.7	38.3
快递业收入平均增长率÷第三产业平均增长率	0.9	3.1

2017年是实施“十三五”规划的重要一年，是供给侧结构性改革的深化之年。近期举行的中央经济工作会议指明了经济工作方向：稳中求进是2017经济工作的总基调。2017年政策的基本方向，是坚持以推进供给侧结构性改革为主线，适度扩大总需求，推动中国经济转型升级。投资、消费、出口的总体扩大和协调平衡仍是拉动经济增长的重要动力。据中国经济学家及前世界银行高级副行长林毅夫和中国银行首席研究员宗良预计：十三五期间，中国GDP可保持6.5%增长，在合理区间内中高速运行。

“十一五”规划和“十二五”规划，明确提出第三产业（服务业）占比提升幅度分别是3个百分点和4个百分点，实际占比提升幅度达到2.3个百分点和6个百分点。另据艾瑞咨询预测2016年全国网购交易规模5.32万亿元。因此，对中国经济和快递业发展进行四项假设：

假设1：十三五期间，国民生产总值保持6.5%增长；

假设2：十三五期间，第三产业在国民生产总值的占比，提升幅度达到5个百分点；

假设3：十三五期间，快递业收入是同期第三产业平均增幅的2.5倍；

假设4：十三五期间，快递平均单价平均增幅为-2%。

可得出：十三五期间，第三产业平均增幅为8.2%，快递业收入平均增幅20.5%，计算出到2020年，快递业收入将达到8380亿元，快递平均单价11.72元/件，快递业务量715亿件。

研究结论

无论宏观经济发展环境分析预测，还是回归模型数据分析预测的结果上看：在“十三五”期间，全国快递业务量均超过700亿件，业务收入超过或有望超过8000亿元，能够实现《国务院关于促进快递业发展的若干意见》对快递业务量和业务收入的规划要求和《邮政业发展“十三五”规划》中制定的发展目标。

伴随“一带一路”、“长江经济带”、“京津冀协同发展”、“西部电商”、“农村电商”、“跨境电商”和城镇化战略的逐步实施和深入开展，以及大数据、“互联网+”、智能物流等新技术，中国快递业必将迎来新的发展机遇。

FIN.

/ 关于流式大数据实时处理技术、平台及应用 /

文 / 中国大数据 编辑 / 协会会员处 李缘 日期 / 2017-09



大数据技术的广泛应用使其成为引领众多行业技术进步、促进效益增长的关键支撑技术。根据数据处理的时效性，大数据处理系统可分为批式(batch)大数据和流式(streaming)大数据两类。其中，批式大数据又被称为历史大数据，流式大数据又被称为实时大数据。

目前主流的大数据处理技术体系主要包括hadoop及其衍生系统。Hadoop技术体系实现并优化了MapReduce框架。Hadoop技术体系主要由谷歌、推特、脸书等公司支持。自2006年首次发布以来，Hadoop技术体系已经从传统的“三驾马车”(HDFS、MapReduce和HBase)发展成为包括60多个相关组件的庞大生态系统。在这一生态系统中，发展出了Tez、Spark Streaming等用于处理流式数据的组件。其中，Spark Streaming是构建在Spark基础之上的流式大数据处理框架。与Tez相比，其具有吞吐量高、容错能力强等特点，同时支持多种数据输入源和输出格式。除了Spark开源流处理框架，目前应用较为广泛的流式大数据处理系统还有Storm、Flink等。这些开源的流处理框架已经被应用于部分时效性要求较高的领域，然而在面对各行各业实际而又差异化的需求时，这些开源技术存在着各自的瓶颈。

在互联网/移动互联网、物联网等应用场景中，个性化服务、用户体验提升、智能分析、事中决策等复杂的业务需求对

大数据处理技术提出了更高的要求。为了满足这些需求，大数据处理系统必须在毫秒级甚至微秒级的时间内返回处理结果。以国内最大的银行卡收单机构银联商务为例，其日交易量近亿笔，需对旗下540多万个商户进行实时风险监控，在确保这些商户合规开展收单业务的同时，最大限度地保障个人用户的合法权益。这样的高并发、大数据、高实时应用需求给大数据处理系统提出了严峻的挑战。银联商务以前使用的T+1事后风控系统存在风险侦测迟滞高(次日才能发现风险，损害已经造成)、处理时间长(十几个小时之后才能完成风险识别)、无法处理长周期历史数据(只能分析最近几日的流水数据)以及无法支持复杂规则(仅能支持累积求和等简单规则)等重大缺陷。为此，亟须研发全新的事中风控系统，以重点实现低迟滞(在1 min内甄别突发风险)、高实时(100 ms内返回处理结果)、长周期(可处理长达10年以上的历史周期数据)以及支持高复杂度规则(如方差、标准差、K阶中心矩、最大连续统计等)等目标。这一目标可以抽象为一个大数据处理科学问题：如何在一个完整的大数据集上，

实现低迟滞、高实时的即席(Ad-Hoc)查询分析处理。

技术解析

现有的大数据处理系统可以分为两类：批处理大数据系统与流处理大数据系统。以Hadoop为代表的批处理大数据系统需先将数据汇聚成批，经批量预处理后加载至分析型数据仓库中，以进行高性能实时查询。这类系统虽然可对完整大数据集实现高效的即席查询，但无法查询到最新的实时数据，存在数据迟滞高等问题。相较于批处理大数据系统，以Spark Streaming、Storm、Flink为代表的流处理大数据系统将实时数据通过流处理，逐条加载至高性能内存数据库中进行查询。此类系统可以对最新实时数据实现高效预设分析处理模型的查询，数据迟滞低。然而受限于内存容量，系统需丢弃原始历史数据，无法在完整大数据集上支持Ad-Hoc查询分析处理。因此，研发具有快速、高效、智能且自主可控特点的流式大数据实时处理技术与平台是当务之急。

实现一个融合批处理和流处理两类系统且对应用透明的系统级方案，需要攻克以下几个技术难点。

- (1)复杂指标的增量计算
- 尽管计数、求和、平均等指标能够依靠查询结果合并实现，然而方差、标准差、熵等大部分复杂指标无法依靠简单合并完成查询结果的融合。再者，当查询涉及热点数据维度及长周期时间窗口的复杂指标时，多次重新计算会带来巨大的计算开销。
- (2)基于分布式内存的并行计算
- 采用粗放的调度策略(例如约定在每天的固定时间将流数据导入批处理系统)会造成内存资源的极大浪费，亟须研究实现一种细粒度的基于进度实时感知的融合存储策略，以极大地优化和提升融合系统的内存使用效率。
- (3)多尺度时间窗口漂移的动态数据处理
- 来自业务系统的数据查询请求会涉及多种尺度的时间窗口，如“最近5笔刷卡交易的金额”“最近10 min内密码重试次数”“过去10年的月均交易额”等。每次查询请求都重新计算结果会对系统性能造成极大的影响，亟须研究实现一种支持多种时间窗口尺度(数秒到数十年)、多种窗口漂移方式(数据驱动、系统时钟驱动)的动态数据实时处理方法，以快速响应来自业务系统的即席查询请求。
- (4)高可用、高可扩展的内存计算

基于内存介质能够大大提升数据分析及处理能力，然而由于其易挥发的特性，一般需要采用多副本的方式来实现基于内存的高可用方案，这使得“如何确保不同副本的一致性”成为一个待解决的问题。此外，在集群内存不足或者部分节点失效时，“如何让集群在不间断提供服务的同时重新平衡”同样是一个待解决的技术难题。亟须研究分布式多副本一致性协议以及自平衡的智能分区算法，以进一步提升流处理集群的可用性以及可扩展性。

“流立方”流式大数据实时处理技术在上述领域取得了一系列突破，该技术提供基于时间窗口漂移的动态数据快速处

理，支持计数、求和、平均、最大、最小、方差、标准差、K阶中心矩、递增/递减、最大连续递增/递减、唯一性判别、采集、过滤等多种分布式统计计算模型，并且实现了复杂事件、上下文处理等实时分析处理模型集的高效管理技术。

平台纵览

基于“流立方”流式大数据实时处理技术，研发了“流立方”流式大数据实时处理平台。其应用框架如图1所示，具有良好的灵活性和适应性。平台的数据装载模块负责从具体业务系统中接入实时流数据，数据抽取模块负责批量抽取历史数据，模型装载模块负责将分析处理模型集中的计算模型和脚本加载到平台中。当收到业务系统发出的实时查询请求时，“流立方”平台能够根据分析处理模型在完整大数据集上实时计算出相应的指标，并进行判断，将结果反馈给业务系统。

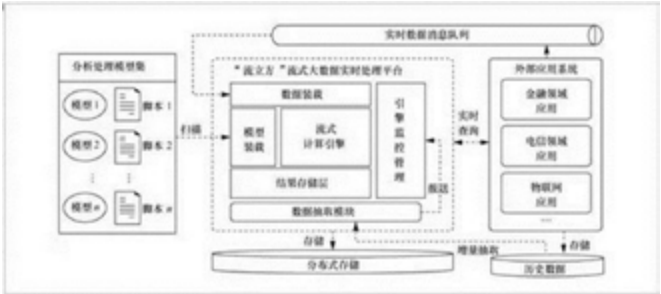


图1：“流立方”平台应用框架

在测试环境为8台服务器(每台服务器配置24核 CPU、256 GB内存)，同时计算16个统计指标(涉及4个维度，包含计数、求和、平衡、最大、最小、标准差、过滤、去重、排序、复杂事件处理等多种算法)的性能测试中，“流立方”平台达到了单节点写入大于43 000 TPS、8节点读取大于100万TPS、平均时延为1~2 ms的优异性能，如图2所示。

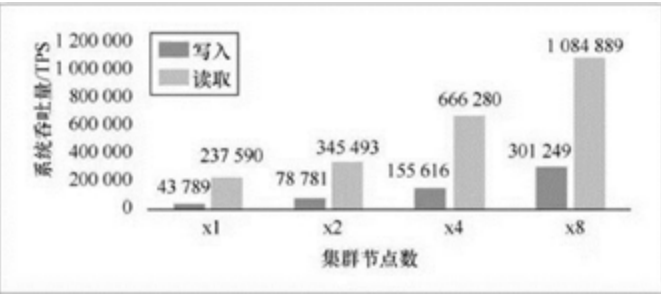


图2：“流立方”平台性能指标

“流立方”平台在解决批式大数据和流式大数据融合实时处理技术难题，实现优异性能的同时，还解决了流式大数据处理平台面临的两大工程化难题。一是作业的编排效率问题。大部分开源流处理平台在完成一个流处理编排时，都需要经过拓扑设计、代码编写、功能测试、打包部署等环节，一般需要一周的时间才能完成。“流立方”平台通过基于“所见即所得”的在线作业编排管理，将上线任务耗时降低到分钟级，大大提升了流处理作业的编排效率。二是流处理作业的灵活变更问题。流处理平台擅长进行逻辑预先定义的增量计算，尽管其计



算效率极高，但计算灵活度受到限制。例如，某业务需要统计过去3个月的数据，现有的流处理平台在该业务上线3个月后才能完全生效，这样的工作方式使流处理技术在实际应用中受到很大的局限。“流立方”平台创新性地引入流媒体播放器的录制与重放思路，在原始数据进入流处理平台时，通过顺序写的方式持久化一份原始数据，在需要上线新的计算作业时，即刻重发指定时间窗口内的原始数据，从而实现快速(分钟级甚至秒级)计算作业上线。

“流立方”平台引入了一系列创新技术，在性能、可用性、可扩展性等多个层面提升了流处理平台的处理能力，满足金融领域在内的众多领域的业务及运维需求。引入数据冲突智能规避技术，解决了流式处理中的热点数据处理问题，从而解决了大颗粒数据维度的处理效率问题;引入Paxos一致性协议，解决内存存储计算时多副本一致性问题，提供了面向运维人员透明的一致性解决方案;引入智能分区技术，基于一致性散列技术，进一步将散列值拆解为散列块，通过散列块的平滑迁移解决存储集群的可伸缩性设计问题，确保对于运维人员的集群变更透明性;引入计算作业的动态运行时加载技术，规避了作业手工打包部署的问题，进一步提升了开发人员的工作效率。

在国内某大型银行卡收单机构组织的招标测试中，测试环节为两台低配置虚拟机，测试数据为该机构的数千万笔交易流水，计算逻辑包括50多条规则，涉及30多个统计指标。在该测试环节下，两家国外著名厂商中，一家厂商的计算时间长达24 h，另一家老牌数据库软件提供商则未能在一天内完成计算。相较于这些国外著名厂商的大数据处理平台，“流立方”平台能够在3 h内完成所有计算，且正确率为100%。

应用场景

“流立方”流式大数据实时处理系统在金融、交通、电信、公安等行业具有广泛的应用场景。以金融风控反欺诈为例，部署“流立方”风控系统仅需在交易前端增加风控探头，将实时交易数据旁路接入系统。“流立方”风控系统根据融合了专家知识和机器学习结果的数百条规则对每笔交易进行风险评估，判断是否允许进行该笔交易，流程如图3所示。该系统平均响应时间在6 ms以下，并发数超过50 000笔/s。同时，实现这一性能仅需要4台服务器。

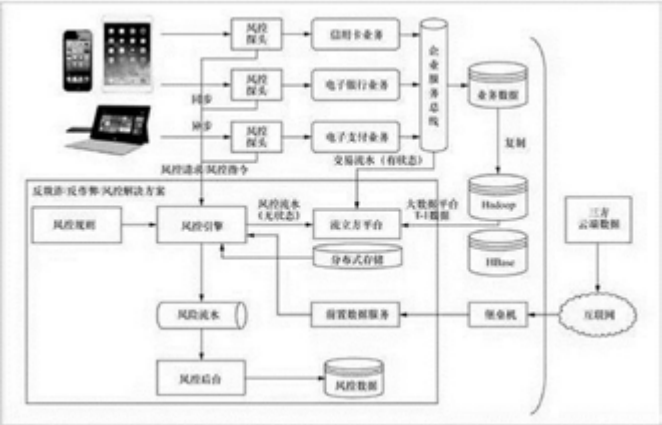


图3：基于“流立方”的金融风控反欺诈流程

基于“流立方”的金融风控反欺诈技术体系包含技术(如设备指纹、代理侦测、生物识别、关联分析、机器学习等技术)、知识(如盗卡反欺诈、伪卡反欺诈、信用卡套现、营销反欺诈等规则与模型)、数据(如虚假手机数据、代理IP数据、P2P失信数

据等标识数据)三大板块。

技术部分中的设备指纹技术通过主被动混合的形式采集设备中软硬相关要素，结合概率论等算法为每一个设备颁发一个全球唯一的指纹编码，这些指纹编码在反欺诈的整个过程中起到非常积极的作用；代理侦测技术通过短时间内扫描IP相关端口来识别那些开启代理的IP，并在这些IP访问金融服务时进行识别;生物识别技术通过采集设备上用户的鼠标点击、触摸、键盘敲击等行为识别操作者是人还是机器以及是否操作者本人的问题;关联分析技术在底层通过图数据库存储不同节点以及关系信息，最终在界面上通过图的形式进行欺诈者关联分析及复杂网络分析；机器学习技术通过有监督、无监督的机器学习算法提升欺诈识别的准确率及覆盖率，并结合流立方技术提供模型的事中预测能力。

基于上述技术体系，研发了银行业务风险实时监控系统、互联网支付业务风险实时监控系统、电商业务风险实时监控系统等金融风控反欺诈系列解决方案。这些方案已应用到银行、第三方支付机构、互联网金融等领域的上百家企业。目前50%以上的线下交易都在“流立方”的保护下进行，基于“流立方”的金融风控反欺诈解决方案每天为我国的金融机构抵御上亿次的攻击。该技术已经成为我国金融安全领域基础设施必不可少的组成部分。

此外，在互联网机器防御系统中，“流立方”同样能发挥巨大作用。如今网络机器人遍布票务、电商、招聘、银行、政府、社交等各类网站，消耗了40%~60%的网络流量。网络机器人不仅消耗网络资源、影响正常客户访问、增加网站运营成本，还会爬取产品、价格信息，形成不正当竞争，甚至混淆网站用户生态，影响营销分析。传统的控制策略通过采取屏蔽频繁访问、设置验证码等方式防御网络机器人，无法应对日益智能化的新型网络机器人。

基于“流立方”的互联网机器防御系统通过在Web服务器上嵌入插件或者独立的嗅探器(sniffer)程序，将全流量的Web访问请求旁路到独立的机器防御集群，进行实时的流量分析及防御决策，并将决策后的结果实时回馈到Web服务器插件中。Web服务器插件在判定当前访问的设备或者IP地址等是机器人时，能够自动改写响应内容，根据不同的风险级别自动拒绝交易或将访问者引导到第三方图形验证码服务商进行机器人验证。

访问者在通过验证后可以继续正常访问Web服务。该系统还创新地将设备指纹以及人机识别服务运用到机器防御系统中，不仅增加了可分析维度，提升了控制颗粒度，同时能够对基于浏览器内核的高级爬虫进行防护。此外，将机器防御规则、数据服务、设备指纹、人机识别以及图形验证码以软件即服务(software as a service, SaaS)的形式提供服务，进一步降低了互联网网站客户的运维门槛，提升了产品竞争力。

基于“流立方”的实时机器防御系统通过多服务器访问流水关联决策、长周期数据决策、复杂规则爬虫识别、设备维

度爬虫识别、人机识别等技术，实现了微秒级(400~800 μs)的识别时延，同时具有机器人识别管控一体化、轻量级接入等优点。根据已经接入机器防御服务的几十家客户的反馈，基于“流立方”平台的防御系统对机器人识别覆盖率在95%以上，准确率为99.9%。该机器防御系统能够拦截这些客户业务系统中占原有访问总流量80%~90%的来自网络机器人的访问流量，使得其业务系统服务器的压力降为原来的10%。由于基于“流立方”的机器防御系统的卓越识别及控制机器人的能力，当前，全国最大的票务平台正在对此服务进行全面的测试，希望能够进一步提升其票务服务能力。

此外，基于“流立方”的流式大数据实时处理平台在智慧交通领域也大有作为。通过实时分析从预埋在全国各地的摄像头采集的车牌信息，配合地理位置信息服务以及基于地理信息系统(geographic information system, GIS)的最短交通距离计算，实现实时套牌车信息抓取，为进一步打击违法犯罪服务提供帮助;通过实时分析交叉路口双向的车流量信息，实时控制每个路口的红绿灯、智能变换潮汐车道及可变车道，从而大大提升城市的通行效率。

“热数据”带来无与伦比的价值，数据从产生开始，其应用价值随时间的流逝呈现指数式下降，如何充分应用“热数据”是一个新生事务，是一个长期任务，也是流式大数据处理技术大有可为之处。“流立方”流式大数据实时处理技术和平台在金融、电信、交通、公安、海关、网络安全等需要引入“事中”感知分析决策模式的行业都具有广阔的应用前景。

结束语

基于批式大数据，可以不断学习新的知识，累积新的经验。然而，在应用这些知识和经验时，流式大数据更能够极大地挖掘“热数据”的潜在价值。这使得流式大数据技术具备更有效的应用推广价值。

流式大数据实时处理是大数据时代信息化的重要抓手。采用“事中”甚至“事前”模式实现感知、分析、判断、决策等功能的智能系统需要流式大数据实时处理平台的支撑。此外，流式大数据实时处理可以为大数据驱动的深度学习提供计算框架支撑。“流立方”流式大数据实时处理平台可为研制融合逻辑推理、概率统计、众包、神经网络等多种形态的下一代人工智能统一计算框架提供支持。

FIN.

/ 大数据告诉你, 近十年来苹果中国消费者变迁 /

文 / 华尔街见闻 编辑 / 协会会员处 李缘 日期 / 2017-09



也许大家都还记得iPhone 4s及iPhone 5开售时，中国消费者疯狂抢购的景象，拥有一台“土豪金”一度也成为江湖地位的象征。但随着国货的崛起及消费个性化的发展，iPhone已经不再是稀缺产品与身份的代表，在经历了iPhone7系列的惨败之后，苹果急需iPhone 8及iPhoneX赢回中国市场。

据统计，大中华地区是苹果在全球的第三大市场，但苹果目前在中国智能手机市场的份额已跌至第五，与此同时，在中国市场的营收也继续“拖苹果后腿”。上月初，苹果公布的第三财季显示，公司在美国 and 全球大多数市场销售业绩均同比增长，中国却连续第六个季度下滑。

不管如何，苹果依然是一个让人尊敬的时代品牌。近日，中国最大的手机零售平台京东发布了以“2009-2017果然精彩”为主题的京东平台iPhone大数据报告，该报告涵盖了近十年来京东销售的iPhone数据，从购买者的年龄、性别、学历以及产品喜好等角度进行分析，揭示了iPhone近十年来在京东平台消费趋势的变化。

Global Smartphone Vendor Marketshare (%)				
Company	2016 Shipments Volume	2016 Market Share	2017 Shipments Volume	2017 Market Share
Samsung	77.6	22.7%	79.5	22.7%
Apple	40.4	11.8%	41.0	11.4%
Huawei	32.0	9.4%	38.4	10.7%
OPPO	18.0	5.3%	29.5	8.2%
Xiaomi	14.7	4.3%	23.2	6.4%
Others	108.8	48.5%	168.8	41.3%
Total	341.5	100.0%	380.4	100.0%

Source: Strategy Analytics, W55 Services

据Strategy Analytics,苹果手机2017年2季度市场份额为11.4%，继续大幅落后三星，但较于2016年的市场份额已经企稳并缓慢回升。

三线及以下城市消费比例上升明显

据京东大数据统计，2010年以前，iPhone只是中国一、二线城市消费者的“专享品”，份额一度超过90%，但受到用户消费能力提高以及渠道下沉的影响，三线及以下城市的消费比例上升明显。目前，一二线城市消费者份额维持在六成左右，三四线与五六线城市各占二成左右，近两年来，甚至出现了五六线城市的份额超过了三四线城市的情况。



年轻的高学历人群是购买iPhone的中坚力量

根据京东公布的iPhone数据显示，在京东上购买iPhone的

用户中，有51%的人是26-35岁，并且16-25岁的年轻用户占比一路飙升，高中生的比例也在增加，iPhone受到了越来越多年轻用户的追捧。



所有在京东购买iPhone的用户中，有90%以上是大学及以上学历，高学历的人群手机消费的目标依旧为iPhone。



除此外，通过该数据还可以看出，iPhone 6成为京东好评率最高的手机，达到98.3%；而iPhone用户复购周期主要集中在2-18个月，超过一半用户重视新鲜感等。



多数男女用户喜欢金色版本的iPhone，4.7寸屏幕成为更多人的首选

iPhone 5s金色版本的推出，极大的满足了中国消费用户对于黄金消费的心理需求，并且迅速成为中国消费者的最爱。京东iPhone大数据显示，有36%的男性用户、34%的女性用

户选择金色版本的iPhone，金色版的iPhone成为了相对较多人的选择。



而且随着智能手机屏幕的不断变大，苹果于2014年正式推出了4.7英寸的iPhone 6及5.5英寸的iPhone 6 Plus，此后的产品升级一致延续了该尺寸类型。通过京东iPhone大数据可以看出，女性用户更喜欢5.5英寸大屏，小屏相对比较受男性青睐，但是4.7英寸的屏幕，成为了消费的主流，男女用户比例均在50%左右。



据赛诺发布的《2017年1-6月iPhone中国销售数据报告》显示，从2009年开售至今，京东已经累计销售超过iPhone 1498.4万台。所有线上销售渠道中，京东以47%的份额占据行业第一，同比增长28%。据了解，京东是中国第一家iPhone线上授权零售商，同时也是iPhone在中国最大的零售平台。iPhone新品发布后，京东也将同步开启预售，不仅会推出白条免息及精准达等优惠，京东App Store充值卡旗舰店也将为果粉提供便利安全的充值服务。

FIN.

/ 大数据分析:《战狼2》火爆荧屏背后的必然 /

文 / 搜狐新闻 编辑 / 协会会员处 李缘 日期 / 2017-09

上个月，去电影院看到《战狼2》分数这么高——9.6分，这在人人审美感官都不同，一部片子总有人喊烂电影的时代，9.6分的国产电影这是多么不容易的成绩啊；最近几天，朋友圈和微博都被《战狼2》刷了个遍，这让笔者产生了对《战狼2》进行全面大数据分析的冲动。



《战狼2》是由吴京执导的动作军事电影，吴京、弗兰克·格里罗、吴刚、张翰、卢靖姗、丁海峰等主演。该片于2017年7月27日在中国内地上映。讲述了脱下军装的冷锋被卷入了一场非洲国家的叛乱，本来能够安全撤离的他无法忘记军人的职责，重回战场展开救援的故事。

将从3个角度进行分析：

1、全网的舆情分析：该电影在全网的传播趋势、热度、关注人群、关注人群画像，以及整体的口碑概况；

2、微博传播分析：基于该电影在微博的某条强大影响力和传播情况的分析，了解该条微博的传播规律，互动粉丝的画像，以及水军的辨识；

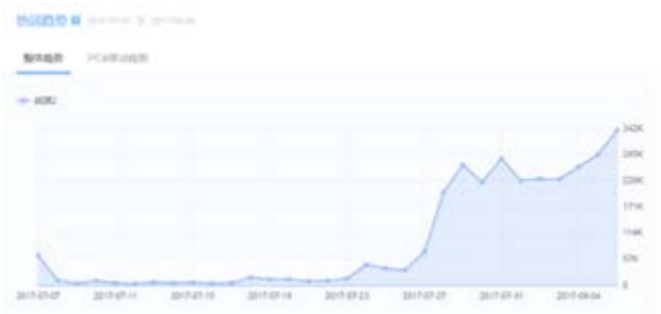
3、文本分析：通过对该电影豆瓣影评的文本数据进行分析，给主角“冷锋”画一个词云画像。

主要分为两大模块，一个是粗线条的、全局性的全网舆情分析，用的是新浪微舆情的热度指数查询功能；另一个是较细致的、微观的单条微博传播分析，用的是新浪微舆情的微博传播分析功能，话不多说，直接上分析。

一、全网热度分析

1.1、“战狼2”的全网传播走势

(1)全网传播走势



通过微指数近一个月的热词趋势，我们可以看出，战狼2的热度从7月27日，即“战狼2”的上映日开始成爆发式直线趋势飙升，自上映以来，大幅度飙升后，有小小幅度下降趋势，但由成上升趋势，并且热度值已达到342k，并且以图线趋势，还再不断上升。

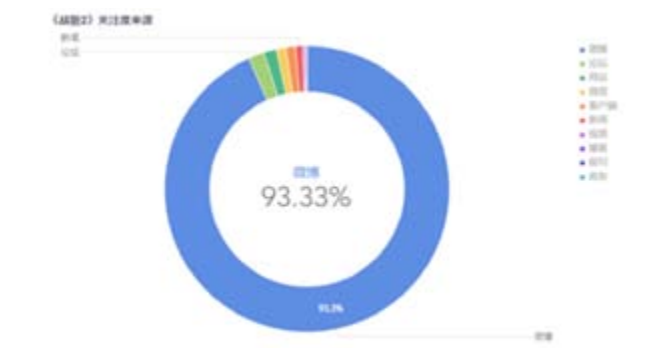
(2)全网关注度来源



从PC&移动近一个月的趋势图，我们可以看出，移动端的热度是PC端热度的三倍，“战狼2”的大多数热度来源自移动端，移动端的当月均值达到了67k，最高热度值达到了271k，这说明移动端(包括自媒体和网民)对该电影的讨论和发声较多。

1.2全网热度来源

(1)全网热度媒体来源



以上图表可以看出，该电影在微博上的信息量占据主导地位，微博仍然是目前娱乐影视的主要营销宣传增强影响力的主要渠道，也是主创团队与粉丝进行互动的主要方式。

除微博外，该电影的主要信息量来自论坛，其次来自网站和微信，论坛作为泛娱乐化社区关于该电影的讨论颇多，论坛是“粉丝文化”的起源地，“找组织来贴吧”，它的宣传造势能力推动了人们对该电影的关注度。

(2)全网热度地域来源

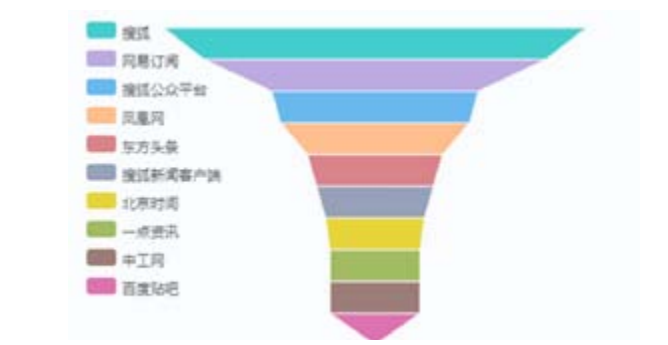


从地域分布来看，与“战狼2”相关的信息主要来源广州、江苏和北京。有点意外，前三竟然没有我大浙江，排名第一的不是帝都不是魔都，竟然是广东。然后笔者去了解了一下，广州为什么信息量在第一？

因为在8月5日的广州路演，吴京首次回应，茅台不是植入，吉普车也不是植入。吴京说：“这次无论是茅台，还是北京吉普，都不是植入的。我就是想宣扬一下我们的国货，让中国制造在世界上横行无阻。”

1.3友好度指数

(1)媒体友好度TOP10



媒体友好度TOP10，搜狐排名第一，百度贴吧第十，毕竟贴吧是网友吐槽水军泛滥发声发泄的地方。

(2)友好度指数

此数据来源微舆情，如图显示，媒体友好指数0.58，相对较为友好，网民指数没有媒体友好指数那么好，但是也是友好，综合友好注数为0.3。

二、重要微博舆情传播

对单条具有影响力的微博的影响力、传播节点等进行分析。官微最具影响力的一条微博是：



1、传播节点分析

吴京的老婆功不可没，转发次数和带动影响力最大，另外还有很多吴京好友的大V明星的互动转发，节点影响力扩散最为剧烈，如：戚薇、刘诗诗、赵丽颖等。

2、转发层级分析

微博层级越高，代表转发的越多，覆盖的微博用户最多，该条微博在传播中形成17个转发层级，经微博反垃圾系统处理后，剩余有效转发数294624条，覆盖微博用户455,370,456人；第3层级转发者共转发95600次，占总转发数的32.45%，第3层级是该条微博传播的主要力量。

此条文博传播关键用户是谢楠，她的粉丝数6195835人，带动了此条微博363639次的转发。

引爆点：



3、转发评论趋势

该微博于7月26日发布后，于7月28日00:00达到转发、转发峰值79458条、评论峰值10175条，此后微博传播速度逐渐降低。

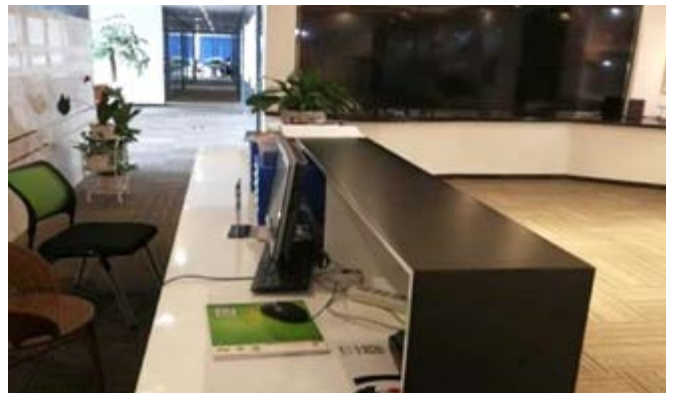
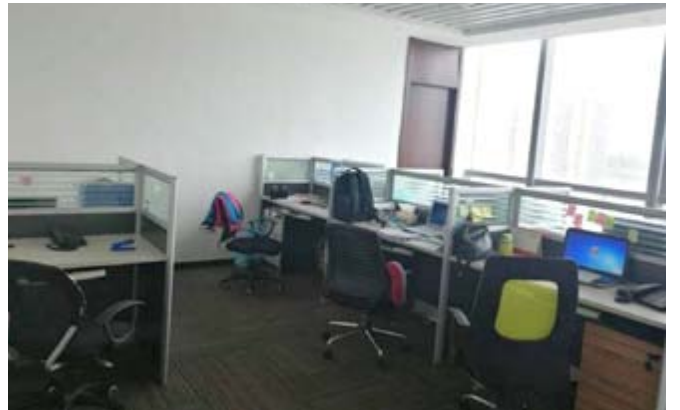
文 / 贵州中科数据分析师事务所 编辑 / 协会会员处 周燕 日期 / 2017-09



豆瓣影评



分析完了,《战狼2》的公众好评传播和大V微博传播力量对“战狼2”的成功起到了绝大作用,当然,公众的好评传播的根本原因还是在于《战狼2》的作品从制作上引起了群主共鸣,从剧情上以“祖国的力量”激起了民心。

FIN.

贵州中科项目数据分析师事务所是中国商业联合会数据分析专业委员会（简称：中数委）事务所会员单位（中数委团证：第112号）。中科事务所经贵州省贵阳市工商行政管理局核准登记，于2015年10月正式成立，是一家专业从事数据分析的服务性机构。

公司致力为中小型企业、国内外银行、投融资公司、政府组织等机构提供投资项目数据分析、行业数据分析、决策数据分析、经营数据分析、运营风险数据分析、投资项目评估、经济效益评价、投融资项目策划、编制可行性研究报告等服务，公司由多位高级数据分析师组成，是一支集经济研究、金融投资、财务分析、工商管理等多领域的复合型团队。能为项目

投资方及融资方提供一份具有经济性、权威性、客观性、公正性、实用性的数据分析报告。

随着大数据时代的快速发展，贵州中科项目数据分析师事务所秉承“诚信、客观、科学、实效、公正”的经营理念，专业的服务精神、先进的大数据分析理论体系为客户提供最有价值的服务，努力成为客户发展历程中值得信赖的合作伙伴，力求成为数据分析行业先领者，为推动中国数据分析事业的发展贡献力量。

贵州中科期待与您的精诚合作，共创辉煌！

FIN.

贵州中科项目数据分析师事务所

办公地址：贵州省贵阳市观山湖区黔灵山路德福中心A2栋11楼11号

联 系 人：孙经理

联系电话：0851-86837900 / 13985185365

CPDA®



如此困难
决策何必

CPDA 数据分析师
解决你的烦恼!

8

天面授 + 365天远程



社群



直播平台

+ Datahooop®
大数据分析专业平台 = my CPDA®



服务号: CPDA大数据圈

全国咨询热线: 400-050-6600

地址: 北京市朝阳区朝外SOHO-C座9层

电话: 010-59000991 / 010-59000559

培训网址: www.chinacpda.com

协会网址: www.chinacpda.org

