



数据分析

China Data Analysis

洞察损益 · 量衡天下

会员·特·刊

中国数据分析行业核心刊物

2014年
第4期
总第二十期

- 01 树行业先锋，做时代领航者
- 02 经营决策数据分析实务专题研讨会报道
- 09 数据挖掘经典算法系列之决策树
- 11 2014年度优秀项目数据分析师事务所评选专题
- 14 基于兰查斯特模型从价格角度分析
黄山品牌在安徽市场发展态势
- 22 新形势下数据分析行业面临的机遇、挑战及应对



行业协会
官方微博



官方微信
wxchinacpda

www.chinacpda.org

欢迎登陆中国数据分析行业网



Datahoop 大数据智能分析平台

Smart Platform

NEW

让大数据发挥 **大价值**

让您的决策比对手更快一步！

了解详情请致电：400-050-6600

1

十多年实践经验积累，集成行业顶尖算法

大数据的核心是分析。只有分析才能让数据发挥价值。而现在大部分大数据平台都没有很好的算法，甚至全盘照抄书上的算法，严重脱离实际。

DataHoop是由中国商业联合会数据分析专业委员会结合11年国内外数据分析行业实战经验自主研发的一款大数据智能分析平台，它结合了行业顶尖专家的经验 and 智慧。内置丰富的数据分析和数据挖掘算法，实现算法参数的自动智能调优和升级，同时包含最完善的行业应用模型，使之可以应用于各行各业。这是目前市面上任何一款软件或平台都无法比拟的优势。

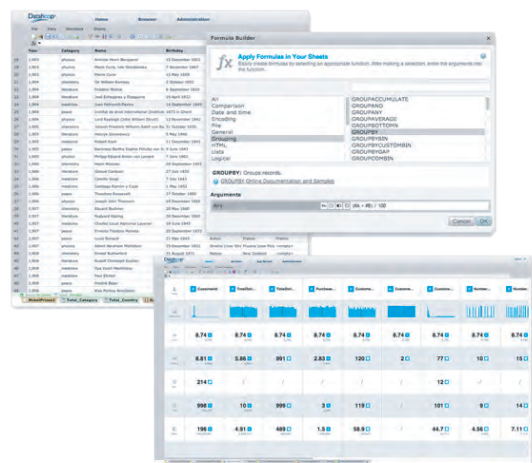


2

支持多种数据接口，真正实现无缝对接

DataHoop数据接口丰富，集成数据转化及预处理功能，提供实时/非实时统一接口，能与企业现有的ERP、CRM、OA、财务软件（金蝶、用友、SAP）以及公司网站等资源实现无缝对接，简单设置就可以对企业现有数据进行数据分析和挖掘，节省了大量的重新开发成本，节省了时间。





3

平台功能支持无限扩展及优化

DataHoop不断集成新的算法和新的功能模块，终身维护算法库，不断调优。这使得它的应用可以随使用者的需求不断扩展。

4

独创的安全系统，提供五重防护

DataHoop私有云尤其重视数据安全，自主研发的安全管理体系可以提供多达五重的防护，通过认证、加密、监控和追踪等手段在传统PC终端和移动终端提供数据保护解决方案。

5

操作简单，无技术要求

DataHoop让数据挖掘和数据分析操作简单，它独创的一键报表生成功能，使得非专业人士也可轻松发现数据价值，模型建立和使用无需编写任何代码。

6

跨平台移动终端支持，可以随时操纵数据

DataHoop提供多终端支持，手机也能访问，老板打开手机就能知道企业状况和解决方案，数据分析人员通过手机就能实现数据分析工作。



Datahoop®



现Datahoop大数据智能分析平台开始内测。

内测面向广大数据分析从业人员，数据分析相关行业从业者，以及数据分析爱好者。

如您希望参与平台内测，请通过行业协会邮件 marketing@chinacdpa.org 申请索取平台的内测账号，也可以通过协会热线400-050-6600进行申请。



卷首语

- 01 树行业先锋，做时代领航者

协会动态

- 02 协会新闻播报

技术专区

- 09 数据挖掘经典算法系列之决策树

优秀事务所评选

- 11 安徽量衡项目数据分析师事务所
12 中颢润（北京）项目数据分析师事务所
12 重庆传晟项目数据分析师事务所
13 上海天元项目数据分析师事务所
13 湖南翰林项目数据分析师事务所

数据分析应用

- 14 基于兰查斯特模型从价格角度分析
黄山品牌在安徽市场发展态势（节选）

- 18 盘点2014年掀起大数据热潮的五个黄金案例

- 20 浅谈数据安全

行业风向标

- 22 新形势下数据分析行业
面临的机遇、挑战及应对
24 浅谈项目数据分析师的培养
25 中国大数据行业亟待解决的五大问题

事务所风采

- 30 贵州新阳光项目数据分析师事务所
30 贵州北辰星项目数据分析师事务所

图说大数据

- 31 图说大数据
—— 数据分析师不是数羊的



2014年第4期 总第二十期

主办单位：中国商业联合会数据分析专业委员会

编辑：王宏莉 张楠 侯健 段欢欢

美工：崔峻珩

地址：北京市朝阳区建国门外大街建华南路17号现代柏联大厦5层

电话：010-65670421 转 115

传真：010-65671802

邮件：xiehui@chinacpda.org



中商联数据分析委员会



微信号：wxchinacpda

欢迎每日关注**微博****微信**

精彩行业信息等着你！

树行业先锋，做时代领航者

正如马云所说，“这个世界不是因为你能做什么，而是你该做什么”，在大数据概念被炒到极热的今天，采集数据、分析数据、应用数据显然已经成为这个时代应该学习的技能。那么对于政府、企业以及数据分析师事务所来说，就会面临这样一个问题：如何有效地将数据分析技能引入实践，真正实现“用数据说话，做理性决策”。同时，对于协会而言，在这个阶段，又该如何更好地发挥行业先锋的作用？

一个概念的引爆可以在一夜之间完成，但一个行业的兴起绝不是一蹴而就的，从行业建立到行业发展是一个长期的过程，协会起着不可替代地指导和监管作用。纵观数据分析行业来到中国的11年间，协会从项目引进、知识普及、行业标准制定、品牌宣传推广到数据分析人才培养、专业事务所建立，走过了一段漫长的市场培育期。在此期间，协会成功的举办了两届行业峰会、五届优秀事务所评选、数百场公益沙龙、几十场创业指导以及十周年行业庆典活动，在2014年开展了项目数据分析师事务所专题研讨会，并且对项目数据分析师课程体系进行了贴近实用性的全面优化。

我们惊喜地看到，所有这些努力都在潜移默化中推进着行业由模糊向着明晰化发展。首先，对“大数据”概念的认知由浮夸向理性回归，企业对数据分析的理解开始向深入发展；其次，应用领域由IT、零售、金融及医疗等敏感行业向全行业发展，企业开始有计划地成立数据分析部门；再次，事务所业务内容从偏重投融资数据分析转向指导企业经营决策方面发展；最后，操作模式由以技术为主导向“技术+咨询”的模式转变，IT不是大数据的核心，它只是进行数据分析的工具，随着企业行为数据化的建立，将数据盘活，进行深度分析后指导企业决策才是数据分析行业的核心所在。

“大数据”时代，是一场科技和产业深刻变革的时代，更是一个可以创造奇迹的时代，无数的机遇摆在面前，谁能够驾驭大数据，谁就可以站在这个时代的前端，成为“大数据”时代的主宰！如果说2013年是中国大数据元年，那么2014年就是数据分析行业元年，因为在这一年里，大数据回归理性，数据分析被引入更多领域和企业。诚如六年前协会在“法定业务”盛行的包围下，依然坚持市场化道路运作一样，今天以及未来的协会都将继续以“市场需求”为导向，继续将行业监管、事务所拓展、技术平台研发、人才培养以及企业数据分析应用指导等服务为己任，真正建立一个引领行业发展的服务型协会！

值此新年之际，我代表协会全体工作人员恭祝数据分析行业同仁、合作伙伴新春快乐、万事如意，感谢大家对行业协会多年的支持和帮助！同时再次恭贺新一届优秀事务所的诞生！我们也期盼在新的一年里更多优秀数据分析人才的涌现，以及更多政府、企业、机构在数据分析应用方面成果的展现！我们愿与您共创数据分析行业美好的明天！

中国商业联合会数据分析专业委员会



协会新闻播报

◎ 文 / 会员处 王宏莉 图 / 崔峻琦

1、“经营决策数据分析实务专题研讨会”成功举办

2014年11月29日至30日，由中国商业联合会数据分析专业委员会针对全国事务所会员单位举办的“2014年经营决策数据分析实务专题研讨会”在北京举办。

会议本着求真务实的精神，围绕“大数据发展机遇中，如何开辟更多的业务领域”这一主旨，邀请了全国各地的事务所代表，用真实的案例深度解析数据分析实务。希望通过此次会议明确数据分析行业未来的发展思路，同时也有助于行业内的交流与沟通。中国商业联合会数据分析专业委员会会长邹东生先生出席此次会议并致开幕词。



会上，各位专家老师不仅探讨了数据分析行业未来发展的模式，还就医疗、零售、电商、电信、微管理等不同领域的数据分析引入和应用进行了深入的研讨和交流。研讨中，邹东生会长表示，大数据时代来临，数据分析行业的机遇与挑战随之而来，我们的思维和方法也必须跟上形势，坚持“技术+咨询”这一发展模式。

协会数据中心主任何林先生对当前行业形势作了简要说明，深刻阐述了数据平台搭建与数据咨询业务技能对事务所发展的重要性，并为大家演示了协会正在开发中的技术平台操作。中颢润（北京）项目数据分析师事务所的王庆生先生表示：面对大数据时代的机遇与挑战，事务所要首先明确自身的市场定位、需求定位和目标客户定位，不断增强团队实力储备，同时加强品牌宣传和客户需求的深度挖掘，从客户角度出发，才能逐步赢得客户的认可。

会后，部分参会人员表示，此次会议明确了事务所未来

发展的方向，专家们从实际案例出发，讲解深入浅出，研讨热烈务实，交流中碰撞出很多新思路 and 火花，对未来发展更有信心。

研讨会观点提要

关于“技术+咨询”商业模式的探讨

在数据大爆炸的今天，收集数据并进行有效分析，从而指导企业决策，帮助企业成功发展，这是大数据的核心价值所在。然而，自大数据概念提出以来，一直活跃在大数据领域的并非数据分析师，而是一些IT人士，甚至很多人认为大数据行业就是IT行业。之所以会有这样的认知偏差，是由国内企业数据化程度普遍较低和对大数据核心价值认知模糊决定的。基于这样的现状，我会在此次研讨会上明确提出了“大数据1.0时代”事务所应该坚持“技术+咨询”的商业模式。

数据分析咨询能力和技术手段是从事数据分析行业的两个必备武器。首先要具有技术能力，这是实现数据收集、存储、结果展示的工具和手段；二是要具备数据分析咨询能力，因为只有数据分析才是真正能够为客户创造价值的核心。二者相辅相成，缺一不可。在数据分析行业的不同发展阶段，这两者的侧重点会有所不同。在欧美等发达国家，数据分析行业已经发展得较为成熟，咨询能力在两者中占有主导作用。

然而在我国，数据分析行业发展不过十余年的时间，严格说还处于数据分析行业的初级阶段。电商、互联网、零售、金融这些天然具有数据资源优势的行业，已经有计划地完成了从数据收集到数据分析的体系建立，但是对于国内多数企业，尤其是中小企业来说，虽然已经意识到数据的重要性，但由于数据缺乏以及从业人员自身原因，这些企业还远远没有完成数据化建立。因此在这个时期，首先帮助企业构建数据化平台变得极为重要，这期间必然会涉及IT技术的应





研讨会事务所代表合影

用。但我们应该清楚地认识到，IT技术只是实现数据收集、数据挖掘的工具或者手段，它并不能直接指导企业决策，真正发挥作用的是数据分析能力，它可以帮助企业发现数据之间的关系，从而指导企业做出正确地决策行为，实现企业盈利目标。

“技术+咨询”既是国际上流行的商业模式，也是国内事务所将数据分析业务向精深化发展过程中必然采用的商业模式。在本次研讨会上，我们看到一些优秀事务所已经在按此种模式运作，并且已经取得了令人欣喜的成果。协会作为行业的领导者，在推进整个数据分析行业发展以及指导事务所开拓业务方面具有不可替代的作用。目前协会正在研发的Datahoop大数据智能分析平台，正是基于这样的初衷，希望通过先进技术的使用，帮助事务所在推动新商业模式过程中遇到技术壁垒而停滞不前时，可以顺利跨过技术难关，将业务重点转移至数据分析咨询服务。相信随着数据分析行业的不断发展，“以咨询业务为主导，以技术为展现工具”的商业模式会被越来越多的事务所采用。

场的嘉宾均来自各行各业中关注数据分析的个中翘楚。

本次活动主旨为“如何用数据创造财富”。三位数据大牛分别带来了业界经验与大家进行交流：猎聘网CDO单艺进行了题为“大数据的价值-岂止于大”的演讲，品友互动CTO沈学华就“大数据时代的精准营销-实时竞价广告上的实践”进行了阐述，快的打车技术副总裁朱磊进行了关于“产业互联网时代的O2O核心-快的大数据的应用”的分享。整个过程中，与会人员反响热烈，纷纷进行互动提问。协会作为特约嘉宾在现场进行了品牌展示，并派发了数据分析行业年度发展报告、会员专刊、峰会特刊等内部刊物，从火爆的领取场面不难看出大家对数据分析行业有着极高的兴趣及关注度。活动圆满结束后，协会及猎聘网相关负责人就数据分析行业人才对接问题进行了更深层次的商谈。

作为中国数据分析行业唯一行业协会，协会一直致力于

2、共建大数据人才

——暨协会参加猎聘网活动实录

2014年12月20日，中国商业联合会数据分析专业委员会（以下简称“协会”）作为特别嘉宾参与了由猎聘网在北京天伦王朝酒店主办的“同道精英汇之大数据时代”主题活动。猎聘网是国内知名的中高端人才招聘网站，本次活动到



猎聘网活动



猎聘网活动

促进业界内外交流、专业人才培养、数据分析技术研发和推广。数据分析行业在中国发展的11年间，协会已培养项目数据分析师近两万名，组建专业事务所百余家，并组织开展了各类型推广活动，包括数据分析行业峰会、项目数据分析师事务所专题研讨会、CPDA课改说明会、项目数据分析师创业指导活动以及数据分析公益沙龙活动等。

随着大数据概念的广为传播，越来越多的企业认识到数据分析在决策体系中的重要性，对数据分析人才的需求也呈现高速增长态势。如何培养合格的数据分析人才并输送至企业，是目前面临的主要问题。只有行业协会、猎头公司、企业、人才四者达成有效联合，才能使数据分析行业在中国更为迅速地发展壮大。2015年，协会将继续与各行各业的企事业单位深入合作，提供人才培养和输送、技术咨询与指导等服务，同时促进行业行为规范、推进项目数据分析师事务所建立，进而推动数据分析行业在中国的整体发展，建立新型的服务型协会！

3、中数委携手百度 共同推进数据分析行业发展

为更好的推进数据分析行业的发展，提高中国数据分析行业品牌的认知度，中国商业联合会数据分析专业委员会

CPDA 项目数据分析师——数据分析行业专业品牌

12年数据分析行业的发展历程：全国15家卓越的人才培养基地；近40名专业数据分析人员组成的师资研发团队；全国100家专业项目数据分析师事务所；全国20000名专业人才队伍服务于各行各业。全面打造数据分析行业专业高端人才！

www.chinacpda.com 2014-11 | 品牌推广

■ 数据分析-职场新利器
■ 项目数据分析师-为企业决策作出决策依据

CPDA认证	行业资质培训	优秀事务所	数据分析服务	热门城市
课程介绍	精准营销	行业介绍	数据掘金	北京
师资介绍	数据挖掘与BI	11年优秀名单	技术支持	上海
学习形式	电商数据分析	12年优秀名单	技术产品	广东
从业介绍	实战实战课程	13年优秀名单	咨询服务	

(以下简称“协会”)与百度就共同促进数据分析行业发展以及数据分析人才培养达成了战略合作。除此之外，百度还为协会量身定制了项目数据分析师(CPDA)品牌专区。

百度品牌专区位于百度网页搜索结果最上方，是专门为著名品牌量身定制的资讯发布平台，是为提升用户搜索体验而整合文字、图片、视频等多种展现结果的创新搜索模式。在品牌展示区上，项目数据分析师(CPDA)官网的丰富资讯将以精选和更为直接的方式展现在用户面前，众多用户也得以更快的了解项目数据分析师的最新信息，更方便的获取考试培训信息，实现了品牌展现与搜索之间的完美融合。

同时在品牌专区设立了专门的展示版块，以协助事务所提升其社会影响力，帮助其在大数据发展机遇中跟上时代步伐，开辟更多的业务领域

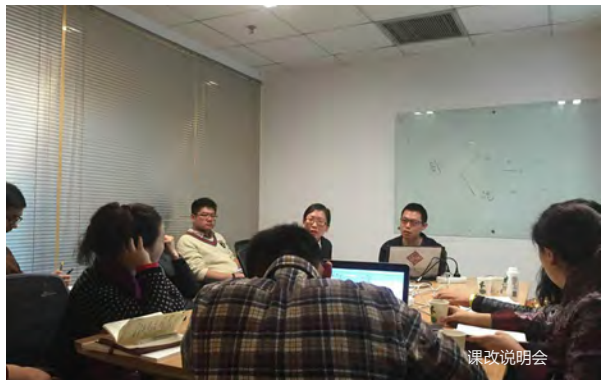
协会一直致力于数据分析行业发展——培养专业人才、支持事务所的发展。为了让更多的人了解数据分析行业，了解项目数据分析师，百度也将积极支持并参与由协会组织的一些线下活动，形成品牌互动，进一步推动整个数据分析行业的发展。

4、将专业进行到底

——暨2014年CPDA课改说明会

2014年12月12日，在中国商业联合会数据分析专业委员会(以下简称“协会”)的会议室中，举行了一场别开生面的项目数据分析师(CPDA)课改说明会，会上就课程的修改方案进行了深度解析。

此次会议由协会组织召开，邀请来自各领域的数据分析专家参会。会议首先由课程编委会成员康琳老师对课改情况做了整体性说明。随后与会的各位老师对课改重点进行了热烈讨论。尤其是协会数据中心何林老师讲到数据工具板块



课改说明会

时，大家尤为热烈。

“项目数据分析师”（CPDA）是协会自主研发的数据分析行业的专业培训课程，旨在为社会培养并输送具有一定专业技能的数据分析人才。该课程自2003年在中国正式开办以来，已经在全国二十多个省市培养了近万名项目数据分析师。目前已具有一套成熟的课程体系及与之匹配的专业教材、远程学习平台及完善的学员服务体系。同时，协会本着专业负责的态度，对CPDA课程不断进行优化，使之更加切合时代发展的需要。

本次课改是自2003年以来的第三次大规模课改，主要在

实效性、前沿工具、知识体系完善以及体系内部的关联紧密性方面做了重点调整，充分扩充“客户数据分析”、“市场分析与预测”两大模块，结合最新的技术，使课程内容更加适应社会发展需求，更加贴近生活，易于应用。

本次课改的成果已在2014年12月27日北京地区第56期面授课中进行展现，并取得了良好的效果，得到了师生们的一致好评。

新版课程内容 NEW		旧版课程内容	
目录	课程概要	目录	课程概要
1. 数据分析概述	<ul style="list-style-type: none"> 数据与数据分析的发展脉络：从小数据到大数据 数据分析的应用领域-数据分析产业链 数据分析人才培养体系 CPDA认证考培系统介绍 	1. 数据分析概述	<ul style="list-style-type: none"> 什么是数据分析？ 数据分析以及在国内外发展现状 数据分析对企业经营和项目投资的重要作用 数据的品质 项目数据分析师未来职业优势和机会 项目数据分析师认证培训课程体系 and 考试方式
2. 战略管理 (企业战略管理的模型与方法)	<ul style="list-style-type: none"> 战略输入 战略匹配 战略决策 	2. 战略管理	<ul style="list-style-type: none"> 企业战略管理是什么？ 战略管理对企业重要的意义 企业战略框架 企业战略管理的模型与方法
3. 数据产生	<ul style="list-style-type: none"> 什么是数据 数据的类型 数据分析流程与建模思想 大数据生成 传统数据采集 	3. 数据采集	<ul style="list-style-type: none"> 什么是数据？数据有哪些类型？ 什么是数据采集？数据采集有哪些作用？ 如何进行数据采集？
4. 数据导入与预处理	<ul style="list-style-type: none"> 大数据导入 传统数据导入 数据清洗 数据加工 	4. 数据处理	<ul style="list-style-type: none"> 什么是数据处理？ 怎样进行数据录入？ 如何进行数据清洗？ 数据的分析前加工
5. 数据分析	<ul style="list-style-type: none"> 远程预习：R语言操作基础 数据分析方法 描述型数据分析 	5. 基础数据分析	<ul style="list-style-type: none"> 基础数据分析方法 对比分析 集中度分析 离散性分析 分布分析 交叉分析
6. 数据挖掘	<ul style="list-style-type: none"> 挖掘概述 Kmeans 购物篮 决策树 朴素贝叶斯 	6. 预测数据分析	<ul style="list-style-type: none"> 什么是市场预测？ 为什么进行市场预测？ 如何进行市场预测？ 定性预测方法 定量预测方法
7. 数据展示	<ul style="list-style-type: none"> 定性数据图标展示 定量数据图标展示 数据展示技巧 	7. 营销决策	<ul style="list-style-type: none"> 什么是营销决策？ 客户数据分析 企业品牌研究 营销组合决策——新产品开发决策 营销组合决策——价格决策 营销组合决策——促销决策 营销组合决策——渠道决策

新版课程内容 NEW		旧版课程内容	
目录	课程概要	目录	课程概要
8.营销决策	<ul style="list-style-type: none"> 营销决策的理论背景 营销决策的实务思路 	8. 生产运营决策	<ul style="list-style-type: none"> 生产管理的内容和数据分析需求 生产计划和产品混合决策 分厂选择模型 经济生产量模型EPQ 设备最优化使用寿命决策 经验曲线模型
9.市场分析 with 预测	<ul style="list-style-type: none"> 市场分析的研究领域 什么是市场预测 为什么进行市场预测 如何进行市场预测 预测方法概述 定量预测方法详解—回归分析与预测 定量预测方法详解—时间序列分析与预测 	9. 采购优化决策	<ul style="list-style-type: none"> 采购决策的5R体系 供应链管理决策 最有采购量EOQ模型 物流需求计划MRO模型 采购优化模型ABC分析法 供应商选择和评估（层次分析法）
10.市场细分 with 客户数据分析	<ul style="list-style-type: none"> 客户数据分析概述 客户细分与市场定位 客户细分方法与模型选择 客户细分技术：聚类分析详解 客户细分技术：主成分分析与因子分析详解 客户数据分析综合案例：从问卷到结论 	10. 数据报表编制	<ul style="list-style-type: none"> 报表编制的基础数据 总投资构成及估算 总成本费用构成及估算 收益构成及估算 现金流量表的编制
11.营销组合数据分析	<ul style="list-style-type: none"> 营销组合决策的意义与基本流程 产品策略 价格策略 促销策略 渠道策略 	11. 经济评价指标分析	<ul style="list-style-type: none"> 为什么要对项目进行经济指标评价？ 静态评价指标 动态评价指标
12.生产采购决策 with 库存优化	<ul style="list-style-type: none"> 生产与采购决策概述 分厂生产策略 生产计划与产品混合决策 供应链管理决策 库存管理决策 供应商决策 生产质量控制 交货期管理 	12. 投资风险分析	<ul style="list-style-type: none"> 不确定性与风险分析 项目盈亏平衡分析 投资收益风险敏感性分析 风险概率分析
13. 投资决策概述	<ul style="list-style-type: none"> 量化投资相关概念 量化投资框架 量化投资原则 现金流量为基准的原则 	13. 投资选择与评估	<ul style="list-style-type: none"> 资本限量决策 公司价值分析
14. 项目投资经济收益分析	<ul style="list-style-type: none"> 现金流量表的解读 经济评价 投资估算 资金构成 成本估算 收入预测 	14. 量化经营实战案例	
15. 投资风险分析	<ul style="list-style-type: none"> 项目不确定性与风险分析 项目盈亏平衡分析 项目敏感性分析 项目风险概率分析 	15. 数据分析项目流程及工作方法	
16. 投资选择与评估	<ul style="list-style-type: none"> 独立项目决策 互斥项目决策 混合项目决策 	16. 数据分析技术指导	
17. 数据分析项目流程及工作方法	<ul style="list-style-type: none"> 项目建议书撰写 项目投资计划书 		
18. 复习课	<ul style="list-style-type: none"> 知识串讲 实操模拟题 		

5、新版微信重磅上线

为了搭建更为宽阔的行业交流平台，为广大粉丝提供详细、准确、全面的行业资讯及课程信息，我们对官方微信进行了全新调整，新栏目的设置更多考虑到服务用户的功能，相信改版后的“CPDA数据分析天地”能给大家带来全新的体验。

改版后的内容主要分为三大板块：

【数·行业】：在这里您可以了解到数据分析行业最前沿的资讯，掌握数据分析行业整体发展动态。同时，此版块中我们将对优秀事务所进行专题展示，未来也可对优秀个人会员进行宣传。因此希望业内同仁积极参与此版块的构建工作，借此窗口进行风采展示。

【数·课堂】：为大家提供完整的人才培养计划，让您在数据分析行业更为专业。目前包含了CPDA认证课程和犀数精讲课程，整个培养体系我们还在不断完善中。同时引人关注的公益沙龙活动也会及时在此版块中公布，欢迎大家积极参与。

【数·应用】：为大家准备了行业内常用软件及各行业应用数据分析的案例，希望对大家在实际应用中有所帮助。同时，由协会自主研发的技术平台也将在此栏目中展现。

改版不仅是为了更好的展现协会风采及动态，更重要的是为数据分析爱好者们提供更为专业的服务，这也是我们迈向服务型协会的第一步。在大数据时代的背景下，数据分析不再只是一门专业技术，它更是我们生活和工作的必备能力，所以我们衷心地希望更多的人士加入数据分析行业大军，同时，无论是对行业还是对协会，我们都欢迎您提出宝贵意见与建议！

6、数据分析软件应用公益沙龙活动完美落幕

2014年12月13日，中国商业联合会数据分析专业委员会（以下简称：协会）在中国科学院国家科学图书馆内成功组织举办了今年第五次数据分析公益沙龙活动。本次沙龙活动主题为主题为数据分析主要软件应用。不同以往，本次主题不再是一如既往的针对某一个行业的数据分析，而是数据分析师们想要深入学习的数据分析软件的使用场景及应用方法。所以虽然是隆冬季节，依旧吸引了大批的数据分析爱好者热情高涨的前来学习探讨。

本次沙龙活动特别邀请了高级数据分析师、软件架构师、现任协会数据中心主任的何林老师与大家分享。何林老师拥有10年的数据分析与R语言编程经验，并主持过国家



级科研项目，参与了国家基因组计划对基因序列进行数据分析与挖掘，服务过新浪网、美国富达投资集团、阿里巴巴集团、银河证券、光大证券、联想集团等单位。渊博的知识积累及丰富的实战经验使得何老师对数据分析软件的应用有着自己独到的见解，所以本次何老师别出心裁的将Excel\SAS\SPSS\Matlab\Stata\SQLR七种数据分析软件比作古龙武侠的“七种武器”来为大家进行了对比讲解。

本次活动由协会培训处王英华老师主持，在简单介绍了数据分析主要软件及使用软件中常遇到的问题后，沙龙分享正式开始。何老师依次讲解了数据分析“兵器谱”及“江湖地位”、七个软件的使用场景和限制以及数据分析师的“修炼”等内容。这种打破常规的宣讲方式一下调动起了数据分析爱好者们的学习兴趣，在分享过程中，大家纷纷拍照记录，热烈探讨，现场出现了一波又一波的互动高潮。这种生动精彩的讲解方式使得数据分析爱好者们更加深刻、客观的认识了数据分析软件的功能作用、优缺点等，并且明确了选择使用软件的方式方法。

两个多小时的活动结束后，依旧有大量学员不愿离开，将何老师团团围住咨询探讨工作中所遇到的问题与困难的解





决方法，何老师都耐心的为大家进行了解答。本年度最后一次公益沙龙活动圆满落幕，协会在2015年将会针对数据分析行业的发展情况及数据分析师的实际需求，开展更多、更新颖、更有针对性的活动。

7、上海地区数据分析公益沙龙成功举办

2014年12月20日，上海CPDA授权管理中心（以下简称：上海管理中心）成功举办了本年度第11期的公益沙龙活动，此次活动的主题为“BI在企业中的应用”。

为了给CPDA学员提供更多的线下交流学习机会，使



其在实际工作中可以更好的运用数据分析，上海CPDA数据分析俱乐部一直坚持每月开展一次不同主题的公益沙龙分享活动。

本次活动特别邀请了有十余年企业信息化和数据分析经验的资深数据分析师王军。王军是上海CPDA第13期学员、项目数据分析师（CPDA）、上海交通大学工商管理硕士，现



任一家美国制造业企业的BI manager，曾在SAP和丹麦马士基物流等知名公司任职。王军老师从制造业企业在实际运营中如何运用数据进行管理和效率提升、企业常用的KPI体系、企业BI系统的架构三个方面系统介绍了BI在企业管理中的角色和作用，并和大家一起探讨了如何才能在企业中更好的发挥数据的作用，让管理者可以根据数据分析更加有效精准的做出决策。

王军老师分享后，来自传统制造业、金融、软件等行业的参会学员纷纷提出了工作中所遇到的问题，与嘉宾老师互动、探讨。同时学员之间也进行了热烈的交流、讨论，并且为了方便更深入的沟通还建立了数据分析公益沙龙微信交流群。

上海管理中心表示，应广大学员及数据分析爱好者的要求，将在2015年增加上海地区数据分析公益沙龙举办的频次，并且会把沙龙主题安排的更加贴合实际，解答大家在实际工作生活中遇到的问题，更希望通过沙龙活动促进数据分析行业在上海地区的蓬勃发展。

FIN

更正声明

2014年第3期 总第十九期《中国数据分析》第12页《数据分析师的利器：R语言》的作者应为何林，因排版失误写成何琳，特此更正。并向作者及广大读者致歉。

数据挖掘经典算法系列之决策树

© 文 / 数据中心主任 何林 图 / 何林 编辑 / 王宏莉

1. 概念

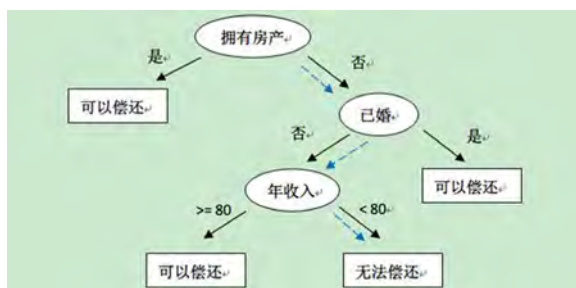
数据挖掘中的决策树又叫做决策树学习，区别于战略管理中的决策树，那个决策树的形状和节点均已知，根据决策树做决策。而数据挖掘中的决策树是未知的，我们的目的就是要通过对历史数据的学习来构建决策树的形状和节点。

数据挖掘中的决策树：从数据产生决策树的机器学习技术叫做决策树学习，通俗说就是决策树。

下面是个颠倒的树：



决策树的形状跟上面那颗颠倒的树很像：



现有的决策树算法及各自特性见下表：

决策树算法简写	全名	输入限制	输出限制	树结构限制	分组合割标准	修剪
C5.0		类型、数值	类型	多叉	信息熵	
CART	分类回归树	类型、数值	类型、数值	二叉	Gini系数和方差	
CHAID	卡方自动交互诊断器	类型、数值	类型、数值	多叉	统计显著性检验	预处理
QUEST	快速无偏有效统计树	类型、数值	类型	二叉	统计显著性检验	不做预处理

本次讲解的是ID3算法，ID3算法是决策树算法的鼻祖，C4.5是ID3的改进，也是最经典的决策树算法，C5.0算法是C4.5算法的商业版。

2. 场景

最典型的应用场景是预测和分类，如新客户分类。

3. 特点

决策树有很好的直观性，生成模式简单，对预测和分类结果也有很好的解释性。

4. 相关概念

熵是无序性（或不确定性）的度量指标。假如事件A的全概率划分是 (A_1, A_2, \dots, A_n) ，

每部分发生的概率是 (p_1, p_2, \dots, p_n) ，那信息熵定义为：

$$entropy(p_1, p_2, \dots, p_n) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 - \dots - p_n \log_2 p_n$$

Information Gain (IG)，信息增益和信息熵描述的信息是一致的；描述的是对于数据集S，将其按照其属性A切分后，获得的信息增益值。注意IG描述的是信息的增益值，当不确定性越大时，信息增益值应该是越小，反之亦然，是负相关的关系。

信息增益的公式如下：

$$\text{Gain}(S, A) \equiv \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} (|S_v| / |S|) \text{Entropy}(S_v)$$

5、算法

下面8条记录是来自某银行信贷部门，姓名做了处理不是真实姓名，如有雷同，纯属巧合。我们根据这个信贷记录构造信贷决策树。

编号	姓名	已婚	年收入 (>=80 千元)	房产	偿还情况
1	李峙锦	是	否	有	可以偿还
2	王政	是	否	有	可以偿还
3	和韵杰	否	否	有	可以偿还
4	杨莹琨	否	否	有	可以偿还
5	权宇豪	是	否	无	可以偿还
6	贺宇翔	否	是	无	可以偿还
7	安龙	否	否	无	无法偿还
8	齐雅格	否	否	无	无法偿还

(1) 目标是偿还情况，先计算“偿还情况”的信息熵：

8条记录，6条可以偿还记录，2条无法偿还记录

$$\text{Entropy}(S) = -(2/8)\log_2(2/8) - (6/8)\log_2(6/8) = 0.811278$$

(2) 计算“房产”属性的信息熵：

8条记录，4条有房产的记录，对应的目标值都是可以偿还记录

$$\text{Entropy}(\text{有房产}) = -(4/4)\log_2(4/4) = 0$$

还有4条无房产记录，对应了2条无法偿还记录和2条可以偿还的记录：

$$\text{Entropy}(\text{无房产}) = -(2/4)\log_2(2/4) - (2/4)\log_2(2/4) = 1$$

计算房产属性的信息熵：

$$\text{Entropy}(\text{房产属性}) = (4/8) * \text{Entropy}(\text{有房产}) + (4/8) * \text{Entropy}(\text{无房产}) = 0 + 1/2 = 1/2$$

$$\text{信息增益 Gain}_{\text{房产}} = \text{Entropy}(S) - \text{Entropy}(\text{房产属性}) = 0.811278 - 1/2 = 0.311278$$

(3) 计算“年收入”属性的信息熵：

8条记录，1条>=80千元的记录，对应的目标值都是可以偿还记录

$$\text{Entropy}(\text{年收入} \geq 80) = -(1/1)\log_2(1/1) = 0$$

还有7条年收入<80千元记录，对应了2条无法偿还记录和5条可以偿还的记录：

$$\text{Entropy}(\text{年收入} < 80) = -(2/7)\log_2(2/7) - (5/7)\log_2(5/7) = 0.863121$$

计算年收入属性的信息熵：

$$\text{Entropy}(\text{年收入}) = (1/8) * \text{Entropy}(\text{年收入} \geq 80) + (7/8) * \text{Entropy}(\text{年收入} < 80) = 0.75523$$

$$\text{信息增益 Gain}_{\text{年收入}} = \text{Entropy}(S) - \text{Entropy}(\text{年收入}) = 0.811278 - 0.75523 = 0.056048$$

(4) 计算“已婚”属性的信息熵：

8条记录，3条已婚的记录，对应的目标值都是可以偿还记录

$$\text{Entropy}(\text{已婚}) = -(3/3)\log_2(3/3) = 0$$

还有5条非已婚记录，对应了2条无法偿还记录和3条可以偿还的记录：

$$\text{Entropy}(\text{非已婚}) = -(2/5)\log_2(2/5) - (3/5)\log_2(3/5) = 0.970951$$

$$\text{信息增益 Gain}_{\text{已婚}} = \text{Entropy}(S) - \text{Entropy}(\text{已婚}) = 0.811278 - 0.606844 = 0.204434$$

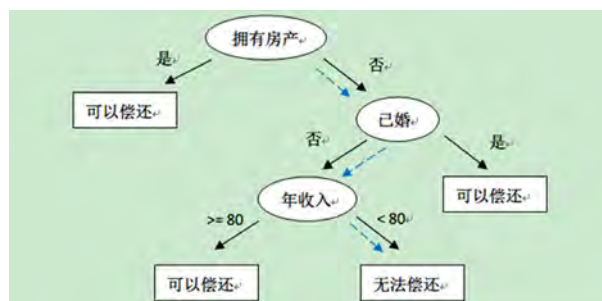
计算已婚属性的信息熵：

$$\text{Entropy}(\text{已婚}) = (3/8) * \text{Entropy}(\text{已婚}) + (5/8) * \text{Entropy}(\text{非已婚}) = 0.606844$$

(5) 根据上面的计算得到：

$$\text{Gain}_{\text{房产}} > \text{Gain}_{\text{已婚}} > \text{Gain}_{\text{年收入}}$$

据此得到决策树如下：



FIN



2014年度 优秀项目数据分析师事务所评选

我国数据分析行业优秀事务所一年一度的评选活动已经圆满结束。优秀事务所评选从2010年开始至今已连续举办五届，旨在通过评优活动彰显优秀事务所的辉煌业绩，号召全国事务所向优秀学习，向优秀挑战，进而超越，同时促进行业内部自检和业务横向交流，达到“以评比找不足，以竞争促发展”的目的。

大数据时代背景下，数据科学快速发展，市场上数据分析的需求也随之有了新的变化。变化既是机会又是挑战。此次评比我们结合事务所运营、业务专业度、团队建设、行业准则及行业参与的五个维度进行评选，同时结合事务所在大数据背景下商务应变能力进行全面综合评估。

此次评选全国各事务所积极踊跃参与，从汇总上来的资料和报告中，我们看到，相较于去年，事务所在业务规模、业务领域、专业程度上均有了大幅提升。大数据时代背景下，我们事务所的经营决策数据分析在业务总量中的比例逐步增大，部分优秀事务所的大数据分析业务已经有了起色，并且前景喜人。分析师培养的投入力度不断加大，行业违规投诉明显减少。这些进步都直接反应了随着大数据时代的到来，数据分析行业蓬勃发展，我们的优秀事务所在新的发展背景下，拥抱变化，迎接挑战，顺应大数据时代的发展趋势，不断提高专业水平，转变经营理念，在市场中胜出。

通过评审专家组的综合评审，有5家优秀事务所从众多参选事务所中脱颖而出，这5家事务所不仅在当地积极带动行业发展并且踊跃扩展事务所业务，取得了丰硕的成果，为行业带来了积极影响，得到了评选评审组的高度认可。

获得2014年度数据分析行业优秀事务所称号的5家项目数据分析师事务所是：

- 安徽量衡项目数据分析师事务所
- 中颢润(北京)项目数据分析师事务所
- 重庆传晟项目数据分析师事务所
- 上海天元项目数据分析师事务所
- 湖南翰林项目数据分析师事务所

安徽量衡项目数据分析师事务所

安徽量衡项目数据分析师事务所是经国家行政管理部核准注册，经中国商业联合会数据分析专业委员会备案（中数委团证第067号）的独立法人机构。事务所现拥有项目数据分析师、注册会计师、注册税务师、物流管理师、

工程师等专业技术人员。业务涉及农业、文化、房产、物流、餐饮、科技、烟草、能源、通讯、金融等国民经济领域，以专业的数据分析服务为核心，致力于为政府机构、企业、国内外银行、风投公司等提供项目投资类数据分析、经营类数据分析、项目数据软件、投资项目评价、市场调研、项目经济效益评价、项目策划、投资咨询、经济信息咨询等专业的、系统的服务，并可为委托方提供可行性报告、商业计划书、项目数据分析报告等多种具有客观性、公正性、实用性的报告。

安徽量衡项目数据分析师事务所本着“专业、专注、专心”的职业态度，根据市场需求，从数据分析的角度为客户做好数据的调研、统计、挖掘、分析等服务。

专家点评:

安徽量衡项目数据分析师事务所作为安徽省第一家事务所，自2011年底成立起，便积极开拓市场，确立中小企业为服务目标，致力于在安徽省内普及数据分析理念、为企业做好经营决策数据分析服务。

安徽量衡项目数据分析师事务所充分洞察到数据分析行业的新变化，敏捷应变，在大数据分析、新媒体（微信）数据分析、经营分析领域积极探索，成果喜人。

安徽量衡提交的数据分析报告业务逻辑清晰，数据结构合理，数据资料较为详实，有足够的支撑，研究论证相对严谨。报告研究论证相对专业度较高，可以使企业客观、系统地了解项目和企业自身的实际情况，帮助企业做出更符合企业长期发展的决策意见。在数据分析深度挖掘上有提升空间。

中颢润（北京）项目数据分析师事务所

中颢润（北京）项目数据分析师事务所是经过北京市工商局登记注册的具有独立法人资格的专业项目数据分析机构，且经中国商业联合会数据分析专业委员会备案核准的常务理事单位（中数委团证号001）。事务所拥有数据采集、数据分析、金融研究、工商管理等方面的复合型人才及专家，聘请大量数据分析行业专家作为资深顾问。事务所运用先进的理念和独到的见解，突破行业的数据分析局限，从决策性数理研究入手，为企业提供专业数据分析和科学决策依据的同时，更精准、全面客观的营销战略成为指导企业规避项目风险、保持强大竞争优势的科学依据。

中颢润拥有包括专业项目数据分析师、资产评估师、注册会计师和专业市场调查人员构成的强大团队，致力于专业数据分析深度服务，坚持投资数据分析报告、经营数据分析报告、决策数据分析报告的权威性、客观性、中立性和实用性，为众多的投融资和企业决策者提供了准确的数据分析成果和决策建议。

专家点评:

中颢润（北京）项目数据分析师事务所作为数据分析行业的常委单位，自成立以来一直坚持以经营决策数据分析业务为主要发展方向，致力于专业数据分析深度服务，坚持投资数据分析报告、经营数据分析报告、决策数据分析报告的权威性、客观性、中立性和实用性。

作为北京地区的项目数据分析师事务所，在面对激烈的市场竞争时，中颢润（北京）项目数据分析师事务所领导人提出要以互联网思维经营事务所，采取微信、微博、线上线下结合等多种新的营销方式，获得了优异的成绩，为全国事务所的经营提供了宝贵的经验，且中颢润事务所在大数据分析领域已有一定的储备，专业实力较强。中颢润的数据分析报告思路清晰，结构严谨，分析细致，数据资料详实，方法得当，论证深入，整体质量较高。为企业做出经营决策提供科学依据。虽然在大数据分析领域有了一定储备，但有待进一步拓展大数据分析业务。

重庆传晟项目数据分析师事务所

重庆传晟项目数据分析师事务所是经过国家行政管理部门核准注册，具有独立法人资格的专业项目数据分析机构，且经中国商业联合会数据分析专业委员会备案核准的事务所会员单位（中数委团证号084）。

重庆传晟项目数据分析师事务所主要致力于房地产政府管理部门全局解决方案、软件开发、咨询服务和档案整理加工，以及房产数据分析统计处理。公司技术实力雄厚，拥有高素质的项目数据分析师团队、软件开发队伍和强大的技术支持力量。主要核心人员从事房地产政府管理部门行业软件开发和咨询工作多年，成功推出了拥有自主知识产权的档案管理系统、数据清理系统、搜索引擎、档案生命周期管理平台、档案整理加工系统、房地产管理部门便捷化服务平台、权证防伪、短信平台等系列产品。公司包括成都和重庆两个研发中心，公司已扎根西南，并不断扩展中。

专家点评:

重庆传晟项目数据分析师事务所一直充分发挥着带动当地项目数据分析行业快速发展的榜样作用,积极拓展业务范围,将企业经营决策数据分析作为事务所发展的长远目标,不断提高自身专业水平和业务能力,获得了当地政府和房地产行业的一致认可。

重庆传晟走的是一条“技术+咨询”的道路,重庆传晟具备较强的信息化技术能力,结合数据分析师事务所得天独厚的数据分析专业实力,这是其独特的核心竞争力。所内从业的数据分析师均具备多年的项目数据分析、软件技术工作经验,秉承以诚为本、客观公正的职业操守,为重庆地区的房地产政府管理部门提供专业、全面、详实、精准的项目数据分析报告。重庆传晟撰写的经营类分析报告宏观分析到位,推论严谨,思路较为清晰,公正客观,可以为需求单位提供最佳决策建议依据。事务所通过自身的实力和专业度获得了良好的行业口碑。

上海天元项目数据分析师事务所

上海天元项目数据分析师事务所是经国家行政管理部门核准注册,经中国商业联合会数据分析专业委员会备案(中数委团证第065号)。事务所是一家专业从事项目数据分析的服务性机构,公司由多位高级项目数据分析师发起成立,并拥有一支集经济研究、金融投资、财务分析、工商管理等多领域的复合型团队。

目前,上海天元项目数据分析师事务所业务网络遍布海内外与全国30多个省市自治区,已与多家国内外金融机构、大型财团、银行、上市公司、商会等建立了业务合作意向及业务往来,业务涉及农业、林业、建材、石油、石化、化工、通信、电子、电力、电网、水利、铁路、民航、煤炭、冶金、轻工、纺织、医药、机械、市政、交通、房地产、卫生、教育、文化娱乐、环保、旅游等诸多行业领域,在为客户提供服务的过程中,秉承诚信、客观、科学、实效、公正的经营理念,坚持高水准的专业品质,为客户提供最有价值的服务,成为客户发展历程中值得信赖的合作伙伴!

专家点评:

上海天元项目数据分析师事务所自成立以来,一直致力为中小型企业、国内外银行、投融资公司、政府组织等机构提供投融资方面专业的、系统的服务,并为项目投

资方以及融资方提供具有经济性、权威性、客观性、公正性、实用性的项目数据分析报告。上海天元撰写的报告内容结构严谨,分析到位。其专业的方法、严谨的作风、客观的态度、公正的原则以及热情的服务,帮助企业做出切实有效的决策。为推动中国项目数据分析事业的发展贡献着自己的力量。

湖南翰林项目数据分析师事务所

湖南翰林项目数据分析师事务所是经国家行政管理部门核准注册,经中国商业联合会数据分析专业委员会备案(中数委团证第080号)的独立法人机构。湖南翰林项目数据分析师事务所是湖南省较早成立规模较大附和资质较齐的项目数据分析师事务所,是一家专业从事项目投融资咨询、各类数据分析及管理咨询的专业机构。

湖南翰林以专业的项目数据分析服务为核心,同时与湖南南方会计师事务所有限公司、湖南新融达咨询(集团)、邵阳南方资产评估有限公司、邵阳南方工程造价咨询有限公司、邵阳科信税务师事务所有限公司、邵阳南方司法鉴定所、湖南邵房房地产评估有限公司等执业联盟单位紧密联合,在会计审计、各类价值评估等业务领域弥补项目数据分析师事务所在传统法定业务上的短板,同时也改善了上述各类中介机构在新兴数据分析业务上的不足,构筑数据分析大平台,编织一张数据分析的大网,拓宽数据分析宽度,增强数据分析的深度,提升企业自主了解自身数据的意识。

专家点评:

湖南翰林项目数据分析师事务所自成立以来,一直注重自身行业品牌及事务所形象的推广宣传,事务所拥有多位企业管理方面的复合型项目数据分析师,在提供科学、专业的数据分析服务时有了最基础的保障。事务所的数据分析报告思路清晰,内容详实,专业度较高,数据资料丰富,市场需求部分研究方法得当,为企业科学决策提供了依据。

FIN



基于兰查斯特模型从价格角度分析 黄山品牌在安徽市场发展态势（节选）

◎ 文 / 安徽量衡项目数据分析师事务所 李勇 编辑 / 王宏莉 图 / 崔峻珩

一、研究背景

（一）安徽省烟草市场发展概况

从销售量来看，安徽省卷烟从2010年的1911035.65百箱上升为2013年的1979623.63百箱，2013年的销售量比2010年、2011年、2012年分别同比增长3.6%、3%、1.2%。增长形势趋于稳定，但其增长绝对值仍然很大，2011年比2010年多增加18082.53百箱，2012年比2011年多增加26989.26百箱，2013年比2012年多增加23516.19百箱。从这种绝对值方面看，安徽省市场还是有待开拓的，至今未达到饱和程度。

从期末库存量占总生产比例来看，2010年至2013年库存量占生产总量的比例总体小于6%，其中2012年和2013年两年均小于5%。

从类别上看，一类烟、二类烟、三类烟的销售量在逐年增加，四类烟、五类烟的销量在逐年减少，这是因为人们生活水平的改变及人们意识的变化导致了人们更加青睐于一类、二类、三类烟。

（二）黄山品牌在安徽市场发展现状

黄山品牌的卷烟全省销量逐年上升，2013年比2010年高出1.28倍，期末库存也越来越多，但销售量占当年生产量的比

例大于96%，且基本稳定，说明生产的数量是合理的。从整体看，2010年至2013年的销售额逐年上升，从单位销量的销售额及单位销量的毛利看，也是逐年增加，说明黄山品牌烟的销量前景美好，最近几年销售盈利仍会增加。

指标	2010	2011	2012	2013
销售量	895296.22	1005854.39	1071140.73	1144226.9
期末库存量	25523.06	35176.07	35842.15	45422.29
生产量	920819.28	1041030.46	1106982.88	1189649.2
销售量占生产量的比例	0.97	0.97	0.97	0.96
销售金额	1.9145E+10	2.26E+10	25432710180	2.832E+10
单位销量的销售额	21383.43	22468.67	23743.57	24749.05
毛利	4275669740	5059232579	5396459952	5.896E+09
单位销量的毛利	4775.7	5029.79	5038.05	5153.15

注：1. 本文图表中涉及香烟数量的单位均为“箱”，即“5万支”；
2. 本文图表中涉及香烟销售额、毛利的单位均为“万元”；
3. 本文图表中涉及香烟价格的单位均为“元/条”。

表1.2.1 黄山品牌2010-2013年销售状况

从全省卷烟销量看，全省卷烟销量和黄山品牌卷烟的销售数量是递增的，并且黄山卷烟占有的市场份额逐年递增的。从2010年市场份额的46.8%、2011年的50%到2013年增加为57.8%。2013年的市场份额比2010年同比增加23.4%，体现了

黄山卷烟在安徽省的知名度越来越高，受到了广大烟民的喜爱以及买烟者的认同。

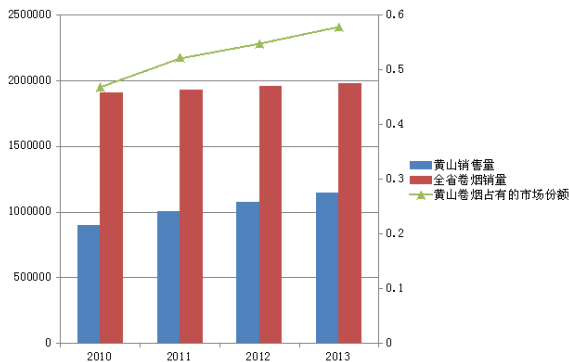


图1.2.1 全省黄山品牌2010-2013年市场前景

二、兰查斯特模型

(一) 概念

“兰查斯特”是英国第一部汽车的设计者，出于兴趣，他对第一次世界大战中交战双方的兵力和折损量进行了分析，提出了著名的“兰查斯特法则”，该法则认为，两个竞争对手之间的差距越大，优势方的损失就越小。

“兰查斯特法则”的提出，是基于概率分析的基础，在多个竞争对手之间进行对比。“兰查斯特法则”认为品牌的市场占有率有三个支点，分别是：品牌独占点（73.9%）、品牌安全点（41.7%）、品牌立足点（26.1%）、即只有当一个品牌的市场占有率超过26.1%，该品牌才有可能从众多强敌环境中脱颖而出，成为未来市场的领导品牌。

(二) 兰查斯特战略模式导出的市场占有率目标

品牌处于竞争中的什么位置，由兰查斯特战略模式导出的市场占有率目标值可以做出判断。

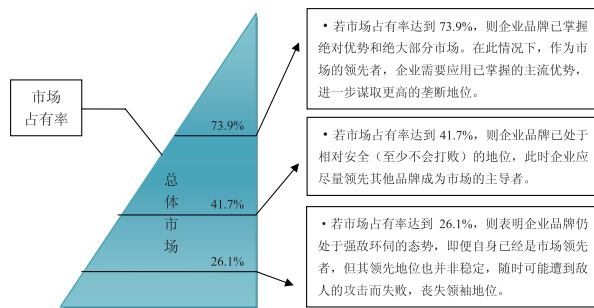


图2.2.1 兰查斯特模型

1) 上限目标值——73.9%

此为绝对优势的独占状态，此时企业在竞争中已经处于绝对安全。

2) 安定目标值——41.7%

在市场竞争中，如果有三家以上公司竞争时，只要谁先取得41.7%的市场占有率，就可以超越其他竞争者，处于优势的位置，不仅成为业界主流，而且很快就能遥遥领先。

3) 下限目标值——26.1%

某企业销售成绩虽然排第一，但在市场竞争中，其位置也不一定是稳定的，随时有被超越的可能。而稳定和不确定可以用26.1%这个数值作为衡量的标准，超过26.1%则表示有可能从势均力敌中脱颖而出形成领先的地位。

三、从价格角度看黄山品牌在安徽重点市场发展态势

首先选出全省卷烟销量最大的前20名，经过计算发现，前20名的总销量占据了香烟市场总销售额的73.69%。

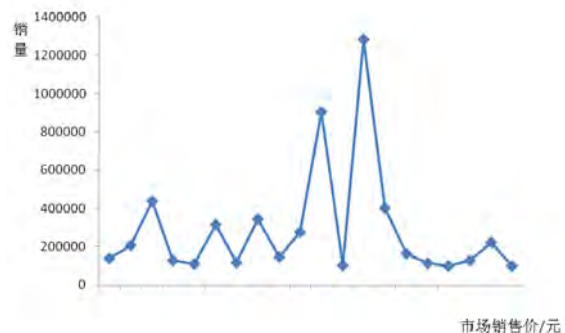


图3.1.1 全国销量前20名香烟的市场零售价与销量关系折线图

从图3.1.1中我们可以看出香烟销量随价格呈双峰分布，其中销量最大的是零售价为50元/条的四类黄山一品（新）硬盒香烟，其次是市场零售价为130的二类黄山牌（新制皖烟）硬盒。由此可知当前消费者的主消费价位在50元-130元/条之间。

全国香烟销量前二十名中黄山品牌占据了其中的9位，除销量第三的不是黄山品牌外，前七强均由黄山品牌占领，黄山品牌香烟占前二十强总数的52.94%。对比不同类别香烟的销量，发现三、四类香烟也是最受消费者欢迎的香烟（见表3.1.1）。

类别	销量	百分比 (%)
1类	1102600	14.18
2类	1304987	16.78
3类	2160430	27.78
4类	2104420	27.06
5类	1103448	14.19

表3.1.1 不同类别香烟的销量及百分比

从图3.1.1可以看出，市场零售价最低和最高的分别为20元/条和450元/条，其销售量几乎均处于谷底，所以为了消除极端值带来的影响，在求平均值时我们直接将其剔除，于是得到均值为95元/条的市场零售价，这说明香烟市场主要还是以中

低档的香烟消费为主导，这与我国橄榄型经济现状有着密不可分的关系，而黄山品牌在中低档香烟的市场上已处于相对安全的地位。

但是居民消费水平也并非一成不变的，近几年普通居民消费水平有了明显的提高，势必会影响居民对香烟价位的选择。为此笔者对2010-2013年安徽省的年销量做出了分析，提取各年份销量最高的前20种不同规格的香烟，对这些香烟市场的零售价格进行合并，随后做出折线图进行对比分析（如图3.1.2）。

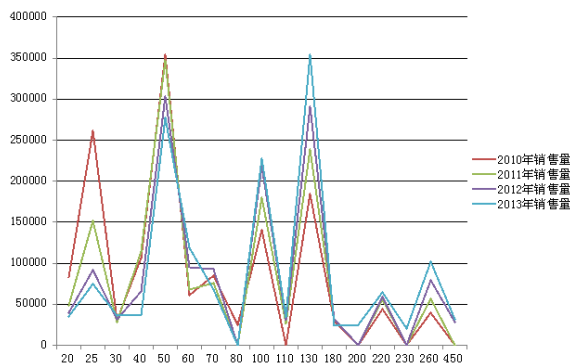


图3.1.2 2010-2013年销量前20名香烟市场零售价格折线图

从图3.1.2中我们可以看到销售价为25元/条香烟的销量自2010年后已经不景气；50元/条香烟的销量一直是一个高峰，但其数值在逐年下降；130元/条香烟的销量在逐年攀升，甚至在2013年超越50元/条香烟的销量成为新高；而100元/条香烟的销量上升趋势强劲，并且有替代50元/条成为第二个高峰值的趋势。这意味着居民对低档香烟的需求减小，对中高档香烟需求在逐渐增大。对比2010年和2013年销量最大的香烟价位，可以发现从25元/条到130元/条有着很大的提升。价位在100到130元/条之间的香烟是目前吸金量最大的。

市场零售价为500-1000元/条的有65种规格的香烟，而其销量仅占据市场总销量的0.89%，这也进一步说明了居民需求水平总体还是以中低档消费为主。在这65种香烟规格中，销量前6名的总销量就占据了高级香烟总销量的90%，也就是说以下六种规格的香烟占了绝大部分高级香烟的销售市场（见表3.1.2）。

在高一级香烟中，中华品牌的中华软包所占份额最大，占高一级香烟的市场的72.10%。黄山品牌价位在500-1000元/条的香烟有六种规格，累计销量占高一级香烟总销量的14.30%。一般而言，销售价格高的香烟，其利润空间也较大，故黄山香烟的规模虽大，效益却不明显。

规格	市场零售价	销量	百分比
中华软盒	650	49698.71	72.10%
黄山(经典皖烟)软盒	600	5883.98	8.54%
黄山新视界硬盒	800	2533.02	3.67%
南京(九五)	990	1643.26	2.38%
黄鹤楼论道74S软盒	500	1159.89	1.68%
黄鹤楼(1916)70S软盒	990	1106.58	1.61%

表3.1.2 高一级香烟销量前六名信息表

根据兰查斯特模型，中华品牌的香烟在高一级香烟市场上已经让顾客对其产生了老大心态，而黄山品牌香烟在高一级香烟市场上连最基本的立足点26.1%都不到，基本无法与中华香烟相抗衡，故其应先从除中华以外的其他品牌香烟中争取到基本的26.1个百分点，再逐渐扩大其在高一级香烟市场中的占有率。而市场零售价为600元/条的黄山（经典皖烟）软盒的销量最好，占高一级香烟市场销量的8.54%，也是仅次于中华软包的销量，这无疑给黄山品牌香烟在扩大高一级香烟的市场占有率方面提供了有力的基点。通过图3.1.2我们可以看出，仅存一个名次差距的第一名和第二名市场零售价也仅差50元，但二者销量却有着极大的差距。这说明高一级香烟销售市场中，各规格的香烟所占份额的差别并不能完全由价格来解释，可能存在品牌、宣传方式等影响。

将不同品牌，不同规格但价格相同的香烟销量进行合并，得到不同价格下各香烟的销量，计算这些销量分别占高一级香烟总销量的对比图3.1.3。

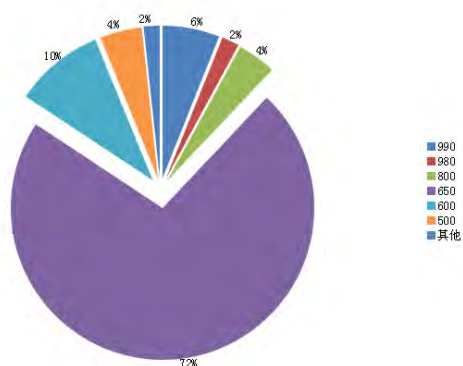


图3.1.3 不同价格香烟销量所占高一级香烟销量百分比

图3.1.3中的结果与前面图3.1.2中第一、二名的结果一致，但第三名的价格发生了变化，从800元/条变成了990元/条。这是由高一级香烟市场上零售价为990元/条的香烟种类以及规格众多所造成的。其中销量最高的价格为650元/条的香烟在高一级香烟市场上有2种，销量第二的600元/条有7种，销量第三的990元/条有17种之多。但这17种销售总量也只是险胜价格为800元/条的黄山新视界硬盒的销量。

由此可见，虽然是高一级香烟市场，但香烟销量并没有随

价格上升而上升，而是停留在高一等级香烟的低价位上。而黄山品牌香烟在高一等级香烟市场上也还有着十分巨大的进步空间，其最强大的竞争对手便是销售价为650元/条的中华品牌的中华软盒。

将黄山品牌香烟的各市场零售价及销量绘制成雷达图（见图3.1.4），我们看到图中有三个明显的离散点，点离中心的距离越远则表示该价位的销量越大。我们可以看出销量最大的是零售价为50元/条的香烟，其次是130元/条，再是100元/条，这与前面得出的市场吸金价位相吻合。而图中其他点的离散程度却不高，这也进一步验证了黄山品牌中高档香烟的市场还有待扩展。

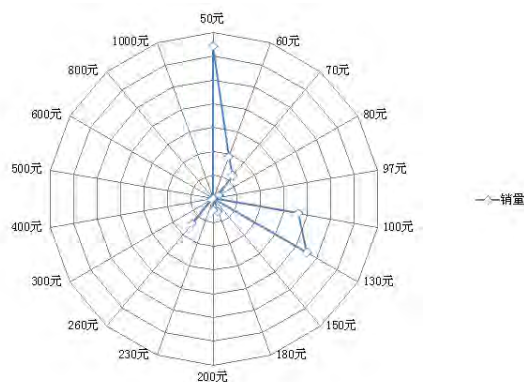


图3.1.4 黄山品牌各价格香烟销量雷达图

最后，参照合肥通过烟草价格对香烟的分类标准，笔者制作了安徽省内各品牌香烟的零售价格的频率直方图（见图3.1.5）。

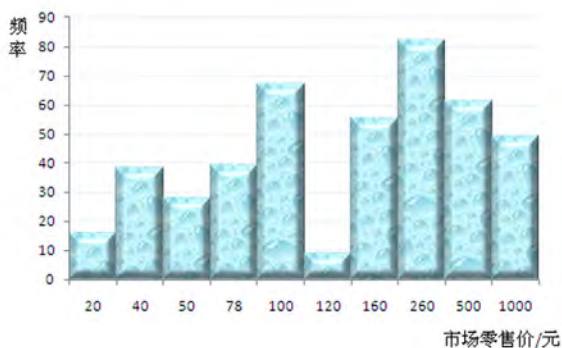


图3.1.5 安徽省各品牌香烟市场零售价频率分布直方图

通过其与黄山品牌香烟不同价位的规格种数的折线图（见图3.1.6）对比，发现在安徽烟草市场上100元/条以上的香烟规格种类繁多，而黄山品牌的规格种类主要还是集中于200元/条以下，这对黄山品牌香烟在拓展高一等级香烟市场的方面也有着一定的参考作用。

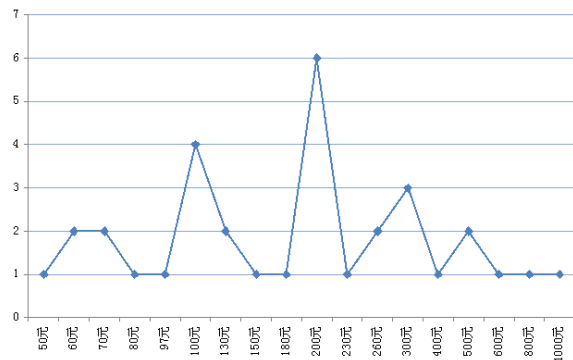


图3.1.6 黄山品牌香烟不同零售价的规格种数折线图

四、小结

最后，对全文做一个简单的总结。

黄山品牌香烟在安徽特别是合肥市场的市场占有率超过了所有其他品牌的香烟，并且达到了一个相对安全的比例52.94%，故黄山品牌应及时认清并利用自己的优势，努力超越其他品牌成为主导者。

在各类别香烟中，黄山在中低档香烟市场的销售率远高于其在高级市场上的销售率，这给了黄山品牌香烟一个警示：虽然规模大，但利润空间不够，效益不足。同时，在高一等级香烟市场上不同品牌、规格的香烟琳琅满目，而黄山品牌在高一等级香烟市场上的规格却还不够。消费者对香烟价格的偏爱也在逐步上升，这提醒黄山品牌在今后的营销生产过程中应当及时调整不同价格香烟的生产规模。

FIN





盘点2014年掀起大数据热潮的五个黄金案例

编辑 / 段欢欢 图 / 崔峻珩



借助大数据的力量进行巫术般地精准营销，爆红的美剧《纸牌屋》将大数据引入了普通人的视野。大数据无疑是当下各领域最热的讨论，简而言之，从不同类型的数据中，快速获得有价值信息的能力，就是大数据技术。年底将至，让我们一起盘点下今年的几大生活数据事件吧！

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一组大小仅为10KB的数据即1000个骚扰号码，随地域、身份以及骚扰趋势的变化而变化。



图一：360手机卫士骚扰电话实时数据

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。

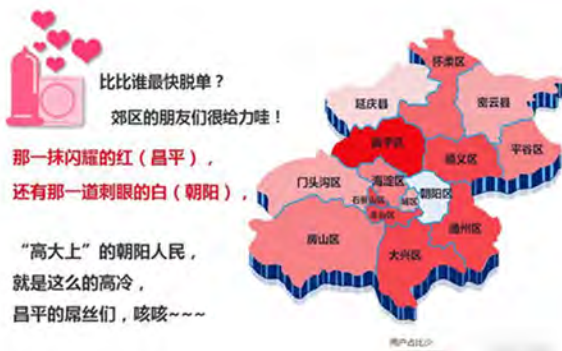


图二：支付宝发布十年对账单

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；激情无限的同时，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。



图三：京东双十一脱单大数据

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

是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。



图四：“37次想你”情侣小心机数据

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页面作为电影《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人知的“小心机”。

IBM社交大数据技术：“上天台”请留步



图五：IBM利用社交大数据预测世界杯排名

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。

FIN

浅谈数据安全

编辑 / 王宏莉 图 / 崔峻珩

大数据时代带来了数据量、数据速率和数据种类的显著增长，随着大数据在不同的领域蔓延，数据安全问题受到越来越多的关注。

数据可以通过不同的系统、应用和环境，从一端流向另一端。这种数据爆炸对业务发展洞察力提供了有意义的帮助，但它也把商业数据暴露给了各种系统、流程和相关人员。而通过某个软件实现端至端的安全保护又是非常昂贵的，况且随着IT技术的不断发展，庞大的数据量在不同的协作系统进行存储、处理、分析，总会存在一些安全漏洞，总会有一些突破安全防护的隐患存在。

数据存在于生活中的各个领域，我们可以从不同的源和不同类型的商业智能工具将数据采集出来并用于分析，最终获得有意义的信息。数据还可以用于协作，而用于协作和处理的工具也有安全性限制，所以，总有暴露敏感数据和内容的概率。一旦大数据的值元素被确定，它就可以被访问、更新、甚至由用户改变，这可能会对企业和个人数据安全都造成严重的威胁。

数据安全和所有人都息息相关，在此我们仅讨论涉及大数据的企业的数据安全问题。大数据企业如通讯、大型电商等企业涉及的数据和客户个人信息都有严密的联系，因此需要在平衡业务需求与数据安全防护之间做到更加精确，对数据安全的保护措施更应全面。

针对以上情况，在此对数据的保护提供一些建议：

1、将大数据分割成小数据：以这种方式，系统将能够更好地处理数据的数量、速度和种类。其结果，企业也将能

够更快和更准确地进行商业决策。



2、识别信息的适用范围：企业需要识别参与此次合作的员工、合作伙伴、供应商或其他第三方，另外也要识别沟通渠道。这有助于给出关于合作环境和利益相关者的详细思路。

3、部署数据控件：将数据控件部署在非常重要的战略位置，这将确保数据的保护与协作。

4、在云计算和移动环境的控制部署：云计算和移动合作是所有应用程序及其部署的重要组成部分，也是风险最高的区域之一。企业需要了解和识别数据是如何在云计算和移动环境中实现共享并在重要位置进行控件部署。

在过去几年中，大多数企业采用单一的软件供应商和单个数据库(SAP, Oracle, PeopleSoft等)为企业服务。此种方式更易于管理但安全问题更为凸显。随着数据



化的发展，系统中安全问题越来越明显，出现安全漏洞的可能性更高。

现在用于大数据安全的工具有很多，我们在此只讨论“处于网络内部”的安全工具及策略。

1. 监控和记录：监控和记录是检测未授权活动的最佳策略之一。目前常见的有系统日志(Linux)、事件日志(Windows)、SNMP（网络事件记录）及可供日志汇总的软件包。SIEM包（安全信息和事件管理软件包）的主要功能是自动检测未经授权的活动，并产生警告。但是，所有SIEM软件需要配置才能正常工作。建议使用预配置SIEM包要时常更新他们，以便通过日志分析，找出安全漏洞。

2. 身份管理：身份和访问管理(IAM)对于大数据保护来说是非常重要的，因为数据是通过使用不同的通道被访问，包括移动设备，SAAS模式或其他服务。身份可以确定是谁正在对敏感数据进行访问。“身份”作为新的安全尺度是非常重要的，将有助于我们应对周边发生的故障。

3. 掩蔽数据：数据掩蔽是保护数据安全的另一种方式。一些数据可以通过加密或断词被屏蔽。

4. 应用安全：最后一步是确保访问敏感信息的大数据的应用安全。要考虑的主要因素包括“权限粒度级别”和“数据加密”。

在当今世界，数据安全问题越来越被重视，早期的大部分流行工具构建时并没有考虑安全因素，随着数据安全问题日益凸显，大数据工具构建者已开始针对数据安全问题进行了不同程度的改善。

正如我们所知道的大数据系统并不像普通单一的供应商系统，因此安全问题的处理更加复杂。目前还没有单一的解决方案、工具或供应商可以全面保护我们的数据。目前，您可以针对需要保护的区域所处位置，使用不同的安全工具。相信随着技术的发展，数据安全问题终将逐步得以解决。

FIN

新形势下数据分析行业 面临的机遇、挑战及应对

◎ 文 / 湖南翰林项目数据分析师事务所 卿启伟 编辑 / 张楠 图 / 崔峻珩

十八届三中全会以来，国务院加快了简政放权、政府职能转变的步伐，尤其是以取消行政审批为突破口，大力推进经济、社会体制等方面改革的深化。2014年8月12日，国务院发布了《关于取消和调整一批行政审批项目等事项的决定》（国发〔2014〕27号），取消了包括注册税务师、注册资产评估师在内的11项职业资格许可和认定事项。2014年10月23日，国务院发布《关于取消和调整一批行政审批项目等事项的决定》（国发〔2014〕50号），取消和下放58项行政审批项目，取消包括土地估价师在内的67项职业资格许可和认定事项，取消19项评比达标表彰项目，将82项工商登记前置审批事项调整或明确为后置审批。国务院下决心取消一系列政府审批项目等事项，有利于政府职能转变，使其更好地履行公共服务、市场监管

的职责；有利于更好地发挥市场在资源配置中的决定性作用，为创业兴业开路，为企业发展松绑，为扩大就业提供助力，为经济社会发展增添新动力；有利于创新社会管理体制，加快政社分开步伐，把适合社会组织管理的事项交由社会组织承担，完善国家治理结构、提升治理能力。社会中介行业进入了崭新的阶段，到了重新洗牌的时候。上述新形势下，数据分析行业面临千载难逢的机遇，同时也存在较严峻的挑战，这一切值得现在及将来涉足这个行业的人士去思考并应对。

一、数据分析行业面临的机遇

1、数据分析行业是一个跨学科的边缘科学行业。它以量化指标为利器，以数据为产品表现形式。现在社会中介行业进入崭新的洗牌阶段，数据分析行业优势得以凸显：

从横向看，只要有需求和满足需求的服务能力，数据分析行业的业务就可以涉足各行各业，不再受行政框框的约束；从纵向看，只要有需求和满足需求的服务能力，它的业务可以深入行业内部，上可以运营大数据，下也可以驾驭微管理，不再受资质等级的限制。

2、数据分析行业是一个理念先进，提前进入新形势的行业。早在3年以前，中国数据分析行业协会网的“行业答疑”栏目中就有针对行业定位的内容，而且现在还可以浏览，相关页面显示：“问：项目数据分析师事务所能从事法定业务吗？未来能从事法定业务吗？答：首先要说明项目数据分析师事务所现在没有法定业务，未来更不会有法定业务。针对这个老生常谈的话题，说明三个问题：1) 中国的经济将越来越市场化，“反垄断法”也出台了，很多原来政府垄断的行业现在也全面市场化了。所谓的法定业务，总让人不自觉地与“垄断”相联系，其实不然，真正的法定业务只可能是应特种行业或特种业务的需要而产生的，它是因为某些行业不能完全放开市场而形成的特殊性质的“垄断”。政府有个很明确的观点：企业的市场行为应该让市场去鉴定和评价。所以，对于有着明确的全市场特征的数据分析行业而言，这种特殊的垄断是不可能形成的。”

可以看出，协会早就预料到中介行业发展的趋势，因而树立的理念是前卫的，至少在几年前就已经朝市场化发展的方向努力了。在新形势到来时，那些被取消准入资格的行业还在为努力调整行业发展思路和状态犯愁时，数据分析行业已经很坦然、很愉悦地对接和融入了新的发展阶段了。

3、大数据时代的到来，不仅为数据分析行业提供了大家均已耳闻目睹的名义上的数据分析概念，也为数据分析行业提供了大家都能够身心体会的实质上的数据分析效益。数据分析行业占得时代先机，前途无量。

二、数据分析行业面临的挑战

1、在传统的中介市场下，其他行业各自占据行政支持领域，某些观念已经深入人心，数据分析行业在传统观念的壁垒中夹缝求生。在新形势重新洗牌的时候，被切割的行政影响力的惯性作用的依旧存在，数据分析行业在传统观念充斥的领域有被视为异类和泊来物加以排斥的危险。

2、数据分析行业是一个跨学科的边缘科学行业，很多方法和模型的使用要借鉴其他行业学科的理论体系，发达国家在数据分析领域已经有了很多成熟、成功的研究方法及案例并可以加以推广，但在我国还处于探索阶段。

3、项目数据分析师的水平良莠不齐，项目数据分

析师事务所的发展参差不齐。部分分析师们不结合事务所的实际，或是不深化分析水平、一味追求“短平快”的投资分析或投资评估业务，失去真正的研究能力；或是被“高大上”“神圣的”大数据分析业务的光环所惑，与其临渊羡鱼，不如退而结网，从适合初创阶段自身特点的基础数据分析业务入手，反而能步步为营，避免出现步履维艰的局面。

4、项目数据分析师的培育机制有待完善，首先表现在网点分布不普及，部分省区还没有建立授权培训点，客观上影响了分析师团队的组成和成长；其次表现在培训内容理论性过强，实战性较弱，考过后的数据分析师存在将理论与实践合理转换的难题。

三、数据分析行业面临机遇和挑战的思考与应对

针对上述机遇和挑战，现在及将来涉足这个行业的人士都会思考并应对，但仁者见仁，智者见智。结合事务所发展的实际情况与同仁们分享经验和体会：

1、高举“数据分析”旗帜

在当今中介市场重新洗牌的时候，正是数据分析行业的分析师们为行业发展举旗呐喊的时候。我们必须旗帜鲜明地宣传项目数据分析师和项目数据分析师事务所。当传统行业褪去笼罩在头上的行政支持光环且印章鉴证力度减弱的时候，项目数据分析师适时出击，用详尽优质的量化数据服务感化市场需求方，会有意想不到的收获。

2、努力形成自己独特的研究体系，提升在市场竞争过程中的核心竞争力。

从两个层面形成体系，一个是协会宏观层面，一个是事务所个体层面。前者由于涉及面广，形成体系需一个过程，后者则可以优先考虑。翰林事务所一直致力于事务所层面的体系建立，譬如业务洽谈流程的建立，相关文书资料的设计（包括委托书内容及格式、责任声明书的版本、业务约定书的格式、报告的封面、封底、目录、参与人员的资质展示、排版、印章的加盖、页眉页脚、表格数据的链接、指标口径的对接等）、内部管理制度的建立、事务所未来发展的规划、事务所网站的建立、事务所宣传资料的编辑印刷等，翰林事务所根据事务所的特点量体裁衣，能够提升事务所在市场竞争过程中的核心竞争力。个体事务所的体系建立后如果被市场证明可行，则可以推广至整个行业，协会宏观层面再提炼后建立相关准则、指南、指引、规范等宏观体系，能够提升数据分析行业在市场竞争过程中的核心竞争力。

3、项目数据分析师一定要树立踏踏实实不断学习的观

念，项目数据分析师事务所的发展一定要结合事务所的实际制订规划，在事务所条件不具备的情况下，不要一味追求“高大上”、“神圣的”大数据分析业务。其实数据分析业务很宽很广，一些适合初创阶段特点的基础的数据分析业务，只要经营得当，也会产生较好效益。要结合实际情况努力开拓数据分析的着陆点，使数据分析业务首先能保证“接地气”，提高项目成功率和回报率。只有解决温饱问题，才会去考虑发展的问题，才有足够的本钱去拥抱大数据时代。

4、数据分析行业协会的顶层设计至关重要。这几年协会对行业的发展已经做了许多实实在在的工作，我们也已经感受到许多实实在在的服务。协会能够凝聚各方智慧和力量，会在行业学术体系建立、行业监管制度建设、会员服务机制的完善等三方面取得历史突破并建立长效机制。基于行业发展的现状以及市场竞争的需要，项目数据分析师及事务所对协会顶层设计的渴求是与日俱增的。

FIN

浅谈项目数据分析师的培养

◎ 文 / 上海天元项目数据分析师事务所 陶长书 编辑 / 侯建 图 / 崔峻珩

进入大数据时代，大数据一词越来越多地被提及，数据概念逐渐深入人心。一方面，信息爆炸时代产生了海量数据，虽然数据本身不会说话，但数据带来的价值展现在每个人面前的都是公平的。如何解读这些数据，如何利用这些数据，用这些数据创造出价值，才是最考量一个数据分析师的知识和眼界的。

我们发现，任何数据，只有站在人的这个视角下分析，才是最有意义的。另一方面，在与客户谈判中，事务所的专业性、权威性、是通过数据分析师来展现的。只有把事务所的优势充分展现在客户面前，客户才会信赖并愿意合作。由此，可以看出项目数据分析师的培养是事务所发展的重中之重。

下面浅谈一下我们事务所对数据分析师进行锻炼和培养的几个方式。

第一，创造良好的学习环境和学习氛围。每天早上一小时的晨会，晨会上大家头脑风暴积极提出问题。下午下班之前一小时的总结，把当天的工作遇到的问题整理、分析、讨论、总结。选取有代表性的可深究的问题，在每周一次的案例展示进一步分解，制定分析方法，最后做出分析结果。我们参照管理咨询公司常用的七步分析法，把问题的解决分成七步来做。不仅调动了整个学习氛围，还锻炼了数据分析师工作的条理性 and 逻辑性，实用性很强。

第二，项目数据分析师参与到事务所业务中去。虽然分析工作是技术工作，但是学以致用，再高深的分析方法和分析工具，如果不能融入到实际业务当中也只能是纸上

谈兵。对于每一位来咨询的客户，我们都要求1-2位分析师和业务人员一起接待。一方面，锻炼分析师的与客户的沟通能力；另一方面，让分析师能够深入了解客户的业务需求，使得理论学习与实际业务联系起来，促进业务能力的提高！

第三，注重平时的积累，厚积薄发。在项目分析当中，孤零零的数据是毫无意义的，必须要对数据的相关行业去做深入了解，把数据放在特定的环境下，去考察才有意义。因此要求我们分析师，平时多研究咨询管理类的书籍，研究成功案例。

第四，注重年青分析师的培养，并积极向数据分析行业的同行学习、交流、合作。年青分析师思路灵活，精力旺盛，但是行业经验不足，所以我们建议这些分析师积极参加一些沙龙、培训。学习回来，让他们用新知识在事务所内部讲解，一方面传递新知识，另一方面加深自己的理解。

总之，功夫在诗外，数据分析不能停留在数据分析工具运用、简单的数据处理层面，必须与实际行业和社会生活相结合，才能分析出对人类有价值的的数据。数据分析切忌闭门造车。

在2014年大数据持续加热的环境下，数据分析师带动着事务所的发展，事务所的发展又为每位分析师提供了学习和快速提升的平台。培养数据分析师，就是在为事务所可持续性发展打下坚实基础。

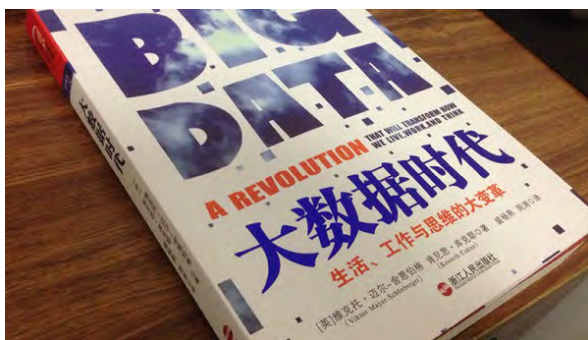
FIN



中国大数据行业亟待解决的五大问题

◎ 编辑 / 张楠 图 / 崔峻珩

《大数据时代：生活、工作与思维的大变革》作者之一维克托·迈尔·舍恩伯格曾表示，如同望远镜让人类能够感知宇宙，显微镜让人类能够观测微生物一样，大数据开启了一次重大的时代转型。



大数据，可谓当下最时髦的词，简单说就是用来描述需要批量处理或分析的大量数据集。

美国是最早发现和使用大数据科学价值的国家。2012年3月，奥巴马政府宣布投资2亿美元拉动大数据相关产业发展，将“大数据战略”上升为国家战略，奥巴马政府甚至将大数据定义为“未来的新石油”。当时美国政府声明

说通过提高美国从大型复杂的数据集中提取知识和管理的能力，来加强整个国家的竞争力，这被认为是跟互联网同一个级别的时代。显然，大数据不止是一个词汇，更是一门技术，是一种思维模式，是一个产业时代。

而中国作为世界上人口最多、GDP排名第二的国家，成立大数据国家队是非常及时的。大数据的精髓在于“大”，它不是抽样而是全样，它不是盲人摸到的象腿或者是象鼻子，而是整个大象本身，大数据的精妙处在于用的人越多越增持，通过这样一个模糊的宏观判断，能够完成一个精准的个体推荐，从而会让整个生产效率得到极大提高。

不过作为一个新生领域，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。

问题一：数据来源错综复杂

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。就现有的数据资源来说，还存在标准化、准确性、完整性低，利用价值不高的情况，这大大降低了数据的价值。

现如今，几乎任何规模企业，每时每刻也都在产生大量的数据，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信息。

未来，数据采集是一个很大的市场，因为分析的数据模型可以根据需求和思维做，但所有的前提是你的数据采集要准，现在的问题一个是采集不到，一个是采集错了，还有一个是采集效率受到网络带宽限制，这几个都做不到的话数据价值很难用起来。

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；就连我们家里的路由器、电视机、空调、冰箱、饮水机、净化器等也开始越来越智能并且具备了联网功能，这些家用电器在更好地服务我们的同时，也在产生着大量的数据；甚至我们出去逛街，商户的WIFI,运营商的3G网络，无处不在的摄像头电子眼、百货大楼的自助屏幕、银行的ATM，加油站以及遍布各个便利店的刷卡机等也都在产生着数据。

随着移动互联网、云计算等技术的飞速发展，无论何时何地，手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。你的一举一动、地理位置、甚至一天去过哪些地方，都会被记录下来，成为海量无序数据中的一个数列，和其他数据进行整合分析。



比如，当你用手机扫描二维码，并将其用微博转发的时候，你的消费习惯、偏好，甚至你的社交圈子的信息，就已经被商家的大数据分析工具捕获。大数据平台在提供服务的同时，也在时刻收集着用户的各种个人信息：消费习惯、阅读习惯甚至生活习惯。这些数据，一方面给人们带来了诸多便利，但另一方面，由于数据的管理还存在漏洞，那些发布出去或存储起来的海量信息，也很容易被监视、被窃取。

大数据散发出不可估量的商业价值。但让人感到不安的是，信息采集手段越来越高超、便捷和隐蔽，对公民个人信息的保护，无论在技术手段还是法律支撑都依然捉襟见肘。人们面临的不仅是无休止的骚扰，更可能是各种犯罪行为的威胁。大数据时代，谁来保护公民的个人隐私？既是每个人都应当思考的问题，也是政府部门不可推卸的责任。

问题二：怎样建立数据挖掘分析模型

步入大数据时代，人们纷纷在谈论大数据，似乎这已经演化为新的潮流趋势。数据比以往任何时候都更加根植于我们生活中的每个角落。我们试图用数据去解决问题、改善福利，并且促成新的经济繁荣。人们纷纷流露出去大数据的高期待以及对大数据分析技术的格外看好。然而在人们鼓吹大数据分析神奇价值的同时，却鲜见其实际正确运用的模式和方法。造成这种窘境的原因主要有以下两点：一是对于大数据分析的价值逻辑人们尚缺乏足够深刻的洞察；其次便是大数据分析中的某些重大要件或技术还不成熟。大数据时代下数据的海量增长，企业员工缺乏这种理解大数据分析逻辑和技术的思路，正是大数据时代下我们面临的严峻挑战。

大数据的大，一般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，这便使原有的以高成本方式获得的描述人类态度或行为的、数据有限的小数据已然变成了一个巨大的、海量规模的数据包。这其实是一种片面认识。其实，前大数据时代也有海量的数据集，但由于其维度的单一，以及和人或社会有机活动状态的剥离，而使其分析和认识真相的价值极为有限。大数据的真正价值在于它的全面，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。

另外，要以低成本和可扩展的方式处理大数据，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。而我国数据处理技术基础薄弱，总体上以跟随为主，难以满足大数据大规模应用的需求。如果把大数据比作石油，那数据分析工具就是勘探、钻井、提炼、加工的技术。我国必须掌握大数据关键技术，才能将资源转化为价值。应该说，要迈过这道坎，开源技术为我们提供了很好的基础。

因此，现在已经有企业开始意识到，要想真正在Hadoop平台上做数据分析、数据挖掘的应用，有两种选择，要么就是汇聚一个懂数据、懂分析、懂编程又要有技巧的技术团队来操作，要么就是选择市场上的成熟的大数据平台。

总而言之，目前尽管计算机智能化有了很大进步，但还只能针对小规模、有结构或类结构的数据进行分析，谈不上深层次的数据挖掘，现有的数据挖掘算法在不同行业中还难以通用。

问题三：数据开放与隐私问题如何平衡

数据应用的前提是数据开放，这已经是共识。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%。目前我国一些部门和机构拥有大量数据但宁愿自己不用也不愿提供给有关部门共享，导致信息不完整或重复投资。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。

下面，我们来看一下美国在数据开放方面的做法。美国政府提供政策和经费保障，使数据信息中心群成为国家信息生产和服务基地，保障数据信息供给不断，利用网络把数据和信息最便捷、及时地送到包括科学家、政府职员、公司职员、学校师生在内所有公民的桌上和家中，把全社会带进了信息化时代。

纵观国内，我国政府、企业和行业信息化系统建设往往缺少统一规划和科学论证，系统之间缺乏统一的标准，形成了众多“信息孤岛”，而且受行政垄断和商业利益所限，数据开放程度较低，以邻为壑、共享难，这给数据利用造成极大障碍。制约我国数据资源开放和共享的一个重要因素是政策法规不完善，大数据挖掘缺乏相应的立法，毕竟我国还没有国家层面的专门适合数据共享的国家法律，只有相关的条例、法规、章程、意见等。无法既保证共享又防止滥用，一方面欠缺推动政府和公共数据的政策，另一方面数据保护和隐私保护方面的制度不完善抑制了开放的积极性。因此，建立一个良性发展的数据共享生

态系统，是我国大数据发展需要迈过去的一道坎。

开放与隐私如何平衡，亦是一大难题。任何技术都是双刃剑，大数据也不例外。如何在推动数据全面开放、应用和共享的同时还可以有效地保护公民、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。

全社会开放与共享数据还很难，这让数据质量大打折扣。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，准确全面应用起来，以推动知识经济和网络经济的发展，促进中国的经济增长由粗放型向精细型转型升级。然而战略观念上的缺失、政府机构协调困难、企业对数据共享的认识不足及投入不够、科学家对大数据的渴望无法满足等都是大数据在当前我国发展应用中不得不面对的困难。



问题四：如何用数据做决策

大数据的技术挑战显而易见，但其带来的决策挑战更为艰巨。大数据至关重要的方面，就是它会直接影响组织怎样作决策、谁来作决策。在信息有限、获取成本高昂且没有被数字化的时代，组织内做重大决策的人，都是典型的位高权重的人，要不然就是高价请来的拥有专业技能和显赫履历的外部智囊。但是，在今时今日的商业世界中，高管的决策仍然更多地依赖个人经验和直觉，而不是基于数据分析。

大数据开发的根本目的是以数据分析为基础，帮助人们做出更明智的决策，优化企业和社会运转。《哈佛商业评论》称大数据本质上是“一场管理革命”，大数据时代

的决策不能仅凭经验，而真正要“拿数据说话”。因此，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与大数据技术工具相适配。这或许是我们最难迈过的一道坎了。

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、垄断型企业，其他行业谈大数据价值为时尚早。随着企业内部的资料量愈来愈大，日后大数据将成为企业支出中的主要因素，特别是数据储存所耗费的成本，很可能造成企业负担，甚至望而却步。因此有远见的高管必须预先做好准备。

问题五：大数据人才的培养

如果说，以Hadoop为代表的大数据是一头小象，那么企业必须有能够驯服它的驯兽师。在很多企业热烈拥抱这类大数据技术时，精通大数据技术的相关人才也成为一个大缺口。

大数据建设的每个环节都需要依靠专业人员完成，因此，必须培养和造就一支懂指挥、懂技术、懂管理的大数据建设专业队伍。

可以说，真正启动大数据在企业和社会的全面应用，面临的不仅仅是技术和工具问题，更重要的是要转变经营思维和组织架构，才能真正地挖掘这座大数据“金矿”。那么在大数据时代，我们要做哪些应对之策，以握战略制胜之点？

整合与开放是基石

大数据服务创业公司Connotate对800多名企业高管进行了调查。结果显示，60%受调查者称：“目前就说这些大数据投资项目肯定能够带来良好回报尚为时过早。”之所以如此，是由于当前大数据缺乏必需的开放性：数据掌



握在不同的部门和企业手中，而这些部门和企业并不愿意分享数据。

大数据是通过研究数据的相关性来发现客观规律，这依赖于数据的真实性和广泛性，数据如何做到共享和开放，这是当前大数据发展的软肋和需要解决的大问题。

2012年美国大选，奥巴马因数据整合而受益。在奥巴马的竞选团队中有一个神秘的数据分析团队，他们通过对海量数据进行分析帮助奥巴马筹集到10亿美元资金；他们通过数据分析使竞选广告投放效率提升了14%；他们通过制作“摇摆州”选民的详细模型，每晚实施6.6万次模拟选举，推算奥巴马在“摇摆州”的胜率，并以此来指导资源分配。奥巴马竞选团队相比罗姆尼竞选团队最有优势的地方：对大数据的整合。奥巴马的数据分析团队也意识到这个全世界共同的问题：数据分散在过多的数据库中。因此，在前18个月，奥巴马竞选团队就创建了一个单一的庞大数据系统，可以将来自民意调查者、捐资者、现场工作人员、消费者数据库、社交媒体，以及“摇摆州”主要的民主党投票人的信息整合在一起，不仅能告诉竞选团队如何发现选民并获得他们的注意，还帮助数据处理团队预测哪些类型的人有可能被某种特定的事情所说服。正如竞选总指挥吉姆-梅西纳所说，在整个竞选活中，没有数据做支撑的假设很少存在。

2012年3月，美国奥巴马政府宣布投资2亿美元启动“大数据研究和发展计划”，将“大数据研究”上升为国家意志。一个国家拥有数据的规模和运用数据的能力将成为综合国力的重要组成部分。国内智慧城市建设目标之一就是实现数据的集中共享。

因此，从社会、国家领域而言，我国亟须在国家层面对大数据给予高度重视，特别需要从政策制定、资源投入、人才培养等方面给予强有力的支持；另一方面，建立良性的大数据生态环境是有效应对大数据挑战、用好大数据的主要出路，需要社会各界、行业协会以及政府部门在国家政策的引导下共同努力消除壁垒，建立和谐的大数据生态系统。

力推合作共赢的商业模式

随着云计算、大数据技术和相关商业环境的不断成熟，越来越多的“软件开发者”正在利用跨行业的大数据平台，打造创新价值的大数据应用，而且这一门槛正在不断降低。因为首先，数据拥有者能够以微乎其微的成本获取额外的收入，提高利润水平；其次，大数据设备厂商需要应用来吸引消费者购买设备，发展合作共赢的伙伴关系



势必比单纯销售设备要有利可图，一些具有远见的厂商已经开始通过提供资金、技术支持、入股等方式来扶持这些“软件开发者”；第三，行业细分市场的数据分析应用需求在不断加大，对于整个大数据产业链来说，创新型的行业数据应用开发者必将是未来整个大数据产业链中最为活跃的部分。

未来，有三种企业将在“大数据产业链”中处于重要地位：掌握海量有效数据的企业，有着强大数据分析能力的企业，以及创新的“软件开发者”。社交网络、移动互联网、信息化企业、电信运营商都是海量数据的制造者，Facebook公司手中掌握着8.5亿用户，淘宝注册用户超过3.7亿，腾讯的微信用户突破3亿，这些庞大用户群所提供的数据，正在等待时机释放出巨大商业能量。可以预测，在不久的将来，Facebook、腾讯、电信运营商等海量数据持有者或者自我延伸成为数据分析提供商，或者与IBM、ZTE等企业密切对接成为上下游合作企业，大数据产业链将在某个爆发时点到来之际，以令人惊讶的速度成长壮大。

大数据的杀伤力需防患于未然

大数据时代，传统的随机抽样被“所有数据的汇拢”所取代，人们的思维决断模式，已可直接根据“是什么”来下结论，由于这样的结论剔除了个人情绪、心理动机、抽样精确性等因素的干扰，因此将更精确、更有预见性。不过，由于大数据过于依靠数据的汇集，一旦数据本身有问题，就很可能出现“灾难性大数据”，即因为数据本身的问题，而导致错误的预测和决策。

大数据的理论是“在稻草堆里找一根针”，而如果“所有稻草看上去都挺像那根针”呢？过多但无法辨析真伪和价值的信息和过少的信息一样，对于需要作出瞬间判断、一旦判断出错就很可能造成严重后果的情况而言，同样是一种危害。“大数据”理论是建立在“海量数据都是

事实”的基础上，而如果数据提供者造假呢？这在大数据时代变得更有害，因为人们无法控制数据提供者和搜集者本人的偏见。拥有最完善数据库、最先接受“大数据”理念的华尔街投行和欧美大评级机构，却每每在重大问题上判断出错，这本身就揭示了“大数据”的局限性。

不仅如此，大数据时代造就了一个数据库无所不在的世界，数据监管部门面临前所未有的压力和责任：如何避免数据泄露对国家利益、公众利益及个人隐私造成伤害？如何避免信息不对等，对困难群体的利益造成伤害？在有效控制风险之前，也许还是让“大数据”继续待在笼子里更好一些。

大数据的经济价值已经被人们认可，大数据的技术正逐渐成熟，一旦完成数据的整合和监管，大数据爆发的时代即将到来。我们现在要做的，就是选好方向，为迎接大数据的到来，提前做好准备。

以未来的视角看，无论是政府、互联网公司、IT企业还是行业用户，只要我们以开放的心态、创新的勇气拥抱“大数据”，大数据时代就一定有属于中国的机会。

大数据人才培养迫在眉睫

大数据相关人才的欠缺将会成为影响大数据市场发展的一个重要因素。据Gartner预测，到2015年，全球将新增440万个与大数据相关的工作岗位，且会有25%的组织设立首席数据官职位。大数据的相关职位需要的是复合型人才，能够对数学、统计学、数据分析、机器学习和自然语言处理等多方面知识综合掌控。未来，大数据将会出现约100万的人才缺口，在各个行业和领域，数据分析中高端人才都将成为最炙手可热的人才，涵盖了大数据的工程师、规划师、分析师、架构师、应用师等多个细分领域和专业。因此需要在行业协会的引导下，促进社会、高校和企业共同努力培养新形势下的数据分析人才，满足日益发展的市场需要。

当大数据被讨论得热火朝天时，我们更需要冷静地思考，如何让技术扎实而有效地落地。虽然距离大数据时代我们还有很长的路要走，但正象孙正义在乌镇演讲时最后说的：“我想说的是，我们要有信心，中国几年后将将成为世界最大的经济体，人类的未来将充满了机遇，会充满了很多的快乐，有很多的光明憧憬等待着我们”。

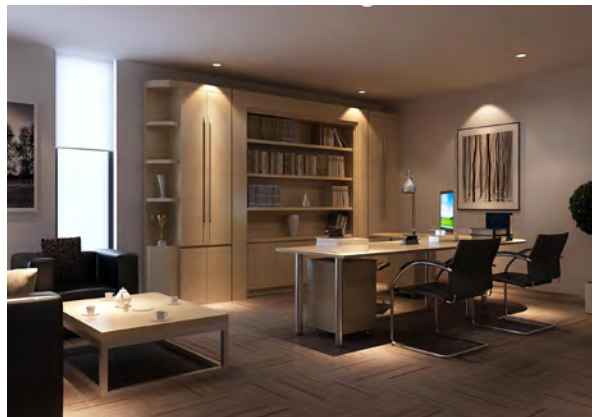
FIN

贵州新阳光数据分析师事务所简介

贵州新阳光数据分析师事务所有限公司的成立依托于贵州新阳光工程招标咨询有限公司（以下简称“新阳光招标公司”），注册资金为人民币壹佰万元。

通过近10年的积累，汇集了招标代理、采购、项目管理以及技术咨询等方面的专家和高级技术管理人才，并拥有由行业权威人士组成的专家库，力争通过资源整合，为政府和企业提供专业的经营决策数据分析，为政府决策、各企事业单位采购、建设项目的绩效评价等提供咨询服务；提供企业诊断、投资预测、财务分析、品牌战略、发展规划、营销策划、商业计划书、项目计划书编制等多元化战略咨询服务。

贵州新阳光数据分析师事务所秉承着以人为本的理念，以实现数据分析价值为己任，锐意进取，务实高效，力求为推动中国数据分事业发展贡献力量。



办公地址：贵州省贵阳市新华路112号贵正大厦6楼

联系方式：包先生 18685153335

贵州北辰星项目数据分析师事务所简介



贵州北辰星项目数据分析师事务所有限公司是经中国商业联合会数据分析专业委员会备案核准，在贵州省贵阳市南明区工商行政管理局登记注册，具有独立法人资格的项目数据分析专业服务机构。

北辰星事务所位于贵阳市云岩区紫林庵商业中心地带，所内专职数据分析师10人，具有项目分析及金融财务相关行业的丰富经验。北辰星事务所成立时，得到了财

务、税务、资产评估、工程造价、建设工程等多个行业的大力支持。

北辰星事务所为政府和企业提供专业的经营决策数据分析、投资项目评估、分析、规划、策划；编制项目数据分析报告；投资商业项目计划书；投资项目可行性报告；招投标代理业务，投资风险分析、项目融资咨询，规划咨询等服务。

北辰星事务所具有良好的职业道德和行业准则，以饱满的热情、专业的眼光、诚恳的态度为宗旨，秉承客观、公正、科学的思想，全面的服务于社会大众。

办公地址：贵州省贵阳市云岩区新建路
金宇紫林广场B1-24-8

联系方式：杜女士 0851-6762255

电子邮件：1009605317@qq.com

图说大数据

数据分析师不是数羊的

咨询业内有一个经典的故事



一个牧民赶着羊群在草原上走，迎面碰到一个人对他说：“我可以告诉你，你的羊群有几只羊。”随即，他用卫星定位技术和网络技术将信息发到总部的数据库。片刻后，他告诉牧民羊群共有1460只羊，并且要求牧民给他一只羊作为报酬，农民答应了。

随后，牧民对他说：“如果你能说出你是干嘛的，你能否把羊还给我？”那人说：“行”。牧民说：“你是一个咨询顾问。”那人很惊讶，问牧民是怎么知道的。牧民说：“有三个理由足以让我知道”

- 1 我没有请你，你自己就找上门来。
- 2 你告诉了我一个早已知道的东西，还要向我收费。
- 3 一看就知道你不懂我们这一行，你抱的根本不是羊，而是一只牧羊犬。

如今，数据分析、数据仓库、数据挖掘、大数据等概念热得发烫，数据分析师被认为是万众瞩目的职场新宠，关于数据分析师技能、职责、职业素养等的讨论不绝于耳。

就像“魏黑”一样，数据分析的质疑声也不断传出和放大：数据分析，挖掘到底能否产生价值，多大价值？

我们把上面数羊的故事改一下



一个牧民有一群羊，他找了一个年轻帮手，问道：“你看着这群羊怎么样？”随即，年轻人走入羊群进行考察，并用各种统计方法和不同工具进行了全面的判断。

然后，他告诉牧民羊群共有1460只羊，仅有10只公羊，其余为母羊，可以繁殖的母羊有1000只，其余为羊仔。根据一些特征，羊群可以分为“淘气型”、“疯跑型”、“活跃乱跳小仔型”。



牧民听后既惊讶又失望，惊讶的是一个没放过羊的人和不一样了解羊群，失望的是他所听到的都是他早已知道的。

目前，数据分析师，数据分析师就有这样的境遇



他们是企业内部组建的“智囊”，被要求执行各类战略、解决业务问题等任务，被寄予厚望。不过，真正体会到价值的企业怕是不多，就像改编的故事一样，数据分析师，数据分析师们很快掌握了企业内部的“经验”，但为企业增值做出的贡献又能有多少呢？

好的数据分析师会这样回答牧民的问题



“羊群共有1460只羊，仅有10只公羊，其余为母羊，可以繁殖的母羊有1000只，其余为羊仔。”



因此，当务之急是卖掉长肥的小羊，马上引进更多的种公羊，以解决当前种羊和母羊比例严重失调的问题；



根据对市场的预估，5月份每卖掉一只小羊将比4月份多赚150元，因此，我们必须把握先机，4月前育肥，5月清栏；



对于“疯跑型”羊，有必要采取两条腿跳绳的方法限制其大范围跑动，对于“活跃乱跳小羊”应采取与成年羊隔离的放养的方式。

看了上面的内容，我们来重新认识一下什么是数据探索



数据探索，就是在大量数据集中发现有用关系的系统性的方法。

在开始之前，你不必知道寻找的是什么，你可以通过拟合不同模型和研究不同关系来探索数据，直到你发现有用的信息为止。

最后发现潜在价值，预见可能发生的某种“坏的未来”，并给出建议。

因此，数据分析师的级别可以分为以下三类

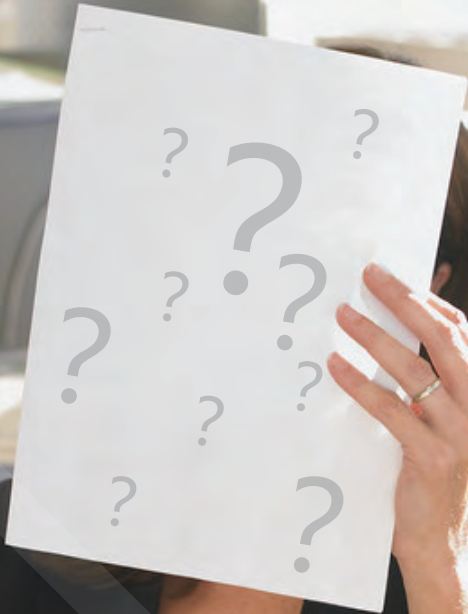
发现经验描述问题 数羊的	被动接受分析原因解决问题 分析羊的	主动预见解决问题创造价值 帮着卖羊的
-----------------	----------------------	-----------------------

你们看是这回事吗？



官方微信：wxchinacpda
微信扫一扫 精彩天天到

CPDA



漫无目的 不如 精准定位

项目数据分析师
引领大数据时代的
知识体系变革

早培训

早升职

早加薪



全新课程
无限精彩

给自己一次机会， 迎接无限精彩的未来！



行业协会
官方微博



官方微信
wxchinacpda

★ 关注微博微信参与互动 赢好礼！

提升量化经营技能，掌握数据分析必备的科学方法论！
正确理解、灵活运用于企业决策全面数据化分析中！
成功融入大数据时代，释放数据分析师全局视野！
行业协会权威认证，承载坚实的启航平台！



www.chinacpda.com



400-050-6600