● 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

China Data Analysis

洞察损益 • 量衡天下

中 国 数 据 分 析 行 业 核 心 刊 物

2015年
第1期
总第21期

协会新址——朝外SOHO C座

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*

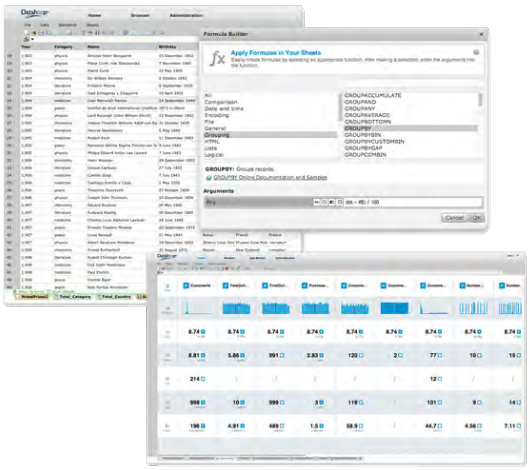
1 十多年实践经验积累，集成行业顶尖算法

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和智慧。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。这是目前世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。

目　录

2015年第 1 期　总第21期

主办单位：中国商业联合会数据分析专业委员会
编　　辑：张　楠　　潘宗祥　　石爱英
美　　工：崔峻珩

地　　址：北京市朝阳区朝外大街乙6号朝外SOHO-C座-931室
电　　话：010-59000991 转 636　/　010-59000067
传　　真：010-59000991 转 607
邮　　件：xiehui@chinacpda.org

做实事，赢未来！

2014年年底，陈年的一篇《凑热闹的公司都会烟消云散》的博文在朋友圈里引起了大量的转发。它告诫所有想在这个浮躁的社会里取得成功的企业是该沉淀一下、踏实做事的时候了。如果依然陶醉在盲目的热闹氛围里，终究是会烟消云散的。同时，也让我们反思这样的一个问题：你的企业生存的价值在哪里，你可以为社会、为你的客户创造怎样的用户体验？我想，只有想清楚了这个问题，你的企业才会逐渐走向成功，而不是昙花一现。

自2013年“大数据”概念火爆后，各种大数据论坛、数据联盟、产业联盟竞相组建，混业经营公司也开始增加数据类业务模块。到2015年3月，两会“互联网+”的提出，更给本就火热的“大数据”概念又添了一把火，云平台、物联网、互联网金融、大数据概念股、工业4.0等名词术语频繁出现在各类媒体的报道中。所有这些，让我们感到了一个时代的热闹，这是好事，因为有越来越多的个人和组织开始关注一个新生事物了。然而一个时代的发展只有热闹是不够的，真正想融入“大数据”时代，不是靠概念的炒作，而是找准时代的核心，真正踏实地创造出能够引导公众正确价值观的“产品和服务”，正像埃德蒙·费尔普斯所说“一个时代的繁荣需要全民的认知、参与以及共同的创新”。

于是，就有了这样的一个新问题，在国内相当一部分企业还没有实现数据化，以及数据分析人才缺失的现状下，有多少人能真正解理大数据的核心是什么？有人会说是“海量的数据”，大数据和小数据区分的一个关键点就是数据量前所未有的庞大，以及数据积累的方式变得快捷多样，然而，数据本身是不会产生价值的，无论在哪个时代，它本身都无法推动这个时代的前进；还有人会说是“处理数据的技术”，诚然，面对每天都在巨增的数据量，必然需要一个先进的技术工具去采集和处理，但是技术本身也不能直接推动时代的发展。只有人，通过一定的思维模式，运用先进的技术手段，将数据进行深度挖掘，分析出数据间的关联性，从而创造出新的价值。这才是大数据时代真正的核心。

作为为“大数据”布局十年多的行业协会，我们看到了巨大的市场需求，同时也清醒的看到了大数据1.0时代带来的问题。随着2015年年检工作的开展，部分早期成立的事务所，在这样一个日趋繁华、需求日益旺盛的大市场中，仍然迷茫、缺少方向，是时候让我们真正沉下心来，努力积淀自己的实力了，否则再多的机遇也留不给“投机者”们。像2014年年底行业研讨会上，众多事务所分析师感慨的那样——2015年注定是不平凡的一年！我们希望无论是协会还是数据分析行业的从业者们，能从各种热闹的场景下回归到产品的研发上，在人才体系培养以及技术咨询应用方面形成自主研发、不断更新的产品体系。让我们“以接地气”的方式为在这个时代下生存的企业和个人做点实在事儿吧。喧嚣结束后，专注经营决策数据分析业务的深化、把握“技术咨询”的先进经验，才是事务所成就辉煌的“王道”！

在即将到来的4月份行业新品发布会上，我们将通过第五版CPDA项目数据分析师认证体系，普及型培训产品数据分析员远程课件，以及积累行业十二年经验研发而成的Datahoop大数据智能分析平台来向大家诠释我们对“大数据核心”的领悟。

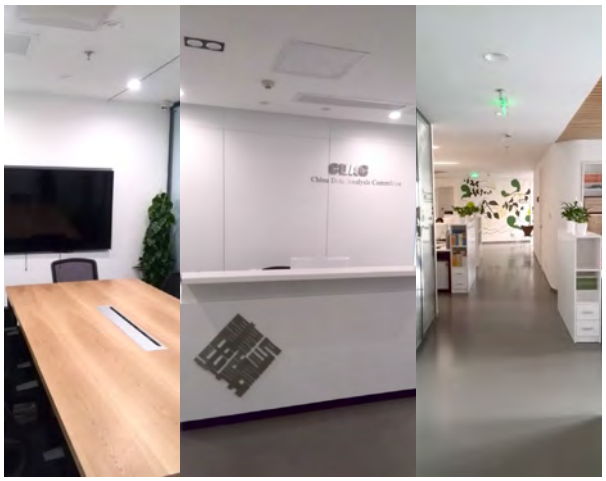
中国商业联合会数据分析专业委员会

行业新闻播报

◎ 文 / 协会会员处　石爱英　　编辑 / 协会市场处　张楠　　图 / 崔峻珩

喜迎中国商业联合会数据分析专业委员会乔迁！

2015年2月5日，中国商业联合会数据分析专业委员会乔迁新居，新办公地点位于北京市朝阳区朝外SOHO-C座9层。这里有全新宽敞的办公环境和高科技一体化的办公设备。此次搬家对于协会来说，是一个全新的起点，我们将在今年不断完善产品线，从人才培养、技术支持到业务咨询，为大众提供一个更为专业的服务平台。协会将引领数据分析行业走向一个新的高度！



中商联数据分析委欢迎各界战略合作伙伴莅临参观、洽谈合作，我们愿与您共同开创大数据时代的未来！

上图是会长办公室安装的温控器，别看它小巧简易，它可是包含了从数据收集到数据应用的高科技产品——将每天收集的室内温度，开关时间等数据进行分析而实现温度的自我调节。让我们感受到数据分析应用可以离我们如此之近！

会议室也运用了高科技手段，一台乐视电视替换了传统的的投影设备，通过AirPlay、DLAN协议来实现PC端、手机端与电视间的互联互通。

协会在办公室设置上与数据应用紧密相连，通过自身体验，来告诉大家大数据应用就在我们身边。

北京、上海两地2015年第1期公益沙龙活动实录

2015年3月14日，北京、上海两地同时开展了数据分析公益沙龙活动。

北京公益沙龙活动：

3月14日，中国商业联合会数据分析专业委员会在朝阳区格林豪泰酒店成功举办了2015年度的首期公益沙龙活动。

此次沙龙分别进行了两个主题的分享，首先由国家电子商务中心数据研究员李欣欣老师讲解了数据分析在商品预测上的重要应用，李老师以生猪价格预测为例进行了演示，引起了学员的极大兴趣，大家纷纷表示，‘生猪价格走势’这类宏观经济市场原来只需要通过建立时间序列模式就可以预测出来的，可见数据分析应用的广泛性及重要性。第二个分享嘉宾是中国商业联合会数据分析专业委员会数据中心数据分析师张婧君老师，分享主题为新媒体数据分析的应用，以现下火爆的《爸爸去哪儿2》举例进行了多维度的数据分析，演示结果明确显示出数据分析对新媒体行业发展有着不可替代的战略意义！

本次沙龙活动完美落幕后，大家反响热烈，纷纷表示期待下期活动的开展。

上海公益沙龙活动：

3月14日，上海CPDA第13期数据分析公益沙龙如期举行，共有30多位数据分析从业者和CPDA学员参加此次沙龙。沙龙的主题是：《数据挖掘方法与实际应用》， 由从事十余年数据分析和挖掘工作的李双老师为大家分享和交流。

李双老师从数据挖掘的概念讲起，清晰的阐述了数据挖掘过程，以及常用的数据挖掘技术，结合他多年的工作经验，将数据挖掘各个技术适合的场景做了简要的归类。他强调数据分析思维对数据挖掘至关重要，以目标为导向，深刻了解消费者的需要，以简单的方法来建立科学合理的挖掘模型。

在数据挖掘商业应用案例交流环节，大家问题较为集中，李老师和大家一起进行了探讨。数据挖掘的实质就是从大量数据中通过算法搜索隐藏于其中信息的过程，信息价值的挖掘对数据挖掘人才综合素质要求很高，需要技术、管理、营销等综合知识。

“数据分析公益沙龙”由中国商业联合会数据分析专业委员会主办，旨在为数据分析爱好者搭建一个自由交流的平台，为大家提供专业技术指导及行业案例展示。自协会成立以来，全国各地已举办百余场沙龙活动，2015年更将“培养数据分析专业人才，指导数据分析落地”为举办理念，希望数据分析真正为企业、为数据分析师所应用。

大数据分析助力节能环保

–暨安徽量衡项目数据分析事务所参加美菱“以旧换新”活动实录

2015年1月14日，中国商业联合会数据分析专业委员会的会员单位–安徽量衡项目数据分析事务所受邀参加了由美菱主办的最美乡村节能达人评选活动。此次活动主题为“以旧换新 美菱先行”，这种让广大农村百姓参与到节能环保主题的活动形式，得到了社会各界的一致好评。

随着今年年初《能效“领跑者”制度实施方案》正式发布，在这份涉及多个耗能产业的节能鼓励政策中，变频空调、冰箱、滚筒洗衣机和平板电视都被纳入其中，相关部门将对符合条件的家电产品制定激励政策。尽管目前相关实施细则尚未出台，但业界普遍认为这意味着新一轮家电节能补贴已经箭在弦上。

在这种情况下，美菱在安徽省打响了家电下乡的第一枪。安徽量衡项目数据分析事务所董事长李勇常年关注家电市场的行情趋势等大数据分析，他对家电下乡有着不同视角的认识。他认为，家电下乡能够受到政府的重视主要是因为它给人民带来了福利，而推动终端节能产品、高耗能行业、公共机构能效水平不断提升的长效机制，则能促进节能减排。

为了研究家电与节能的发展关系，安徽量衡项目数据分析事务所选择人均用电量,社会消费品零售总额及家电销售、四大家电产量指标，并通过安徽省近5年的家电市场指标进行相关性分析，得到这5类指标有极高的正相关性。也就是说，随着电器的不断增加，人均用电量是不断的向前增加。为了节约资源，家电的节能环保是必经之路。通过大数据分析，李勇表示乡下家电相对于城里的家电来说，污染更重，浪费更多，而群众的环保意识也相对较为薄弱。所以，不仅要“家电下乡”，还要“节能下乡”。

此次数据分析在节能环保领域的“大显身手”，不仅为节能环保产业带来了新的机遇，也为我国的节能环保事业带来了新的方法与策略。中国商业联合会数据分析专业委员会在全国各省市组建了百余家专业项目数据分析师事务所，他们在当地与各行业深入合作，运用数据分析为各个领域带来了巨大的经济和社会价值。

重庆大数据支撑产业发展2015年研讨会

2015年3月12日，由重庆市大数据应用产业联盟（重庆传晟项目数据分析师事务所发起成立）、重庆科技发展战略研究院、重庆市经济信息中心联合举办的“重庆大数据支撑产业发展2015年研讨会”在重庆市科学技术研究院盛大召开。来自政府机关、科研机构、高校、地产、制造、电信、医药、农业、互联网、金融等机构的近两百名嘉宾出席会议，重庆市科委高新处副处长雷治政先生出席此次会议并致开幕词。

本次会议围绕着“大数据支撑产业发展”这一主题，以重庆大数据产业发展典型行业应用交流，重庆大数据产业链调查与建设研讨，重庆大数据产业发展建言建议三大版块展开。此次会议也为“重庆大数据产业创新发展行动专项方案设计”项目提供初步的行业调研资料，为后续深度调研的开展，最终项目的切实落地都意义重大。

会上，来自教育、农业、金融、医药、互联网、制造行业的各位专家老师，就相关行业的大数据运用现状和运用方向做了主题演讲。此次会议以主题演讲加自由发言的模式进行研讨交流，与会嘉宾就各行业大数据运用难点与场景模式展开了激烈讨论，北京航天大学、重庆大学、重庆大数据应用产业联盟的数位专家从国内外落地案例对比分析到运用场景构架设想，对各位嘉宾的提问做出了详细解答。

重庆传晟项目数据分析师事务所大数据运维总监高峡先生，做为本次重庆市大数据应用产业联盟代表以“大数据应用场景”为主题进行了主题演讲。在其演讲中就大数据4V属性进行了介绍讲解，并对零售、医疗、地产、交通、教育行业进行了落地案例分析介绍，就各行业大数据模式创新、应用产业链发展规划也提出了初步解决方案。就目前大数据起航发展，高峡先生从数据治理、整合资源、量化体系、价值挖掘四个方面进行了分析阐述。

本次研讨会议，引发了各行业嘉宾就大数据运用的激烈探讨，在交流中解决行业难点的同时也碰撞出了更多的大数据合作运用思路。此次会议受到了各界的高度关注，不论从各大政企机构积极参与踊跃发言，还是从各大媒体关注与重庆新闻联播的报道，都可以看出此次会议为重庆大数据应用环境营造起到了极大的推动作用。

2015年度“项目数据分析师证书”年检及继续教育通知

编者按：“项目数据分析师证书” 实行定期年检制度，时间为三年一检。中国商业联合会数据分析专业委员会对项目数据分析师年检主要是考察数据分析从业人员的继续教育情况，即检查项目数据分析师是否有参与协会组织开展的远程继续教育，旨在不断提高和保持其专业胜任能力。

中国商业联合会数据分析专业委员会定于 2015 年 3 月起面向全国项目数据分析师办理证书年检及远程继续教育。请到期的各位学员尽快与“协会”客服处联系办理相关手续。

怎么知道我是否需要年检了?

翻看自己项目数据分析师证书左下角显示的有效期，如下图所示：凡是有效期至2015年的证书都需在到期前一个月联系协会客服处办理年检!

【特别注意】

1、年检需提交项目数据分析师证书原件；

2、“项目数据分析师”证书已到有效期限的，须在“协会”办理年检手续，证书有效期已过且未做年检，证书将自动失效；

3、证书年检时学员须在三年内至少参加一次“协会”组织的继续教育学习，超过年检日期一年以上不参加继续教育的，证书将不再给予年检，同时项目数据分析师学员证书自动失效。

请登录项目数据分析师官方网址www.chinacpda.com查询证书年检详情。

基础知识：CPDA学员与协会会员的区别

很多人不是很清楚CPDA学员和协会会员的区别，本文特就此问题进行详细解答：

1、学员及会员的概念

学员是指：正在学习CPDA项目数据分析师课程以及已经取得《项目数据分析师证书》的都为CPDA学员。

会员是指：在取得《项目数据分析师证书》后，提交申请，协会对其相关资质条件审核通过后，方可成为会员。

注：取得CPDA证书不会自动成为会员，需要经过一系列入会流程才可能成为会员。

2、学员证书年检及会员证书年检的区别：

CPDA学员在考核合格后获得的《项目数据分析师证书》每三年年检一次，学员可享受的免费服务为：参加协会组织的各种形式的继续教育及协会举办的各类活动等。

会员证书年检每年一次，会员可享受在学术及职业规划方面的全方位服务。

如何看CPDA人才培养与其它认证培训项目的区别

◎ 文 / 协会CPDA 培训处　　编辑 / 协会市场处　潘宗祥　　图 / 崔峻珩

编者按：近期网络上出现了一些对CPDA培训课程恶意抵毁的不良言行和虚假信息，为了让广大关注行业的人们了解真实的数据分析行业和课程、肃清不良风气，特撰此文。

自“大数据”概念在2013年被引爆后，我们惊喜地看到，无论是政府、企业还是个人，均对“大数据”给予了前所未有的热忱关注，利用“大数据”进行“精准分析”，形成企业运营进程中的“量化决策”支持，使“大数据”的魅力得以突显。由此，百度上搜索“数据分析”后，可以一下子跳出5400多万条信息，在知名招聘网上搜索“数据分析”，会出现成千上万的招聘岗位，也可见一斑。各行各业都对数据分析人才呈现不断上升的渴求趋势，懂“数据”、精“分析”的人才必然是“大数据”时代的宠儿。

这个时代需要什么样的数据分析师?

顾名思义，数据分析师就是一群“玩数据”的人，通过对数据的收集、清洗、深度挖掘以及分析，从而使得数据产生商业价值，让数据变成生产力。分析历史、预测未来、优化决策，这是大数据时代赋予数据分析师的职责。

国内大数据概念刚刚萌芽，人才市场还不那么成熟，大家对数据分析人才还只是一个懵懂的认知，认为会统计、懂IT技术的，是大数据的核心人才。其实不然，真正的数据分析师需要了解影响决策的各个环节，从数据的收集、整理展现、分析和商业洞察、到决策行为的转化等。所有这些都要求分析师必须是一个知识和能力全面发展的人才，不仅需要有数学及统计学相关的背景，还要具备丰富的经济、管理学知识，了解行业、了解企业，并对量化的信息加以深度整理、逻辑、分析进而做出具化的决策建议。当然，还要求优秀的数据分析师要保持不断学习、深化知识提升状态。随着大数据时代的深化，相信企业对数据的理解将很快从“数据”转向真正的“分析”，一个优秀的数据分析师必将是这个时代不可多得的人才。

众多的培训项目，为什么选择CPDA?

据《中国数据分析行业发展报告》显示：至2015年，大数据将在全球创建440万个工作岗位，其中有190万个工作岗位在美国。中国能够理解与应用大数据的创新人才更是稀缺资源。据国外一份薪金调查显示，数据分析师2014年的平均年薪是17.5万美元，预计2015年涨幅高达9.3%。相信通过上述数据，已经入行的数据分析从业者们在考虑如何成为一名优秀的数据分析师，而还未入行的人们则在考虑是否趁这个时机来个职业转型。现在，我们就告诉您有这样的一个机会——通过学习CPDA课程可以实现您的愿望。

从品牌角度讲，CPDA历经十多年风雨考验，已经成为了数据分析行业第一认证品牌

“项目数据分析师”从2003年至今，已发展了12年之久，培养了近万名分析师，分布在全国十几个省份，受到业界一致好评，是目前唯一一个由全国主管行业协会认可的课程体系，也是由协会和工信部教育考试中心共同推出的专业人才培养项目。深入的培训（8天面授、一年远程教育）、严格的考核（工信部教育考试中心主持考核，取证要看实力哟），使证书的含金量逐年提升，已成为当之无愧的“行业第一考”。

“项目数据分析师”是中国数据分析行业的主要从业人群，我国的“项目数据分析师事务所”均由取得CPDA资质证书的学员组成的。目前全国已有百余家专业事务所，利用他们在CPDA所学到的专业知识为IT、金融、医疗、零售、物流等

领域的企业提供着决策支持服务，得到社会各界高度关注！CPDA不仅是一个证书的名称，也是持证者成就事业成功的象征，是加入一个高端从业人群的象征。

优秀的课程成就CPDA优秀的口碑

任何一门优秀的课程，一定要有实用的课程，注重理论与实践的高度结合。数据分析是实战性行业，只有理论没有实践经验，无法在社会上得到认可。CPDA的课程优势正是由此形成：

1、CPDA课程源于实战。行业协会是课程的设计者和完善者，协会监管着全国从业事务所的业务和数据分析指导工作，每年从业事务所和分析师经历的“成功”和“失败”的案例和经验，滋养着行业的从业课程。没有实战体会，由何得到真正“落地”的课程?

2、CPDA课程不断改变和完善。CPDA课程已经历了五次大的调整和改进，教材已更新了三版，越来越多的模型、决策分析方法的引入，使课程从数据分析思维建立到实际操作能力迅速提升，结合时代发展的变化，CPDA不断将最新的观点、最热门的操作技术融入课程体系，2015年新课程即将全面上线，相信一定使学员们惊呼“这真是太实用了”！

3、师资不掺水！CPDA培训工作从十几年前就坚持一个传统：不选没有实战经验的老师、不聘请外地师资（只从北京总部统一派遣）、坚持每次课程的满意度数据收集和分析。十几年下来，优秀的师资使学员和我们的老师建立了深厚的友情，学员对课程的整体满意度没有低过90%，这是真真实实的口碑，是数据分析结果。目前，协会已经在全国十五个省市建立了授权中心，接受学员的咨询和报名。

数据分析行业在中国还是新兴行业，知识概念更新很快，只有不断学习、更新知识，才能真正具备一定的从业水平。这也正是CPDA重视课程研发、重视学员体验的根本。因此为了使CPDA证书真正代表行业标准，协会每三年会对项目数据分析师证书进行一次年检，我们希望学员铭记：一个真正地认证不是一劳永逸的！它需要在时代中历练，在实践中升值。协会关注每一位正在向优秀数据分析师迈进的学员，因此有了定期动态了解、职业发展指导等工作。也正是报着“帮助学员适应大数据时代发展”的行业精神，协会为CPDA学员提供终身的免费教育服务，通过课程更新、会员专刊派送、微信公众号、公益沙龙讲堂等形式促进分析师在业内的学习与成长，同时，通过线上社交媒体为学员建立交流平台，鼓励学员自主交流、共享结果。我们也为学员提供了服务热线以及专职老师，可以随时答疑解惑，帮助学员解决实际工作中的问题。

CPDA与协会其它培训产品的关系

一个时代的发展是全民参与的结果，同样，一个行业的普及也需要全民共同的努力。在大数据时代，无论是个人还是企业，每天都会面临很多选择，这个时候帮助我们做出决定的就是“分析思维”，通过已经掌握的信息、数据，经过大脑的分析处理，从而影响我们下一步行为的动向。所以说，数据分析已经是一件离我们生活和工作很近的事情，它会像语言、驾驶、计算机操作一样成为生存的基本技能，面对这样的发展趋势，普及性教育已经成为必然。也许不是每个人都能成为项目数据分析师，但是每个人都该懂点数据分析。

CPDA项目数据分析师认证课程为专业人才的培养提供了行业标准，是各企业和事务所专职岗位的必备“上岗证”，然则，不是所有人都能具备数据分析师所需要的资质条件，不同的企业又面临自己的需求，因此，针对这种情况，协会鼓励社会培训机构在人才培训上以丰富、多元化的视角推进数据分析的人才培养工作，中国不仅需要专业的执业分析师，也需要在企业中从事具体量化分析工作的基础数据分析人才。

由此观点，协会于2015年推出了犀数品牌的社会类培训课程（包括数据分析员基础人才培养体系）。如果你只是想学习一下数据分析基本的思维过程以及初级的操作手法，你可以考虑数据分析员的培训课程，它以远程的形式进行教学，可以不分时间、地点进行便捷学习，同时也可以帮助你认识行业、转型职业。如果你是一个已经在行业中小有经验，然而又对某个领域，或者某种技术想精深化学习的人士，你可以考虑精讲课程，目前已经有了针对电商、R语言以及精准营销方面的课程。

总之，只有数据分析人才遍布各个领域，才意味着大数据时代真正地来临。

数据分析具备客观、理性的特征，数据分析的培训更需要“诚信为本”

我们知道，在一个行业高速发展的阶段，良莠不齐的形象会缤纷展现，随着社会对数据分析人才的需求呈几何倍数上涨，一些培训机构看到“项目数据分析师”在市场发展态势良好，也想来分享这块蛋糕，从行业协会的角度看，作为整个行业的管理机构，我们欢迎越来越多的有志之士和正规培训机构加入行业大军，通过强强合作、正当竞争，来共同促进行业的发展。

但同时我们也发现这其中一些不和谐的作法：恶意贬低CPDA的品牌、以虚假的协会（如国外注册的伪行业组织）欺骗学员等等。遵守国家的法律法规，以合法的竞争取得社会的认可才是正道，否则将会对数据分析行业人才培养、行业口碑形成损害，进而影响的是全行业权益。对上述那些趋利的、混淆市场的不当行为，我们会高度关注。

在此，协会也向广大数据分析爱好者发出呼吁，在培训认证市场较混乱的今天，请用理性思维去看待行业发展，同时提高辨识能力，切莫被虚假或者夸张宣传而误导。

欢迎各行业数据分析爱好者以及寻找“朝阳”培训项目的机构加入我们，让我们同共为“大数据”时代美好的明天奋斗！

NBA：少数人的道路

◎ 文 / 帕特里克·罗兹　　编辑 / 协会市场处　张楠　　图 / 崔峻珩

闹铃声响起,该去训练了。 贾内尔·斯托克斯（下面简称斯托克斯）径直去到篮球馆,更衣,和队友一起开始热身。这是他在高中的第三年。孟菲斯,田纳西州本地有很多他这样思想的人。很快,他必须做出选择,一个将会影响他未来的选择。 在他家的桌子上有来自阿肯色大学、康涅狄格大学、佛罗里达大学、肯塔基州大学、孟菲斯大学、密西西比大学和田纳西州大学提供的篮球奖学金邀请。

一个高中运动员能获得一个大学的体育奖学金已经很罕见了，更不用说来自这么多学校的邀请。正如我们看到的,他在篮球界真的是太突出了。大多数人高中以上就不打篮球了，只有很小一部分的拥有世界级天赋的人才继续在NBA中打篮球。也就是说,什么样的机遇能够使他开始自己的NBA生涯?

高中

在美国,职业篮球享有盛誉,在2014年最受欢迎的运动中排名前五。任何一个孩子在高中打篮球都想要打职业篮球。不幸的是,他在和全国约54000名与他有同样想法的高中生球员竞争。几乎每一个临近学校的篮球场都被这些年轻人挤满了，他们超越着自己的极限、狂热的磨练着自己的技能，希望可以得到大学的录取通知单(对大多数人来说,这是进入NBA舞台的第一步。

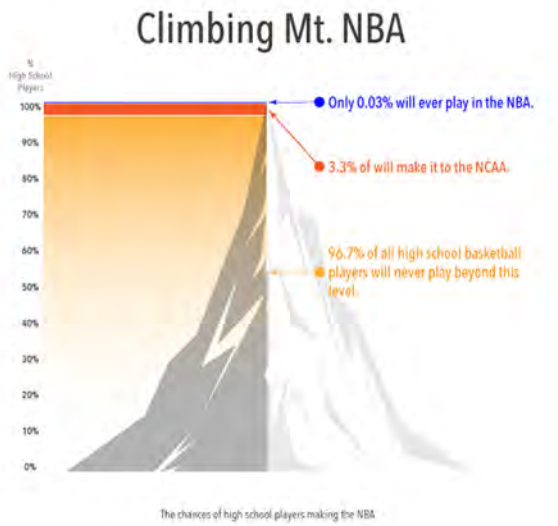
像ESPN和Rivals这样的体育节目电视网，经常统计球员数据，每年都做球员的详细的前景报告。

30个人中大约有1个(大约3.3%的比率)，会因着自己精湛的技能而收到NCAA(全国大学体育协会)成员的邀请。尽管在这些人当中,更少比率的人会被一所拥有强大篮球队的学校招募，而且只有更加少数的人会同时收到多个著名学校的邀请。换句话说,只有精英中的精英才可以自由选择他们想要去的学校(当然有奖学金)。

在斯托克斯的例子中,在高中的出色发挥把他直接归入精英中的精英行列。 结果就是，他被公认为排名前20有前景的球员，而且成为极少数同时受到上述多个学校邀请的球员。

做决定的时刻到来了

斯托克斯坐在厨房的饭桌前，面前摆着收到的各个学校的邀请函。其中有三个因着不同的原因真的吸引到了他：肯塔基大学以能让球员最快进入NBA而出名；孟菲斯则提供了一个精



英培养计划,更不用说他还可以在他家乡的学校打球了。 但对于肯塔基大学和孟菲斯大学，同时面临一个问题,他目前在这个高中的状况让他现在不能直接去大学打球。他在高中刚开始的学期转学了，没有资格参加高年级的比赛。因此他选择立即毕业并尽快开始自己的大学生涯。因为这样那样的NCAA的规则,肯塔基大学和孟菲斯大学没有办法给他提供奖学金,但田纳西州大学可以。

NBA要求球员在选秀时至少19岁。在一系列高中生跳过大学、直接进入NBA,造成巨大的社会舆论影响之后，05年这个选秀规则实行。

他填写了入校的一些文件内容,通知了Cuonzo Martin(田纳西大学当时的篮球教练)，并在当地一家餐馆安排了一个记者招待会。 那一周,他让公众知道他做的决定,并在2012年1月作为新人参加比赛。

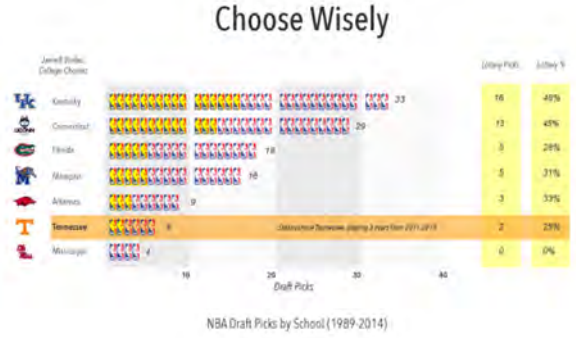
大学时光

在田纳西大学的时候，斯托克斯有立竿见影的影响力。他在作为新人的菜鸟赛季平均出场25分钟,通常达到两位数的得分。在接下来的三年大学时光里,他帮助田纳西大学平均每年超过20场胜利（赢得20场以上的胜利）。在大三的时候,他带领球队打入NCAA锦标赛的16强。生活如此美好,他开始考虑是时候进入NBA了。

此时,斯托克斯满脑子想的是他是否会在NBA首轮选秀中被选中。当各种杂音一直环绕在你周围的时候,通常也是你决定离开学院进入NBA选秀的时刻。

在大学篮球中，有一个很大的争议，就是打一年就结束的问题:大部分优秀的球员都会在入学一年后考虑进入NBA。

正如上面所提到的，因为斯托克斯想马上进入NBA,所以他选择了田纳西大学。 虽然他在田纳西打出了很好的大学生涯,但他如果选择多上一年高中的话,他可以进入肯塔基大学或孟菲斯大学。这两个大学在过去的25年里已经输送给NBA很多球员,远远超过田纳西大学



NBA Draft Picks by School (1989-2014)

为什么这个很重要? 因为球员如果能够成功完成大学时期精英球员培养计划，则证明他有能力处理来自NBA的压力。这些名校之间的竞争是非常激烈的，帮助这些球员在竞争中获得最大优势。 当然,并非总是如此——也有NBA的精英球员来自不出名学校的例子(也有直接高中毕业就进入NBA的天才，比如科比.布莱恩特)。 也就是说（综上所述）,大部分NBA精英球员来自名校。

选秀

2014年，斯托克斯在田纳西大学度过了三年之后，决定参加选秀。当亚当–萧华(NBA总裁)迅速说出第一轮的选秀选手的时候,斯托克斯也在焦急地等待中。

不幸的是,第一轮选秀的名单并没有出现斯托克斯这个名字。尽管如此,他在第二轮第8顺位被犹他爵士队选中。选秀中的交易是很常见的,他被交易到孟菲斯灰熊。这对这个年轻人来说是个好消息,因为他可以在他的家乡孟菲斯打球。

为什么在尽量靠前的顺位被选中是这么重要的呢?历史告诉我们,NBA非常善于评估人才。当然,任何人都可以举出与这个论点相反的例子,但大部分来说他们总会做出准确的选择。在CraDribbles网站中有一篇文章很好的分析了关于一个球员是在第一轮或者第二轮选中对应的职业生涯命运。在那篇文章中,他们调查NBA球员发挥重要比赛时间(场均25分钟左右)。82%的球员来自第一轮被选中的球员。而在82%的球员当中，有超过一半的球员都是乐透区选秀球员。从1999年到2013年这个时间之间的数据来看，就是这个论点的体现

NBA draft status of players averaging at least 25 minutes per game from 1999-2013

简而言之,如果你想在重要时间出场——因此获得更高的工资,享受更长的职业生涯,等等——那么你应该足够好的走出大学,第一轮被选中。 当然,一些个二轮秀也同样在NBA取得了很好的成绩,但是他们是例外而不是惯例。 如上所述,NBA非常

善于评估人才。

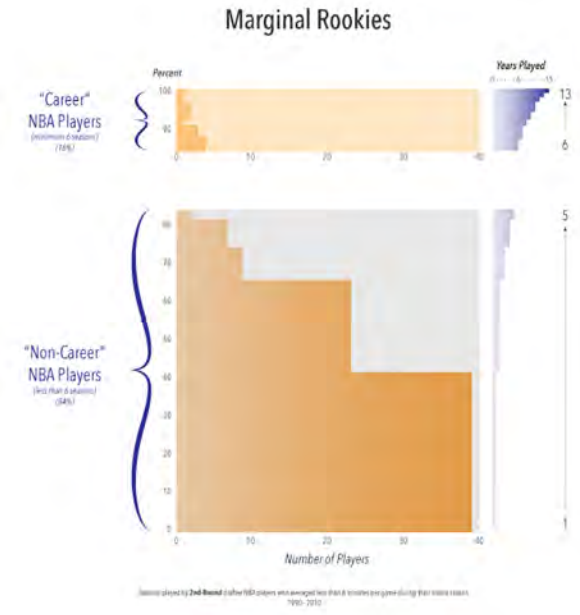
职业生涯

在NBA的第一个赛季前的夏天,斯托克斯可以与孟菲斯灰熊队签一份为期三年的合同。 事实上,两年的合同——对于第二轮被选中的球员已经是很大的惊喜了。 他在NBA的第一年新秀年薪是72.5万美元，到第二年上升到84.5万美元。他的梦想已经成真了。

首轮秀被选中拥有保障合同。二轮秀则需要谈判是否能拥有一份合约。

现在我们来看关键问题:斯托克斯能在NBA有一个不错的职业生涯吗?

让我们来定义“职业生涯”:在过去的20年里，球员的NBA生涯长度平均是6年(用的是1990年–2010年所有球员的数据)。 这实际上相对于早期NBA的球员已经是很大的进步了(和19世纪40年代、50年代的球员相比)，那时候NBA球员职业生涯还不到3年。是由很多原因造成的:伤病、能力下降、薪水问题和个人问题等等。出于本文的研究目的,我们就使用6年这个NBA的平均“职业生涯”。对于斯托克斯来说,这意味着他将从2014年到2020年在NBA征战。 我甚至还没有说他已经开始成名或者获得正常上场时间或甚至说在每一场比赛中都出场。他只是在NBA摸爬滚打了6年,甚至作为一个板凳球员去待这么久。



截至2015年2月13日,斯托克斯还是在灰熊队的编制内。对他来说不幸的是，他只有很少的上场时间,大概场均只有5.6分钟上场。通常,他是整场坐在板凳席上的。更糟糕的是,他大部分时间被下放到NBA发展联盟——一个明确的信号,他要在NBA挣扎的活下去。他还能够暂停在发展联盟，继续和灰熊队队友们一起打球。但此时,斯托克斯NBA职业生涯已经处于危险之中。他已经落后了,但还没有被淘汰。

那么,他会有6年的职业生涯吗? 和这些随机选择的球员相比,我们来看谁也是从类似的情况开始的。在相同的时间范围内(1990年–2010年),我们将选择那些场均比赛不到6分钟的第二轮被选秀的球员。这里我们有一个明确的和完整的包括95名球员的样本。如果缩小样本范围，则与本文的深度不符。

大约有16%的球员和斯托克斯有类似情况下，完成了6年的职业生涯，甚至更长久一些。因此,我也给他相同的几率。换句话说,在六个平行宇宙,只有一个斯托克斯有可能完成自己的职业生涯。我个人当然希望是在我们这个宇宙当中。

结论

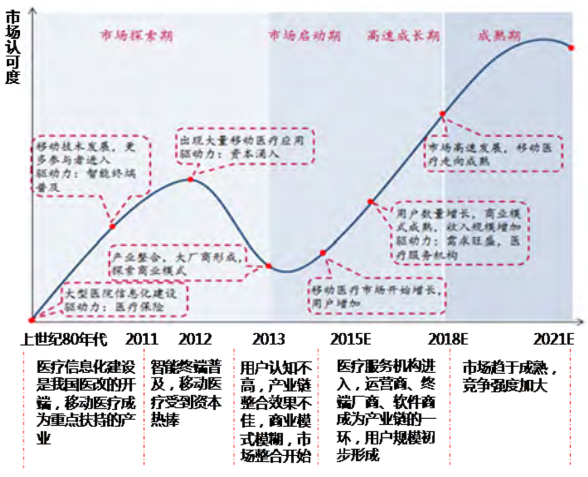
在NBA，对于像斯托克斯这样的边缘球员的生涯是很艰辛的。他们不断的在NBA发展联盟、欧洲职业球队、亚洲职业球队等等之间游走。然而，如果他们希望在最富有的、有最多天才球员的地方（NBA）打球，他们必须忍受这些。斯托克斯已经战胜了最艰难的时刻，只是走得太远,0.03%的高中球员能够实现他们的梦想。现在他已经在NBA了,他有大约16%的机会去打职业联赛。他的名字是否能够进入球队大名单中——还不能确定——他的能力还有待观察。这是一个危险的旅程;事实上,这是“少数人可以走的道路”。

医疗行业大数据的应用

◎ 文 / 数据分析师 李冠　　编辑 / 协会市场处　潘宗祥　　图 / 崔峻珩

当前，大数据的应用覆盖面非常广泛，几乎涵盖了我们所熟知的所有领域，包括零售、互联网、电信、教育等行业。除了较早前就开始利用大数据的互联网公司，医疗行业可能是让大数据分析最先发扬光大的传统行业之一。

医疗行业早就遇到了海量数据和非结构化数据的挑战，而近年来很多国家都在积极推进医疗信息化发展，这使得很多医疗机构有资金来做大数据分析。因此，医疗行业将和银行、电信、保险等行业一起首先迈入大数据时代。麦肯锡在其报告中指出，排除体制障碍，大数据分析可以帮助美国的医疗服务业一年创造3000亿美元的附加价值。



医疗大数据发展历程及问题

纵观医疗行业大数据发展的历程，自上个世纪80年代，医疗系统就开始了信息化建设，到2011年，移动医疗开始登上舞台。预计2015年以后，大数据便将替代移动医疗，发挥它强大的作用。

医疗大数据类型及来源

医疗大数据同样也有结构性数据和非结构性数据之分。所谓结构性数据简单来说就是数据库，即将数据存储在数据库里，可以用二维表结构来逻辑表达实现的数据。基本包括高速存储应用需求、数据备份需求、数据共享需求等。非结构化数据库是指其字段长度可变，并且每个字段的记录又可以由可重复或不可重复的子字段构成的数据库，用它不仅可以处理结构化数据（如数字、符号等信息）而且更适合处理非结构化数据（全文本、图像、声音、影视、超媒体等信息）。

医疗中的结构化数据主要有：临床数据、实验数据、电子病历、基因数据、药品销售数据、财务数据等。它们可以从医院的信息系统（HIS）、实验室信息系统（LIS）、电子病历数据库、财务信息系统等途径获得。非结构化数据主要有：患者的情绪数据、患者评价数据、体验数据、医院管理数据、影像数据等。它们可以从心理分析数据库、满意度调研、体验数据

库、影像数据库及交互网站中获得。

医疗大数据中的数据分析

医疗大数据尚且处于起步阶段。它具有数据的海量性、主观性、标准复杂性、结构多样性、数学表征不显著等多种特点。然而它当前也存在着许多问题有待我们去解决。一个医疗系统或一个药企会有成TB或PB的数据，当前让这些数据为医疗行业的发展提供帮助，为医院提供决策作用，还存在一定的困难，一是这些医疗数据分散在各个医疗机构当中，无法集成利用；二是数据利用度很低，医院信息系统积累了大量数据，但多数情况下仅在管理层面有限应用，很少涉及到临床专业层面；三是数据不一致、不完整、不准确，由此产生了许多垃圾数据；第四是分析能力不足，缺少实时数据和维度的建模分析，灵活性很差。

经长时间的研究，中颢润事务所的研究员们总结了医疗数据分析的一个大致流程，供大家参考。对医疗数据进行挖掘过程中首先要定义问题。定义问题是指提前明确定义所要研究的具体目标是什么。例如在对病案资料进行数据挖掘前必须要明确其最终目的是搜集疾病的关联要素还是要有效提高疾病的诊断率。数据挖掘的目的不同，所要构建的模型也是各有不同。所在，在数据挖掘前要与相关专家人士进行交流，明确数据挖掘的最终目标以及研究对象等各个要素。其次，还要对数据进行预处理。医院的数据库内积累了大量的数据，所以不论采取哪种数据挖掘方法都不会利用到全部的已有数据，而是从中选择一部分。在此过程中，所选择的数据必须要有一定的挖根生。数据预处理的意义主要在于它是决定数据挖掘过程是否会成功的主要因素，而且这一阶段所要耗费的时间相当长。之后是数据挖掘流程，而这又同时是整个研究的又一个决定性流程，这一流程包括多个环节，例如数据模型的选择、建模以及实验等。医疗数据挖掘的算法种类是非常多的，现今人们应用较多的主要是减维、模式识别、可视化、决策树、遗传算法与人工神经网络；统计方法主要有回归分析法、判别分析法、聚类分析法、主元分析法、相关分析法等。最后则是对最终结论进行评估。下面举几个实例简单说明一下：

实例一：人工神经网络评价药效

应用BP神经网络可以对药效进行评估。大概的思路是将服药后病人的各项指标值作为网络输入量，把对应的模糊综合评价结果微量作为网络输出；其次利用训练样本对BP网络进行训练，不断修正网络连续权向量，当输出误差达到要求时，结束网络训练，得到相应的输出量值；最后利用训练好的BP神经网络，对药效指标进行评价。

实例二：实验数据分析

当前在国内已普遍应用的是临床实验数据T检验技术，此项技术可应用于治疗效果、药效、新药上市、术前术后等数据的分析，具体是对观察组和实验组分别记录各项指标，通过统计学的T检验分析得出两组是否有显著差异，如有，则证明治疗方法有效。

实例三：药品销售数据

药品零售行业也可利用刚刚介绍的关联分析来增加自己的销售量。例如：购买眩晕类药物、维生素类药品的患者再购买理血药的可能性比较大；购买心脑血管和消化系统类药物的多为老年患者，同时再购买理血药的可能性大；购买金银花、胖大海等清热类的再购买五官科用药的可能性比较大。这些都可以通过关联分析确定它们是否有关联，从而这些可以搭配采购。

智慧医疗

关于智慧医疗，当前也是一个热门话题。随着智慧医疗的推广，医院的工作将更有效率，同时对于患者来说，就医会更加方便、快捷，毋庸置疑，它是一个双赢的工程。中颢润也将这类当前前端的技术整理了两个实例，在此介绍给大家。

实例一：远程医疗

杭州和南京建立了远程医疗监控系统，村民只需带一个腕表，就可以测心率和血压，通过监测仪传送到屏幕上，再通过远程设备传送到县级医院，专家可以根据情况进行诊断，如果情况紧急可派村里的卫生所及时出诊并送至县医院，这也规避了家里无人的风险。同时，也可用手机在线交流，交进行缴费。通过高级算法以及云计算的方式来加速基因序列分析，让发现疾病过程变得更快、更容易、更便宜。将病例集中在云端，医生可通过语义搜索查找任何病历中的相关信息。

实例二：个性化医疗

另一种在研发领域有前途的大数据创新，是通过对大型数据集的分析发展个性化治疗。针对不同的患者采取不同的诊疗方案，或者根据患者的实际情况调整药物剂量，可以减少副作用。个性化医疗目前还处在初期阶段。麦肯锡估计，在某些案例中，通过减少处方药量可以减少30%~70%的医疗成本。

百度为某医疗集团的数据项目就体现了这一点。首先，患者先自行拍照，医院根据图像识别技术及云计算进行分析，之后与实时病例库作对比分析（以往的类似病例），进而智能推荐，包括治疗方案，医生，用药模型等，数百名专家提供解答。

医疗大数据是一个非常庞大也具有深度的研究范围，当前中颢润已涉足医疗大数据领域，并初步取得了一些成果。然而对医疗大数据的探索不能就此止步，中颢润今后还将努力钻研，探索此领域的奥秘，也希望广大仁人共同努力，碰撞出新的火花，共同推进医疗大数据的发展。

55个最实用的大数据可视化分析工具

◎ 编辑 / 协会市场处　张楠　　图 / 崔峻珩

俗话说的好：工欲善其事，必先利其器!一款好的工具可以让你事半功倍，尤其是在大数据时代，更需要强有力的工具通过使数据有意义的方式实现数据可视化，还有数据的可交互性;我们还需要跨学科的团队，而不是单个数据科学家、设计师或数据分析员;我们更需要重新思考我们所知道的数据可视化，图表和图形还只能在一个或两个维度上传递信息， 那么他们怎样才能与其他维度融合到一起深入挖掘大数据呢?此时就需要倚仗大数据可视化(BDV)工具，因此，笔者收集了适合各个平台各种行业的多个图表和报表工具，这些工具中不乏有适用于NET、Java、Flash、HTML5、Flex等平台的，也不乏有适用于常规图表报表、甘特图、流程图、金融图表、工控图表、数据透视表、OLAP多维分析等图表报表开发的。为了进一步让大家了解如何选择适合的数据可视化产品，本文将围绕这一话题展开，希望能对正在选型中的企业有所帮助。下面就来看看全球备受欢迎的可视化工具都有哪些吧!

一、Excel

Excel作为一个入门级工具，是快速分析数据的理想工具，也能创建供内部使用的数据图，但是Excel在颜色、线条和样式上课选择的范围有限，这也意味着用Excel很难制作出能符合专业出版物和网站需要的数据图。

二、Google Chart API

Google Chart提供了一种非常完美的方式来可视化数据，提供了大量现成的图标类型，从简单的线图表到复杂的分层树地图等。它还内置了动画和用户交互控制。

三、D3

D3(Data Driven Documents)是支持SVG渲染的另一种JavaScript库。但是D3能够提供大量线性图和条形图之外的复杂图表样式，例如Voronoi图、树形图、圆形集群和单词云等。

四、R

R语言是主要用于统计分析、绘图的语言和操作环境。虽然R主要用于统计分析或者开发统计相关的软件，但也有用作矩阵计算。其分析速度可比美GNUOctave甚至商业软件MATLAB。

五、Visual.ly

如果你需要制作信息图而不仅仅是数据可视化，Visual.ly是最流行的一个选择。

六、Processing

Processing是数据可视化的招牌工具。你只需要编写一些简单的代码，然后编译成Java。Processing可以在几乎所有平

台上运行。

七、Leaflet

Leaflet是一个开源的JavaScript库，用来开发移动友好地交互地图。

八、Openlayers

Openlayers可能是所有地图库中可靠性最高的一个。虽然文档注释并不完善。且学习曲线非常陡峭，但是对于特定的任务来说，Openlayers能够提供一些其他地图库都没有的特殊工具。

九、PolyMaps

PolyMaps是一个地图库，主要面向数据可视化用户。PolyMaps在地图风格化方面有独到之处，类似CSS样式表的选择器。

十、Charting Fonts

Charting Fonts是将符号字体与字体整合(把符号变成字体)，创建出漂亮的矢量化图标。

十一、Gephi

Gephi是进行社会图谱数据可视化分析的工具，不但能处理大规模数据集并且Gephi是一个可视化的网络探索平台，用于构建动态的、分层的数据图表。

十二、CartoDB

CartoDB是一个不可错过的网站，你可以用CartoDB很轻易就把表格数据和地图关联起来，这方面CartoDB是最优秀的选择。

十三、Weka

Weka是一个能根据属性分类和集群大量数据的优秀工具，Weka不但是数据分析的强大工具，还能生成一些简单的图表。

十四、NodeBox

NodeBox是OS X上创建二维图形和可视化的应用程序，你需要了解Python程序，NodeBox与Processing类似，但没有Processing的互动功能。https://www.nodebox.net/code/index.php/Home

十五、Kartograph

Kartograph不需要任何地图提供者像Google Maps，用来建立互动式地图，由两个libraries组成，从空间数据开放格式，利用向量投影的Python library以及post GIS，并将两者结合到SVG和JavaScript library，并把这些SVG资料转变成互动性地图。

十六、Modest Maps

Modest Maps是一个很小的地图库，在一些扩展库的配合下，例如Wax、Modest Maps立刻会变成一个强大的地图工具。

十七、Tangle

Tangle是一个用来探索，Play和可以立即查看文档更新的交互工具。

十八、Crossfilter

Crossfilter既是图表，又是互动图形用户界面的小程序，当你调整一个图表中的输入范围时，其他关联图表的数据也会随之改变

十九、Raphael

Raphael是创建图表和图形的JavaScript库，与其他库最大的不同是输出格式仅限SVG和VML. http://raphaeljs.com/

二十、jsDraw2DX

jsDraw2DX是一个标准的JavaScript库，用来创建任意类型的SVG交互式图形，可生成包括线、矩形、多边形、椭圆、弧线等图形。http://jsdraw2dx.jsfiction.com/

二十一、Pizza Pie Charts

Pizza Pie Charts是个响应式饼图图表，基于Adobe Snap SVG框架，通过HTML标记和CSS来替代JavaScript对象，更容易集成各种先进的技术。

二十二、Fusion Charts Suit XT

Fusion Charts Suit XT是一款跨平台、跨浏览器的JavaScript图表组件，为你提供令人愉悦的JavaScript图表体验。它是最全面的图表解决方案，包含90+图表类型和众多交互功能，包括3D、各种仪表、工具提示、向下钻取、缩放和滚动等。它拥有完整的文档以及现成的演示，可以帮助你快速创建图表。

二十三、iCharts

iCharts提供可一个用于创建并呈现引人注目图表的托管解决方案。有许多不同种类的图表可供选择，每种类型都完全可定制，以适合网站的主题。iCharts有交互元素，可以从Google Doc、Excel表单和其他来源中获取数据。

二十四、Modest Maps

Modest Maps是一个轻量级、可扩展的、可定制的和免费的地图显示类库，这个类库能帮助开发人员在他们自己的项目里能够与地图进行交互。

二十五、Raw

Raw局域非常流行的D3.js库开发，支持很多图表类型，例如泡泡图、映射图、环图等。它可以使数据集在途、复制、粘贴、拖曳、删除于一体，并且允许我们定制化试图和层次。

二十六、Springy

Springy设计清凉并且简答。它提供了一个抽象的图形处理和计算的布局，支持Canvas、SVG、WebGL、HTML元素。

二十七、Bonsai

Bonsai使用SVG作为输出方式来生成图形和动画效果，拥有非常完整的图形处理API，可以使得你更加方便的处理图形效果。它还支持渐变和过滤器(灰度、模糊、不透明度)等效果。

二十八、Cube

Cube是一个开源的系统，用来可视化时间系列数据。它是基于MongoDB、NodeJS和D3.js开发。用户可以使用它为内部仪表板构建实时可视化的仪表板指标。

二十九、Gantti

Gantti是一个开源的PHP类，帮助用户即时生成Gantti图表。使用Gantti创建图表无需使用JavaScript，纯HTML–CSS3实现。图表默认输出非常漂亮，但用户可以自定义样式进行输出(SASS样式表)。

三十、Smoothie Charts

Smoothie Charts是一个十分小的动态流数据图表路。通过推送一个webSocket来显示实时数据流。Smoothie Charts只支持Chorme和Safari浏览器，并且不支持刻印文字或饼图，它很擅长显示流媒体数据。

三十一、Flot

Flot是一个优秀的线框图表库，支持所有支持canvas的浏览器(目前主流的浏览器如火狐、IE、Chrome等都支持)。

三十二、Tableau Public

Tableau Public是一款桌面可视化工具，用户可以创建自己的数据可视化，并将交互性数据可视化发布到网页上。

三十三、Many Eyes

Many Eyes是一个Web应用程序，用来创建、分享和讨论用户上传图形数据。

三十四、Anychart

Anychart是一个灵活的基于Flash/JavaScript(HTML5)的图表解决方案、跨浏览器、跨平台。除了图表功能外，它还有一款收费的交互式图表和仪表。

三十五、Dundas Chart

Dundas Chart处于行业领先地位的NET图表处理控件，于2009年被微软收购，并将图表产品的一部分功能集成到Visual Studio中。

三十六、TimeFlow

TimeFlow Analytical Timeline是为了暂时性资料的视觉化工具，现在有alpha版本因此有机会可以发现差错，提供以下不同的呈现方式：时间轴、日历、柱状图、表格等。

三十七、Protovis

Protovis是一个可视化JavaScript图表生成工具。

三十八、Choosel

Choosel是可扩展的模块化Google网络工具框架，可用来创建基于网络的整合了数据工作台和信息图表的可视化平台。

三十九、Zoho Reports

Zoho Reports支持丰富的功能帮助不同的用户解决各种个性化需求，支持SQL查询、类四暗自表格界面等。

四十、Quantum GIS(QDIS)

Quantum GIS(QDIS)是一个用户界面友好、开源代码的

GIS客户端程序，支持数据的可视化、管理、编辑与分析和印刷地图的制作。

四十一、NodeXL

NodeXLDE 主要功能是社交网络可视化。

四十二、OpenStreetMap

OpenStreetMap是一个世界地图，由像您一样的人们所构筑，可依据开放协议自由使用。

四十三、OpenHeatMap

OpenHeatMap简单易用，用户可以用它上传数据、创建地图、交流信息。它可以把数据(如Google Spreadsheet的表单)转化为交互式的地图应用，并在网上分享。

四十四、Circos

Circos最初主要用于基因组序列相关数据的可视化，目前已应用于多个领域，例如：影视作品中的人物关系分析，物流公司的订单来源和流向分析等，大多数关系型数据都可以尝试用Circos来可视化。

四十五、Impure

Impure是一个可视化编程语言，旨在收集、处理可视化信息。

四十六、Polymaps

Polymaps是一个基于矢量和tile创建动态、交互式的动态地图。

四十七、Rickshaw

Rickshaw是一个基于D3.JS来创建序交互式的时间序列图表库。

四十八、Sigma.js

Sigma.js是一个开源的轻量级库，用来显示交互式的静态和动态图表。

四十九、Timeline

Timeline即时间轴，用户通过这个工具可以一目了然的知道自己在何时做了什么。

五十、BirdEye

BirdEye是Decearative Visual Analytics，它属于一个群体专案，为了要提升设计和广泛的开源资料视觉化发展，并且为了Adobe Flex建视觉分析图库，这个动作以叙述性的资料库为主，让使用者能够建立多元资料视觉化界面来分析以及呈现资讯。

五十一、Arbor.Js

Arbor.Js提供有效率、以力导向的版面配置演算法，抽象画图表组织以及筛选更新的处理。

五十二、Highchart.js

Highchart.js是单纯由JavaScript所写的图表资料库，提供简单的方法来增加互动性图表来表达你的网站或网站应用程式。目前它能支援线图、样条函数图。

五十三、Paper.js

Paper.js是一个开源向量图表叙述架构，能够在HTML5 Canvas 运作，对于初学者来说它是很容易学习的，其中也有很多专业面向可以提供中阶及高阶使用者。

五十四、Visualize Free

Visualize Free是一个建立在高阶商业后台集游InetScoft开发的视觉化软体免费的视觉分析工具，可从多元变量资料筛选并看其趋势，或是利用简单地点及方法来切割资料或是小范围的资料。

五十五、GeoCommons

GeoCommons可以使用户构建富交互可视化应用来解决问题，即使他们没有任何传统地图使用经验。你可以将实社会化数据或者GeoCommons保存的超5万份开源数据在地图上可视化，创造带交互的可视化分析作品，并将作品嵌入网站、博客或分享到社交网络上。

传统的数据可视化工具仅仅将数据加以组合，通过不同的展现方式提供给用户，用于发现数据之间的关联信息。近年来，随着云和大数据时代的来临，数据可视化产品已经不再满足于使用传统的数据可视化工具来对数据仓库中的数据抽取、归纳并简单的展现。新型的数据可视化产品必须满足互联网爆发的大数据需求，必须快速的收集、筛选、分析、归纳、展现决策者所需要的信息，并根据新增的数据进行实时更新。因此，在大数据时代，数据可视化工具必须具有以下特性：

(1)实时性：数据可视化工具必须适应大数据时代数据量的爆炸式增长需求，必须快速的收集分析数据、并对数据信息进行实时更新;

(2)简单操作：数据可视化工具满足快速开发、易于操作的特性，能满足互联网时代信息多变的特点;

(3)更丰富的展现：数据可视化工具需具有更丰富的展现方式，能充分满足数据展现的多维度要求;

(4)多种数据集成支持方式：数据的来源不仅仅局限于数据库，数据可视化工具将支持团队协作数据、数据仓库、文本等多种方式，并能够通过互联网进行展现。

数据可视化技术在现今是一个新兴领域，有越来越多的发展、研究等数据可视化分析，在诸如美国这些国家不断被需求。企业获取数据可视化功能主要通过编程和非编程两类工具实现。主流编程工具包括以下三种类型：从艺术的角度创作的数据可视化，比较典型的工具是 Processing.js，它是为艺术家提供的编程语言。从统计和数据处理的角度，R语言是一款典型的工具，它本身既可以做数据分析，又可以做图形理。介于两者之间的工具，既要兼顾数据处理，又要兼顾展现效果，D3.js是一个不错的选择。像D3.js这种基于Javascript的数据可视化工具更适合在互联网上互动的展示数据。

这么大的湖，我哪里知道湖里到底有多少条鱼呢？

◎ 编辑 / 协会市场处　张楠　　图 / 崔峻珩

如何计算湖中鱼的总数？这个也算是个老问题了，各种方法都有，答案也是千奇百怪。

今天就借助这个案例，我们来了解一些基本的抽样方法和概率分布。

有人给出极端的回答：把湖水抽干。认为只有抽干了，才能知道到底有多少条鱼，而不是通过一些方法来预估、估计湖里鱼的数量。

湖里的鱼条数是一个问题，但是我们也希望通过这一个问题，来解决很多类似的问题。

如果真把湖水抽干，用数数的方法，再重新填满湖泊，这样又费时又费力，岂不是很不划算？

这里，介绍一种方法：“抓与重抓”

首先，抓捕一批鱼，比如说1000条，然后给鱼打上标记，之后再放回湖里。过一段时间，等这些鱼均匀地分布在湖中后，再抓一批上来，假设他们又抓了1000条，其中有50条做过标记，这意味着湖里有大约50/1000=5%的鱼都做了标记。因此，可以得出结论，湖里大概有1000/5%=20000条鱼。

用到两个重要的概念：1.抽样，2.分布。

1.抽样。用抽取的样本数据来计算总体。

2.二项分布理论。概念如下：

设随机变量X的可能值是0，1，2，…，n，而概率函数是

其中。这种分布叫做二项分布。

例如，设一批产品共N个，其中有M个是次品，即次品率p=。对这批产品进行放回抽样，即每次任取一个产品，检查其质量后仍放回去，如此连续抽取n次，则在被抽查的n个产品中的次品数X服从二项分布B(n,p)。

而对应鱼的这个例子，是总数不知道，次品对应为标记的鱼数量，进行放回抽样，是不是很类似了？

看似很简单的运用一个比例方法，其实用到了很多量化、概率的思想。当我们把他总结出来的时候，对于其他类似问题就可以回答的很标准、很贴切了。

总结：用案例生动直观的引导记住概念理论，是不错的方法。

接下来，我们再用统计学中的正态总体参数的区间估计，来估算鱼总量的大致范围，因为范围会有误差，所以尽量减小误差。计算误差时只需改变样本方差的算法即可，其他都不变。

这里用到的正态总体参数的区间估计公式如下：

假设前提：总体服从正态分布，而样本方差等于方差。总体均值的置信水平为1–a的置信区间是

本例中的样本方差是用群体内的样本比例乘以非样本比例，即做标记的比例（0.05）乘以没做标记的比例（0.95），结果是0.0475。将样本方差除以样本数，取平方根，结果是。从而得到做标记的鱼占鱼的总量的90%的置信区间（这里是份额，不是总量），[0.05–（0.007*1.645），0.05+（0.007*1.645）]，结果是3.8%~6.2%。由于做标记的鱼是1000条，因此湖中鱼的总数是16256~25984（1000/0.062=16256，1000/0.032=25984）。

看似是一个比较大的范围，但假设我们以前的不确定性水平很高，校准估计的范围也只是2000~5000条，所以这一范围已经大为缩小了。而且我们用的是90%的置信区间，如果置信区间可信度达到95%，得到的区间就会更小一些。

如果我们当初放养了5000条，现在仅仅是想知道鱼的总数是增加了还是减少了，那么任何大于6000的数字都表示鱼的总数增加了，超过10000条当然更好。

如果把初始范围和相关阈值都考虑在内，不确定性显然已经大为减小，误差也在可接受范围之内。实际上，我们完全可以在第一次抓捕中只抓250条鱼，然后放掉，再抓250条，也就是说抽样俩个只有前面的1/4。假设做过标记的鱼在第2次重抓时所占比例也是5%左右，那么我们对鱼的总数超过6000条仍然很有信心，也就是6000仍然在90%置信区间内。

这种通过抽样来揭示全貌的方法特别有用，这种方法已经用于评估美国人口普查局统计遗漏的人数、未知的潜在顾客数量等问题。未能看到整体全貌，并不意味着不能对它进行估算。

从本质上说，抓与重抓是两次独立抽样，比较两次抽样的重合程度，可以估计群体总数。对应你的行业，你能想到什么例子呢？？

附加：很简单的问题，可是能反应很多统计知识，并利于我们去理解。这里举的例子很简单，方法用到抽样、分布、参数估计等等，概念很拗口，但是结论很简单，有兴趣的去学习这些才会更有意思。

数据挖掘经典算法系列之朴素贝叶斯

◎ 译 / 协会数据中心　编辑 / 协会市场处　张楠　图 / 崔峻珩

1.概念

贝叶斯定理（Bayes theorem）是一种把类的先验知识和从数据中收集的新证据相结合的统计原理。

需要了解的前提概念有两个：

1.条件概率；2.概率乘法定理。

(1).如果我们在事件B已经发生的条件下考虑事件A的概率，则这种概率叫做事件A在事件B已发生的条件下的条件概率，记作。

我们用掷骰子的案例来简化这个概念：

掷骰子，其中“所得点数为奇数”记为事件A；”所得点数大于1”记为事件B。求以下事件的概率。（1）事件A、事件B各自发生的概率；（2）事件A、事件B同时发生的概率；（3）在已知掷的点数大于1的条件下，点数为奇数的概率。

用事件发生的计数情况，

可以直接得到:P(A)=$\frac{1}{2}$，P(B)=$\frac{5}{6}$，P(AB)=$\frac{1}{3}$。

第三问：已知掷的点数大于1，那么总数变为5，奇数点只有两个，则P(A|B)=$\frac{2}{5}$。

在考虑已经发生的事件A情况下，总数是在变化的，这就和原来的总体有区别了，条件概率的含义也就明白了。

(2).在明白条件概率的基础上，就可以得到概率乘法定理：

设事件A的概率P(A)>0，则在事件A已发生的条件下事件B的条件概率P(B|A)=$\frac{P(AB)}{P(A)}$。

那么　P(AB)=P(A)*P(B|A)=P(B) *P(A|B)

从而不难得到 $P(B|A)=\frac{P(A|B)*P(B)}{P(A)}$，即贝叶斯公式。

另一种解释用先验概率和后验概率来命名，内容是一样的。

P(A)是A的先验概率，因为它不考虑任何B方面的因素。

P(A|B)是已知B发生后A的条件概率，也被称为A的后验概率。

则后验概率=先验概率*调整因子

而对于朴素贝叶斯，为什么要叫做朴素？Naive的直译，意思为简单的、朴素的、天真的。因为此算法的前提假设是类别之间不相关，也就是相互独立，这种假设是很强的假设，有时候无法证明，所以就加上了“朴素”两个字。

2.场景

经常用于分类，最典型应用场景：过滤垃圾邮件

3.特点

（1）.算法逻辑简单，方便实现；

（2）.具有自我学习功能，也就是说使用的越多，那么分类的效果就越好。

4、算法示例

某个医院早上收了六个门诊病人，确诊如下：

症状	职业	疾病
打喷嚏	护士	感冒
打喷嚏	农夫	过敏
头痛	建筑工人	脑震荡
头痛	建筑工人	感冒
打喷嚏	教师	感冒

现在又来了第七个病人，是一个打喷嚏的建筑工人。请问他患上感冒的概率有多大?

$$P(感冒|打喷嚏 * 建筑工人) = \frac{P(打喷嚏 * 建筑工人|感冒) * P(感冒)}{P(打喷嚏 * 建筑工人)}$$

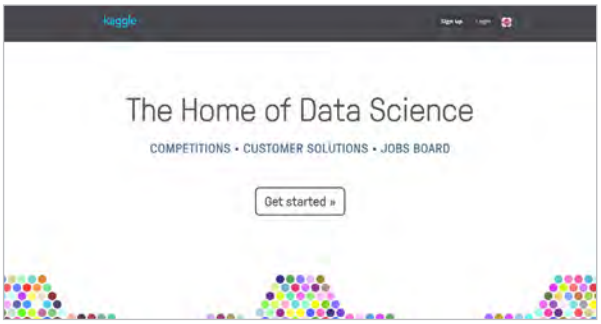
根据贝叶斯的朴素假设，打喷嚏和建筑工人独立，所以

$$上式 = \frac{P(打喷嚏|感冒) * P(建筑工人|感冒) * P(感冒)}{P(打喷嚏) * P(建筑工人)} = \frac{\frac{2}{3} \times \frac{1}{3} \times \frac{3}{6}}{\frac{3}{6} \times \frac{2}{6}} = \frac{2}{3}$$

朴素贝叶斯算法就是这样，根据历史记录得到各个特征的统计概率，如上式中的P(打喷嚏|感冒)、P(建筑工人|感冒)、P(感冒)、P(打喷嚏)、P(建筑工人)都是通过计数法得到的概率。

统计算法在Kaggle数据科学竞赛的成功

◎ 文 / Lillian Pierson P.E.　编辑 / 协会市场处　张楠　图 / iStock Photo



最近，数学建模平台Kaggle举办了一个大数据联合竞赛来预测股票价格的短期变化。 联合举办的另一个平台BattleFin——也是致力于众包投资分析人才的发现和培养。参赛选手的新闻数据和情绪数据由RavenPack公司提供,然后要求使用这些数据来构建模型，进而预测价格变化。运用这些模型和预测数据，交易员和投资者在做投资决策的时候将用获得的信息来改进风险预警，进行投资。

Steve Donaho博士是大数据联合竞赛的赢家,其他三个获胜者都是kaggle请来的。 事实上,Donaho博士在Kaggle比赛中的出色表现为其赢得了一个在250987名选手中靠前的名次。在其中一个时点上,Donaho在整个Kaggle平台选手中是排名第一的。 这次成功充分说明了Donaho博士在数据科学方面的创造力,聪明和灵敏性。在统计视图网站的独家采访中,Donaho博士讨论了他在数据科学方面的兴趣和Kaggle比赛的成功。

1、通过Kaggle比赛，你认为最有用的统计机器学习算法是什么?对于你自己而言，通过使用这些特定的方法，最大的收获是什么?

在过去的几年里，我发现GBM算法(广义boost回归模型)在R软件中是非常有用的,广泛适用于各种不同问题。 我使用GBM算法的第二个用处是完成了好事达保险公司的一个购买预测比赛，第三个用到的地方是在德勤保险客户流失预测的比赛中。之前，我开始使用XGBoost算法，它在本质上是类似于GBM算法的,但是计算要更快一些，而且对功能进行了一定的改进。而最近,我也被数字运营商Criteo、Tradeshift、Avazu举办的在线学习算法比赛所吸引。对于量很大的数据,在线学习技术能迅速给出不错的结果，并且不用使用很多的内存。

2、当你参加Kaggle比赛的时候，你采用什么标准方法?

我通常在比赛的开始花相当多的时间只是筛选数据,并且在我应用任何学习算法之前深入了解它。 有时这会给我创造一个很好的竞争优势——例如在好事达的比赛中,我发现某些组合的产品永远不会发生在美国各州。 排除这些组合，给我和我的伙伴节省了很多时间，形成了一个很大的竞争优势。另外在开始阶段，我会先试试一些简单的方法，我称它为“改善基线”:我先选择一个简单的想法,然后调整不同几个方面，来看我可以节省多少资源。我做这些有几个原因:1)有时我发现一些相对简单的解决方案,执行效果很好(复杂的不一定是更好的)；2)在实践中我发现客户喜欢简单的解决方案,这样他们能够把握掌控它；3)如果一个解决方案是做得很好,我想了解是什么驱动它成功,对于简单的模型算法这是很容易发现的。如果你直接从复杂的解决方案开始,很难知道是什么驱使着成功,并且不知道这样的复杂性是否有必要。

3、 是什么启发你开始参加Kaggle比赛?

我第一次听说Kaggle是在2011年一篇《华尔街日报》文章中。 数据科学比赛听起来很有趣。 在我的正常工作状态下有一周的休息时间,所以我参加了比赛，而距比赛结束也仅剩下一周的时间。 我用一个假名BreakfastPirate来签约,因为我认为我应该拿不到好的名次。比赛结束的时候，我发现在第一次比赛竟然得了第十名,而比赛过程中的状态让我感觉很棒，使我沉浸于其中。

也许有一部分读者的真正激情在于分析——在这种情况下,这些人应该被告知:数学,计算机等等只是用来帮助他们分析的支撑技能和工具。

4、你为什么参加Kaggle比赛？从中你获得了什么呢?

首先,它很有趣！我是一个完完全全的数字爱好者。 我热衷于把我的全身心投入到一组新被设定的数据中,不断的挖掘它、分析它。对于我们了解行业也是很有意思的,在工作之前，我并没有了解过零售销售、航班到达时间,非洲土壤成分,流感预测,点击率预测等等方面。第二,它迫使我去学习新技术和新算法。 我经常筛选获胜者发布的解决方案,我学习最聪明的、新的方法。Kaggle在过去的一年中，无疑是最具有竞争力的比赛。如果我看到选手赢得比赛用的是我之前并没有使用过的算法,我就要强迫我自己学习这种算法,以保持竞争力。这就是我开始使用XGBoost算法的原因。 第三,它是数据科学家们分享想法的一个社区。 是的,它是一个比赛,但在留言板上我们还可以分享很多的想法,这就变得相当的有趣了。

5、什么是你进入数据科学领域的初衷?

当我还在高中时,我得到的唯一职业建议是，“你擅长数学，你应该成为一个工程师。 “所以我去大学学习，让自己成为一个工程师。 我知道我喜欢电脑所以我主修计算机和电子工程。 当我攻读学士学位的时候,我发现我对软件比硬件更感兴趣。 所以我继续努力，攻读了计算机科学硕士学位和博

士学位。当我马上要完成博士学位的时候,我意识到，“我真的不喜欢电脑，正如我不喜欢身边所有的同学一样。 我真正想做的是分析数据,而电脑只是在我追逐数据分析梦想过程中的一个工具。”我花了这么多年和完成这么多学位，才明白我的能力不是数学。 我真正的能力是有良好的分析技能,并且我喜欢分析事物。不幸的是,在我高中的时候,分析能力不是很容易被定义,所以没有人能够说，“你有良好的分析技能,对于喜欢分析的人们这有一系列的职业道路。”但愿如今学校在知人识才和超越方面做得更好。 “你擅长数学。你应该成为一个工程师。 “但是以防万一,也许有读者真正的激情在于分析——在这种情况下,这些人应该被告知:数学,计算机等等只是用来帮助他们分析的支撑技能和工具。他们需要的不是结果本身,而是明白如何达到目的。

更多关于Steve Donaho的资料: Steve Donaho博士有20年关于海量数据的架构解决方案经验。 他已经在众多领域的Kaggle比赛中保持了前十的地位，这些类型包括股票市场情绪分析、保险、名称解析,零售销售预测,医药销售预测,和航班到达时间预测。在开始他的Kaggle分析之旅之前,他是Mantas公司的分析主管(现在是Oracle金融服务公司的一部分),为金融服务行业提供商业智能。 在Mantas公司的时候，他是一个发明家，也是一个先驱者，创造了四个关于分析方法的专利。 他的关于检测欺诈和内幕交易的算法已经出版了,并在多个数据库知识发现(KDD)会议中提出并讨论。他专业知识的领域包括欺诈检测、洗钱检测、金融市场、银行和经纪、医疗保健、电信和客户分析。

XiData
犀·数

大数据时代 de 通行证

咨询热线：
010-59000076

报名QQ：2853092077

数据分析员
招生开始啦！

www.cdachina.cn

微信：wxchinacpda

一、行业背景

中国商业联合会数据分析专业委员会（以下简称“协会”），是经国务院国有资产监督管理委员会审核同意、中华人民共和国民政部正式批准和登记的中国数据分析行业唯一的行业协会，在数据分析行业发展及专业知识的培养方面有着不可超越的优势。目前数据分析职业培训分为数据分析员（初级）和项目数据分析师（高级）两种。

二、专业背景

协会针对企业基础数据分析岗位推出的职业技术证书——“数据分析员”职业资格认证培训工作全面开展，证书由工信部颁发。这是继工信部授权协会开展全国项目数据分析师考试资格认证之后，推出的基础数据分析技能认证考试。

三、认证前景

大数据时代，越来越多的企业意识到依靠数据分析做出的决策给企业带来的好处，但是在数据分析人才选聘中，应聘者没有数据挖掘和数据分析技巧，企业也无相关系统培训，导致用人单位和应聘者之间无法良性沟通，而“数据分析员”的认证培训课程既可以解决学生无相关工作经验不能马上上岗问题，又解决了用人单位急需基础数据分析人才的困境。

因此，为顺应社会需求，协会与工信部联合推出数据分析行业的初级技能证书 ——《数据分析员》证书

- 解决企业中基层数据分析人才的培养问题
- 界定数据分析行业中的层次，顺应市场需求
- 促进毕业生就业，使学员适应公司不同岗位

四、培训及考核

数据分析员培训内容分为基础知识和技术技能知识两部分，形式为远程教学。

- 数据分析基础
- SQL数据库原理
- 数据挖掘技术概论
- EXCEL 、SPSS两种数据分析工具实操（多种分析模型的使用）
- 参加工信部组织的全国统一考试，通过后由工信部考试中心颁发《数据分析员》证书

五、招生对象

1、从事数据分析行业或对初级数据分析技术感兴趣的在职工作人员。

2、在读的计算机、会计、统计、营销等专业的在校学生。

六、收费标准

全国统一收费标准为**1200元** ，包括学习视频课件、考试费、证书费。

七、报名方式

报名地址：中国商业联合会数据分析专业委员会
北京市朝阳区朝外大街乙6号朝外SOHO C座931室

联系方式：肖老师 、赵老师　010-59000076 / 010-59000991转630 、631

联系邮箱：zhaojy@chinacpda.org

数据分析师事务所激活大数据行业

◎ 文 / 云科技时代 宁川　　编辑 / 协会市场处　潘宗祥　　图 / 崔峻珩

着大数据的流行，越来越多的企业开始部署大数据采集技术，在生产、制造、管理和经营过程中，积累了海量数据。如今，企业面临的问题，是不知道该怎么挖掘这些数据的价值。数据价值的挖掘在很大程度上不仅仅是对现有数据集的分析，还可以通过引进其它数据源，在不同的数据集之间建立关联，从而为业务创造新的价值。

在2015年3月12日IBM大数据白皮书的媒体沟通会上，IBM大中华区全球咨询服务部副合伙人、大数据与分析中国区负责人谢国忠分享了一个把不同大数据集相关联并产生创新型商业价值的案例。国内有很多风力发电厂，而风力发电和气候波动有关联。IBM与国内某风力发电企业合作，通过传感器采集数据预测风力，在天气数据与机组发电功率之间建立了精准的预测模型，让该电厂获得了巨大的收益。

如果企业不知道如何利用自己的数据集，或如何与其它数据集相关联而产生创新型商业价值，该怎么办？一个咨询服务性行业由此应运而生，这就是项目数据分析师事务所。项目数据分析师事务所是由项目数据分析师发起成立的数据分析行业中介服务机构，早期主要业务范围包括投资项目评估、经济效益评价、项目数据分析研究、项目融资、投资项目策划、投资中介等。我国最早的项目数据分析师事务所成立于2005年，而今天的项目数据分析师事务所的业务范围已经由早期的投融资项目分析扩大到了多个行业的数据分析服务。

早在2003年底，工信部电子行业职业技能鉴定指导中心根据国家财政部、国家发改委关于规范长期投资项目数据分析方法及与国际接轨的总体精神，设立了“项目数据分析师”培训项目，并制定出项目数据分析师培训、考试及管理办法。2004年1月1日，深圳作为项目数据分析师全国试点开始考培工作，我国首批“项目数据分析师”诞生，拉开了中国数据分析行业的序幕。2005年4月，全国第一家项目数据分析师事务所经工商局审批成立，从此在西安、深圳、成都、北京等地诞生了多家项目数据分析师事务所。数据分析专业事务所的出现，是我国数据分析行业的一个里程碑事件。

2008年4月，数据分析行业的全国性行业组织——中国商业联合会数据分析专业委员会正式成立。中国商业联合会数据分析专业委员会（CDAC）是以项目数据分析师事务所等企业为主体，与项目数据分析业相关的科研院所、大专院校、经营性企业、服务性企业和相关团体与个人自愿组成的全国性行业组织。2009年8月，数据分析行业的第一个行业标准在行业专家及全体事务所的支持下正式发布。

截止到2015年初，全国数十个省市已组建了100多家专业的项目数据分析师事务所，并在数据分析技术所涉及的各个领域发挥着重大的作用。

从涉及的领域上来看，数据分析行业已经从单一的投融资业向多元化多领域发展转变。“项目数据分析师”（CPDA）是中国商业联合会数据分析专业委员会自主研发的数据分析行业专业培训课程，该课程自2003 年在国内正式开办以来，已经在全国二十多个省市培养了近万名项目数据分析师。通过专业的课程培训及考试后，学员可获取主管行业机构和主管部委认证机构颁发的《项目数据分析师证书》和《项目数据分析师职业技术证书》，而获得了“项目数据分析师”证书的学员在具备创业条件后，经过协会的备案以及工商部门的注册即可成立“项目数据分析师事务所”。

据了解，在数据分析服务领域，项目数据分析师事务所是中国数据分析行业唯一被认可从事数据分析服务的专业机构。除了此类型专业机构，还有混业经营的其它机构提供数据分析服务，包括市场调研公司、IT 公司、网络公司和系统集成公司等，而对这些商业机构来说，数据分析是其业务组成部分之一。

在2014 年，大数据应用的典型行业仍然是以互联网为基础的电子商务。除了电子商务行业相对成熟外，数据分析已经逐渐向金融、电信、物流、医疗以及交通、教育等领域拓展并取得了初步成效。然而，大数据的理念、技术和应用模式等整体还处于初级阶段，谈概念的多、真正用的少，传统企业在理解什么是大数据，制定大数据战略以及整合多种数据源方面还存在诸多问题，大数据技术与模式的改造和适配还要经历长期过程。

虽然我国项目数据分析师事务所已经有十年的历史，但我国大数据分析的黄金时代才刚刚开始。在未来十年甚至更长的时间里，以中国商业联合会数据分析专业委员会为代表的行业组织，与中关村大数据产业联盟、中国企业大数据联盟、长三角大数据联盟、深圳大数据产业发展促进会等数据共享联盟，以及新近出现的一线大数据联盟等，将真正推动大数据与传统企业业务的结合。

项目数据分析师事务所发展概况

◎ 文 / 节选自《2014年度中国数据分析行业年度发展报告》　编辑 / 协会市场处　张楠　图 / 崔峻珩

项目数据分析师事务所是中国数据分析行业唯一被认可从事数据分析服务的专业机构，接受中国商业联合会数据分析专业委员会的监管。项目数据分析师事务所的设立需经中商联数据分析委的严格审批，在运营接受中商联数据分析委的监督和检查。

截止2014年底，中商联数据分析委共收到设立项目数据分析师事务所的有效申请558份，经考察和资质审核共有54家项目数据分析师事务所通过审批，审批通过率不足1/10。审批通过率不高的部分原因是因为中商联数据分析委“保质控量，宁缺毋滥”的审批原则，另外一方面因为市场被“大数据浪潮”催热，有很多盲目追热，导致申请数量的激增。

1、事务所地域分布情况

经审批通过的事务所遍布全国超过20个省市自治区，数量较多的区域主要有北京、陕西、山东、安徽和重庆，其中北京地区的事务所数量将近占全国事务所总数量的三分之一，其余地区事务所数量相对较少。

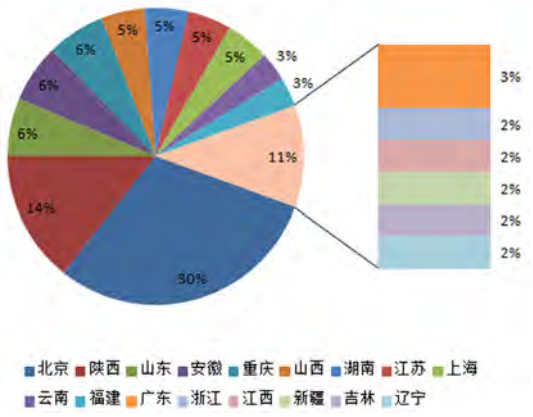


图1–1 事务所以北京最为集中

2、事务所团队建设情况

(1) 经营规模

从调研数据来看，目前的项目数据分析师事务所规模普遍偏小，人数最多的一家事务所为45人（但其兼职人员占到一

半），人数低于12人的事务所数量超过总调研事务所数量的一半，占比达52%，且多为9–12人。

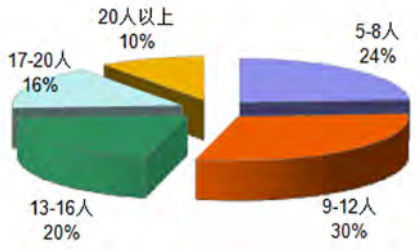


图1–2 事务所普遍规模偏小

(2) 人员学历构成

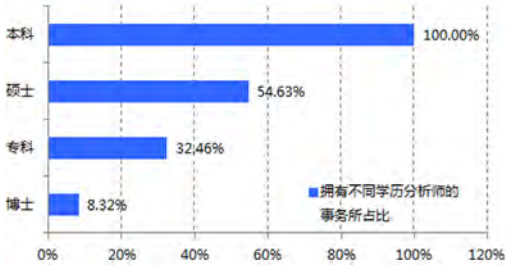


图1–3 所有的事务所都有本科学历的分析师

各事务所的项目数据分析师的学历主要由专科、本科、硕士和博士构成，由调研数据可看出，每个事务所都有本科学历的分析师，有32.46%的事务所有专科学历的分析师，有54.63%的事务所有硕士学历的分析师，有8.32%的事务所有博士学历的分析师。博士硕士学历的分析师比去年增加11.78%，事务所整体人员水平较以前提高很多。

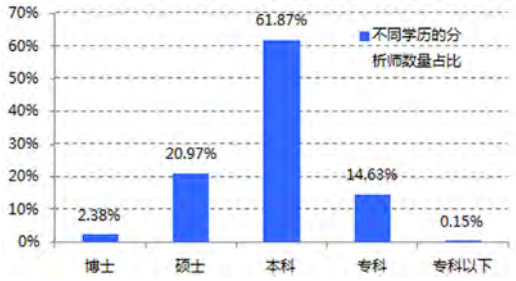


图1–4 从业项目数据分析的人员学历以本科为主

(3) 人员从业年限

大部分事务所都有从业1–5年的分析师，各事务所中有从业3–5年的分析师的事务所占比达79.96%，各事务所中有从业1–3年的分析师的事务所占比达76.42%，同时也有17.58%的事务所拥有还未从事项目数据分析工作的刚拿到证书的分析师。

从就业人数的分布来看，目前的项目数据分析师的从业年限主要集中在1–5年（占比87%），其他就业年限的分析师数量较少，目前的项目数据分析师从业队伍相对年轻。

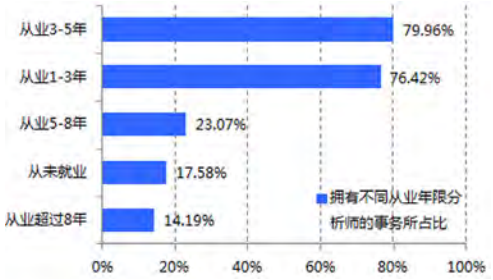


图1–5 大部分事务所都有从业1–5年的分析师

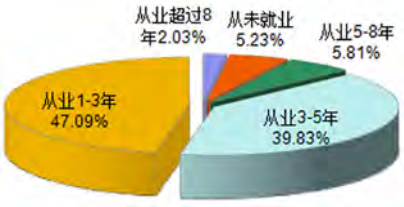


图1–6 从业人数中以从业1–5年的分析师为最多

(4) 人员专业背景

从事务所分析师的专业情况看，拥有财务、税务相关专业分析师的事务所占比最大，高达85.22%；而有统计、数学相关专业分析师的事务所占比69.83%；有工商管理、企业管理相关专业分析师的事务所占比为56.05%；拥有计算机相关专业分析师的事务所占比为38.40%；拥有经济贸易相关专业分析师的事务所占比为31.44%。

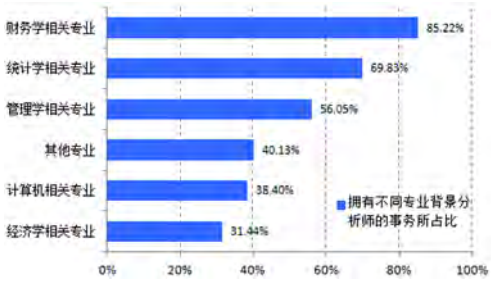


图1–7 超过85%的事务所拥有财务、税务相关专业的分析师

从从业人数来看，有财务税务相关专业的人数占比为30.79%，有统计、数学相关专业的人数占比为19.49%，而计算机相关专业的人数从之前几乎没有到去年增加了14.69%，其余类相关专业的人数占比相对较低。

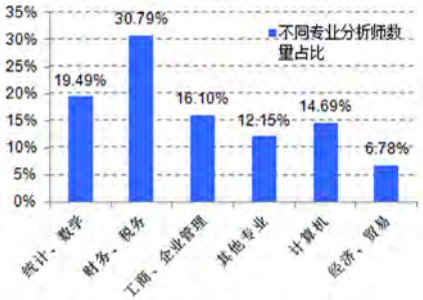


图1–8 从业分析师中财务、税务相关专业背景人数最多

(5) 人员行业背景

从事务所分析师的从业背景来看，大部分事务所都有财务注会、咨询行业以及企业高管的背景，拥有这些从业背景分析师的事务所比例分别为66.67%、56.86%和49.02%，同时也有一部分事务所拥有税务审计、风险评估背景、计算机背景的分析师，还有少部分事务所拥有一些其他行业背景的分析师。

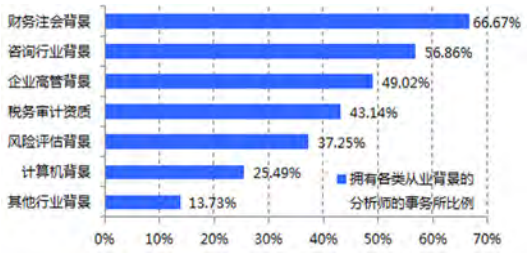


图1–9 大部分事务所都拥有财务、咨询行业背景的分析师

从从业人员构成背景看，事务所以有咨询行业背景的人员为最多，占比为27.47%，财务、注会资质的人数次之，占比为21.60%，拥有企业高管背景的人员占比为14.81%。

其他从业背景中有电商行业背景、通信行业数据分析背景、工程造价分析背景、计算机行业背景等，占比不足10%。

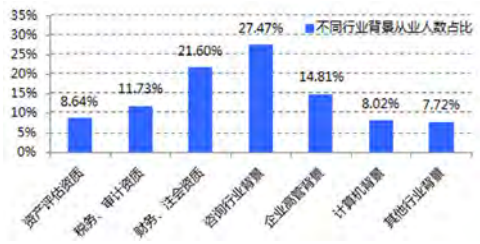


图1–10 拥有咨询行业背景的分析师数量最多

(6) 人员综合素养

从项目数据分析师所拥有的技能证书看，大部分的分析师只有3种以下技能证书。

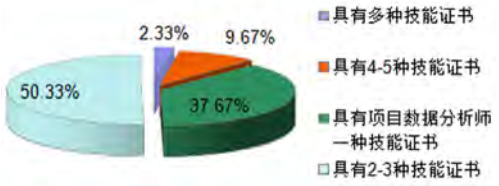


图1–11 拥有2–3种技能证书的分析师占比最大

3、事务所业务情况

(1) 业务类型

各事务所业务定位以企业经营决策为主，兼顾其他业务的事务所占比为36.51%；以投融资业务为主，兼顾可研、评估类业务的事务所占比为14.29%；以投融资业务为主，兼顾企业经营决策业务的事务所占比为12.70%；以可行性研究报告业务为主的事务所占比为11.11%；只做过银行信贷评估、商业计划书的事务所占比为4.76%；只做投融资业务的事务所占比3.17%，只做企业经营决策类业务的事务所占比1.59%。

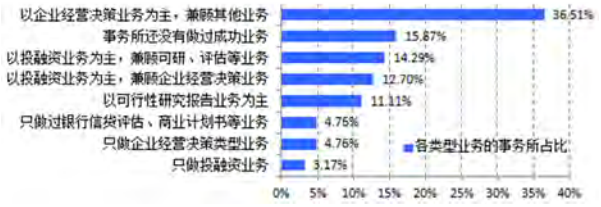


图1–14 以企业经营决策业务为主、兼顾其他业务的事务所占比最高

(2) 新增行业情况

2014年各事务所新增的业务中，涉及的行业主要有房地产行业、制造业和农林牧渔业。建筑业和电子商务业也有一定业务，其他行业的业务相对较少。

新增行业	新增数量
房地产业	18
制造业	14
农、林、牧、渔业	12
建筑业	12
电子商务	11
信息传输、计算机和软件业	10
文化、体育和娱乐业	9
金融业	8
采矿业	6
教育	6
批发和零售业	5
科学研究、技术服务和地质勘查业	4
医疗卫生、社会福利业	4
水利、环境和公共设施管理业	4
交通运输、仓储和邮政业	3
公共管理与社会组织	3
住宿和餐饮业	3
居民服务和其他服务业	3
电力、燃气及水的生产和供应业	2
租赁和商务服务业	2
其他行业	2
国际组织	0

表1–1 事务所涉及的房地产业、制造业的业务相对较多

(3) 合作企业类型

从事务所合作过的企业类型来看，目前事务所主要合作的企业类型仍然是国内中小型私企，有55.56%的事务所与中小型私企进行过合作；有44.44%的事务所与股份制企业进行过合作；与政府、部委等机构有合作的事务所占比为31.75%；与国有企业和国内大型私企有过合作的事务所占比约为28.57%；与外资企业有过合作的企业占比为14.29%。

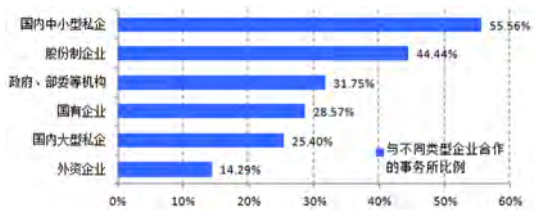


图1–21 目前与事务所合作的企业主要是中小型私企

(4) 服务项目类型

从事务所已经提供的服务项目类型来看，市场进入分析类和竞争分析类的项目数量占比最大，精准营销类、产品定价类、客户满意度以及用户使用习惯及态度分析类项目相对较少，品牌分析类和广告效果评估类项目更少，目前还没有事务所从事过产品测试类和渠道暗访类的项目。

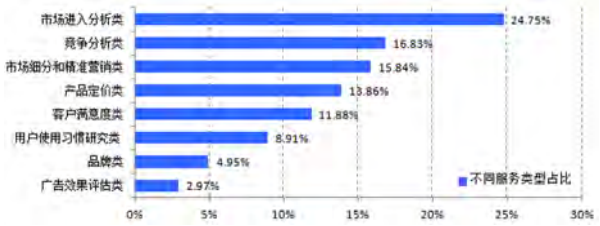


图1–22 不同服务类型的项目数量差距不大

(5) 经营分析内容

目前事务所涉及的经营决策服务类的内容主要集中在基础的数据预测分析、数据录入整理和数据的分类与收集上，提供过该种内容服务的事务所占比均为31.75%；其他类型的一些内容服务相对较少。

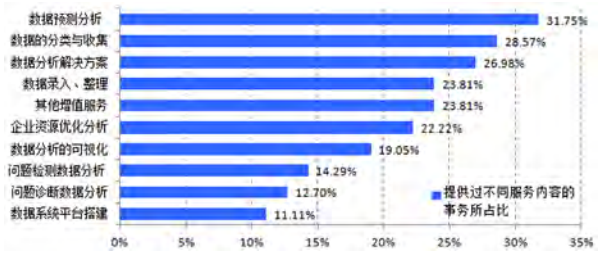


图1–24 事务所的服务内容仍然比较基础

4、事务所自身认识情况

(1) 事务所自身对优势的定位

事务所优势	得分
事务所的专业性（专业团队、数据分析研究）	18
事务所人才涉及范围较广、结构合理	15
业务资质较多，业务范围较广	10
不同的区域性优势	9
数据分析行业的专注性（只做数据分析）	8
规模不大、业务转型灵活	5
已确立企业经营的数据分析方向	4
政府资源较多	3
有比较专注的行业	3
经营管理制度健全	3
市场优势明显	3
注重国内市场宣传	3
长期从事投融资业务，经验较充足	3
有较充足的企业管理经验	2
人员对新事物接受程度较高	2
与当地金融机构有良好的合作关系	2
事务所宣传推广做的比较好	2
自身品牌的可塑性强	1
有专业的网络推广公司做宣传	1
各项证书、资质齐全	1
在建设自己的数据平台	1
组织优势，有工会组织、党支部	1
已形成专业的报告产品	1
业务量、财务较稳定	1
找到对口行业快速打开市场	1

表1–2 专业性优势是事务所最主要的优势

从事务所的优势列表可以看出，事务所的优势主要集中在事务所专业性、事务所资源以及事务所的内部管理上。

(2) 事务所自身面临的问题和困难

从事务所的调研情况看，行业认可度低、市场难打开、业务难开展是目前事务所普遍存在的一个问题。

事务所资源的问题和困难	得分
地区企业对数据分析行业认可度不高，业务难开展	14
事务所知名度较小，客户资源非常有限	9
市场开拓比较困难	8
分析师专业水平有待提高	8
目前以投融资业务为主，转型到企业经营较困难	7
人才储备不足、招聘困难	5
数据积累不足、数据收集难度大	5
经营性核心技术缺乏，缺乏系统的培训，业务难开展	4
事务所宣传力度较小、缺乏有效宣传	3
目前事务所还没建立自己的核心竞争力	3
市场竞争激烈、业务开拓难度较大	3
工作形式和内容偏于模板化，人才积累受限	1
事务所刚起步，行业经验较少	1
对于市场扩展人员没有系统的培训方案	1
对经营决策数据分析认知不清	1
事务所决策定位不准	1
事务所与市场接合度不高	1
品牌还有待进一步树立稳固	1

表1–3 行业认可度不高业务难开展是事务所面临的最主要困难

(3) 事务所对量化经营与投资的认识

最后从调研问卷的情况看，事务所对经营决策类业务和投融资业务的认识主要体现在以下几点：

	企业经营类业务	投融资类业务
业务价值	经营决策类业务可以更好的提升事务所从业人员的业务能力和事务所的专业水平	量化研究类业务从广义上说应该归属在经营决策类以内，一份高水平的量化研究类的报告，可以为项目方的投资决策和风险规避提供重要的参考
业务特点	经营决策类业务能跟企业建立长期的业务联系，服务时间较长，有利于数据分析服务的价值体现，也有助于事务所业务的稳定	量化（投融资）研究类业务所涉及的领域虽然也很广阔，但量化（投融资）研究类业务比较单一、是短期的、一次性的服务
业务开展	经营决策类业务将会是各事务所未来发展的主要方向，但目前开展经营决策类业务也存在比较客观的难题，市场认知度不高，需要协会及各事务所同心协力共同宣传影响市场的认知度	在经营决策类业务得到较好地发展之前，量化研究类业务不能彻底抛弃
业务趋势	随着数据分析行业的发展，会有更多企业认识到精准决策的重要性，必然会产生海量的企业经营决策业务需求，因此经营决策类业务未来必将达到一个高峰，行业协会目前倡导全国事务所向经营决策类业务转型具有一定的前瞻性，是正确的引导。	量化（投融资）研究类业务目前不是很好开展，因为投融资行业，在市场中信誉度不高，很多投融资客户需要数据分析报告时，都附带要求提供其他公司的资质，并不是为了项目而要数据分析，损坏了数据分析行业的科学性和价值

表1–4 企业经营类业务与投融资业务的比较

“两会”给数据中心领域带来了哪些新的机会？

◎ 编辑 / 协会市场处 张楠 图 / 崔峻珩

2015“两会”指的是全国政协第十二届三次会议和十二届全国人大三次会议，从2015年3月3日开始，到2015年3月15日结束。“两会”代表将从人民中得来的信息和要求进行收集及整理，传达给党中央，向政府有关部门提出选民们自己的意见和要求，并进行广泛讨论。“两会”上的很多提案往往关注民生、关注社会热点问题，很多提案和人们的生活息息相关，所以会引起广泛关注。那么“两会”的召开具体能给数据中心领域带来哪些新的信息呢？“两会”会给数据中心领域带来哪些新的市场机会呢?本文越俎代庖，代各路专家、媒体来大话分析一下。

“两会”讨论的都是一些人们广泛关注的行业和市场。和相比其他领域，数据中心只能算是一个小众市场，只能算做是科技领域里的很小部分，而科技包含非常广泛，包括：信息安全、信息技术、集成电路、电子技术、网络技术等等，所以我们只能从“两会”上的科技热点来捕捉数据中心领域的新机会。先来看看来自科技领域的“两会”代表，这些人代表了全体科技人的发声，从他们提交的方案和言语，往往能够得到只言片语，将影响着整个科技产业的发展方向。

我们注意到今年将有28位来自IT、家电、电信、互联网行业科技领域代表参加本次“两会”，其中互联网行业的代表人士已经增加至6人，反映这个行业影响力或者重要性正在逐渐提高，实际也的确是这样。BAT三家的老板成为了国务院和中央领导的座上宾，经常出现在各种国家科技研讨会议上，并不断发声。

在今年“两会”总理在政府工作报告中，曾三次提及互联网发展，并且提出要制定“互联网+”行动计划。互联网不仅要自己行业发展，还希望通过互联网带动其它行业共同发展。所以互联网领域必然成为2015年投资主线，今年热度或将持续，政府将大力支持互联网和其它传统行业的有机结合，借互联网模式盘活更多的行业，这样必然给互联网带来更多的机会，而数据中心是互联网行业赖以生存的部分，互联网所有的业务都要通过数据中心来实现，所以不难现象2015年数据中心行业将继续火热，甚至超过以往，数据中心的新建和扩容将再增速，以便更好地支撑互联网行业的发

展，支撑传统行业与互联网的融合，所以数据中心领域的投资在2015年必然会持续增加。

大数据已经火了一段时间了，但是在“两会”上依然是热点，众多代表都提到了对“大数据”的发展，其广阔的发展前景毋庸置疑，大数据将成为未来世界政治、经济、政府管理的重要依据。国家发改委主任徐绍史就强调：对国家政策措施的调整要有更富的“工具箱”，但现在一些统计数据不匹配、相背离的现象很多，因此需要应用大数据、云计算以及多种分析评价方法来分析经济运行情况，来掌握新常态下经济的新特点，探寻新常态下经济运行的新规律。

全国人大代表、浪潮集团董事长兼CEO孙丕恕提议政府从国家层面统筹规划，出台指导意见和行动规划明确政府数据开放工作的主管部门，尽快着手制定全国统一的政府开放数据标准，从技术操作层面为全国政府开放数据的大共享奠定基础。“大数据”只有大才有意义，那么从国家层面进行推动，必然给“大数据”更广阔的发展舞台。同样数据中心是大数据最好的土壤，没有数据中心，“大数据”一切都是空谈。而数据中心要承载这些“大数据”，原有的系统是无法完全支持的，就需要不断引入一些新技术，比如云计算或者虚拟化技术，才能部署“大数据”，所以数据中心要进行技术改造，以便可以尽快落实“大数据”技术。“大数据”需要海量的数据统计与计算，需要占用数据中心大量的计算和网络资源，所以数据中心依然要不断扩容，提升计算和网络处理的能力。

随着互联网和4G网络的普及，网络空间已成为继陆、海、空、天外的“第五空间”。不过网络给经济社会带来快速发展的同时，个人、企业等信息资料泄露也愈演愈烈。信息安全是2014年“两会”的热点，今年“两会”上依然是热点。很多代表关注国家信息安全，对于如何保持好个人网络信息安全和国家信息安全，提出了各种想法和建议。

信息已经是一种宝贵的资产，涉及个人利益和国家利益，数据成为个人和国家的最重要资产，所以安全问题尤为重要。有代表就提议对个人信息保护方面专门立法，通过法律来全面系统地就个人信息保护的基本原则、个人信息主体的权利、监管机构及其职责进行详细阐述。信息安全立法对于保护信息安全是非常必要的，这也同时给数据中心安全技术带来了蓬勃发展，因为保护信息安全需要部署一些安全设备和技术，自然给数据中心安全带来受益。同时信息安全保护如何取证，留证都是数据中心安全技术需要考虑的事情，在这里产生出了巨大的市场机会。

今年的“两会”给科技行业带来了新的活力，同时也给数据中心市场带来了春风。无论是“互联网+”，“大数据”还是“信息安全”这些议题都离不开数据中心市场的健康、高速发展，数据中心是这些议题最终实施落地的最佳部分。总体上看，“两会”给数据中心带来了新的市场机会，为数据中心持续的高速发展提供了动力，尤其是数据中心的安全技术，支撑“大数据”部署的新技术，这些都将成为数据中心市场里的新热点。

探索多渠道数据服务平台 以专业服务赢得社会信任和市场认可

——湖南翰林项目数据分析师事务所发展介绍

◎ 文 / 湖南翰林项目数据分析师事务所　　编辑 / 协会会员处　石爱英　　图 / 崔峻珩

随着中国经济新常态的出现，数据分析行业也进入了新的发展阶段。湖南翰林项目数据分析师事务所紧跟形势发展，探索多渠道数据服务平台，以专业服务赢得社会信任和市场认可。我们具体向以下方面发展：

一、我们积极向邵阳市财政局推荐我们公司，介绍公司的经营状况和技术力量，得到了财政局的认可。获得了代理记帐资质，为湖南翰林项目数据分析师事务所今后从事会计类基础数据代理及财务数据分析提供了有力的支持，也为开拓审计法定业务以外的增值业务创造了条件。有了财政局的文件和资质，2014年我们顺利地以湖南翰林项目数据分析师事务所的名义为邵阳市检察院基建财务数据进行专业数据服务，解决了该单位多年以来的办公楼基建财务数据长期挂往来账未转固定资产以及总分账有差异等遗留问题，为该单位执行新的行政事业单位会计制度提供了固定资产期初基础数据，以我们的专业服务赢得该单位的认可。

二、在市场经济条件下，信息获得不对称， 项目数据分析师事务在数据采集方面遇到瓶颈，相关信息资料难以取得。2013年国务院《征信业管理条例》颁布，企业的信息采集纳入法制规范管理，湖南翰林项目数据分析师事务所紧跟形势发展，成立了湖南翰林企业征信有限公司，专门采集企业信息，由湖南翰林项目数据分析师事务所所长卿启伟担任公司董事长（法人代表）。

目前我们已经建立了翰林企业征信网站，翰林企业征信系统也已经建成，处于系统测试阶段，我们正在与有关科技公司开发征信数据分析产品，并正在整理材料向人民银行征信管理部门办理征信机构备案手续。湖南翰林企业征信有限公司的成立，主要面向中小企业融资及经营服务数据分析市场，充分彰显数据价值，助推企业发展。为湖南翰林项目数据分析师事务所数据分析提供了数据资源，为事务所的发展插上了腾飞的翅膀。

三、随着中国经济新常态下城镇化的推进，房地产开发及棚户区改造方面的数据服务需求越来越大，但这方面很多领域还保留相关行政审，我们经过多方努力，整合了我们的资源，收购了本地一家房地产评估公司---邵阳华信房地产评估有限责任公司，湖南翰林项目数据分析师事务所所长卿启伟具备房地产评估师资格，有多年房地产评估工作经验，经协会审核合格后担任邵阳华信房地产评估有限责任公司董事长（法人代表），由同样具备房地产评估师资格的湖南翰林项目数据分析师事务所的数据分析师卿上顺担任技术总监。湖南翰林项目数据分析师事务所的其他员工同时成为邵阳华信房地产评估有限责任公司的员工。

四、 经过多次考察，邵阳市本地最大的本科院校邵阳学院已经与湖南翰林项目数据分析师事务所签订协议，确定湖南翰林项目数据分析师事务所为邵阳学院会计学、财务管理、资产评估学等专业的校外实习基地。湖南翰林项目数据分析师事务所所长卿启伟多次应邀为邵阳学院会计系的学生进行培训讲座、体会分享，向在校学生宣传了数据分析行业，得到学校师生的好评和欢迎。

除上述外，湖南翰林项目数据分析师事务所与联盟执业单位的业务平台继续巩固和加强，包括湖南南方会计师事务所、邵阳南方资产评估公司、湖南新融达咨询集团、邵阳科信税务师事务所、邵阳南方造价咨询公司、邵阳南方司法鉴定所等中介机构。我们坚持以湖南翰林项目数据分析师事务所为基础，以全国优秀事务所为招牌，探索多渠道数据服务平台，以专业服务赢得社会信任和市场认可，在中国经济新常态下，搏击数据分析市场大潮流，为事务所的发展夯实基础，为数据分析行业发展添砖加瓦！

联系地址：邵阳市大祥区万基银座小区1单元4楼
联系方式：卿启伟　13187299268
网　　址：www.hlsjfx.com

河南明豫项目数据分析师事务所

河南明豫项目数据分析师事务所是在中国数据分析行业权威，中国商业联合会数据分析专业委员会的指导下成立的。是河南首家项目数据分析师事务所。依托中部地区最大的物资集散地——郑州，致力于打造富有品牌影响力的专业性数据服务机构。

事务所拥有数据采集、数据分析、金融投资、工商管理及软件开发等各方面的专业人才，并聘请国内、省内相关行业专家教授作为资深顾问，所内分析师均有中国商业联合会数据分析专业委员会颁发的项目数据分析师（CPDA）资质证书。事务所立足于建筑、房地产及农业，整合多方资源，业务还可涉及高新技术、投资、软件开发等各个行业领域，致力于为企业及事业单位提供企业项目可行性分析，投资项目评估、分析、策划，项目数据分析报告及经济效益评价等专业信息类服务，在创立之初就受到多家企业关注，并逐步发展为紧密的合作伙伴。

随着大数据时代的到来，数据分析在现代企业经营管理决策中起着越来越重要的作用。明豫利用事务所的人才优势，从各项量化指标入手，保证数据的真实性，并以科学的定量分析为主，预测分析等方法为客户特供专业化的数据分析服务，用全面、精准的数据分析报告帮助客户规避决策风险，为客户投资经营保驾护航。

事务所着力打造“公正，客观，务实，创新”的品牌形象，始终坚持公正中立的立场，秉持科学客观的态度，秉承严谨务实的的作风，运用不断创新的理念，全面提升执业水平，积极地为企业发掘每一个可利用的信息，为企业提供准确的数据分析结论和决策建议，提高企业核心竞争力，实现企业价值。做到成为中部中小企业的长期战略伙伴，为中国数据分析事业的发展贡献一份力量，力求成为业内公认、社会信赖、理论新颖、服务优质的中国数据分析领跑者。

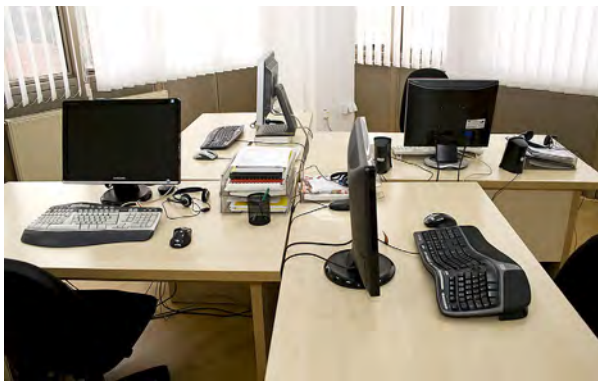
联系地址：郑州市管城回族区乾元街24#院东1单元2楼
联系方式： 孙春光 18539912993 / 13523080467

湖南中楚项目数据分析师事务所

湖南中楚项目数据分析师事务所位于湘楚腹地长沙，是经过湖南省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，公司致力于为客户提供深度数据分析、数据挖掘、市场研究服务。

事务所于2013年成立以来，在数据分析、挖掘方面均取得了长足发展。已与湖南地区零售行业建立起长期合作关系，为零售超市的发展和决策提供强有力的数据支撑；成功实施了一些科技型企业的数据分析与预测；在医疗领域我们进行了深度的研究与分析，并取得一些研究成果，目前与医院机构的接洽正在有条不紊的进行中，本年度有望建立合作关系。公司自始至终注重人才队伍的建设与发展，并定期进行学习与交流，在数据可视化、挖掘语言的掌握等能力方面也取得了明显提升，确保为客户提供更好的数据服务。

中正无偏，楚天阔。在协会的大力支持下，中楚正以饱满的热情，稳健的步伐开创湘楚大地数据分析的新局面，也



热忱欢迎业内同行的广泛交流，期待与各行业的紧密合作，助力本土企业发展。

联系方式：刘凡 13975190282 / 0731-85211248
网　　址：www.cncpda.com
企业微信：cncpda
新浪微博：湖南中楚项目数据分析师事务所

迎接数据科学的拐点

◎ 文 / 周庭锐　　编辑 / 协会市场处　潘宗祥　　图 / 崔峻珩

2014年，我们一起见证了数据科学（data science）在全球范围内的迅猛勃兴，而中国，毫无例外地参与了这场盛会。数据科学是大数据时代里，面对波涛汹涌、排山倒海而来的海量数据，通过超高速计算，对人事时地物甚至隐含概念进行探索挖掘、特征识别和统计分析的科学，这种科学曾经是极少数武功高强科学家的专利，但是由于世界上还存在一些为数更少但能明察秋毫洞烛未来的资本家，他们很快就理解到这种新兴科学的商业价值，于是在风险资本的推动下，让这些科学家摇身一变成为新兴科技的企业家，这其中一部分人，例如扎克伯格、贝佐斯，已经是纳斯达克天空的闪耀星星，但是还有更多的人正站在风口上，盼着忽然而至的东风，让他们刹那间就能翩然飞翔。阿里巴巴可能是其中相当璀璨的明星之一，从事电子商务，在中国蹲伏了14年之后，成长成一只年营业额高达2480亿美元的庞然巨物，并一举在纳斯达克IPO套得218亿美元，轻易秒杀稍早京东商城18亿美元的意气风发。事实上阿里巴巴所赖以为生的并不是电子商务，而是电子商务背后带来的数据流，通过对这些数据流的拦截与计算，衍生出包括对零售与物流商家的种种数据服务，以及针对卖家与买家设计的各式各样金融商品。阿里巴巴成功上市的意义并不仅仅只是让华尔街一众白皮肤的投资人狂野高喊“爸爸”而已，它代表着中国电子商务发展的一个里程碑。而从“网络外部性”的效应看，这很可能也是中国电子商务发展的分水岭，从此再难有第二家能够做到这种规模的“纯”电商。

非纯电商的O2O实践

事实证明，我们在2014年开始看见大量“不纯”电商的兴起，他们誓言要让“纯”的传统电商从中国版图里消失。例如地产商万达在他的商业卖场里广泛布下wifi的天罗地网，通过记录商场顾客的手机号码，捕捉他们在卖场里从两脚移动到心智选择的消费轨迹。

实际上万达的策略目标远不仅止于此，在部分城市里万达已经开展O2O实验，希冀融合对线上线下消费行为的分析与理解，做真正全渠道的生意;类似的故事正在全国范围里此起彼落地上演：步步高开始经营云猴电商开放平台;大润发凭借自己商超模范生的身份，一方面在天空放出飞牛来攻击阿里巴巴的天猫，另一方面还同时拉开“千乡万馆”计划，渗透进小区服务中心，打算在地面打造1万家网络体验馆，形成以小区为核心的O2O闭环，垄断线上线下所有的消费数据。这股风潮连物流业者顺丰也坐不住了，嘿客一出手就是500家门店全国同时开业，充分利用他们在城市物流的优势，通过二维码构筑绵密的、超低运营成本的城市自动售货亭（city kiosk）网络。实际上，中国所有行业里的玩家几乎全都急不可待地抢着加入这场竞争。先不提那些不间断地侦测用户行为的服务类运营商，例如百度、腾

讯、360、三大电信公司等等，他们早就从方方面面各种生活切入点布下天罗地网来拦截广大用户每天所思所想所问所求的信息；现在连传统制造业都已经大张旗鼓地对电商发起绝地大反攻。例如海尔的空气盒子、小米内建的用户监测系统、众多家电厂商正在寻思的结合wifi的智能家电，在打着捕捉用户数据的如意算盘，希望通过对用户行为的持续监测与随之而来的深刻理解，实现他们隐秘不宣的赢利策略。这是2014年在中国市场里如火如荼的实况。2014年是中国厂商意识到消费者数据价值的元年，而展望2015年，厂商对消费数据真正的深刻解析才刚要开始。

对大数据解析的不足

到此为止，绝大多数中国厂商对于数据解析的理解，还停留在算算百分比、看看时间轴上的数量趋势、至多加上点均值计算，如此而已。就连行业里的模范生淘宝天猫所提供的数据魔方都仍然停留在这点水平上。相对而言，百度可以算是个异类了，在各种预测上算尽了作为坐拥大数据大牛公司的基本责任。坐拥大数据的目的是为了预测，通过过去的已知来预测即将发生的未知，而不仅仅只是描绘现况而已。真正坐拥大数据是为了从这些浩如烟海、快速积累、驳杂不纯的数据里寻出蛛丝马迹，作为制定企业策略的依据。这些蛛丝马迹，例如小至消费者在商场里被摄像头捕捉到的一颦一笑、通过手机声纹监听搜集到的抑扬顿挫、上传微博论坛照片颜色的调性、最近贴上微博文本里的轻声叹息、甚至可穿戴设备监测到的皮肤电位差、脉搏速度、体温、行为活跃度等等，都可以通过数学模型转化为对人们情绪的估计，像这些隐微的估计才是数据科学家们最感兴趣的课题。

知微才能见著

对消费行为蛛丝马迹的掌握是成就伟大企业的基本能力。商人察言观色掌握客户情绪的目的是顺水推舟，和气生财，所谓“龙之为虫也，可扰押而骑也。然其喉下有逆鳞径尺，人有撄之，则必杀人。人主亦有逆鳞，说之者能无撄人主之逆鳞，则几矣”（《韩非子·说难》），就是这个道理。何况数据挖掘更不限于情绪估计而已。例如通过网络爬虫爬取电商销售数据、商品热卖排行，我们可以非常容易地通过数学模型来估计品牌、定价、促销、具体商品的种种细部设计等等因素，是如何对销售构成影响的，通过这样的计算我们甚至可以确保实现零库存的可能。不论是情绪还是影响销售的因素，我们都在寻找一些幽隐微小但是意义重大的信息来帮助我们做出正确的企业决策。营销的基本制胜原则在于高筑竞争壁垒，而有效的竞争壁垒必然是基于竞争对手难以捕捉的，甚至是无法觉察的信息。那些百分比、具体数字、甚至是均值，都是很容易被看见、被理解的强信息（strong signals），这样的信息我们知道，竞争对手也知道，所以策略价值不高；反而那些被杂讯淹没的、数量微小的、刚刚萌芽的、很容易被忽略的弱信息（weak signals），更可能是具备重大策略价值的信息。**数据科学提供我们这样的机会，从海量数据中去发掘隐微但是具有重大策略意义的信息，然后在竞争对手还毫无所知的情形下，突出奇兵，直接占领整个利基市场。**

钱学森的大数据思想：开放的复杂巨系统

◎ 文 / 云科技时代　宁川　　编辑 / 协会市场处　张楠　　图 / 崔峻珩

我国著名科学家钱学森是中国科学院及中国工程院院士、中国载人航天奠基人、中国两弹一星功勋奖章获得者，曾被誉为“中国航天之父”“中国自动化控制之父”等。

钱学森还是一位系统学家。关于系统科学，钱学森曾明确指出，系统科学是从事物的整体与部分、局部与全局以及层次关系的角度来研究客观世界的。客观世界包括自然、社会和人自身，能反映事物这个特征最基本和最重要的概念就是系统。所谓系统是指由一些相互关联、相互作用、相互影响的组织部分构成并具有某些功能的整体。

钱学森提出了系统新的分类，将系统分为简单系统、简单巨系统、复杂巨系统和特殊复杂巨系统，生物体系统、人体系统、人脑系统、地理系统、社会系统、星系系统等都是复杂巨系统的代表，其中社会系统作为最复杂的系统又被称作特殊复杂巨系统。这些系统都是开放的，与外

部环境有物质、能量和信息的交换，所以又称作开放的复杂巨系统。

今天的“大数据”系统就是在新时代下出现的开放复杂巨系统。大数据是指由数量巨大、结构复杂、类型众多的数据构成的数据集合，所涉及的信息资料库规模巨大，无法在合理的时间内通过目前的主流软件工具达到撷取、管理、处理并整理为可帮助组织进行决策的信息。

大数据系统有4个V特性：数据量大（Volume）、数据种类多样（Variety）、实时性强（Velocity），蕴藏的商业价值大（Value）。特别是随着物联网的发展，越来越多的传感器被部署在城市、服务业、金融业、工业、农业、能源等领域，借助条形码、二维码、RFID等可唯一标识产品，传感器、可穿戴设备、智能硬件、视频采集等源源不断地产生着海量数据，相关领域的数据规模已经达到TB甚至是PB级。有统计表明，2013 年中国产生的数据总量是2012 年的两倍，相当于2009 年全球的数据总量。

按着复杂巨系统的概念，大数据系统不仅数据规模巨大且结构复杂，而且元素或子系统种类繁多、本质各异、相互关系复杂多变，在宏观与微观层次存在着复杂的关联度，相互作用机制不清楚，不能通过简单的数据分析的方法描述其宏观行为，而且大数据系统本身又是一个开放的系统。

开放的复杂巨系统的主要特性包括：开放性，系统对象及其子系统与环境之间存在着物质、能量、信息的交换；复杂性，系统中子系统的种类繁多，子系统之间存在多种形式、多种层次的交互作用；进化与涌现性，系统中子系统或基本单元之间的交互作用，从整体上演化、进化出独特的、新的性质，如通过自组织方式形成某种模式；层次性，系统部件与功能上具有层次关系；巨量性，数目极其巨大等。

钱学森把关于复杂巨系统的理论研究称作复杂巨系统学，并于1992年提出建设人机结合、从定性到定量的综合集成研讨厅体系的设想。综合集成研讨厅体系的概念就是将专家群体（各领域的专家）、数据和各种信息、计算机、网络等信息技术有机结合起来，把各种学科的科学理论和人的认识结合起来，基于网络构成的系统。该综合集成研讨厅

由研讨终端、中心研讨厅、研讨厅骨干网（Internet或WAN）、研讨厅管理服务系统、研讨厅信息资源库、以及分布各地的感兴趣的和相关的研讨群体与技术支持群体组成。

钱学森的“开放的复杂巨系统”思想对于研究今天的大数据有重大指导意义。人机结合的综合集成研讨厅体系，将专家体系、机器体系和数据体系有机结合起来组成智能系统，是研究和解决“开放的复杂巨系统”相关复杂问题的有效途径。

在钱学森之后，又有学者中科院院士戴汝为从智能系统角度提出了人工社会，在综合集成研讨厅体系基础之上，以人与网络计算机为单元，通过以复杂问题为牵引的交互和组织，形成了开放的人工社会，以求解复杂问题为目的，以人为计算中心。戴汝为曾师从钱学森，并于80年代后期率先在国内开展了人工神经网络研究。

结合钱学森与戴汝为的学术思想，将有助于当今社会对于大数据现象的思考、研究及指导实践。

大数据时代

怎么成为高薪金领?!

Answer here

Tel:400-050-6600

www.chinacpda.com

大数据时代的高薪攻略

微信扫一扫 : wxchinacpda

主办方：中国商业联合会数据分析专业委员会

创先机 赢未来

—— 暨2015大数据新品发布会

4月16日 • 北京广播大厦酒店

• CPDA项目数据分析师 •
• 犀数数据分析员 •
• DATAHOOP大数据智能分析平台 •

助您
全新解读
“大数据时代的核心 —— 数据分析”

关注微博微信，了解最新行业资讯

行业协会
官方微博

官方微信：
wxchinacpda

CDAC

会议时间：2015年4月16日
会议地点：北京市朝阳区建国门外大街甲14号 北京广播大厦酒店
报名方式：市场处 张女士　**报名电话**：010-59000067 / 18612766693
报名 QQ：2853092057　**报名邮箱**：zhangn@chinacpda.org