● 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

第三届中国数据分析行业峰会

The 3rd China Data Analysis Industry Summit. 2015

—— 触摸大数据本质

2015年 第2期 峰会专辑

09 促进企业数据化——百家企业扶持项目正式启动

10 2015第三届中国数据分析行业峰会 峰会指南

18 R语言最新技术导读

35 从"物"往"人"的零售进化

36 大数据治疗领导者"拍脑袋决策"流行病

41 北京市不同区县酒店分布及价格水平探究性分析

关注行业微博微信 了解数据分析行业前沿知识

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*

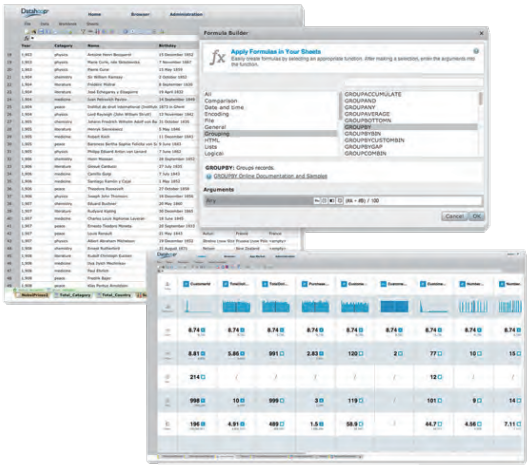
1 十多年实践经验积累，集成行业顶尖算法

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。

《中国数据分析》会员特刊
2015年第二期 总第22期 (峰会专辑)

主办
中国商业联合会数据分析专业委员会

编委
何林　石爱英　曹莹　黄平　崔欢欢　常琳

主编
张楠

出版时间
2015年7月

美工
崔峻珩

联系我们
中国商业联合会数据分析专业委员会
地址: 北京市朝阳区朝外soho C座9层 100020
电话: +86-10-59000991
传真: +86-10-59000991转 607

投稿

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

感谢您对《中国数据分析》的支持!

投稿邮箱：marketing@chinacpda.org

目 录 CONTENTS

P01 卷首语
突破行业瓶颈，推动企业发展

P04 协会动态
项目数据分析师获猎聘网VIP服务
"CPDA项目数据分析师"蝉联工信部优秀课程五连冠
"创先机,赢未来——2015大数据新品发布会"在北京拉开帷幕
全国各地数据分析公益沙龙活动火热进行中
中颢润(北京)项目数据分析师事务所成为协会副主任单位
重庆传晟携手重庆理工大学共建大数据实验室

P09 行业风向标
促进企业数据化——百家企业扶持项目正式启动

P10 峰会指南
+ 关于峰会　+ 合作伙伴　+ 峰会日程安排
+ 与会嘉宾　+ 路线指引　+ 联系方式

P18 "数"业专攻
R语言最新技术导读
用R语言进行时间序列的分位数回归
互联网的发展驱动大数据的发展
数据科学中的“数据智慧”
6张图带你看懂“块数据”
DeepMind背后的人工智能：深度学习原理初探
Spark成为大数据分析领域新核心的五个理由

P35 运"数"有道
从"物"往"人"的零售进化
大数据治疗领导者“拍脑袋决策”流行病
零售企业如何借助数据分析进行品牌定位
北京市不同区县酒店分布及价格水平探究性分析

P54 事务所风采
贵州华鑫成项目数据分析师事务所
北京鼎盛恒信项目数据分析师事务所
河南智宸项目数据分析师事务所
云南誉诚俊安项目数据分析师事务所
湖南中楚项目数据分析师事务所

突破行业瓶颈，推动企业发展

2015年是数据分析行业在中国发展的第12年，也是大数据概念在中国火速发展的一年，在刚刚过去的半年间，我们不断感受着“数据浪潮”的冲击：3月国务院总理李克强在两会上首次提出了“互联网+”的工作计划，加速了企业与云计算、大数据、物联网的结合；4月国务院办公厅印发的《政府信息公开工作要点》，鼓励企业、第三方机构、个人对政府公共数据进行深入的分析和应用；同期，行业主管机构主办的“创先机，赢未来——2015大数据新品发布会“，展现了企业数据化进程中技术与人才两大难题的解决方案；5月国际大数据产业博览会暨全球大数据时代贵阳峰会以“大数据+：数据驱动产业变革”为主题，围绕大数据未来发展趋势，从大数据发展过程中的关键和共性问题、全球大数据产业商机等方面进行了深入探讨。所有这一切，都向我们传达着这样一个理念：企业发展，要与大数据结合！

然而，如何让大数据落地，为其所用，是当前企业遇到的最大问题：还没有建立数据化管理的企业面临的是如何解决数据有效存储的问题；有数据没有应用的企业，考虑如何打通内部数据、搭建“大数据“分析平台；而正在进行数据应用的企业则考虑如何有效解决数据算法和分析难点，从而转化为商机。因此，面对上述情况，协会决定为企业踏实地办两件实事：一是启动长达5年时间的“百家企业战略转型扶持”项目，二是7月10日在江苏昆山举办第三届中国数据分析行业峰会。

对于前者，协会希望通过自有的技术开发能力及数据深度分析能力，为企业提供技术平台支持，降低数据应用门槛，帮助企业构建内部数据的完整性以及与外界数据的有效交互，同时号召全国项目数据分析师事务所与协会一同帮助企业完成“量化决策”模式的转变。对于后者，本届主题定义为“触摸大数据本质”，将以主题演讲和深度对话的模式，通过案例或场景的展现，与来自一线企业、数据分析师事务所、重点院校的行家学者一起为参会者答疑解惑，共同探讨大数据产业发展及大数据人才储备、技术应用等的前沿话题，探索各行业之间面临的数据分析应用问题的解决途径，突破企业在数据化进程中的困境。相信这将是一场覆盖面最广的行业盛会！

2015年，中国“数据化“的趋势已经势不可挡，扑面而来的”数据热浪“来势汹汹，身为行业主管协会，在夯实自身实力的基础上，带领事务所突破业务瓶颈，帮助企业实现数据落地，是行业赋予我们的使命！在此，协会发出号召，欢迎广大业内人士、企业、团体加入行业建设队伍中来，为中国大数据与世界接轨共同努力！

最后，预祝此次峰会圆满成功！恭祝全体数据分析同仁、支持和帮助行业成长的各位参会代表身体健康、万事如意！

中国商业联合会数据分析专业委员会

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

给自己一次机会，迎接无限精彩的未来！

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！

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！

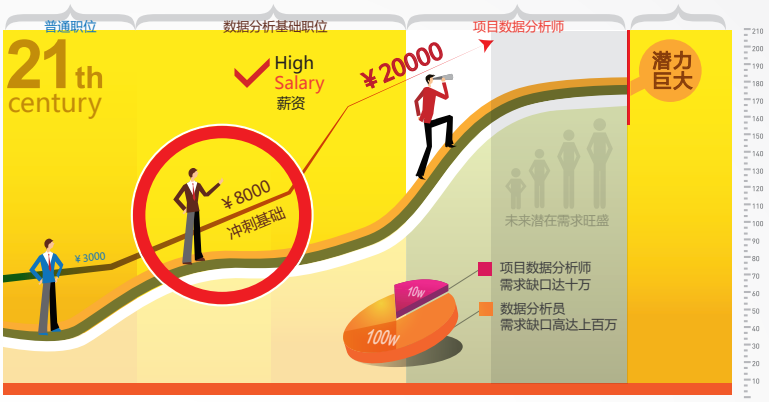
www.chinacpda.com

400-050-6600

犀·数 数据分析员

专业背景

协会针对企业基础数据分析岗位推出的职业初级技能技术证书——“**数据分析员**”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。



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，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

数据分析员课程架构

数据来源
- 互联网数据
- 内部数据
- 调研数据

数据处理
- 清洗
- 加工

分析方法
- 描述分析
- 推断分析
- 有差异吗？—— 方差分析
- 谁在影响它？—— 回归分析
- 明天会怎样？—— 时间序列
- 物以类聚 —— 聚类分析
- 你在几线城市？—— 因子分析

报告展示
- 优秀作品展示
- 数据可视化

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
- EXCEL 、SPSS两种数据分析工具实操
- 数据挖掘技术概论
- 考试通过后由工信部考试中心颁发《数据分析员》证书

联系地址：北京市朝阳区朝外soho C座9层 100020
联系方式： 赵老师 肖老师 010- 59000076 / 010-59000991转630、631
企业 QQ： 2853092077 电子邮件：zhaojy@chinacpda.org

项目数据分析师获猎聘网VIP服务

为保证项目数据分析师(CPDA)未来享有更优质的专业职业服务，中国商业联合会数据分析专业委员会（简称“协会”）与中高端人才互联网招聘平台猎聘网达成数据分析人才的战略合作。通过此次合作，猎聘网将为全国已获取CPDA证书的项目数据分析师提供VIP金卡会员服务， CPDA项目数据分析师将免费获取猎聘网的各项特权，及享受一对一的专业化猎头服务。

从2003年至今，项目数据分析师职业已发展12年。协会培养的近万名分析师分布在全国十几个省份，受到各行各业的一致好评。CPDA数据分析师课程是目前唯一受到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。通过深入的培训和严格的考核，CPDA证书的含金量逐年提升，目前已成为当之无愧的“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，不仅是大多数企业内部数据分析岗位，我国的项目数据分析师事务所均由取得CPDA证书的学员组成。目前全国已有百余家专业事务所，各事务所正在为IT、金融、医疗、零售、物流等领域的企业提供着决策支持服务，已得到社会各界高度关注。

项目数据分析师CPDA不仅是加入高端从业人群的象征，更是持证者成就事业的象征。此次猎聘网的关注，一方面我们可以看到项目数据分析行业的人才或已紧缺，另一方面对于CPDA项目数据分析师而言，未来的机遇也将更具多面冲击性和多面选择性。

“CPDA项目数据分析师”蝉联工信部优秀课程五连冠

CPDA“项目数据分析师”课程被工业和信息化部教育与考试中心评为2014年度行业教育培训优秀课程方案。中国商业联合会数据分析专业委员会被评为2014年度全国信息技术人才培养工程优秀培训基地。

2015年4月1日，工业和信息化部教育与考试中心培训工作会议在武汉召开，中国商业联合会数据分析专业委员会副秘书长蒋文伟先生代表协会出席了本次会议，并在会上领取了2014年度《数据分析行业教育培训优秀课程》荣誉证书以及《全国信息技术人才培养工程优秀培训基地》奖牌。CPDA“项目数据分析师”课程已经连续五届蝉联该奖项。

随着大数据热潮的兴起，与数据相关的各类培训项目竞相面市，或偏重技术层面，或偏向某一行业，很难有一款类似CPDA“项目数据分析师”课程这样既注重思维体系的建立、又注重全行业实战技能培养的全面课程。

作为中国最早的数据分析行业培训课程，CPDA课程在12年的发展过程中，一直注重课程体系的不断更新和完善，无论是教程的更新还是师资团队的提升建设，均受社会的良好口碑。“一分耕耘一分收获”，CPDA课程取得这样的成绩，是无数业内精英倾注极大心血成就的。五连冠的取得当之无愧!

目前CPDA“项目数据分析师”课程已经完成了三次大规模革新，其成果将在2015年新品发布会上进行全新展示。加入数据分析人才建设队伍，完善数据分析人才培养，企业才能真正进入数据决策时代。

协会期待着与更多优秀机构的合作。

“创先机，赢未来——2015大数据新品发布会”在北京拉开帷幕

2015年4月16日，由中国商业联合会数据分析专业委员会(简称“协会”)主办的“创先机，赢未来——2015大数据新品

发布会”在北京召开。来自全国各地的项目数据分析师事务所、CPDA授权中心、以及数据应用企业代表们参加了本次发布会。发布会吸引了新华网、新浪、搜狐、36大数据、砍柴网、云科技时代、中国软件网等众多媒体的关注。

协会会长邹东生在会上表示：大数据时代之所以产生，一个很重要的原因是数据体量的增大，使得企业利用“大数据”进行精准决策变得可行有效。正是因为在企业决策方面带来的变革，因此使得这个时代具有了划时代的意义。然而“大数据”只是时代的特征，它不是一个行业。数据本身不能带来价值，要通过分析，给企业决策带来帮助才能产生价值，因此数据分析才是一个行业，是在这个时代真正产生价值的行业。发挥大数据的价值，数据分析就要发挥作用。

数据分析作为一个独立的产业，在中国发展不过12年的时间，在这期间，行业由模糊逐渐转向明晰。越来越多的政府、企业和个人开始认识到大数据的核心价值在于数据分析。然而如何将数据分析与企业实际情况结合，这是中国数据分析行业面临的主要问题。

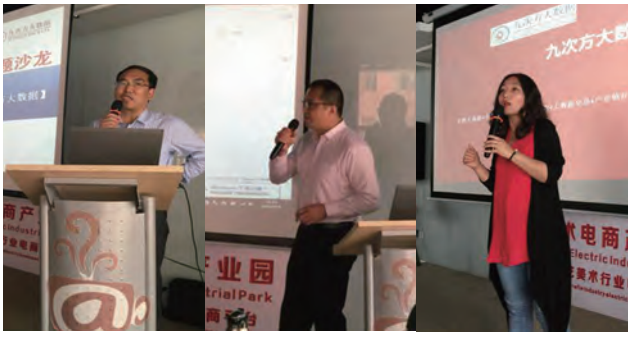
关于这一点，邹东生认为：与欧美发达国家相比，中国的数据分析行业还处于起步阶段，企业数据化程度偏低和数据分析人才缺失是目前面临的两大问题。

因此首先要帮助企业构建数据化平台，实现数据的有效

存储和联动。同时，随着时代发展越来越成熟，数据可以被越来越多加以存储的时候，发现数据背后决策性规律并加以应用的人将成为企业的灵魂人物，数据分析人才的储备和体量的增大对于一个行业的发展以及未来大数据与企业决策对接起到至关重要的作用。因此，只有解决好上述这两大问题，才能帮助中国企业在大数据时代真正腾飞。

协会自2008年成立以来，一直在项目引进、知识普及、行业监督、技术支持、数据分析人才培养及数据分析专业事务所建立等方面发挥着积极地的作用。正是看到了富化数据的产品时代面临的问题以及巨大的市场机遇，协会借由发布会上对Datahoop大数据智能分析平台以及两款数据分析人才培养计划的展示，让公众认识大数据时代的核心，让数据分析变得落地可行，让更多的机构加入到行业建设的队伍中来，从而真正激活中国的数据分析行业市场。

河南：互联网金融+大数据高端主题沙龙



2015年5月9日，河南 CPDA学友会于郑州组织召开互联网金融+大数据主题沙龙。中央电视台特约评论员、著名财经金融评论家余丰慧教授、九次方金融大数据常务副总裁胡媛媛女士和北京财富经济合伙人林元君先生受邀讲坛，十余名学友会成员及郑州当地企业管理人员参加。沙龙呈现了一场学术专业高端的讲座和充满互动思考讨论。

三位专家分别从互联网金融的本质、大数据的定义和起源、大数据在互联网金融方面的应用等三方面详细的阐述了在信息技术极度发达的今天，大数据和数据分析作为最科学、最客观的工具，已经被广泛应用于互联网金融等各行各业。

讲座结束后，三位专家和参会者进行了深度交流，活动气氛一度到达顶峰。

CPDA学友会成员敏锐的行业洞察力及对数据分析知识的专业性得到了三位专家和全场参会人员的一致肯定。

北京："数据分析在行业预测中的实际应用"数据分析公益沙龙

2015年5月16日下午中国商业联合会数据分析专业委员会（简称“协会”）在北京财富中心千禧公寓二层咖啡吧举办了“数据分析在行业预测中的实际应用”为主题的2015第二期数据分析公益沙龙活动，60余位来自金融业、保险业、电子商务、集团公司等不同行业的数据分析师及对数据分析感兴趣的高校在校学生参加了本期沙龙活动。

本期沙龙邀请了中国国际电子商务中心内流通首席分析师郭召芬女士及中颢润（北京）项目数据分析师事务所资深项目数据分析师王庆生先生。

郭召芬以“商业网点规划中的分析模型应用”为主题，在时间序列模型、回归模型、聚类分析、vorinoi模型、shift–share模型的应用方面进行了深入的分析与讲解。王庆生老师以“互联网数据分析——学习与应用”为题，综合实际案例，生动形象地从流量分析、销售分析、客户分析、营销效果分析以及市场预测性分析五大方面告诉大家处理这些用户数据的方法，通过众多数据模型的筛选、建立，实现量化经营决策。

讲座主题的专业性与讲师授课的趣味性深深吸引了每一位听众，大家认真的用手机记录、备分整个讲座过程。在互动环节中，二位专家就讲座中涉及到的知识点及实际工作中遇到的数据分析问题与参会者展开交流，对大家提出的问题进行了详细的解答。

河南："当大数据遇见电子商务"主题沙龙

2015年6月7日下午，“当大数据遇见电子商务”沙龙活动于河南电信文化营业部成功举办。活动由中国商业联合会数据分析专业委员会、项目数据分析师（CPDA）河南授权中心主办，河南MBA&EMBA同学会协办，郑州电信文化营业部五众领航店和CPDA学友会共同承办。沙龙特邀国家电子商务中心数据研究员李欣欣老师为大家做专题讲解，河南CPDA学友会和各行各业数据分析从业者、爱好者共50余人参加了此次沙龙活动。

李欣欣拥有多年国民经济行业数据研究经验，曾参与国家商务部城乡市场监测信息服务体系项目，大量撰写行业分析报告及分析文章，其中多篇文章得到国务院领导和商务部领导批示。

在活动中，首先河南MBA&EMBA同学会会长、河南CPDA学友会筹备组组长韩晓强代表同学会，学友会以及河南CPDA授权中心对李欣欣的到来表示感谢，并对各位嘉宾的到来表示欢迎，祝愿他们学有所获。他还介绍了数据分析协会以及河南CPDA学友会的成立宗旨和发展理念，得到了广大数据分析爱好者的共同认可。

随后，李欣欣通过丰富的图表展示，介绍了传统实体店零售和网络零售的区别，说明电子商务的出现在零售革命中的重要作用，以阿里巴巴集团和通用电气公司为例，介绍了电子

商务的应用带来的商业变革。电子商务企业的发展离不开大数据的支持，李欣欣用轻松诙谐的语言，介绍了大数据应用给电子商务企业带来的机遇与挑战。同时说明了在未来电子商务领域中，大数据人才的稀缺。鼓励大家尽早树立“大数据”应用的思维。

互动环节也是精彩纷呈，大家兴致高昂，纷纷提出关于数据分析领域的若干问题，李欣欣一一解答。期间的抽奖活动，极大的活跃了沙龙现场的气氛。

活动在大家热烈的掌声中圆满结束。许多参与者对李老师的精彩分享不禁感叹、意犹未尽。通过分享，广大学员对大数据的分析应用有了初步了解，体会到树立大数据思维重要性，同时表示有兴趣成为数据分析协会大家庭中的一员。CDAIS 2015

中颢润(北京)项目数据分析师事务所成为协会副主任单位

中国商业联合会数据分析专业委员会经过慎重考虑和严格审议，接受“中颢润（北京）项目数据分析师事务所”的申请，吸收其成为我会副主任单位会员，并颁发会员资质证书。

“中颢润（北京）项目数据分析师事务所”（简称：“中颢润”）发展至今已连续四届获得中国数据分析行业优秀事务所。自成立以来，中颢润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，坚持数据分析报告的客观性、专业性和实用性，在业界获好了良好的社会口碑。

在此协会祝愿中颢润在今后的发展中，可以发挥副主任单位会员职责，协助协会开展数据分析行业的推广工作，以自己精湛的业务能力服务行业，帮助更多的项目数据分析师事务所成长，为数据分析行业在中国的发展做出更多的贡献！CDAIS 2015

重庆传晟项目数据分析师事务所携手重庆理工大学共建大数据实验室

2015年5月21日，重庆理工大学大数据实验室建成并投入运行，这是重庆市首个专业的大数据实验室，将推动整个重庆市大数据技术人才培养与行业大数据实验分析研究发展。

重庆理工大学大数据实验室由重庆传晟项目数据分析师事务所负责承建，重庆理工大学数学与统计学院运营使用。实验室主要由大数据分析管理系统、问卷调查系统、项目数据分析系统三部分组成：

1、大数据管理分析系统包含：大数据分布式文件存储以及大数据分布式文件映射处理、大数据机器学习、大数据分布式数据库、分布式网络爬虫；

2、问卷调查系统包含：在线问卷设计以及发布、客户关系、人力资源、网站社区、心理测试、市场调研、教育调研、行业服务、媒体出版各问卷模板调研；

3、项目数据分析系统包含：数据多维模型设计、数据多维模型数据处理、数据多维模型报表、数据分析可视化。

重庆传晟项目数据分析师事务所还将和重庆理工大学数学与统计学院合作，共同构建行业数据分析典型应用模型和编写大数据技术人才培养课程课件，力求把实验室打造成重庆市专业的大数据技术人才培训中心。

重庆理工大学大数据实验项目，也是“重庆市大数据行动方案”关注的重点项目之一，目前主要研究领域是金融高频数据与地产数据，预计未来将结合政府大数据规划方案，提升打造成为大数据工程技术研究中心。CDAIS 2015

中国商业联合会数据分析专业委员会 主办

赢了产品却输在服务？ 赢了资本却输在运作？
赢了平台却输在客户？ 赢了内容却输在体制？
赢了开发却输在时间？
……
还不明白？
你赢的只是数据，输在分析。
中国商业联合会数据分析专业委员会 开启

财富时代的机遇

“2015—2020 百家企业 经营模式创新与战略扶持”

若你心有所向，何不从善如流？

不知道数据分析行业的微博微信？ 你OUT了！

行业协会官方微博

微信: wxchinacpda

WEB: www.chinacpda.org TEL: 010-59000065 Email：helin@chinacpda.org QQ：2853092062

促进企业数据化
——百家企业扶持项目正式启动

文 / 中商联数据分析专业委员会市场部 张楠　　图 / 崔峻珩

随着2015年大数据概念在中国的火速发展和“互联网+”工作计划的提出，企业加快了对云计算、大数据、物联网改革的步伐。目前，很多具有前瞻性的企业已经高度重视数据化服务的建设，部分企业已经开始构架自己的“大数据”分析平台，然而在平台构架、搭建过程中，面临“技术难度大”、“搭建费用高”、“技术服务公司对数据算法和分析不够专业”以及“存储数据有效性“、”数据与经营决策间的对接“等问题，更多的企业在大数据时代，心有所向，力所不及！

针对以上问题，中国商业联合会数据分析专业委员会（以下简称"协会"）作为数据分析行业全国性协会组织本着如下的目的，开展了为期5年的战略扶持计划：

1）响应中国商业联合会号召，加大对会员单位的服务力度，加快中国零售商企业适应大数据时代的步伐，促进其取得发展先机；

2）免费为企业搭建数据分析平台，提供技术支持，降低数据应用门槛，协助企业构建内部完整数据从而达到与外界数据之间的有效交互，促进企业数据分析实现可视化和专业化；

3）推广企业成功案例，引导更多企业重视数据分析，壮大数据分析市场，维护行业事务所（企业）经营发展空间。

百家企业扶持计划活动开展时间为2015年6月1日至2020年5月31日。在此期间，中商联数据委将对加入活动企业的大数据程度进行评估，为企业数据保全提供方案；根据企业需求及硬件情况，为企业引入Datahoop平台接入服务，使企业数据得以有效存贮；根据企业数据情况，对数据进行基础数据处理及数据展现分析，帮助企业利用先进的大数据平台迅速提高企业决策分析能力；协助企业进行业务数据梳理工作，根据企业决策的需求急迫性，逐步引入经营数据建模分析及运营方案，其中包括：客户行为分析、促销策略、客户价值分析、用户流失分析、商品定价策略、销售数据分析、品牌舆情分析、盈利预测、生产作业过程中的量化分析等。

由于“百家企业扶持计划“具有高度严谨、全面的分析特性，对参与企业的要求是：

1、数据化程度不高，但企业有传统方式的基础数据保存（如ERP或CRM系统等），具备一定的数据基础；

2、企业对数据及数据分析工作高度重视，有意愿迅速将企业数据进行科学存贮及应用；

3、企业领导牵头，可派专人或专门对接的团队，确保数据化服务工作有效开展；

4、北京或近京地区优先，便于更好的了解企业数据需求，更好的提供对接服务。

企业报名后，协会将对报名企业进行筛选，符合条件的将进行合作洽谈和签约。

+ 关于峰会

“2015触摸大数据本质之第三届中国数据分析行业峰会”将于2015年7月8日—10日在江苏省昆山皇冠国际会展酒店举行。中国数据分析行业峰会被业界誉为“知识和思想盛宴”，自2010年起至今已成功举办两届，得到了社会各界的聚焦和广泛关注。本届峰会将与中国商业联合会主办的第十届中国零售商大会共同举办，涉及零售、金融、信息技术、消费、教育、医疗等诸多行业，邀请数十多位行业资深专家、学者、知名行业领袖共同探讨大数据产业发展及大数据人才储备、技术应用等的前沿话题，分享大数据智慧及大数据应用“落地”案例，探索各行业之间面临的数据分析应用问题的解决途径。此次峰会将大力促进大数据在各行业领域的快速广泛应用，提升企业生产及管理精细化和智能化的水平，从而推动中国大数据产业向纵深发展！

+ 合作伙伴

战略伙伴：中顾润(北京)项目数据分析师事务所 Zhong Haorun (Beijing) Certified Projects Data Analyst Firm　北京犀数科技有限公司

合作伙伴：Baidu 百度　活动家　上海天元项目数据分析师事务所 Shanghai Tianyuan Certified Projects Data Analyst Firm　cirrus DI 雲圖科學計算　银联智惠　北京大学·商务智能研究中心 WWW.BIRC.ORG.CN　湖南翰林项目数据分析师事务所　電子工業出版社 http://www.phei.com.cn　中国国际电子商务网 www.ec.com.cn　贵州经纬求实项目数据分析师事务所　猎聘网 Liepin.com　大数据文摘

现场支持：31会议网 www.31huiyi.com

新媒体支持：云科技时代　互联网参会指南

合作媒体：CCTV　中国新闻社 CHINA NEWS SERVICE　新华社 XINHUA　千龙网 www.QianLong.com www.BeiJingNews.com.cn　国际商报

第三届中国数据分析行业峰会日程安排

（7月10日 9:00 - 17:00）

会议地址：昆山皇冠国际会展酒店8层会展1厅

9：00-9：05	介绍论坛议程及嘉宾 主持人：蒋文伟 中国商业联合会数据分析专业委员会副秘书长
9：05-9：15	触摸大数据本质 演讲嘉宾：邹东生 中国商业联合会数据分析专业委员会会长
9：15-9：35	大数据 用起来 演讲嘉宾：何 林 北京犀数科技有限公司首席数据科学家
9：35-10：05	自动计算——披着商业模式外衣的数据科学 演讲嘉宾：周庭锐 台湾云图科学计算股份有限公司创始人兼首席数据科学家
10：05-10：20	茶歇
10：20-11：00	“高峰对话”：数据分析对商业模式创新的革命性影响
11：00-11：20	“大数据+”，跨界改变竞争力 演讲嘉宾：江 青 中国统计信息服务中心大数据研究实验室主任
11：20-11：35	数据分析才人在企业构架中的重要性 演讲嘉宾：徐晓颖 项目数据分析师上海授权中心负责人
11：35-11：50	大数据助推零售业发展 演讲嘉宾：王庆生 中颢润（北京）项目数据分析师事务所副总经理
11：50-12：00	“2015-2020年百家企业经营模式创新与转型战略合作”项目启动仪式
12：00-13：30	午餐、休息
13：30-13：50	从数据到智能—百度大数据行业应用探索与实践 演讲嘉宾：沈志勇 百度研究院大数据实验室数据科学家
13：50-14：10	社交网络中我的标签是否会影响我的社交圈 演讲嘉宾：黄丹阳 北京大学商务智能中心研究员
14：10-14：30	商务领域方面的大数据建设情况和思路 演讲嘉宾：李正波 中国国际电子商务中心内贸信息中心副总经理
14：30-14：50	中国高端乘用车市场专题研究之高端车潜在用户搜索行为研究 演讲嘉宾：朱雪宁 北京大学商务智能中心研究员
14：50-15：10	数据价值三定律在商业智能、互联网金融与安全云上运用的案例分享 演讲嘉宾：龙凯 银联智惠联合创始人兼CTO
15：10-15：30	理智看待大数据 用心寻找盈利点 演讲嘉宾：卿启伟 湖南翰林项目数据分析师事务所所长
15：30-15：45	茶歇
15：45-17：00	中国数据分析行业内部会议

+ 与会嘉宾

邹东生

嘉宾介绍：邹东生，中国商业联合会数据分析专业委员会会长、北京大学光华管理学院MBA校友导师、北京大学光华管理学院MBA、北京市青联第十届委员。同时，也是中国数据分析行业发起人、奠基人，具有丰富的企业经营管理咨询经验，是资深数据分析专家，曾主持编写《投资数据分析》、《经营数据分析》等书。

演讲主题：触摸大数据本质

演讲内容：IDC统计过去2年产生的数据超过人类历史数据总和，5年后，每天甚至每个小时产生的数据都是以前人类历史的总和，这是一个数据大爆炸的时代。大数据作为一个新的能源站上了历史舞台。阿里巴巴集团董事局主席马云说：未来最大的能源是大数据，大数据是未来的核心。协会自2003年见证并参与了大数据的发展历程，通过大数据行业12年的观察和研究，针对大数据的常见误区，剖析大数据发展的三大挑战，指出大数据的重大战略意义，触摸大数据的本质。

王庆生

嘉宾介绍：王庆生，资深数据分析师，现就职于中颢润（北京）项目数据分析师事务所，任副总经理。从事数据分析工作7年，先后就职于大赢家集团北京分公司，奥维云网，中颢润项目数据分析师事务所，主攻金融、线上线下零售以及生产制造行业数据分析方向。擅长挖掘算法，主要应用于金融大数据的挖掘，以及零售业基于O2O模式下的线上线下融合的精细化管控压缩成本以及精准营销提升客户体验并实现销售的提升。

演讲主题：大数据助推零售业发展

演讲内容：目前，零售业面临电商的冲击导致规模萎缩，增长率下滑。同时不仅电商冲击，跨业竞争也越来越严重，一些互联网企业纷纷加入传统行业的角逐。面对当前形势，零售业纷纷寻求改变，打造O2O模式，但成功者寥寥无几。面对竞争与改善的失败，零售业陷入困境。

在大数据时代下，通过系统性的，体系化的数据分析方法，不仅能在策略制定的完备方面给出指导意义，同时，在运营期间也从产业链与供应链的综合分析，给出科学的决策。如O2O模式建立的体系框架，基于数据运营的数字化经营与营销的体系搭建，面对市场的纷繁变化，如何科学的应对，精确的应对，做到成本控制的同时，达到效益的最大化。

在理论背景下，利用协会开发，中颢润提供支持的大数据平台datahoop，分析一家奢侈品零售企业的现状，并在线上线下均为该企业创造价值的案例进行分享，让各企业更直观的感受大数据时代数据分析的威力。

何　林

嘉宾介绍：何林，中国商业联合会数据分析专业委员会数据中心主任、北京犀数科技有限公司首席数据科学家、资深数据分析师、软件架构师和高级咨询师。北京大学数学专业与管理专业双硕士，历任高级程序员、技术总监、高级数据分析师和咨询总监，主导开发了多个大型信息系统，主持研发了DataHoop大数据智能分析平台，担任数据分析项目总监，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

演讲主题：大数据 用起来

演讲内容：石油被发掘出来很长一段时间内，仅被用作照明的油料，中国的四大发明之一火药在很长时间里仅仅只是用来制作烟花爆竹。随着数据指数级爆炸性增长，大数据成为了一种新的重要能源。协会数据中心利用DataHoop的大数据尖刀，揭开大数据的神秘面纱，剖析大数据的价值，体验大数据的读心术，见证大数据的预见力，体会大数据作为一种新的战略能源给企业带来的强大变革力量。

龙　凯

嘉宾介绍：龙凯，银联智惠联合创始人兼CTO

北京大学计算机专业学士、旧金山大学计算机科学与技术专业硕士、斯坦福大学管理科学专业硕士。

在硅谷学习和工作近10年，曾任硅谷中国工程师协会副主席。

2012年底在上海参与创建银联智惠信息服务（上海）有限公司，为中国银联子公司。

互联网思维的海归创业团队，基于银联体系海量交易数据做增值服务，包括针对线下商户的精准营销和BI业务，另外还给以p2p为主的金融机构提供信贷类数据增值服务。

演讲主题：数据价值三定律在商业智能、互联网金融与安全云上运用的案例分享

李正波

嘉宾介绍：李正波，管理学博士，高级经济师。现任中国国际电子商务中心内贸信息中心副总经理。长期从事三农问题、宏观经济、商务大数据以及商品流通等领域的研究工作。先后参与自然基金、社科基金、福特基金等国家级课题研究和商贸流通评价、奢侈品市场政策、内贸流通“十二五”规划、“十二五”扩大消费战略等省部级研究项目，多次参加商务领域重要文件起草工作。参编著作三部，译著一部。曾在《管理世界》、《中国农村经济》、《中国发展观察》等核心期刊发表论文10余篇。

演讲主题：商务领域方面的大数据建设情况和思路

演讲内容：随着云计算、大数据、物联网等技术的快速发展，数据分析与应用在经济、社会、政治、等领域的作用日益重要。商务大数据主要是汇集商务领域的数据资源，利用大数据技术，深度挖掘数据之间的关联关系，进行全方位、智能化的分析计算和可视化的结果展示，为商务领域的政府、企业等单位开展决策分析等工作提供更加全面的数据支撑和技术保障。《商务大数据框架与实施路径》重点介绍商务大数据的构建思路、总体架构、试用方向等。

沈志勇

嘉宾介绍：沈志勇，博士，百度研究院大数据实验室数据科学家，负责大数据行业应用探索方向，联合国百度大数据联合实验室决策委员会成员。本科毕业于北大数学学院概率统计专业，随后于中科院软件所获得博士学位，研究方向为数据挖掘。曾任惠普中国研究院研究员，研究领域包括机器学习与数据挖掘。

演讲主题：从数据到智能——百度大数据行业应用探索与实践

演讲内容：大数据这个概念在近几年获得了IT技术界乃至全社会的广泛关注，各行各业都希望大数据相关的理念和技术在自己的生产和业务中发挥重要的作用。以百度或者BAT为代表的互联网企业在大数据领域积累了大量的设施、技术与人才，目前正在积极寻求相关技术往互联网以外行业的转移。百度大数据实验室作为百度在大数据方向的重要布局，目前已经在医疗、教育、金融、体育、零售、公共安全等方向的大数据应用进行了积极的探索与实践。在这次分享中会简单介绍这方面的工作，尤其是其中零售相关的项目。

朱雪宁

嘉宾介绍：朱雪宁，博士，北京大学商务智能中心研究员。参与研究高端车用户搜索行为研究、“实物期权在对外投资中的应用”、KDDCUP数据挖掘竞赛。

演讲主题：中国高端乘用车市场专题研究之高端车潜在用户搜索行为研究

演讲内容：利用奇虎360大数据平台，我们对100万在线用户的13亿搜索序列文本做了分析，并对高端车用户以及商学院人群做了对比分析。我们希望通过提取有效指标、数据分析以及统计模型的方式，试图理解高端车潜在用户搜索平台上表现出的“忠诚”以及“叛逆”行为，从而对在搜索引擎中的广告投放的策略提出可行建议。

江　青

嘉宾介绍：江青，中南财经政法大学MBA合作导师、高级工商管理硕士，高级公关员、大数据研究应用实践者，中国统计信息服务中心大数据研究实验室主任。历任教师、广播电台电视主持人、制片人、专栏作家、报社记者/主任、协会副秘书长等职，对政府、企业等标准化运营、数字决策、公共关系、媒体管理等具备丰富实战经验，曾任政府机构、知名企业智库顾问等，多年来致力于研究探索“组织标准化运营”、“品牌营销”、“领导者数字决策”、“公共关系及媒体传播”、“大数据研究应用”等，承担多个部委、地方政府、企业等项目，现致力于推动CSISC厦门大数据研究服务基地产业化、丝绸之路大数据创新监管、国统大数据智库等项目。

演讲主题：大数据+，跨界改变竞争力

演讲内容：互联网时代，传统观念遇到强有力的挑战，大数据思维越来越冲击当下的社会，传统行业利用大数据实现跨界将成为新的风景？企业在大数据时代也应该有所作为，近年来的实践表明，大数据是企业营销的致胜竞争力。企业营销离不开数据参考已经成为一种常态，新时代下的数据分析人才应该具备哪些素质？7月10日，中南财经政法大学MBA合作导师、中国统计信息服务中心大数据研究实验室主任江青将现场为您解读。

周庭锐

嘉宾介绍：周庭锐，台湾云图科学计算股份有限公司创始人兼首席数据科学家，台湾科技大学管理学院教授，亚洲零售与流通学会常务理事，瑞典斯德哥尔摩大学学术顾问，北京盛德大业国际管理咨询股份有限公司董事首席顾问。英国华威大学（The University of Warwick）博士。曾任教中国人民大学、南澳大学（University of South Australia）、西南交通大学、台湾高雄第一科技大学。曾任台湾商业发展研究院特聘研究员兼行销与消费行为研究所所长。具20年以上企业管理咨询顾问经验，客户包括湖南卫视、上海天娱传媒、四川新都化工、上海欣旺壁纸、深圳星河集团、台湾统一集团、7-11、远流集团、台湾糖业等数十家公司。除消费行为、品牌策略、零售管理等研究领域外，擅长利用基于Hadoop与Spark的分布式高速计算技术，进行线上线下消费行为大数据分析。主持三项中国国家自然科学基金项目。于知名学术期刊发表中英文学术论文超过150篇，包括Journal of International Marketing, Journal of Business Research, Journal of Brand Management等。

演讲主题：自动计算——披着商业模式外衣的数据科学

演讲内容：“自动计算”是大数据应用里的一个不常被讨论，但是却隐匿在任何商业模式之下几乎无所不在的课题。事实上，小至无人飞行器的悬空拍照，大至 PM2.5的明日预报推算，处处隐含着“自动计算”的影子。周庭锐教授分别从“应用对象”、“计算方法”、与“数据结构”三个视角出发，由简至繁讨论三种不同层次的自动计算方法。在最抽象最复杂的充分柔性自动计算模型里，周教授提出一个平台式框架设计，能够通过人工智能的形式，以单一工作平台同时处理多种数据类型，最终以机器脚本自动生成统计分析报告。这个方法涉及：数据特征的自动识别、自动化数据清洁、统计方法的识别与选择、统计检定与最优模型、文本物件与统计物件的匹配、和自动化分析报告的生成。这个方法目前已经实际使用于零售终端POS交易明细的模糊匹配、零售渠道消费者潜在画像的智能识别、移动用户手机APP使用行为分析、和临床医药数据分析等领域。

黄丹阳

嘉宾介绍：黄丹阳，博士，北京大学商务智能中心研究员

演讲主题：社交网络中我的标签是否会影响我的社交圈

演讲内容：随着互联网技术的日益成熟，网络文本数海量增加，如何针对这些海量数据进行分析，并进行统计推断也成为了当前的研究热点。我们考虑一个社交网络，不仅能够观察到用户的网络结构并且能够观测到一系列的标签（例如新浪微博）。这些标签是用户自己定义的并反映了用户的爱好、职业、生活方式等特征。这些特征往往是线上个性化推荐所关注的。本研究建立了用户个性化标签和社交网络结构之间的模型，并能够应用于用户个性化标签推荐。这个模型基于传统的社交网络IP模型。为了能够使得模型在大规模网络数据上快速计算，本研究给出了条件极大似然估计的方法，从而减少模型错误识别的风险，同时大大降低了运算花费。本研究将根据新浪微博数据实例展现大规模社交网络的模型建立与推断方法 。

卿启伟

嘉宾介绍：卿启伟，湖南翰林项目数据分析师事务所所长、湖南翰林企业征信有限公司董事长、邵阳华信房地产评估有限公司董事长、湖南新融达土地评估公司邵阳地区负责人、邵阳科信税务师事务所副所长、湖南南方会计师事务所项目经理、邵阳南方资产评估有限公司项目经理、邵阳南方司法鉴定所司法鉴定人，拥有项目数据分析师、中国注册土地估价师 、中国注册资产评估师 、中国注册房地产估价师、中国注册税务师 、会计师 、司法鉴定人等执业资格。公开发表及获奖的文章有：《试论评估程序在资产评估中的影响》发表于《湖南省注册会计师》杂志；《细节决定成效---试论房地产现场查勘望、闻、问、确、记五步曲》获得湖南省房地产中介专业委员会论文一等奖，并发表于《中国房地产估价师与经纪人》杂志等

演讲主题：理智看待大数据 用心寻找盈利点

演讲内容：在大数据与我们扑面而来时，我们需要一种理智的心态去看待大数据。大数据是人类文明发展到一定阶段的产物，但它仅仅只能作为一种生产要素存在，还远远成不了随心所欲按需分配的社会公共产品，数据本身是不会产生价值的，那种认为跨入大数据行业就会大捞一把的想法是很天真的。大数据时代不是比谁拥有的数据最多，而是比谁驾驭和运用这些数据并将其对接市场需求的能力和手段最多。大数据给了每个涉足这个行业的人公平的机会和充分展示自己聪明才智的平台。临渊羡鱼，不如退而结网。

数据分析行业完整的盈利点应该是“数据资源+分析手段+市场需求=可盈利数据产品”，数据资源是数据分析的物质基础条件，分析手段是我们这些数据分析师的必备谋生之道，数据分析的市场需求是我们数据分析业务的着陆点。我们要寻找适合自己的盈利模式，至于哪种盈利模式适合自己，需要经营者用心寻找，用心体会，决不可脱离实际盲目跟从，好高骛远，也不可守株待兔，墨守旧局，固步自封，丧失发展良机。

朱叶青

高峰对话特约嘉宾

嘉宾介绍：朱叶青，1992年毕业于北京大学生命科学学院获得生物化学专业理学学士；

2001年毕业于北京大学光华管理学院获得MBA学位；

2010年毕业于麻省理工大学斯隆商学院高级经理人培训班。

2000年进入GE公司

2006年3月被任命为GE消费者金融集团亚太区 IT服务和投资管理总监

2010 – 2013年起担任GE金融 亚太区 董事总经理。

2014年1月创立天得一清投资管理有限公司，并担任公司总裁。

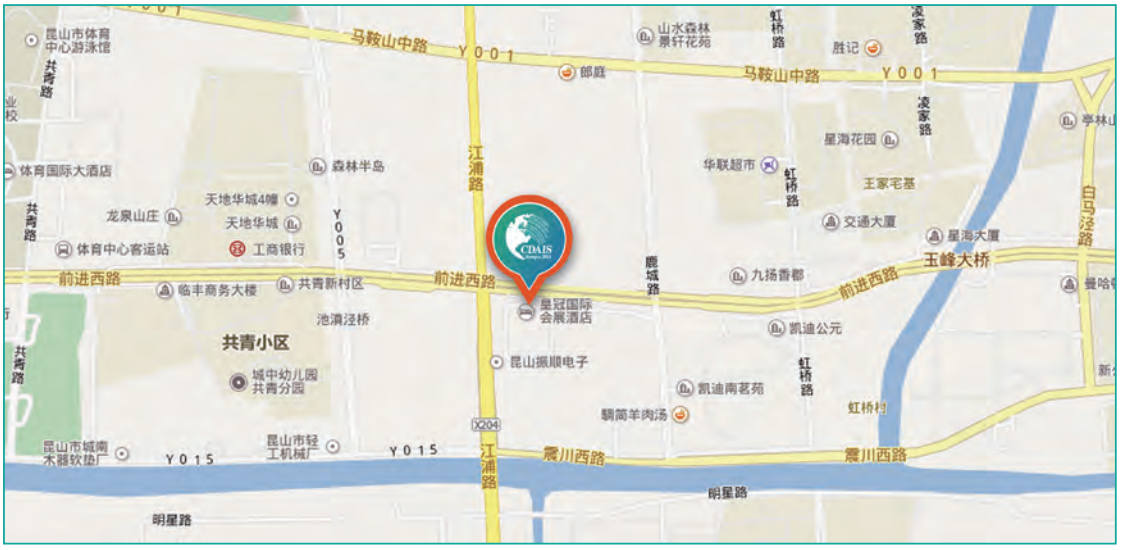
朱叶青先生拥有深厚的IT、互联网、消费者金融服务和投资专业背景，对全球很多著名金融公司和IT公司有非常深入的研究。

他现任科技部金融科技创新联盟和中国信息技术服务与外包产业联盟专家委员会委员，同时担任银监会和国家开发银行特聘专家。

2008年由工业和信息产业部和中国外包服务网评选为中国软件出口（外包）特别推动人物大奖。

2011年荣获中国服务行业领军人物奖和中国金融服务卓越行业精英奖。

+ 路线指引



会场地址：昆山皇冠国际会展酒店8层会展1厅　昆山前进西路1277号 (前进西路与江浦路交叉口)

乘车贴士：1. 无锡机场昆山城市候机楼 —— 昆山皇冠国际会展酒店

步行360m到电视台站上车，乘坐昆山25路到华润中心站下车，步行530米

2. 昆山南站 —— 昆山皇冠国际会展酒店

昆山南站上车，乘坐28路到华润中心站下车，步行530米

+ 联系方式

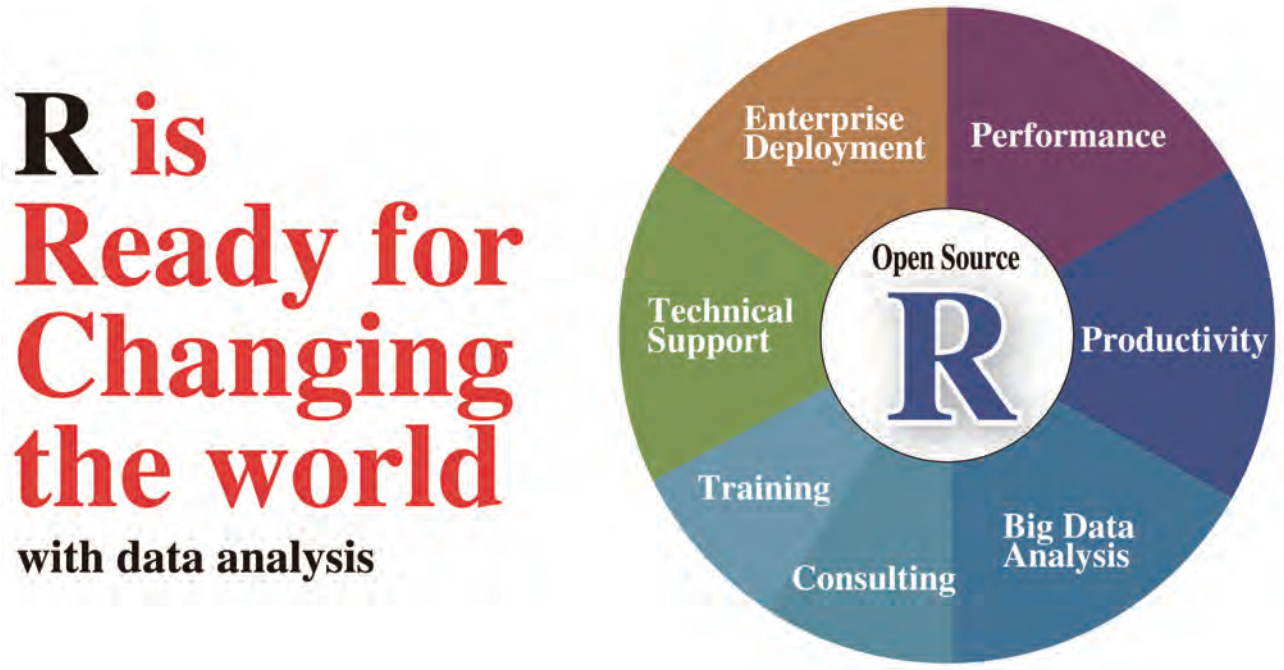
CDAIS Jiangsu 2015

微信：wxchinacpda

现场嘉宾如想了解及沟通大数据更多信息，

可加群和关注“CPDA数据分析天地”微信公众账号

峰会现场联系人：张楠 18612766693　　魏云龙 13810356524



R语言最新技术导读

文 / 中商联数据分析委数据中心 曹莹　　编辑 / 张楠　　图 / 崔峻珩

“R语言是在大数据时代被工业界了解和认识的语言，R语言被时代赋予了挖掘数据价值，发现数据规律，创造数据财富的任务。”

R语言因其免费、开源的软件，庞大的社区和资源，面向数据的思维模式，自由的语法和接近工业级分析数据的能力，无疑成为2015年最热门的语言。R应用的热门领域包括统计分析、金融分析、数据挖掘、互联网、生物信息学、生物制药、全球地理科学和数据可视化等。

一、Quantmod：金融分析

首先，Quantmod作为金融分析包，对于股票信息的获取必须是操作简单、灵活快捷的。R语言中Quantmod包提供了从网络、数据库中获取上证、上证A股、上证B股、深证成指等股票数据的函数，同时，金融分析包也为使用者提供了可以获取和查看上市公司财务报表，期权交易数据、美联储经济数据等方面的代码服务。

另外，使用Quantmod包中几乎涵盖了所有热门技术分析图。比如：平均趋向指标ADX、平均真实波幅指标ATR、布林线指标BBands、顺势指标CCI、Chaikin资金流量指标CMF、Chande动量摆动指标CMO、指数平均数指标EMA、包络线指标Envelope、弹性成交量加权移动平均线指标EVWMA、移动平均收敛发散指标MACD、动量指标Momentum、合约终止线Expiry、抛物线指标SAR、简单移动平均指标SMA、随机动量指标SMI、双指数移动平均指标DEMA、区间震荡线DPO、变动率指标ROC、相对强弱指标RSI、交易量指标Vo、威廉指标WPR、零滞后指数移动平均ZLEMA等。

最后，金融分析包不仅提供了基本的条形图、蜡烛图和线图的作图工具，而且还可以对图形进行背景、颜色、缩放和存储等个性化处理。

二、highfrequency包：高频分析

在金融市场中， 高频数据是指逐笔交易数据(transaction by transaction data) 或逐秒记录数据(tick by tick data)。纽约股票交易所的交易行情数据库包含了综合磁带系统报告的所有证券的交易和报价记录(Trades and Quotes– NYSE TAQ)。

R中针对高频数据的highfrequency包依赖于xts和zoo两个

包，主要实现的功能有组织高频数据、高频数据的清洗及整理、高频数据的汇总、高频数据的相关模型（波动性模型及预测， HAR模型、HEAVY模型）等。

三、Xgboost包：速度快效果好的boosting模型

在数据分析的过程中，我们经常需要对数据建模并做预测。在众多的选择中，randomForest, gbm和glmnet是三个尤其流行的R包，它们在Kaggle的各大数据挖掘竞赛中的出现频率独占鳌头，被坊间人称为R数据挖掘包中的三驾马车。

Boosting分类器属于集成学习模型，它基本思想是把成百上千个分类准确率较低的树模型组合起来，成为一个准确率很高的模型。这个模型会不断地迭代，每次迭代就生成一颗新的树。对于如何在每一步生成合理的树，R包使用由Friedman提出的Gradient Boosting Machine。它在生成每一棵树的时候采用梯度下降的思想，以之前生成的所有树为基础，向着最小化给定目标函数的方向多走一步。在合理的参数设置下，需要生成一定数量的树才能达到令人满意的准确率。在数据集较大较复杂的时候便需要几千次迭代运算。

现在，xgboost工具更好地解决这个问题。xgboost的全称是eXtreme Gradient Boosting。Xgboost具有速度快，效果好，功能多这三个突出的优点。

四、Rcurl包：网络爬虫

Rcurl网络爬虫包基于HTTP协议，提供3个切入口（getURL、getForm、postForm）和170多个参数

在万维网飞速发展的网络背景下，搜索引擎在人们的生活工作中无疑扮演着重要的角色，而网络爬虫则是搜索引擎技术的最基础部分。在搜索引擎成为主流检索工具的今天，互联网上的网络爬虫各式各样，但爬虫爬取网页的基本步骤大致相同：

1）人工给定一个URL作为入口，从这里开始爬取。

万维网的可视图呈蝴蝶型，网络爬虫一般从蝴蝶型左边结构出发。这里有一些门户网站的主页，而门户网站中包含大量有价值的链接。

2）用运行队列和完成队列来保存不同状态的链接。

对于大型数据量而言，内存中的队列是不够的，通常采用数据库模拟队列。用这种方法既可以进行海量的数据抓取，还可以拥有断点续抓功能。

3）线程从运行队列读取队首URL，如果存在，则继续执行，反之则停止爬取。

4）每处理完一个URL，将其放入完成队列，防止重复访问。

5）每次抓取网页之后分析其中的URL（URL是字符串形式，功能类似指针），将经过过滤的合法链接写入运行队列，等待提取。

6）重复步骤3）、4）、5）。

五、recharts包：图形高大上

recharts包是R对echarts的接口。ECharts基于Canvas，纯Javascript图表库，可以提供直观，生动，可交互，可个性化定制的数据可视化图表。与R结合后可以非常方便地将数据和模型的结果进行动态展示，是R中图形可视化的一大利器。

六、风险控制矩阵Risk Matrix

是一种有效的风险管理工具。可应用于分析项目的潜在风险，也可以分析采取某种方法的潜在风险。

步骤：

1、列出该项目的所有潜在问题

2、依次估计这些潜在问题发生的可能性，可按低，中，高，也可按数字0 – 10

3、依次再估计这些潜在问题发生后对整个项目的影响，也可按低，中，高或0 – 10方法

4、可得出风险矩阵图便于分析

5、找出预防性措施

6、建立应急计划

风险矩阵图给出四种分类：

1、如潜在问题在红色区域，则应该不惜成本阻止其发生，（如果成本大于可接受范围，则放弃该项目）。

2、如潜在问题在橘红色区域，应安排合理的费用来阻止其发生。

3、如潜在问题在黄色区域，应采取一些合理的步骤来阻止发生或尽可能降低其发生后造成的影响。

4、准备应急计划，该部分的问题是反应型，即发生后再采取措施，而前三类则是预防型。

用R语言进行时间序列的分位数回归

文 / 中商联数据分析委数据中心 常琳　　编辑 / 张楠　　图 / 崔峻珩

时间序列分析作为数理统计的一个重要分支，因其良好的预测能力，已被广泛的应用于自然科学和社会科学的各个领域之中，而分位数回归具有良好的非参数性质，将分位数回归应用到时间序列中可以拓宽时间序列模型的应用范围，使其能更好的为社会生产服务。本文结合我国社会消费品零售总额数据，利用时间序列分位数模型进行建模分析，预测。通过与实际数据对比，体现了模型良好的预测能力。

一、为什么需要分位数回归?

传统的线性回归模型描述了因变量的条件均值分布受自变量X的影响过程。其中，最小二乘法是估计回归系数的最基本方法。如果模型的随机误差项来自均值为零、方差相同的分布，那么回归系数的最小二乘估计为最佳线性无偏估计；如果随机误差项是正态分布，那么回归系数的最小二乘估计与极大似然估计一致，均为最小方差无偏估计。此时它具有无偏性、有效性等优良性质。

但是在实际的经济生活中，这种假设通常不能够满足。例如当数据中存在严重的异方差，或后尾、尖峰情况时，最小二乘法的估计将不再具有上述优良性质。为了弥补普通最小二乘法（OLS）在回归分析中的缺陷，Laplace提出了中位数回归（最小绝对偏差估计）。在此基础上，Koenker和Bassett把中位数回归推广到了一般的分位数回归上。分位数回归相对于最小二乘回归，应用条件更加宽松，挖掘的信息更加丰富。分位数回归由于不仅能够度量回归变量对分布中心的影响，而且也能度量回归变量对分布上尾和下尾的影响，因此在某些情况下，比经典的最小二乘法回归法更具有优势。随着分位数回归理论及算法的不

断发展，使得分位数回归越来越广泛的被应用到各个领域中。

二、什么是分位数回归?

1.分位数回归的基本思想

定义a：设随机变量X的分布函数为$F(X)$，那么，对任意$0 < p < 1$的p，称使得$F(X) = p$的X为此分布的p分位数，或者称为下侧p分位数。

定义b：给定样本，$x_1, x_2, \cdots x_n$,那么m_p可定义如下：

$$m_p = \begin{cases} \frac{1}{2}\left(x_{(np)} + x_{(np+1)}\right), \text{若 np 是整数；} \\ x_{([np+1])}, \text{若 np 不是整数；} \end{cases}$$

从决策理论的角度考虑，不同的损失函数所对应的内涵是不同的，在分位数回归模型中，定义损失函数为

$$\rho_\tau(u) = u(\tau - I(u))$$

其中$0 < \tau < 1$，$I(u)$为示性函数

$$I(u) = \begin{cases} 1, u < 0 \\ 0, u \geq 0 \end{cases}$$

$$\rho_\tau(u) = \begin{cases} (\tau - 1)u, u < 0 \\ \tau u, \quad u \geq 0 \end{cases}$$

可以看出损失函数为分段函数（见图1），且$\rho_\tau(u) \geq 0$。

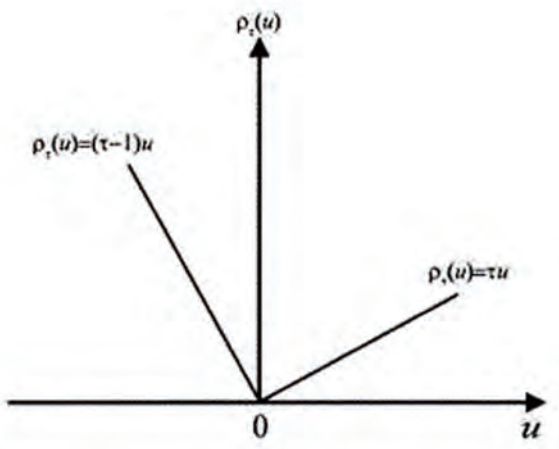


图1损失函数示意图

对于最小二乘法，求解$\min_{\mu \in R} \sum_{i=1}^{n} (y_i - \mu)^2$可得到$\hat{\mu}$为样本均值。当给定x时，若把y的条件均值记作$\mu(x) = x^T\beta$，即y是x的线性函数，则$\beta$可由$\min_{\beta \in R} \sum_{i=1}^{n} (y_i - x^T\beta)^2$估计得到，这就是经典的最小二乘回归问题。

类似的，求解$\min_{\alpha \in R} \sum_{i=1}^{n} \rho_\tau(y_i - \alpha)$可得$\hat{\alpha}(\tau)$，即为样本$\tau$分位数。当给定$x$时，$y$的条件分布记作$F_y(y|x)$，则其逆函数表示为$Q_y(\tau|x) = \inf\{y: F_y(y|x) > \tau\}$。定义$Q_y(\tau|x) = x^T\beta(\tau)$为样本条件$\tau$分位函数，其中$x$为p维向量，则$\hat{\beta}(\tau)$可由

$\min_{\beta \in R^p} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^T \beta)$估计得到。

$$\min_{\beta \in R^p}[\sum_{y_i \geq x_i^T \beta} \tau |y_i - x_i^T \beta| + \sum_{y_i < x_i^T \beta} (1-\tau)|y_i - x_i^T \beta|]$$

其中$\tau \in (0,1)$，β为系数向量，它随着τ的变化而不同。分位数回归的本质是通过τ取0，1之间的任何值，它也能在一定程度上代表所有数据的信息，但更侧重于特定区域的数据。

2.分位数回归的优点

(1)分位数回归对模型中的随机误差项不做任何要求，使得模型具有较强的适应性。

(2)分位数回归由于对所有的分位数进行回归，因此对于数据中出现的异常点具有鲁棒性。

(3)分位数回归对于因变量具有单调不变性。

(4)分位数回归能更详尽的提取出因变量和自变量之间的关系。

三、时间序列建模

1.序列模型识别及定阶

为时间序列选择合适的模型并确定模型的阶数叫做模型识别，即判断该序列所适合的模型类型和模型的阶数。三类平稳时间序列的自相关系数和偏自相关系数具有不同的统计特性。如果一个时间序列是由某一类模型所生成的，理论上它就应该具有相应的自相关特性，因而我们可以计算出平稳时间序列的样本自相关系数和样本偏自相关系数，将其特性与不同类型序列的理论自相关系数和偏自相关系数的特性进行比较，进而初步判定序列X_t所适合的模型类型。三类平稳时间序列的自相关系数和偏自相关系数的统计特性如表1，可以依据表中性质来初步确定时间序列模型的类型。

模型（序列）	AR(n)	MA(m)	ARMA(n,m)
自相关系数	拖尾	m阶截尾	拖尾
偏自相关系数	n阶截尾	拖尾	拖尾

表1 平稳时间序列的统计特性表

(1)若时间序列的自相关系数ρ_k在m步截尾（即$k > m$时，$\rho_k = 0$），并且偏自相关系数φ_k收敛到零，则可判断时间序列为MA(m)序列。实际计算的样本自相关系数ρ_k不会在m步后全为零，而是在零的上下波动。

(2)若时间序列的偏自相关系数φ_k在n步截尾，且自相关系数ρ_k收敛到零，则可判断时间序列为AR(n)。

(3)若自相关系数ρ_k和偏自相关系数φ_k都不截尾，但都收敛到零，则时间序列很有可能为ARMA(n,m)。

2.参数估计

对序列进行模型识别后，接下来就是估计模型中的参数，以便进一步识别和应用模型。参数估计方法有矩估计法，最小二乘估计法，极大似然估计法和分位数法。使用分位数估计方法的时间序列模型，我们称之为时间序列分位数模型。这里我们详细介绍用分位数回归对三种平稳时间序列的系数求解过程。

(1)AR(p)序列

对于p阶自回归模型AR(p)：$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + \varepsilon_t$，其中$t = 1,2\cdots n$，$\{\varepsilon_t\}$为白噪声序列。
依据分位数回归的思想，构造损失函数为:

$$S(\beta(\tau)) = \sum_{i=1}^{n} \rho_\tau(y_t - (\beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p}))$$

则AR(p)序列的模型参数估计值为:

$$\hat{\beta}(\tau) = arg \min_{\beta \in R^{p-1}} S(\beta(\tau))$$

(2) MA(q)序列

对于q阶平均移动模型MA(q)：$y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}$，令$e_t = y_t - \hat{\mu}$，其中$\hat{\mu} = \frac{1}{n}\sum_{t=1}^{n} y_t$，
则有MA(q)模型参数的估计值为：

$$\hat{\theta}(\tau) = arg\min_{\beta \in R^{1q}} \sum_{i=1}^{n} \rho_\tau(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q})$$

(3)ARMA(p,q)序列

对于ARMA(p,q)序列模型：

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}$$

先估计模型参数$\beta\beta$：

$$\hat{\beta}(\tau) = arg\min_{\beta \in R^{1p-1}} \sum_{i=1}^{n} \rho_\tau(y_t - (\beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p}))$$

然后令$e_t = y_t - (\hat{\beta}_0 + \hat{\beta}_1 y_{t-1} + \hat{\beta}_2 y_{t-2} + \cdots + \hat{\beta}_p y_{t-p})$，
再估计模型中参数$\theta\theta$：

$$\hat{\theta}(\tau) = arg\min_{\beta \in R^{1q}} \sum_{i=1}^{n} \rho_\tau(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q})$$

四、实证分析

在各类与消费有关的统计数据中，社会消费品零售总额是表现国内消费需求最直接的数据。社会消费品零售总额是国民经济各行业直接售给城乡居民和社会集团的消费品总额。它是反映各行业通过多种商品流通渠道向居民和社会集团供应的生活消费品总量，是研究国内零售市场变动情况、反映经济景气程度的重要指标。

因此，如若依据已有的数据序列能够比较准确的估计出近几年的社会消费品零售总额是很有必要的。依据预测值，政府可以制定出相应的调控措施，使得国家经济能更加有效平稳的发展。之前介绍了分位数的相关理论，讲述了时间序列模型的相关理论知识。本文结合国家统计局网站上所能查到的1952年–2008年中国的社会消费品零售总额（单位：亿元）建立时间序列模型，利用分位数回归估计模型参数，并对其进行预测。

所有计算通过R软件完成。

1.数据预处理

首先，依据数据画出数据的时间序列如图2。

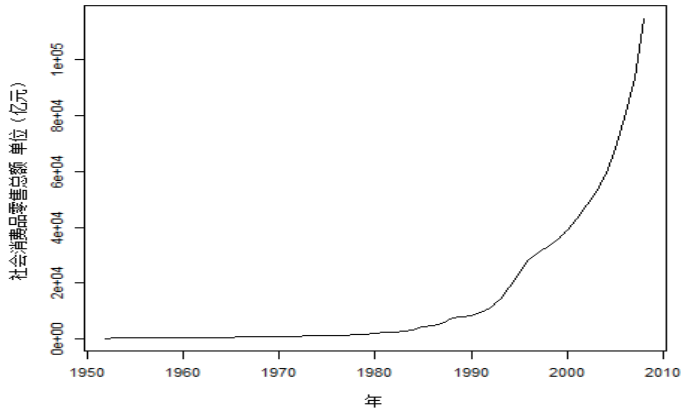


图2 社会消费品零售总额时序图

依据图2中的时间序列图，可以发现序列呈指数增长的趋势，显然是不平稳的。欲对时间序列数据建模，则首先应对数据进行预处理，这里需要对数变换与差分运算的结合使用，如果序列X_t含有指数趋势，则可以通过取对数将指数趋势转化为线性趋势。然后再进行差分以消除线性趋势，最终得到一个转化后的中国社会消费品总额时间序列图如图3。

从图3中可以看出，经过处理后的序列基本是在一个常值附近随机波动，而且波动有界。依据时序图检验法基本认定序列是平稳的。然后对处理后的时间序列图进行单位根法平稳性检验，得到检验的p值为0.02146，从而拒绝原假设，认为此时的序列是平稳的。至此经过数据的预处理，我们得到了一个平稳的时间序列。然后，检验序列的随机性，检验结果显示在1到12各阶延迟下P值均小于0.05，即拒绝序列白噪声的原假设，说明序列不是纯随机的，还有信息可以提取。

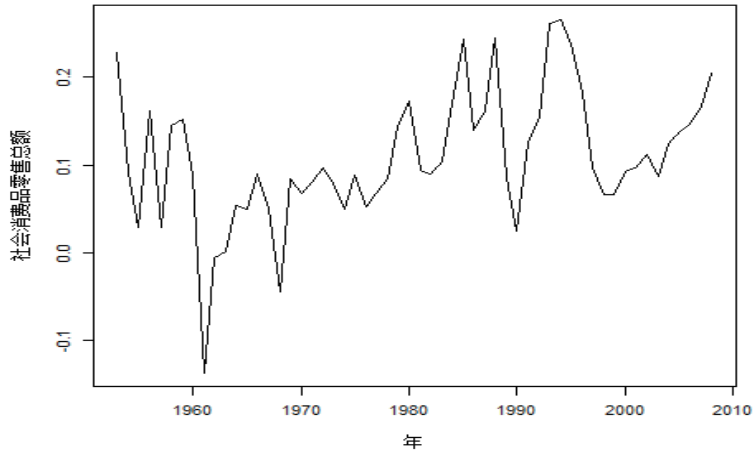


图3处理后社会消费品零售总额时序图

2.模型识别

接着，我们进行建模。根据数据，计算平稳数据的自相关系数和偏自相关系数，如图4。

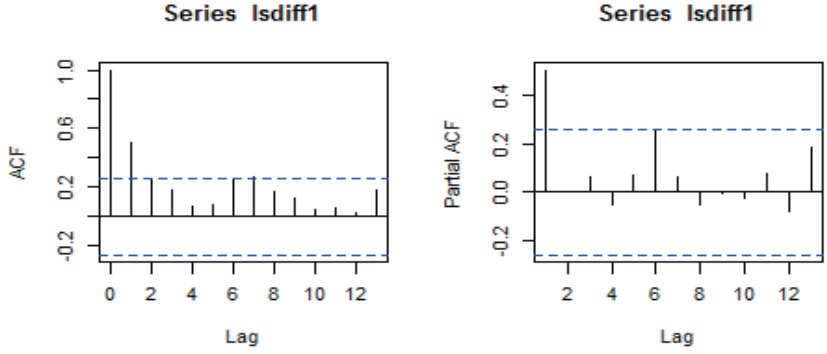


图4 处理后序列的自相关图与偏自相关图

依据平稳时间序列模型具有的自相关系数和偏自相关系数的特点(见表1)来选择合适的模型，根据图4可以看出来序列自相关系数拖尾，偏自相关系数一阶截尾。根据准则，可以认定的模型为AR(1)模型，模型的表达式为：$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$

3.参数估计

经过模型识别后，得到AR(1)模型，下面用分位数回归对模型系数进行估计。

这里选择五个分位点τ =0.05，0.25，0.5，0.75，0.95，利用分位数回归对模型进行参数估计。

分位点	β_0	β_1
τ =0.05	-0.07722	0.661251
τ =0.25	0.037951	0.321688
τ =0.5	0.053807	0.535082
τ =0.75	0.077717	0.612028
τ =0.95	0.123333	0.754962

表2不同分位点下AR(1)对应系数估计值

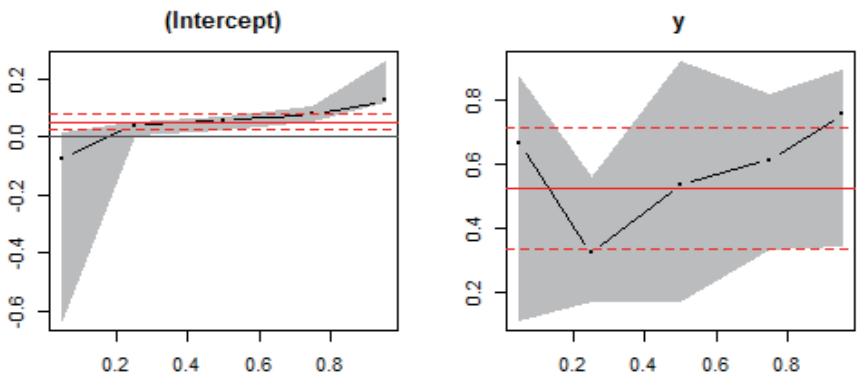


图5不同分位点下的系数估计值的比较

通过计算，运用分位数回归估计出模型系数(见表2)以及不同分位点下的系数估计值的比较见图5。可以看出各个分位点对应着不同的系数。至此，我们得到了时间序列分位数模型。

4.模型检验

确定了拟合模型的口径之后，我们还要对该拟合模型进行必要的检验。

由残差序列检验结果（表3）知，各阶延迟下LB统计量的P值都显著大于0.05，可以认为这个拟合模型的残差序列属于白噪声序列，该模型显著有效。

延迟阶数	LB统计量	P值
6	3.5378	0.7389
12	9.16	0.6892

表3 残差序列白噪声检验结果

5.预测和结果分析

接下来，将用得到的模型进行预测和结果的分析。通过模型识别，参数估计，得到不同分位点下的时间序列模型，这里仅给出分位点τ =0.25,0.5,0.75的模型拟合图，如图6。通过图形可以看出，模型的拟合效果还是比较好的，能比较精确的模拟出原数据的增减趋势，除个别点外，误差也比较小，因此基本可以认定模型能比较准确的预测未来几年的社会消费品零售总额。

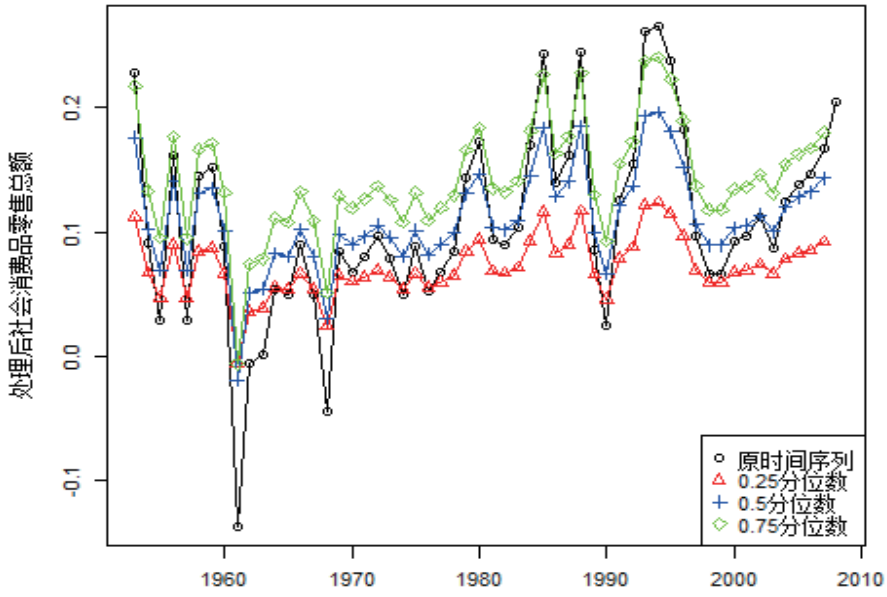


图6 分位点为0.25,0.5,0.75处模型拟合效果图

模型是对1952年–2008年中国社会消费品零售总额建立的，并预测了2008年以后五年的零售总额（见表4）。这里采取分位点0.05,0.25,0.5,0.75,0.95进行建模预测。为了说明模型的预测效果，表中给出了2009年–2013年中国统计局网站上公布的社会消费品零售总额数据。

年份	实际值	最小二乘法	τ=0.05	τ=0.25	τ=0.5	τ=0.75	τ=0.95
2009	132678.4	141934	121706.4	127391.3	135206.9	140677.3	151615.4
2010	156998.4	176094.4	117079.2	136812.1	155714	172162.1	211553.7
2011	183918.6	218904.9	105636.4	145403.3	177219.6	210558.4	307757.9
2012	210307	272401.4	91357.18	154015.6	200420.9	257417	462030.9
2013	237809.9	339152.7	76824.88	162961.9	225892.3	314628.2	710322.6

表4不同分位点对应模型的预测值

通过对模型的拟合效果图和预测结果进行分析，从表4中可以看到，在经济正常发展的这五年中，0.5分位数模型预测值较实际值最为接近，尤其是2010年，0.5分位数模型预测值是155714亿元，而实际数值为156998.4亿元，预测百分误差是0.82%。从预测百分误差来看，模型的预测精度还是相当的高。可以看到模型的拟合和预测效果还是不错的。在大部分年份，我国的经济发展还是比较正常的，而且一旦市场经济趋向异常时，我国政府会积极进行调控确保经济健康平稳发展。由于模型对于预测正常年份的数值非常准确，因此模型的应用前景还是很广泛的。但是难免有些年份，国家会控制经济发展的速度或者降低经济发展速度，以调控国家经济构成，以使国家经济构成更加合理。我国每年都提前会公布下一年的国家预算，根据预算的增减幅度和结构构成，我们可以初步估计出国家对于下一年经济的发展所持的态度，以后一年中，国家所采取的经济政策基调也基本已定。若政府所持基调是控制经济过热发展那么我们可以选择分位点较低的时间序列模型进行预测，反之选择分位点较高的时间序列模型进行预测。政府可以根据模型进行预测，然后根据预测值进一步的优化和改进调控的方向和幅度。因此，使得模型具有较强的应变能力。

分位数回归理论提出至今已有三十多年的历史了，其理论也已越来越成熟，并被广泛的应用于多种学科之中。尽管如此，分位数回归在理论和方法上依然还有很大的提升空间。希望有兴趣对分位数回归做进一步研究的学者能积极的推动分位数理论和应用的进一步发展，以便使其能更好的服务各学科和社会生产实践之中。

互联网的发展驱动大数据的发展

文 / 上海天元项目数据分析师事务所　　编辑 / 张楠

自上个世纪70年代驶入“信息高速公路”，互联网的发展始终在“创新与改变”中跨越前行。每一天开启网络，迎接我们的都是可能或正在发生的各种改变。眼下，当人们还在津津乐道云计算的时候，大数据时代突然火爆起来，成为业界人士舌尖上滚烫的话题。

如果说印刷革命引爆了人类社会知识的生产与传播，开启了知识传播的大众时代，那么互联网新媒体开启的“大数据”时代，则是一场更为深广的革命。

在“大数据”时代，以互联网为代表的媒介技术使得信息的生产呈几何级数式增长，人类主宰信息的能力远远落在后面。我们走在街上，隔三差五看到的街边小店打折促销关门熄火的信息、听到“都是马云惹的祸”“都是淘宝惹的祸”这些话语早已不那么新奇了，甚至已经习惯了，因为在淘宝的影响下，我们已经不经意的参与到了互联网的世界。不仅是使街边小店关门解散的淘宝，还有对银行产生巨大冲击的余额宝、与传统交通产生激烈竞争滴滴打车、快的打车...这些都是互联网形势下新的经济产物，尤其是今年三月两会提出的“互联网 + ”后，这类现象还将更多。

据CNNIC（中国互联网络数据中心）数据显示，截至2014 年 12 月，我国网民规模达 6.49 亿，全年共计新增网民3117 万人，互联网普及率为 47.9%，可以说有接近一半的中国人在使用互联网。而在2005年，互联网普及率仅为8.5%，仅经历了十年时间，互联网的普及率增加了接近40%，这样看来，街边小店陆续关门的现象就怪不得马云了。

不仅PC互联网，手机互联网的发展更是令人惊叹！

据中国互联网络数据中心2014年8月发布的《中国移动互联网调查研究报告》显示，截至2014年6月底，我国网民中使用手机上网的网民规模为 5.27亿，人群占比由 2007年的24.0%提升至 83.4%，不到十年时间足足提升了超过60%！手机网民在整体网民中占有惊人的比例！由此可见，报纸、超市、游戏机、电话...这些都可以不需要了，看新闻、购物、聊天，一部小小的手机就可以搞定！

现在，互联网已经成为人们生活和工作中形影不离的工具，在未来两三年内，互联网还将继续渗透到我们的生活和工作中，并将在更多方面改变和改善我们的生活和工作。

随着“互联网+”的继续迅猛发展，物联网的发展已不再遥远。我国现在大概有5亿台电脑，有十几亿人口，未来手机保有量约20亿部左右，在未来的几年内，我们每个人平均会拥有30到50个智能设备在和互联网连接，也就是可能在几年内，我们所有人拥有的智能设备数将会达到或超过400亿到500亿，而且所有的这些智能设备，它可能在我们睡觉的时候也在工作，不断的采集各种信息，这些信息都将成为我们今天所说的大数据。因此，我们可以认为互联网的发展将成为大数据发展的最大驱动力！

数据科学中的“数据智慧”

文 / 郁彬　翻译 / 吕翔 张心雨 施涛等　编辑 / 张楠　图 / 崔峻珩

在大数据时代，学术界和业界的大量研究都是关于如何以一种可扩展和高效率的方式来对数据进行储存，交换和计算（通过统计方法和算法）。这些研究领域无疑非常重要，然而，只有当我们对数据智慧（Data Wisdom）也给予同等程度的重视时，大数据（或者小型数据）才能被转换为真正的知识和有用的，可被采纳的信息。换而言之，我们要认识到必须拥有足够数量的数据才有可能对复杂度较高的问题给出较可靠的答案。“数据智慧”对于我们从数据中提取有效信息和确保没有误用或夸大原始数据是至关重要的。

“数据智慧”一词是我对应用统计学核心部分的重新定义。这些核心部分在伟大的统计学家（或者说是数据科学家）John W. Tukey和Geogre Box的文章中有详细阐述。

要让统计圈以外的人了解，“数据智慧”是非常必要的重命名，因为它比“应用统计学”这个术语能更好概括其核心成分。这样一个有信息量的名称可以使人们意识到应用统计作为数据科学一部分的重要性。

引用维基百科中对“智慧“这一词条解释的第一句话，我想说：“数据智慧“是将领域知识、数学和方法论与经验、理解、常识、洞察力以及良好的判断力相结合，批判性地理解数据和依据数据做决策的一种能力。

“数据智慧”是数学、自然科学和人文主义这三方面能力的融合，是科学和艺术的结合。在缺乏有实践经验者的指导下，个人很难仅仅靠从读书中获得“数据智慧”，想要学习它的最好方法就是和拥有它的人一起共事。当然，我们也可以通过问答方式来帮助形成和培养“数据智慧”的能力。我这里有十个基本问题，我鼓励人们在开始从事数据分析项目或者在其过程中可以经常问问自己。这些问题刚开始时是按照一定顺序排列的，但是在不断重复的数据分析过程中，这个顺序完全可以被打乱。

这些问题也许无法详尽彻底的解释“数据智慧”，但是它们体现了“数据智慧”的一些特点。

1.要回答的问题

数据科学的问题最开始往往来自于统计学或者数据科学以外的学科。例如，神经科学中的一个问题：大脑是如何工作的？或银行业中的一个问题：该对哪组顾客推广新服务？要解决这些问题，统计学家必须要与该领域的专家进行合作。这些专家会提供有助于解决问题的领域知识，早期研究成果，更广阔的视角，甚至可能是对该问题的重新定义。与这些（往往可能很忙的）专家建立联系需要很强的人际交流技巧。

而这种交流对于数据科学项目的成功是必不可少的。在有充足数据来源的情况下，经常发生情况的是在数据收集前要回答的问题还没有被精确定义。正如 Tukey 所说的那样：“我们在‘探索性数据分析（Exploratory Data Analysis）’的游戏中。”我们寻找需要回答的问题，然后不断重复统计调查过程

（就像上文提到的 George Box 的文章中所述）。由于误差的存在，我们谨慎的避免对于数据中出现的模式进行过度拟合。例如，当同一份数据既被用于问题的建模又被用于问题的验证时，过度拟合就会发生。一条黄金准则就是将数据分割，在分割时考虑到数据潜在的结构（如相关性，聚类性，异质性）使分割后的每部分数据都对原始数据具有代表性。用其中一部分来探索问题，而另一部分用来通过预测或者建模来回答问题。

2.数据收集

什么样的数据与（1）中要回答的问题最相关?

实验设计（统计学的一个分支）和主动学习（机器学习的一个分支）中的方法对解决这个问题有所帮助。即使是在数据已经收集好了以后，考虑这个问题也是很有必要的。因为对理想的数据收集机制的理解可以暴露出实际数据收集过程的缺陷，能够指导下一步分析的方向。

下面的问题会有所帮助：

数据是如何收集的？在哪些地点？在什么时间段？谁收集的？用什么设备收集的？中途操作人员和设备被更换过吗？总之，试着想象自己亲身在数据收集现场。

3.数据含义

数据中的某个数值代表了什么含义？它测量了什么？它是否测量要测量的？哪些环节可能会出差错？在哪些统计假设下可以认为数据收集没有问题？（对数据收集过程的详细了解在这会很有帮助）

4.相关性

收集来的数据能完全或部分地回答要研究的问题吗？如果不能，还需要收集什么其他数据？第2个问题中提到的要点在此处同样受用。

5.问题转化

如何将（1）中的问题转化成一个数据相关的统计问题，使之能够很好回答与原始问题呢？有多种转换方式吗？比如，我们可以把问题转换成一个与统计模型有关的预测问题或者统计推断问题吗？在选择模型前，列出将每一种能解决与实质性问题的转化方式的优点和缺点。

6.可比性

各数据单元是否是可比的，或经过标准化处理而可视为可交换的？苹果和橘子是否被组合在一起了？数据单元是否相互独立？两列数据是不是同一个变量的副本?

7.可视化

观察数据（或其子集），制作一维或二维图表，并检验这些的数据的统计量。询问数据范围是什么？数据正常吗？是否有缺失值？多使用颜色和动态图，注意有意料之外的情况记住，我们大脑皮层的30%都是用来处理图像的，所以可视化在挖掘数据模式和特殊情况时非常有效。通常情况，为了找到大数据的模式，可视化在建立某些模型之后使用最有用，比如，计算残差并进行可视化展示。

8.随机性

统计推断的概念，比如p值和置信区间，都依赖于随机性。那数据中的随机性是什么含义呢？我们要对统计模型的随机性尽量明确地定义。哪些所研究的领域中知识支持所用统计模型中的随机性的描述？一个表现统计模型中随机性的最好例子，就是因果关系分析中Neyman–Rubin 的随机分组原理（在AB检验中也有使用）。

9.稳定性

你会使用哪些现有的方法？不同的方法会得出同一个定性的结论吗？对数据进行随机扰动，例如，可以通过添加噪声或二次抽样实现（一般来说，应确定二层样本有原样本的结构，如相关性，聚类特性和异质性，这样二层样本能较好地代表原始数据）。结论依然成立吗？我们应该只相信那些能通过稳定性检验的方法，稳定性检验简单易行，能够抗过度拟合和过多假阳性发现，具有可重复性（要了解关于稳定性重要程度的更多信息，请参看文章）。

可重复性研究最近在科学界吸引了很多注意，请参照《Nature》特刊。《Science》的主编Marcia McNutt指出“实验再现是科学家用以增加结论信度的一种重要方法”。同样，商业和政府实体也应该要求从数据分析中得出的结论，当用新的同质数据检验时是可再重复的。

10.结果验证

人们怎样能知道数据分析是不是做的好呢？衡量标准是什么？可以考虑用其他类型的数据或者先验知识来衡量有效性，不过可能需要收集新的数据以确认结果的有效程度。

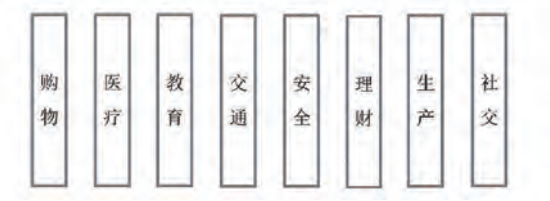
在数据分析时还有许多其他问题要考虑，但我希望上面的这些问题能使你对如何获取“数据智慧”产生一点感觉。作为一个统计学家，这些问题的答案需要在统计学之外获取。要找到可靠的答案，有效的信息源包括“死的”（如科学文学、报告和书籍）和“活的”（如人）。出色的人际交流技能使得寻找正确信息源的过程简单了许多，即使是在寻求“死的”信息源的过程中也是这样。因此，为了获取充足的信息，人际交流技能将更加重要，因为在我的经验中，知识渊博的人通常是你最好的指路。

6张图带你看懂“块数据”
——为什么说得“块”者得天下？

编辑 / 何林 曹莹　　图 / 崔峻珩

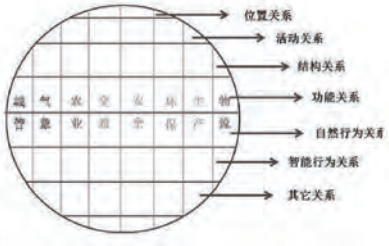
要了解块数据，首先你得知道什么是条数据

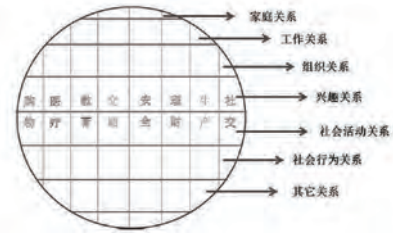
条数据可以定义为在某个行业和领域呈链条状串起来的数据。但这些数据被困在一个个孤立的条上，相互之间却不能连接起来。下图中的简图道出了目前的数据困境，无论是医疗还是教育，社交还是购物我们的数据被牢牢封锁在了领域内，无法融会贯通，更无法落地到服务为民。



相对于条数据

块数据，就是以一个物理空间或行政区域形成的涉及人、事、物的各类数据的总和。块数据是从数据到“数聚”的过程，这是块数据的起点。数据是分散的、分割的、碎片化的，当这些分散的、分割的、碎片化的数据聚合在一起的时候，就开始产生“块”。那么，这种“块”是一种什么东西呢？“块”就是一个多维的无限的变量。这个多维是思维范式，无限是跨界，变量是一种不确定性和不可预知性，这是大数据时代我们认识世界的基础，也是改造世界的方法。





如果还觉得抽象，再为您举个超级简单的“例子”：一个百货商场每天卖出许许多多的商品，每个商品有从原材料到加工成商品的生产过程的数据，也有品牌设计、广告营销和销售数量的销售数据，以及产品售后服务、商户反馈等服务数据，这些都是以产品为中心的“条数据”，而百货商场销售的商品种类、数量，男女老少在商场的购物、娱乐情况、天气、公交和停车场对商场经营情况的影响等等，事实上，商场的影院播映一部聚人气的大片时，商场的销售量也会随之上升，这些数据可以称为块数据，这个“块”是指这个商场，商场这个物理空间产生的数据总和就是商场的块数据。当块扩大到社区和城市层面时，在这个块上形成的数据总和就是本书所指的块数据。

一句话总结，块数据就是大层面、大空间、各个领域融会贯通的数据，它呼吁数据开放、增强数据的有效利用率，为各行各业所用并更好地服务于民生。我们国家庞大的数据更应该转化为创新的驱动力量为我所用。

知道了什么是块数据，再来看看块数据有哪些特征

块数据是从解构到重构的过程，这是块数据的机制。一旦数据被集聚，就会形成“块”，这种“块”对物质、能量、要素、权力、意识就会被解构。大数据时代人们获取信息的方式、交往和交友的方式、生活方式、意识形态、社会组织模式都将发生深刻的变革，这种变革的本质就是解构。每一次解构的结果都会产生新的重构，比如权力被权利所替代，这就是解构中的重构。下图从五个维度分析了块数据的特征。

块数据的价值

块数据是从多维到共享的过程，这是块数据的价值。大数据时代带给我们最大的好处是什么？如果概括起来解释就是多维和共享，就是每一个人在大数据时代能够快速 分享人类最先进的文明成果，这种多维和分享是在任何时间、任何地点、任何人、任何事、任何方式获得任何信息，这就是共享的魅力。共享是大数据时代对人类最大的贡献。我们过去不知道的事现在可以知道，我们过去不能获得的信息现在可以获得，过去少数人拥有的东西，现在大多数人都能拥有，这就是共享，共享正在成为一个新时代的标志。

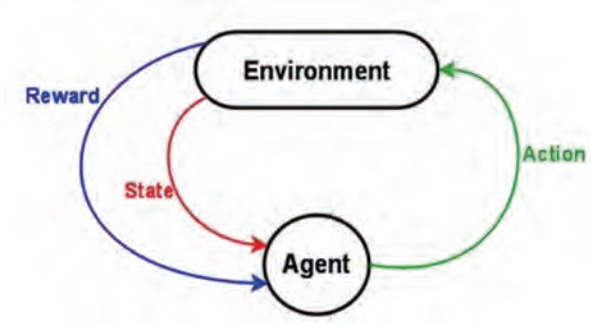
所以说得“块”者得天下，得“块”者得未来，一点都不夸张。

当然，大数据也面临着诸多的挑战，我们需要从制度上和法律上给予约束，规避风险。

DeepMind 背后的人工智能：深度学习原理初探

文 / 张天雷　　编辑 / 张楠

去年11月，一篇名为《Playing Atari with Deep Reinforcement Learning》的文章被初创人工智能公司DeepMind的员工上传到了arXiv网站。两个月之后，谷歌花了500万欧元买下了DeepMind公司，而人们对这个公司的了解仅限于这篇文章。近日，Tartu大学计算机科学系计算神经学小组的学者在robohub网站发表文章，阐述了他们对DeepMind人工智能算法的复现。



在arXiv发表的原始论文中，描述了一个单个的网络，它能够自我学习从而自动的玩一些老的电视游戏。它仅仅通过屏幕上面的图像和游戏中的分数是否上升下降，从而做出选择性的动作。

在训练的一开始，这个程序对游戏一点都不了解。它并不知道这个游戏的目标，是保持生存、杀死谁或者是走出一个迷宫。它对这个游戏的影响也不清楚，并不知道它的动作会对这个游戏产生什么影响，甚至不知道这个游戏中会有哪些目标物品。通过在这个游戏中尝试并且一遍一遍失败，这个系统会逐渐学会如何表现来获得比较好的分数。同时需要注意的是，这个系统对所有不同的游戏使用了同样的系统结构，程序员没有对不同程序给予这个程序任何特殊的提示，比如上、下或者开火等等。

最终结果显示，这个系统能够掌握一些游戏，并且比一些人类玩家还要玩得好。这个结果可以看作对AGI（Artificial General Intelligence）迈近的一小步，非常吸引人。文章给出了如下的细节，从任务、机器学习基础、深度学习模型和学习过程四部分阐述了他们的工作。

一、任务

这个系统获得了某个游戏屏幕的某幅图像，如下图是从一个最简单的游戏Breakout中获取的一幅图片。在简单的分析之后，它已经选择了如何做出下一步。这个动作已经被执行了，并且这个系统被告知了分数是否增加了、减少了或者没有变。基于这个信息，以及已经进行了的大量的游戏，这个系统会学习如何玩从而提高游戏的分数。

二、机器学习和人工神经网络

在深入深度学习的实现过程之前，文章先介绍了机器学习和人工神经网络的概念。

机器学习的一个非常通常的任务是这样的：给出了一个目标的信息，从而能够知道它属于哪个种类。在深度学习的过程中，程序想要决定在目前游戏状态下如何进行下一步动作。机器学习算法从例子中进行学习：给出了许多的目标例子和它们的种类，学习算法从中找出了那些能够鉴定某个种类的目标特征。学习算法会产生一个模型，能够在训练集中最小化错误分类率。这个模型之后会被用来预测那个未知目标的种类。

人工神经网络ANN（Artificial Neural Networks）是机器学习的一个算法。它是由人类的大脑结构产生的灵感。这个网络由许多节点组成，如同大脑由神经元组成，并且互相之间联系在一起，如同神经元之间通过神经突触和神经树联系在一起。对于每个神经元，都会对其应该传递的信号的情况做特殊规定。通过改变这些连接的强弱，可以使得这些网络计算更加快速。现在神经网络的结构通常由如下部分组成：

a. 神经的输入层（获得目标的描述）

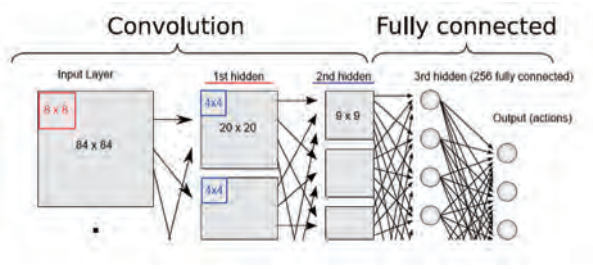
b. 隐藏层（主要部分，在这些层中学习）

c. 输出层（对于每个种类都一个神经节点，分数最高的一个节点就是预测的种类）

在学习过程结束之后，新的物体就能够送入这个网络，并且能够在输出层看到每个种类的分数。

三、深度学习

在这个系统中，一个神经网络被用来期望在当前游戏状态下每种可能的动作所得到的反馈。下图给出了文章中所提到的神经网络。这个网络能够回答一个问题，比如“如果这么做会变得怎么样？”。网络的输入部分由最新的四幅游戏屏幕图像组成，这样这个网络不仅仅能够看到最后的部分，而且能够看到一些这个游戏是如何变化的。输入被经过三个后继的隐藏层，最终到输出层。



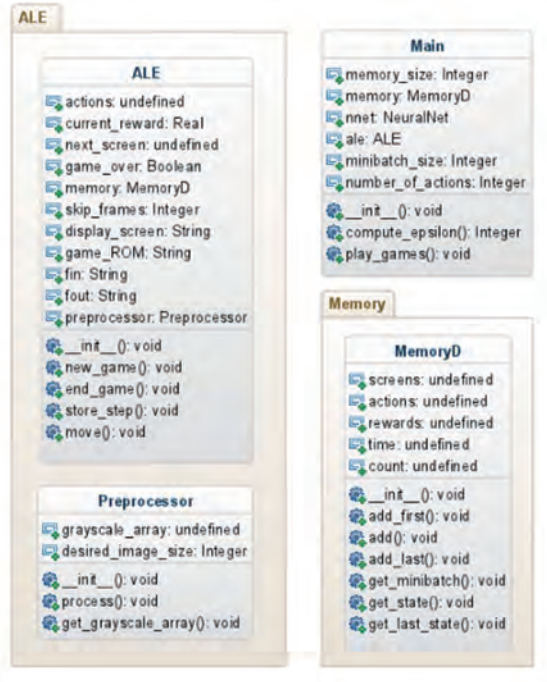
输出层对每个可能的动作都有一个节点，并且这些节点包含了所有动作可能得到的反馈。在其中，会得到最高期望分数的反馈会被用来执行下一步动作。

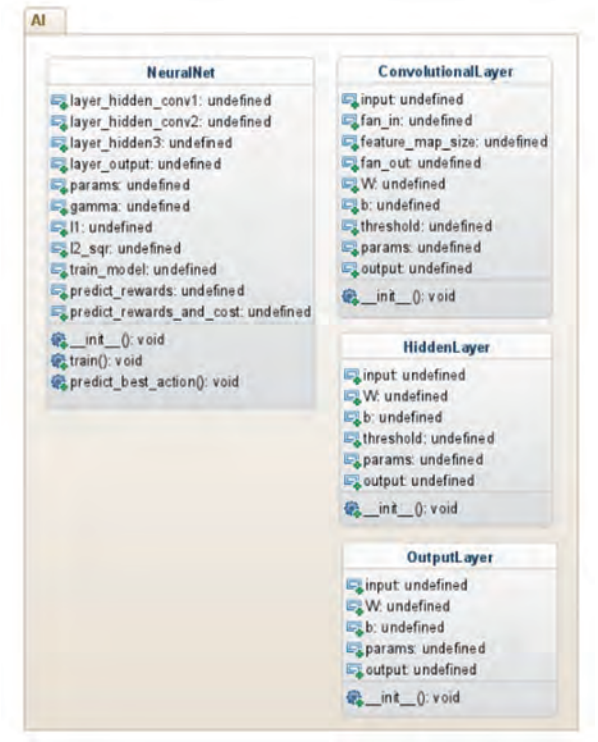
四、学习过程

系统通过学习过程来计算最高期望分数。确切地说，在定义了网络的结构之后，剩下唯一会变化的就只有一件事：连接之间的强弱程度。学习过程就是调整这些方式地权重，从而使得通过这个网络的训练样例获得好的反馈。

文章将这个问题当作一个优化问题，目标是获取最好的反馈。可以通过将梯度下降与激励学习方法结合起来解决。这个网络不仅仅需要最大化当前的反馈，还需要考虑到将来的动作。这一点可以通过预测估计下一步的屏幕并且分析解决。用另一种方式讲，可以使用（当前反馈减去预测反馈）作为梯度下降的误差，同时会考虑下一幅图像的预测反馈。

关于代码的更多细节，可以参考他们报告中所展示的代码架构图：





五、总结

文章最后给出了DeepMind深度学习的整个流程：

1.构建一个网络并且随机初始化所有连接的权重

2.将大量的游戏情况输出到这个网络中

3.网络处理这些动作并且进行学习

4.如果这个动作是好的，奖励这个系统，否则惩罚这个系统

5.系统通过如上过程调整权重

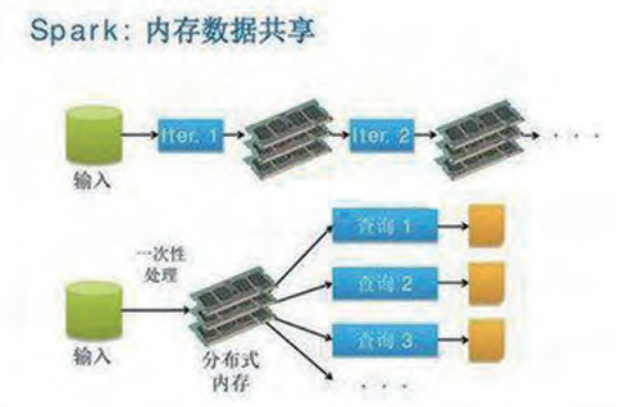
6.在成千上万次的学习之后，超过人类的表现。

这个结果可以看做是在AGI方向的从传统机器学习迈出的一小步。尽管这一步可能非常小，这个系统可能都不知道或者理解它做的事情，但是这个深度学习系统的学习能力远远超过之前的系统。并且，在没有程序员做任何提示的情况下，它的解决问题的能力也更加宽广。他们的代码可以在GitHub主页上找到。CDAIS 2015

Spark成为大数据分析领域新核心的五个理由

编辑/ 中商联数据分析委数据中心 崔欢欢　　图 / 崔峻珩

在过去几年当中，随着Hadoop逐步成为大数据处理领域的主导性解决思路，原本存在的诸多争议也开始尘埃落定。首先，Hadoop分布式文件系统是处理大数据的正确存储平台。其次，YARN是大数据环境下理想的资源分配与管理框架选项。第三也是最重要的一点，没有哪套单一处理框架能够解决所有问题。虽然MapReduce确实是一项了不起的技术成果，但仍然不足以成为百试百灵的特效药。



依赖于Hadoop的企业需要借助一系列分析型基础设施与流程以找到与各类关键性问题相关的结论与解答。企业客户需要数据准备、描述性分析、搜索、预测性分析以及机器学习与图形处理等更为先进的功能。与此同时，企业还需要一套能够满足其实际需求的工具集，允许他们充分运用目前已经具备的各类技能及其它资源。

就目前而言，并没有哪种标准化单一处理框架足以提供这样的效果。从这个角度出发，Spark的优势恰好得到了完美体现。

尽管Spark还仅仅是个相对年轻的数据项目，但其能够满足前面提到的全部需求，甚至可以做得更多。在今天的文章

中，我们将列举五大理由，证明为什么由Spark领衔的时代已经来临。

1. Spark让高级分析由理想变为现实

尽管多数大型创新型企业正在努力拓展其高级分析能力，但在最近于纽约召开的一次大数据分析会议上，只有20%的与会者表示目前正在企业内部部署高级分析解决方案。另外80%与会者反映其仍然只具备简单的数据准备与基本分析能力。在这些企业中，只有极少数数据科学家开始将大量时间用于实现并管理描述性分析机制。

Spark项目提供的框架能够让高级分析的开箱即用目标成为现实。这套框架当中包含众多工具，例如查询加速、机器学习库、图形处理引擎以及流分析引擎等等。

对于企业而言，即使拥有极为杰出的数据科学家人才(当然这一前提同样很难实现)，他们也几乎不可能通过MapReduce实现上述分析目标。除此之外，Spark还提供易于使用且速度惊人的预置库。在此基础之上，数据科学家们将被解放出来，从而将主要精力集中在数据准备及质量控制之外的、更为关键的事务身上。有了Spark的协助，他们甚至能够确保对分析结果做出正确的解释。

2. Spark让一切更为简便

长久以来，Hadoop面临的最大难题就是使用难度过高，企业甚至很难找到有能力打理Hadoop的人才。虽然随着新版本的不断出炉，如今 Hadoop在便捷性与功能水平方面已经得到了长足进步，但针对难度的诟病之声依然不绝于耳。相较于强制要求用户了解一系列高复杂性知识背景，例如 Java与MapReduce编程模式，Spark项目则在设计思路上保证了每一位了解数据库及一定程度脚本技能(使用Python或者Scala语言)的用户都能够轻松上手。在这种情况下，企业能够更顺畅地找到有能力理解其数据以及相关处理工具的招聘对象。此外，供应商还能够快速为其开发出分析解决方案，并在短时间内将创新型成果交付至客户手中。

3. Spark提供多种语言选项

在讨论这一话题时，我们不禁要问：如果SQL事实上并不存在，那么我们是否会为了应对大数据分析挑战而发明SQL这样一种语言?答案恐怕是否定的 ——至少不会仅仅只发明SQL。我们当然希望能够根据具体问题的不同而拥有更多更为灵活的选项，通过多种角度实现数据整理与检索，并以更为高效的方式将数据移动到分析框架当中。Spark就抛开了一切以SQL为中心的僵化思路，将通往数据宝库的大门向最快、最精致的分析手段敞开，这种不畏数据与业务挑战的解决思路确实值得赞赏。

4. Spark加快结果整理速度

随着业务发展步伐的不断加快，企业对于实时分析结果的需要也变得愈发迫切。Spark项目提供的并发内存内处理机制能够带来数倍于其它采用磁盘访问方式的解决方案的结果交付速度，传统方案带来的高延迟水平会严重拖慢增长。

从"物"往"人"的零售进化

文 / 周庭锐博士　编辑 / 曹莹　图 / 崔峻珩

实体零售业的巨大痛苦之一是：摸得到顾客的钱，看不到顾客的脸。我说看不到顾客的脸，其实不是真的看不到，而是记不住。或许少数店员记住了少数顾客的脸，但是从公司的立场考虑，零售业者对于大面积的顾客、乃至于顾客的家庭，不仅所知甚少，甚至根本无从把握。

今天的实体连锁零售企业已经完全不像数百年前零售业的老祖宗那样，以单一的门店，服务于为数甚少的社区型顾客。在那个时候，店老板自己充当店员，对于隔壁邻居家的底细如数家珍，对于他们的消费行为了若指掌，所以门店的单品管理非常准确高效，即使在处在相当落后的原始技术条件下，依然可以运行得低损耗高流转，从一家小小门店成就为若干年之后的富商巨贾。

但是今天的实体连锁体系，动辄数百数千门店家数，服务几十万几百万顾客，这些顾客的面貌早已模糊，更遑论去理解他们的个别消费行为。或许读者要抗议，我们有POS系统，我们时时刻刻进行着单品的ABC分类，我们的时间带别分析、品项分析、甚至购物篮分析，随时随处在刻画着我们门店顾客的画像。

这样的说法，可以说对，也可以说错。纵观零售理论发展史，零售管理的聚焦，从最早期的关注于“地”（零售吸引力理论），逐渐进化为关注于“物”（单品管理），最后进化到关注于“人”（POS系统与购物篮分析），这个从“物”到“人”的进化，是革命性的成就，无奈门店管理者的脑袋实在很难转化，在今天，可以说，80%以上的零售人员，即便已经有了许多年的POS使用经验，脑袋里念兹在兹的管理重点还是“物”，而不是“人”。

为什么零售管理的制高点是“人”而不是“物”？道理显而易见。零售买卖的真相是“人来买物”，而不是“物等人来买”。所以成就业绩的关键是在顾客进店之前，已经知道了他/她打算进来买些什么东西；而不是先进了一堆货，再来琢磨着如何把这些商品贩卖出去。从这一点来说，今天的零售门店远远不如百年前的村前小卖部。

今天的零售门店并不缺顾客的消费资料。零售门店有POS系统，更理想些有店内会员卡，最厉害的门店甚至在店内布置了RFID（无线射频）、Wifi热点、或Beacon（蓝牙基站）设备，来追踪门店顾客在店内的消费行踪。比较大的问题是，门店主管有没有这样的认识，理解到千辛万苦获取这些数据的目的是什么？

门店搜集到的消费行为数据可以做些什么用途呢？当然单品管理是其中最基础的工作，包括鲜度管理、畅销品滞销品识别、进货补货等等工作的优化；再进一步的用途显然是购物篮分析，可以理解顾客的每一次采购中，摆在心中的购物清单都是些什么？理解顾客的购物篮，让门店管理者知道原来单品的ABC分类是可能误导单品决策的，因为一些非热卖商品，或许恰好是另一个热卖商品的配套，缺此则失彼。

然后呢？可惜大部分的门店管理者到此为止就没有然后了。知道了前述这些消费行为信息，足够让门店管理者预测任何一个顾客进店来想买些什么吗？能够预测任何一位我们拥有顾客ID的人，明天会不会进来我们的门店，然后购买任何一样特定商品吗？

美国大型零售商Target曾经闹过这么一件大新闻：有位美

国父亲有天忽然收到来自Target寄来的，关于怀孕妈妈相关商品的促销信函，而收信人显然指明就是他钟爱的未成年女儿，为此他勃然大怒，向Target的管理当局提出妨碍名誉的严正控诉。但是最后查明，他的女儿确实怀孕了！Target比这位父亲更早地知道了这位未成年女孩怀孕的事实。

由于人类的消费特征有趋向固定的倾向（也就是固化的消费偏好），对于Target来说，他们这种超大型多品类百货商店的困境就在于，消费者很难认同他们的品项完整性真的能够跟专卖店比美。如果消费倾向是固定的，那么Target将会愈来愈难以获取足够的顾客基数，来维持如此庞大门店体量的开销。好消息是，人类一生之中存在几个特别容易改变消费倾向的时间点：离家独立、结婚、怀孕、退休。如果Target能够预测到他的顾客正处于这些时间点里的任何一点，他们就可能通过种种营销手段，来改变这位顾客的消费知觉，甚至改变他/她的消费偏好。

Target通过大数据技术，描绘了每一位顾客的潜在画像，进而推估了每一位顾客在任何时刻的背景特征，然后利用这个特征来进行针对性的营销工作。在这个例子里，Target其实是通过这位女孩的历史消费明细，得知她忽然之间开始严格使用不含化学添加物的沐浴精、洗发水等，因此捕捉到她可能已经怀孕这个事实。当然这背后存在科学研究的依据，美国人倾向在怀孕三个月左右开始避免接触任何带有化学添加物的产品，认为会危及胎儿的健康。

从这个例子里我们可以看到门店管理人员的思想聚焦从“物”转移到“人”的威力。事实上“顾客潜在画像”已经是现代零售管理里最为重要的一件工作，我们希望能够从顾客的消费明细、购买清单里，窥见他/她的性别、大概年纪、家里是否有婴幼儿、是否有车有房、是否养了宠物等等。我个人最近的工作之一就是在发展一套通用性的“消费者潜在画像智能识别系统”，能够单单只凭POS数据和会员卡号，就能判断任何一位顾客的潜在画像，进而接入单品管理系统，实现高度进化的零售管理。这是大数据目前在零售行业里的前沿，而且已经开始在台湾大型线上商超实际应用。CDAIS 2015

大数据治疗领导者“拍脑袋决策”流行病

文 / 中南财经政法大学MBA合作导师、中国统计信息服务中心大数据研究实验室主任 江青　编辑 / 曹莹

从古到今，凭借直觉和经验“拍脑袋决策”已成为统治阶级乃至当代领导者的一种流行病。殊不知，现实世界是复杂的，外部环境是多变的，这种“拍脑袋决策”所造成的后果往往是高风险，很有可能会因此而付出沉重的代价。

“拍脑袋”已成为领导者的决策流行病

大家知道，人不是机器，无论你的愿望有多么好，所做出的决定必定受到个人意识的作用，有很多其实是错误的。例如：河南省卢氏县委原书记杜保乾为“突出山城特色，体现南国风光”，竟然用自己的“构想”代替城建部门的规划，刨掉了原有郁郁葱葱的泡桐树，建起了棕榈树、云杉、四季桂、竹子、柳树、黄杨树、泡桐等7条不同风格的绿色街道，当然这些只适宜南方环境的树种，成活率极低，栽了死，死了栽，

栽了再死。为建造这些“形象工程”，仅资金耗费就达1300多万元，而且这些资金都是国家拨给卢氏县的扶贫款。再如，广东省珠海市当时的主要领导决心建造“全国最先进的机场”，于是拍板于1995年投资40亿元（总造价69亿元）建设珠海机场。他们原本指望机场靠营业收入来偿还银行的贷款和拖欠的工程款，不料却陷入了巨大的亏损之中，拖欠的巨额债务根本没有能力偿还。

还有一个典型的案例：某城市繁华地段有一个食品厂，因经营不善长期亏损，该市政府领导决定将其改造成为一个副食品批发市场，这样既可以解决企业破产后下岗职工的安置问题，又方便了附近居民，为此进行了一系列的前期准备，

包括项目审批、征地拆迁、建筑设计规划等。不曾想，外地一开发商已在离此地不远的地方投资兴建了一个综合市场，其中就有一个相当规模的副食品批发场区，足以满足附近居民和零售店的需求。面对这种情况，市政府陷入了两难境地：如果继续进行副食品批发市场建设，必然亏损;如果就此停建，则前期投入全部泡汤。而在这种情况下，该市领导又盲目作出决定，将食品厂厂房所在地建成一居民小区，由开发商进行开发，但对原食品厂职工没能做出有效的赔偿和再就业等方面的考虑，使该厂职工陷入困境而长期上访，对该市的稳定造成了隐患。

而最典型的案例则是上世纪九十年代，全国各地竞相出现的开发区建设热，不少地方罔顾实际从省到市到乡纷纷上马，各地大大小小名目繁多的开发区多达一万个。结果却是事与愿违，不少开发区既荒芜了大片土地，又损失了巨额资金。河南省灵宝市豫灵镇借债建开发区，结果欠下一亿多元债务，按照该镇的经济实力，还清债务得需要100年。

综上几个案例不难看出，这些不良后果的起因，大都是领导者盲目自信“拍脑门” 决策的“流行病”所造成的。那么，如何避免类似悲剧的发生呢？人们又把希望寄托在 “调查研究、民主决策”和“聘请管理咨询专家” 论证上。然而，所谓的调查研究、民主决策，不过是在当地的小圈子里搞些调查，找几个部门或民意代表开个座谈会，再通过常委会举手表决而已。且不说这种小范围的“调查研究”和民主决策的实效如何，光时间也消费不起。那么，聘请管理咨询专家是否就能有效避免决策失误呢?

专家之谜

首先让我们回顾一下上世纪50年代三门峡工程决策失误的一段往事。

1952年8月,中苏两国政府商定将黄河综合规划列为苏联的技术援助项目,苏联政府同意派水利专家来华指导。三门峡工程是在黄河上修建的第一座大型水利枢纽工程，国家为其投入了大量的人力、物力和财力，给予了很高的期望。

1958年11月，三门峡工程完成对黄河的截流。1960年9月实现关闸蓄水拦沙。1961年2月,当坝前水位达332.58米(尚未到设计高度)的时候,泥沙淤积就迅速发展。下半年,15亿吨泥沙全部铺在了从潼关到三门峡的河道里,潼关的河道抬高,渭水河口形成拦门沙,渭河航运窒息,从无水患的渭河两岸也不得不修起了防洪堤。而关中平原的地下水无法排泄,田地迅速出现盐碱化甚至沼泽化,粮食因此减产。这一年,潼关以上的黄河、渭河大淤成灾。

1962年3月,水库内的淤积已经开始迅速发展,潼关河床在一年半的时间内暴长4.5米,成了名副其实的“悬河”。最严重的是河床的泥沙淤积向上游延伸,严重危害着关中平原的安全和以西安为中心的工业基地。为此，水电部不得不在郑州召开会议,将美妙的“黄河清”暂时放在一边,而把三门峡水库的运用方式由当初定的“拦蓄上游全部来沙”改为“滞洪排沙”。水位不得不降低。而失去了大水头,第一台15万千瓦的发电机组发电不足一个月便丧失了用武之地,只好改装5万千瓦小机组。同时耗费大量的人、财、物力打通排水洞以泄泥沙。如此一来,投进水库不下百亿元的资全打了“水漂”。

这是新中国成立后的一起重大决策失误，而这项决策正是在苏联专家的帮助下完成的。

改革开放以来，我国的各类专家咨询机构相继建立，国务院各部委及各级地方政府大多成立了专家咨询委员会，并制定出专家咨询委员会工作制度（条例），规范了决策咨询的适用范围、工作程序、咨询形式、激励机制等。

然而，有资料援引中组部“建立决策咨询机制” 研究组在13个省、直辖市和自治区的调查发现，目前公共决策专家咨询中存在很多缺陷，如： 缺少对专家的激励约束和咨询效果的评估；决策者有选择地对重大事项进行决策咨询，有选择地确定决策咨询机构或专家个人；决策咨询过程和结果并没有形成完善的记录和档案管理，导致决策咨询缺乏责任追究机制；决策咨询工作走过场，智库或智库成员成为决策者的附庸和利益代言人等等。

由此可见，管理咨询专家的“高、大、上”与效果不成正比，专家们拥有粉丝但解决不了实际问题。那么，领导者如何才能摆脱“拍脑袋决策”的流行病，从过分依赖专家的“迷宫”里走出来，找到一条更为科学、高效、智慧的管理决策之道呢?

数据帮你治疗“领导流行病”

先轻松一下，看看当年我们的林彪同志在指挥辽沈战役中，是如何准确判断并一举擒获“出身黄埔军校并留学法国著名的圣西尔军校”的敌首廖耀湘的。

【数据积累，运筹帷幄之中】

林彪从红军带兵时起，身上就有个小本子，上面记载着每次战斗的缴获、歼敌数量。每次打完仗，林彪就亲自往上面添加数字，并为之沾沾自喜…令人感觉到这个23岁任军长，25岁就当军团长的人，似乎有点小气。

1948年辽沈战役开始之后，在东北野战军前线指挥所里面，每天深夜都要进行例常的“每日军情汇报”：每支部队歼敌多少、俘虏多少，缴获的火炮、车辆多少，枪支、物资多少……司令员林彪的要求很细，俘虏要分清军官和士兵，缴获的枪支要统计出机枪、长枪、短枪，击毁和缴获的汽车也要分出大小和类别。其实，这几乎是一大堆千篇一律枯燥无味的数据!

林彪几乎终日倒骑着椅子面对墙上的地图观察和思考。他要计算到进攻时有全胜的把握，还要留出退路。而这些精确的部署都来自于那些看上去十分乏味的数据准备。

【数据分析，找到最有价值的信息】

1948年10月14日，东北野战军以迅雷不及掩耳之势，仅用了30小时就攻克锦州，并且在全歼了守敌十余万之后不顾疲劳挥师北上，与从沈阳出援的二十余万敌精锐廖耀湘集团在辽西相遇。一时间形成了混战，战局瞬息万变，谁胜谁负实难预料。

一天深夜，值班参谋正在向林彪等指挥员们读着某师上报的其下属部队的战报，说他们部队碰到了一个不大的遭遇战，歼敌一部分，其余逃走。这时，林彪突然叫了一声“停!”他的眼里闪出了光芒，问：“刚才念的在胡家窝棚那个战斗的缴获，你们听到了吗?”

大家带着睡意的脸上出现了茫然，不都是差不多一模一样的枯燥数字吗?林彪扫视一周，见无人回答，便接连问了三句：“为什么那里缴获的短枪与长枪比例比其它战地略高?”、“为什么那里缴获和击毁的小车与大车的比例比其它战地略高?”“为什么在那里俘虏和击毙的军官与士兵的比例比其它战地略高?”

人们还没有来得及思索，等不及的林彪司令员大步走向挂满军用地图的墙壁，指着地图上的那个点说：“我猜想，不，我断定!敌人的指挥所就在这里!”

林彪可以如此笃定，就取决于他每晚必做的功课，这些战报在他脑中汇集成为一个庞大的数据库，当出现差异，他可以及时获取到准确信息，找出价值所在。

【准确定位，精细画像，一举拿下廖耀湘】

得出结果之后，林彪口授命令，追击从胡家窝棚逃走的那部分敌人，并坚决把他们灭掉。各部队要采取分割包围的办法，把失去指挥中枢后变得混乱的几十万敌军切成小块，逐一歼灭。

而此时的廖耀湘，正庆幸自己刚刚从偶然的一场遭遇战中安全脱身并与自己的另外一支部队汇合。没想到，紧追而来的解放军迅速把他的新指挥部团团围住，拼命攻击。同时，漫山遍野的解放军战士中不断有人喊着：“白净脸，矮又胖，金丝眼镜湖南腔，不要放走廖耀湘!”

一场激战过后，穿着满身油渍伙夫服装的廖耀湘只好从俘虏群中站出来，无奈地说“我是廖耀湘”，沮丧地举手投降。

廖耀湘对自己精心隐蔽的野战司令部那么快就被发现并灭掉，觉得实在不可思议，他认为那纯属是一个偶然事件，输得不甘心。当他得知林彪是如此得出判断并迅速出击之后，这位出身黄埔军校并留学法国著名的圣西尔军校的新六军军长说：“我服了，败在他手下，不丢人。”

没错，这就是利用数据分析管理决策的神奇功效！与凭直觉和经验“拍脑袋决策”相比，哪个更智慧、更全面、更精准、更可靠呢?

现如今，随着互联网、云计算、大数据的到来，我们的领导者不用再像当年林彪那样装个小本本，听参谋念，一笔笔地记，然后捧着一堆数据对着地图彻夜不眠了。你只须配备好大数据班子人才，然后坐在电脑桌前，鼠标轻轻一点，你所需要的动态信息、智慧的现状分析以及智库参考即可得来。聪明的领导者们，大数据时代已经到来，你准备好了吗？！

零售企业如何借助数据分析进行品牌定位

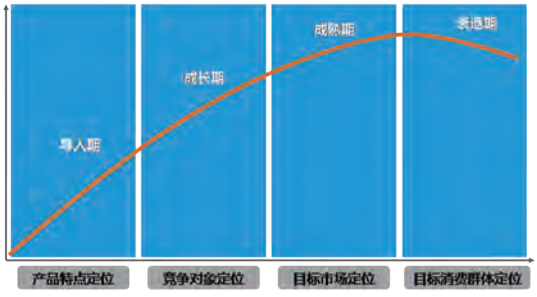
文/ 中颢润（北京）项目数据分析师事务所 张佳丽　　编辑 / 黄平

零售业，与每个人生活紧紧相关，正经历着翻天覆地的变革，消费客群的转移，消费行为的改变，实体零售商的客流不断减少，销售额以10亿级规模在缩水。目前来看，零售行业整体增长减缓的趋势仍在继续。数据显示，2014年1-8月，中国社会消费品零售总额16.61万亿元，同比增长12.1%，较去年同期下降0.7个百分点。而大型零售企业的增速也在明显放缓，限额以上的企业商品销售总额8.3万亿元，增速为9.7%，低于社会消费品零售总额2.4个百分点。

在如此严峻的市场环境下，有些品牌的销售额却一路高歌猛进。如王老吉的销售额由2002年的1.8亿蹿升至2013年的150亿，对于这样快速的成功，中颢润认为主要是源于“王老吉是饮料不是药”的品牌精准定位。在参与市场竞争的过程中，品牌已经成为企业以最低市场风险、最低营销成本来获取更大市场回报的工具。几乎所有的企业都在探寻那些杰出品牌长盛不衰的原因，如同从诞生至今已有百年的可口可乐。

事实上，建立长久不衰的品牌的必要条件是鲜明的品牌定位。中颢润认为零售企业进行品牌定位时可以从产品特点、目标市场、竞争对手、目标消费群体四个方面进行定位，同时还需要结合该企业所在行业的成熟度，考虑采用何种品牌定位。

图1：行业成熟度对应的定位方式



产品特点定位

在行业处于导入期，消费者对于品类没有了解，需要进行产品教育，产品特点传达得准确，有助于消费者快速了解品牌，同时也可以使产品在消费者心目中占有一定的位置。为此我们需要从产品特征、包装、服务等多方面作研究，通过市场调查掌握市场和消费者消费习惯的变化，在必要时对产品进行重新定位。

针对一个新品牌的面市，目标顾客的反应肯定有很大的差异——漠视、关注、尝试和充当传播者的都有。由市场实践分析发行顾客这四种行为状态的比例依次是60、20、15、5，这基于一个前提，即企业在一个有效期内应有各种有效和中等强度的媒体和推广策略 否则这些数字将没有意义。企业在应用时仍应依照实际的市场调查结果来制订相应的推广计划，它依然是有一定的指导意义的。因为这四种行为表现涵盖了顾客对新品牌的态度，而且就是这些显著的态度决定了企业的推广策略。

竞争对象定位

当行业处于成长期，消费者对于产品有很好的了解，竞争越来越激烈，导致产品的同质化，因此需要进行竞争性品牌定位。成长期的产品，其性能基本稳定，大部分消费者对产品已熟悉，销售量快速增长，竞争者不断进入，市场竞争加剧。企业为维持其市场增长率，需要了解市场中的竞争环境、所处的竞争地位以及竞争对手的分析。

对竞争环境的分析主要从潜在进入者、购买者、供应商和替代品角度进行定性分析，了解潜在进入者的威胁程度、供应商和购买者讨价还价的能力以及替代品的危险系数。对于竞争地位的分析主要采用兰查斯特模型，通过市场占有率情况分析该品牌所处的阶段（见图2）。而对于竞争对手的分析则需采用定标比超分析，通过市场增长率与市场份额的相关分析，了解哪些品牌可以作为本企业的标杆，哪些企业需要赶超，哪些企业需要提防其增长过快冲击本企业的市场，从而对企业构成威胁（见图3）。

图2：竞争地位分析模型

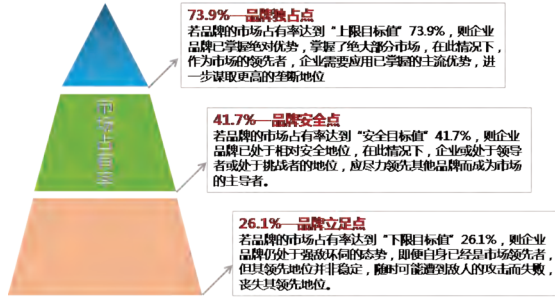
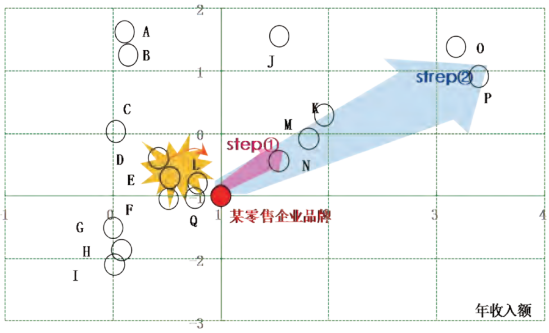


图3：竞争对手趋势图



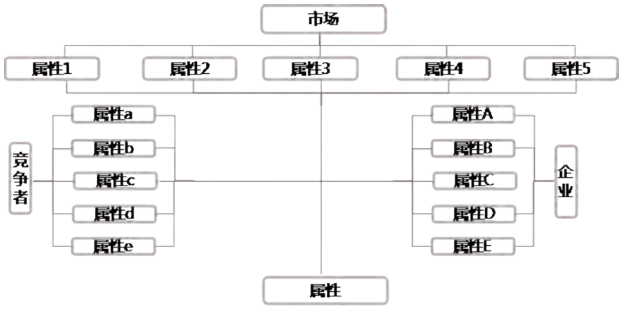
该零售企业要以P为标杆，赶N，防L

目标市场定位

当行业处于成熟期，消费者对于各个品牌及产品都很了解，竞争导致市场的同质化日益严重，市场进入细分时代，需要重新界定和进行区隔。在营销时代关键的不是对某一产品本身做些什么,而是你在消费者的心目中做了什么。单凭质量上乘和价格低廉已难以获得优势。当今,成功品牌的竞争优势已主要来源于市场定位。

市场定位的实质就是寻找差异性，是企业着意配置和挖掘的特色产品，而要想着意配置和挖掘出产品的特色，企业在实践中有必要使用一些专业性的工具使之具体化。企业在进行市场定位是需要同时考虑消费者、竞争者和企业自身情况三个因素，并找出每个因素中产品或服务的几个主要属性。我们围绕定位的影响因素采用市场定位模型对市场进行精准的定位。市场定位模型主要是基于市场研究的基础找到细分的目标市场，同时还需要找到竞争者的市场定位点并研究竞争者的定位对消费者购物行为的影响，企业在找到细分市场和竞争者定位后，可以分析自身的实力，然后进行自己的定位，从而可以找到市场的最佳位置（见图4）。

图4：市场定位模型



目标消费群体定位

当行业处于衰退期，产品创新越来越差，市场成熟度越来越高，品牌之间已经形成了各自号召力，消费者也比较固定，此时需要进行情感维护，对消费群体进行精准营销，提高其忠诚度。

我们可以用人口学的（年龄、性别、教育程度）、心理学的（价值观、文化取向）和行为学的（消费行为模式、一般行为特征）指标来定义或者标示出这些人的特点。通过研究发现市场上任何产品的用户基本上都可以分成四大群：第一群叫做创新个性型，领导风气，勇于尝新；第二种类型的人叫做易感流行型，他很敏感，特别会被时尚所影响，并介绍给其他人； 第三类人是舆论主宰型和物有所值型，他们特点就是，看看广告、报纸上、人家说什么好； 第四群人我们把他叫做简洁务实型，重视价格，注重实惠，不为时潮所动。

我们通过对于这四类消费群体的分析，可以找出来不同类型消费者最终的用意是什么？我们通常用决策树的分析方法。在决策树上，我们会看到不同的路径。不同的路径代表一个品牌陈述自己存在理由和说服消费者的不同的逻辑思维线路。如果我们再做一个定量的研究，我们就会发现每个路径它们有不同百分比的。

定量研究不仅可以帮助我们分析价值路径，还可以提供各相关价值之间的联系程度，从而我们可以绘制品牌张力图。即通过对于不同群体的价值选择的分析，我们就会发现什么，有的价值在不同群体当中是相对稳定的，我们把那种价值叫做恒定价值，有一些价值则在不同群体之间所处的位置有很大的区别，对于一个品牌的价值来说，如果它只能表现恒定的价值，表明这种品牌是一种老成持重的品牌，而如果一个品牌，它要是一个有活力的品牌，就应能表现某些快速变动的价值。所以，如果我们要做一个综合的品牌，那我们就要选择一种恒定的价值跟一些活跃价值的组合。

总之，关于品牌定位理论的应用和实践对于国内企业开展品牌经营，加入全球化的品牌竞争是十分必要而且有效的。不需要很长的时间，品牌竞争即将全面卷入各行各业，因此，国内产品制造商必须提前着手规划、建设自己的品牌。

北京市不同区县酒店分布及价格水平探究性分析

文/ 中商联数据分析委数据中心 崔欢欢　　编辑 / 石爱英　　图 / 崔峻珩

目前北京分为14个区2个县，分别为门头沟区、平谷区、石景山区、房山区、怀柔区、通州区、大兴区、顺义区、昌平区、丰台区、西城区、东城区、海淀区、朝阳区、密云县、延庆县。其中东城区为原东城区和崇文区、西城区为宣武区和原西城区。北京市目前有酒店5千多家，酒店的分布和价格水平一方面能在很大程度上反映该区县经济水平，另一方面可以让消费者当处于某一区域时，对该区域价格水平有个总体的认知。

下面主要从总体酒店分布及价格水平、各区县不同价格区间下酒店个数、价格峰值主要分布区域及酒店名称、北京主要连锁酒店名称及分布等进行探究。

一、总体酒店分布及价格水平

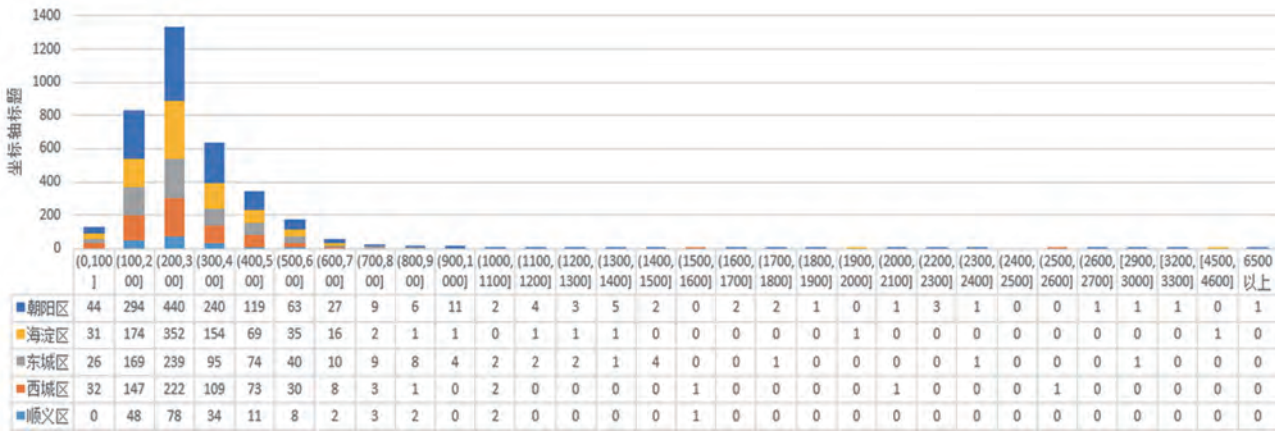
a、主要酒店价格分布及平均价格水平

总体上，北京市酒店主要分布在市中心，例如朝阳区、海淀区、东城区、西城区等，在郊区或距市中心较远的区县，酒店个数明显下降；酒店价格平均水平在(100,400]之间,其中除门头沟区、房山区、通州区酒店价格在200元以下,其他各区均明显在200元以上，这与其他城市酒店价格水平有着明显的区别；酒店价格与酒店个数所呈现出的趋势有着明显的差异，例如：由于顺义区首都机场的原因，使得顺义区平均价格水平较高，作为旅游区的密云县和平谷区虽为远郊，其酒店价格水平较其他近郊和五环内区的价格水平并没有明显差异。

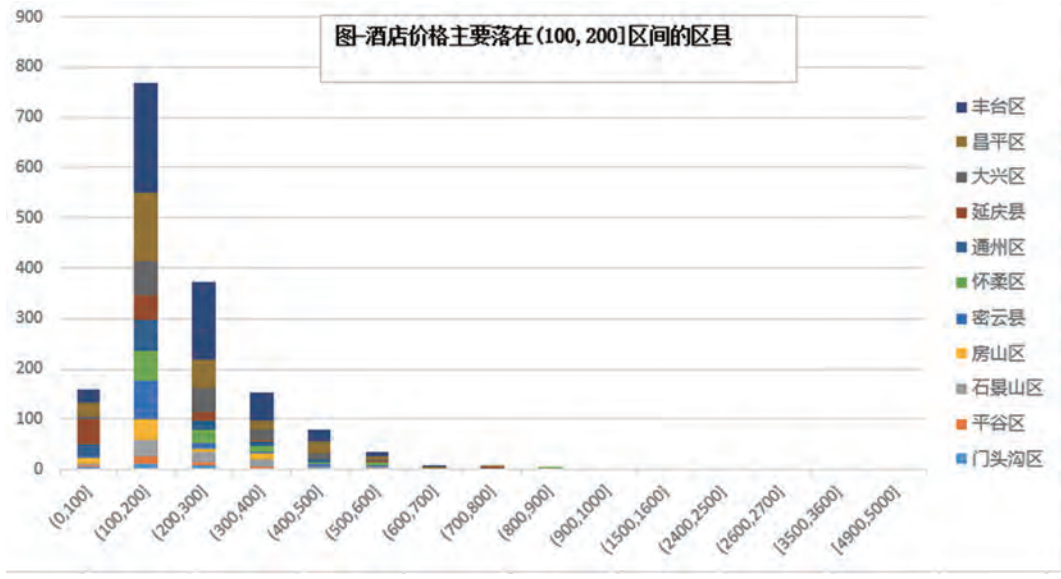
图-北京市各区/县酒店分布及价格水平

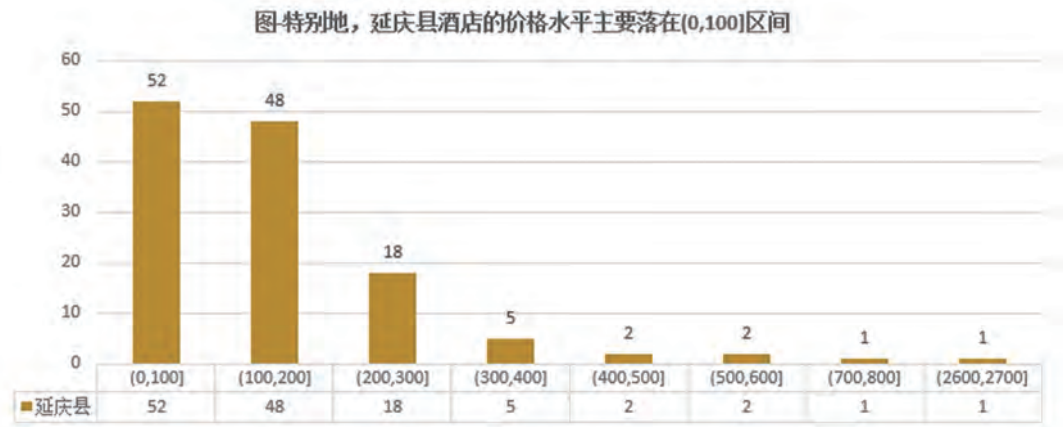
b、主要价格区间内的区县

图：价格水平主要在(200,300]区间的北京各区



	(0,100]	(100,200]	(200,300]	(300,400]	(400,500]	(500,600]	(600,700]	(700,800]	(800,900]	(900,1000]	(1000,1100]	(1100,1200]	(1200,1300]	(1300,1400]	(1400,1500]	(1500,1600]	(1600,1700]	(1700,1800]	(1800,1900]	(1900,2000]	(2000,2100]	(2200,2300]	(2300,2400]	(2400,2500]	(2500,2600]	(2600,2700]	(2900,3000]	(3200,3300]	(4500,4600]	6500以上
朝阳区	44	294	440	240	119	63	27	9	6	11	2	4	3	5	2	0	2	2	1	0	1	3	1	0	0	1	1	1	0	1
海淀区	31	174	352	154	69	35	16	2	1	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0
东城区	26	169	239	95	74	40	10	9	8	4	2	2	2	1	4	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0
西城区	32	147	222	109	73	30	8	3	1	0	2	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0
顺义区	0	48	78	34	11	8	2	3	2	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0





北京市各区县主要价格水平区间为(100,200]、(200,300]、其中：

1）门头沟区、平谷区、石景山区、房山区、密云县、怀柔区、通州区、大兴区、昌平区、丰台区、10个区县的主要价格区间为(100,200]；

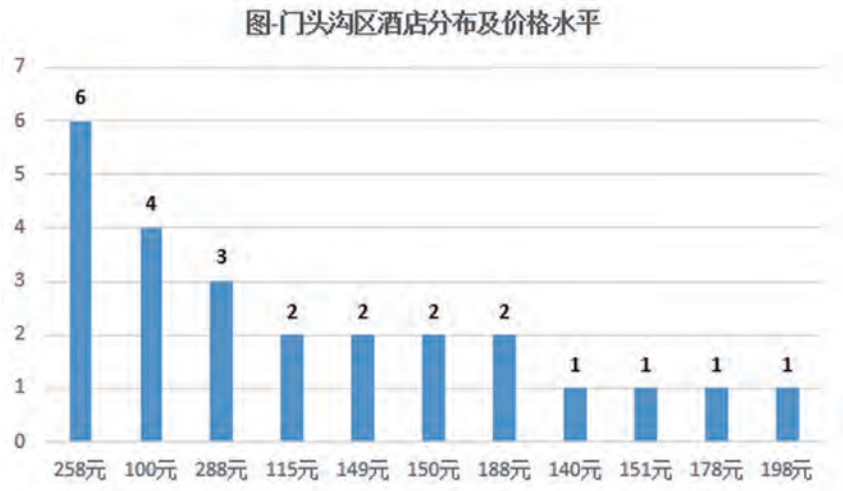
2）顺义区、西城区、东城区、海淀区、朝阳区、5个区县的主要价格区间为(200,300]；

3）另外特别的：延庆县40%的价格区间为(0,100]，37%的价格区间为(100,200]

2、分区县酒店分布及价格水平

1)、门头沟区

酒店价格：	100	115	140	149	150	151	178	188	198	258	288
酒店个数：	4	2	1	2	2	1	1	2	1	6	3



门头沟渠酒店的价格水平差异度不是很大，均为300元以下，最低价为100元

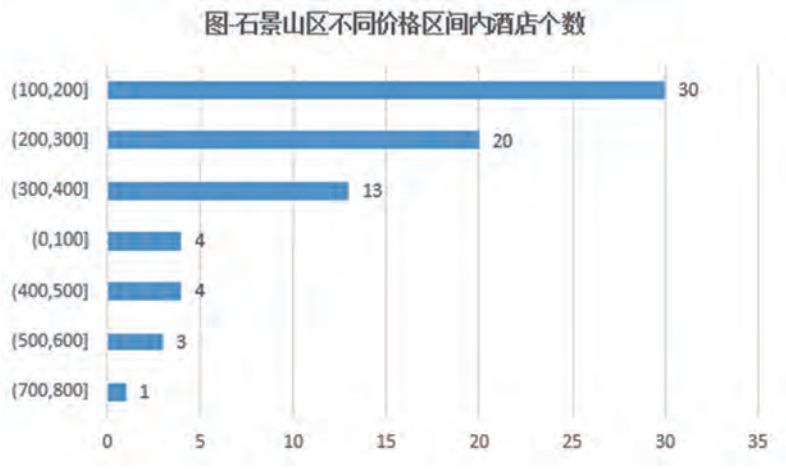
2）、平谷区:

酒店价格：	50	80	100	116	129	139	150	158	168	169	180	252	280	312	320	390	600
酒店个数：	1	1	2	5	2	2	1	1	2	1	1	2	2	3	1	2	1
酒店价格：	780	828	2580	3589													
酒店个数：	1	1	1	1													

平谷区的酒店价格水平差异很大，从50元至3589元不等，主要是由于平谷区渔阳国际度假村酒店价格为3589为最大值,

3）、石景山区

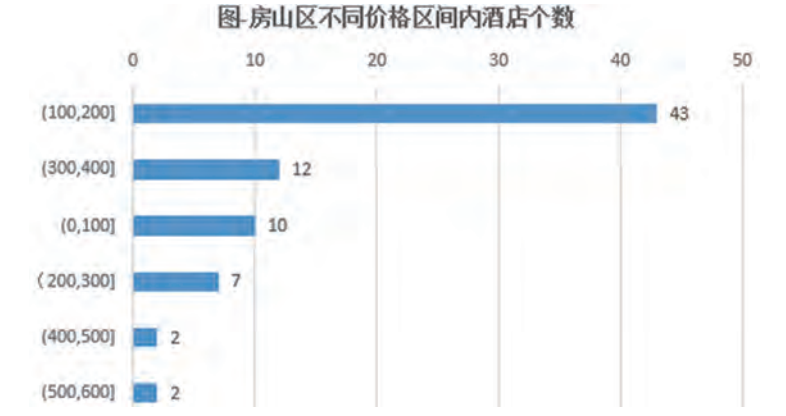
酒店价格：	70	99	100	115	134	135	145	148	150	159	160	166	167	168	170	178	179	197	198	201	208	217	218	228
酒店个数：	1	2	1	1	2	1	2	2	1	1	2	1	1	3	1	2	4	2	4	1	2	2	1	4
酒店价格：	238	296	298	329	350	388	398	450	488	578	712													
酒店个数：	2	3	5	2	3	3	5	3	1	3	1													



石景山区酒店价格水平差异度不是很大，40%的酒店价格位于(100,200]之间，边界区域相对分布较小

4）、房山区

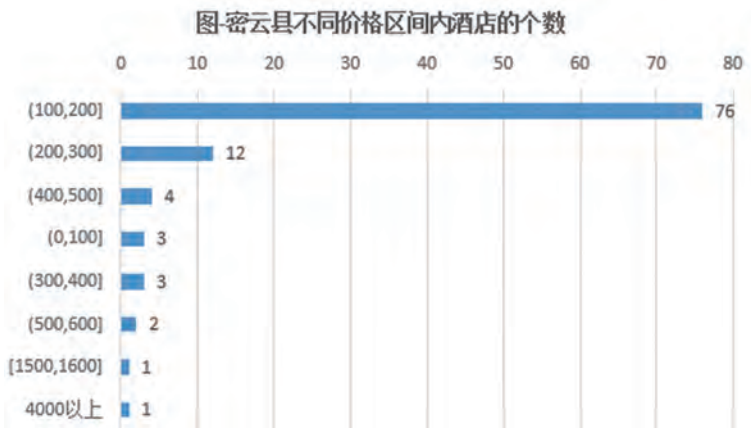
酒店价格：	88	97	98	100	108	117	118	120	128	130	138	139	142	147	150	157	168	169	170	178	180	186	189	220
酒店个数：	2	1	4	3	1	2	3	8	1	3	6	3	2	2	3	1	2	1	1	1	1	1	1	2
酒店价格：	249	290	320	323	356	380	450	560																
酒店个数：	1	4	3	2	1	6	2	2																



房山区酒店价格水平以50%的概率落在(100,200]，价格的差异不是明显

5）、密云县

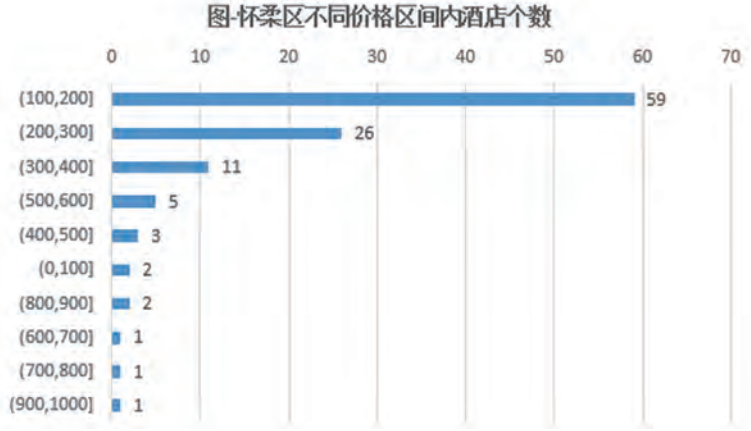
酒店价格：	89	95	119	120	125	132	137	142	145	150	151	160	179	185	188	198	200	230	240
酒店个数：	2	1	1	6	2	2	2	2	2	1	1	1	1	30	20	3	2	2	1
酒店价格：	260	268	280	288	350	360	450	488	588	1500	4940								
酒店个数：	1	1	3	4	2	1	2	2	2	1	1								



密云县酒店价格的差异度很大，但接近3/4的酒店在(100,200]区间内，这与密云县为远郊相符，但由于密云县为旅游区，北京水镇酒店为4940元/天的价格，其古北水镇民宿也高达1500元/天

6）、怀柔区

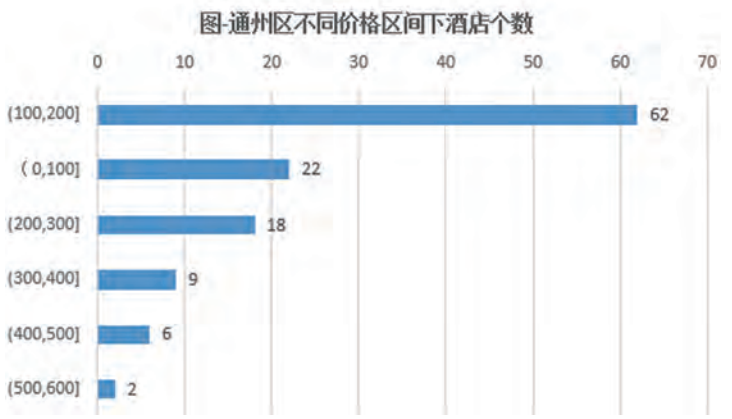
酒店价格：	80	113	120	130	140	148	150	155	158	160	168	169	170	178	180	188	198	208	226	248	260	266	278	280
酒店个数：	2	1	20	1	4	4	8	1	2	3	3	1	1	1	3	2	4	1	2	4	4	3	2	3
酒店价格：	288	304	318	360	390	399	450	510	567	580	678	710	836	890	980									
酒店个数：	7	1	2	2	3	3	3	2	2	1	1	1	1	1	1									



怀柔区的价格主要为(100,200]，酒店水平差异不是很明显

7）、通州区

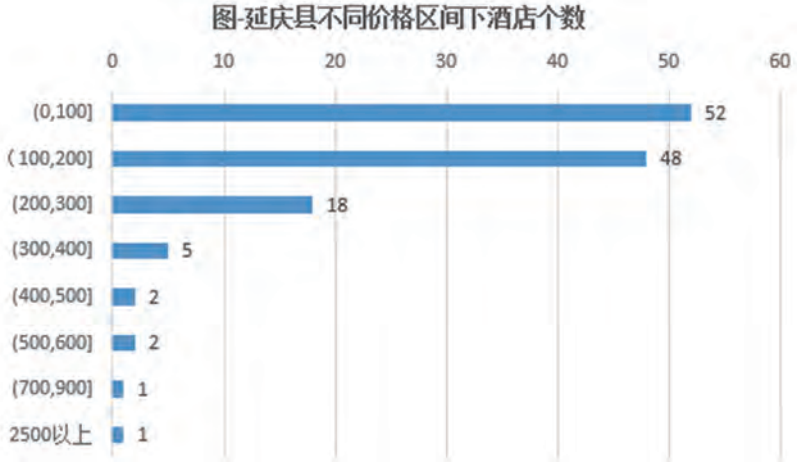
酒店价格：	70	80	88	89	90	98	99	108	109	110	111	118	120	123	128	129	130	137	138	139	148	157	158	159
酒店个数：	2	2	2	1	1	3	11	5	1	2	2	2	2	2	9	3	2	2	3	3	2	3	1	2
酒店价格：	160	166	167	168	169	180	189	198	208	218	248	250	280	288	298	308	318	332	400	488	499	598		
酒店个数：	1	4	3	1	3	1	1	2	2	3	3	1	5	3	1	2	2	2	3	4	2	2		



通州区的价格以超过50%的概率落在(100,200]，酒店水平差异不是很明显

8）、延庆县

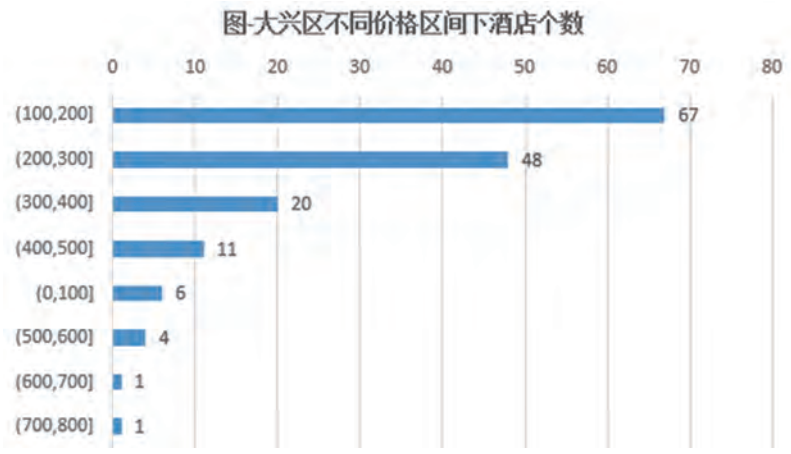
酒店价格：	60	70	80	98	100	120	130	134	140	147	148	150	151	170	180	198	200	248	258
酒店个数：	7	6	18	1	20	26	2	1	6	2	2	2	1	1	1	2	2	2	6
酒店价格：	280	288	290	298	340	378	480	598	757	2680									
酒店个数：	1	4	2	3	2	3	2	2	1	1									



延庆县酒店价格以40%的概率落在 (0,100]，37%的概率落在 (100,200]，除长城脚下公寓价格为2680元/天外，其余均在1000元以下

9）、大兴区

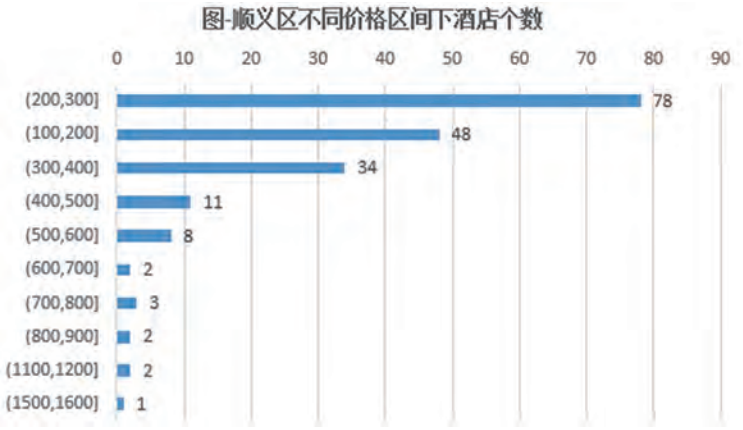
酒店价格：	90	99	110	116	118	119	120	130	138	140	148	155	158	160	167	168	172	177	178	179	180	183	187	188
酒店个数：	1	5	3	2	4	4	3	1	3	4	2	1	4	1	3	9	1	1	5	1	1	2	1	3
酒店价格：	189	198	199	208	209	218	220	227	228	229	238	240	249	258	259	268	280	288	298	304	308	338	370	378
酒店个数：	2	1	5	6	1	4	3	1	1	3	2	6	1	3	3	3	3	2	6	2	5	2	2	3
酒店价格：	380	388	398	408	410	420	448	559	569	677	750													
酒店个数：	2	1	3	2	3	3	3	2	2	1	1													



大兴区的价格水平均在1000元以下，其中以40%的概率落在(100,200]内

10）、顺义区

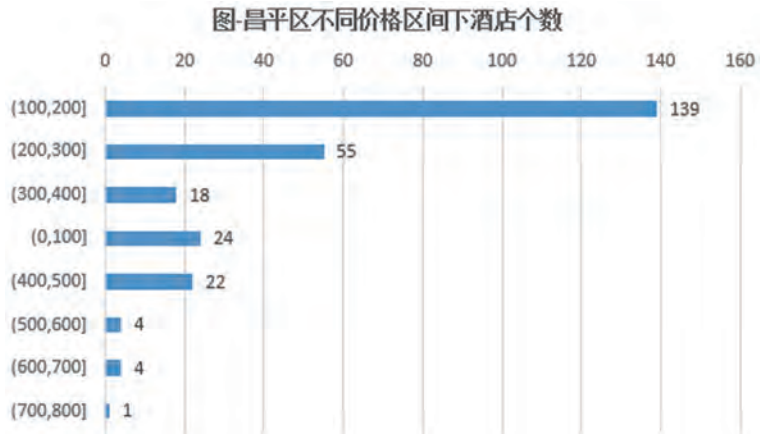
酒店价格：	128	130	132	138	150	153	155	158	160	168	177	178	179	187	188	189	190	198	208
酒店个数：	4	3	2	10	2	3	1	3	4	2	1	2	1	1	2	2	3	2	4
酒店价格：	218	219	220	225	229	233	238	240	245	246	260	268	270	276	278	279	280	288	298
酒店个数：	3	1	3	1	2	2	4	6	4	1	6	13	4	4	6	2	3	4	4
酒店价格：	299	318	320	348	358	368	370	388	397	399	430	450	468	488	498	500	538	550	588
酒店个数：	1	3	4	4	3	5	3	7	3	2	1	3	2	2	2	1	2	4	2
酒店价格：	638	670	720	740	789	818	860	1148	1180	1551									
酒店个数：	1	1	1	1	1	1	1	1	1	1									



顺义区的酒店价格均在100元以上，且以42%的比率落在(200,300]之间，其次为(100,200]，酒店价格峰值主要在1100以上

11）、昌平区

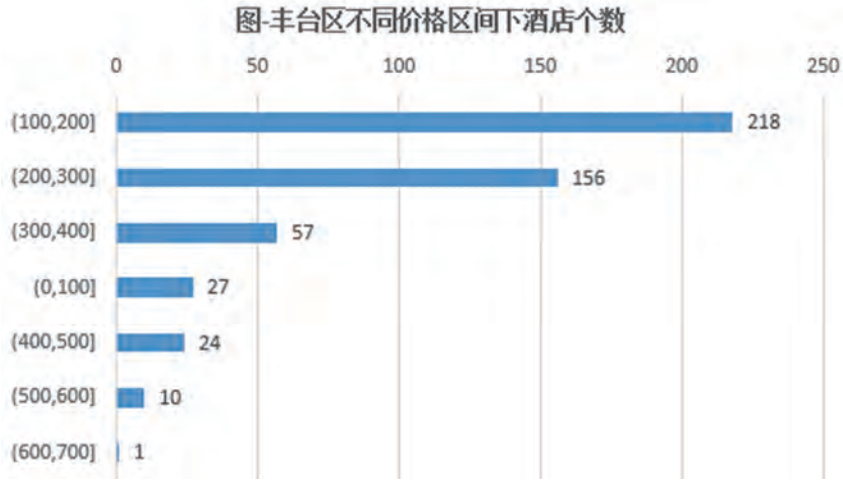
酒店价格：	55	59	60	66	68	96	98	99	100	108	109	111	118	119	120	126	127	128	130	132	138	139	140	142
酒店个数：	1	1	1	1	1	2	4	7	6	4	2	2	10	5	16	2	2	6	1	4	5	12	3	2
酒店价格：	145	147	148	150	151	158	159	160	165	167	168	169	170	178	179	180	187	188	189	197	198	199	207	208
酒店个数：	3	2	3	2	2	4	1	2	2	3	7	2	1	3	1	1	2	8	1	2	6	5	2	3
酒店价格：	210	218	219	220	228	229	230	238	239	246	248	258	268	278	288	298	300	310	319	320	328	336	380	398
酒店个数：	2	3	4	3	4	1	4	1	2	4	6	3	1	2	1	3	6	2	1	2	3	3	4	3
酒店价格：	428	430	438	450	468	470	478	480	500	598	600	618	680	686	758									
酒店个数：	2	2	3	3	2	2	2	2	4	2	2	1	2	1	1									



昌平区的价格水平以50%的比率落在(100,200]，其次为(200,300]，酒店价格水平均在1000元以下。

12）、丰台区

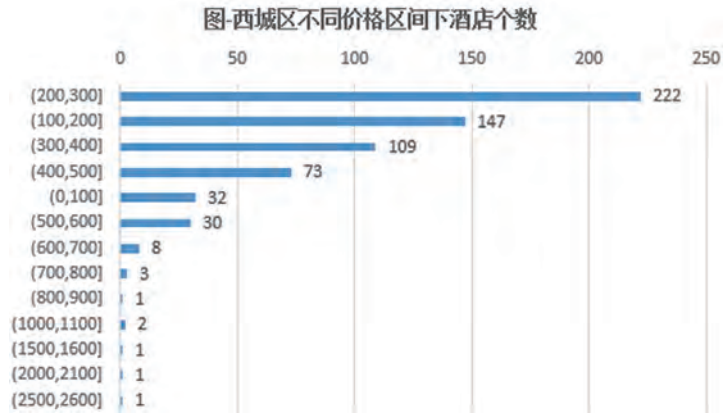
酒店价格：	50	65	78	79	88	89	91	98	99	108	109	115	118	119	120	125	127	128	130	131	133	138	139	140
酒店个数：	2	1	1	1	2	3	2	4	11	4	4	1	5	4	7	2	2	11	2	2	2	15	2	2
酒店价格：	146	148	149	151	156	157	158	159	160	161	167	168	169	170	171	176	177	178	179	180	183	187	188	189
酒店个数：	7	5	6	1	1	2	5	4	4	4	2	11	10	3	1	1	8	4	4	1	1	4	15	7
酒店价格：	190	197	198	199	200	208	209	210	217	218	220	227	228	229	230	234	237	238	239	240	245	246	248	258
酒店个数：	2	3	22	13	2	10	8	1	2	7	1	4	9	3	3	2	1	12	4	2	4	3	2	4
酒店价格：	259	265	268	269	275	278	279	280	284	287	288	298	299	301	303	315	318	320	322	328	334	338	348	350
酒店个数：	4	2	9	5	2	7	2	2	3	2	20	14	2	3	1	4	5	2	2	5	3	1	6	1
酒店价格：	355	368	376	388	398	438	448	458	468	488	492	498	499	500	511	528	540	546	568	688				
酒店个数：	3	5	1	3	12	3	5	2	2	2	2	2	2	4	2	2	2	2	2	1				



丰台区的价格水平均在700元以下，且以43%的概率落在(100,200]，以超过30%的概率落在(200,300]之间

13）、西城区

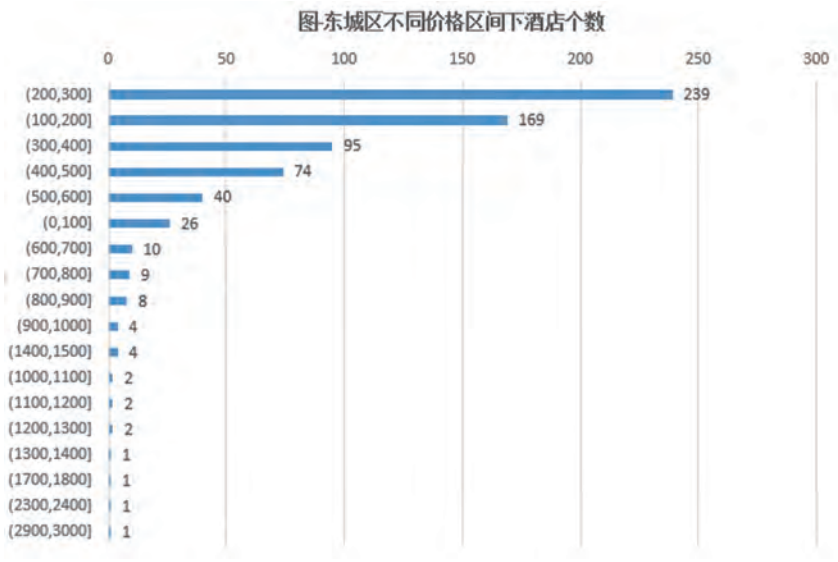
酒店价格：	28	50	60	65	70	78	80	88	90	99	108	110	111	118	119	128	129	130	138	139	142	146
酒店个数：	1	2	8	2	3	4	6	2	2	2	3	4	2	2	2	2	5	3	5	2	2	2
酒店价格：	148	149	150	158	159	160	168	170	174	178	179	180	188	189	190	197	198	199	202	207	208	209
酒店个数：	12	4	3	2	4	6	17	8	6	10	5	3	6	3	3	4	15	2	2	4	3	2
酒店价格：	217	218	219	220	227	228	229	230	235	237	238	239	241	242	244	246	247	248	249	250	256	257
酒店个数：	2	1	3	7	1	13	5	1	1	2	19	1	2	2	2	2	4	9	6	3	4	1
酒店价格：	258	259	260	267	268	269	271	273	275	277	278	279	280	282	284	287	288	289	298	299	300	303
酒店个数：	12	4	2	4	11	7	3	2	3	4	8	9	2	2	2	2	6	2	21	12	2	2
酒店价格：	308	311	313	317	318	319	322	328	332	338	340	348	349	350	353	358	359	360	368	370	380	387
酒店个数：	1	1	2	1	2	1	2	5	2	8	2	7	1	2	3	18	1	7	4	2	6	3
酒店价格：	388	389	390	396	398	400	418	420	428	430	432	438	440	448	450	454	458	466	468	476	478	488
酒店个数：	2	2	5	3	12	2	4	3	1	4	3	12	3	2	11	2	2	2	1	2	2	4
酒店价格：	494	498	508	517	518	530	538	550	568	578	580	588	598	600	603	635	638	683	688	690	699	700
酒店个数：	2	13	2	2	4	2	2	4	2	2	2	2	4	2	1	1	1	1	1	1	1	1
酒店价格：	712	723	800	870	1010	1065	1556	2056	2631													
酒店个数：	1	1	1	1	1	1	1	1	1													



西城区酒店价格水平差异性较大，其中以35%的概率落在(200,300]内，其次以23%的概率落在(100,200]，价格水平的峰值主要在1500以上，酒店分别为：北京金融街丽思卡尔顿酒店（2631元 / 天）、北京金融街洲际酒店（2056元 / 天）、北京金融街威斯汀大酒店（1556元 / 天）

13）、东城区

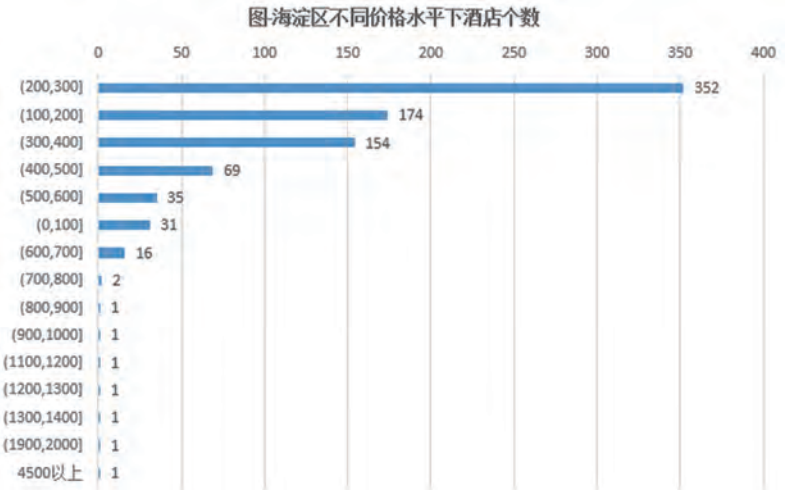
酒店价格：	60	70	78	88	98	100	108	118	120	128	138	139	148	149	150	151	155	158	159	160	161	163
酒店个数：	3	2	2	7	10	2	2	3	3	4	10	6	13	6	3	2	2	6	4	1	2	1
酒店价格：	168	169	173	174	175	178	179	183	187	188	189	190	192	196	197	198	199	207	208	216	218	220
酒店个数：	4	2	2	2	2	8	14	3	1	18	5	3	1	2	4	18	12	4	10	2	15	6
酒店价格：	227	228	230	238	239	240	246	248	251	257	258	260	263	266	268	269	275	277	278	280	283	288
酒店个数：	10	15	5	20	4	8	4	12	2	5	9	3	2	4	19	1	3	2	6	8	2	10
酒店价格：	290	292	298	299	300	307	308	309	310	313	318	320	322	328	329	332	338	339	342	348	350	358
酒店个数：	2	4	20	15	7	1	5	1	1	1	6	1	1	10	1	2	8	1	2	4	10	2
酒店价格：	359	360	370	376	380	388	389	398	400	418	420	428	438	443	445	448	458	460	462	468	474	477
酒店个数：	2	1	3	1	5	2	2	18	4	6	3	9	6	2	2	8	2	1	2	4	1	2
酒店价格：	478	480	486	488	491	496	498	499	500	508	518	529	530	536	538	558	568	569	580	598	625	640
酒店个数：	6	4	2	2	2	2	4	2	2	4	4	2	2	1	4	4	2	2	2	13	2	1
酒店价格：	650	665	675	680	690	698	699	745	758	788	790	791	795	798	800	818	843	850	855	856	858	897
酒店个数：	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	2	1	1	1	1
酒店价格：	945	948	990	1000	1013	1050	1136	1157	1242	1300	1318	1458	1472	1481	1500	1725	2361	2990				
酒店个数：	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1				



东城区酒店价格水平差异性较大，其中以34%的概率落在(200,300]内，其次以24%的概率落在(100,200]，价格水平的峰值主要在1700以上，酒店分别为：北京华尔道夫酒店 （2990元 / 天）、北京东方君悦大酒店（2361元 / 天）、北京雅诗阁来福士中心服务公寓（1725元 / 天）

14）、海淀区

酒店价格：	35	38	39	50	60	75	78	79	80	85	88	90	98	100	108	109	118	120	125	128	129	130
酒店个数：	1	1	1	2	1	2	2	2	1	1	6	3	2	6	7	2	1	4	2	5	11	2
酒店价格：	138	142	145	148	150	158	160	164	165	168	169	170	174	177	178	179	180	181	187	188	189	190
酒店个数：	8	3	2	15	5	9	1	3	1	11	3	2	1	1	5	4	4	1	2	20	1	2
酒店价格：	192	193	196	197	198	199	200	202	207	208	209	210	215	217	218	219	220	227	228	230	233	237
酒店个数：	1	1	1	1	13	18	1	2	2	10	1	3	2	5	5	3	3	13	15	5	2	5
酒店价格：	238	239	240	243	245	246	247	248	249	255	256	257	258	259	260	265	267	268	269	272	274	275
酒店个数：	10	5	4	2	2	4	2	17	11	2	3	7	29	11	3	3	5	22	5	1	2	7
酒店价格：	278	279	282	284	286	287	288	289	290	294	296	297	298	300	303	307	308	313	318	320	322	328
酒店个数：	8	6	3	5	2	2	38	3	4	10	2	2	31	3	2	3	3	2	7	2	4	10
酒店价格：	330	332	338	339	340	341	348	349	350	351	356	358	360	365	368	370	378	380	388	389	390	398
酒店个数：	1	1	10	1	6	1	9	1	3	1	1	14	3	3	10	6	4	11	11	2	3	13
酒店价格：	399	400	408	419	428	430	438	448	450	454	455	458	460	468	480	485	488	490	498	512	516	518
酒店个数：	3	3	3	3	11	2	4	5	2	2	2	2	1	4	8	2	6	4	8	1	2	2
酒店价格：	519	528	538	545	550	558	565	568	588	598	608	611	615	620	624	625	628	638	650	667	668	676
酒店个数：	2	2	2	2	5	7	2	2	4	2	1	1	1	1	1	1	1	1	1	1	1	1
酒店价格：	688	689	698	748	758	888	945	1127	1280	1334	1942	4600										
酒店个数：	1	1	2	1	1	1	1	1	1	1	1	1										

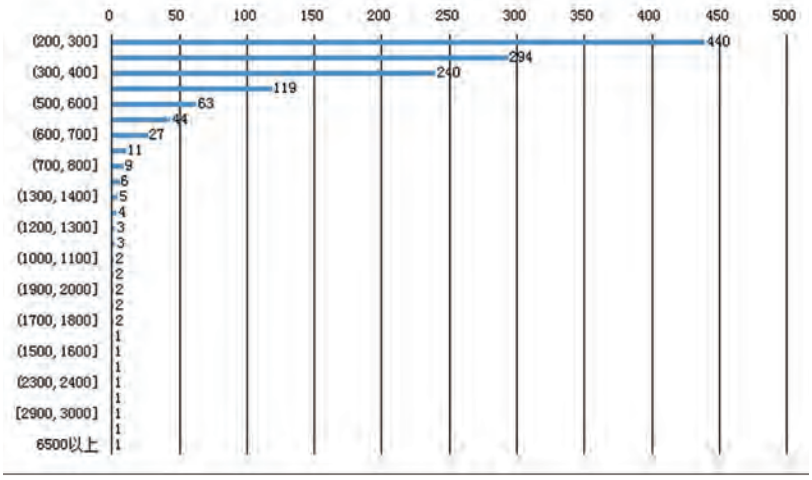


海淀区酒店价格水平差异性较大，其中以42%的概率落在(200,300]内，其次以21%的概率落在(100,200]，价格水平几乎全部为700元以下，价格的峰值分布不均匀价格最高的为北京颐和安缦酒店（4600元/天）、主要是近颐和园附近

15）、朝阳区

酒店价格：	25	28	29	30	65	68	70	78	79	80	88	90	98	99	100	108	109	110	118	120	128	130
酒店个数：	9	2	1	2	1	2	6	2	1	2	2	1	2	9	2	3	2	2	3	2	5	4
酒店价格：	137	138	139	142	148	150	151	156	158	159	160	161	165	167	168	169	170	177	178	179	180	181
酒店个数：	1	21	6	2	26	2	2	1	12	5	3	1	5	4	15	4	18	1	7	12	8	2
酒店价格：	187	188	189	190	196	197	198	199	200	205	206	207	208	209	210	216	217	218	219	220	227	228
酒店个数：	2	30	12	1	3	8	37	21	1	1	2	9	18	3	1	2	5	14	5	5	8	6
酒店价格：	229	230	236	237	238	239	240	244	245	246	247	248	249	251	255	256	257	258	259	260	264	265
酒店个数：	6	6	1	6	19	16	3	2	5	18	3	14	3	2	5	12	4	26	5	11	4	8
酒店价格：	267	268	269	270	275	277	278	279	280	284	287	288	289	294	296	297	298	299	300	305	308	309
酒店个数：	2	28	4	2	10	4	8	6	10	3	4	25	4	5	2	3	45	12	5	3	8	2
酒店价格：	310	313	318	319	322	327	328	329	330	332	338	340	341	348	349	350	358	360	363	365	368	369
酒店个数：	1	4	8	1	8	2	11	1	3	2	13	1	3	27	7	11	11	1	1	1	19	12
酒店价格：	370	376	378	379	380	388	389	390	398	399	400	406	408	409	412	414	419	420	426	428	430	436
酒店个数：	3	3	3	1	6	18	1	2	32	7	3	3	3	2	2	2	3	9	2	11	2	1
酒店价格：	438	450	458	459	460	466	468	469	478	479	480	482	488	489	490	498	499	500	519	520	528	538
酒店个数：	7	6	9	2	6	2	6	2	2	2	8	2	5	1	4	9	2	4	2	2	2	4
酒店价格：	547	548	550	558	560	567	568	578	580	587	591	598	599	600	608	625	628	634	638	650	655	656
酒店个数：	2	1	2	6	2	2	4	4	2	3	2	16	3	4	1	3	1	1	2	1	1	1
酒店价格：	660	666	673	680	689	690	698	700	703	708	710	723	730	770	778	788	803	818	858	888	889	898
酒店个数：	1	2	1	3	2	1	5	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1
酒店价格：	920	932	949	950	966	978	980	988	990	998	1006	1035	1111	1148	1150	1200	1231	1238	1278	1340	1375	1380
酒店个数：	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
酒店价格：	1390	1495	1500	1553	1610	1612	1725	1730	1826	1955	2000	2013	2220	2223	2300	2688	2990	3220	6900			
酒店个数：	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1			

图-朝阳区不同价格区间下酒店个数



朝阳区酒店价格水平差异性较大，其中以34%的概率落在(200,300]内，其次以24%的概率落在(100,200]，以19%的概率落在(300,400]；格水平的几乎全部为700元以下，最低价格为25元，最高价格为6900元，其中价格水平较低的主要为学生公寓、求职公寓等、价格为6900的酒店为北京嘉里大酒店(原香格里拉嘉里中心大酒店)

三、酒店价格主要峰值分布探究

由以上分析可知，大部分酒店价格水平均在1000一下，对于价格峰值分布在(1000,1500]、(1500,2000]、以及(2000,3500]、3500以上进行分析。

1)、(1000,1500]

主要分布区域：

城　　区：	朝阳区	东城区	海淀区	密云县	顺义区	西城区
酒店数量：	16	11	3	1	2	2

主要价格:

```
 [1] 1006 1035 1111 1148 1150 1200 1231 1238 1278 1340 1375 1380 1390 1495 1500 1013 1050 1136 1157 1242
[21] 1300 1318 1458 1472 1481 1127 1280 1334 1180 1010 1065
```

酒店及价格:

城区	酒店名称	最低价. 天.	地址
朝阳区	北京东隅酒店	1006	朝阳区酒仙桥路22号，近酒仙桥。
朝阳区	北京上东盛贸饭店	1035	朝阳区东四环北路2号，近芳园南街、霄云路。
朝阳区	北京五洲皇冠国际酒店	1111	朝阳区北四环中路8号，近亚运村北京国际会议中心东侧。
朝阳区	北京丽都皇冠假日酒店	1148	朝阳区将台路6号，与京密路交汇处。
朝阳区	北京朝阳悠唐皇冠假日酒店	1150	朝阳区三丰北里3号（外交部东南侧），近朝外市场街。
朝阳区	北京外国专家大厦	1200	朝阳区北四环中路华严北里8号，健翔桥辅路东南侧。
朝阳区	北京雅诗阁服务公寓	1231	朝阳区建国路乙108号，近招商局。
朝阳区	北京亚太大厦酒店公寓	1238	朝阳区雅宝路8号，近雅宝路。
朝阳区	北京千禧大酒店	1278	朝阳区东三环中路7号，近财富中心。
朝阳区	北京JW万豪酒店	1340	朝阳区华贸中心建国路83号，近新光天地。
朝阳区	北京昆仑饭店	1375	朝阳区新源南路2号，燕莎友谊商城正对面。
朝阳区	北京新云南皇冠假日酒店	1380	朝阳区东北三环西坝河太阳宫桥东北角云南大厦，近西坝河大中电器。
朝阳区	北京瑞居酒店	1390	朝阳区工体北路50号　（工人体育场内21看台对面），近工人体育场西路。
朝阳区	北京健一公馆	1390	朝阳区东四环中路红领巾桥北路西侧，近朝阳区绿化局。
朝阳区	北京紫檀万豪行政公寓	1495	朝阳区建国路23号，近中国紫檀博物馆。
朝阳区	北京丽都服务式酒店公寓	1500	朝阳区将台路6号，近京顺路。
东城区	京奥克伍德华庭酒店·绿	1013	东城区东直门外斜街8号，近新源里东街。
东城区	京新红资红色主题四合院酒	1050	东城区东四六条九号，近地铁5号线张自忠路站。
东城区	王府半岛酒店（原王府饭	1136	东城区王府井金鱼胡同8号，近王府井步行街。
东城区	北京丽晶酒店	1157	东城区金宝街99号，近北京金宝大厦。
东城区	北京万豪酒店	1242	东城区建国门南大街7号，近北京古观象台。
东城区	北京励骏酒店	1300	东城区金宝街90-92号，近国旅大厦。
东城区	北京涵珍园国际酒店	1318	东城区交道口南大街秦老胡同20号，近中央戏剧学院。
东城区	北京八十一酒店·演乐酒店	1458	东城区东四南大街演乐胡同70号，近东四南大街。
东城区	北京万豪行政公寓	1472	东城区东城区霞公府街1号北门，近王府井大街/商业街。
东城区	北京新世界酒店	1481	东城区祈年大街8号，近东打磨厂街。
东城区	北京金隅喜来登酒店	1500	东城区北三环东路36号，近地铁5号线和平西桥站A口。
海淀区	北京钓鱼台大酒店	1127	海淀区三里河路49号，近月坛南街。
海淀区	中关村软件园国际会议服务	1280	海淀区东北旺西路8号中关村软件园四号楼AB座，近百度大厦。
海淀区	北京香格里拉饭店	1334	海淀区紫竹院路29号，紫竹桥西北侧。
密云县	北京古北水镇民宿	1500	密云县北京市密云县古北口镇司马台村古北水镇旅游。　密云风景区 】
顺义区	北京临空皇冠假日酒店	1148	顺义区天竺镇府前一街60号，近京顺路。
顺义区	北京首都机场朗豪酒店	1180	顺义区首都机场三号航站楼二经路1号，近T3航站楼。
西城区	北京西单美爵酒店	1010	西城区宣武门内大街6号，近君太百货、汉光百货。
西城区	北京金融街酒店式公寓	1065	西城区金城坊街1号，全国政协礼堂南200米。

2)、(1500,2000]

酒店名称	城区	最低价（天）	地址
北京首都机场希尔顿酒店	顺义区	1551	顺义区首都机场3号航站楼，近三经路与二纬路的交汇处。
北京万达索菲特大饭店	朝阳区	1553	朝阳区建国路93号万达广场C座，近Soho现代城。
北京金融街威斯汀大酒店	西城区	1556	西城区金融大街乙9号，近月坛北桥。
北京北辰洲际酒店	朝阳区	1610	朝阳区北京市朝阳区北辰西路8号院4号楼，近亚运村鸟巢。
北京瑞吉酒店（原北京国际俱乐部饭店）	朝阳区	1612	朝阳区建国门外大街21号，建国门桥东北角。
北京雅诗阁来福士中心服务公寓	东城区	1725	东城区东直门南大街1-2号，近东直门内大街。
金茂北京威斯汀大饭店	朝阳区	1725	朝阳区东三环北路7号，近地铁亮马桥站。
北京海航大厦万豪酒店	朝阳区	1725	朝阳区霄云路甲26号，近鹏润大厦。
北京怡亨酒店	朝阳区	1730	朝阳区东大桥路9号，近日坛北路。
北京希尔顿酒店	朝阳区	1826	朝阳区东三环北路东方路1号，燕莎桥/三元桥交界。
北京盛捷中关村服务公寓	海淀区	1942	海淀区海淀中街15号，近中关村海龙大厦。
北京康莱德酒店	朝阳区	1955	朝阳区东三环北路29号，近呼家楼。
北京燕莎中心凯宾斯基饭店	朝阳区	2000	朝阳区亮马桥路50号，近燕莎友谊商城。

3)、(2000,3500]

酒店名称	城区	最低价（天）	地址
北京四季酒店	朝阳区	2223	朝阳区亮马桥路48号，近燕莎桥。
北京国贸大酒店	朝阳区	2688	朝阳区建国门外大街1号，近建国门外大街。
北京盘古七星酒店	朝阳区	2220	朝阳区北四环中路27号，盘古大观，近奥运村国家体育馆。
北京中国大饭店	朝阳区	2300	朝阳区建国门外大街1号，近国际贸易中心。
北京丽思卡尔顿酒店（华贸中心）	朝阳区	2990	朝阳区华贸中心建国路甲83号，近新光天地。
北京柏悦酒店	朝阳区	3220	朝阳区建国门外大街2号，近国贸立交桥西南角。
北京瑜舍	朝阳区	2013	朝阳区三里屯路11号院1号楼，近地铁团结湖站。
北京金融街丽思卡尔顿酒店	西城区	2631	西城区金融街金城坊东街1号，太平桥大街路口。
北京金融街洲际酒店	西城区	2056	西城区金融街11号，月坛北桥东北角。
北京华尔道夫酒店	东城区	2990	东城区金鱼胡同5-15号，近乐天银泰百货。
北京东方君悦大酒店	东城区	2361	东城区长安街1号东方广场，北京东方广场内。
北京远见桃花海度假别墅	平谷区	2580	平谷区金海湖镇祖务村祖务西路1号，近京东大溶洞（胡关路)平谷城区
北京长城脚下的公社	延庆县	2680	延庆县京藏高速路水关长城出口，近石佛寺村。 十三陵水库、居庸关长城风景区

4)、3500以上

酒店名称	城区	最低价（天）	地址
北京嘉里大酒店(原香格里拉嘉里中心大酒店)	朝阳区	6900	朝阳区光华路1号，光华桥西侧。
北京水镇大酒店	密云县	4940	密云县北京市密云县古北口镇司马台村古北水镇旅游
北京颐和安缦酒店	海淀区	4600	海淀区颐和园宫门前街1号，近颐和园路。
北京渔阳国际度假村	平谷区	3589	平谷区东高村镇大旺务村688号，近大旺务中路. 平谷城区

北京市各区县价格水平的峰值分布有明显的差异，其中：

1 ）、有”相对上界”：

门头沟区均在300元以下、石景山区均在800元以下、房山区均在600元以下、怀柔区均在1000元以下、通州区均在600元以下、昌平区均在1000元以下、丰台区均在700元以下；

2 ）、无”相对上界”：

平谷区最高价格水平为3589元 / 天、密云县最高价格水平为4940元 / 天、延庆县最高价格水平为2680元、顺义区最高价格水平为1551元、西城区最高价格水平为2631元、东城区最高价格水平为2990元、海淀区最高价格水平4600元、朝阳区最高价格水平为6900元。

贵州华鑫成项目数据分析师事务所

贵州华鑫成项目数据分析师事务所有限公司，成立于2014年10月，地址：贵阳市观山湖区长岭北路贵阳国际会议展览中心D区3栋4楼3号。是贵州省首批成立的专业从事项目数据分析的服务性机构，公司由多位资深项目数据分析师发起成立，并拥有一支集项目投资、财务分析、工商管理等多领域的复合型团队。

贵州华鑫成项目数据分析师事务所是新成立的事务所，业务水平还需提升，但我们会以专业的方法、谨慎的作风、客观的态度、公正的原则以及热情的服务，为行业协会、中小型企业、国内外银行、投融资公司、政府组织等机构提供投资项目数据分析、投资项目评估、经济效益评价、项目数据分析研究、数据处理、项目融资、投资项目策划、社会经济咨询、投资中介等专业的、系统的服务，并为项目投资方以及融资方提供一份具有经济性、权威性、客观性、公正性、实用性的项目数据分析报告。

贵州华鑫成数据分析事务所将秉承诚信、客观、科学、实效、公正的经营理念，高水准的专业品质，为客户提供最有价值的服务，努力成为客户发展历程中值得信赖的合作伙伴，力求成为项目数据分析行业先领，为推动中国数据分析行业的发展贡献力量 。

联系方式：江睿　13985154312 / 0851-6577704
办公地址：贵州省贵阳市瑞金北路14号成黔大厦3楼308室

北京鼎盛恒信项目数据分析师事务所

北京鼎盛恒信项目数据分析师事务所有限公司是经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。事务所接受中国数据分析行业监管机构——中国商业联合会数据分析专业委员会（以下简称协会）的监管，事务所于2011年10月成为中国商业联合会数据分析专业委员会常务理事单位。

鼎盛恒信以专业的项目数据分析服务为核心，拥有数据采集、数据分析、市场预测、企业管理等各方面的复合型人才及专家，聘请大量数据分析业专家作为资深顾问，并同时依托中逸集团的强大专业团队，集中了注册会计师、注册资产评估师、注册造价工程师、注册税务师等各类相关专业技术人才。鼎盛恒信坚持用专业的分析方法、科学的分析理论、严谨的工作态度为广大客户提供包括经济信息咨询、数据处理、投资咨询、市场调查在内的各项项目数据分析服务，力求为众多的投融资和企业决策者提供具有经济性、权威性、客观性、公正性、实用性的投资数据分析报告、经营数据分析报告、决策数据分析报告等各类基于数据分析形成的分析结果及建议，作为投融资及改善管理的有力依据。

鼎盛恒信及其依托的中逸集团，已在专业咨询领域经历了10多年的执业服务，业务领域广泛，经验丰富。鼎盛恒信秉承诚信、客观、科学、实效、公正的经营理念，愿与各级政府、海内外企业、金融机构、投融资机构真诚合作，为投资决策建言献策、把控投资风险；同时，我们也愿与业界同仁相互交流，在行业协会的帮助指导下，把项目数据分析事业做大做强，推动中国项目数据分析行业的发展，力争把鼎盛恒信打造成业界最具影响力的项目数据分析公司。

联系方式：58362096 / 58362045
办公地址：北京市西城区太平桥大街丰汇时代大厦A座606室
网站主页：www.zycpa.com.cn/

河南智宸项目数据分析师事务所

河南智宸项目数据分析师事务所有限公司（以下简称“智宸事务所”）现已工商核名成功，注册资金200万元，目前协会会员资质审核正在紧张进行中，预计9月底正式成立。

智宸事务所现已汇集通信、零售、交通运输等各方精英，并聘请众多数据分析行业专家及高校教授作为资深顾问团队。团队核心人员均为研究生以上学历，拥有多家大型企业数据分析行业经验。

智宸事务所将以可靠的数据来源、专业的分析技术、公正的原则、诚信的精神，为河南的各类企业（主要方向为通信业、零售业、交通运输业）、银行金融机构、政府组织机构提供投资项目数据分析（评估、分析、规划、策划），企业经营数据分析，企业战略分析等具有经济性、权威性、客观性、公正性、实用性的项目数据分析报告及其他行业咨询类报告。

目前事务所已与多家企业、政府、高校建立了合作关系，同时已加入河南MBA&EMBA企业家联合会，蓄势待发！

智宸的理念：精准的数据视角+专业的行业经验+优质的全程服务

智宸的使命：致力于成为中原地区最大、最专业的数据分析机构，为河南的经济发展提供独立、权威、公正、诚信、优质的“数据分析”服务，同时肩负推广河南地区数据分析行业发展的重任。

联系方式：孙经理 0371-55339936 / 17737707797
办公地址：郑州市金水区农业路16号省汇中心B座2408室

云南誉诚俊安项目数据分析师事务所

云南誉诚俊安项目数据分析师事务所有限公司于2015年3月经云南省红河州弥勒市工商局核准注册，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（中数委团证第103号），事务所拥有数据分析师、注册税务师、注册会计师、注册管理会计师、工商管理等方面的专业人才及团队。

公司一直倡导：“信誉至上、诚信为本、努力不止”的经营理念；坚持“激励、信任、沟通、协作”的团队精神；以“关注客户的关注点”为服务目标，积极运用先进的数据分析技术，从而快捷、高效、严谨地为客户提供专业、全面、合理的项目数据分析报告，为企业做出最佳经营决策提供科学依据。

公司业务范围（服务种类）：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。

联系方式：鲁云萍　0873-6288450　/　18988258667

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有限公司位于湘楚腹地长沙，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
公司网址：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

上海天元项目数据分析师事务所有限公司

Shanghai Tianyuan Certified Projects Data Analyst Firm Co.,Ltd.

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）的一家专业从事项目数据分析的服务性机构。自成立以来，在协会的支持与肯定下，凭借自身的实力及不懈努力，潜心向行业优秀事务所学习，不断完善，已连续两年获得中国数据分析行业**优秀事务所**的荣誉称号！

我事务所业务涉及农业、林业、建材、化工、电子、电力、水利、煤炭、冶金、轻工、纺织、医药、机械、交通、房地产、教育、文化娱乐、环保、旅游等诸多行业领域，致力为企事业单位提供全面、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。

上海天元项目数据分析师事务所有限公司

Shanghai Tianyuan Certified Projects Data Analyst Firm Co.,Ltd.

地 址：上海市徐汇区天钥桥路327号创机商务中心

电 话：13917778657 / 021-24193019 王经理

网 址：www.shtianyuan.com

构筑数据分析大平台，树立数据分析新旗帜

——湖南翰林项目数据分析师事务所有限公司

湖南翰林项目数据分析师事务所有限公司是一家专业从事项目投资融资咨询、各类数据分析及管理咨询的专业咨询机构。事务所2013年和2014年连续两年荣获中国数据分析行业全国优秀事务所。

事务所由具有项目数据分析师、资产评估师、房地产估价师、土地估价师、注册税务师、会计师、司法鉴定人等多种执业资格的复合型专业人员申请发起，系中国数据分析行业事务所会员单位，接受中国数据分析行业监管机构——中国商业联合会数据分析专业委员会的监管。事务所位于湖南省湘中城市邵阳，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。

湖南翰林项目数据分析师事务所坚持“三年打基础、三年上台阶、三年大发展”的发展规划，努力探索适合于本所实际情况的发展路径，结合实际情况努力开拓数据分析业务的着陆点，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，建立适合于经济新常态下的数据分析盈利模式。在此基础上整合资源构建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。

办公地址：湖南省邵阳市大祥区西湖路南端市人民银行对面万基银座小区1单元402室
联 系 人：卿启伟 13187299268
办公电话：0739-5189006

银联智惠
UnionPay Smart

银联智惠大数据与互联网金融研究院

银联智惠大数据与互联网金融研究院背靠银联系统多年积累的个人消费数据、P2P数据等各成员单位的数据优势和行业经验，一方面在为政府在行业立法和监管提供专业方案与建议，为市场的规范成熟做贡献；另一方面，在引入斯坦福优质学术资源后，研究院全面开展与外部的互惠合作。从而形成基于B2B模式服务产业公司，从活动需求到咨询管理再到投资合作的不断深入的增值服务。

活动
- 培训考察
- 会议展览

→

咨询
- 名师指点
- 系统解决方案

→

投资
- 产融
- 基金

我们的目标：

研究院旨在为大数据与互联网金融业者的建立结识、交流、合作的平台家园。

我们的产品：

目前已经开发并不断投入的商业信息及社交产品包括书籍、杂志、报告等印刷物、微信公众号、线下线上沙龙、论坛、展览、考察、总裁班、专业类培训、在线教育、电视节目、俱乐部等。

近期预告：

大数据与互联网金融总裁班：
聚焦大数据与互金，激发传统企业新活力.

大数据与互联网金融专业培训：
互联网金融企业法律风险、P2P风险管控与大数据分析、互联网金融平台运营与营销推广等.

美国大数据与互联网金融高端商务考察：
探知模式创新与风控技术，回归理性与高效发展

详情请扫描二维码，获取更多行业报告、总裁班、考察等信息。

http://www.unionpaysmart.com
TL: +86 21 31216300 苏小姐

中颢润(北京)项目数据分析师事务所
Zhong Haorun (Beijing) Certified Projects Data Analyst Firm

全国量化研究专家
专注大数据落地与创新

中颢润(北京)项目数据分析师事务所

中颢润是全国最早开展量化研究的机构，领先的大数据研究模型方法、先进大数据分析智能平台，为企业规避经营风险提供大数据解决方案，为政府构建智慧城市提供量化服务支持，此外，先后通过了美国、英国、法国、德国等大使馆认证及涉外调查等多项资质，获得多家国际500强企业认可。

随着大数据时代的快速发展，通过数据进行科学分析已经成为全社会的共识，中颢润正在以专业的服务精神、先进的大数据分析理论体系和优质的服务过程，逐步成为大数据时代下的量化研究专家，成为中国数据分析行业的领跑者！

专业
Professional

中颢润(北京)项目数据分析师事务所
Zhong Haorun (Beijing) Certified Projects Data Analyst Firm

Tel : 010-65188088-106 \ 13001995337 \ 13651113534
Add : 北京市建国门内大街七号光华长安大厦三座二层(100005)
Web : www.chinacpda.cn