



# 数据分析

CHINA DATA ANALYSIS 数据分析·因你而不凡



《中国数据分析》行业特刊  
2020年第03期 总第43期（季刊）  
咨询热线：**400-050-6600**  
<http://www.chinacpda.org/>  
投稿邮箱至 [xiehui@chinacpda.org](mailto:xiehui@chinacpda.org)

China  
Data  
Analysis



中国商业联合会数据分析专业委员会 主办

## 乘风破浪，唯奋勇者强；危中有机，唯创新者胜！

中流击水，浪遏飞舟。在决胜全面小康、决战脱贫攻坚的关键时刻，突如其来的新冠肺炎疫情加大世界经济衰退下行压力，可谓艰难。但惟其艰难，才更显勇毅。而随着数据要素市场在逆势中释放出经济增长新动能，大数据也迎来了重大发展契机，当前数据作为新型生产要素，已成为信息时代国家重要战略资源。近年来，我国大数据从无到有、从有到优，大数据应用加速向各行各业渗透，市场主体愈发重视数据这一重要企业资产和资源所蕴含的巨大商业价值。同时大数据与各行业融合发展，助推数字化经济转型，实现企业级数据智能驱动业务，成为驱动发展的强大引擎。

发展是第一要务，人才是第一资源，创新是第一动力。“功以才成，业由才广。”夯实大数据人才基础，为数据发展注入人才，激发人才创新活力和潜力是重任。得人才则兴，失人才则衰。站在历史新起点，中国数据分析行业才能更加蓬勃发展。

近二十年来，协会在培养大数据人才的道路上深耕细作，为了更好地完善人才发展，不断推进素质教育，创新教育方法，提高人才培养质量，努力形成有利于创新人才成长的育人环境。百余位具有一线实战经验的老师组成的科研团队，带着助力中国数据分析人才更专业、更优秀的信念，结合自身实战经验不断探索，综合各行业对数据人才的需求，制定出一整套科学的人才培养体系，时刻以培养具有大数据思维的实战型人才为己任，在十几年间为各行各业培养了数万名优秀的数据分析师专业技术人才。

“大疫当前，百业艰难，但危中有机，唯创新者胜。”在不久前的企业家座谈会上，习近平总书记勉励企业家要做创新发展的探索者、组织者、引领者，勇于推动生产组织创新、技术创新、市场创新，重视技术研发和人力资本投入，有效调动员工创造力，努力把企业打造成为强大的创新主体，在困境中实现凤凰涅槃、浴火重生。在危机中育新机、于变局中开新局，“富有之谓大业，日新之谓盛德。”对于行业而言，行业本身就是“大业”，而维持发展“大业”，不能固步自封而要不断创新，这种创新意识和创新能力正是一个行业的“盛德”。

疫情期间，加速了线下教育的在线化，“线上+线下”OMO教育模式将成为大势所趋。协会“双师教学”模式以OMO模式将线上和线下优质师资资源进行整合；“自主选课”模式也从统一灌输转向了因材施教；充满黑科技的智慧教室采用多形式的互动教学方法，成功地开展个性化、探索式教学，使课堂教学有机联动。

创新不仅是实现高质量发展需要弘扬的品质，同样是应对疫情冲击、走出经营困境的不二法门。近期，由协会牵头，结合目前国内外大数据发展现状和市场需求，调研了近百家公司数据分析师事务所和大数据公司，联合全国十余位大数据专家、优秀数据分析师事务所共同对事务所从业、分析师执业、数据分析业务及费用参考规范进行研讨。行业规范的同时，更能稳定市场秩序和社会效益，同时帮助相关企业在在危机中寻求创新，困境中寻找出路，逆势生长。

在疫情防控常态化的今天，加快推进生产生活秩序全面恢复，面对危机和困难，不言放弃，寻求机遇和创新。无论企业还是个人，不负时光，不负心中炽热，奋勇向前，所向披靡，梦想一定不会被辜负！

乘风破浪，唯奋勇者强；危中有机，唯创新者胜！

中国商业联合会数据分析专业委员会





## 本期目录 CONTENTS

### 卷首语

- 01 乘风破浪，唯奋勇者强；危中有机，唯创新者胜！

### 协会动态

- 05 从底层逻辑探索Python的强大之处  
06 互联网时代下的大数据分析之道  
08 疫情下大数据应用与未来发展趋势  
09 如何利用Social Listening洞察商业情报  
11 学霸亲授 考试通关秘笈  
12 数据分析师事务所被冒名，行业发展任重道远

### 政策导向

- 13 大数据时代如何善用智能设备保护用户隐私  
15 工信部：2020年底建成国家工业互联网大数据中心  
16 山东出台健康医疗大数据管理办法 10月1日起施行  
17 山西省加快大数据融合应用  
18 提高大数据时代社会治理智能化水平

### 行业动态

- 19 COVID-19之后的数字银行发展历程  
20 在制造业中实施大数据战略时要考虑的挑战  
21 10个步骤巧用数据分析法 轻松摸透竞争对手  
24 2020年中国大数据安全市场现状分析：市场规模将超1万亿  
25 阿里发布《2020农产品电商报告》：  
去年全平台农产品交易额2000亿 数字农业将成风口  
26 从综艺IP到私域IP，获取别人的“品牌重塑”经验  
28 数据分析师：指标体系搭建四原则

### 学"数"交流

- 32 数据分析应用实战：阳台能晒多长时间太阳？  
38 SPSS单样本非参数检验——卡方检验  
40 机器学习在传统服装行业的尝试  
42 数字经济时代科技创新浅谈  
44 浅谈医改下，新冠疫情时，运营数据对提升药企应对的意义

### 事务所专栏

- 46 数据分析师事务所从业规范  
48 数据分析师执业规范  
49 用大数据思维助力数字化转型的全新服务行业



### 主办单位

中国商业联合会数据分析专业委员会

### 编委成员

李苗苗

### 出版时间

2020年09月出版 总第43期

### 美工设计

崔峻珩

### 联系我们

中国商业联合会数据分析专业委员会  
地址: 北京市朝阳区朝外SOHO-C座9层  
电话: 400-050-6600 / 010-59000991  
传真: 010-59000991转 607

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

感谢您对《中国数据分析》的支持！

投稿邮箱: xiehui@chinacpda.org

官方网址: www.chinacpda.org

## / 6.24 CPDA数据分析师沙龙精彩回顾 /

### ——从底层逻辑探索Python的强大之处 /

来源 / 中国商业联合会数据分析专业委员会 编辑 / 协会会员处 李苗苗 日期 / 2020-06



6月24日，我们为全国的CPDA数据分析师带来了一场精彩的主题为《如何快速入门Python》的线上沙龙活动。

本次活动特邀中国商业联合会数据分析专业委员会科研中心课程设计师、特聘讲师张磊老师，带领大家结合实际业务问题，从底层逻辑探索Python的强大之处。活动吸引了各行业、各领域的近400名CPDA数据分析师。

在1个多小时的活动中，小伙伴们积极参与到了各环节的互动中，与张磊老师在线上进行互动交流、深入探讨学习。让想快速入门Python的同学们，体验了一把“引爆思维”的满足，现在就让我们一起来回顾吧！

#### 线上沙龙回顾一：思维层面带你构建Python的核心知识体系



近年来，机器学习、数据分析在各行各业的应用越来越广泛，人们也越来越能感受到它们的强大威力和巨大的潜力。

CPDA数据分析师通过系统的课程学习，已充分掌握了数据分析的完整流程，并能通过Datahoop算法平台实现各个模块的便捷操作，以最短的时间看到模型的实现结果。然而，很多同学也想进一步了解，如何通过“高深莫测”的Python代码实现这些算法建模？这正是本次沙龙活动的初衷所在。

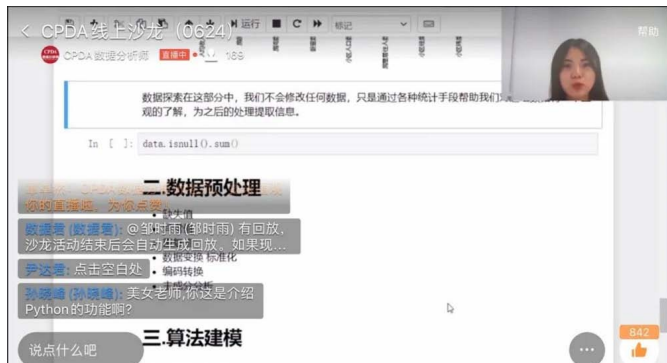


#### 线上沙龙回顾二：人生苦短，我用Python

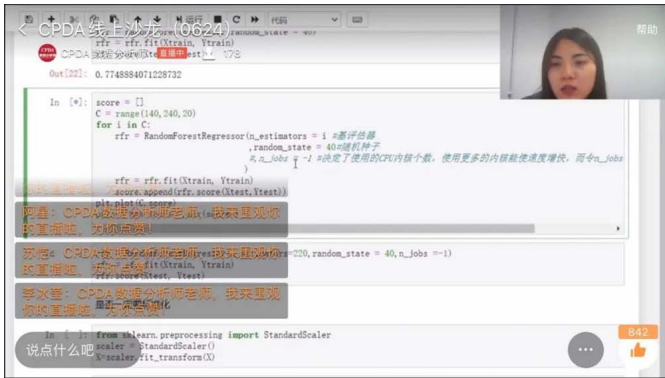


紧接着，张磊老师采用娓娓道来的方式，从5个方面，剖析Python为何流行的原因，并带领大家一同感受Python的独特魅力。同时她也强调，除了高效简洁易读这些天生的优势，对于数据分析它有众多优秀的第三方库供大家使用，不仅能高效处理大数据，能三行代码实现算法建模，对新手来说也十分友好，并没有我们想象的那么复杂。

#### 线上沙龙回顾三：基于多元线性回归算法进行房价预测分析

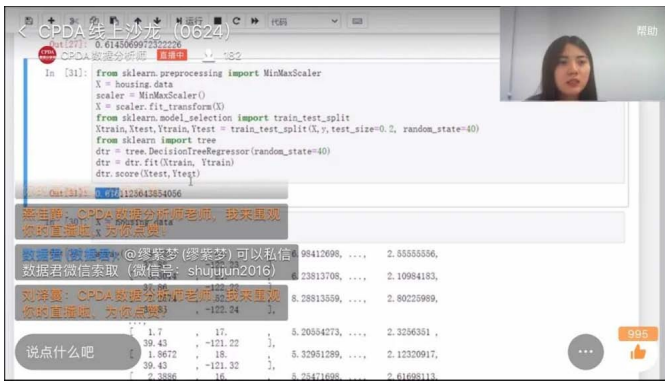






张磊老师带领大家进入Python房价预测实战的演练，通过线性回归、决策树、随机森林等监督算法回归预测加利福尼亚房价，以及如何进行调整参数使模型的效果更好，最终经过模型比较选择了随机森林算法。

最后，张磊老师就大家关心的个性化问题进行了现场答疑，小伙伴们纷纷表示本次沙龙活动不仅有丰富的理论干货，还有关于Python的实战案例经验分享，让CPDA数据分析师们大呼过瘾，收获颇丰！



## / 7.3 CPDA数据分析师沙龙精彩回顾 ——互联网时代下的大数据分析之道 /

来源 / 中国商业联合会数据分析专业委员会 编辑 / 协会会员处 李苗苗 日期 / 2020-07



7月10日，我们为国内CPDA数据分析师带来了一场精彩的主题为《数据分析如何在游戏行业落地应用》的线上沙龙活动。本次活动特邀国内顶尖游戏出海公司数据总监、CPDA数据分析师兄张炳出老师为广大CPDA数据分析师们分享，如何更精准地将数据洞察和分析方法以简单的方式呈现给决策者的干货内容，并为企业提供行之有效的决策性建议。活动吸引了各行业、各领域的300余名CPDA数据分析师。

在近2个小时的活动中，小伙伴们积极参与到了各环节的互动中，张炳出老师将其十余年的从业经验倾囊相授，并在线上与学员们进行互动交流。通过真实的数据化工作场景，让想了解数据分析业务基础及分析思路的同学们，体验了一场“烧脑”的实战盛宴，现在就让我们一起来回顾吧！

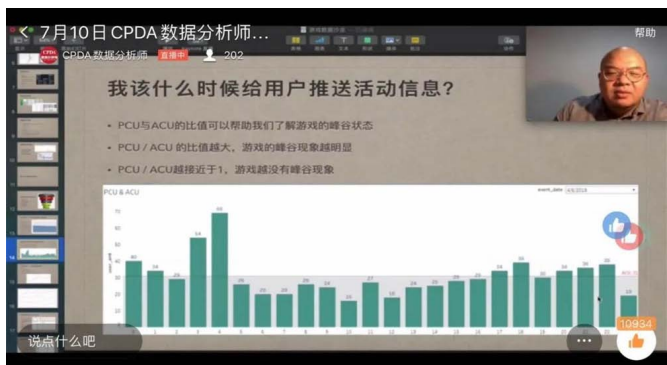
### 线上沙龙回顾一：游戏数据的采集、处理与存储





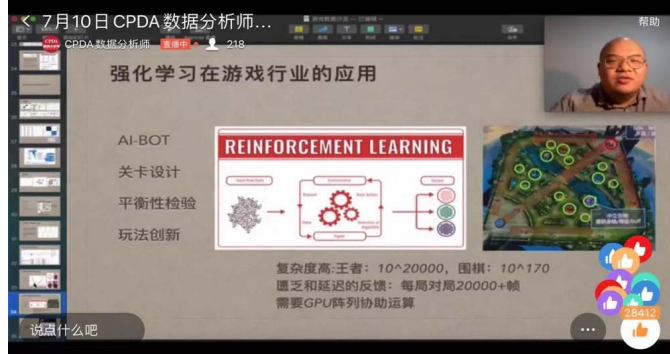
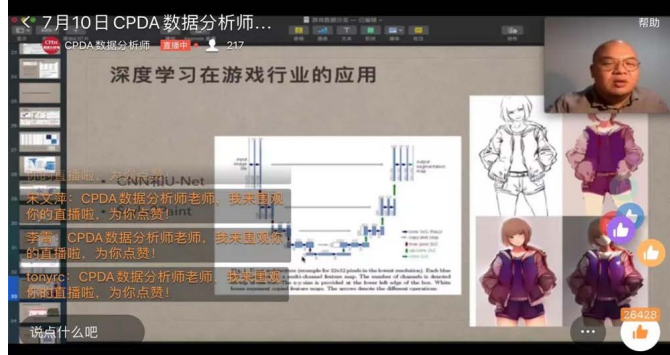
沙龙活动一开始，张炳出老师就为大家讲解了游戏行业的数据来源、采集方法、SDK和埋点的概念以及经验。并强调，游戏数据有非常明显的大数据特征，随即又为我们介绍了如何高效、合理、完备地处理数据的经验。

线上沙龙回顾二：游戏行业数据分析的方法和经验



紧接着，张炳出老师为我们讲解如何更加有效及深入地分析基本业务指标，并通过指标的组合发现新问题和一些指标需要规避的陷阱。通过介绍关于数据驱动运营的真实案例（如发现产品数值设计不均衡、玩家成长路线不合理等），向我们阐述了用高维度数据向低维度面板投射产生综合性更强的图表的BI理念；以及经典分析方法的业务运用，如从价格需求弹性系数设计ARPU弹性、RFM模型做精准客服、KANO模型判断功能排期等。

线上沙龙回顾三：数据分析如何在游戏行业落地应用



最后，张炳出老师为大家讲解了在CPDA课程中讲解过的各种机器学习算法在游戏行业业务中的落地应用。如逻辑回归发现游戏设计上的缺陷、聚类算法进行用户分群精准营销、NLP应用到游戏舆情监控、卷积神经网络做线图上色、对抗网络制作AI BOT等。



# / 7.24 CPDA数据分析师沙龙回顾 ——疫情下大数据应用与未来发展趋势 /

来源 / 中国商业联合会数据分析专业委员会 编辑 / 协会会员处 李苗苗 日期 / 2020-07



7月24日，我们为全中国CPDA数据分析师带来了一场精彩的主题为《如何用Python文本挖掘解析疫情数据》的线上沙龙活动。

本次活动特邀高松老师为广大CPDA数据分析师们进行专业分享。本期活动的最大亮点是基于真实数据信息，为大家详细展示Python编程方式是如何对大文本数据实现高效分析与挖掘。同时，为大家分享了老师在教学工作中的宝藏经验。活动吸引了数百位各行业、各领域的CPDA数据分析师。

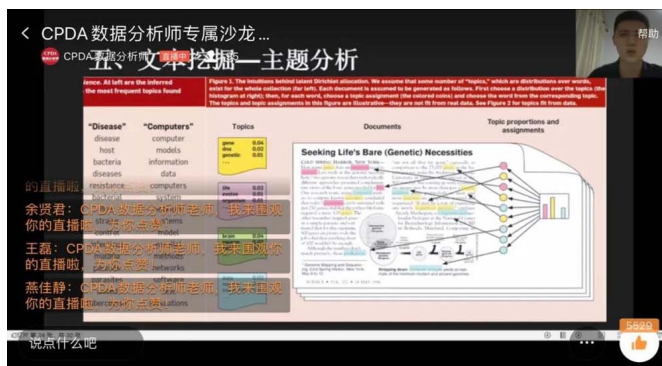
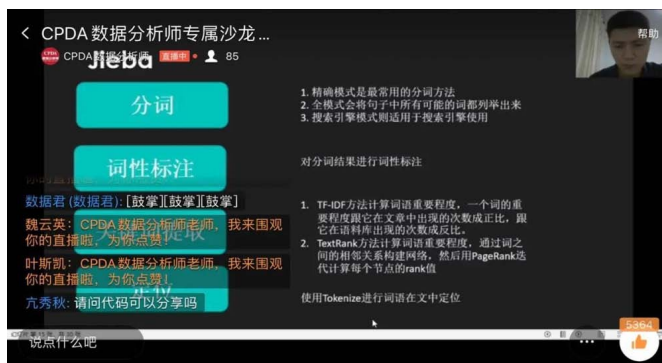
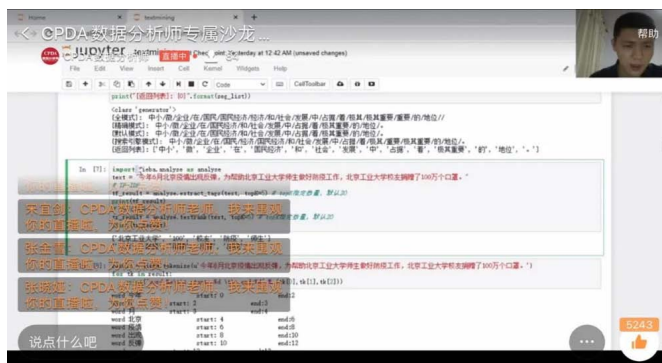
高松老师将其多年的从业经验倾囊相授，并在线上与学员们进行互动交流。通过真实的数据化工作场景，让了解数据分析业务基础及分析思路的同学们，体验了一场“真实”的实战盛宴，现在就让我们一起来回顾吧！

## 回顾一：自然语言处理技术阐述



活动开始，高松老师通过一些典型文本案例，对文本挖掘的底层技术进行分步介绍，包括基础的原理介绍、深层次理解分析、基础应用、高级应用以及便捷算法等。

## 回顾二：Python编程方式处理大文本数据展示



紧接着，高松老师为我们循序渐进的讲解在处理文本挖掘工作中的基本流程，文本挖掘与数据挖掘工作的不同之处，数据爬虫编码的推导过程，以及作实际过程中的一些需要规避的雷区等。

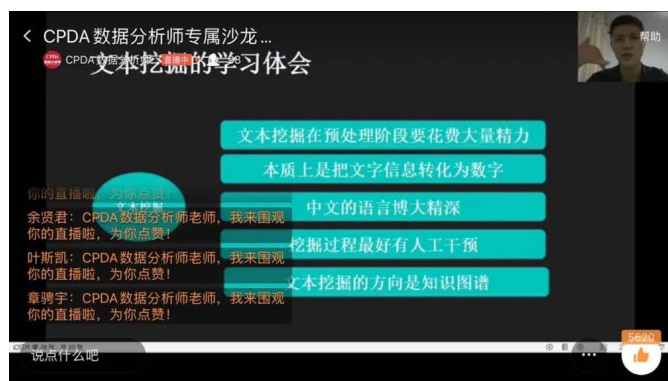
基于对自然语言处理技术的认知，高松老师带领大家执



行一个千位级别的关于新冠疫情新闻文本处理的完整工作过程。通过学习这一案例指引大家更大批量的文本信息工作，向大家分享了十多个文本预处理工具及优势、Jieba工具的精巧使用、Pyecharts数据可视化应用、基于TF-IDF计算每个词语的权重值、文本主题分析和工具使用等。

老师表示，Python编程方式运用得当，将会大大缩短我们日处理文本的工作量，以较高的工作效率获得更有价值的指导性意见。

### 回顾三：文本挖掘之私囊倾受



最后，高松老师为大家分享了其在数据分析工作中的多年经验积累，讲解了文本挖掘在预处理阶段的付出与注意事项、人工干预对文本挖掘过程的影响、未来文本挖掘的重点方向，文本挖掘的学习讲义等等。

在本期沙龙活动的答疑环节里，高松老师就大家关心的个性化问题进行了现场答疑，小伙伴们纷纷表示本次沙龙活动不仅有丰富的理论干货，还有关于个人教学工作上的资源分享，让CPDA数据分析师们受益匪浅！

## / 8.14 CPDA数据分析师沙龙回顾 ——如何利用Social Listening洞察商业情报 /

来源 / 中国商业联合会数据分析专业委员会 编辑 / 协会会员处 李苗苗 日期 / 2020-08



8月14日，我们为全国CPDA数据分析师会员们带来了一场主题为《如何利用Social Listening洞察商业情报》的专属沙龙活动。

本次活动邀请到新新浪微热点数据pm，达观数据VOC产品负责人高长宽老师为大家做互动分享。本期活动的亮点对生活中息息相关且信息数量庞大的社会化媒体展开详细的数据分析推导，通过层层剖析与若干应用技术的分享，将工作中实际遇到的竞品分析、市场概览、售前支持、客户服务于关系维

系、购买信号等问题拉到沙龙互动活动中，逐一攻破。同时，为大家分享了老师在实践工作中的丰富经验。活动深受CPDA数据分析师学员的肯定与好评，纷纷表示受益良多。

沙龙活动持续一个多小时，小伙伴们积极参与，气氛活跃。高长宽老师将其多年的从业经验倾囊相授，并在线与学员们进行互动交流。现在就让我们一起来回顾吧！

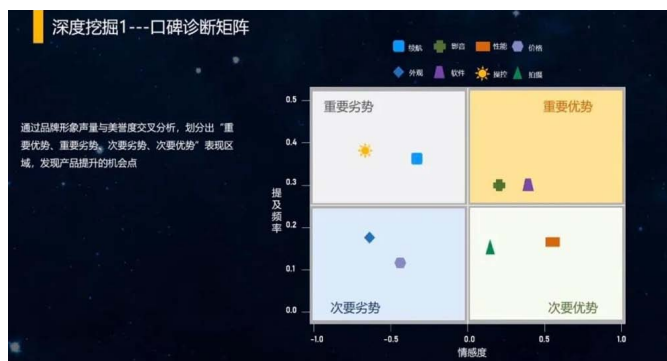
### 01. 什么是社会化媒体，怎样挖掘重要信息。





活动开始，高长宽老师为大家详细介绍社会化媒体的定义、特征、类别、主客观信息（动机），包括从海量且庞杂的社会化媒体数据中心抽取各类主客观信息，再将其萃取成有价值的商业情报的三个步骤：数据→信息→情报。其中，信息加工包括数据清洗、检索，挖掘，业务知识包括信息甄别，信息的相关性判断等。

## 02. 什么是社会化聆听, 常见分析维度&应用场景有哪些。



接下来，高老师为大家进一步讲解社会化聆听的定义内容以常见分析维度、深度挖掘、应用场景。

社会化聆听的常见分析维度：信息来源分布，关键词、用户画像、口碑热词（正负观点词云，促进改善产品）、KOL识别、全网声量走势、关联词分析、典型意见挖掘、情感分析等。

在深度挖掘发放中，老师特别提到口碑诊断矩阵和商业模型波士顿矩阵。通过品牌形象声量与美誉度交叉分析，划分出“重要优势、重要劣势、次要优势”表现区域，发现产品提升的机会点。

## 03. 社会化聆听设计的NLP技术



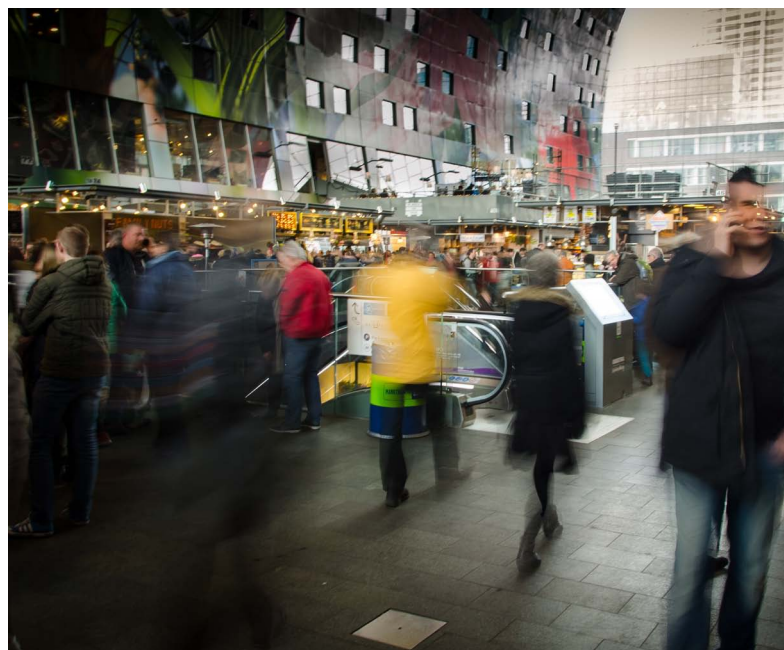
社会化聆听的数据分析工作是丰富的，涉及的NLP技术也较多。高长宽老师为大家讲解了四类NLP技术，包括观点抽取、意图识别、语义检索等。自然语言的等级从初级到高级，分为字词级分析，句子级分析，篇章级分析。老师特别提醒道，观点抽取（Opinions Extraction）是社会化聆听很重要的手段，并对此做了进一步介绍。

## 04. 社会化聆听在旅游行业的应用



最后，高长宽老师以客群划分举例说明社会化聆听在旅游行业的应用，利用社会化聆听从外部数据中心获取市场调查步骤，以及可应用的较为重要的情感模型分享。老师表示，抓住消费者的情感，才能抓住了他们的心，最后为工作服务。

在本期沙龙活动的答疑环节里，高长宽老师就大家关心的个性化问题进行了现场答疑，小伙伴们纷纷表示本次沙龙活动不仅有丰富的理论干货，还有关于个人教学工作上的资源分享，让CPDA数据分析师们受益匪浅！





# / 8.28 CPDA数据分析师沙龙回顾 /

## ——学霸亲授 考试通关秘笈 /

来源 / 中国商业联合会数据分析专业委员会 编辑 / 协会会员处 李苗苗 日期 / 2020-08



8月28日，我们为全国的CPDA数据分析师会员们带来了一场主题为《学霸亲授 考试通关秘笈》的沙龙活动。活动邀请到CPDA数据分析师全国第二季度统考排名第2位的优秀学员罗海涵与大家做考试经验交流。

由于正值CPDA数据分析师第三季度全国统考前夕，很多同学都是带着问题来参加的。

互动中的问题虽然很多，但是主要集中在投资分析、相关系数、科目二和科目三的考题区别、工具和模型的使用、模拟练习与操作等。师兄罗海涵做完主题分享后，对大家的问题逐一作了交流分享，也将其个人的成功经验，学习方法传授给大家。

对于广大待考的同学们来说，本次活动犹如及时雨一样，解答了他们心中的一些困惑与疑问。很多同学表示，这样与高手过招的机会很难得，希望以后多多举办。现在就让我们一起来回顾吧！

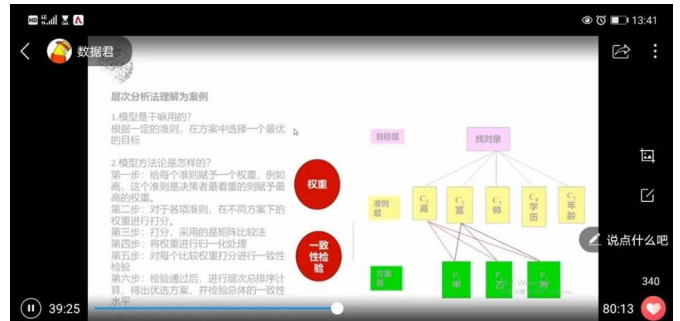
### 01、专注专业&高度重视



活动开始，学霸师兄罗海涵将自己报名学习CPDA数据分析师的心得、目的以及制定的学习规划与大家分享。

尤其是在分享到协调工作、学习和家庭的时间问题上，师兄更是提出了自己的建议，并提醒大家一定要把学习和考试重视起来。

### 02、深度理解，学会沉淀



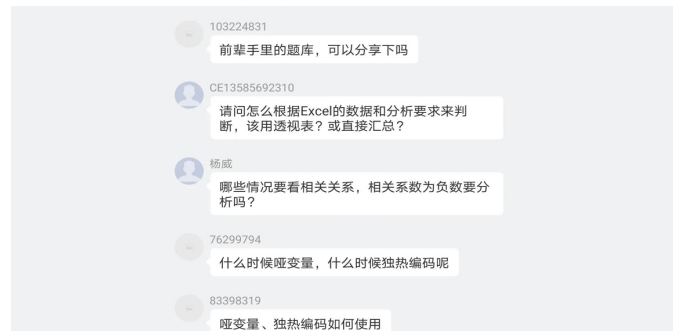
在进一步分享如何系统化学习的时候，学霸师兄分享了自己的思维导图。思维导图中将理论知识、主要算法，和模型应用全部囊括在内。师兄表示，学习的过程中，一定要把学到的知识沉淀下来，学以致用，才能融会贯通。在寻找最优学习模型中，师兄还将自己的工作方法进行了分享讲解。

### 03、融会贯通之模型与案例分享



为便于理解与沟通，学霸师兄以巴斯模型和层次分析法为例，为大家详细说明两大模型的应用、方法论、模型公式、以及方法步骤。生动活泼的沟通方式，让大家更好的消化与理解。

### 04、工作经验分享&有序备考





学霸师兄是从事移动通信行业的。匹配的岗位和行业特性，让师兄的专业知识得到更好的发挥。师兄分享了在工作中从特征筛选法，走向机器学习建模的8个思考，精准营销的相关性分析，以及提升业务分析的6个数据挖掘方向。同时也表示，要掌握数据处理的能力，认真学习EXCEL、SQL、PYTHON等工具。

对于大家非常关注的备考问题，师兄给出了时间规划建

议和练习指南。建议参加考试的同学们至少提前一个月开始准备，合理规划练习时间。练习方面，一定要注重模拟练习题的反复温习和实操。

在本期沙龙活动的答疑环节里，师兄罗海涵就大家关心的个性化问题进行了现场互动交流。小伙伴们纷纷感谢师兄的分享，表示受益匪浅！

## / 数据分析师事务所被冒名，行业发展任重道远 /

来源 / 中国商业联合会数据分析专业委员会 编辑 / 协会会员处 李苗苗 日期 / 2020-08

近日，中国商业联合会数据分析专业委员会在其官网发布《关于部分单位伪造中国商业联合会数据分析专业委员会颁发的执业资质的声明》，声明中提到专业委员会多次收到企业及个人反映，一些单位伪造数据分析师事务所身份对外发布不具法律效力的“会员执业资质证书”，这一行为严重损害了中国商业联合会数据分析专业委员会的行业信誉之余，扰乱了行业的良好秩序，并且很大程度上侵害了其他个人或团体的合法权益，有的甚至产生利益纠纷。

我国的数据分析行业从无到有，直至今天不断发展壮大，是由国家相关单位、全国性行业组织的积极肯定与大力扶持，业内专家以及众多从业者倾力付出的结果。据了解，中国商业联合会数据分析专业委员会自2005年在陕西成立全国第一家数据分析师事务所起，也曾出现过查询真伪《会员执业资质证书》的不良事件，但是今年上半年情况尤为严重。据统计，2020年上半年中国商业联合会数据分析专业委员会陆续收到类似投诉、核查真伪资质证书的事件已达百余起。如此激增的不良事件，引起专业委员会的高度重视。

### 中国商业联合会数据分析专业委员会

中商联数据分拆委[2020]05号

#### 关于部分单位伪造中国商业联合会数据分析专业委员会颁发的执业资质的声明

近期，我会收到众多企业及个人反映，某些单位伪造我会颁发的《会员执业资质证书》承接业务、欺骗客户。针对此情况，我会调查核实后声明如下：

1.中国商业联合会数据分析专业委员会是国务院国有资产监督管理委员会、中华人民共和国民政部批准设立的专业委员会，指导、促进数据分析行业的行业自律，对符合要求的数据分析师事务所，授予数据分析行业《会员执业资质证书》，相关资质可在中国数据分析行业网站（www.chinacpda.org）查询。

2.凡未经过我会审批颁发资质证书且在中国数据分析行业网站无资质信息的单位，其所持我会颁发的《会员执业资质证书》均为伪造。

3.对伪造我会颁发执业资质的单位，一经我会查证核实，将严厉追究其法律责任。

4.作为中国数据分析行业的专业协会组织，我会呼吁所有行业内单位严格遵守行业规定，遵循诚实信用原则，维护正常、健康的市场秩序。

我会将加强行业监督力度，欢迎并鼓励社会公众监督，监督举报电话：400-050-6600。

特此声明。

中国商业联合会数据分析专业委员会  
2020年7月29日



中国商业联合会数据分析专业委员会表示，此类以商业利益为目的的冒名伪造行为屡见不鲜，一方面是看到专业委员会的行业影响力及地位，另一方面是觊觎数据分析行业《会员执业资质证书》的含金量。他们的欺诈行为主要以颁发伪造资质证书、承接业务，以及侵害客户的个人权益为主。

对此，专业委员会提醒广大业内人士以及关注数据分析领域的有志青年，在选择数据分析师事务所或是希望从事相关业务时，一定要擦亮双眼谨慎选择。

一旦遇到对方自称拥有中国商业联合会数据分析专业委员会认可的数据分析师事务所从业资质时，可登录中国数据分析行业网站（www.chinacpda.org）快速辨别真伪，从而保护自己的合法权益。

冒名欺诈层出不穷，选择数据分析师事务所切勿盲目选择！



以“人才培养和数字化转型”为服务核心，推动国内大数据产业全面发展。

随着互联网的快速普及，信息技术和人类生产生活交汇融合，全球大数据技术产业迎来了爆发式的增长需求。上至国家建设数字中国的战略要求，下至企业数字化转型的迫切需

求，无不透露出大数据风口的来临。

综合国内外大数据环境发展的趋势，也可以从侧面折射出数据分析师事务所频繁遇到冒名欺诈的一些原因。然而当前正是我国大数据应用技术腾飞崛起的加速阶段，亟需有资质有专业且成熟的领航者。

作为中国数据分析服务领域的领航者，中国商业联合会数据分析专业委员会推动了数据分析技术的普及和应用，培养了数万名专业的数据分析师人才，在全国各省市、自治区的成立了近百所CPDA数据分析师事务所，促进了中国数据分析行业健康发展，践行以“大数据思维”助力政务数字化发展、企业数字化转型、教育数字化应用，围绕数据的深度分析、业务场景构建、深层次的咨询等，以大数据思维帮助企、事业单位实现数字化转型并提供行之有效的战略决策。

## / 大数据时代如何善用智能设备保护用户隐私 /

来源 / 第一财经 编辑 / 协会会员处 李苗苗 日期 / 2020-08

今年4月，我国出台《网络安全审查办法》，自6月1日起实施。这表明我国在网络日益发展的今天开始逐步关注用户隐私与网络安全等方面。欧洲专利局的数据显示，到2025年，物联网设备的数量将达到750亿。随着智能设备的日益普及，公共交通、家用电器、智能手机、智能制造等领域将产生大量的数据。在未来，共享智能设备产生的数据将会成为社会进步和经济发展的核心驱动力。然而，对于如何共享数据、数据所

有权状态，以及如何保护数据，我们知之甚少。本团队通过走访不同行业的多家企业得出如下观点。

### 1. 务实是数据收集的关键。

德勤2019年发布的中国智能制造分析报告指出，大部分企业正致力于数据纵向集成，其中81%的受访企业已完成计算机化阶段，而63%的企业已部署了相应的数字化工厂。然而，





仅有33%的企业其智能制造利润贡献率超过50%。

我们走访中同样发现了这一问题，一些企业只是一味地储存数据而无法将数据合理利用，这只会占用企业的人力物力资源，而很难为企业带来盈利。一位接受访问的云服务提供商建议：“一些不必要的数据可以直接删除，有用的数据存储在本地，重要的数据要上传到云服务器。”

## 2. 隐私是保持不败的基础。

近年来，企业对用户数据保护不当造成的隐私泄露事件，并不罕见。2017年，支付宝年度账单首页因默认勾选同意《芝麻服务协议》而被指泄露隐私。此外，南都个人信息保护研究中心发布的《2017个人信息保护年度报告》显示，在对1500多个手机应用与网站的隐私政策测评中，八成以上的平台隐私政策透明度低，且普遍存在暗藏格式条款等弊病。

首先，在平衡合作与隐私的前提下，建立数据管理方案显得尤为重要。“我们不会与其他公司共享用户数据。”一位智能家电公司的高管表示。事实上，一些智能算法可以利用公司间共享的数据来推断竞争对手的运营情报。

其次，从保护用户数据的角度来看，用户数据的泄露会给企业形象和用户评价带来严重的影响，并可能面临法律风险。2017年3月，京东与腾讯的安全团队联手协助公安部破获的一起特大窃取贩卖公民个人信息案，其主要犯罪嫌疑人乃京东内部员工，其在网络黑市贩卖盗取的涉及交通、物流、医疗、社交、银行等个人信息50亿条。

再次，上下游企业之间的数据泄露可能会产生严重的后果。波耐蒙安全研究公司在2017年的调查中发现，有56%的企业曾经遭遇过与供应商有关的数据泄露事件。平均有470家外部公司可以访问公司的敏感信息。这些数据泄露为企业带来大量损失，也降低企业的声誉，损害消费者的权益。因此，发现共享过程中数据泄露的原因，从源头上杜绝数据泄露，对企业的发展至关重要。

因此，如何平衡数据共享和数据保密成为企业保持竞争优势的重要问题。我们建议企业：建立数据共享和保护的集体信息管理标准和技术协议；评估数据保密等级，平衡数据共享与保密；预测信息泄露风险。

## 3. 公共政策是控制的调节者。

各国政府都意识到了数据保护的重要性。虽然信息技术在不断发展，但仍存在产业发展滞后、网络安全体系有待进一步完善等问题。例如区块链技术作为一种高度安全的信息共享技术，也存在一系列的安全风险，这些风险体现在它的底层代码、加密算法、共识机制、智能合约、数字钱包的安全性等方面。还要结合公共政策及应对突发公共事件的新情况，灵活、机动地配合做好相应的安排。例如，我国应对当下突发的新冠肺炎疫情，就很好地发展和实践了数据保护与数据共享相结合的底层逻辑，为人类社会之共同抗疫，贡献了中国实践和经

验。

2016年11月7日，中国颁布《网络安全法》，自2017年6月1日起施行。该法的宗旨是保障网络安全，维护网络空间主权、国家安全和社会公共利益，保护公民、法人和其他组织的合法权益，促进经济社会信息化发展。该法主要涉及网络安全支持与推广、网络运行安全、关键信息基础设施运行安全等内容。

现行的法律法规在一定程度上可以防止数据泄露。然而，数据在结构、内容方面具有高度多样性的特点。例如，用户个人数据、手机通话数据、社交媒体数据、传感器数据等数据在结构和内容方面都非常不同，这使得政府部门难以针对所有的用户数据进行统一的规范化管理，也因此导致当前的法律对数据的保护具有一定的局限性。

第一，法律法规对所有权的保护主要集中在成果、财产等方面。第二，泄露数据的扩散速度快，不易追踪。第三，数据容易远距离转移，转移活动可能发生在不同的地区或国家。

随着互联网法律碎片化的发展，跨区域、跨国家的维权司法体系日趋复杂。温州通领集团与美国莱伏顿公司和同行业竞争对手开展了长达六年的知识产权诉讼，终于在美国获得“五连胜”。

在这样的环境下，我们建议政府机构优化一个公共可访问的匿名数据池，该数据池可以提供的数据包括各市县和地区的产业发展情况、能源消耗、教育程度、健康指数、气象信息、交通信息等。这些数据将为企业和公共机构提供支持，以全方位进一步赋能我国经济竞争力，并促进人工智能、云计算和其他许多先进技术在我国之稳步发展和应用，并向世界贡献中国智慧。





## / 工信部：2020年底建成国家工业互联网大数据中心 /

来源 / 中国商业联合会数据分析专业委员会 编辑 / 协会会员处 李苗苗 日期 / 2020-08



The screenshot shows the official website of the Ministry of Industry and Information Technology (MIIT) of the People's Republic of China. The page displays a notice titled "关于印发《工业互联网专项工作组2020年工作计划》的通知" (Notice on Issuing the Work Plan of the Industrial Internet Special Working Group for 2020). The notice is dated July 10, 2020, and is issued by the Industrial Internet Special Working Group Office. The document number is MIIT Letter (2020) 153. The notice outlines the plan to build a national industrial internet big data center by December 2020. The website header includes navigation menus for news, public information, and services, along with a search bar and social media links.

7月10日，工信部在网站发布《工业互联网专项工作组2020年工作计划》。《计划》提出，在2020年12月建成国家工业互联网大数据中心。

#### 工信部工业互联网专项通知

《计划》提出共54项具体举措，分为十大任务类别，包括：提升基础设施能力、构建标识解析体系、建设工业互联网平台、突破核心技术标准、培育新模式新业态、促进产业生态融通发展、增强安全保障水平、推进开放合作、加强统筹推进、推动政策落地。

#### “提升基础设施能力”类任务含五部分重点工作：

(一) 升级建设工业互联网外网络。推动基础电信企业建设覆盖全国所有地市的高质量外网。打造20个企业工业互联网外网优秀服务案例。

(二) 支持工业企业建设改造工业互联网内网络。鼓励工业企业升级改造工业互联网内网。鼓励工业企业与基础电信企业合作，利用5G改造工业互联网内网。

(三) 推动工业互联网IPv6应用部署。开展IPv6在交通行业推广应用的研究，推进IPv6在应急管理领域推广应用的研究。

(四) 加大无线电频谱等关键资源保障力度。开展5G工业互联网专网频率使用兼容性试验，规范物联网、工业互联网频率使用。

(五) 建设工业互联网大数据中心。由工业和信息化部（信发司）牵头，建成国家工业互联网大数据中心，引导各地建设一批工业互联网大数据分中心，完成时间为2020年12月。

工信部还提出建设工业互联网平台，2020年将推动40万家企业上云上平台，应用工业互联网平台开展研发设计、生产制造、运营管理等服务。

此外，工信部计划研究制定《工业互联网创新发展行动计划（2021-2025年）》，并于2020年12月出台。

## / 山东出台健康医疗大数据管理办法 10月1日起施行 /

来源 / 山东政府网 编辑 / 协会会员处 李苗苗 日期 / 2020-08

8月25日，山东省政府官网发布山东省人民政府第335号令，《山东省健康医疗大数据管理办法》(以下简称《办法》)经省政府常务会议通过，现予公布，自2020年10月1日起施行。

《办法》明确，健康医疗大数据，是指在疾病防治、健康管理等过程中产生的，以容量大、类型多、存取速度快、应用价值高为主要特征的健康医疗数据集合，以及对其开发应用形成的新技术、新业态。健康医疗大数据资源目录由基础信息、公共卫生、计划生育、医疗服务、医疗保障、药品管理、新型业态等组成。

根据《办法》，公民、法人和其他组织可以通过健康医疗大数据平台获取无条件开放数据。除了涉及商业秘密、个人隐私或者依据法律、法规规定不得开放的健康医疗数据以及有条件开放数据外，其他都属于无条件开放数据。同时，政府举办的医疗卫生机构和国有健康服务企业，应当依法将其提供服务产生的健康医疗相关数据汇聚到健康医疗大数据平台。

另外，《办法》还提出构建“互联网+医疗健康”服务模式，推进网上健康咨询、预约诊疗、检查检验结果互认和临床研究数据整合共享，发挥优质医疗资源引领作用，推动覆盖全人口全生命周期的预防、治疗、康复和健康管理一体化电子健康服务。鼓励支持健康医疗业务与大数据技术深度融合，构建健康医疗大数据产业链，与健康养老、家政服务、餐饮旅游、休闲健身、环境保护、商业保险等协同发展，创新健康医疗大

数据应用新业态。

健康医疗大数据是国家重要的基础性战略资源，是提高全民健康医疗水平的重要基础。国家卫健委确定在全国建设五个健康医疗大数据区域中心，并确定在山东省建设北方中心，同时要求山东省在健康医疗大数据管理特别是立法方面积极探索

政府高度重视，成立了北方中心建设领导小组，出台了《关于促进和规范健康医疗大数据应用发展的实施意见》(鲁政办发〔2017〕55号)，在基础设施建设、数据汇聚应用、标准规范建立、双招双引推动等方面做了大量工作。目前北方中心已成为当前唯一通过国家评估和授权建设的国家健康医疗大数据中心。



## / 山西省加快大数据融合应用 /

来源 / 山西日报 编辑 / 协会会员处 李苗苗 日期 / 2020-08

日前，山西省工信厅出台《山西省促进大数据发展应用2020年行动计划》，推动大数据与实体经济深度融合。今年，山西省将发挥能源、区位、气候等优势，推动传统工业高质量转型，加强统筹协调，加速推进5G、数据中心、工业互联网等新型数字基础设施建设。

推动大数据融合应用。以山西省优秀大数据企业为抓手，推动传统产业数字化、网络化、智能化转型，促进数字经济与实体经济深度融合。选树一批大数据优秀产品和试点示范项目，加快国家级大数据产业发展试点示范项目建设，发挥示范引领作用，引导大数据融合应用产品创新、服务创新、模式创新。

提升企业管理能力。贯彻落实工信部《关于工业大数据发展的指导意见》，组织开展数据管理能力成熟度评估(DCMM)宣贯培训，遴选“有数据、用数据、数据准要素化”的企事业单位进行试点评估，构建山西省DCMM评估服务体系，提高全省企事业单位数据管理水平和综合竞争力。

推动大数据在扶贫领域应用。动态完善扶贫大数据项目库，优秀案例加大宣传推介力度，并给予资金奖励。组织大数据企业深度对接，探索推动涉农扶贫领域数字化转型升级的路径。联合国内数字化服务平台，开展数字扶贫相关培训、讲座，加强贫困地区数字化人才能力提升，帮助了解、掌握相关数字化技能和数字化应用路径。

## / 提高大数据时代社会治理智能化水平 /

来源 / 中国社会科学网 编辑 / 协会会员处 李苗苗 日期 / 2020-08



中国特色社会主义进入新时代，社会主要矛盾发生转化，对中国特色社会主义社会治理体系提出了更高要求。“治理”这一概念内含了协调社会生活和社会关系的多种方法，大数据时代的社会治理智能化更加突显了治理方法的科技引领性。随着5G、人工智能、工业互联网等新型基础设施的全面铺开，围绕数据采集和信息整理的相关服务，诸如大数据中心、云计算中心等将随之蓬勃发展。大数据技术在加强和创新社会治理中的广泛应用，将深化我们对智能社会运行规律及其治理规律的认识。在这样的背景下，积极推进社会治理体系智能化，运用先进的理念、科学的态度、专业的方法、精细的标准提升社会治理效能，有助于增强社会治理预见性、精准性、高效性。

当前，加强和创新社会治理，要求完善党委领导、政府负责、民主协商、社会协同、公众参与、法治保障、科技支撑的社会治理体系。“体系”即复杂巨系统，在系统工程的意义上，大数据技术是社会治理体系获得科技支撑的物质基础，能够为社会治理体系的演化持续赋能。

进入大数据时代，我们建构智能化社会治理体系具有两大方法论优势：一是系统集成。大数据技术能帮助我们总体认识和把握社会矛盾和社会发展，在更加多元的维度上实现

“1+1>2”的治理效能。二是深度学习。以数据密集型的大脑式计算方法打造智能化核心驱动力量，实现社会治理科学化、精细化。

在社会化大生产快速发展的今天，地理上的以及行业的边界日益被人口、生产要素的大规模流动所打破，社会网络空间的非线性动力学特征越来越鲜明。但是，面对全球化背景下社会总体图景的变迁，社会治理体系的系统效应尚未充分发挥出来，社会治理体系碎片化和力量分散的问题依然没有得到根本解决。智能科技介入将极大地提升社会治理体系的整体性和协同性，提高预测、预警、预防各类风险的能力，进而实现标本兼治和持续有效的长程管理。

### 第一，以新基建为契机，锻造强大的信息能力。

大数据技术不是对客观世界的随机采样，而是能反映总体情况。正如美国数据科学家迈克尔·舍恩伯格所言，“大数据发展的核心动力来源于人类测量、记录和分析世界的渴望”。大数据技术的进步能让我们更加完整和精准地认识世界，一个大规模应用数据、组织生产和共享经济的时代正在开启。近来新基建被广泛重视，它是以新发展理念为引领，以技术创新为驱动，以信息网络为基础，面向高质量发展需要，提供数字转



型、智能升级、融合创新等服务的基础设施体系。可以说，新基建将极大地带动社会治理体系的数字化升级。受益于当下的产业数字化和数字产业化，社会治理体系具备了大数据采集与计算的物质条件。

新一代移动通信技术将驱动社会进入万物互联时代，5G与云计算、物联网、人工智能等领域的深度融合，推动形成了新一代信息基础设施的核心能力。尤其是5G网络，与4G相比具有高传输、低延迟、泛连接的特点，应用场景涉及增强型互联网、3D视频、云办公、增强现实以及自动驾驶、智慧城市、智能家居等。以5G网络为主体架构的新一代“信息高速公路”将为海量的数据和信息传递提供一条宽阔的高速传输信道，这是社会治理体系形成强大信息能力的保障。

## 第二，建设城市大脑，提升协同治理水平。

中国特色社会主义社会治理体系强调整体性和协同性，然而在市场经济条件下，社会结构的层次呈现多元化和弥散化，给协同治理带来了挑战。社会治理体系智能化以人工智能技术为支撑。人工智能如同云端大脑，依靠“信息高速公路”传来的数据进行深度学习和系统演化，完成机器智能化进程。目前方兴未艾的城市大脑建设正是这一进程的典型代表。就当前的实践看，城市大脑主要是为公共生活打造的数字化界面，包括交通出行、数字旅游、卫生健康、应急防汛等若干系统的应用场景，每天生成的协同数据多达上亿条。来自四面八方的在线数据不是“静态的”、而是“运动的”，城市大脑汇集的主题场景皆为现在进行时。随着数据采集的颗粒度越来越细，城市大脑能够高效便捷地掌握社会治理场景的确切信息和事件资料。

此外，在城市大脑建设中，多种网络实现有效连通，信息访问、接入设备的协同运作打破了过去部门、企业、团体之间的数据孤岛，从而推动形成了立体化、网络化的社会治理体系。城市大脑主要通过模型、算力和算法对大数据进行分析，并根据治理目标剔除干扰，提取信息，进行研判和预测，从而获得智能化解决方案，从技术上保障了社会治理效能最大化。

## 第三，立足民生改善，做好智能公共服务。

民为邦本，本固邦宁。国家建设得好不好，一个重要指标就是看老百姓的日子过得怎么样。提高人民物质文化生活水平，是改革开放和社会主义现代化建设的根本目的。新一轮大数据智能化的浪潮将会对包括智能制造、数字媒介、科技创新、医疗、教育、交通等多个领域产生广泛而深远的影响。提升智能化公共服务水平，就是要围绕民生改善，充分利用好泛在于社会各领域的的数据资源，加快释放“数字红利”，有效调配公共资源，协调多方利益。比如，坚持以人民为中心的发展思想，建立覆盖城乡的无线政务应用服务平台，开发并完善一体化移动办公系统，在幼有所育、学有所教、劳有所得、病有所医、老有所养、住有所居、弱有所扶等重要民生问题上解难

题、办实事。再如教育领域，身临其境的学习体验将普遍化。

在传统教育模式下，边远地区的学生较少接触高质量的教育资源，随着5G、VR/AR、AI等技术引入教学场景，就会给他们带来与大城市的学生一样的学习体验。当社会因教育资源均等化而实现整体素质的跃迁，社会生产力和文明程度的提高也会相应地优化社会治理体系，促进社会和谐与安定团结。

## 第四，着眼风险防控，开发适用于社会治理体系的CPS。

社会治理涉及面广、问题复杂，不仅需要定性判断，还需要作科学的定量分析。社会治理体系可利用5G网络超大连接的优势，收集社会治理策略执行后的各项微观数据，并且进行实时监测、量化分析、动态预测、评估效果，客观评估社会安全总体形势。然后，根据监测到的大数据科学研判发展趋势，依据相关信息反馈，调整社会治理方略和社会政策。

CPS即“信息物理融合系统”，是一个综合计算网络和物理环境的多维复杂系统，能够实现大型工程系统的实时感知、动态控制和信息服务。过去它被广泛应用于企业生产中，如今也可用来提升社会治理的运行效率。把CPS引入社会治理体系，对特定场景进行数学建模，并将模型映射到虚拟空间，实现决策过程的“人机结合”，有助于决策者对政策效果实施多方位和可能性推演，从而在社会治理中做到态势感知、风险预判、及时响应、预备处置。5G网络结合AI场景应用，为基于CPS的智能化社会治理体系提供了多主体协同行动能力，让防范重大风险社会问题真正成为可能。

## 第五，加强自组织建设，打造学习型社会和智慧社区。

大数据技术的确可以充分整合政府部门和基层社会数据，排除社会安全隐患、管理好人口流动，解决社会矛盾问题，提升社会管理质量。但是，社会作为一个整体不是在被动地接受管理，社会系统运行具有高度的复杂性和不确定性，社会治理体系无论多么完善，都不能削弱社会的自组织能力，相反它应该内嵌于社会自组织中才能发挥长效作用。

因此，谈及社会治理体系智能化，千万不能简单地理解为“全场景智能化监控”。在系统科学的意义上，自组织主要是社会有机体自我演化出的一种秩序结构，它主要受规则支配，而且这种规则能够随环境变迁不断变化。建设学习型社会，其内涵不止于知识的传播，更为重要的是群体交流和群体智慧的形成，这是社会自组织的基础。在5G增强网络的矩阵中，虚拟空间与现实世界的结合将更加紧密，社会群体之间深入互联互通，形成广泛的交融和理解。基层群众通过各类移动平台（App）使利益诉求和公共意见产生碰撞，大家在交流中生成社区智慧，以共识为基础的“想法流”把各方智慧和力量凝聚起来，从而增强社会网络韧性、促进社会和谐。

## / COVID-19之后的数字银行发展历程 /

来源 / 中国数据分析行业网 编辑 / 协会会员处 李苗苗 日期 / 2020-08



### 对感染的恐惧推动了数字银行业务

大约有三分之一的零售银行客户计划在COVID-19之后增加对在线和移动银行服务的使用，因为越来越多的服务（例如贷款和抵押要求和付款）正在变为虚拟，并且新客户可以使用数字化登机服务，由于人们不愿接触现金或信用卡，因此移动支付应用程序正迅速普及，随着大流行的继续蔓延，预计这些数字将上升，尤其是在第二波浪潮导致另一轮限制和关闭的情况下。

### 数字银行业务比替换以前在分行中提供的服务要多得多

开放式银行可以通过使用标准API捕获客户帐户中的数据，为新的创新数字服务提供动力，从而为银行提供竞争优势，例如通过访问客户数据的全面视图，银行可以允许从其他位置的帐户进行付款，由基于机器学习的应用程序提供支持的数字个人顾问可以分析所有储蓄和投资账户以及客户的个人目标，以推荐最佳机会，金融公司还可以使用提供的数据来确定与客户最相关的服务，最具竞争力的服务价格以及最有利的投资条款和条件，基于高级风险模型和市场产品的贷款和抵押，只是想一想，通过向客户展示如何提高他们的收入或减少他们的花费来说服客户转移到您的银行是多么容易。

这些服务也不限于B2C银行最近推出了为小型企业设计的数字工具该工具可帮助简化交易并生成现金流量预测，并易于与财务顾问联系以获取快速建议。

### 迈向更高效的系统架构

随着越来越多的人和企业依靠数字应用程序来提供银行服

务，在线交易数量持续增长，对现有的IT计算资源造成压力。查询数量的大量增加导致瓶颈，这可能降低应用程序的性能并影响客户服务水平，当客户等待太久才能完成交易或获得贷款批准时，或者如果客户知道可以从另一家银行获得更好的条件，则他们更有可能转换，因此银行面临着扩大其昂贵的传统基础架构以提供预期的用户体验质量的需求，或者寻找能够以最佳的TCO弹性扩展以所需的速度来管理此数据的现代解决方案。

### 在许多情况下大型金融服务机构受到缠结和过时的系统的限制

这些系统过于复杂，无法以最佳方式管理，处理和分析来自不同来源的大量数据，最近在超过60%的受访者表示担心银行将由于“银行核心架构的当前状态”而难以开放其API，但是有一些创新旨在帮助银行处理更多数据，更经济高效，而无需翻新和更换现有平台，例如某些内存计算平台可以将所有数据模型（结构化，非结构化和半结构化），服务业务逻辑和分析共同定位在内存数据核中，并将其分布在水平可扩展的体系结构中，以实现异常快速的处理时间，同时消除了对过剩产能进行投资的需求。

### 这些现代的数据和分析平台支持敏捷开发和应用程序的部署

这些应用程序涉及在可变数据流，事务性和历史数据上同时运行的交互式查询，分析甚至机器学习，这些数据存储在外部数据湖和数据仓库中，应用程序能够以100倍的速度访问内部或外部的数据湖上的数据，无论是在内部还是在混合环境中，这不仅可以加快查询和BI报告的速度，而且还可以提高时间关键型应用程序和决策服务（如风险）的准确性分析，欺诈检测，贷款审批，动态定价和个性化服务。

### 数字银行将成为新规范

数字银行将成为新的规范，但是金融服务将需要采用新技术来利用其数据并从中获利，以保持其市场领先地位，数字银行业务的全部潜力只有在拥有高性能数据管理基础架构的情况下才能实现，该基础架构能够以优化的TCO快速可靠地完成对时间紧迫的应用程序的查询和分析，COVID-19可能导致分支机构关闭，但它可以加快开放银行业务的采用，带来便利和新的应用程序，这些应用程序将改变客户的银行业务方式。



## / 在制造业中实施大数据战略时要考虑的挑战 /

来源 / 中国数据分析行业网 编辑 / 协会会员处 李苗苗 日期 / 2020-08

### 利用工业大数据可以进行预测性维护和实时监控

可以发现故障的可能性，并可以降低维护成本，例如在化工厂 深入了解管道中的流体/气体流量有助于预测维护时间，制造业正在迅速采用大数据策略以提高效率和生产率，先进的分析功能可帮助解码复杂的制造流程，用自动化算法替代人为制定的决策，并提高生产效率和速度。

### 什么是大数据？

大数据是指在给定的时间范围内无法使用传统方法存储和处理的海量数据，能够利用海量数据对行业有益，有大数据技术可用于处理如此大量的数据，例如Hadoop是一个框架，旨在使用简单的编程模型通过商品硬件在分布式数据处理环境中存储和处理数据，它可以高速和低成本地存储和分析存在于不同机器中的数据，还有其他可用的技术，例如MongoDB，Rainstor, Hunk等。

宝马集团依靠对生产数据的智能使用来实现高效流程和优质质量，这是从大数据中创造价值的最佳范例。

制造汽车会在整个价值链中产生大量数据，宝马集团使用其智能数据分析数字化集群来选择性地分析这些数据并增强其生产系统，智能数据分析的结果为提高生产和物流各个领域的质量做出了有效的贡献。

### 大数据在许多方面对制造业产生了积极影响

- 1、工艺改进可提高产量并提高生产效率。
- 2、诸如“神经网络技术”和“机器学习”之类的方法比较了各种生产要素的影响。
- 3、供应链管理的改善导致交货时间缩短，风险降低。
- 4、个性化生产使企业能够满足个性化或特定需求以及更多需求。
- 5、实施中面临的主要挑战。

### 确定需求大数据战略是关于收集信息并使用它们来改变业务运营方式的

一家制造公司在生产中执行许多流程，因此了解大数据策略对改进特定流程的需求至关重要，当然首先需要确定需要解决什么问题，否则我们可以无意识地探索大量数据，并希望最终我们能在其中找到一些东西，在大多数情况下，制造公司会花费大量时间和资源来捕获随机数据并对其进行处理，因此在大多数情况下并没有带来任何好处，确定实际需求并收集可帮助您实现目标的数据是一项挑战。

数据选择内部产生大量数据，例如客户交易数据，内部

供应链数据和整个公司的许多绩效数据

对于许多公司而言，单单处理这些数据有时是一个挑战，但这并不是可以创造所有价值的地方，了解其他可用数据来源非常重要，例如我们可以将天气和气候数据，交通模式数据，价格比较数据等外部数据带入市场，以了解市场上还提供了哪些其他价格，确定要使用的数据，如何获取数据，如何将其汇总为可在整个公司中使用的集成形式是一个挑战。

### 转型能力大数据实施中最困难的部分是转换功能

重要的是要了解数据的真正影响通常需要大量的策略，团队的努力和所花费的时间，使具有适当技能的人有能力使用最新的数学技术和最新的统计方法来处理数据并带来收益，需要招募在行业中以相同方式工作多年的人员，建立一支由熟练专业人员组成的高效团队是一项真正的变革管理挑战，在许多情况下，公司会聘用现有人员并对他们进行新方法，新流程和新技术的培训，这需要用在不同环境中拥有丰富经验的人员来补充现有团队。

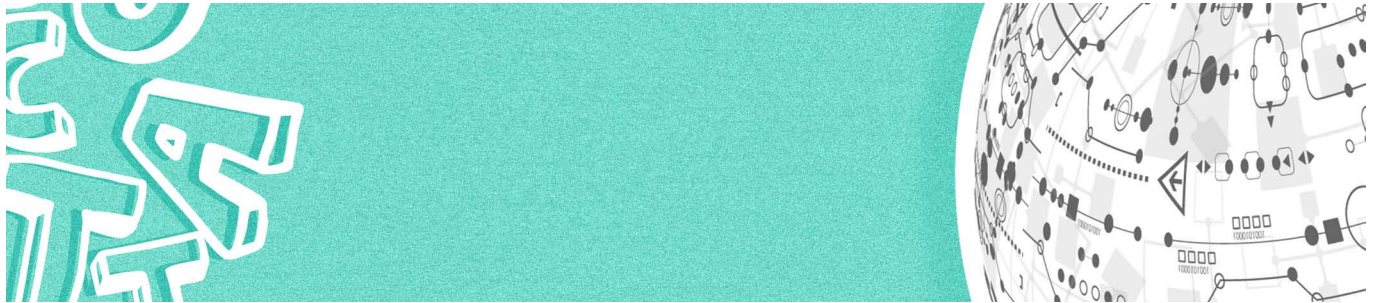
随着数据的增长制造公司正在应用分析技术以更快的速度和效率获得可观的价值,但是在花费数百万美元用于数据分析之后，由于仍未解决的挑战，公司仍无法看到收益，这些挑战发生在实施的所有级别，例如捕获正确的数据，快速处理和分析数据，同样大数据缺乏情感智能，公司还必须想出对大数据产生情感影响的方法。





# / 10个步骤巧用数据分析法 轻松摸透竞争对手 /

来源 / 中国数据分析行业网 编辑 / 协会会员处 李苗苗 日期 / 2020-08



分析竞争对手的目的是为了解对手，洞悉对手的市场策略等。我们可以用竞争对手分析的五个层次来说明，能准确地确定竞争对手，这是分析的最低层次，能分析出对手状况则是第二层次，最高层次是通过竞争分析制定策略后能够引导对手的市场行为。确定了你的竞争对手并收集到足够数据后，我们就要对他们进行深度分析了。

### 竞争对手分析路径

竞争对手分析共分为 10 个步骤，如图 1 是分析路线图。这个分析路线图侧重于企业前端即营运端的分析。

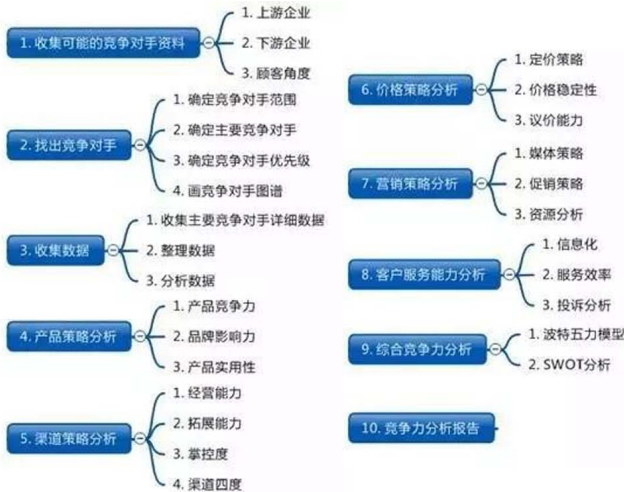


图1 竞争对手分析路线图

一个企业的策略如果是根据竞争对手策略来制定的话，这个企业是没有持续性的，每个企业策略应该具有企业自身的特色。分析竞争对手的目的是为了解对手，洞悉对手的市场策略等。我们可以用竞争对手分析的五个层次来说明，如图 2 所示。

能准确地确定竞争对手，这是分析的最低层次，能分析出对手状况则是第二层次，最高层次是通过竞争分析制定策略后能够引导对手的市场行为。



图2 竞争对手分析层次

### 画竞争对手图谱

路线图第二步找出竞争对手中，画竞争对手图谱是为了将各个层面的核心竞争对手和潜在竞争对手标注出来，以便在渠道策略、资源投放、生产规划等方面更有针对性和差异化。如图 3 所示，这是用思维导图做出来的竞争对手图谱，当然也可以用地图形式来画。图中类别之间还可以进一步细分，也可以继续标注出每个竞争对手的习惯反应类型等。



图3 竞争对手图谱

### 量化竞争对手的四度

在第五步渠道策略分析中需要量化竞争对手的四度，这四度就是渠道广度、渠道宽度、渠道长度和渠道深度。渠道的广度指公司产品覆盖的区域多少，渠道的宽度是指有几种类型的通路，渠道的长度指产品平均经过几个中间渠道到达消费者手中。渠道的深度指通路上渠道商数量的多少。

我们用四度举例来说明竞争对手分析中的三种方法：

#### · 排行榜方法

适用于对单个指标的若干个对象强弱分析时使用，例如我们对渠道广度这个指标分析时，可以对竞争对手进行排行，从而看出竞争对手间渠道广度强弱，如图 4 所示。

	渠道广度-城市数
对手D	325
对手A	294
本企业	293
对手E	270
对手H	234
对手K	197
对手B	171
对手J	166
对手G	153
对手F	132
平均值	223.5

图4 竞争对手渠道广度分析

#### · 四象限分析法

四象限法适合对两个指标进行分析，如图 5 所示，这是个对竞争对手渠道广度和渠道深度的分析图。和竞争对手相比本企业覆盖了 293 个城市，广度不错，但是深度只有 2283 个客户，低于 2644 的平均值，偏低，和竞争对手 A 相比渠道深度差距相当大。

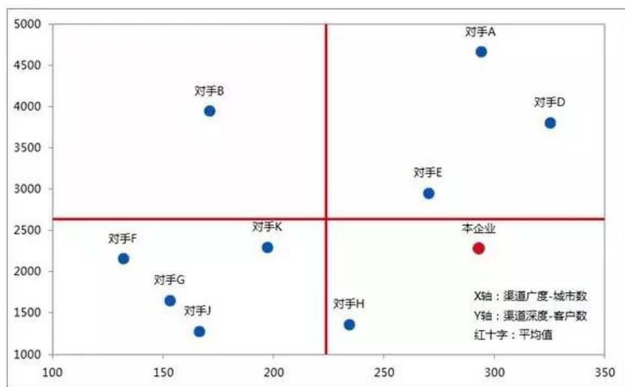


图5 全国竞争对手渠道广度-深度分析

另外，如果需要对三个指标进行分析时，可以考虑用四象限气泡图来辅助分析。

#### · 雷达图分析法

当需要分析的指标有 4 个或 4 个以上时，一般的图表就不能达到很好的展示效果，此时可以考虑使用雷达图。但是雷达图由于只有一个坐标轴，不能同时展示不同的量纲数据和不同的数量级数据，例如渠道广度值介于 100-400，渠道深度介于 1000-5000，所以需要进行去量纲处理。我们可以采用排名的方法实现去量纲或去数量级的目的。如图 5 所示，这是渠道四度的雷达图展示，坐标轴为名次，一共 10 个对象，每个点是一个维度的排名值。

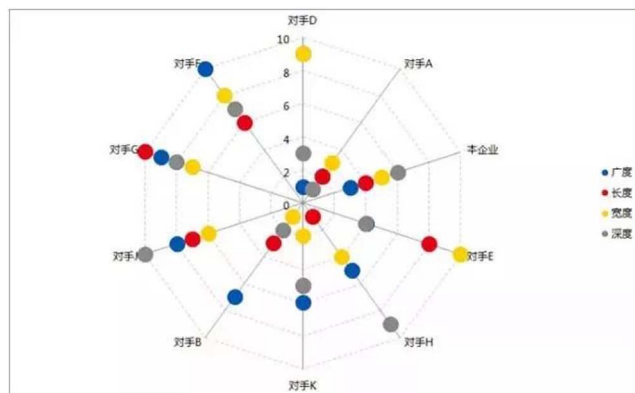


图6 竞争对手四度分析

### 波特竞争力分析模型 (Porter 5 Force Analysis)

波特竞争力分析模型是哈佛商学院教授迈克尔·波特提出的，用于竞争战略分析。他把竞争力归纳为五力，分别是供应商的议价能力、购买者的议价能力、潜在竞争者进入的能力、替代品的替代能力和行业内竞争者现有的竞争能力。五力的组合决定了行业的利润水平，如果企业处在一个供应商议价能力低，购买者议价能力也低，有行业壁垒潜在竞争者不易进入，没有替代品，同时行业竞争也不充分的行业中，这个企业一定是高利润高垄断的高帅富企业。

#### 传统零售业的波特五力分析

**供应商的议价能力：**无论是自营化的连锁超市，还是平台化经营的百货、购物中心，基本上都是零售商占主导地位，供应商的议价能力不强。属于店大欺客（户）的状况，特别是例如电器连锁、KA 大卖场等，供应商的议价能力更低。

**购买者的议价能力：**越充分竞争的市场，消费者选择的余地就越大，零售商间的竞争赤裸裸的体现在价格上，从而造成了顾客的议价能力逐渐加强。

**潜在竞争者进入的能力：**传统零售业是一个需要高投入，投资周期长，要求规模化的行业，潜在竞争者直接进入的能力并不强。

**替代品的替代能力：**目前传统零售的最大替代者是电子商务，电子商务对传统零售的冲击逐渐增强，所以替代品的替代能力很大。当然替代的边界在哪儿，目前没有人知道。

**行业竞争力：**零售业是一个充分竞争的行业，高线城市

大都饱和，低线城市还有一些机会。波特五力分析模型除了对行业整体的分析，还可以与具体竞争对手进行对比分析，可以通过专家打分的方式进行量化处理。

#### · SWOT 分析模型

SWOT 是经典的战略分析工具，始于麦肯锡。分别由优势-Strengths、劣势-Weaknesses、机会-Opportunities 和威胁-Threats 四部分组成。它是对企业所处的外部环境以及企业内部环境的一种综合分析方法。SWOT 分析可以用在公司战略、竞争对手分析、市场定位、甚至个人的职业规划等方面。

用 SWOT 分析竞争对手就是将收集到的竞争对手情报进行综合分析，并最终形成分析结论和策略。SW 为内部关键因素，OT 是外部关键因素。对于零售企业或零售品牌来说，建立 SWOT 分析模型前我们需要回答如下问题：

#### 优势

S1. 我们最擅长什么？是产品设计开发？渠道布局？营销手段？还是价格杀手？

S2. 我们在成本、技术、定位和营运上有什么优势吗？

S3. 我们是否有其他零售商不具有或做不到的东西？例如有的零售商有企事业单位发放购物券优势。

S4. 我们的顾客为什么到我们这儿来购物？我们的供应商为什么支持我们？

S5. 我们成功的原因何在？

#### 劣势

W1. 我们最不擅长做什么？产品、渠道、营销还是成本控制？

W2. 其他零售商或品牌商在那些方面做得比我们好？

W3. 为什么有些老顾客离开了我们？我们的员工为什么离开我们？

W4. 我们最近失败的案例是什么？为什么失败？

W5. 在企业组织结构中我们的短板在哪儿？

#### 机会

O1. 外部在产品开发、渠道布局、营销规划和成本控制方面我们还有什么机会？

O2. 如何吸引到新的顾客？如何做到与众不同？

O3. 在外部因素中和公司短期、中期规划目标的机会点有哪些？

O4. 竞争对手的短板是否是我们的机会吗？

O5. 行业未来的发展如何？是否可以异业联盟？

#### 威胁

T1. 经济走势、行业发展、政策规则是否会不利于企业的发展？

T2. 竞争对手最近的计划是什么？是否会有潜在竞争对手出现？行业内最近倒闭的企业是什么原因？

T3. 企业最近的威胁来自于哪儿？有办法规避吗？

T4. 上下游的客户中是否有不和谐的地方？资源状况如何？

T5. 舆情是否不利于公司发展？

行业不一样、企业不一样这 25 个问题也会不一样，每个企业可以根据自己的特性进行调整。我们需要通过这些问题来对 SWOT 进行量化处理。如图 6 所示，结合收集到的竞争对手情报，对 25 个问题分别进行打分，然后设定不同问题的权重，最后就得到 SWOT 以及 SW、OT 的综合得分。

	内容	权重	得分-10分制	加权得分	内-外部合计得分
S	S1	0.15	5.00	0.75	5.71
	S2	0.20	7.10	1.42	
	S3	0.05	7.90	0.40	
	S4	0.10	7.80	0.78	
	S5	0.05	6.10	0.31	
	合计	0.55	34.90	3.65	
W	W1	0.15	5.10	0.77	5.63
	W2	0.10	5.20	0.52	
	W3	0.10	3.90	0.39	
	W4	0.05	2.70	0.14	
	W5	0.05	4.90	0.25	
	合计	0.45	38.80	2.06	
O	O1	0.10	8.50	0.85	5.63
	O2	0.10	6.10	0.61	
	O3	0.15	5.20	0.78	
	O4	0.10	8.10	0.81	
	O5	0.10	8.00	0.80	
	合计	0.55	35.90	3.85	
T	T1	0.15	4.00	0.60	5.63
	T2	0.05	6.50	0.33	
	T3	0.05	3.90	0.20	
	T4	0.10	4.10	0.41	
	T5	0.10	2.50	0.25	
	合计	0.45	34.00	1.78	

可以对自己企业和不同竞争对手分别打分，就能很好的发现彼此的 SWOT 现状。SWOT 除了用在战略、竞争对手分析、职业生涯规划等方面外，一些战术制定也可以用 SWOT 进行梳理，例如谈判策略制定，新产品上市等。



# / 2020年中国大数据安全市场现状分析：市场规模将超1万亿 /

来源 / 中国商情报 编辑 / 协会会员处 李苗苗 日期 / 2020-08



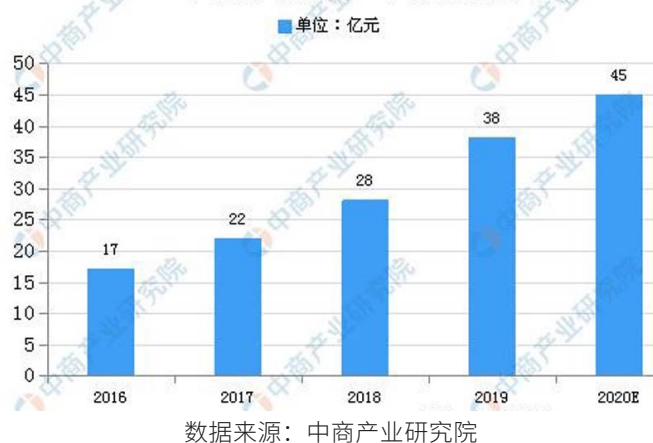
中商情报网讯：近年来，中国大数据产业规模持续高速增长，数据显示，2018年我国大数据产业规模突破6000亿元；随着大数据在各行业的融合应用不断深化，2019年中国大数据市场产值达到8500亿元。未来，大数据技术应用将进一步加深，中商产业研究院预测，2020年大数据产业规模有望突破10000亿元。



## 大数据安全市场规模高速增长

大数据安全是用以搭建大数据平台所需的安全产品和服务，以及大数据场景下围绕数据安全展开的大数据全生命周期的安全防护。数据显示，2019年我国大数据安全市场规模近40亿元，2016年-2019年复合增长率为30.8%。中商产业研究院预测2020年中国大数据安全市场规模将达45亿元。

### 2016-2020年中国大数据安全市场规模统计预测



## 大数据安全技术的现有能力和水平存在差距

近年来，国内外的大数据平台安全、数据安全、隐私保护相关的技术稳步发展，但大数据平台的漏洞扫描与攻击监测技术相对薄弱，数据共享安全、非结构化数据库的安全防护以及数据泄露溯源技术的改进迫在眉睫。

## 大数据成为网络安全重点目标

随着互联网与大数据的逐步发展，大数据的安全问题逐渐暴露，针对大数据的勒索攻击和数据泄露问题日趋严重，全球大数据安全事件呈频发态势。催生了大数据相关的安全技术、解决方案及产品的研发和生产。

## / 阿里发布《2020农产品电商报告》： 去年全平台农产品交易额2000亿 数字农业将成风口 /

来源 / 央广网 编辑 / 协会会员处 李苗苗 日期 / 2020-08



7月14日，阿里巴巴发布《2020农产品电商报告》（以下简称《报告》）。报告显示，2019年，阿里平台农产品交易额为2000亿元，稳居全国最大农产品上行电商平台。2020年疫情期间，截至4月25日，淘宝、天猫平台累计为全国农民售出超过25万吨滞销农产品。

《报告》中指出，未来五年是数字农业发展的窗口期和机遇期，阿里巴巴积极打造数字农业新基建，在数字化生产、物流、营销、金融等方面全面发力，积极探索数字农业新业态，助力中国由农业大国向数字农业大国转变。

### 阿里稳居全国最大农产品上行电商平台

报告是今年阿里巴巴第一次全面公布关于农产品电商的大数据。其中显示，2019年，阿里平台农产品交易额为2000亿元，稳居全国最大农产品上行电商平台。

在农产品电商销售额中，浙江、广东、江苏三省占据前三位。而山西、山东、河北三省增速最快，增幅超过40%。

随着人们生活质量的提高，吃得好变成更重要的话题，肉蛋禽、奶类、蔬菜等基本生活所需不再是农产品网购的最爱，人们更喜欢选择休闲食品、滋补食品和水产品。

广东不愧是最喜欢吃的省份。《报告》显示，农产品电商消费排名上，广东省位居第一。或许是受了邻居的影响，它身旁的广西壮族自治区，农产品电商消费增速高达35%。

### 抗疫助农，阿里全力帮助湖北

《报告》也全面展现了今年上半年，特别是疫情期间，阿里平台的线上线下渠道为解决全国多地农产品滞销而做出的努力。尤其是对湖北的特殊行动。

疫情期间，借助阿里的渠道联动，有超过550种湖北农产

品被推广，第一季度线上交易金额达到10亿元。

3月2日，疫情以来的首单大宗秭归脐橙，通过阿里巴巴紧急打造的数字供应链运出湖北发往全国。

3月31日，淘宝上线“湖北加油”专区，2020年将销售120万吨湖北农产品。

4月1日，阿里巴巴旗下盒马事业群、数字农业事业部牵头，从湖北采购价值10亿元的小龙虾。

农产品直播在疫情期间发挥了作用，2020年一季度，农产品直播销量比上年同期增长1.4倍。

### 用数字化搭建中国乡村新基建

《报告》也指出了未来的趋势，中国迫切需要数字农业新基建。

去年10月10日，阿里巴巴数字农业事业部成立，在数字化生产、物流、营销、金融等方面全面发力，“盒马村”、淘宝直播成了数字农业的新业态。盒马村是指根据订单为盒马种植农产品的村庄，是阿里巴巴数字农业基地的典型代表。今年疫情期间，全国“长出”几十家盒马村，在多地农产品出现滞销时，盒马村无一滞销。

近期，在数字农业上，阿里巴巴又连出大动作，宣布搭建全国农产品数字化流通网络，“产地仓+销地仓”模式构成了这个系统的骨架；和农业农村部战略合作，共推数字农业；和山东淄博战略合作，建立盒马市，全国五大产地仓落地淄博。这些，都是中国乡村数字农业的新基建。

未来，中国将从农业大国转变成数字农业大国。



## / 从综艺IP到私域IP，获取别人的“品牌重塑”经验 /

来源 / 中国数据分析行业网 编辑 / 协会会员处 李苗苗 日期 / 2020-08

疫情的出现打乱了品牌的正常传播节奏，线下商业受阻，用户从线下涌入线上，如何才能抓住他们的注意力，对于任何品牌都是一场营销大考，疫情背景下品牌应该如何破局？

对于大众品牌而言，线上营销自然绕不过娱乐综艺营销，这也是最有效地快速破圈方式，但如何才能将综艺IP势能更长效地为品牌所用，从而将娱乐营销作为流量入口搭建自身的内容流量池？

同样发生的还有大环境下用户需求的快速变化，移动互联网语境下，年轻消费趋势高频转变，这也让品牌始终面临着年轻化传播的压力，不仅仅是老字号，所有品牌都该将年轻化作为常态化的营销动作，如何实现品牌年轻化？



以上三大问题都成为招商银行今年上半年疫情期间的营销挑战，时趣帮助招商银行通过一系列#一招就行电台#传播，完成了与年轻群体的有力互动，并给出了后疫情时代中品牌年轻化传播、综艺IP营销更多可能性和样板。

前有多变多元的年轻人，后有新冠疫情突如其来，招商银行应该如何切中年轻人的痛点？在当下传播环境中实现品牌突围？

### 疫情下营销受阻，哪里去找新蓝海？

疫情的突发打乱了品牌常规的营销规划，线下商业受阻

让招商银行不得不寻求有效的线上营销方式。而从另一个方面来看，疫情的持续也直接影响大众情绪痛点，这也给予了品牌足够的营销内容切入点。

可以发现，疫情虽能堵住了线下的常规营销渠道，但在内容传播、社会化营销层面上，却更有可能让大众用户产生情感共鸣从而获得破圈，这便是疫情所带来的大众情绪共鸣点。那么，具体有哪些营销机会点呢？

一方面，疫情让消费者彼此物理隔离，这刺激了大众的在线生活，无论是线上平台的月活用户数，还是用户平均时长相比去年下半年都有显著提高，其中文娱类产品更是有着快速增长，因此，品牌使用文娱类内容手段实现持续输出，可以吸引用户的持续关注；

另一方面，消费者不仅开始重视健康，更在隔离的环境下向往人与人之间的陪伴与温暖，治愈和暖心成为了年轻大众的情绪爆发点，而与之相对的担忧、抑郁都是长期困扰年轻人群的健康痛点，这也为品牌在内容洞察及创作上提供了重要的突破口。

对于招商银行来说，疫情可以说既是挑战亦是机遇，但关键点在于，如何将品牌价值与年轻人的情绪实现无缝对接，深化年轻人的品牌形象？

招商银行的切入点在于选择与芒果TV声音互动陪伴真人秀《朋友请听好》综艺进行营销合作，借助综艺的传统渠道影响力和疫情下的关注热度，输出招商银行的暖心、陪伴品牌价值，实现招商银行品牌年轻化的目的。



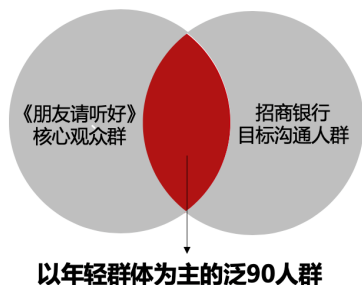


但综艺合作对于品牌而言相对被动，在传播上受节目本身节点影响较大，因此综艺营销往往只是品牌营销动作中的第一步，用来作为话题入口和流量入口。对于招商银行也是如此，《朋友请听好》为的是开辟流量新赛道，为后续持续拉新和获客提供基础土壤，

如何把综艺所带来的势能放大破圈？如何形成更加广泛的社会话题？如何为品牌构建持续的内容流量闭环？这才是招商银行疫情下营销的重中之重。

### 如何将综艺势能为品牌所用？

对于大多数品牌而言，借势综艺的方法除了冠名植入等硬广手段外，就是根据综艺节目内容做相关延展，但这样始终会有内容短缺的问题所在。破局方法在于通过综艺节目而构建品牌自身的内容池，再通过内容传播实现源源不断的流量引入及转化。



针对招商银行面临的挑战，时趣通过打造品牌自身内容IP的方式实现流量闭环。整体营销策略分为三步走：1、通过综艺流量搭建品牌IP内容；2、激活用户参与及KOC传播；3、社交化破圈传播。

IP的构建是一件系统工程，需要一整套的营销动作进行支持，所以尽管传播策略是清晰的，但每个执行步骤中都将面临不小的挑战：

#### 1、品牌自身IP如何在短期内独立孵化？

所有IP的孵化运营，都需要长时间的积累与沉淀，一方面需要受众的积累，另一方面需要内容的持续输出。对于招商银行而言，自有渠道及《朋友请听好》的合作可以解决受众来源问题，但原创内容必须通过品牌自身进行搭建，另外，这个自建的IP内容还需要独立于综艺本身，让其有足够的延展性在未来可以持续传播。

招商银行品牌形象与《朋友请听好》声音互动节目均以“温暖、陪伴”为内核，针对年轻人的生活境遇及痛点，招商银行通过与综艺调性匹配度高的音频电台方式，作为构建自身品牌IP的形式选择。

在内容创意上，时趣帮助招商银行采取了「一招就行电台」的整体IP名称，并且跟随综艺的播出节奏，每1期或2期，围绕1个消费者典型的烦恼情景，输出治愈系的电台内容，并在微信中打造短视频版的「一招就行电台」。



根据对年轻人情感的分析，我们将普遍烦恼情绪主要分为六大类主题，不同主题面向不同的目标人群进行输出。

具体来说，这六大烦恼主题分别为：失去激情、面对失败、同辈压力、亲情事业两难、爱情选择、生活迷茫，这也是电台内容策划的出发点，分别面向职场新人、考研学子、在校大学生、北漂青年、都市情侣、女性创业者六类精准目标人群。



完成内容的初步框架搭建后，便可以在IP内容框架中持续实现品牌资产的增值。通过陪伴式电台内容的输出方式，让招商银行品牌调性变得更加具有亲和力，将品牌定位于“朋友”的角色，以真挚、平等的态度，站在消费者的视角上与之对话，方能获得消费者的接纳。

#### 2、品牌IP如何吸引用户参与分享？

IP内容需要在用户互动反馈中不断发酵才能实现社会化传播的目标，有了「一招就行电台」的内容框架和策划，接下来的重点在于如何刺激用户进行参与和分享，刺激KOC的传播扩散。

传统的内容IP更多在于单向内容输出，即所谓的PGC内容，优点在于内容品质的把控，但缺点在于无法覆盖更多垂直人群，并且减少了用户参与感，因此招商银行通过PUGC的方

式，在专业团队策划内容主题、拍摄主题视频的同时，开放用户投稿创作的通道，实现目标用户的广泛参与，并快速催化IP内容的成熟和丰满。

用户共创内容主要来自于微信留言及邮件信箱，用户内容的“上墙”与“选中”激发用户群体间的互动，培育忠诚粉丝群，也代表着对「一招就行电台」内容的认可与共鸣。



用户留言参与“上墙”

我们抓取了用户评论共创内容的关键词，发现用户的个人主题创作内容大多基于自身的亲身经历，并且在「一招就行电台」内容的引导下获得了正向的情绪反馈，也就是说，电台内容有效地帮助目标用户缓解生活烦恼，输出品牌正能量的价值主张。

- ◆ 聚焦个人经历，反应真实个体声音：“自己”“我们”“生活”“工作”……
- ◆ 反应群体焦虑情绪，引发情感共鸣：“失眠”“失望”“放弃”“遗憾”……
- ◆ 传递正能量，正向情绪引导获得治愈：“努力”“成长”“希望”“加油”“温暖”“喜欢”“帮助”……



用户UGC内容的数据分析

### 3、如何使品牌IP内容破圈传播？

构建自身IP内容体系、激活用户参与感后，最后的问题在于渠道平台传播。在当下的媒介碎片化时代，零碎片段化的信息很容易被淹没，而不同平台拥有不同的话语体系与平台文化，导致精细化传播始终是品牌面临的一个难题。

「一招就行电台」相关物料在主流平台中已经实现的不错的曝光量，比如说除了招商银行3个微信公众号外，腾讯视频的全网播放量超16万，微博官微发布21条微博总阅读破92万，互动量近3000条，但进一步的破圈引起社交话题依旧需要佳



在KOL合作上，招商银行采用了行业垂直KOL+面向大众KOL的组合拳，并利用KOL共创的内容生产模式，全方位激起大众讨论。其中，垂直行业KOL从品牌建设差异化进行战略切入，以树立标杆实现圈层热议，奠定“出圈”的基础；大众化的KOL更多是以情感生活切入、传递温暖，输出有价值的内容文字，使读者受益，造成广泛共鸣。

值得注意的是在与大众类KOL的合作过程中，需要进行品牌与KOL的内容共创，而不能简单作为硬广告投放制造曝光量，那样的话不仅有损招商银行品牌的“朋友对话感”，而且也会降低转化率，事实上只有对用户真正有价值的内容才有益于品牌的长期建设。

在微信之外，「一招就行电台」的重点传播阵地还有抖音、小红书、微博。时趣根据不同平台采用不同的内容传播策略：

针对抖音的短视频传播特性，我们每周策划话题性趣味主题，混剪节目亮点片段，融入招行品牌宣传。实现借节目热度吸引关注的同时，减少受众排斥心理，达到品牌宣传的目的。

基于小红书种草平台特性，每周以笔记形式，结合节目话题，从情感角度切入，用更符合平台女性群体特征的表达方式与受众沟通。在引发人群关注的同时，塑造招行温暖的品牌调性。



基于微博社交媒体平台特性，在节目播出的第一时间，从节目内容中提炼热门话题，借节目和明星热度吸引粉丝关注，扩大传播受众基础。

通过自建品牌IP「一招就行电台」，招商银行不仅打破了传统的综艺合作的传播套路，而且探索了与年轻客群对话的新尝试，在疫情环境下传递“温暖”、“倾听”的品牌内

涵。从最终效果来看，招商银行在无明显加持、非总冠名的情况下，传播效果超过了75%的其他节目赞助商，并且树立了年轻用户群中的品牌IP认知，长期来看增加了招商银行品牌资产。

## / 数据分析师：指标体系搭建四原则 /

来源 / 中国数据分析行业网 编辑 / 协会会员处 李苗苗 日期 / 2020-08



### 指标体系的定义和选取原则

定义：在业务的不同阶段，分析师牵头、与业务方协助，制定的一套能从各个维度去反应业务状况的一套待实施框架。在建立指标体系时，要注重三个选取原则：根本性、可理解性、结构性。

根本性：核心数据一定要理解到位和准确，切记不能选错

可理解性：所有指标都要配上业务解释性，如日活的定义是什么，打开还是点击还是进程在就可以

结构性：能够充分对业务进行解读，如新增用户只是一个大数，我们还需要知道每个渠道的新增用户，每个渠道的新增转化率，每个渠道的新增用户价值等

### 指标体系建立的四步法

在建立指标体系之前，我们先了解一下指标的构成，在我们工作过程中遇见的指标多为派生性指标，即，原子性指标+修饰词+时间段，修饰词指标本身是可选内容，而原子性指标和时间段为必选内容。

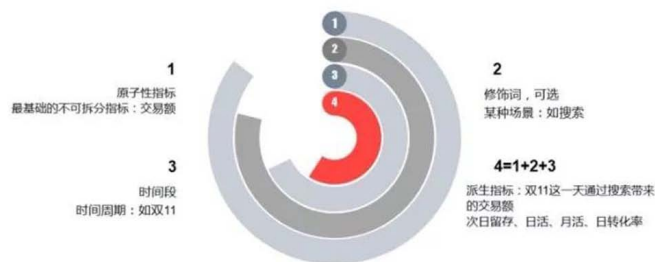
在这里，原子性指标指的是不可拆分的指标：

如：交易额、支付金额、下单数之类；修饰词多是某种场景的表现

如：通过搜索带来的交易等；时间段即为一个时间周期；如：双十一期间，618活动期间等。三者叠加就形成派生指标

如：“双11这一天通过搜索带来的交易额”，如果不需要修饰词，就是“双11这一天带来的交易额”

同样，像此类日活、月活、次日留存、日转化率等都属于派生指标。



在合理筛选完指标后，就要着手建立对应的指标体系，主要分为四个步骤：厘清业务阶段和方向、确定核心指标、对指标进行核心维度拆解、指标的宣贯、存档、落地。



### 1. 厘清业务阶段和方向

对于一家公司往往分为三个阶段：

业务前期（创业期）

业务中期（上升期）

业务后期（成熟发展期）

针对不同的阶段关注的核心指标也是有差别的。



业务前期，最关注用户量，此时的指标体系应该紧密围绕用户量的提升来做各种维度的拆解，如渠道；

而在业务中期，除了关注用户量的走势大小，更加重要的是优化当前的用户量结构，比如看用户留存，如果留存偏低，必然跟产品模块有关，是不是某个功能流量惩戒效果太差，这也是在分析产品的健康度，为产品体检；

成熟发展期，更多关注的就是产品变现能力和市场份额，要关注收入指标ROI等，各种商业化模式的收入，同时做好市场份额和竞品的监控，以防止新起势力抢占份额等。

### 2. 确定核心指标

此时最重要的是找到正确的核心指标，举个例子帮助大家理解。

例：某款产品的日活口径是打开APP，通过不断的买量，日活也一直在上升，然而分析时发现，打开APP的用户中，3秒跳出率高达30%，这是非常不健康的，那么当前的核心指标日活实际上已经有了问题，更加好的核心指标应该是停留时长大于3秒的用户数。

每个APP的核心指标不太一样，所以一定要多花时间去考虑这件事，这个非常重要，不只是看日活和留存那么简单。就像趣头条这款APP，它的日活和留存指标一定非常高，但仅关注这种指标肯定会出现问题，并且它的真正核心指标绝对不是单纯的日活和留存。

### 3. 核心指标维度拆解

核心指标的波动必然是某种维度的波动引起，所以要监

控核心指标，本质上还是要监控维度核心指标。通用的拆解方法都是先对核心指标进行公式计算，再按照业务路径来拆解。假设，当前的核心指标是停留时长大于3秒的用户数。

公式：停留时长大于3秒的用户数=打开APP的用户数\*停留时长大于3秒的占比



分析“打开进入APP的用户数”时，我们要关注渠道转化率，分析用户从哪里来；同时用户通过哪种方式打开的，如通过点击桌面图标、点击通知栏、点击Push等；并且，这类用户的用户画像是什么，用户画像也更多是在这个时候才最有作用，更多要基于场景和相应的指标来分析。

“停留时长大于3秒的占比”该指标要重点关注如，停留时长的分布，停留1秒的用户有多少、2秒的用户有多少、3秒的用户有多少，具体分布情况是怎样的；停留大于3秒的用户特征和行为特性是怎么样的情况；停留小于3秒的用户特征，并且要分析是否有作弊或刷量的可能性。

又或是，比如电商平台注重交易额，在真正达成交易之前，用户要打开APP、选择商品、确认订单、支付订单等整个交流漏斗模型。每一个环节的关键指标都可以通过公式的形式进行拆解，在根据拆解公式逐个分析对应的影响因素。

### 4. 指标宣贯、存档、落地

宣贯：就是在完成整个指标体系搭建后，要当面告知所有相关业务人员，最好开会并邮件。一方面为下一步工作做铺垫，另一方面是为了让所有相关人员知晓已完成，以防甩锅

存档：对指标口径也业务逻辑进行详细的描述存档，如xxx功能如渗透率=该功能的日点击人数/日活。只有完成这一步，之后的人在查阅时才能看懂是什么意思

落地：就是建立核心指标的相关报表，实际工作中，报表会在埋点前建好的，这样的话一旦版本上线就能立刻看到数据，而且也比较容易发现问题

整个指标体系的搭建更多工作是由数据分析师来完成的，产品经理需要配合分析师选择并确认指标，这也是在建立之初最重要的一点。有的公司没有分析师这个岗位，就需要其他技术同时来配合完成了。



### PP指标体系案例

#### 1. 当前业务发展阶段

知乎当前处于业务发展期和成熟期之间，有2个论点：当前知乎的业务正在一个快速调整期间，内容向娱乐大众化转型；商业化进行较大的探索，但不是做的很重，也就是说无论是最核心的内容还是商业模式，都在探索当中。

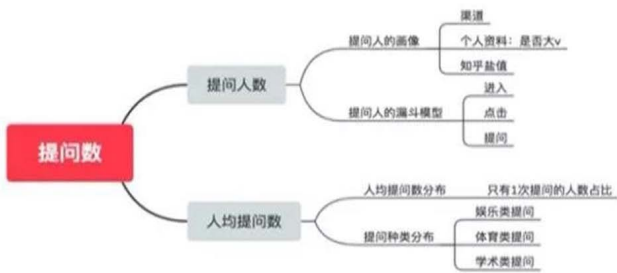
在工作中，这块内容可通过每年的业务规划了解。



#### 2. 核心指标拆解

我们单就首页推荐这个子模块进行分析，对于首页推荐子产品，最重要的指标就是问答数。

问答数=提问数+回答数=提问人数\*人均提问数+回答人数\*人均回答数



在整个过程都是要跟业务进行大量的交流，既不能被业务牵着走，也不能脱离业务，同时要有自己的独立思考。

有些同学会觉得评论、点赞、收藏数应该也是核心指标，而实际上：

· 评论点赞多跟产品的健康度没有直接关系，评论点赞多的本质原因是因为提问回答的内容比较精彩，这是一个相关性而不是因果性关系

· 很多做内容的同学会觉得评论很重要，只要评论做上去了，日活就能涨上去，数据相关性上是这样的，但业务逻辑性不对

所以，评论、点赞、收藏这些都是一个二级功能，更底层的理解实际上是增加APP的社交属性。

#### 3. 会议、存档、建表

会议：产品（负责使用）、研发（负责打点）

存档：对不太好理解的指标要进行单独的解释，比如什么是日活

建表：确定好打点之后，就要建表，确保数据第一时间出来，能及时获得数据反馈并发现问题

字段	业务语言	技术语言	业务人员	研发人员
是否大V	被关注人数超1000	Action日志中 Follow_uv>10000	小A	小B

#### 总结

指标体系建立本质上是考验相关业务人员和分析师的逻辑性，不同业务阶段的指标体系不一样，并且在选择核心指标时一定要正确。在做核心指标的拆解时，通用模式就是先公式拆解，在按业务模块、路径来划分。



# / 数据分析应用实战：阳台能晒多长时间太阳？ /

作者 / 广东 CPDA 数据分析师 刘程浩 编辑 / 协会会员处 李苗苗 日期 / 2020-08

前段时间比较忙，我和老婆差不多在第N时间才加入现在所在小区的住户微信群。感慨一下，现在不同于几十年前住单位福利分房了，那时候串门联络感情非常方便。那时候大家都在一个大单位工作，住的都是同一个单位的单身汉或者家属，平时白天都可以不用关门，打招呼串门很方便。而且那个年代职工新搬进单位的宿舍，提着一篮子熟鸡蛋遛一圈，就能把所有邻居都认识了。

但在商品房+微信的时代，邻居们都来自五湖四海，平时都不怎么见面。要想熟络一下，认识认识同一楼层的邻居都不一定有时间。但我发现，微信群里共同聊一个话题，倒是个不错的熟络方法。

今天上班路上，就看到群里正在热烈讨论一个话题，“我家的阳台能不能晒到太阳，或能晒多久的太阳？”如下图，里面不乏许多大神级的解读。



阳台要有太阳晒，这是个很老题材的话题了，其实在看房阶段，决策是否要买这个单元之前，怕是大家都已经在楼盘工地、样板间里来回穿梭，考察好几十回了。但由于很多房子是预售，看的时候还是工地，真正搬进去的房子在那个时候不仅进不去，且房子周围都是脚手架和堆满的建材，比较危险。因此很难真实的考察到很多细节，最多远远地看一下楼距，估计一下自己看中的房子有多高。

如果说一开始买的楼层很高，或者买的是小区边上户型，那这个问题就不是问题，因为确实没有小区内其他的楼宇遮挡自己的阳台采光。而其他楼层的户型就不同了，如果来看房的时候赶上阴天，或工地正在施工进不去……所以很大程度就要看经验或者自己预判。

当然了，有门路的可以找熟人咨询；而没有熟人的，那就只好在样板间、售楼大厅研究模型，或者研究下售楼书上的宣传图纸。

等掰着手指可以验房和交房的时候能走进屋里看了，大家会发现很多细节和之前看到的模型和图纸多少有些出入。甚至时间久了，看到真正房子时，却又记不得当时考虑的因素。

因此，要想回答自己家里的阳台一年下来每天能晒个多长时间，还是要重新考虑和计算的。下面介绍3种比较常用的方法：

## 方法一：实地测量和计时法

由于我国在北半球，考虑到每年的夏至、冬至是一年中白天时间最长和最短的两个特殊日子，因此可以在那2天专门在家测算和观察下，太阳有没有照进阳台，照进来多大面积，一共照了多长时间，非常精准。其余的日子里，阳台被光照射的时间就是这2个日子观测到的中间值。非常简单对吧？不过用这种方法，对于大多数年轻人来说，还是有一些比较严格的条件：

1、一般来说你得比较有时间（例如财务自由），或者已经退休在家，总之来说得比较有时间。因为夏至和冬至这2天，遇上周末不用上班的可能性不大。比方说今年的夏至，就是星期一得上班。如果不是财务自由或退休在家，你总不能和老板或者和客户请假，要待在家里看太阳吧？所以，比较有时间的人，才能够用得到这种方法。

2、得天公作美，或者说天气得配合。如果夏至那天刚好遇上大暴雨或阴天，冬至那天刚好下雪或也是多云，你就没法去观察和测算了。再加上这2个时节，下雨或阴天，经常一搞就好几天，你也说不准哪天请假方便。难道你敢请假一请就请个一星期，不至于吧。

3、当然还有一种互联网的方法，那就是家里安装摄像头。很多人的家里都装了摄像头，不过很多人的摄像头都是拍摄室内。因此，除非阳台上装了一个摄像头能够拍摄到一天下来阳台的光照情况，这样就可以通过在家里回看视频统计太阳晒阳台的时间了。

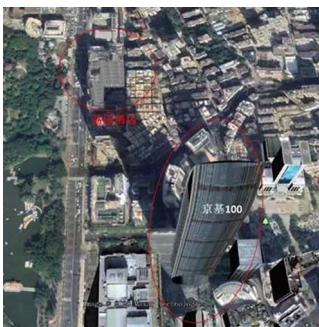
## 方法二：卫星图片观察法

如果你有能看到卫星图片这方面的人脉，那就太方便了。只要调出夏至和冬至的卫星图片（或者之前之后几天的，只要是晴天的都行），看看太阳照射下，自己住的楼层没有被其他楼层的影子遮挡的时间有多长就行了。当然，还有更加先进的图像处理算法，可以在雨天模拟出晴天的效果！这种人脉资源怕是相当的难得，因为一般来说，只能找商业卫星公司。如果你没有这方面的人脉，还有一种大致估计的方法：自己可以下载谷歌地球，在谷歌地球上可以根据时间线找到最靠近夏至和冬至这2天的卫星图片，然后观察下自己住的楼层没有被遮挡就大概知道了。

例如，我找到了深圳的著名建筑京基100，和它旁边的荔



园酒店。这张照片拍摄于2020年3月16日早上8点，从京基100的影子上可以推算在春分8:00的时候，京基100的影子大概率是不会对荔园酒店的底层构成遮挡的。既然春分时节都不会遮挡，那么夏至日早上8:00更加大概率不会遮挡。而冬至日嘛，这个还不太好说。



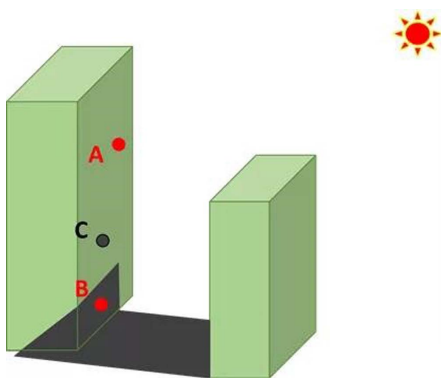
因为这种方法，还是有些问题：

1、谷歌地球的照片不是每天、每时都更新的，因此想要看到更加详尽的照片，特别是具体某一点某一时刻的照片，真得比较难；

2、这种通过影子的估算和估计，需要读图人有一些比较基础的看图能力和地理学知识。不然，影子往哪里偏移，影子在偏移的过程中该做多少压缩或延伸，如果都没个概念的话就麻烦了。

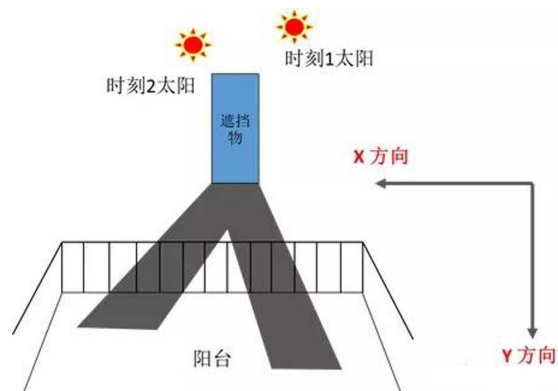
### 方法三：太阳高度角方法

我们考虑一般化场景：如下图，一般来说如果能够住在左边建筑的A点或以上位置就是不错的，因为A点不容易被对面建筑的影子遮挡，而住在B点的话会被遮挡。但对于住在C点而言，到底会不会晒到太阳呢？



为此，如果选用太阳高度角这种方法就会比较复杂，需要一定的地理学或天文学的知识，以及几何、三角函数的基础。所以，我建议只要时间充足，可以考虑方法一；家里有矿可以考虑方法二；像我这样平时比较忙，又只能搬砖挖矿的，就倾向于方案三了。

回到正题，方法三的思路从下图看出要从2个切入点入手：



假设你站在阳台上，阳台对面有个遮挡物会造成阳光的遮挡，从而在阳台上落下影子；不同时刻的太阳位置，落在阳台上的影子的位置是不一样的；

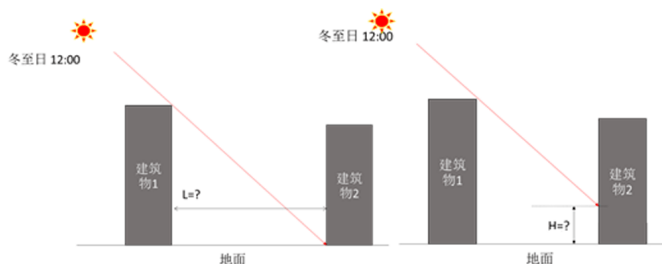
第一个切入点，Y方向作为0-1的先决条件：由于阳台比较窄，因此只要阳台出现影子，就说明被遮挡1，反之就不被遮挡0；第二个切入点，一旦阳台被遮挡，那么X方向上最长的影子移动距离是多少？

下面先研究Y方向，阳台是否会被遮挡

从中学的地理知识可以知道，北半球两栋建筑物之间的距离，至少要保证在冬至日的正午时刻，相互间不产生阳光遮挡。为啥这么说呢？

第一、正午时刻太阳高度角在一天中最大，此时影子最短。如果正午时刻对方建筑被自己的影子所遮挡，那么当天其他时刻，自己的影子都大概率会遮挡住对方（考虑到对方建筑可能较宽，所以用“大概率”）；

第二、冬至日那天，太阳直射点在南回归线，北半球太阳高度角在一年中所有的正午时刻中是最低的。如果这一天的正午对方建筑被自己的影子所遮挡，那么一年到头所有的日子，大概率下自己的影子都会遮挡住对方。用下图来表示就是这样：



在上图左中，冬至日12:00时，如果那条红色的光线刚好照到了建筑物2的墙角，此时两栋建筑物的距离L，就刚好是建筑物2不被对面建筑1影子遮挡的最小距离。这个时候，只

要建筑物2距离建筑物1的水平距离 $\geq L$ ，那么至少全年的正午时刻，在建筑2门口肯定能晒到太阳。

可是现实中很多时候，如上图右，开发商为了节省用地，两栋建筑物之间的距离会小于L。或者说考虑到1楼和2楼作为入户大堂不住人，照不照太阳影响不大，因此允许两栋楼之间的距离小于L。

此时，上图右的问题就变成H=?时，在这个高度以下的建筑物长时间晒不到太阳。

为了回答L=? 以及H=? 我们用一个场景进行建模，选择下图中建筑1和建筑2作为研究对象

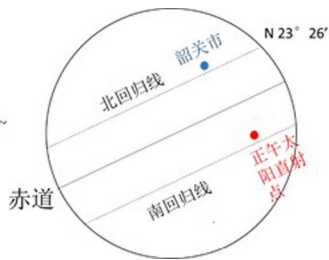


第一步，需要判断的是，太阳到底出现在建筑1一侧，还是建筑2一侧。因为太阳的位置将决定谁会被遮挡住。

这一步的判断其实不难。

首先看上图，由于是拍摄于2019年11月8日早上11:00，此时建筑物的影子方向朝北，此时说明太阳就在南方，因此太阳的位置就在建筑1那侧。

韶关经纬度:  
113.8°E (东经113.8°)  
24.8°N (北纬24.8°)  
呵呵!这个初二就会拉,本人高中地理班的哦~  
相信我!  
查看更多步骤...  
韶关的经纬度是多少?\_百度知道  
zhidao.baidu.com



还有一种方法就是考察纬度的因素。

我查了一下所在的韶关市市区的地理坐标，大致是 N 24.8°，E113.8°。

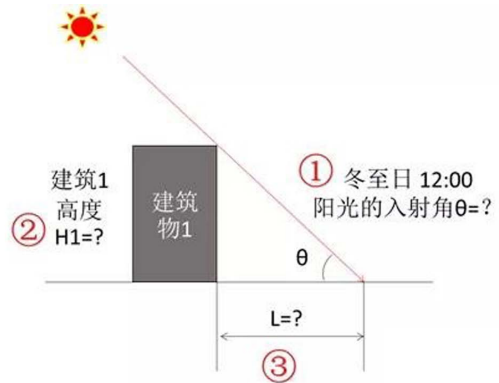
由于在11月份，太阳的直射点在南半球，而韶关市在北回归线以北，因此也可以看出，太阳靠南，在建筑1那侧。

这样一来就清楚了，建筑1和建筑2之间，我们要考察的是建筑2是否会被建筑1的影子所遮挡。

第二步，简化建模，考察一下建筑1和建筑2之间的距离L是否达到了互不遮挡的最小值?

这时候，我们将建模所需要的参数，放到图中来看就比

较清楚了。



上图中一共有3个参数：

参数1，冬至日12:00，阳光的高度角 $\theta=?$

参数2，建筑1的高度  $H1=?$

要求出的参数3：此时建筑1的影子长度（或者最小楼距） $L=?$

首先求 $\theta$ ，这个容易求出来  $\theta = 41.7^\circ$ 。一般的，天文学中关于某天某时刻的太阳高度角的计算公式是这样：

$$\theta = \arcsin(\sin(Dec)\sin(Lat) + \cos(Dec)\cos(Lat)\cos(t))$$

上面的等式中，Dec：赤纬，天文学术语，指的是太阳直射点的纬度，如果在南半球则取负数。冬至日太阳直射点在南回归线，也就是取 $Dec = -23.5^\circ$

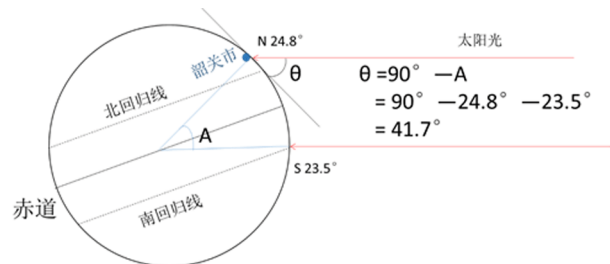
Lat：观测点所在的纬度，例如韶关的纬度就是 $N 24.8^\circ$ 。

t：时角，天文学术语，解释起来有些费劲，百度百科是这么解释的：天体的时角定义为天子午圈与天体的赤经圈在北极所成的球面角。换成不太严谨但容易理解的话说，就是地球每小时自转 $15^\circ$ ，在观测时点上，地球已经自转了多少个 $15^\circ$ 。

计算公式是  $t = (\text{观测时点} - 12) \times 15^\circ$

例如，现在是正午12点，那么时角就是 $0^\circ$ ，之所以要减12，那是因为规定了正午12点的视角为 $0^\circ$ 的缘故。不过由于取的是正午时间，因此 $\cos t = 1$ ，整个公式就简化了。经过数学变形，得到冬至日正午时刻韶关市区的太阳高度角为 $\theta = \pi/2 - (Lat - Dec)$ ，也就是 $41.7^\circ$ 。

用高中课本中简单的平面几何公式就可求出。



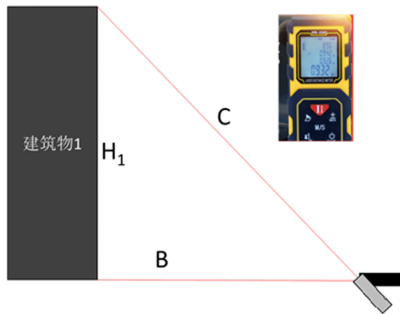


其次求H1

求H1的方法有很多种。高手都在民间，所以我就用了一种相对容易的方法。

工具是激光测距仪，原理用的是勾股定理。原理和实操都很简单，下图大家一看就秒懂

$$H_1 = \sqrt{C^2 - B^2}$$



求出了θ, H1后, 那么L就容易求了

利用公式:  $L = \frac{H_1}{\tan \theta}$  代入数值求出L;

在进入下一个环节之前, 我们不妨根据一般性推导出一个临界公式, 也就是测算当在建筑2所住的楼层阳台顶部外沿高度要达到的临界值 (见下图)。

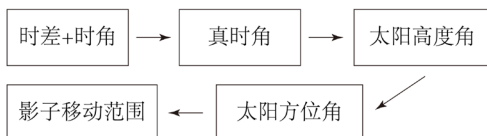


经测算, 一般情况下, H的临界值为  $H_{\text{临界}} = H_1 - L_1 \times \tan(\theta)$

当阳台顶部外沿实际高度高于临界值H临界时, 至少能保证每天正午能晒到太阳 (也就是Y方向判定不被遮挡), 此时可以测算出在冬至日那天能晒到多久的太阳; 当阳台顶部外沿实际高度低于H临界时 (Y方向上判定被遮挡), 那就一年到头大概率会晒不到太阳。

下面研究X方向上, 影子横向移动多宽

这个就比较需要系统的立体几何以及天文学的基础知识了。总体的研究思路流程如下:



上面的流程中, 有一些名词已经之前解释过了, 这里不多做解释; 下面讲一下没介绍过的。

1、时差: 不是旅游倒时差的那个时差, 而是标准时区的经度和观测地的经度之间存在角度距离。例如北京时间所在经度为E 120°, 而韶关经度为E113.6°, 相隔了6.4°, 从时间上看差了大约半个小时。也就是说当北京时间早上6:00的时候, 韶关真正的时间是早上5:30多一些。之前写太阳高度角的时候, 按道理也要考虑这个因素的, 但一方面前面内容不用涉及到太阳升起和日落, 另一方面误差也不大, 因此就没在之前考虑了。时差的计算公式为:

$$\text{时差} = \frac{\text{观测点经度} - 120^\circ}{15}$$

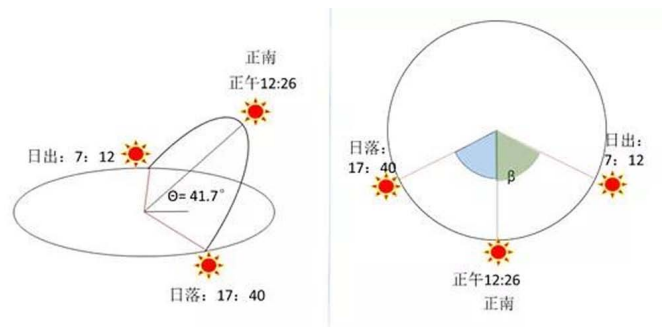
2. 真时角: 就是“时角+时差”

3. 太阳方位角: 天文术语, 按百度百科的解释, 太阳方位角通常被定义为从北方沿着地平线顺时针量度的角。它的计算公式为:

$$\beta = \arccos \left( \frac{\sin(\text{Dec}) - \sin(\text{Lat}) \sin(\theta)}{\cos(\text{Lat}) \cos(\theta)} \right)$$

这个看着不太好理解, 下面我画个图大家就容易理解了。

首先, 由于存在时差, 韶关的冬至日是要到12:26分, 太阳才到达正南方, 此时影子最短。此时的太阳高度角为41.7°, 太阳位于正南方, 此时方位角为β= 180°; 其次, 同样因存在时差, 冬至日早晨是7:12分日出, 17:40分日落。经计算当天太阳是在正南方偏东约64°升起, 也就是俗话说的约东南方升起 (此时β=-116°, 负数代表上午。但是站在观测者的角度上看, 负数度数看着很别扭, 所以取|β|的补角64°, 就会比较自然和生活化); 同理, 当天太阳在正南方向偏西约64°的方向落下, 即俗话说的约西南方向落下。



太阳方位角的变化, 就决定了影子的活动范围

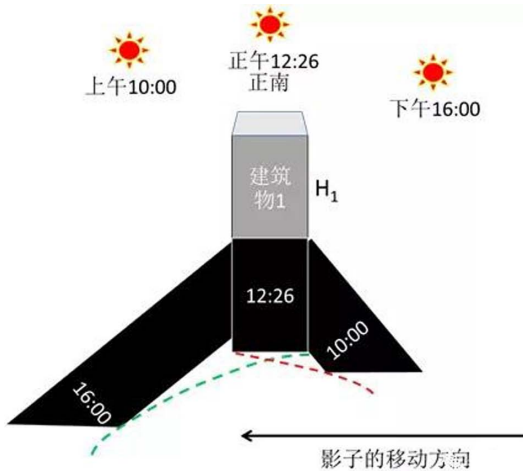
4. 影子活动范围, 指的是太阳在不同的方位角出现时, 影子的位置。在这里, 我们需要确定一个时间范畴。理论上一旦太阳出现在地平线之上, 只要有遮挡物就一定会有影子, 但有些时段的影子我们并不care。例如我们想晒衣服, 早上10点之前的太阳, 阳光并不强烈, 此时晾干的贡献比晒干还大;

亦或者傍晚时的太阳，也是如此。只有日上三竿到正午，正午到下午那段时间的太阳是我们生活中所需要的。所以我们想要研究阳台能否晒到太阳或者能被太阳晒多久，应该取一个我们都常用来晾晒的时段，在这里我就取上午10:00到下午16:00。

我计算了一个韶关市区冬至日不同时段的太阳高度角-方位角一览表。一起来看一下（第二列粘贴时发生错误，请忽略）

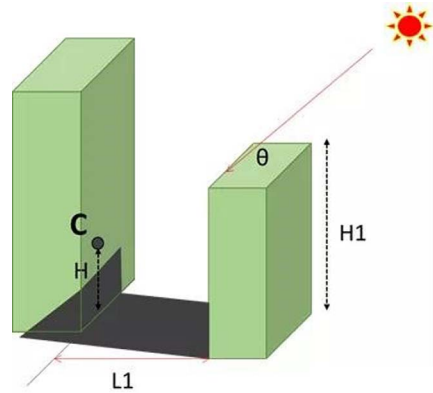
时点	太阳高度角 $\theta$	生活化的方位角
10:00	1. 18°	南偏东 39.15°
11:00	1. 46°	南偏东 25.06°
12:00	1. 36°	南偏东 7.95°
12:26	1. 7°	正南方
13:00	1. 07°	南偏西 10.36°
14:00	1. 66°	南偏西 27.13°
15:00	1. 01°	南偏西 40.78°
16:00	1. 19°	南偏西 51.35°

有了这个表，并基于表中的数据我们就可以制作一下3个关键时点的影子变化边界，我将这3个时点的影子的边界连接上去后，得到了下图的2条虚线。



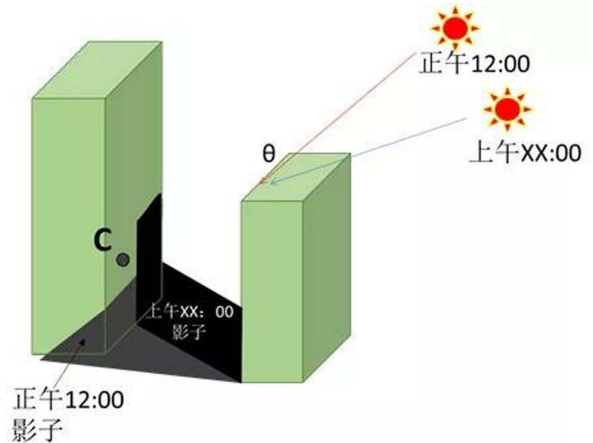
如果你住在1楼，若从10:00到12:26，阳台位于红色虚线之下（外），或者下午12:26-16:00，阳台位于绿色虚线之下（外），那么阳台都是可以晒到太阳的。1楼晒得到太阳，那么1楼以上的住户肯定全都晒得了太阳。但实际上经常是这种场景，那就是受之前所说的楼距过小的影响，没办法住在上面2条虚线范围之外。这个时候我们就回到之前讲到的一般化场景。

我们想知道下图中C点（假设为阳台的顶部），从10:00到16:00之间是否能够晒到太阳？这个就要花点时间来测算了。

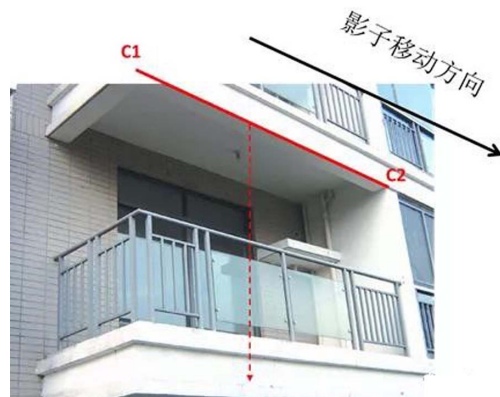


关于C点能否在冬至日正午晒到太阳，从前面推导出来的先决条件H在正午时刻的临界高度值 $H_{\text{临界正午}} = H1 - L1 \times \tan(\theta_{\text{正午}})$  计算出的 $H_{\text{临界正午}}$ ，如果这个值 $\leq H$ ，那么C点在冬至日正午肯定可以晒到太阳。至于早上10点到下午16:00能否都能晒到太阳，则还需进一步分场景计算。

有一点可以肯定的就是，一天中其他时间点的太阳高度角肯定比正午时 $\theta$ 值要低，这就意味着除正午外其他时间点的影子上沿高度肯定超过 $H_{\text{临界正午}}$ 值。这个是显而易见的，如下图：影子越长，一旦影子受到遮挡，影子的上沿就越高。



于是，我们得到一个判断C点是否被影子遮挡的临界条件状态，不过所需要要换成具体的场景来看。首先C点是阳台的天花完整外沿，它是有长度的。我们设左侧为C1，右侧为C2，每天影子的变化方向是由C1→C2，也就是自西向东。

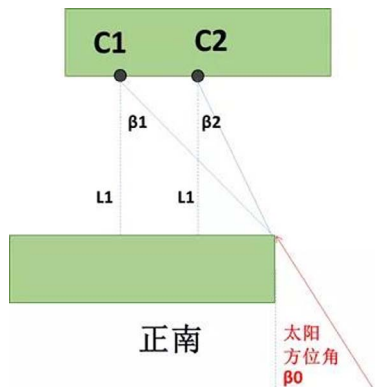




另外，当我们把分析的视角从轴测图转向平面图时，就很容易分析了。下图中，是代表冬至日当天，上午10:00到正午12:26分时，需要测量的量包括如下：

上午的太阳方位角 $\beta_0$

C1点和对面建筑外立面构成一个夹角 $\beta_1$ ，C2点和对面建筑外立面构成一个夹角 $\beta_2$ ，由下图易知， $\beta_1 > \beta_2$



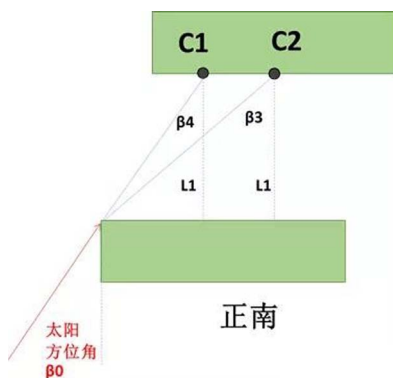
$\beta_0$ 、 $\beta_1$ 、 $\beta_2$ 之间存在如下关系：

当 $\beta_0 > \beta_1$ 时，整个C点阳台不受遮挡；

当 $\beta_2 < \beta_0 \leq \beta_1$ 时，C点阳台开始被遮挡或部分被遮挡，此时还要兼顾H和H临界。这个时候，如果在此期间 $H \leq H$ 临界，只要逆运算 $\beta_1$ 和 $\beta_2$ 分别等于太阳方位角时的时刻，就知道C点阳台被影子遮挡的时间；如果此期间 $H \geq H$ 临界，则C点阳台完全不受遮挡；

当 $\beta_0 \leq \beta_2$ 时，如果在此期间 $H \leq H$ 临界，C点阳台完全受到遮挡；如果此期间 $H \geq H$ 临界，则C点阳台完全不受遮挡；

同样的，如果是在下午这段时间内



$\beta_0$ 、 $\beta_3$ 、 $\beta_4$ 之间存在如下关系：

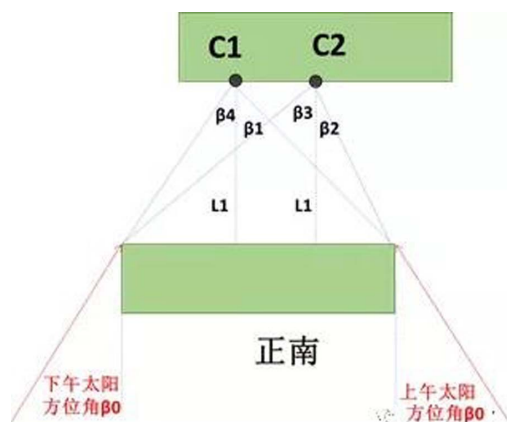
当 $\beta_0 > \beta_3$ 时，整个C点阳台不受遮挡；

当 $\beta_4 < \beta_0 \leq \beta_3$ 时，C点阳台开始被遮挡或部分被遮挡，此时还要兼顾H和H临界。这个时候，如果在此期间 $H \leq H$ 临界，只要逆运算 $\beta_3$ 和 $\beta_4$ 分别等于太阳方位角时的时刻，就知道C点阳台被影子遮挡的时间；如果此期间 $H \geq H$ 临界，则C点阳台完全不受遮挡；

当 $\beta_0 \leq \beta_4$ 时，如果在此期间 $H \leq H$ 临界，C点阳台完全受到遮挡；如果此期间 $H \geq H$ 临界，则C点阳台完全不受遮挡。

分析了那么久，举个实例分析下：

假设小王住在在韶州市某小区，他所住的楼层阳台顶部外沿距离地面28米，他的阳台的南面斜侧，有另一栋小区住宅楼，楼高40米。小王所住的楼和阳台对面的楼楼距为 20米，对面楼的立面和小王阳台顶部外沿的两端，形成的夹角如下图：



其中， $\beta_1=47^\circ$ ， $\beta_2=25^\circ$ ， $\beta_3=52^\circ$ ， $\beta_4=29^\circ$

请分析，假设冬至日全天晴朗，小王的阳台在冬至日那天能够照射到多久的太阳？

解答：

韶州市冬至日正午高度角为 $41.7^\circ$ 为已知。

首先判断实际楼距是否达到理论不遮挡的最小楼距。理论楼距 $L = 40 / \tan(41.7^\circ) = 43.88$ 米；而小王所住楼和阳台对面楼实际楼距只有20米，因此必然小王所在的楼宇的低楼层必然长期会受到影子遮挡；

接下来，要看看小王坐住的楼层阳台的遮挡情况；

前面计算冬至日那天不同时点的太阳高度角、方位角、临界高度等等概念不再一一赘述，直接上图表进行分析：

时刻 <sup>o</sup>	生活中的方位角 <sup>o</sup>	H 临界 <sup>o</sup>	阳台外沿高度 <sup>o</sup>	$\beta_1$ <sup>o</sup>	$\beta_2$ <sup>o</sup>	$\beta_3$ <sup>o</sup>	$\beta_4$ <sup>o</sup>	判断 <sup>o</sup>
日出 7:12 <sup>o</sup>	64.02	40.00 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	不受遮挡 <sup>o</sup>
7:30 <sup>o</sup>	62.09	38.74 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	不受遮挡 <sup>o</sup>
8:00 <sup>o</sup>	58.56	36.65 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	不受遮挡 <sup>o</sup>
8:30 <sup>o</sup>	54.59	34.57 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	不受遮挡 <sup>o</sup>
9:00 <sup>o</sup>	50.09	32.49 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	不受遮挡 <sup>o</sup>
9:30 <sup>o</sup>	44.98	30.41 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	被遮挡 <sup>o</sup>
10:00 <sup>o</sup>	39.15	28.37 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	被遮挡 <sup>o</sup>
10:30 <sup>o</sup>	32.52	26.42 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	不受遮挡 <sup>o</sup>
11:00 <sup>o</sup>	25.06	24.68 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	不受遮挡 <sup>o</sup>
11:30 <sup>o</sup>	16.82	23.28 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	不受遮挡 <sup>o</sup>
12:00 <sup>o</sup>	7.95	22.39 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	不受遮挡 <sup>o</sup>
12:30 <sup>o</sup>	1.23	22.14 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	不受遮挡 <sup>o</sup>
13:00 <sup>o</sup>	10.36	22.57 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	不受遮挡 <sup>o</sup>
13:30 <sup>o</sup>	19.09	23.61 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	不受遮挡 <sup>o</sup>
14:00 <sup>o</sup>	27.13	25.12 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	不受遮挡 <sup>o</sup>
14:30 <sup>o</sup>	34.37	26.93 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	不受遮挡 <sup>o</sup>
15:00 <sup>o</sup>	40.78	28.91 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	被遮挡 <sup>o</sup>
15:30 <sup>o</sup>	46.41	30.96 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	被遮挡 <sup>o</sup>
16:00 <sup>o</sup>	51.35	33.04 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	被遮挡 <sup>o</sup>
16:30 <sup>o</sup>	55.69	35.12 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	不受遮挡 <sup>o</sup>
17:00 <sup>o</sup>	59.54	37.20 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	不受遮挡 <sup>o</sup>
17:30 <sup>o</sup>	62.97	39.31 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	不受遮挡 <sup>o</sup>
日落 17:40 <sup>o</sup>	64.01	40.00 <sup>o</sup>	28 <sup>o</sup>	47 <sup>o</sup>	25 <sup>o</sup>	52 <sup>o</sup>	29 <sup>o</sup>	不受遮挡 <sup>o</sup>

在上表的最右列，我用了if...else，或者IF()的函数，很容易就判断出了小王阳台被遮挡的时间段。上表这5个标红的时间段就是阳台被遮挡，当然我在这里用了半小时为间隔，如果用10分钟或者更细的话，可以计算得更细致一些。

根据之前设置的上午10:00到下午16:00的有效照射时段来看，小王的阳台在上午不到半小时，下午有1个半小时是会有太阳遮挡的。总计被遮挡的时间有2小时左右，占整个6小时有效时段的1/3，可以说也还是不错的了。是不是很方便？

小结：上面讲了长长的分析过程，看的让人眼花缭乱，也有可能让人头晕脑胀。下面需要做一下总结：

第一步：如果你时间充裕，或者多装几个摄像头，通过记录任何日子下的阳台被太阳光照时间，了解阳台是否被遮挡；或者，如果你有商业卫星公司的资源，可以查看特定日子，例如冬至日及前后几天的卫星图片，也可以准确的知道在一年中白昼最短的日子里，阳台能够被太阳光照的时间；其他日期里，阳台受光照的时间只会比冬至日久。如果以上两个条件都不具备，那么可以考虑烧脑的下一步。

第二步：考虑太阳高度角及太阳方位角，测算冬至日当天阳台被有效光照的时间。以下是动作分解：

1、通过百度地图或谷歌地图、高德地图……，精确测量所在小区的经度和纬度。这次我用了韶关市区的经纬度，主要是为了取整方便；

2、得到上面的数据后，利用赤纬、时角、时差分别计算出冬至日当天各个整点时段的太阳高度角 $\theta$ 、太阳方位角 $\beta$ 。做好做一个表格，方便对比；

3、测量自己住的楼房和对面楼房的楼距、对面楼房的高度、阳台外沿高度、阳台外沿两端和对面楼房形成的夹角 $\beta_1$ 、 $\beta_2$ 、 $\beta_3$ 、 $\beta_4$ ；

4、上一步测量完之后，计算阳台外沿高度的临界值H临界；并根据第2步得到的各时点的太阳方位角 $\beta$ ，和 $\beta_1$ 、 $\beta_2$ 、 $\beta_3$ 、 $\beta_4$ 进行对比，判断阳台是否受到遮挡，如果受到遮挡，那么反算出被遮挡的时间段；

5、利用第2步得到的各时点的太阳高度角 $\theta$ ，计算各时段影子上沿在自己住的楼房外立面的高度H，并用之与H临界进行对比，修正第4步的结果。

## / SPSS单样本非参数检验——卡方检验 /

作者 / 河北 CPDA数据分析师 曾洲 编辑 / 协会会员处 李苗苗 日期 / 2020-08

今天我们聊聊非参数检验，之前给大家介绍过单样本非参数检验，今天我们一起来看一下单样本非参数检验中的一种——卡方检验。

卡方检验一般是用来对样本总体分布进行检验的非参数检验方法。比如说人的血型和性格是否有关系，如今的人口结构和很多年前是否还是一样等等情况都可以使用卡方检验进行分析。

我们先来看一下卡方检验的原假设，也就是 $H_0$ 是什么：

$H_0$ ：样本总体的分布与假设的分布（也称期望分布或者理论分布）无显著性差异。

了解了卡方检验的原假设，我们再来看一下卡方检验的基本思想以及理论依据：如果一个随机变量X所在的样本总体中随机抽取任意数量的检验样本，这些抽取出来的检验样本落在X的k个互不相交的子集中的观测频数服从一个多项分布，这个多项分布在k趋于无穷时近似服从卡方分布。所以，我们在这个理论下，就可以对变量X的总体分布的检验从对各个观测频数的分析入手。

在分析的时候我们一般是需要得出Pearson卡方统计量。如果卡方值较大，说明期望频数与观测频数分布差距较大，这时候就需要拒绝原假设；反之，我们就需要接受原假设。

现在我们来看一下本次分析需要用到的数据：

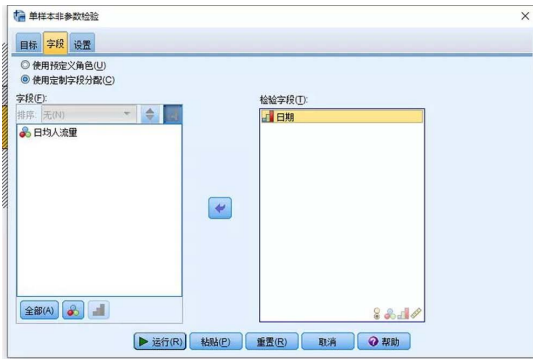
日期	日均人流量
星期日	44.00
星期一	78.00
星期二	90.00
星期三	94.00
星期四	89.00
星期五	110.00
星期六	84.00

上图中的数据是某个医院的人流量相关数据，我们需要对一周内的日均人流量进行分析，了解是否一周中每天的病人流量是相同的。

在进行分析时，我们首先要对日均人流量数值进行加权处理：



然后我们选择非参数检验中的单样本：



我们在字段选项中将日期选入检验字段中，设置选项中选择卡方检验。最后我们点击运行即可得出本次分析的结果：

假设检验汇总

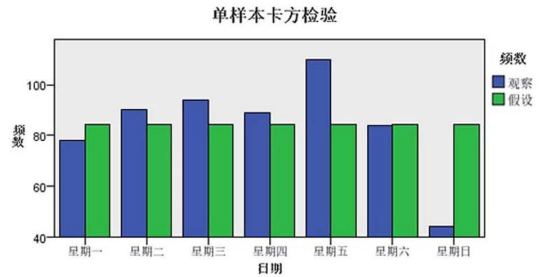
原假设	测试	显著性水平	决策者
1 日期的类别以相同概率发生。	单样本卡方检验	.000	拒绝原假设。

显示渐进显著性。显著性水平是 .05。

在结果中我们可以看到本次检验的原假设、检验方法、显著性水平P值以及最后的决策。本次分析中显著性P值为

0.000，小于显著性水平0.05，说明一周中每天的人流量是不一致的。

我们双击结果图，可以进入到模型中看更加详细的分析结果：



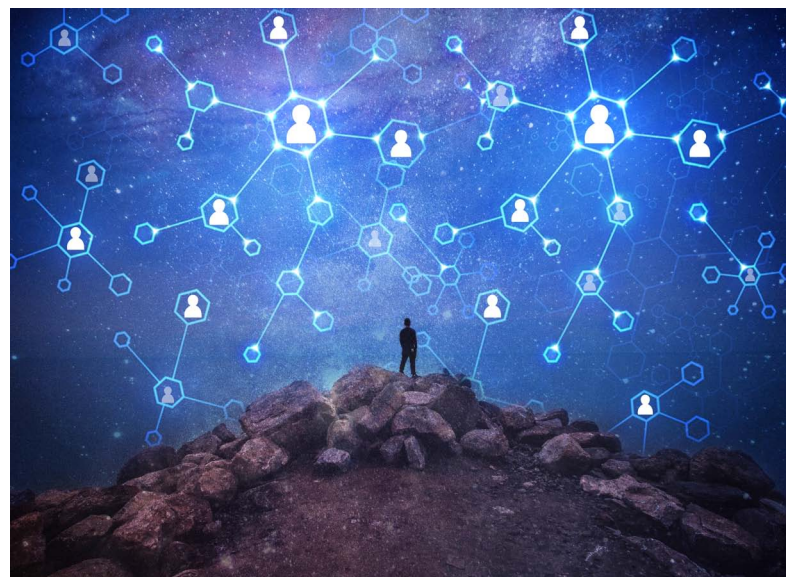
总计 N	589
检验统计量	29.389
自由度	6
渐进显著性 (2-sided 检验)	.000

1. 有 0 个单元格 (0%) 的期望值小于 5。最小期望值为 84.143

从上图可以看到，卡方统计量为29.389，显著性P值为0.000，小于显著性水平0.05，说明通过了显著性检验，也就是说一周中每天的人流量有显著差异。

同时我们还可以看出在周五的时候人流量最多，周日的时候人流量最少（周末医院门诊较少，人流也会降低），其余几天的人流量差别不是特别大。

还有一点需要大家注意的是当我们进入到模型中以后，我们还可以通过把鼠标放在聚类条形图上来了解各个字段的频率及差别（残差）。





# / 机器学习在传统服装行业的尝试 /

作者 / 成都 CPDA学员 兰将军 编辑 / 协会会员处 李苗苗 日期 / 2020-08

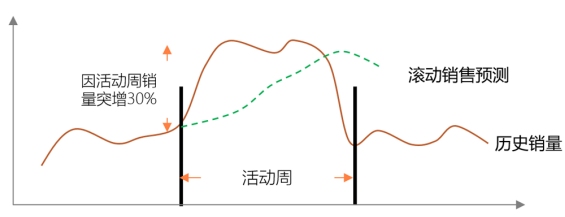
我在传统服装行业深耕有10余年了，最近几年才开始慢慢接触机器学习及大数据分析等领域，也有幸接触了公司涉及这些方面的项目，并结实了一群共同学习CPDA的同仁们。在此给大家分享一下机器学习在传统服装行业门店端销售预测所带来的价值和面临的挑战。

首先我们来描述一个业务场景，公司4-5个人管理400-500家门店的补货，门店经常出现欠品或者爆仓的情况。我们知道商品的欠品会直接带来销售损失；同样爆仓的情况会增加门店仓库压力，影响卖场环境体验，如何高效地完成门店商品补货就是急需解决的问题。

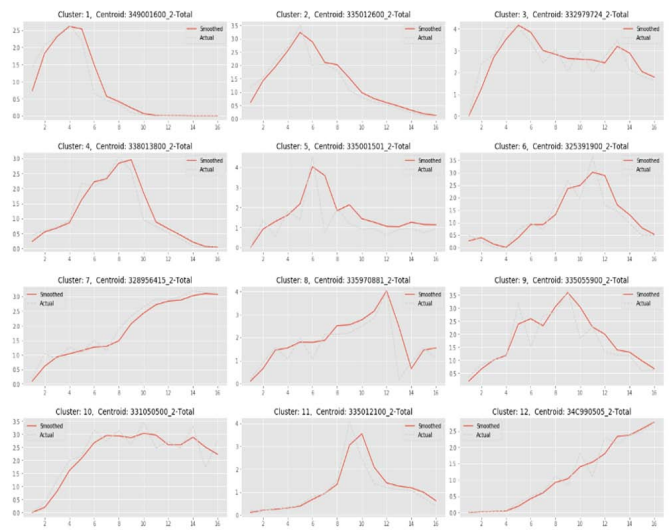
我们先来看传统情况下我们的商品补货逻辑：

计算门店SKU过去15天的平均历史销量加上陈列量通过保证商品周转天来进行补货，此销售预测方法的特点是简单、易懂、完全自动化，能够快速响应近期销售变动，更加灵敏。此种方法会带来以下的影响：

1、促销活动前，无法提前考虑促销活动的影响，销售预测过低，有滞后性；促销活动后，无法自动剔除促销活动的影响，销售预测过高，有滞后性。



2、无法考虑产品销售曲线，始终存在系统性偏差，销售曲线呈上升趋势时预测过低，销售曲线呈下降趋势时预测过高。



这种原始的补货方法在多个业务场景面临很多挑战，比如主推商品、商品调价、团购活动、进入活动周、活动周结束等等特殊情况，现在业务面对这些只能进行手工调整安全周转天数，去修正系统最终补货量。

以商品降价为例，用户所需操作如下：

商品降价开始前手动调高各门店-SKU的安全周转天数，增大补货量；商品降价结束后手动调低各门店-SKU安全周转天数，减少补货量；商品回到正常销售后手动将各门店-SKU安全周转天数调回常规水平。该方法虽然可以从后端修正最终补货量，但调整频繁、操作繁琐，效率低；天数的调整依靠经验、无数据支持，无法保证有效性，效力低。

那么我们来通过机器学习怎么来进行门店销售的预测呢？

首先我们来考虑影响门店销售的一些因素：

1、商品的生命周期；2、商品的调价；3、主推商品；4、活动周与非活动周；5、天气及温度对销售的影响。影响门店销售的因素如此之多，那么我们如何来保证我们商品不缺货不爆仓呢？我们先从最重要的销售预测开始。以下是销售预测训练的模型结构：

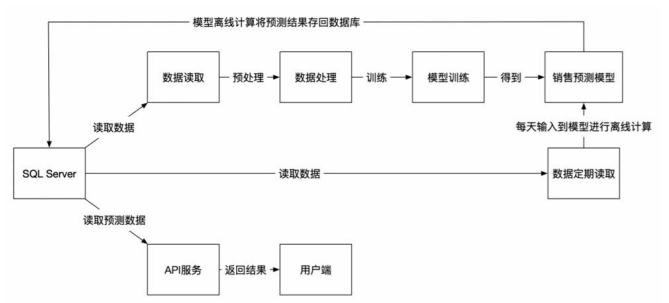
系统总体 (如图)分为三个部分：

模型训练模块：模型通过读取 SQL Server 数据，将门店销售数据进行预处理，导入到模型中进行训练，得出销售预测模型，以便后续使用。

模型离线预测模块：系统会每天从 SQL Server 读取新增的销售数据，输入到销售预测模型中，预测出之后两周的销售数据，并将结果存储到数据库中，以便之后使用。

API 服务模块：当 API 模块收到用户请求时，通过读取数据库相应信息，

返回模型预测结果给用户端。用户端即可使用该接口进行销量预测。



我们使用了XgBoot基础模型，通过AutoML技术进行自动超参数调优，找出最优化的模型结构和参数，来对未来两周的销售数量进行预测。为什么选择两周销量我们主要是考虑了门

店与仓库的距离，还有仓储的运量进行评估。

算法的实现我们主要考虑了以下的一些内容：

构造训练数据：使用过去70天的交易数据构造训练数据，使得无销量样本和有销量样本比例近似为1:1。

构造特征：主要考虑了门店特征：例如门店面积、门店口岸、所在城市等；商品特征：例如商品品类、面料、颜色、吊牌价等；天气特征：是否有雨、温度等；促销折扣特征：例如促销类型、折扣力度。

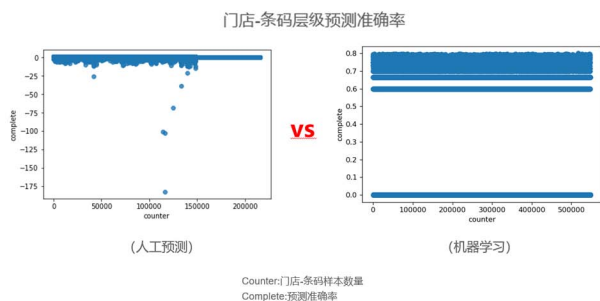
模型训练：模型 `xgboost.XGBRegressor()` ,模型主要参数暂略。

测试结果：使用了所有门店销量进行预测,通过和真实销量数据的比对，所有活跃门店每周销量大于2的SKU 准确率在63%~95%之间,总体平均准确率为77%。

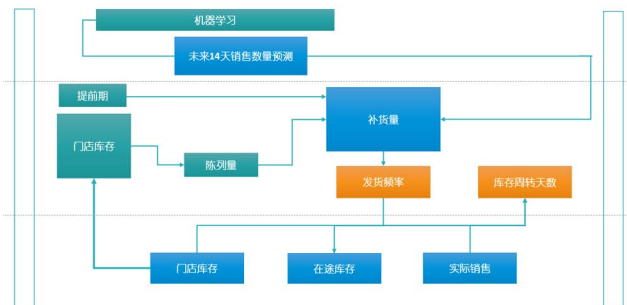
以下为门店未来两周预测的准确率(部分):

shopid	准确率	shopid	准确率	shopid	准确率	shopid	准确率
256	80.52%	1218	76.20%	4893	78.54%	7466	78.86%
262	81.82%	1219	73.65%	4902	78.59%	7477	75.80%
263	72.46%	1220	78.54%	4903	79.50%	7482	75.80%
266	85.93%	1227	68.34%	4996	72.81%	7486	73.84%
268	85.31%	1228	74.31%	4997	78.08%	7487	78.00%
270	84.29%	1230	77.90%	5019	78.90%	7492	74.15%
271	85.02%	1236	77.79%	5045	71.22%	7493	74.29%
272	85.87%	1252	74.84%	5085	78.90%	7494	71.49%
274	78.46%	1253	76.98%	5133	73.18%	7495	86.14%
276	73.48%	1257	73.24%	5134	71.41%	7497	75.80%
278	75.50%	1258	78.16%	5135	79.21%	7498	80.53%
280	75.65%	1260	79.00%	5173	78.93%	7500	78.99%
281	82.28%	1261	78.95%	5222	77.03%	7501	76.05%
283	88.13%	1271	80.29%	5262	76.45%	7503	75.80%
284	83.95%	1276	71.71%	5263	76.34%	7504	75.45%
286	89.41%	1277	71.95%	5269	78.18%	7505	72.35%

通过上线测试后，我们发现机器学习相较于人工原始的预测有了较大提升：



我们在完成了较高的准确率预测，我们如何将预测值SKU的销售数量应用至我们的实际补货环境中呢，我们根据我们的实际业务，做了以下流程：



1、销售预测通过机器学习计算出门店-商品条码层级未来14天的销售数量，偏差值为30%-40%。

2、计算门店补货量：

补货量= (陈列量+预测销量) -当前库存 (含在途) 。

3、根据仓库-门店的提前期进行提前补货。

通过运行上线后的观察，我们发现机器学习和原来人工补货有了较大提高：机器学习补货相较于原有补货准确率提升40%左右；机器学习补货方案相较于原有补货方案周转天数降至14天。

经过一段时间的尝试，对门店的销售预测达到了一些预期，并应用到了我们生产业务当中去，但是还有很多机会点等待实现，例如商品仓库补货与门店调拨的最优方案，销售人员能力与客流对销售的影响；还有一些困难点，例如服装的流行趋势，新品上市的表现也是当前模型的盲区；数据的不规范性，给我们数据清洗和聚类分析带来挑战。

在这里我们可以预见服装行业基础资料的规范性和庞大的线上线下数据是可以给业务带来价值的，我们同样期待机器学习在供应链返单中带来更多有价值的机会点。也同样希望身在服装行业的数据分析师们能多交流，为提升传统服装零售行业的发展做出贡献。





## / 数字经济时代科技创新浅谈 /

作者 / 大连 CPDA学员 王森 编辑 / 协会会员处 李苗苗 日期 / 2020-08

在全球新一轮技术革命的推动下，随着大数据、云计算、物联网、区块链等数字科技创新加速和应用加深，我国的数字经济呈现出一日千里的蓬勃发展态势。电子商务、社交媒体、共享出行、餐饮外卖等各类互联网平台企业不断涌现并迅速壮大，成为数字经济时代的典型代表，走在了世界的前列。数字经济的发展不仅深刻影响和改变了传统产业组织和经营模式，而且重构了市场供求体系，优化了资源配置，促进了数字科技企业和传统企业跨界融合创业创新，拓展了消费市场发展空间，创造了更多具有数字经济特征的新就业渠道和就业岗位。这些，均是数字经济科学技术创新所带来的巨大效应。

2019年底爆发的新冠病毒席卷全球，不论是我国，还是欧、美等世界各国、各国际组织，在政治、经济、文化诸方面都面临着史无前例的巨大考验。全球经济按下了暂停键，世界经济急剧衰退，大多数国家宏观经济整体下滑，毫不夸张地说，最凶险的全球经济危机来了。数字经济在这惊涛骇浪的特殊时代，将面临着新的挑战，而机遇，也来了。

### 一、数字经济时代的来临

18世纪末期，英国的第一次工业革命使传统的手工劳动被机械生产所代替，将过去以农业和手工业为基础的经济社会转变为以工业和机械制造带动的经济社会。同时，进一步加快了全球的新一代工业变革。第二次工业革命将工业生产进一步形成生产线生产，开创产品批量生产的新模式。随着科技进步和技术创新，世界经济形势也迎来了第三次工业革命，生产线高度自动化是引线，再结合了电子和信息技术的广泛应用后，制造生产过程不断实现无人操作。机械和自动化设备逐渐替代了人类作业。到现在，我们正在经历第三次工业革命，而随着近年电子信息技术、AI人工智能和大数据技术的兴起，有人便提出在未来10年，“工业4.0”将要来临，而这一概念，最初是在2011年德国举办的“汉诺威工业博览会2011”上被提

出。“工业4.0”在德国被认为是机械化、电气应用、自动化之后的第四次工业革命。而数字经济便是从第三次工业革命向第四次工业革命转动的重要齿轮。

从我们国内的角度来看，数字经济是时代变迁的必然产物，前两次的工业革命，我们没有赶上。在将来到的数字经济时代，我们已做好准备。数字经济是复合资源和高新技术高度融合的新经济模式。改革开放以来，我们不断向西方先进技术和自主创新。在方方面面追逐世界最先进高新技术。同时，为了更好地整合全国的数据资源，国家不断加强全国的数据管理工作；通过部署优势政策和建设各地高新区等办法，加强各地区对高新技术的扶持力度。在数字经济时代中，准确、复合和高价值的数据资源与先进且不断更新的技术和高新技术人才，是数字经济时代的主要抓手。

### 二、数字经济的挑战和机遇

近年来，国家对科技创新和数字化建设的重视程度不断加深。数字经济应运而生，数字经济是数据资源和高新技术高度融合的新经济模式，在数字经济时代，数据资源和原实物财产都是有价资源，数据资源的价值更加不可忽视。另外，高价值数据资源主要掌握的官方机构手中，而技术或技术人才主要集中在企业及高校当中，必须找到恰当的融合方式，利用官方机构资源优势，发挥企业和高校的技术和技术人才的研发和挖掘数据的能力，对数据资源进行分析并得出用于决策的依据和理由。让政府或企业在做规划时，决策更加理性，角度更加全面、预测更符合规律。

数据资源的全球流动及全球化配置特点，技术的不断更新，应用市场的网络化特点，直接打破过去经济发展模式，任何地方都可以发展起来，只要有合适政策及领军人才，都可以汇集数据，发展以数据为基础的数字经济。想要在数字经济时代高速发展，如何合理利用数据资源是至关重要的。



可以说，不论是国家还是地方，掌握了更先进的技术和人才，就会在新的数字经济时代获得优势，而科技创新就是保持优势的最重要的手段。

### 三、智慧城市的推进

智慧城市，是数字经济的重要组成部分，是数字经济科技创新的表象体现。智慧城市的数量和质量，一定程度代表着数字经济的先进程度，也为科技创新提供了集各种有利条件于一体的“孵化器”。

国际/国内	年份	国家（或经济体）	描述
国际	2006	欧盟	欧洲 Living Lab 组织采用新的工具和方法、先进的信息技术调动“集体的智慧和创造力来解决社会问题，并提出了欧洲智慧城市网络。
		新加坡	“智慧国 2015”计划，通过物联网等新一代信息技术，建设经济、社会发展一流国际化城市。
	2008	美国	在纽约招考的外国关系理事会上，IBM 提出了“智慧地球”理念。
			美国迪比克与 IBM 合作，建立美国第一个智慧城市。
	2012	瑞典	德哥尔摩在普华永道 2012 年智慧城市报告中，斯德哥尔摩名列第五
国内	2012	中国	首批国家智慧城市试点名单（90 个）
	2013		国家“智慧城市”技术和标准试点城市名单
	2014		经国务院同意，发改委、工信部、科技部、公安部、财政部、国土部、住建部、交通部等八部委印发《关于促进智慧城市健康发展的指导意见》，要求各地区、各有关部门落实本指导意见提出的各项任务，确保智慧城市建设健康有序推进。
			2018

### 四、数字经济科技创新的不足

（一）科技创新在数字经济的有序进行需要企业内部和整个社会环境的共同作用和支持，而目前的环境存在的缺陷和不足，制约了科技创新的持续发展。从企业内部环境来说，大多数企业尚未形成良好的创新氛围。技术创新是高风险的，许多企业以眼前利益为主，重模仿，轻自主研发；重引进，轻吸收再开发。这样的经营方式虽然在初期能为企业带来一定的经济效益，但也造成了企业的创新意识薄弱、创新氛围缺失等问题。从外部环境来讲，党和国家虽然高度重视科技创新，但未能提供足够丰富的政策支持。我国目前有关科技创新方面的政策法规仍不完善，科技创新法尚未形成。

（二）作为知识的载体，科技创新人才是科技创新成功与否的关键。但就目前而言，多数企业存在科技创新人才短缺的问题。一方面，企业所需要的技术创新人才对工作的要求也较高，容易受企业规模和影响力等因素的影响。相当一部分人才选择在校和科研机构工作，而中国民营企业占比巨大，这些企业无法对这部分人才产生足够的吸引力。这导致国内高新技术人才的分布不均，很多企业出现人才短缺的情况；另一方面，部分创新人才对自身职业规划非常明晰，而企业的科技创新工作是基于利益驱使，这往往导致人才极强的流动意识，造

成创新人才的流失。这种流失在给企业带来技术外泄风险的同时，也造成企业的创新过程停滞不前，创新成果无法实现，技术创新能力难以稳步提升。

（三）科技创新具有高投入的特点，充足的资金投入是企业进行技术创新的根本保证，但我国企业的技术创新资金投入不够充足。首先，我国体制导致企业的类型并不相同，占比最大的中小型企业，虽具备经营灵活的优势，但资金薄弱，而技术创新需要较大的前期投入，以及较长的投资回收期，这些企业无法完全靠自身来支付技术创新的高昂成本，不得不减少对技术创新的资金投入。其次，我国投融资体系尚不健全，较发达国家的成熟体系有较大提升的空间。融资渠道单一，中小型企业大多以技术资产为主，有形资产比重较小，可提供的担保资产相应较少，无法满足银行的贷款要求，影响着企业科技创新的资金投入。

### 五、数字经济科技创新的建议

（一）整合高质量数据资源，营造自主科技创新环境。数字经济时代数据是根本，技术是手段。对数据的收集、清洗、归类同样需要一定的技术。通过各级政府提供有利的创新环境，完善相关政策法规，为企业技术创新的持续发展提供保障；保证知识产权不受侵害，维护企业的创新利益；开展企业和高校的专项科技创新竞赛活动，提高自主创新意识；进一步加大研发费用加计扣除力度，减少企业研发负担，营造良好创新环境的形成将极大促进技术创新能力的提高。同时，加强“政企”合作关系建设，发挥政府机关丰富的数据资源和企业的科技创新能力，促进数字经济全面发展。

（二）引进各地人才、培养本地人才。高新技术人才掌握着数字经济时代的最重要的技术手段，不同地域的人才对技术的理解不尽相同，在被各领域引入后都会丰富自身经历，并开发出新的技术。引进各地人才是提升科技创新重要手段之一。各地政府应给予外来人才诸如落户、购房、奖励基金、补助等利好政策；对于本地人才，应更加注重专项、专业技能的培养，为本地人才创造“为家乡建设”而学习和工作的精神环境。通过提升参与度和成就感，通过建立完善的激励机制来减少创新人才的流失。提高技术创新人才的工资待遇，其对企业满意度不断上升的同时也调动了自身的创新积极性。

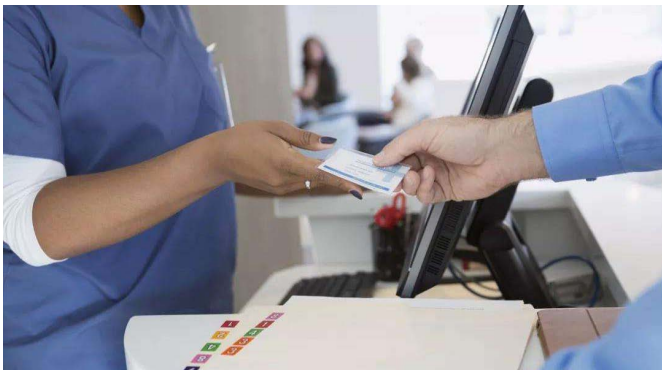
（三）加大政府扶持、完善内、外部融资。为了有足够的创新资金，需要政府加大对企业技术创新的扶持力度，降低企业的融资难度。国家可设立更丰富、更有针对性的创新专项资金、进行研发经费投入、减免税收等举措加大对企业技术创新的扶持力度。而对于企业融资难的问题，一方面通过建立针对具有高新技术的企业融资政府担保体系、政府贴息等方式来帮助企业顺利获得银行贷款；另一方面要拓宽融资途径，通过企业内部融资、民间融资、风险投资等方式，获取更多的创新资金，形成一个创新资金投入保障体系，为企业科技创新的良好循环提供保证。

## / 浅谈医改下，新冠疫情时，运营数据对提升药企应对的意义 /

作者 / 武汉 CPDA 数据分析师 杨红鹰 编辑 / 协会会员处 李苗苗 日期 / 2020-08

新冠病毒疫情时，因成本压力、现金流压力、消费者消费意愿减退等原因，多数企业面临缺资金、缺市场、缺产品等问题，就多数企业而言，活下去才是终极目的，而药企，还得面对医改深化带来地前所未有的挑战。

为稳定经济，各国已纷纷出台财政和货币政策，从消费、投资、出口等方面提供支持，并加大政府支付，我国也出台了六稳、六保政策，但企业自身的有效应对才是谋定的主体。



一边是疫情推动下的消费升级、迭代，一边是企业需迅速拉动市场，企业的应对不仅需涉及战略、业务的调整，模式的创新，还需优化自身产业链，降本增效，其中风险规避、成本管控、数字化转型等都需提上日程；

大数据、云技术、5G技术、人工智能等高速发展，为运营数字提供了无限可能，我以为，产业链数字化是未来企业经营的趋势，也是企业适应新形式，实现转型、创新的关键路径之一，并将成未来商业主流；

何为数字化产业链？大致可理解为在企业的研发、采购、生产、销售、物流等关键环节，通过运营数据去清晰问题所在，去突破其中难点和痛点，从而助力业务发展。

在医药行业，因药品特殊性，在药企现今的经营中，数据分析常用于传统的销售业绩分析、经营评估，但未真正实现全面数字化转型，价值空间有较挖掘。

数字化转型，是将数字化应用融入各产业链的业务运营，需不断探索和实践。如战略层面有项目盈亏分析，能计算投入的盈亏情况、生产方面有经济生产批量，得出最经济的单次产量、而在销售环节，RFF、聚类、逻辑回归等模型，还可精准完成客户识别及分类、客户画像、流失分析等种种任务。

在产品的设计期有KANO模型，可明确用户对产品性能需求度的排序、有PSM价格敏感度，可计算出最多消费者（含潜在客户）能接受的价格区间和最优价格；

算法模型很多，应用方法也因物而变，这些只是一部

分，我们需将事项转化为具体的问题，再根据实际需求去探索。如应用较广的词云，操作不难，用处却很广大。

在市场研究工作中，我曾在可行项目分析中，用其了解用户对产品的需求态度，在销售业务中，也用于部分消费者洞察等，在此，我想分享该应用的一个尝试，推动大家深入探讨。

词云是消费者洞察的有效展示，词频统计暗含玄机：

在产品销售中，我们常需了解消费者对产品的评价、关注点等，如用传统访谈的方法耗财、耗时、耗力不说，还难保持持续跟踪和监测，更何况用户本身就存在言行不一的情况，如果问题设计不到位、或问题有预设，参考价值可能大打折扣。

此时词云就有了新的意义，适合统计上万的数据，甚至样本量越多越好。但要了解消费者真实体验，数据从何而来呢？除了我们官网后台的数据外，知名电子商务平台的评价也是便捷的来源。

举例说明，针对某电商平台相关品类的领导品牌湿纸巾，我下载了1万+评价，因其有大量客服回复、图片、拷贝刷屏等脏数据，我先对其完成数据清洗，再得出词云如下（为确保该品牌隐私，已将数据做了调动和修改处理）：



根据词频统计，可清晰看到，用户对酒精、湿巾、消毒、好用、方便、回购、携带、包装等词语提及频次较高，这应是大众用户对产品的主要印象；

一般，我们对微信软文，或影片的评价等也就在这步，可以看出正向居多还是负向居多。

但深入细分：

1) 酒精相关词语累计近400次，疫情、消毒、杀菌及刚需和场景相关词语出现490+次，方便、便携、随用等相关词语词频达230+次，湿纸巾词频100+，小包装字眼出现频率达83+以上；

解读：可初步理解为消费者已对酒精=防疫灭菌基本建立起稳固认知，而消毒湿巾便捷好用基本意见一致，小包装更灵活，多用于复学、复工、就餐、外出等各种场景。

因此在产品传播时除新场景开发外，不建议对消费者再做酒精杀菌教育；

酒精	191	很浓	39	度	22
了	176	给	38	我	22
用	124	有	38	浓	22
方便	116	收到	38	快递	21
不错	110	孩子	36	价格	21
买	82	着	36	顺丰	21
购买	74	好评	35	带	21
味	74	就	34	小包装	21
也	70	携带	33	上学	21
湿巾	70	好用	32	再	20
挺	66	厚实	31	整体	20
还	61	出门	28	携带方便	20
第二次	58	特别	26	快	19
味道	57	疫情	26	真的	18
非常	57	东西	26	满意	17
是	54	多次	25	物流	17
:	48	评价	25	比较	17
消毒	46			值得	17
				材质	16
				菜	16

2) 其次快递、小包装相关词语出现率分别为110+，90+，价格紧随其后，可以理解为用户对产品属性的关注度依次为酒精、灭菌、湿巾、便捷、快递、小包装、价格等；

3) 评论中对品牌提及度极少仅几次，可理解为用户心智中尚未建立领导品牌，该品类还有介入机会；

快递	21	比较	17
价格	21	值得	17
顺丰	21	材质	16
带	21	来	16
小包装	21	发货	16
这个	21	还是	16
上学	21	放心	16
再	20	活动	16
整体	20	会	15
携带方便	20	没有	15
快	19	期间	15
真的	18	手感	15
满意	17	使用方便	14
物流	17	安全性	13
比较	17	必备	13
		和	13

另外，在日常生意中，为避免意淫用户需求，我们常需调查获取用户真实需求，但如果将词云结合KANO使用，先筛选出该品类最受用户关注的属性（避免意淫用户需求），再用KANO对这些品类属性满意度排序，是否能更有效契合用户潜意识呢？

事实表明，此方法可有效为消费者洞察提供真实世界的部分信息，若深一步探索，或能将资料转化为更为有效的情报。

运营数据目的是为经营服务，将其中一些难点、重点问题，落实到具体的事项上，寻求清晰的可能，或明确解决方向，因此，数据只是材料，加工还取决于个人运用和实际需求。

有时大道至简，可能不一定那么复杂，需要某种模型和算法，比如对某些政策的分析；

近年医改政策多如雪片，以医保为杠杆的三医联动在国家顶层设计的背景下分步推进，药品各终端面临前所未有的冲

击和变局，药企经营更是处于极大变数中，而医保政策后期走势是业内讨论和关注的重点，未来医保监管真会收紧吗，存在着较多质疑。

除了从政策推进、法规制定、监管方式试点推进的角度分析外，我还需要一些数支持。

基于此，我尝试将医保基金的2018年、2019年的收支变化等数据做了简单运算，政府医改的底层逻辑就自然浮出水面，如图：

原因	基金显示的问题	现实和解决方法	政府计划和行动
取消个人账户	城乡居民参保积极性下降 (参考降低0.3%)		
总体支出大于收入；	统筹基金累计结存同比下降 15.99% (存量危机)	2019年医改下，收支已接近平衡，说明医改控费有效。如要 保障覆盖和医保扶贫，范围和力 度还需加大；	<ul style="list-style-type: none"> <li>从增量改革转向存量改革；</li> <li>推出医改系统集成；</li> </ul> 行动： 1) 坚持带量采购，推进区域 性、全国性联盟采购； 2) 医保基金与企业直接结算， 支付标准与集采价格协同；
实施医疗救助和医保扶贫 医保目录扩大，待遇提高； 异地就医便捷，统筹级别提高	城乡支出增速最大达14.23%；	城乡医保存在较大刚需，而国 家加大医保扶贫，控费可能；	
个人账户累计结存同比增长 13.63%； 在其增长额中，有13.89%由 基金监管贡献；	基金监管力度加大是个增收 增长要素之一，有较大提 升空间。个人账户累计结存 占统筹累计结占60.97%；	加大监管力度范围和深度，能 够有效提升个人账户累计结存， 实现共济；	存量改革； 监督检查常态机制； 依法追究欺诈骗保等；

数据显示，到2019年统筹基金累计结存同比下降15.99%，存量危机；其中，2019年城乡支出增速最大达14.23%，且国家还要加大医保扶贫等力度，增速还将增长；而国家不断地将重大疾病用药、慢病用药、急病用药纳入报销范畴，人口又是老龄化趋势，医保基金将持续承压；

疫情后，国家经济遭受重挫，医保基金来源也将受到一定影响，因此如果要达到国家要求的保基本、保覆盖的基本需求，医保基金调整更是迫在眉睫，而调整最有效的方式就是结构调整和费用管控。

这就是医保局所说的存量改革。从2019年的基金收支情况来看，已近平衡，说明医改集成的控费效果显著；所以后期政府推进品种结构调整，强化医保控费，坚持带量采购等就好理解了。

再看个人账户累计结存，有至少13.89%的份额源自基金监管，而现今检查的力度、广度、深度还有较大提升空间；个人账户总额已占统筹基金总额的60.97%，其力量可观。结合国家的态度和行为，试想若加大监管力度和深度，就能提升个账的累计结存，并能有效实现门诊共济，因此，建立监督检查常态机制，依法打击骗保等实施的决心可以看见了。

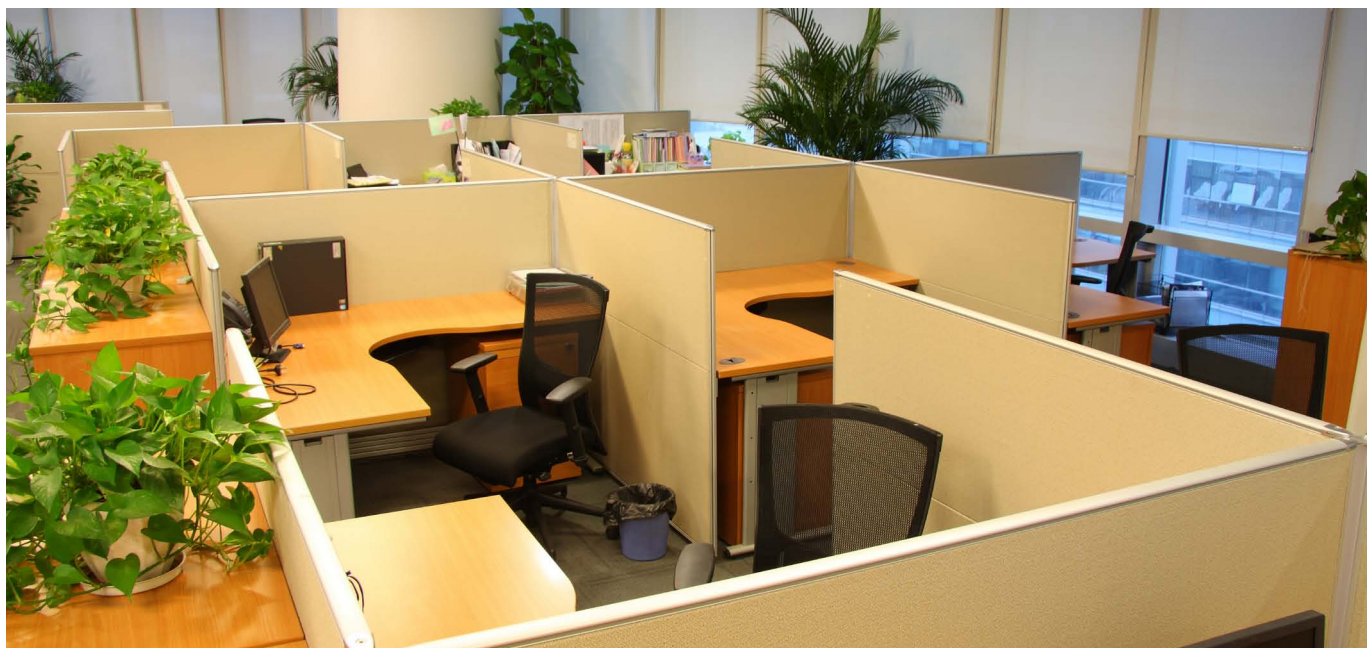
取消地方医保目录增补，也是医保基金控费的一个重要举措，当增补品种不再享受医保基金支付时，相关品种销售将遭受较大冲击。药企也需转换销售思维，去展示产品的效果，去布局更多的触点，去提升产品的服务，以此推动销售。那么基于用户行为的数据沉淀，和消费偏好等的分析，在洞察消费者，拉近品牌和用户的距离上，将显得尤为重要。

在大数据时代，我们应养成数据分析的思维和意识，去探索如何有效利用海量数据，攻克经营中遇到的痛点和难点，从而有效实现转化应用，最终为战略决策、经营调整所用。



# / 数据分析师事务所从业规范 /

来源 / 中国商业联合会数据分析专业委员会 编辑 / 协会会员处 李苗苗 日期 / 2020-08



## 第一章 总 则

第一条 为了规范数据分析师事务所合规经营管理，优化数据分析师事务所内部治理，提升数据分析师事务所竞争力和可持续发展能力，加强对数据分析师事务所从业行为监督和指导，根据中国商业联合会数据分析专业委员会（以下称“我会”）《工作条例》、《会员管理办法》以及《中国数据分析行业自律公约》，经我会与相关从业机构共同协商，制定本规范。

第二条 本规范所称数据分析师事务所，是指在我会登记备案的数据分析师事务所，包括副主任单位事务所、常务委员单位事务所和一般事务所。

第三条 数据分析师事务所须严格遵守国家法律、法规和行业自律公约，不得有国家法律、法规及监管部门规定的禁止行为。

第四条 数据分析师事务所执行数据分析业务时，应当遵守本规范以及依据本规范制定的其他业务准则。

第五条 我会负责组织本规范的制定，并监督执行。

## 第二章 数据分析师事务所业务执行规范

第六条 依法经营，诚实守信。数据分析师事务所应当严格遵守国家有关法律、法规和行业规章，积极倡导诚实守信的良好风气，切实规范自身的经营行为，在平等、自愿、公平和诚实信用的原则下开展业务，不得损害国家利益、社会公共利益、行业利益和客户的合法利益。

第七条 不以欺诈牟利，不取不义之财，数据分析师事务所在执行数据分析业务活动过程中，要避免和客户发生利益冲突，在不违反结论客观原则的前提下始终将客户的利益放在第一位，不得为自身谋取任何潜在利益。

第八条 数据分析师事务所应当制定统一的业务管理制度，明确业务承接、执行等环节的规范要求，在全所范围内执行统一的业务管理标准。

第九条 数据分析师事务所在执行数据分析业务中，应具备风险识别预警能力，向客户进行必要的风险提示，同时通过数据分析能够识别业务管理过程中存在的风险点，具有前瞻性，可提出综合性解决方案。

第十条 数据分析事务所应对收集的数据在专业范围内进行真实性、准确性和完整性的判断，数据分析和结论应以客观事实为依据，合乎逻辑，无不实陈述。严禁弄虚作假，不得向客户提供含有虚假的数据和具有片面、误导性分析的结论或意见。不能与客户或其他市场参与主体合谋篡改数据从而歪曲分析结果。

第十一条 数据分析师事务所提交数据分析成果时，应当自觉使用合法合规信息，不得以任何形式使用或泄露国家保密信息、客户信息以及未公开重大信息。

第十二条 数据分析师事务所应当结合自身发展战略和经营管理需要，不断提高数据分析师事务所在业务管理、财务管理、人力资源管理等方面的信息化水平，并运用信息化手段加

强对执业质量和管理状况的实时监控。

第十三条 数据分析师事务所应当加强对人员专业技能的培训，须组织执业人员认真参加我会举办的执业教育培训，积极参与我会各种线上、线下大数据课程培训，充分利用我会提供的各种学习机会，不断提升大数据专业技能。

第十四条 数据分析师事务所应重视并加强对事务所的品牌宣传，提升行业声誉及数据分析师事务所的社会影响力。同时也可充分利用我会资源平台及各类渠道，进行对外宣传推广、提高品牌影响力。

第十五条 合理定价，不搞恶性竞争，相互压价；公平交易，有序竞争，不以不正当手段破坏市场秩序。数据分析师事务所应遵循公平竞争原则，维护正常的市场秩序，遵守商业道德，杜绝恶性竞争和垄断市场等行为，不得以减免或承担相关费用为条件进行不正当竞争；不得以诋毁行业内其他单位的商业声誉、泄露其商业秘密等不正当手段争揽业务；不得利用政府行政资源干预手段不计成本地争揽客户，扰乱正常的经营秩序。

第十六条 数据分析师事务所在为客户提供服务时，应严格参照我会制定的《数据分析服务内容 & 费用组成参考》中数据分析服务的组成内容提供规范服务。

### 第三章 数据分析师事务所职业道德规范

第十七条 数据分析师事务所应遵循独立、客观、公平、审慎、专业、诚信的执业原则，自觉弘扬行业优秀文化，加强从业人员职业道德修养，规范行为，履行社会责任，遵守社会公德。

第十八条 数据分析师事务所在具体的从业过程中，要始终保证工作的独立、客观和公正。

第十九条 数据分析师事务所应提倡学知识、学科学、学技术，以不断强化职业道德和业务技术为核心，努力提高从业人员的道德修养和综合素质。

第二十条 数据分析师事务所应相互尊重、团结互助、加强合作、共谋发展，共同维护行业声誉，不得在公众场合及媒体上发表贬低、损害同行声誉的言论，不得以不正当手段与同行竞争。

第二十一条 数据分析师事务所应履行保密义务，对客户提供的数据要严格保密、对数据分析过程及其结果要科学、严谨。

### 第四章 规范管理

第二十二条 我会负责本规范的日常监督管理。

第二十三条 鼓励、支持、保护一切组织和个人对违反本规范的行为进行社会监督和舆论监督。提倡行业机构间互相监督。对背离行业道德的现象和行为，任何人都可以向我会举报，经核实后，将按照我会《会员管理办法》对违规行为给予处理。

### 第五章 附 则

第二十四条 本规范由我会负责解释。本规范的内容如与国家法律法规和政策相抵触的，以国家法律法规和政策的规定为准。

第二十五条 本规范自公布之日起施行。

中国商业联合会数据分析专业委员会

二〇二〇年六月三十日





# / 数据分析师执业规范 /

来源 / 中国商业联合会数据分析专业委员会 编辑 / 协会会员处 李苗苗 日期 / 2020-08

## 第一章 总 则

第一条 为了规范数据分析行业数据分析师的执业行为，提高数据分析师专业素养，以及维护数据分析师良好执业形象，加强数据分析行业管理，根据中国商业联合会数据分析专业委员会（以下称“我会”）《工作条例》、《会员管理办法》以及《中国数据分析行业自律公约》，经我会与相关从业机构共同协商，制定本数据分析师执业规范。

第二条 本规范所称数据分析师，是指取得CPDA数据分析师证书，并且在我会登记备案的数据分析师事务所从事数据分析业务的执业者。

第三条 数据分析师须严格遵守国家法律、法规和行业自律公约以及所在经营机构内部管理制度，不得有国家法律、法规及监管部门规定的禁止行为。

第四条 数据分析师执行数据分析业务时，应当遵守本规范以及依据本规范制定的其他业务准则。

第五条 我会负责组织本规范的制定，并监督执行。

## 第二章 数据分析师业务专业素养

第六条 数据分析师应在独立、客观、公平、审慎、专业、诚信的执业原则下开展业务，不得损害国家利益、社会公共利益和行业利益。自觉弘扬行业优秀文化，加强自身职业道德修养，规范自身行为，履行社会责任，遵守社会公德，遵守所在事务所的内部管理制度，规范执业行为。

第七条 数据分析师应恪守诚信原则，具备严谨、负责的态度，保证数据的客观、真实、准确，保持中立立场，客观评价数据分析过程中存在的问题，为客户提供有效的参考依据。

第八条 数据分析师在执行数据分析业务时应积极主动地发现和挖掘隐藏在数据背后的真相，始终对数据和结论保持敏感。

第九条 数据分析师应具备勇于创新的精神，通过不断创新，提高自身分析能力，使自己站在更高的角度来分析问题，为整个研究领域乃至社会带来更多的价值。

第十条 数据分析师在执行数据分析业务时应具备缜密的逻辑思维和清晰的逻辑推理能力，在面对复杂问题时需考虑错综复杂的成因，分析所面对的复杂环境因素，理清问题的整体及局部结构，进而理清结构中相互的逻辑关系，客观、科学地做出判断。

第十一条 数据分析师应团结互助、相互尊重，加强合作、共谋发展，共同维护数据分析行业良好形象，不得在公众场合及媒体上发表贬低、损害数据分析行业声誉的言论，不得以不正当手段与同行竞争。



## 第三章 数据分析师业务执行能力

第十二条 数据分析师在执行数据分析业务活动过程中，要始终保证工作的独立、客观和公正，要避免和客户发生利益冲突，在不违反结论客观原则的前提下始终将客户的利益放在第一位，不得为自身谋取任何潜在利益。

第十三条 数据分析师在执行数据分析业务中，应保证数据的安全、稳定、准确、完整，并具备对数据进行多维度分析的能力，以实现为客户提供运营策略方向支持，实现经营决策的战略目标。

第十四条 数据分析师在执行数据分析业务和作出结论时，应以客观事实为依据，合乎逻辑，无不实陈述。严禁弄虚作假，不能与客户或其他市场参与主体合谋篡改数据从而歪曲分析结果。不以主观好恶或个人偏见行事，不事先设定立场。

第十五条 数据分析师在提交数据分析成果时，应自觉使用合法合规信息，不得以任何形式使用或泄露国家保密信息、客户信息以及未公开的重大信息。

第十六条 数据分析师在执行数据分析业务时，应懂得管理理论，在搭建数据分析框架时确定分析思路。同时应熟悉行业知识、业务及流程，不应脱离行业认知和业务背景而得出分析结论。

第十七条 数据分析师在执行数据分析业务中，应具备风险识别预警能力，向客户进行必要的风险提示，同时通过数据分析能够识别业务管理过程中存在的风险点，具有前瞻性，可提出综合性解决方案。

第十八条 数据分析师应当按照我会的相关规定参加执业教育，并积极参与我会组织的专业技能培训课程和相关活动，



认真学习和领会国家与行业路线方针政策，不断提高大数据专业能力、执业水平及合规意识。

第十九条 数据分析师应遵守所在事务所的管理制度，履行岗位职责，充分尊重和维护所在事务所的合法权益。数据分析师离职后，应当履行与原所在事务所签署的劳动合同或协议的有关约定，承担相应的保密、竞业限制、培训赔偿等义务。

#### 第四章 规范管理

第二十条 我会负责本规范的日常监督管理。

第二十一条 对违反本规范行为的数据分析师，根据违反规范的情节轻重，我会可分别对其采取批评教育、内部通报、向公众曝光，以维护行业的规范管理及公正性。

第二十二条 鼓励、支持、保护一切组织和个人对违反本规范的行为进行社会监督和舆论监督。

#### 第五章 附 则

第二十三条 本规范由我会负责解释。本规范的内容如与国家法律法规和政策相抵触的，以国家法律法规和政策的规定为准。

第二十四条 本规范自公布之日起施行。

中国商业联合会数据分析专业委员会  
二〇二〇年六月三十日

## / 用大数据思维助力数字化转型的全新服务行业 /

作者 / 中国商业联合会数据分析专业委员会 编辑 / 协会会员处 李苗苗 日期 / 2020-08



2020年伊始,一场突如其来的新冠肺炎疫情打乱了所有人的生活。几个月来,大数据助力全民战“疫”,让我们看到了大数据应用的广阔前景,特别是在此次新冠肺炎疫情防控工作中,众多城市利用大数据优势,对疫情预警、追踪、监测、宣传,并服务于疫情防控决策和公众参与,对进一步提升政府社会治理能力、公共服务水平和应对大型公共安全突发事件能力均有重大意义。

疫情危机孕育新机遇,数字中国建设不断向前推动。同样,在疫情常态化控制时代的复工复产中,大数据更加发挥强劲的发展动能和更加广阔的发展空间:在数字经济、技术创新、网络惠民等方面不断取得重大突破,有力推动数字化转型迈上新台阶。

据相关信息了解,在大数据应用价值不断凸显的今天,数据

分析师事务作为一个全新的第三方服务行业,已经遍布在全国各省市,正在践行以“大数据思维”助力政务数字化发展、企业数字化转型、教育数字化应用,围绕数据的深度分析、业务场景构建、深层次的咨询等,以大数据思维帮助企、事业单位实现数字化转型并提供行之有效的战略决策。

面对机遇和挑战,随着科技的不断发展,大数据应用的广阔前景,商业智能与数据分析的结合愈加丰富,数据分析师事务所作为一个全新的朝阳行业,更是发挥其数据分析思维潜力及拥有广阔的发展空间,为实现科学化、精准化、高效化挖掘数据的价值,为企、事业单位做出切实可行的决策,并助力众多行业企业在疫情常态化控制时代,完成复工复产和数字化转型!



# 后疫情时代， 我们起扛！

懂  
行业

知  
业务

重  
分析

善  
融通



关注CPDA数据说

我们只培养解决  
企业关键需求的  
大数据人才！

咨询热线：400-050-6600

CPDA®  
数据分析师  
Certified Projects Data Analyst