



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
2019年第04期 总第40期(季刊)
咨询热线:400-050-6600
<http://www.chinacpda.org/>
投稿邮箱: xiehui@chinacpda.org



这个时代抛弃你，连声招呼都不打！

三年前，AI为更多人所熟知的事件是AlphaGo战胜了世界围棋冠军柯洁的一场“人机大战”。此后，柯洁只能被称为“人类围棋世界第一”。一时间，街头巷尾，议论纷纷，热钱涌入，风口造势，各行业言必称AI。不管你接不接受，科技在改变这个世界，每个人都无法改变和逃离。

据工信部中国信通院副所长张雪丽介绍，截止至2018年9月，北京则以445家的总数，成为全球人工智能企业最多的城市。更厉害的是，中国靠全球1/5的企业获得了3/4的融资，足以可见投资者对于中国AI领域的期待与展望。

人工智能的背后是基于大数据和数据处理的强大支撑，通过对于数据的清理、结构化、深入分析得出相对应的结果，进而将这些结果转化为“智能”进行输出。这其中数据分析师们成为整个生态中最硬核的部分！人工智能不会像人类那样推断出结论，它通过试验和错误学习，“神经网络”、“深度学习”让机器愈加智能。于是，许多分析师可能担心它们会被自动化和人工智能所取代。恰恰相反，机器无法理解场景，只有人类才有能力以复杂的术语对数据进行语境化。在人工智能时代，广大数据分析师完全可以用人工智能来简化数据分析的基础工作，将匹配和清洗的过程交给人工智能，从而大大减少分析师为准备数据所花费的时间。减少了数据探索的时间，就可以花费更大的精力、更多的时间在业务洞察方面。数据分析师将会做到机器无法做到的事情——提供场景和解释数据。只有数据分析师们变得精通业务、了解行业，将分析结果结合到具体的场景中，就能够为公司业务提供有意义的分析，从而做出更好的数据驱动决策。可见数据分析遇见人工智能可谓是如虎添翼！

当今这个不断变化的时代，抛弃一个固执己见、默守陈规的人连声招呼都不打！站在2019年末，祝愿所有的CPDA数据分析师都能以开放的心态和理性的兴奋，充分拥抱这个变化中的新时代。

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 这个时代抛弃你，连声招呼都不打！

协会动态

- 04 安徽省高职数字经济人才培养研讨会议顺利召开
05 苏州CPDA数据分析线下沙龙活动圆满结束

政策导向

- 07 山西晋城：《关于促进大数据发展应用的若干意见》
08 《海南省大数据开发应用条例》正式实施

行业动态

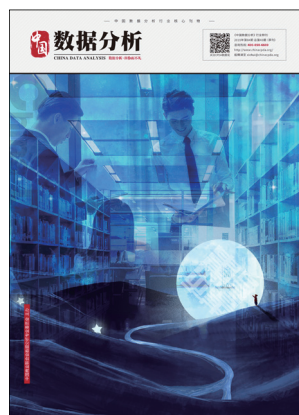
- 09 大数据时代，拥抱大数据分析技术
10 大数据在国内实现真正落地，还需要迈过三道坎
11 大数据的世界之旅
12 得数据者得"天下"，学会数据分析走出"霸王步"
13 人工智能是数字营销革命中不为人知的趋势

学数交流

- 15 运用数据分析方法解读武汉市政务大数据发展
17 罪案大数据分析
22 基于数据分析方法的运营商营业厅选址研究
24 空调机组报价预测——基于线性回归的数据分析案例
26 基于数据分析方法的电信移动业务收入预测探究

事务所专栏

- 28 高致病性禽流感H5N1病毒在职业人群感染的数据研究
30 广西尚辰合讯数据分析师事务所



主办单位

中国商业联合会数据分析专业委员会

编委成员

李苗苗 / 杜天天

出版时间

2019年12月出版 总第40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



CPDA数据分析师培训



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

/ 安徽省高职数字经济人才培养研讨会议顺利召开 /

来源 / CPDA数据分析师网 编辑 / 协会会员处 李苗苗 日期 / 2019-12

2019年11月22日，由合肥市大数据产业创新战略联盟主办，安徽警官职业学院承办，CPDA数据分析师合肥授权管理中心协办的“安徽省高职数字经济人才培养研讨会议”在安徽警官职业学院顺利召开。

本次研讨会议，由安徽警官职业学院、安徽职业技术学院、安徽财贸职业学院、安徽商贸职业技术学院、徽商学院等20多家高职院校代表，以及企业、行业协会、大数据产业联盟、媒体朋友等共同出席，共同探讨建立产学研用深度融合的高职数字经济人才培养新模式，打造开放交流合作的新平台，培育区域数字经济人才生态圈。与会专家们会上还做了精彩主题分享，让大家多方位多角度认识数字经济人才在数字经济中的重要作用。

本次会议中，为加快落实合肥市数字经济发展“一号工程”，促进安徽警官职业学院与合肥市其他高职院校，以大数据为代表的数字经济相关专业高质量发展，探索数字经济人才培养与引进体系，打造产教融合下的“新人才、新模式、新高地”，实现校企共赢，在合肥市数据资源局与安徽警官职业学院的共同努力下，合肥数字经济应用人才高职培养基地正式落户安徽警官职业学院。安徽警官职业学院胡来龙院长和合肥市数据资源局王伟副局长为合肥数字经济应用人才培养基地进行揭牌。



揭牌仪式

在主管部门和学校的共同努力下合肥市首个数字经济应用人才培养高职基地已顺利落成，未来更多的产业资源、人才培养政策等多项举措会陆续落地。

胡来龙院长作为院方代表与CPDA数据分析师合肥授权管理中心负责人吴昊，在现场签订数字经济应用人才培养校企合

作与高职基地运营协议



签订协议

合肥数字经济人才培养专家委员会秘书长、CPDA数据分析师合肥授权管理中心负责人吴昊，中国商业联合会数据分析专业委员会培训处主任李伟，星环信息科技（上海）有限公司、星环大学的教育产品总监孙国库，安徽警官职业学院专业负责人劳雪松等4位嘉宾，分别做主题分享。



CPDA数据分析师合肥授权管理中心负责人 吴昊

合肥数字经济人才培养专家委员会秘书长、CPDA数据分析师合肥授权管理中心负责人吴昊做了合肥市数字经济人才企业需求情况分享，对数字经济人才定义、合肥市数字经济人才现状、合肥市数字经济人才企业需求，以及数字经济人才课题研究成果分别做了阐述。分享了近一年合肥市以大数据为代表的数字经济人才岗位能力要求、薪资匹配，以及人才供需侧重点等信息。



中国商业联合会数据分析专业委员会培训处主任 李伟

中国商业联合会数据分析专业委员会培训处李伟主任做数据分析师职业人才培养经验分享，对大数据分析人才需求与现状、数据分析师行业协会、CPDA数据分析师培养经验做了阐述，为与会专家老师提供了协会职业人才培养经验参考。



安徽警官职业学院专业负责人 劳雪松

安徽警官职业学院专业负责人劳雪松老师做大数据技术

与应用专业建设的思考主题分享，劳老师结合多年的教学管理经验，对高职院校在数字经济行业中的地位、大数据专业建设思路、大数据专业人才培养方案、人才培养的痛点与应对思路等做了详细的分析与阐述，分享观点与到场其他院校的大数据专业老师形成了极大的共鸣。

嘉宾分享环节结束后，与会20多家省内高职院校相关院系老师充分肯定了分享嘉宾提出的人才现状与应对思路，统一认识与行动思路。对安徽警官学院承建合肥数字经济应用人才基地，给予了充分的肯定，希望未来研究的成果、资源能够与各院校同步、分享。



研讨会圆满结束

本次高职数字经济人才培养研讨会议的召开，促进了安徽警官职业学院以大数据为代表的数字经济相关专业高质量发展，同时也激励了各院校积极探索、参与合肥市数字经济人才培养工作，共同打造产教融合环境下的“新人才、新模式、新高地”，实现校企共赢。



/ 苏州CPDA数据分析线下沙龙活动圆满结束 /

来源 / CPDA数据分析师网 编辑 / 协会会员处 李苗苗 日期 / 2019-12

2019年11月2日，对于苏州大数据圈来说是个比较有纪念意义的日子。

目前，苏州CPDA数据分析师的培训、授课已经开展了九期，很多小伙伴都觉得只有上课的时候才能跟老师和同学们交流一下，我们还是缺少一个可以与大家面对面、深入交流的机会与平台。在这样的声音中，CPDA苏州授权中心在苏州工业园区的苏悦广场举办了第一届线下沙龙活动



上午9:00准时开始签到入场，小伙伴们都很积极，现场新老朋友交流气氛极为热烈。本次沙龙，我们邀请到了既是老学员又是讲师的洪赦虹与顾晓峰为我们做精彩分享。



首先是我们的第九期学员，也是一位资深讲师，给大家分享了数据分析与抽样的理论与应用知识。深入浅出的讲解，让晦涩难懂的问题变得简单易懂，便于理解。

• 洪赦虹 Daisy Hung



- 曾任友达光电 工程师
- 凌越信息有限公司 销售
- 西胜国际股份有限公司 品保经理
- 亚博制程管理顾问有限公司 资深顾问师
- 苏州朗坤企业管理咨询有限公司 资深顾问师



一个小时的分享很快就过去了，经过中场的茶歇，下半场的分享由我们第六期的学员顾晓峰为大家进行分享。顾晓峰也是我们江苏省第一家数据分析师事务所的创始人，分享的内容也是相当精彩。在分享的过程中，顾晓峰给大家讲解了数据分析与企业决策的思路。



活动最后，参加人员都表示在活动中受益匪浅，并感谢两位老学员的精彩分享，同时提出希望以后可以多多举办这样的活动。



未来，CPDA苏州授权中心还会邀请更多的行业大咖、资深从业者，从自身行业、岗位角度出发，为大家带来更多的数据分析实战经验，希望对大家有帮助。

数据分析师不仅仅是一个岗位，而是每个人、每个岗位都可以应用到的一个技能，是可以影响我们生活的一种思维模式。让我们期待下一次沙龙吧！

希望，你的未来在你眼中，不再是一片没有尽头的茫然和漫长，而是一个个可以被期待的时间节点。

/ 山西晋城:《关于促进大数据发展应用的若干意见》 /

来源 / 山西省人民政府官网 编辑 / 协会会员处 李苗苗 日期 / 2019-12

2019年7月26日,山西省晋城市人民政府办公室印发了《关于促进大数据发展应用的若干意见》(以下简称《意见》),具有很强的政策性、指导性和操作性,标志着晋城市推进大数据应用实现重大突破进入了新阶段,将对增强全市广大干部群众的大数据应用体验感产生深远影响。

晋城市人民政府办公室关于印发促进大数据发展应用的若干意见的通知

各县(市、区)人民政府、开发区管委会,市人民政府各委、办、局:

《关于促进大数据发展应用的若干意见》已经市人民政府同意,现印发给你们,请结合实际,认真组织实施。

晋城市人民政府办公室

2019年7月26日

《关于促进大数据发展应用的若干意见》

为深入实施大数据发展“1183”工程,加快建设数字强市和智慧城市,推进大数据应用实现重大突破,打造中等智能城市全国样板,根据国家相关政策和省委、省政府及市委决策部署,结合我市实际,提出以下意见:

一、总体要求

1.全市大数据发展应用要全面贯彻市委市政府决策部署,坚持“一盘棋”思想,遵循统筹规划、集约建设、共享开放、有效应用的原则,依托“一云”“一网”“一中心”,整合汇聚政务服务、公共服务和便民服务资源,突出在线政务服务、智能便民服务、智能决策指挥、智能政务办公等重点应用,加快建设全市统一的门户网站、移动APP,推动实现“网上办”“掌上办”,增强广大人民群众的大数据应用获得感和幸福感。

二、主要任务

2.加快建设全市统一的大数据云平台。引进阿里巴巴一流的云计算技术,提升计算分析、系统迁云、数据集成、可视展示、政务办公和便民服务能力,搭建统一的数据整合、共享和交换平台,为全市各级各部门提供统一的网络、计算、存储、运维和安全等技术服务和平台支撑,促进政务大数据治理和公共服务在线化、智能化、集约化。(责任单位:市大数据应用局)

3.加快推进政务数据应用系统迁移上云。各级各部门要将已建、在建的非涉密政务数据应用系统,限期迁至市大数据云平台,构建互联互通、业务协同、信息共享的“大系统”。各

级各部门未经批准,不得再自行建设业务数据基础平台和非涉密共享交换平台。〔责任单位:各县(市、区),市直各相关部门〕

4.加快推进政务数据整合共享。政务数据是政府资源,归政府所有。各级各部门以“共享为原则、不共享为例外”,按照“谁主管、谁提供、谁负责”的原则,围绕建设人口、法人、宏观经济、空间地理、社会信用、电子证照等基础数据库的相关要求,依规编制政务数据资源目录,将采集的政务数据统一存储到市大数据云平台,纳入全市数据共享交换平台系统。对已纳入国家和省直部门的业务数据,由相关部门按业务对口协调系统接口和数据落盘;对没有建设数据应用系统的单位,通过全市数据桥接系统,以文件、电子表格、文档等方式及时进行上传。〔责任单位:各县(市、区),市直各相关部门、晋城税务局〕

5.加快推进电子政务外网全覆盖。按照“一网通办”要求,对未整体接入的部门要在2019年12月底前全部接入市级电子政务外网,实现网络通、业务通。除国家明文规定外,各部门不得新建业务专网,已经建成的要逐步分类并入市电子政务外网。(责任单位:市公安局、市教育局、市民政局、市司法局、市人社局、市规划和自然资源局、市生态环境局、市水务局、市卫健委、市应急管理局、市市场监督管理局、市统计局、市医疗保障局、市能源局等相关部门)

6.加快建设智能便民服务移动APP。制定统一的便民服务项目接入、数据、接口和交互标准,完善统一的可信身份认证体系,打造集政务服务、公共服务、便民服务、信息资讯、政民互动为一体的统一移动服务门户,为实现“一号申请、一次登录、一网办理”提供应用支撑。(责任单位:市大数据应用局)

7.加快梳理便民服务资源。各级各部门组织本单位及下属企事业单位,梳理面向企业和个人的政务服务、公共服务和便民服务事项及相关业务应用系统情况,包括部门网站、政务新媒体、微信公众号和小程序等,确定可提供给全市统一便民服务APP利用的服务资源。〔责任单位:各县(市、区),市直各部门〕

8.加快推进便民服务资源接入汇聚。各级各部门现有的APP按照要求逐步接入统一的智能便民服务APP;在建APP要按照标准规范进行建设接入,为市民提供综合查询、预约办事、线上服务、费用收缴等服务。2019年底前重点推进互联网政务服务、不动产、公积金、交通出行、教育、医疗、社保、旅游、水电气暖等高频服务事项接入,其余的服务应用建

成后逐步叠加，不得建设独立的移动APP。〔责任单位：各县（市、区），市直各相关部门〕

9.加快推进大数据应用集约化。各级各部门按照要求，逐步将部门网站接入全市统一的“晋城在线”门户网站集约化智能管理平台，构建整体联动、集中发布的政务新媒体矩阵。全面推广统一的办公自动化软件和智能政务办公平台应用，实现高效协同“指尖上”办公。〔责任单位：各县（市、区），市直各相关部门〕

10.加快推进数字产业发展。按照国家和省市大数据产业政策，鼓励引进大数据相关企业、科研机构 and 高端领军人才，支持社会资本参与投资大数据发展应用，给予相应的优惠扶持

和激励，大力发展数字经济。〔责任单位：各县（市、区），市直各相关部门〕

三、保障措施

11.加强组织领导。市大数据应用局负责全市大数据发展应用的统筹规划、指导协调、推动落实和监督考评工作，提出资金建议计划，严格项目审核、建设、监管和验收。各级各部门要明确责任单位和责任人，细化工作措施，狠抓工作落实，高标准完成各项任务。12.加强督查督办。定期汇总通报各级各部门的工作进度和完成情况，对工作不力、进展拖延的单位纳入“13710”督办系统进行重点督办。



/ 《海南省大数据开发应用条例》正式实施 /

来源 / CPDA数据分析师网 编辑 / 协会会员处 李苗苗 日期 / 2019-12

11月1日，《海南省大数据开发应用条例》（以下简称《条例》）将正式实施。海南日报记者从省大数据管理局了解到，这是我省关于大数据的首个地方性法规。

《条例》的实施，标志着我省从大数据采集到交易的各个环节都有了明确的法律依据，为服务海南自由贸易试验区和中国特色自由贸易港建设，促进大数据产业发展，维护大数据安全提供了重要保障。

据悉，大数据是信息化发展的新阶段，是国家基础性战略资源，是数字经济的关键要素。2018年4月，习近平总书记在视察海南省政务数据中心时，对我省信息化建设工作表示肯定的同时，希望海南继续在大数据建设上积极探索，创造更多经验。

省委、省政府近年来高度重视大数据工作，把发展大数据作为支撑海南自贸区、自贸港建设的重要抓手。顺应这一时

代要求，《条例》应运而生。

今年9月27日，海南省第六届人民代表大会常务委员会第十四次会议审议通过了《海南省大数据开发应用条例》。《条例》分为六章，共五十七条，从总则、大数据开发与共享、大数据开发与产业促进、数据安全与保护、法律责任等方面做出规定，覆盖了大数据采集、汇聚、存储、管理、开放共享等各个环节。

《条例》具有很强的政策性、指导性和可操作性，紧紧围绕大数据发展趋势，推动我省大数据开发应用，促进大数据产业的发展，培育壮大数字经济，服务海南自贸区、自贸港建设，标志着我省在推进大数据开发应用和产业发展方面进入了新阶段。

/ 大数据时代，拥抱大数据分析技术 /

来源 / CPDA数据分析师网 编辑 / 协会会员处 李苗苗 日期 / 2019-12



2008年9月,《自然》杂志出版了“大数据”专刊,由此揭开了大数据时代的序幕,预示着数据分析时代的到来。大数据、人工智能和移动互联网成为第四次工业革命的三大显著特征。尽管人们今天掌握的数据量比历史上任何时候都要大,但大数据的高增长、类型的多样性并不能直接转化为价值。由于数据的价值密度低、更新速度快、时效性差,人们只有采用更为科学的方法和更为有效的数据分析技术,方能从纷繁复杂的海量数据中获取价值。如果把大数据比作石油,那么数据分析技术,特别是分析者的技能则是炼油的手段,唯有此才能实现价值增值的最终目的。

自2009年开始,中国人民大学中国就业研究所便开始尝试将数据分析方法应用于我国宏观就业市场的短期分析工作。我们对智联招聘全站数据进行量化分析,定期发布中国就业市场景气指数(简称CIER指数),以及时反映我国就业市场职位空缺与求职人数的动态,监测就业市场景气程度变化。目前CIER指数已经被政府部门、研究人员和商业机构广泛接受,作为衡量就业市场的“温度计”,成为我国短期宏观劳动力市场分析、判断和监测的重要工具。

大数据不仅在宏观经济和社会发展层面有着巨大影响,也为企业提升微观人力资源管理的水平带来了挑战和机遇。一方面,以大数据、人工智能为代表的新技术革命促进了新经济和新就业形态的产生,“零工经济”“合伙制”“自适应组织”和“新雇佣经济”的出现,对传统人力资源管理方式提出了更大的挑战,成为了人力资源专业人士必须面对的现实;另一方面,随着时间的推移,移动互联网、社交媒体、可穿戴设备等在工作场所的应用越来越普遍,组织可以轻松获得大量的员工行为数据,并有机会从大量数据中形成观点和结论。特别是随着计算技术、分析技术的发展,微观人力资源管理运用数据分析成为可能。人力资源从业者如何利用数据分析技术,充分挖掘员工数据中的潜在价值,从而帮助组织提升竞争力,已成为人力资源管理实践界人士的重要议题。

令人欣喜的是,以谷歌、IBM等为代表的跨国公司和以百

度、腾讯等为代表的中国公司已经率先进行了一定的探索。在谷歌,人力资源部坚持采用数据分析方法,把分析作为制定人力资源决策的基础。IBM则利用其认知计算平台和沃森分析系统,运用人工智能的方法,挖掘数据分析在人力资源管理领域的应用潜力。百度公司提出建立基于数据分析的智能化人才管理体系。腾讯公司建立活力实验室,利用已有的技术优势和数据优势,尝试进行组织活力分析。可以说,今天各级各类组织的人力资源部比以往任何时候都需要并有可能了解数据、分析数据,挖掘数据价值。

尽管迄今为止,越来越多的企业试图将数据分析方法应用于人力资源管理中,但总体而言这项工作仍处于起步阶段。应该看到,包括大数据在内的各类数据分析,在微观人力资源管理层面的应用与普及依然进展缓慢。究其原因,一方面源于企业对数据分析的价值或重要性认识不足,另一方面也在于行为数据更敏感,大大限制了数据收集和使用的进度。还有一个更为关键的因素就是人力资源管理工作缺乏数据分析技能,包括分析性思维等方面的素质和能力。因此,要想更好地利用数据分析技术,进一步提高人力资源管理水平,未来还有很长的一段路要走,还有很多的工作要做。

《HR的分析力-人力资源数据分析实践指南》一书三位作者对数据分析及其在人力资源领域的应用具有非常丰富的实践经验。他们在书中深入讨论了为什么要进行人力资源数据分析,用较大的篇幅向读者介绍了应该如何针对不同的分析问题采取不同的研究设计,以及当数据集之中出现数据缺失、过期数据,甚至当人力资源分析人员无数据可用时应该如何开展工作。

本书提出的推动建立人力资源数据分析文化,完善分析项目管理和运营模式,建立数据分析团队,提升数据分析能力,正确处理与技术供应商的合作关系,一些领先组织开展人力资源数据分析的成功案例和经验分享,人力资源数据分析实践指南和建议等,对当前我国大力推动人力资源分析实践,寻找各类问题的解决方案都富有启发,可供人力资源从业者参考和借鉴。

随着第四次工业革命的到来,包括人力资源在内的各个领域正面临着前所未有的知识更新和能力更新的挑战,唯有以更加积极的态度拥抱数据分析时代的到来,才是应对这一挑战的正确选择。更新知识结构,强化数据分析技能已成为趋势,也必将成为现实。我相信,本书的出版将有助于推动我国人力资源数据分析理论研究、人才培养和管理实践的不断进步和发展。

/ 大数据在国内实现真正落地，还需要迈过三道坎 /

来源 / CPDA数据分析师网 编辑 / 协会会员处 李苗苗 日期 / 2019-12



全球来看，对大数据认识、研究和应用还处于初期阶段。特别是对我国来说，大数据真正落地，还需要迈过三道坎。

第一，数据是否足够丰富和开放？丰富的数据源是大数据产业发展的前提。而我国数字化的数据资源总量远远低于美欧，每年新增数据量仅为美国的7%，欧洲的12%，其中政府和制造业的数据资源积累远远落后于国外。就已有有限的数据资源来说，还存在标准化、准确性、完整性低，利用价值不高的情况，这大大降低了数据的价值。



同时，我国政府、企业和行业信息化系统建设往往缺少统一规划和科学论证，系统之间缺乏统一的标准，形成了众多“信息孤岛”，而且受行政垄断和商业利益所限，数据开放程度较低，这给数据利用造成极大障碍。制约我国数据资源开放和共享的一个重要因素是政策法规不完善，一方面欠缺推动政

府和公共数据的政策，另一方面数据保护和隐私保护方面的制度不完善抑制了开放的积极性。因此，建立一个良性发展的数据共享生态系统，是我国大数据发展需要买过去的第一道坎。

第二，是否掌握强大的数据分析工具？要以低成本和可扩展的方式处理大数据，这就需要对整个IT架构进行重构，开发先进的软件平台和算法。这方面，国外又一次走在我们前面。特别是近年来以开源模式发展起来的hadoop等大数据处理软件平台，及其相关产业已经在美国初步形成。

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

第三，管理理念和运作方式能否适配数据化决策？大数据开发的根本目的是以数据分析为基础，帮助人们做出更明智的决策，优化企业和社会运转。哈佛商业评论说，大数据本质上是“一场管理革命”。大数据时代的决策不能仅凭经验，而真正要“拿数据说话”。因此，大数据能够真正发挥作用，深层次看，还要改善我们的管理模式，需要管理方式和架构的与大数据技术工具相适配。这或许是我们最难迈过的一道坎了。

/ 大数据的世界之旅 /

来源 / CPDA数据分析师网 编辑 / 协会会员处 李苗苗 日期 / 2019-12



我们现在处于技术时代，创新是常态。伴随几项发展而来的一个领域是大数据领域/这个领域的多样性和广博性理所当然地成为了“性感”的形容词，在处理和数据分析数据集时，它会把你带入一个全新的维度。

大数据是指需要先进技术处理的大而复杂的数据集。这组数据的特征集中在数量上，最终术语“大”表明了它的数量。其他特征涉及到速度，指的是实时可用的大数据生成的速度，与小数据相比，生成的频率也更高。

这种数据类型的其他特性考虑可伸缩性的扩展值和可变性。Variety侧重于数据绘图的类型，包括文本、图像、音频等。

大数据应用

在这个有技术支持的环境中，大数据在很多领域都很适用。以下是大数据的一些重要领域。

金融领域

在金融领域，大数据是一股巨大的力量。它实际上负责各种金融科技系统的概念化和开发，这些系统是银行、保险、金融交易和投资等行业的新趋势。

在分析已建立的系统和检测可能出现的异常(如欺诈)时，它可以派上用场。这是该行业面临的重大问题之一，这组复杂的数据可以帮助检测洗钱计划、庞氏骗局，甚至非法访问客户的账户。

当涉及到保险和金融交易时，分析方面在发现风险和潜在的投资市场方面起作用。

媒体和通讯

在媒体和传播领域，大数据的作用至关重要，尤其是当

你专注于广告和营销的时候。在这种情况下，它有助于为特定的目标受众创建内容，从而有效地传递消息。它还可以通过度量传递的内容的性能来检查响应。

制造业

在制造业中，在设计设备时，大数据是非常重要的，以确保在行业的各个过程中成本最小化。它还专注于提高产出质量，并使其成为一个有利可图的企业。

政府

大数据在政府中扮演着重要的角色，在财务方面更是如此。它还跟踪寻求改善其服务的各部委和政府部门的产出。此外，它还在情报方面发挥作用，维护有关国家的安全。

数据科学

如果没有数据分析师，没有数据科学家的努力，大数据将仍然是一组复杂的、没有基础的数据。这是一个能够收集大量数据并具有分析和解释能力的专业人士。它们在很多情况下都派上了用场，许多处理大数据的公司需要这样的专业人员来解读数据。数据科学家可能扮演许多角色，例如数据分析师，其主要角色集中在分析方面。还有机器学习工程师的角色，科学家利用他们设计和修改算法的技能来处理这类复杂的数据。

大数据处理是未来的趋势，它在很多领域都扮演着重要的角色。数据科学家可以方便地设计系统来分析和解释信息，同时关注所涉及的算法。作为一个日益增长的广阔领域，你可以通过在这个广阔的领域进行适当的培训来满足你对这个领域的兴趣。

/ 得数据者得"天下"，学会数据分析走出"霸王步" /

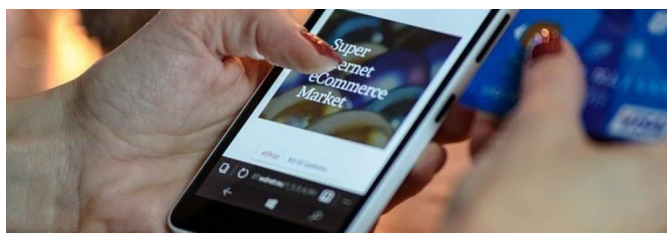
来源 / CPDA数据分析师网 编辑 / 协会会员处 李苗苗 日期 / 2019-12

在这个信息爆炸的时代，“数据分析”这个词想必经常出现在你的视线里，但你了解“数据分析”正在偷偷改变这个世界吗？

在逛网上购物的时候，你是否发现当你浏览过一件衣服以后，接下来几天同类型的商品就会出现在首页下拉的推荐页面中，刺激你再次浏览并促成消费。

在刷某音的时候，如果你经常停留于某种类型的短视频内容或经常搜索观看某个类型的内容，那么这类型的内容就会经常出现于你的推荐中，因为满足了你的内容喜好，于是你越看越嗨，根本停不下来。

又或者，当你点开外卖软件，下单一家商家的美食，那么当下一次你再打开这个外卖软件，这家店铺就会再次出现在你的主页美食推荐的靠前位置，当你回忆起对这家店铺的美食体验，又怕踩坑新店时，你或许就会决定再次下单。



其实上述这些日常生活中你经常接触到的情形，就是平台利用“数据分析”而形成的一种商业运作模式，他们根据用户日常的浏览操作数据，来提取用户喜好，从而刺激消费，不但提高了用户的使用体验，还能够更好地引导用户关注和购买商品。

除此之外，“数据分析”不仅仅只应用在你熟悉的互联网平台中，时下，它几乎影响着全球各行各业，发挥着及其重要的作用。



▲壳牌石油（图源：网络）

皇家荷兰壳牌公司集团可能大家都不陌生，对于业务延伸至全球的壳牌来说，很多看似不起眼的问题，都可能造成巨大的损失，例如各种设施中的零件位于哪里？什么时候需要再进货？机器发生故障时能否及时修理等等。

在智囊团的建议下，壳牌建立了一个数据分析平台，利用模型预测3000多种不同的石油钻井机部件出现故障的时间，每年成功为壳牌节约数百万美元。

另外，其中一个名为Databricks的工具通过Apache Spark来捕获流数据，能够更好地计划购买机器部件的时间、保存时效以及库存物品的放置地点。该工具托管于微软Azure云中，帮助壳牌将库存分析从超过48小时减少到不到45分钟，每年减少了数百万美元的库存转移和重新分配成本。

如果数据分析对于能源行业来说，是一个预测工具，可以规避风险，减少损失，那么对于零售行业来说，它就变成了可以提高销售概率的“武功秘籍”。



▲胡椒博士碳酸饮料（图源：网络）

胡椒博士集团有限公司（Dr Pepper Snapple Group）是一家美国软饮料公司，总部位于德克萨斯州普莱诺。多年以来，销售人员外出工作时都需要拿着一个装满客户数据、销售记录和促销信息的大活页夹，又笨重又不方便，为了争取像沃尔玛(Wal-Mart)和塔吉特(Target)这样的大型客户，他们还不准备很多的资料进行分析。

如今，集团给销售人员都配备了一个轻便的ipad，它可以准确地告诉销售人员需要访问哪些商店，能够为他们提供什么，以及一些针对性的关键指标，通过使用数据分析工具，当月执行销售业绩比上个月提升了50%。

这些销售人员正在使用的平台名为MyDPS，配备了机器学习和其他分析工具，当员工加载应用程序时，这些工具就可以提供推荐，提供每日的运营记分卡。算法还会给员工提供执行预期的计划，包括他们是否按能够完成计划，以及如果没有按计划完成，该如何纠正。

胡椒博士集团的首席信息官Tom Farrah表示：“他们是光荣的接单员。现在，他们正在成为智能销售人员，手中掌握着帮助其实现目标的信息。”

对于零售业来说，数据分析正在改革和优化传统的销售方式，但对于农业来说，数据分析已经颠覆了整个行业。



▲孟山都公司（图源：网络）

孟山都公司（Monsanto Company）创立于1901年，是一家跨国农业生物技术公司。从前传统的农业种植，都是参照农民的经验决定种植作物、种植地点、种植时间等等，但经验往往不够准确，经常出现失误，带来一定的损失。

现在，孟山都公司正在利用数据科学，通过数学和统计模型，绘制出最佳的时间线来种植雄性和雌性植物，并推测出最佳的种植地点，从而为种植者提供指导性的建议。

孟山都全球IT分析主管Adrian Cartier表示，计算机算法能够在几天内，处理超过900亿个数据点。2016年，孟山都因此节省了约600万美元，减少了4%的供应链足迹。

孟山都成功的关键是在它和供应链业务之间建立了一种“从摇篮到坟墓”的合作方式，他们将农业和供应链与数学和统计领域结合起来，创造出更多的价值。

这些大公司的故事告诉我们，在这个讲求科学和效率的时代，无论是哪个行业和个人，如果能够合理地利用计算机科学，有效地进行数据分析，都可能有意想不到的收获。

/ 人工智能是数字营销革命中不为人知的趋势 /

来源 / CPDA数据分析师网 编辑 / 协会会员处 李苗苗 日期 / 2019-12

数字营销的世界总是在变化的。新趋势每天都在出现。有各种各样的方法和原则被不断地引入在线营销。然而，人工智能的应用将使数字营销领域发生重大转变。

许多数字营销人员认为，人工智能将在这一行业的未来扮演重要角色。这绝对不是一个利基领域。Gartner发现，到明年，30%的公司将在一项或多项销售功能中使用人工智能。让我们谈谈这个话题，再深入一点。

人工智能与数字营销

现代计算机系统能够创造出能够自我学习和适应的软件——人工智能。这一过程包括：

认知过程——使用指导方针来尝试和预测结果。

统计采购-依据所输入的信息或数据的使用说明。

自我纠正——如果结果不准确，改变方法的能力。

基本上，有两种不同类型的人工智能：

强大的人工智能-它有潜力处理不同的任务，适应各种问题提供。

窄或弱人工智能-它的目的是明确的行动与有限数量的简单预编程的任务。苹果的Siri就是这类人工智能的完美例子。

在过去的几年里，数字营销人员一直没有信心将人工智能技术应用到他们的营销策略中。然而，现在一些成功的公司已经将人工智能应用到他们的营销计划中。亚马逊、阿里巴巴、京东已经应用人工智能向潜在买家展示相关商品，这取决于他们之前的交易、浏览和搜索。它被证明可以增加回头客的比例，并促使顾客快速购买。作为在线营销的一部分，人工智

能已经证明是一项值得的投资。

虽然人工智能技术可以大大改善你的数字营销努力，仍然需要专业的帮助和传统的数字营销方法来取得成功。

人工智能将如何彻底改变网络营销

你知道人工智能有能力收集数据，分析数据，并在正确的时间应用它。这种潜力在数字营销领域非常有价值。因此，为了做出数据导向的决策，越来越多的营销人员依赖人工智能技术。

人工智能的好处数不胜数。它们不仅让公司更容易接触到新客户。它们还可以简化现有客户参与模型的自动化。一项研究显示，到2020年，85%的客户与公司的互动不会涉及到人。人工智能技术将处理客户服务咨询。

收集数据并进行分析

常见的商业策略之一是客户关系管理。通过收集关于客户的信息，企业可以采取对客户友好的措施。例如，通过结合大数据、人工智能和CRM，可以提高从各种平台收集的客户数据的速度。这使您能够识别客户的需求，并根据数据预测运行营销活动，以最大限度地获得成功和投资回报。

对客户的实时协助

所有的客户都希望得到最好的服务和快速的响应。如今，企业经常为了这个唯一的原因而雇佣人工智能聊天机器人。它们可以作为一种独特的自动化工具来使用，从而产生一种真实的人类代表正在实时地与客户交谈的印象。人工智能是提供这种优势的最佳技术。

这些聊天机器人有潜力响应必要的客户查询，同时跟踪和履行订单。如果你在Facebook上创建了一个公司页面，你可以找到这个聊天机器人功能，不需要额外的费用。企业有机

会提供更好的客户服务标准，而不需要雇佣真正的员工。这些机器人24小时保持活跃，并减少聊天等待时间，以改善客户体验。

人工智能在其他领域的应用将继续增加。因此，品牌所有者和营销人员将更好地洞察客户的需求和期望。

网站建设者与ai启用的技术

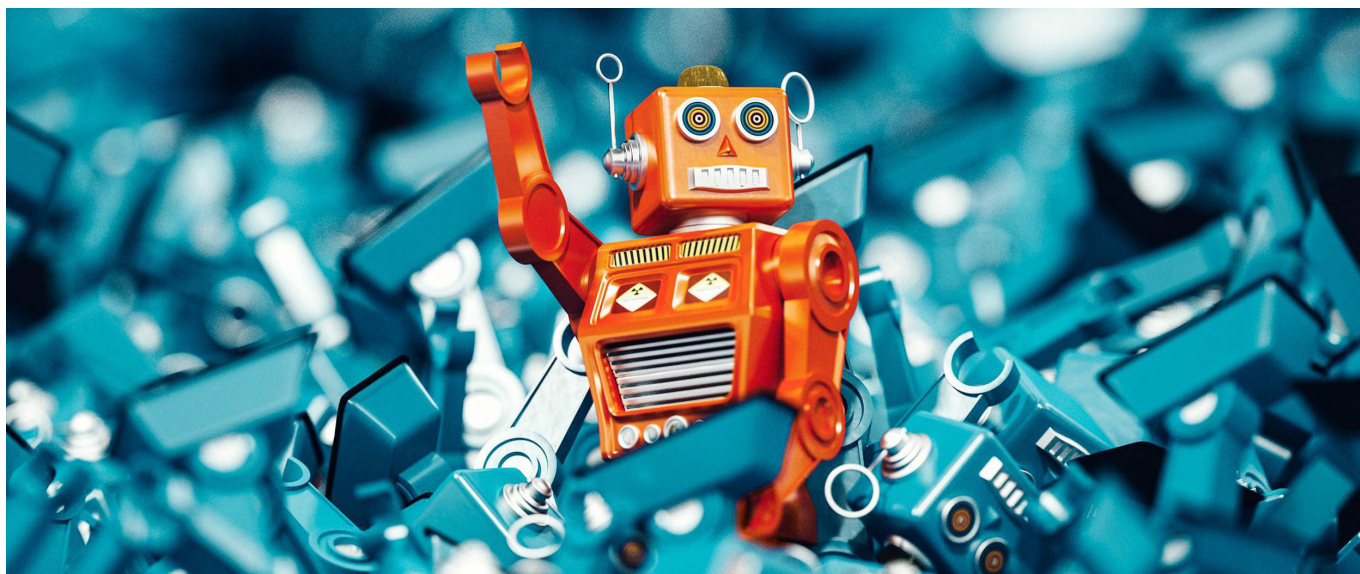
在网络营销领域，人工智能可以帮助简化网站设计，让那些几乎不懂或根本不懂技术的人能够创建一个漂亮的网站。这类产品中最著名的是Wix的ADI(人工设计智能)，它在向用户询问一些问题后创建一个网站。这并不是唯一可行的选择，许多其他的项目，如电网在市场上出现。它可以通过分析你的数据和你的互动来创建你的网站。然而，你仍然需要处理你的内容，而AI技术将处理网站页面、内容块和颜色。

更好的标准广告与人工智能

要运行一个成功的营销计划，一个主要的先决条件是接触目标受众。企业发现使用人工智能技术更容易处理在线营销活动。该技术不仅可以收集信息，还可以预测潜在客户的行为。有了准确的数据，你可以根据目标受众的喜好制作出最好的广告。如果你将它与市场调查、A/B测试和焦点小组相结合，你将最有可能创造一个令人难以置信的良好表现的活动。

人工智能正在对数字营销行业产生巨大影响

我们可以毫不犹豫地说，在推广你的品牌和业务时，人工智能是一项令人惊叹的技术。你可以通过应用人工智能技术进行数据分析、与客户互动以及开展活动来提高营销效率。自动化的下一步已经颠覆了数字营销的格局。这就是为什么你一定要考虑使用人工智能技术，并尽可能地利用它们。



/ 运用数据分析方法解读武汉市政务大数据发展 /

作者 / CPDA数据分析师 武汉 刘影、刘君钊、李雪璐、余诗路、罗欢、李典、梅平、闵鹏、郭原、龚晟

编辑 / 协会会员处 李苗苗 日期 / 2019-12

习近平总书记在中共中央政治局第二次集体学习时强调，审时度势精心谋划超前布局力争主动，实施国家大数据战略加快建设数字中国。总书记指出，要让百姓少跑腿，让数据多跑路。

一、概况与数据样本

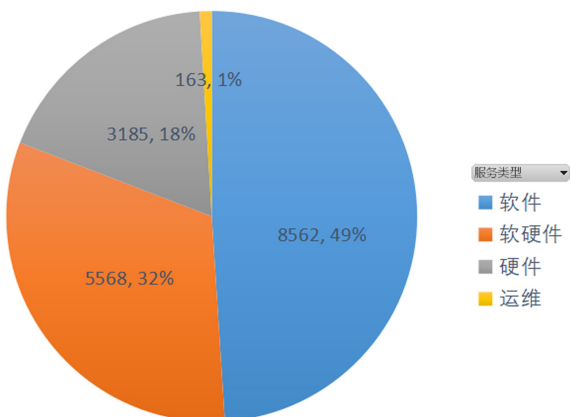
一个城市的大数据建设，往往是个循序渐进的过程，那么整个城市市场的大数据建设到了什么程度，未来是什么样的发展趋势，如何更好地贯彻落实大数据战略，助力城市大数据发展，成为一个有价值的课题。本文选择武汉市作为分析样本，进行武汉市城市政务大数据市场分析。数据样本为湖北政府购买服务信息平台公开的大数据类中标信息，本文通过对样本的数据分析，研究武汉市的大数据市场。

根据《湖北省大数据发展行动计划（2016-2020年）》，到2020年，湖北大数据产业产值规模有望达到4000亿元，带动相关产业新增销售收入过万亿元，成为经济社会发展的新引擎。截止2019年年初，中国大数据行业相关企业超过4700家，且数量仍在增长。

对湖北省政采公开招标的大数据项目做数据分析，数据维度为湖北省武汉市地区公开招标的项目（计入地点在武汉的项目，不含高校教学建设项目），从2014年至2019年10月，选择武汉市已中标的53个大数据项目，项目金额累计18280万元，平均项目规模345万元，招标的甲方累计37家，中标的乙方累计为49家。

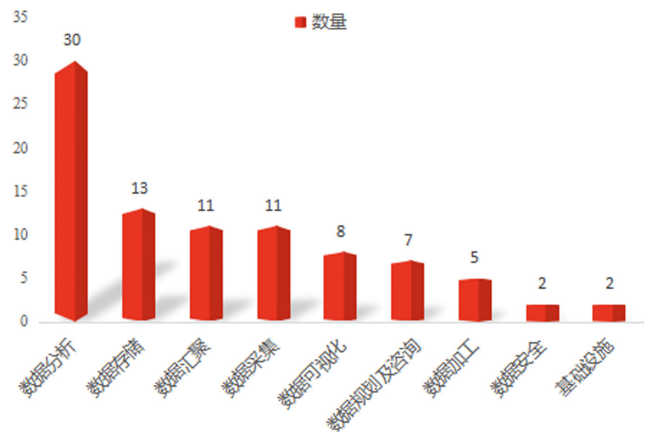
二、武汉市城市政务大数据发展数据分析

整个城市的大数据建设既包含基础设施的建设，又包含分析系统的建设。本文的已中标项目服务类型分析中，将大数据服务器、设备等建设项目分类为硬件类；平台建设、系统建设、分析咨询等项目分类为软件类。



按照项目总金额占比，纯软件类的项目占比49%，纯硬件类项目占比18%，软件硬件均包含的项目占比32%，说明当前的硬件条件能满足城市大数据发展的要求，软件、平台与分析类的需求更大。

按照大数据项目不同的任务阶段，可分为数据汇聚、数据加工、数据存储、数据采集、数据清洗、数据分析、数据可视化、数据备份、数据安全、数据规划及咨询、基础设施建设共11个阶段，因为存在一个项目覆盖多个任务阶段且项目金额无法拆分的情况，按照项目中覆盖的任务阶段做计数分析：

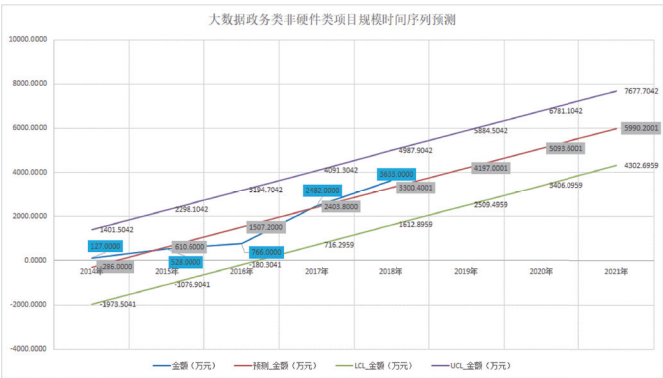


可以看到，处于数据分析阶段的项目排名第一，其次是数据存储、数据汇聚和数据采集。这四个阶段的实施和应用，都需要有很强的计算资源和基础设施做支撑，结合硬件占比金额较小的情况，可以推测出，当前武汉市的城市大数据市场的硬件建设、数据安全建设、数据基础设施建设已初具规模，整体市场更多的在数据分析及基于大数据的智慧城市应用上。

样本数据的53各项目中，涉及甲方37家，乙方49家，乙方的数据量超出甲方32%，且与总项目数几乎相等，结合项目所涉及的大数据项目阶段，可看出甲方所在的大数据发展阶段不同。而乙方的多样性表明，大数据项目需要大量的定制化和业务细分。大数据项目复杂度高，各类需求的排列组合有数千种，需要解决的问题也各不相同，但最终的目的都是通过数据分析解决实际中的问题。因此，大数据市场应当紧紧围绕业务场景和实际需求进行开发，如果标准产品无法适应政务大数据和城市大数据的发展需要，应当深入调研需求部门遇到的难题，做定制化的解决方案。

武汉市城市政务大数据市场未来发展潜力巨大，对需要通过省政采招标平台的大数据政务类非硬件类项目规模做时

间序列预测，应用ARIMA模型，预计2019年全年规模为4197万元，预计2020年全年规模为5093万元，预计2021年全年规模为5990万元。从分析图中可以看出非硬件类项目规模呈线性增长趋势，测量值与预测值的差异主要是2016年的数据特殊，偏离较大，整体不影响预测趋势。



模型拟合度

拟合统计	平均值	标准误差	最小值	最大值	5
平稳 R 方	.905	.	.905	.905	.905
R 方	.905	.	.905	.905	.905
RMSE	530.254	.	530.254	530.254	530.254
MAPE	89.982	.	89.982	89.982	89.982
MaxAPE	325.197	.	325.197	325.197	325.197
MAE	329.520	.	329.520	329.520	329.520
MaxAE	741.200	.	741.200	741.200	741.200
正态化 BIC	13.190	.	13.190	13.190	13.190

百分位数					
10	25	50	75	90	95
.905	.905	.905	.905	.905	.905
.905	.905	.905	.905	.905	.905
530.254	530.254	530.254	530.254	530.254	530.254
89.982	89.982	89.982	89.982	89.982	89.982
325.197	325.197	325.197	325.197	325.197	325.197
329.520	329.520	329.520	329.520	329.520	329.520
741.200	741.200	741.200	741.200	741.200	741.200
13.190	13.190	13.190	13.190	13.190	13.190

智慧城市的发展建设和城市大数据的发展建设密切相关，完成整个大数据产业链的建设将对武汉市的大数据发展起到重要作用。产业链的建设可根据大数据任务阶段进行对应配置：

- 1.数据采集：结合城市物联网建设，对传感器、视频流制订数据规范；
- 2.数据清洗：根据数据规范，对原始数据做预处理，使之符合数据交互的标准；
- 3.数据迁移：根据统一的数据规范和数据交互标准，使可

用数据能够在安全的条件下做交互和迁移；

- 4.数据同步：在不同的存储条件下，保障数据的一致性和时效性；
- 5.数据安全：为数据分析系统、数据共享系统做好数据安全的保障工作，做到每一条数据的痕迹和使用日志都记录在案；
- 6.数据存储：稳定、快速响应的存储设施；
- 7.数据分析：可以与云存储、云平台、软件服务、数据服务、系统集成、人工智能相结合，围绕各业务部门、各管理部门的实际职责做分析，要产生实际的社会价值或者经济价值；
- 8.数据可视化：与实际使用的业务部门、管理部门；
- 9.城市人工智能：以大量的数据为基础，具备自主学习、提升城市管理效率的能力。

三、结论与展望

智慧城市的发展建设和城市大数据的发展建设密切相关，武汉市城市政务大数据市场发展潜力巨大。当前武汉市的智慧城市建设的基础设施建设日趋完善，在城市大数据的未来展望中，将更多的在数据安全的基础上，解决数据交互、数据融合，使处在不同数据孤岛中的数据能融合到大数据共享交换平台或者城市大数据中心。

在万物互联，万物智能的新时代，打破数据壁垒、让多维度数据有机结合在一起，通过全生命周期的城市大数据分析，让智慧城市建设与数字经济发展更深入的融合，将为武汉市的腾飞提供新的动力。



/ 罪案大数据分析 /

作者 / CPDA数据分析师 广州 陈達安 编辑 / 协会会员处 李苗苗 日期 / 2019-12

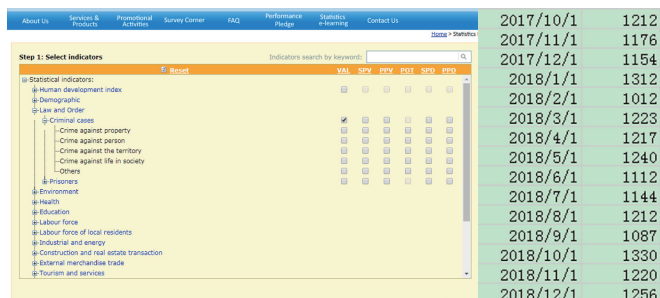
一.背景

在澳门的警务部门中，不断有警员反映近年罪案不断增多，工作量及压力大增，由于办案警员要在半年内提交总结报告予检察院，每件案件需经过多名警员共同侦办，包括录取口供，侦查，取证，公文往来等多个步骤，一宗案件需花费大量的人力物力及时间调查，若案件数量及工作量继续大增，社会治安环境下降，同时办案警员工作将超出负荷，令警队士气大减，对警队形象及市民造成影响。

故此，利用R语言以及时间序列相关分析，结合2000年-2018年18年的犯罪统计数字进行分析，并利用ARIMA寻找恰当的模型对未来两年案件数目进行预测。

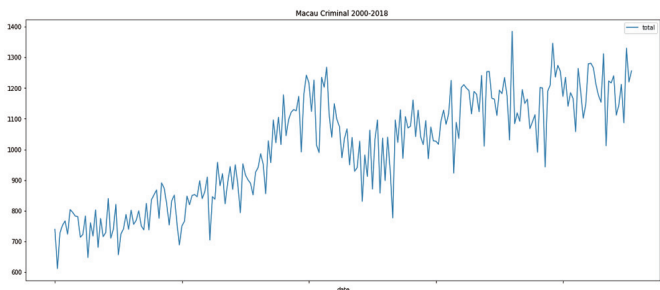
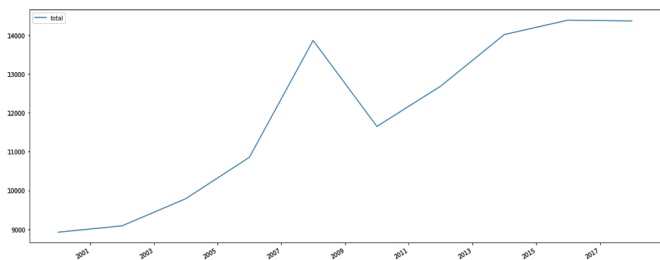
二.数据的来源及解释

澳门统计局的犯罪统计(2000-2018年的刑事案件数量)



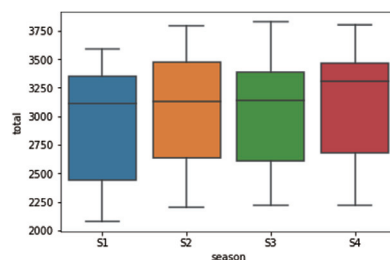
三.数据探索性分析

1. 折线图-查看趋势



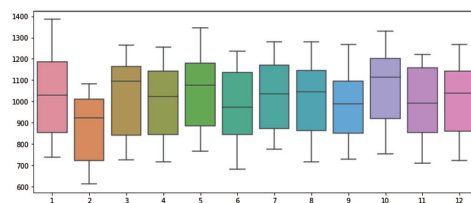
2. 箱型图-查看数据分布

结果:从折线图中可看出犯案数字有上升趋势，其中可能含有季节性因素，但仍需进一步分析。

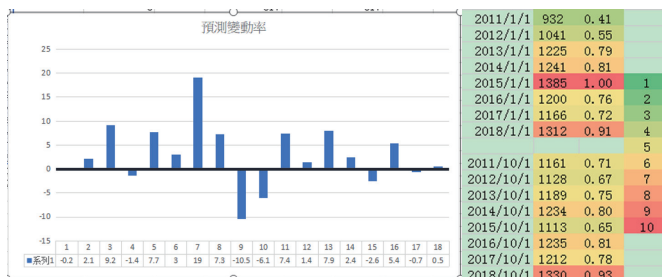


从季度箱型图中可得出：第1季中位数及第4季中位数偏高，数据分布集中在顶部，应作进一步分析。

3.月度箱型图



4.棒型图-查看变动率及极值



从月箱型图中可看出：2月份罪案数字整体比其余月份低，推断由于2月为农历新年，全城休息庆祝，澳门市民回乡过年，故2月罪案较低为正常现象。

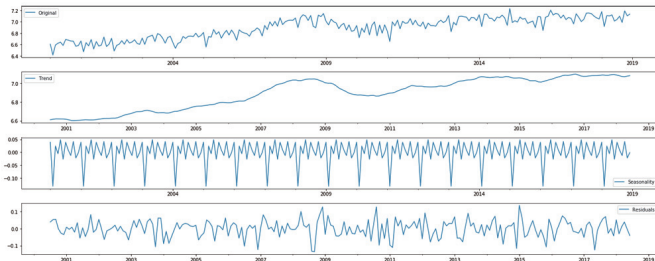
1月份犯罪数字出现最大值，但Q1,Q3,中位数与其余月份分布大致相同，1月偏高并有极值，2月减至最低，说明情况异常，值得进一步探索，而1月的极值说明被强行拉高，配合棒型图，说明2007年时罪案数字变化最大，突发增多，出现失控的情况。

按照经验中得出，普遍的犯罪为侵犯财产的罪行，由于警队宣传与市民防盗防骗意识相关，故警队的宣传仍有改善空间，必须在未来的计划中在农历新年期间加大宣传，以免再次

出现失控的现象。

10月份中位数及整体有偏高及整体上升趋势，10月为国庆及十一黄金周，内地旅客会大批涌到澳门购物，由于人多拥挤，容易制造罪犯犯案的机会，而针对10月的行动，必须在旅游区，巴上加大大宣传，提醒旅客，以免财物被盗。

5.分解序列-分解周期性因素及趋势



从图可得知序列有明显的趋势性及季节性因素，噪声有波动，由于序列中呈现线性趋势，故使用较广泛的ARIMA时间序列模型进行预测。

四.建模并撰择模型

模型1--使用R-AUTO.ARIMA自动建立模型，RMSE=76

模型2--使用DATAHOOP WINTER指数平滑建立模型，RMSE=80

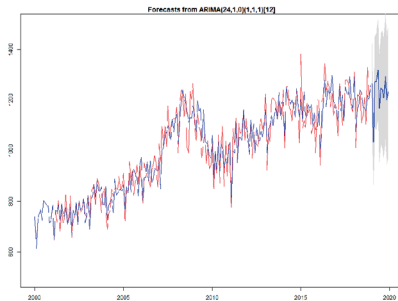
模型3--通过分析后自建立模型，各评价指标最好，参数为：

`arima(x = tsdata, order = c(24, 1, 0), seasonal = list(order = c(1, 1, 1), period = 12))`

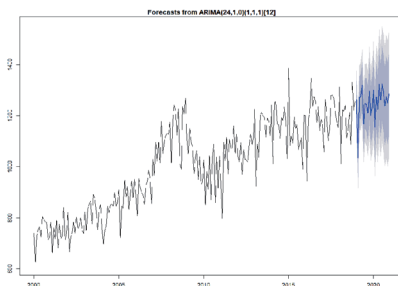
置信水平为0.05, RMSE=53 (撰择自设参数建立模型)

五.预测结果

1.拟合结果如下:



2.预测结果如下:



3.预测年度与月度的数值结果如下:

```
> x.fore
```

	Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2019	1263.986	1190.8340	1337.137	1152.1099	1375.861	
Feb 2019	1035.085	957.5968	1112.573	916.5770	1153.593	
Mar 2019	1271.191	1188.6865	1353.696	1145.0111	1397.371	
Apr 2019	1274.330	1187.0690	1361.591	1140.8759	1407.784	
May 2019	1320.476	1229.6891	1411.262	1181.6297	1459.322	
Jun 2019	1167.500	1069.8224	1265.177	1018.1151	1316.885	
Jul 2019	1248.203	1148.7488	1347.657	1096.1009	1400.305	
Aug 2019	1242.964	1136.9983	1348.929	1080.9035	1405.024	
Sep 2019	1210.039	1101.0846	1318.994	1043.4074	1376.672	
Oct 2019	1295.828	1182.0277	1409.629	1121.7854	1469.871	
Nov 2019	1201.038	1084.5776	1317.499	1022.9269	1379.150	
Dec 2019	1232.128	1112.9648	1351.292	1049.8835	1414.373	
Jan 2020	1301.754	1177.9436	1425.565	1112.4022	1491.106	
Feb 2020	1154.714	1028.5674	1280.860	961.7896	1347.638	
Mar 2020	1273.263	1142.9297	1403.597	1073.9352	1472.591	
Apr 2020	1224.621	1089.1065	1360.135	1017.3694	1431.873	
May 2020	1324.259	1184.3316	1464.187	1110.2583	1538.260	
Jun 2020	1261.211	1117.2960	1405.126	1041.1119	1481.311	
Jul 2020	1325.455	1178.7828	1472.127	1101.1393	1549.771	
Aug 2020	1286.791	1136.1460	1437.436	1056.3994	1517.183	
Sep 2020	1239.083	1085.0416	1393.124	1003.4972	1474.668	
Oct 2020	1274.228	1116.0449	1432.411	1032.3078	1516.148	
Nov 2020	1248.774	1087.2768	1410.272	1001.7852	1495.764	
Dec 2020	1285.491	1123.2813	1447.701	1037.4125	1533.570	

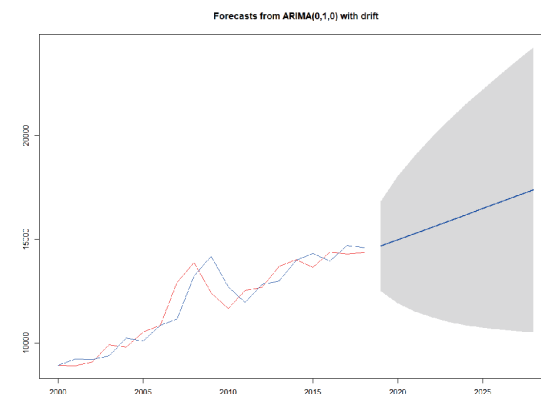
	Lo80	预测值	Hi80
2019	13583	14762	15929
2020	13462	15199	16928

4.检验预测结果及对比

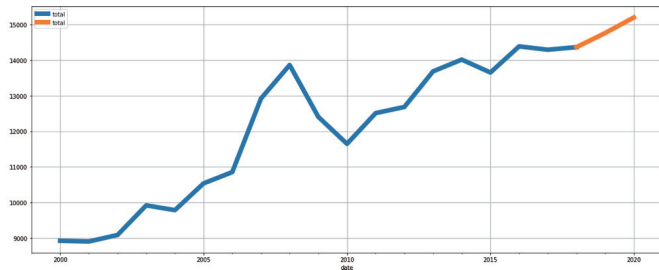
		预测值	真实值	差值
Jan	2019	1264	1320	-56
Feb	2019	1035	893	142
Mar	2019	1271	1151	120
Apr	2019	1274	1173	101
May	2019	1320	1215	105
Jun	2019	1168	1168	-1
Jul	2019	1248	1283	-35

		Lo 80	真实值	Hi 80
Jan	2019	1191	1320	1337
Feb	2019	958	893	1113
Mar	2019	1189	1151	1354
Apr	2019	1187	1173	1362
May	2019	1230	1215	1411
Jun	2019	1070	1168	1265
Jul	2019	1149	1283	1348

5.年度趋势预测

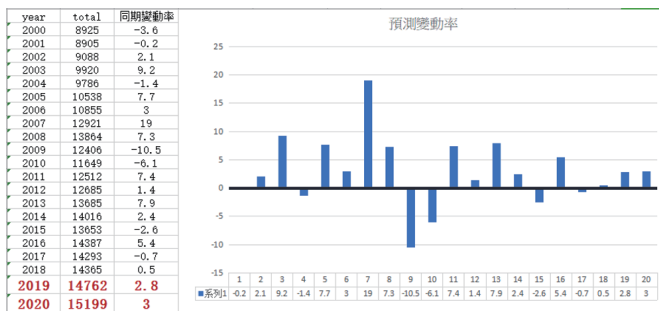


6.月度趋势预测



7.预测犯罪案件宗数及同期变动率

司法與罪案			
罪案			
總數			
	宗	同期變動率	
2019 1月	1,320	0.6	
2月	893	-11.8	
3月	1,151	-5.9	
4月	1,173	-3.6	
5月	1,215	-2.0	
6月	1,168	5.0	
7月	1,283	12.2	



a、根据图一撰择最优模型的拟合结果显示，模型拟合结果好，没有较大偏差。

b、根据图三，得出预测的月度与年度的预测值及预测区间。

c、根据图四，在给置信度0.05的前提下，真实值落在我们估计区间之内概率为99.5，而真实值落在比正常预测的估区间窄(80%置信水平)的范围附近，说明模型的预测结果较为准确。

d、根据图五，图六，进行年度与月度的预测，均显示犯罪数字在未来两年呈上升趋势，并创新高。

e、而图七的显示今年2019总犯案数字约14700宗，而至2020年将达15100宗，预测未来两年的犯罪变动率每年上升3%，创历年新高。

六.总结

为了使数据具有可比性，取2010-2018的数据正规化，将罪案数字缩放到[0-1]之间作，并分10级作对比。

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010											
1月	740	0.17	763	0.22	821	0.27	824	0.27	763	0.20	910	0.39	887	0.36	1028	0.54	1113	0.73	1109	0.64	1027	0.54
2月	612	0.00	648	0.05	657	0.06	738	0.16	689	0.10	705	0.12	794	0.24	957	0.45	992	0.49	1040	0.55	831	0.28
3月	729	0.15	761	0.19	725	0.15	836	0.29	751	0.18	846	0.30	953	0.44	1096	0.63	1180	0.73	1149	0.69	982	0.48
4月	753	0.18	719	0.14	741	0.17	852	0.31	766	0.20	838	0.29	917	0.39	1022	0.53	1242	0.82	1099	0.63	912	0.39
5月	767	0.20	803	0.25	788	0.23	868	0.33	848	0.31	958	0.45	901	0.37	1105	0.64	1217	0.78	1075	0.60	1063	0.58
6月	734	0.14	681	0.09	740	0.17	776	0.21	820	0.27	882	0.35	890	0.36	1016	0.52	1124	0.66	973	0.47	871	0.34
7月	804	0.25	775	0.21	802	0.25	891	0.36	850	0.31	921	0.40	852	0.33	1178	0.78	1225	0.79	1058	0.55	1033	0.54
8月	795	0.24	716	0.13	756	0.19	874	0.34	853	0.31	823	0.27	926	0.41	1045	0.56	1014	0.52	1067	0.59	1096	0.63
9月	783	0.22	729	0.15	769	0.20	823	0.27	846	0.30	891	0.36	941	0.43	1097	0.63	990	0.49	950	0.44	858	0.32
10月	781	0.22	840	0.29	800	0.24	754	0.18	898	0.37	944	0.43	986	0.48	1126	0.66	1235	0.81	1039	0.55	1037	0.55
11月	714	0.12	711	0.13	751	0.18	832	0.29	840	0.29	870	0.33	952	0.44	1130	0.67	1203	0.76	929	0.41	899	0.37
12月	723	0.14	740	0.17	738	0.16	851	0.31	862	0.32	950	0.44	856	0.32	1126	0.66	1268	0.85	941	0.43	1040	0.55

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020												
1月	932	0.41	1041	0.55	1225	0.79	1241	0.81	1395	1.00	1200	0.76	1166	0.72	1312	0.91	1					
2月	777	0.21	1016	0.52	923	0.40	1011	0.52	1094	0.61	943	0.43	1058	0.58	1012	0.52	2					
3月	1096	0.63	1208	0.62	1088	0.62	1253	0.83	1119	0.66	1190	0.75	1284	0.84	1223	0.79	3					
4月	102	0.02	970	0.46	1052	0.55	1058	0.52	1092	0.62	1210	0.68	1152	0.65	1250	0.75	4					
5月	1129	0.67	1073	0.60	1022	0.76	1167	0.72	1195	0.74	1364	0.95	1102	0.63	1240	0.81	5					
6月	971	0.46	1028	0.54	1211	0.77	1164	0.71	1150	0.70	1236	0.81	1147	0.69	1112	0.65	6					
7月	1107	0.64	1027	0.54	1200	0.76	1111	0.65	1154	0.71	1274	0.86	1279	0.86	1144	0.69	7					
8月	109	0.04	1017	0.52	1115	0.55	1068	0.59	1254	0.76	1150	0.73	1266	0.80	1108	0.65	8					
9月	1076	0.60	1059	0.62	1116	0.65	1182	0.74	1190	0.69	1217	0.73	1278	0.85	1087	0.61	9					
10月	1161	0.71	1128	0.67	1120	0.75	1234	0.80	1113	0.65	1235	0.81	1212	0.78	1330	0.93	10					
11月	1142	0.56	1082	0.61	1119	0.73	1175	0.72	991	0.49	1041	0.63	1108	0.67	1150	0.72	11					
12月	1128	0.56	1183	0.66	1103	0.54	1154	0.63	1154	0.63	1274	0.74	1154	0.72	1255	0.83	12					

(一) 观看2000-2018的案件有不断上升趋势，特别在2016-2019年，警级数处于第8,9级，而根据预测的情况，今年与下年的情况将会再创新高，2019年为接近14700宗[上下限为13583-15929]，2020年为15100宗[上下限为13462-16928]，每年上升3%，显示社会将面临罪犯高企的情况，而前线警员将处于工作量继续增多的困境，这样大大对社会及警队造成负面影响。

(二) 2000-2018年间的数据显示1月份为犯罪高频月份，案发率及增长率偏高，其中于2015年1月出现最大值1385宗，2013年-2018年平均案件为1245宗，偏向高值，2018年1月的案件宗数为1312，级数为第9级，接近极值，由于案件发生在农历新年前夕，建议在12月至1月期间加强向市民作宣传，以防犯罪率再持续上升，在新年前夕重新部署警力安排，加大宣传及警力巡逻，最后检查部署成果及绩效。

(三) 数据中显示10月的罪发率偏高，相信为国庆十一黄金周影响，综观过去每年10月份的案发率逐渐上升，过去整个2011-2018年10月份的案发率逐渐上升，2018年10月的案件宗数为1330，级数为第9级，接近极值，这一规律可作为警示信号，表示今年有可能再创新高，鉴于会有大量内地旅客到澳门旅游区购物，建议在口岸加强宣传及增大告示内容，并派便衣及巡逻警员到旅游及购物区巡逻，防止小偷有机可乘，以产生阻吓作用，在适当部署后，查看日后的部署效果及绩效。

(四) 减在少变动率及犯案率直接降低工作量的前提下，建议首要任务在1月及10月加强宣传及打击犯罪，(1)10月的黄金周应在主要购物区加强宣传及告示，告示牌应放署足够大幅以便每位旅客都能清晰看到，内容表达直接明确，主要警力应部署在店铺的营业及旅客的出门时段，时间为早上11时至晚上11时。(2)在1月前应提前准备对社区和旅客作宣传，并在1月期间针对社区及旅游区部署适当警力，对应社区则撰择在晚上及人少的地方加强巡逻，以防发生入屋盗窃。

(五) 由于罪案直接影响警员的工作量及压力，建议先鼓励，表扬及赞赏前线警员及对他们的工作作出认同及肯定，以提升警队士气，往后将案件分类重新调配，适当分配到次紧急的侦查部门，以缓解某部门工作量过多的情况。

(六) 未来两年的犯罪率将会持续上升，并以每年约3%

的增长增幅，对于警员，每增加一宗案件则增加一担压力，减少罪犯发生及缓解前线警员压力的任务，重要且紧急。

在此本人希望借此分析报告对市民及警务部门作出警示，市民自身做好防护措施，小心保管财物能直接断绝罪案发生，除保障自身的财产利益外，更能维持社区稳定安全；而警务部门在市民的配合下，工作成效事半功倍，双得益彰。

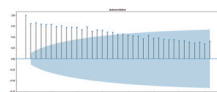
在未来让市民与警务部门积极配合，令澳门各社区安定繁荣，长治久安。

七.模型详细结果

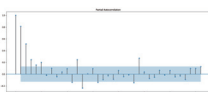
模型1 --自建模型

平稳性检验

PACF自相关系数



ACF偏自相关图



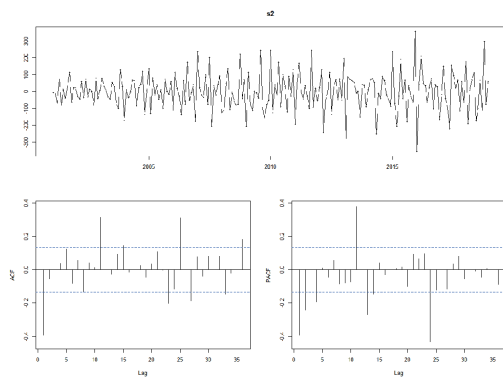
ADF平稳性检验

```
> adfres(tsddata)
Title:
Augmented Dickey-Fuller Test
Test Results:
PARAMETER:
Lag Order: 1
STATISTIC:
Dickey-Fuller: 0.2558
P VALUE:
0.6974
Description:
Wed Sep 18 11:10:52 2019 by user: Stan
```

自相关系数(ACF)呈现拖尾性(lag=36)，即序列与之前的时刻有强烈的相关性，且相关项的范围较大，另ADF检验的P值为0.69，>0.05 说明序列为非平稳序列。

消除趋势为平稳序列

在进行一阶差分及二阶差分后，发现而ACF 及PACF在lags12处有季节性周期且为强相关，因此进行一阶12季节差分，结果如下：



Augmented Dickey-Fuller Test

```
data: s2
Dickey-Fuller = -8.0355, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
warning message:
In adf.test(s2) : p-value smaller than printed p-value
```

经在对比一阶差分，二阶差分等等，发现一阶差分后12步有较好的结果，P值<0.01及Dickey-Fuller值最少，经检验后趋势基本消除了，序列较平稳。

参数撰择

经过了差分变换，通过的自相关图和偏自相关图来找到合适的 ARIMA 模型。

通过观察ACF，PACF，ACF决定Q值，PACF决定P值，拖尾值为零，截尾为撰定的值。

由于PACF是截尾，ACF为拖尾或截尾，故PACF在4,12,24,36阶后截尾，故P值撰4,12,24,36，ACF1阶后截尾或拖尾，故Q=1或0，季节性12LAG后撰项与ACF PACF一致，为1阶截尾，故SEASON P Q为1。

根据尝试各种组合，得出最优模型的参数如下：

ARIMA(24,1,0)(1,1,1)(12)

RMSE=53，AIC=2432

Ljung-Box 残差检验通过，显示残差之间不存在自相关性。

> accuracy(fit1)

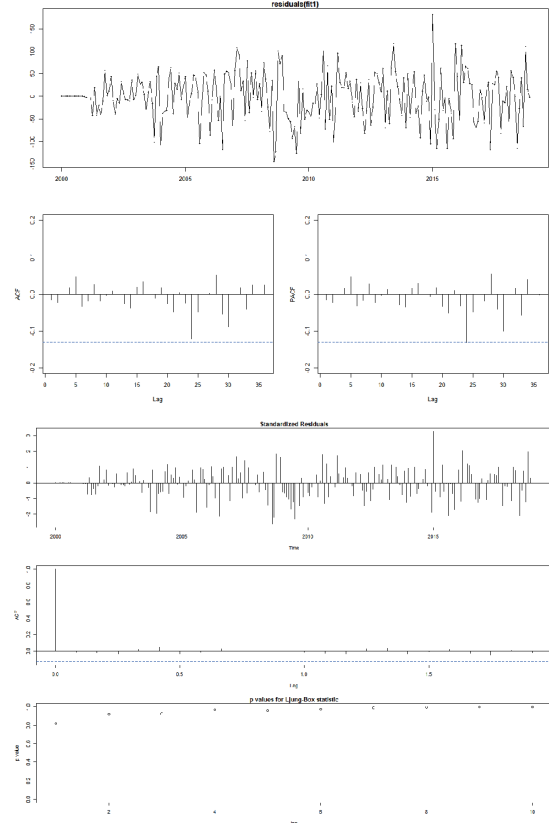
```
Training set ME RMSE MAE
```

```
MPE MAPE MASE ACF1
```

```
call:
arima(x = tsdata, order = c(24, 1, 0), seasonal = list(order = c(1, 1, 1), period = 12))
```

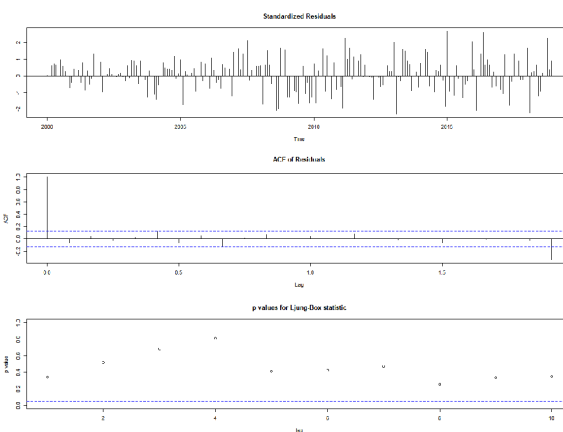
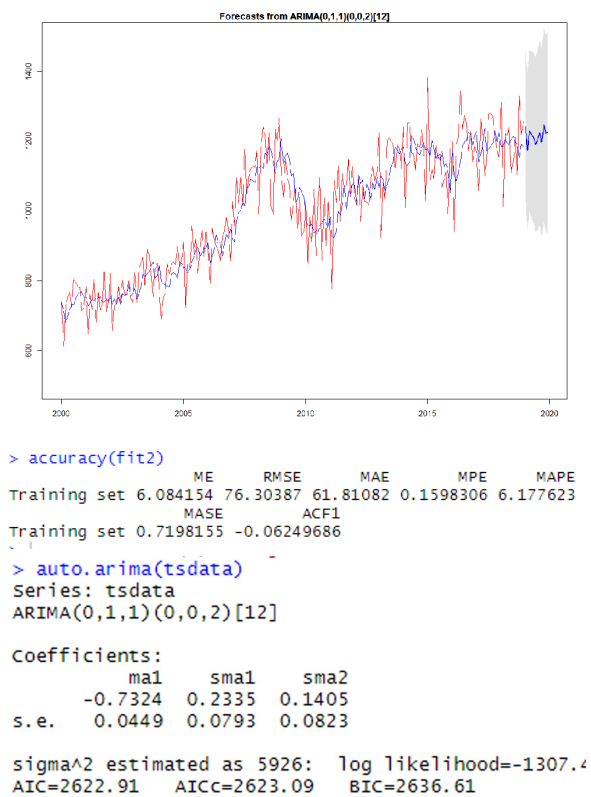
```
Coefficients:
ar1 ar2 ar3 ar4 ar5 ar6 ar7
s.e. 0.0658 0.0789 0.0817 0.0815 0.0826 0.0844 0.0841
ar8 ar9 ar10 ar11 ar12 ar13
s.e. -0.0254 0.0530 -0.0633 -0.0342 -0.3080 -0.1993
ar14 ar15 ar16 ar17 ar18 ar19 ar20
s.e. 0.0835 0.0836 0.0835 0.0828 0.1321 0.0943
ar21 ar22 ar23 ar24 sar1 sma1
s.e. -0.1093 0.0987 0.0519 0.0797 -0.0513 0.0273 -0.0104
s.e. 0.0866 0.0850 0.0885 0.0833 0.0846 0.0839 0.0835
s.e. 0.0748 -0.0157 -0.2375 -0.3283 0.2617 -0.9999
s.e. 0.0833 0.0819 0.0779 0.0730 0.1532 0.0925
```

sigma^2 estimated as 3086: log likelihood = -1189.49, aic = 2432.98

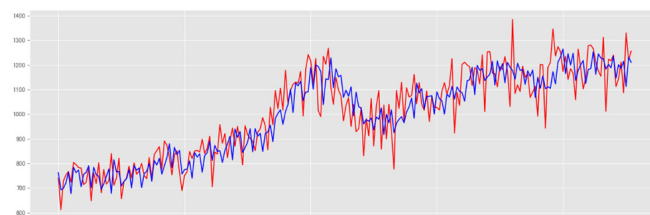




模型2---R-AUTO ARIMA:



结论:AIC=2623, RMSE=76.3, Ljung-Box 残差检验显示存在自相关性,且检验不通过。自建模型较优
模型-3使用DATAHOOP 的WINTER指数平滑

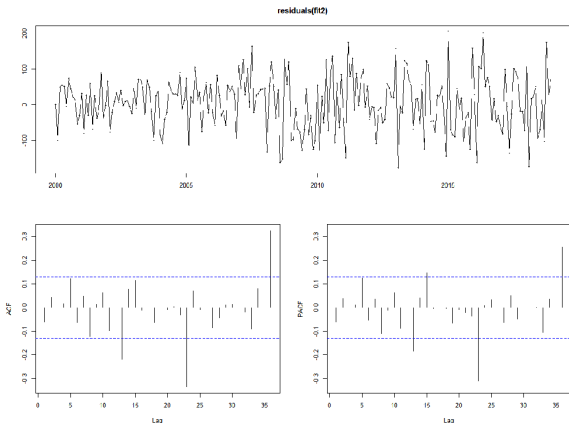


原始模型结果及检验

RMSE

80.50659475079976

结论:, RMSE=80, 自建模型较优。



/ 基于数据分析方法的运营商营业厅选址研究 /

作者 / CPDA数据分析师 武汉 黄卓、官娅芳、陈琼、周融、刘宽、乐磊 编辑 / 协会会员处 李苗苗 日期 / 2019-12

我们是CPDA武汉第16期学员，通过参加中国数据分析专业委员会数据分析师（CPDA）培训并获得认证，使我们建立了良好的数据分析思维、掌握了一定的数据分析方法和技能，并结合实际工作进行了相关的应用研究。基于数据分析方法，本文利用综合评价和层次分析建立了运营商营业厅的选址模型，为运营商营业厅选址提供了可供参考的量化评价方法，在一定程度上可以拓展相关领域的分析视角。

一、研究背景

2019年11月27日，携号转网服务正式在全国范围内运行，323.2万用户已完成携号转网。随着电信市场竞争环境的巨大变化，电信运营商之间的用户资源竞争愈演愈烈。为抢占更多市场，获取更大收益，运营商必须进行消费升级，提升服务水平。营业厅作为运营商为用户提供产品和服务的重要渠道，能为用户提供更完善的服务、更直接的体验感知，越来越多的用户通过营业厅办理业务、体验服务。营业厅的选址不仅关系到用户与营业厅渠道互动的便利性，也将影响到运营商竞争力的提升和运营目标的实现，如何进行科学选址已成为运营商需要优先考虑的关键问题。

二、运营商营业厅的选址原则

营业厅是运营商开展营销服务的实体渠道。营业厅选址需要综合运营商的整体布局、渠道建设、客户需求及竞争环境等多种因素，具有长远性和系统性。因此，运营商营业厅的选址应具备以下几个方面的原则：一是整体布局原则，营业厅的选址要符合企业的战略规划，着眼全局；二是客户中心原则，营业厅选址要围绕客户需求，尽量选择在客户集中的区域，以提高为客户服务的便利性；三是供需平衡原则，即根据拟选区域的营业厅数量和工作量饱和度来判断是否需要在该区域内新增营业厅或扩大营业厅规模；四是经济效益原则，营业厅建设的最终目的是为企业带来经济效益，营业厅选址必须考虑投入产出，以实现利益最大化；五是重点覆盖原则，既要考虑重点对象也要考虑重点区域，既要避免出现盲点，又要避免出现重叠；六是协调共生原则，营业厅选址应充分了解拟选区域内社会网点及现有营业厅数量、行业竞争情况等方面信息，合理布局，以提升不同渠道间的协同作用。

三、建立运营商营业厅选址模型

（一）确定运营商营业厅选址方法

大数据时代，运营商因其特殊的行业特征，在进行营业厅选址分析时，必须综合备选区域内的用户情况、区位优势、经济水平、人口数量、竞争状况等各种商业社会数据进行分析。营业厅选址决策是在多重资源约束条件下的优化分析，既要尽可能的增加市场覆盖面，为用户提供更为优质的服务，还要实现一定的经济效益。因此，本文从运营商营业厅选址的影

响因素出发，建立运营商营业厅选址决策的评价指标体系，采用既有定性分析又有定量分析的层次分析法（AHP）确定出各指标的权重，然后再利用综合评价法建立综合评价模型。

（二）设立营业厅选址综合评价指标体系

评价指标体系是指由多个不同角度、不同层次的具有内在联系和特性的指标所组成的有机整体。本文根据工作实际，结合实地调研和深入分析得出影响营业厅选址因素主要可以分为：社会因素、服务需求性、可见性、竞争情况、周边配套、运营特征等六大类，设立的评价指标体系如表1所示。

表1 营业厅选址综合评价指标体系

营业厅选址综合评价指标体系	社会因素	区域规划情况
		商圈及企事业单位数量
		经济发展及消费水平
	服务需求性	人口规模及密度
		运营商存量用户数
	可见性	地理区位
		营业厅可视度
	竞争情况	自建自营厅数量
		业务代理网点数量
		自建他营厅数量
	周边配套	交通便利度
		配套设施建设情况
	运营特征	选址面积
		租金或建设成本

（三）运用层次分析法（AHP）确定评价指标权重系数

层次分析法（AHP）是20世纪70年代，由美国运筹学家T.L.Saaty提出的一种多准则决策方法，常被用于多目标、多准则、多要素、多层次的非结构化决策问题，其主要分析步骤包括：建立层次结构模型、构造判断矩阵、层次排序及一致性检验。

1、建立运营商营业厅选址的层次结构模型

建立层级结构模型即根据决策目标将影响因素指标体系按照因果关系分成多个层次，如目标层、准则层、方案层等。本文结合营业厅选址综合评价指标体系，将运营商营业厅选址的层次结构模型分为目标层A、中间因素层Bi和底层指标层Ci，层次结构模型如图1所示。

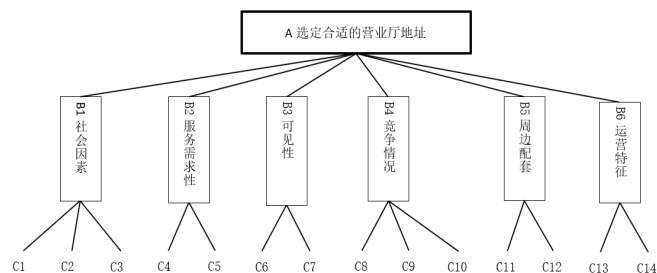


图1 运营商营业厅选址的层次结构模型

注：C1-C14分别对应评价指标体系中的14个指标。

2、构造判断矩阵

根据T.L.Saaty提出的1-9标度方法，按照层次结构模型逐层构建判断矩阵。本文基于专家评定和市场调研确定出各要素权重，并获得7个判断矩阵，如表2-8所示。

表 2 判断矩阵 A-Bi

A	B1	B2	B3	B4	B5	B6
B1	1	1/3	3	1/2	1/2	1/3
B2	3	1	5	1	2	2
B3	1/3	1/5	1	1/3	1/5	1/5
B4	2	1	3	1	5	4
B5	2	1/2	5	1/5	1	1/2
B6	3	1/2	5	1/4	2	1

表 3 判断矩阵 B1-Ci

B1	C1	C2	C3
C1	1	2	1/2
C2	1/2	1	1/3
C3	2	3	1

表 4 判断矩阵 B2-Ci

B2	C4	C5
C4	1	2
C5	1/2	1

表 5 判断矩阵 B3-Ci

B3	C6	C7
C6	1	1
C7	1	1

表 6 判断矩阵 B4-Ci

B4	C8	C9	C10
C8	1	4	2
C9	1/4	1	1/2
C10	1/2	2	1

表 7 判断矩阵 B5-Ci

B5	C11	C12
C11	1	1
C12	1	1

表 8 判断矩阵 B6-Ci

B6	C13	C14
C13	1	1/4
C14	4	1

3、层次排序及一致性检验

根据已建立的判断矩阵计算出每个判断矩阵的最大特征值 λ_{max} 及其对应的特征向量W，并引入平均随机一致性指标RI（见表9）进行一致性检验。如果检验可行，特征向量（归一化后）即为权向量；如果不能通过，需重新构造判断矩阵。本文采用根法求解，各判断矩阵的特征向量W和 λ_{max} 计算结果如表10所示。有五个矩阵排序一致性指标CI为0，具有完全的一致性，另两个矩阵排序一致性指标CI接近于0，具有满意的一致性。所有判断矩阵的一致性比率CR均小于0.1，满足一致性要求。

性要求。

表 9 随机一致性指标 RI 的数值表

阶数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
R.I	0	0	0.52	0.89	1.12	1.26	1.36	1.41	1.46	1.49	1.52	1.54	1.56	1.58	1.59

表 10 各判断矩阵的特征向量 w 和 λ_{max}

A	w	B1	w	B2	w	B3	w	B4	w	B5	w	B6	w
B1	0.090	C1	0.297	C4	0.667	C6	0.500	C8	0.571	C11	0.500	C13	0.200
B2	0.271	C2	0.163	C5	0.333	C7	0.500	C9	0.143	C12	0.500	C14	0.800
B3	0.042	C3	0.540					C10	0.286				
B4	0.304												
B5	0.122												
B6	0.171												
λ_{max}	6.559		3		2.000		2.000		3.000		2.000		2.000
CI	0.112		0.003		0.000		0.000		0.000		0.000		0.000
CR	0.088		0.005		0.000		0.000		0.000		0.000		0.000

底层指标Ci对目标层A的总排序如表11所示。由表11可知，对运营商营业厅选址影响最大的因素依次是人口规模及密度（C4）、自建自营厅数量（C8）、租金或建设成本（C14）等。在有多个营业厅地址备选时，可依据表11的总排序权重作出选择。

表 11 C层对A层的总排序

	B1	B2	B3	B4	B5	B6	C层总排序
C1	0.090	0.271	0.042	0.304	0.122	0.171	0.0267
C2							0.0147
C3	0.540						0.0486
C4		0.667					0.1807
C5		0.333					0.0903
C6			0.500				0.0210
C7			0.500				0.0210
C8				0.571			0.1737
C9				0.143			0.0434
C10				0.286			0.0869
C11					0.500		0.0610
C12					0.500		0.0610
C13						0.200	0.0342
C14						0.800	0.1368

（四）综合评价分析

根据线性综合评价法以及表11中得出的总排序权重，可得出如下综合评价模型：

$$G(A) = \sum_{i=1}^{14} w_i C_i = 0.0267C_1 + 0.0147C_2 + 0.0486C_3 + 0.1807C_4 + 0.0903C_5 + 0.0210C_6 + 0.0210C_7 - 0.1737C_8 - 0.0434C_9 - 0.0869C_{10} + 0.0610C_{11} + 0.0610C_{12} + 0.0342C_{13} + 0.1368C_{14}$$

公式中，G（A）代表目标评价值，w_i代表第i个指标的评价值，C_i代表第i个指标的权重。因第8、9、10个指标为竞争性指标，与目标值成负相关，所以要从目标评价值中减去。运营商可根据本文中设计的模型和分析结果选取出营业厅选址的最优方案。

四、结语

本文用数据分析思维和方法对运营商营业厅选址的研究和探讨，需要在工作中进一步验证价值，提升研究的有效性和科学性及工作实际应用性。通过数据分析方法，充分挖掘数据价值，产生数据效益，促进行业的发展。

/ 空调机组报价预测——基于线性回归的数据分析案例 /

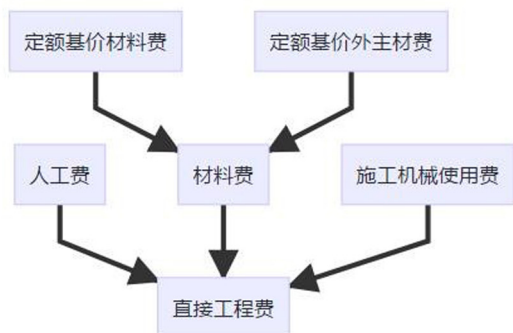
来源 / CPDA数据分析师 贵州 郑雨 编辑 / 协会会员处 李苗苗 日期 / 2019-12

导论：

建设项目的工程费用中，占比最大的费用通常是主材设备费，主材设备价格的合理性直接决定了工程费用的合理性。在日常工程造价工作中如何改善主材设备价格管理水平，是提高工作效率和质量的重要环节。

背景

从工程成本管理角度讲，直接费作为直接构成工程实体项目的现金流出，是建安工程费的重要组成部分，其中直接工程费又是直接费中的核心内容。



材料费，视具体工程项目情况，在工程费用中占比约为45%~70%，尤其在安装工程费用占比大的项目，这个比例更高。材料费的取定，除了定额基价中已经确定的材料价格外，很大一部分需要结合市场和项目需求自主定价。

为了取得较为准确的材料价格信息，造价咨询方需要向厂家、经销商询价。这部分工作，尤其是涉及到安装工程的主材设备询价，工作量常常是巨大的。

例如一个含有强弱电、消防、空调、通风、给排水等常规系统的建筑工程项目，仅安装部分不同类型的主材设备数量就可能在600种以上；一个星级酒店，安装部分不同类型的主材设备数量可能超过3000种。询价、接受和辨别主材设备报价，是一个繁琐而细致的工作。

如何更好的开展这项具体工作，使我们采用的主材设备价更为准确、合理，并能在厂商报价基础数据资料上发掘新的信息，是一个有价值的问题。

思考

笔者主要从事安装造价工作，在清单拦标价编制和全过程造价控制工作中，经常面对数以百计的主材设备需要确价。除了很少一部分能在造价信息上找到较有权威性的参考价格之外，其他绝大多数主材设备，都需要通过询价自主定价。

在明确项目定位和业主意图的基础上，主要的定价方式如下：

- 参考企业近期类似工程材料价格
- 参考互联网第三方网站材料信息
- 发布询价函，提请厂家或供货商针对项目报价

由于工程项目具有单件性特点，针对具体项目的厂商询价一般较为准确，且更具有说服力。

然而实际工作中，往往由于工作任务时间紧迫，无法逐一针对眼前项目的主材设备进行全面询价，必须充分利用以往的询价成果对新出现的主材设备进行推断预测。

实践探索

(1) 推断预测方法的确定

根据设计原理及工作经验，组合空调机组的价格，主要与风量、制冷量、电机功率三个自变量呈线性相关关系。

想要通过上述三个自变量推断价格这个一因变量，无疑应该用多元线性回归分析。

(2) 基础数据收集

某项目需要采购组合式空调机组一台，

参数要求如表一：

序号	设备名称	品牌	风量	制冷量
1	空调机组	新晃	30000	99.4
	机外余压	电机功率	含税报价	税率
	350	15		13

表一：待询价设备参数要求

近期已询价项目，该品牌报价如表二：

序号	设备名称	品牌	风量	制冷量	机外余压	电机功率	含税报价	税率
1	空调机组	新晃	7000	87	350	4	23606	13
2	空调机组	新晃	5000	6.2	350	11	19073	13
3	空调机组	新晃	7000	6.2	350	11	23606	13
4	空调机组	新晃	9000	111.8	350	5.5	27089	13
5	空调机组	新晃	6000	7.5	350	3	19073	13
6	空调机组	新晃	2500	6.2	350	7.5	14906	13
7	空调机组	新晃	15000	49.7	350	7.5	40880	13

表二：已报价设备信息

以“近期已报价设备信息”作为训练集，试构建模型，并推断待询价设备“含税报价”。

(3) 查看变量关系

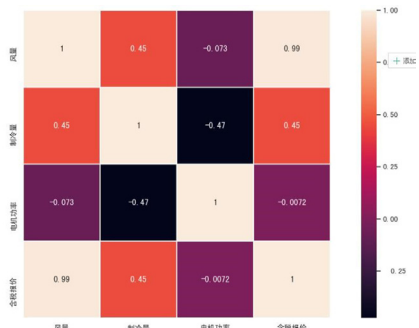
采用多元线性规划有一个前提，就是验证自变量间是否有相关性。

如果自变量间存在较强的相关性，则会产生多重共线性问题，容易导致回归结果混乱，不能合理解释变量间逻辑关系等问题，影响推断预测的准确性。

如果自变量间没有明显的相关性，则可以进行多元线性

回归分析。

以“风量”、“制冷量”、“电机功率”三个自变量及“含税报价”因变量，计算相关矩阵系数结果如图一：



图一：三变量相关性热力图

由图可知：自变量“电机功率”与因变量“含税报价”相关性不明显，可以剔除此自变量。以“风量”、“制冷量”、两个个自变量及“含税报价”因变量，并计算相关矩阵系数结果如表三：

列名	风量	制冷量	含税报价
风量	1	0.452686686	0.9923276
制冷量	0.452686686	1	0.449871939
含税报价	0.9923276	0.449871939	1

表三：两自变量相关性矩阵

由表可知：自变量“风量”与“制冷量”间，存在较强的相关性，为防止产生多重共线性问题，接纳与因变量“含税报价”相关性强的自变量“风量”，将“制冷量”剔除。

(4) 构建回归模型

已“风量”为自变量、“含税报价”为因变量，构建一元线性回归模型，结果如下：

R方	调整R方	F	
0.984714066	0.981656879	322.0980929	
P值	AIC	BIC	
9.86267E-06	120.0681641	119.9599844	
模型	系数	p值(t检验)	
风量	2.131278422	9.86267E-06	
常数项	8353.165893	0.000355915	
置信区间下限	置信区间上限	预测区间下限	预测区间上限
1.826013048	2.436543797	-6.474027038	10.73658388
5848.474592	10857.85719	8344.560588	8361.771199

表四：单变量一元线性回归模型结果

由表四可知：

自变量“风量”系数2.131278422，常数项系数8353.165893。

自变量“风量”与常数项P值（t检验）均<0.05，通过检验；模型R方0.9847，拟合效果好，模型可以采纳应用。

已此模型预测价格与实际报价比较，误差均<±10%，对设备价格推断预测具有一定价值，如下表五：

序号	风量	含税报价	含税报价_predict	误差率 (预测价-报价) / 报价
1	7000	23606	23272.11485	-1.43%
2	5000	19073	19009.558	-0.33%
3	7000	23606	23272.11485	-1.43%
4	9000	27089	27534.67169	1.62%
5	6000	19073	21140.83643	9.78%
6	2500	14906	13681.36195	-8.95%
7	15000	40880	40322.34223	-1.38%

表五：模型预测误差率统计表

(5) 推断预测

将待报价设备（见表一）作为预测集，引入上述回归模型进行设备价格预测，结果如表六：

序号	设备名称	品牌	风量	制冷量	机外余压
1	空调机组	新晃	30000	99.4	350
		电机功率	含税报价	税率	预测
		15	nan	13	72291.51856
		置信区间下	置信区间上	预测区间下	预测区间上
		72278.81236	72304.22477	72266.10615	72316.93097

表六：模型预测结果

由表可知：待报价设备预测价格为72291.52元。

为验证报价水平，再次向厂商询价此设备，报价为69954.00元，模型预测误差率为3.34%。

结论

运用线性回归分析处理单一变量或多变量的主材设备价格推断预测问题，在合理选择变量、预先排除多重共线性影响的前提下，可以得到较好的预测效果，从而提高询价报价效率，保证报价质量水平。

应用条件

需要注意的是，用此种方法虽然可以快速得到所需主材设备的近似报价，但也有一定先决条件。

· 同品牌同类型主材推断

主材设备除规格外的材质、类型、型号要求必须一致；

· 合理的时效性

由于建筑材料均具有一定的时效性，推断主材设备报价时，要特别注意基础数据采集的时间与推断时刻间材料价格在市场上是否有波动；

· 合理的自变量范围

由于在实际的生产制作中，风口规格不是没有上限的，极端大小的风口规格可能和一般常见规格尺寸存在较大的价格差异，有的产品甚至需要单独开模制作，厂商报价不一定是呈现明显线性变化的。这就导致我们不能忽略用来推断的方程中，构造方程自变量X的取值范围。因为更大或更小的规格尺寸，并不知道该厂商是否能够生产，且不知道厂商报价是否还是在同一水平上。

· 相应的采购规模

由于订货规模对厂商的报价也是一个敏感性因素，有必要记录构造函数时的报价批量信息。如果采购数量与以往偏差较大，也要考虑使用模型函数进行推断的准确性。应用时，宜在推断基础上加以适当调整。

启发与备注：

数据分析不是外行看似望而生畏的建模技术，不是一味追求高级算法的炫技表演，而是要在产业环境下、业务场景中，运用数据思维发现工作中的“真北问题”（真正的痛点并且具有指向性可量化的问题），并用可以量化的“数据语言”提出问题，采用最简单最合适的方法解决问题，进而产生效益。

/ 基于数据分析方法的电信移动业务收入预测探究 /

作者 / CPDA数据分析师 武汉 付俊、黄耘、丁琳、张嫚、罗光耀 编辑 / 协会会员处 李苗苗 日期 / 2019-12

本文探究电信移动业务收入预测，由于电信移动业务收入具有时间序列的特点，本文采用了综合ARPU法和三次指数平滑法的数学模型，结合实际进行了模型修正，从而对电信移动业务收入进行了实证分析，通过实证分析与精度检验发现，数据分析模型在电信移动业务收入预测方面具有较高的可靠性。

1.引言

当前电信服务市场的竞争异常激烈，业务收入的增长是衡量电信运营商经营绩效的重要考核指标。在电信运营商的实际工作中，全年业务收入一般基于存量收入保有率、存量收入、增量收入等历史数据的经验值进行预测，较少基于数据模型进行业务收入的预测。然而，电信业务收入明显符合时间序列的特点，其每月的态势具有稳定性或规则性，时间序列的核心思想是最近的过去态势，在某种程度上会持续到未来，即预测值是以前观测值的加权和。因此，本文拟定使用时间序列中的指数平滑法对电信移动业务收入进行预测探究，并选取某电信运营商某个分公司（以下简称“A分公司”）2016年和2017年两年的电信移动业务收入数据作为预测数据集和测试数据集，以检验指数平滑法的准确性，从而验证该数学模型在实践中应用的可靠性。

2.电信业务收入前人研究综述

所谓预测，就是根据事物以往的历史资料，应用科学的判断方法或者计量方法，对不确定性事件或者未知事件可能的演变情况，事先做出推测和估计，以指导和调节人们的行动。经查阅相关文献，业务收入定量的预测方法包括：专家预测法、回归分析预测法、时间序列法、综合ARPU法、增长率法、GDP分析法等。本文拟定采用时间序列之指数平滑法和综合ARPU法相结合进行业务收入预测。

2.1 时间序列之指数平滑法概述

时间序列是某个对象的过去与现在的数列，是按时间先后顺序排列而成的一个数列。指数平滑由布朗提出，他认为时间序列的态势具有稳定性或规则性，所以时间序列可被合理地顺势推延；他认为最近的过去态势，在某种程度上会持续到未来，所以将较大的权数放在最近的资料。指数平滑法的基本原理为：指数平滑法是移动平均法中的一种，其特点在于给过去的观测值不一样的权重，即较近期观测值的权数比较远期观测值的权数要大。根据平滑次数不同，指数平滑法分为一次指数平滑法、二次指数平滑法和三次指数平滑法等。它们的基本思想都是：预测值是以前观测值的加权和，且对不同的数据给予不同的权数，新数据给予较大的权数，旧数据给予较小的权数。

2.2 综合ARPU法概述

ARPU是单用户价值法，ARPU注重一个时间段内运营商从每个用户所得到的收入。综合ARPU法适用拥有电信企业（研究对象）在规划期内各年的综合ARPU值、规划期内各年通话出账用户数，使用方法为规划期内电信企业（研究对象）各年收入值=规划期内各年的综合ARPU值×规划期内各年通话用户数。

电信运营商业务收入的变化是有一定趋势的，而且有很强的季度型特点，由于存量用户占比较大，新增用户占比较小，因此，业务收入的历史时间越近，对未来的影响就越大；业务收入的历史时间越远，对未来的影响就越小。同时，当前电信运营商每月出帐用户数和ARPU值主要受外部行业环境的影响较大，如提速降费导致ARPU值不断降低，移动电话的普及率超100%，出帐用户数的增长越来越缓慢。由上可知，电信运营商的业务收入特点非常切合指数平滑法的数学模型和综合ARPU法。

3.指数平滑法模型的原理

3.1 传统模型的原理

指数平滑法是生产预测中常用的一种方法。也用于中短期经济发展趋势预测，所有预测方法中，指数平滑是用得最多的一种。根据平滑次数不同，指数平滑法分为：一次指数平滑法、二次指数平滑和三次指数平滑法等。由于电信移动业务收入符合三次指数平滑法数学模型的特点，模型不仅考虑了变量的线性增量，而且还考虑了变量的抛物线增量，因此，采用三次指数平滑模型进行电信移动业务收入的预测。预测电信移动业务收入三次指数平滑法的数学模型如下：

$$S_t^{(1)} = \alpha X_t + (1 - \alpha) S_{t-1}^{(1)} \quad 4.1.5-1$$

$$S_t^{(2)} = \alpha X_t^{(1)} + (1 - \alpha) S_{t-1}^{(2)} \quad 4.1.5-2$$

$$S_t^{(3)} = \alpha X_t^{(2)} + (1 - \alpha) S_{t-1}^{(3)} \quad 4.1.5-3$$

$$a_t = 3S_t^{(1)} - 3S_t^{(2)} + S_t^{(3)} \quad 4.1.5-4$$

$$b_t = \frac{\alpha}{2(1-\alpha)^2} \left[(6-5\alpha)S_t^{(1)} - 2(5-4\alpha)S_t^{(2)} + (4-3\alpha)S_t^{(3)} \right] \quad 4.1.5-5$$

$$c_t = \frac{\alpha^2}{(1-\alpha)^2} \left(S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)} \right) \quad 4.1.5-6$$

$$\text{预测方程: } X_{t+T} = a_t + b_t T + \frac{1}{2} c_t T^2 \quad 4.1.5-7$$

模型中，

$S_{t-p}^{(n)}$ 是第 $t-p$ 周期的 n 次指数平滑值；

α 是平滑系数， $0 \leq \alpha \leq 1$ ；

$X_t^{(1)}$ 是第 t 周期的实际值；

a_t 是第 t 周期的水平值；

b_t 是第 t 周期的线性增量；

c_t 是第 t 周期的抛物线增量；

X_{t+T} 是第 $t+T$ 周期的预测值， $(T \geq 1, T$ 为正整数)。

3.2 传统模型的修正

考虑到指数平滑法对数据的转折点缺乏鉴别能力，在电信移动业务收入预测中，由于2月当月只有28天，通话时长相对其他月份少2天或3天，而预测时按整月计算，这将导致2月ARPU的预测结果出现偏差。因此，对于上述情况，须专门针对2月的ARPU预测值进行修正，修正公式为2月ARPU预测值=1月ARPU预测值-（1月ARPU预测值/31）*2。

4.指数平滑法的模型训练与检验

4.1 模型的训练与测试

模型的训练与测试，拟定按照三次指数平滑法的数据采集和数据预处理，特征构造初始平滑指数及平滑系数，根据指数平滑法模型以及2016年历史数据，预测2017年的电信移动业务收入。

4.1.1 数据采集

选取A分公司2016年和2017年两年的移网业务收入数据作为训练集和测试集（数据详见表4-1），按照综合ARPU法将业务收入拆解为出帐用户数和ARPU两项，同时选择指数平滑法进行分项预测。具体如下：

表 4-1 A 分公司 2016-2017 年出账用户数及 ARPU 值表

数据分类	日期	ARPU 值（元/月）	出帐用户数（户）	数据分类	日期	ARPU 值（元/月）	出帐用户数（户）
训练集	2016 年 1 月	89.9	173141	测试集	2017 年 1 月	87.4	189273
训练集	2016 年 2 月	87.6	175043	测试集	2017 年 2 月	81.7	190448
训练集	2016 年 3 月	91.1	177387	测试集	2017 年 3 月	87.1	192925
训练集	2016 年 4 月	88.8	178766	测试集	2017 年 4 月	86.4	193836
训练集	2016 年 5 月	90.3	179886	测试集	2017 年 5 月	86.4	195203
训练集	2016 年 6 月	88.7	181093	测试集	2017 年 6 月	84.2	196356
训练集	2016 年 7 月	90.2	182839	测试集	2017 年 7 月	84.3	198388
训练集	2016 年 8 月	88.9	184634	测试集	2017 年 8 月	84.1	200086
训练集	2016 年 9 月	87.6	186114	测试集	2017 年 9 月	81.9	200497
训练集	2016 年 10 月	87.6	186946	测试集	2017 年 10 月	81.0	202139
训练集	2016 年 11 月	85.8	187411	测试集	2017 年 11 月	77.8	205735
训练集	2016 年 12 月	86.8	188391	测试集	2017 年 12 月	80.0	206535

4.1.2 数据预处理

经查，训练集与测试集数据值完整，无异常值、无缺失值，数据可用。

4.1.3 特征构造—初始平滑指数和平滑系数确定

初始平滑指数和平滑系数这两个参数的选择，对预测模型的准确度影响非常大，是模型能否应用在实际工作中的关键。

初始平滑指数的确定。利用2016年前三个月的实际收入，计算它们的算术平均值，得到模型的初始平滑指数。由此得到ARPU和出账用户数对应的初始平滑指数，如表4-2。

表 4-2 ARPU 和出账用户数对应的初始平滑指数

指标名称	ARPU（元/月）			出账用户数（户）		
	一次指数	二次指数	三次指数	一次指数	二次指数	三次指数
	平滑	平滑	平滑	平滑	平滑	平滑
初始平滑指数	89.6	89.6	89.5	175190	175190	175143

平滑系数α值的确定。由于有足够的历史数据，为更精确的得到α，使得α值更贴切实际，现基于A分公司2015年和2016年全年的历史数据，基于三次指数平滑法的数学模型，通过线性规划求解实际值与真实值之差的平方和最小值，从而求出ARPU和出账用户数对应的平滑系数α值为0.13。

4.1.4 收入模型预测

根据修正后的三次指数平滑法，A分公司2017年ARPU及出账用户数如表4-3中“ARPU预测值”和“出账用户预测值”对应列；运用综合ARPU法，通过对应月份的ARPU值×出账用户数，可以得出A分公司的对应月份业务收入，如表4-3中的“预测收入”列。

4.2 收入预测模型的检验

依据2017年实际收入与预测收入开展模型检验，收入预测误差如表4-3：

表 4-3 2017 年度预测及误差统计表

2017 年月份	ARPU 预测值（元）①	出账用户预测值②	预测收入（万元）③=①×②	实际收入（万元）	收入预测误差
1 月	86.7	189673	1644	1655	-0.65%
2 月	80.9	190985	1545	1555	-0.68%
3 月	86.0	192342	1654	1681	-1.56%
4 月	85.7	193745	1660	1675	-0.89%
5 月	85.3	195195	1665	1687	-1.29%
6 月	84.9	196691	1671	1654	1.04%
7 月	84.6	198232	1676	1672	0.28%
8 月	84.2	199820	1682	1683	-0.09%
9 月	83.8	201454	1687	1643	2.69%
10 月	83.3	203134	1693	1637	3.41%
11 月	82.9	204860	1698	1600	6.14%
12 月	82.5	206632	1704	1652	3.12%
合计			19979	19793	0.94%

由上可知，2017年度按月预测出账收入与实际出账收入对比情况，当月预测出账收入与实际出账收入误差基本控制在±3.5%之内。2017年度累计全年度的预测收入19979万元，实际收入为19793万元，预测值与实际值误差为-0.94%，远低于项目要求的1.5%准确度。

5.指数平滑法模型的展望

基于平滑指数法的模型原理，本文选取了收入波动不显著的某运营商A分公司的历史数据，运用了修正的三次指数平滑法模型进行了预测。从模型的训练和测试、模型检验过程来看，模型当月预测值和真实值的误差基本控制在±3.5%之内，模型效果拟合度高，可见修正的三次指数平滑法用于预测电信业务收入的方法可行。但在实际工作中，要保证业务收入的准确性，仍需要考虑各种外部因素对业务收入带来的影响，须重点考虑流量资费越来越低、语音需求处于萎缩、固网资费捆绑促销、携号转网政策全国开放、跨网漫游启动等因素，从而确保业务收入的预测达到预期。

/ 高致病性禽流感H5N1病毒在职业人群感染的研究 /

来源 / 北京中辉智诚数据分析师事务所 史景红 编辑 / 协会会员处 李苗苗 日期 / 2019-12

首先我和大家简单介绍一下我的背景，我在北京中辉智诚数据分析师事务所从事数据研究工作，具有医学背景，专业是流病与统计，我们这个专业主要从事的就是医药卫生方面的项目研究、数据分析、疾病控制等等。

我现在将与大家分享的一个项目，是我们曾经做的一个调查研究的项目。通过分享，我希望可以达到一个目的：就是对卫生健康领域方面的研究的某个方面有个初步了解。

2003年年底—2013年年初，高致病性禽流感H5N1感染人类，在中国的病死率是68%，发病时间多集中在冬春季，大多数病例有直接或者间接接触病死禽的暴露史。

在中国禽类养殖的这些职业人群中，他们感染高致病性禽流感的情况研究尚少，我们想了解高致病性禽流感H5N1病毒在职业人群的感染情况，以及家鸭感染高致病性H5N1病毒的情况。

为了达到这个目的，我们是怎么做的呢？通过什么方法解决这个问题呢？通过现况调查研究，根据我们的研究目的，我们查阅了大量文献，组织专家研讨会，制定调研方案等。这里，重点说一下数据获取和分析部分。



一、数据获取和处理

1、调查问卷

数据获取有多种方式，譬如互联网现在比较流行的八爪鱼抓取，或者查阅统计年鉴。我们这个项目用的数据获取办法是调查问卷。根据我们的项目目的，我们设计了2份调查问卷，一份是针对养殖人员的，包括基本信息、病死家禽接触史、病死野禽接触史、接种疫苗等情况，另一份是针对养殖户的，包括基本信息、养殖方式、死禽等信息的获取。在设计调查问卷的时候，注意编码的唯一性，这就像我们的身份证号一样，我们都用的是打印的条形码，有号码可以扫也有数字可以录入。关联几个数据库利用的就是编码的唯一性。

2、调研安排

项目	工作内容
调查方法	入户调查 采集样本（血标本，环境标本）
调查对象	养殖户
调查地点	洞庭湖地区
调查时间	冬春季
项目成员及职责	项目负责人：负责整个项目统筹，包括沟通业务需求、撰写研究方案、调查员培训、控制项目进度与质量，以及对项目成员的协调和管理。 监督员：向项目负责人汇报工作进展，负责对调查员及其调查质量直接负责。 调查员：负责实际调查访问，及时向监督员报告进度、反馈问题、接受监督员和项目负责人的监督管理。 采样人员：负责标本采集，及时向监督员报告进度、反馈问题、接受监督员和项目负责人的监督管理。 样本检测员：负责标本整理、检测、记录，及时向监督员报告进度、反馈问题、接受监督员和项目负责人的监督管理。 数据处理人员：负责调查问卷审核、数据录入、数据检查、对各地调查数据质量进行评价；负责标本检测结果数据录入、数据检查。 数据分析人员：负责对处理好的数据按照研究方案进行数据分析。 报告撰写人员：负责撰写分析报告。
项目质量与进度控制	质量控制：监督员跟访，要求拍照、当场审核。通过电话对调查对象进行抽查。 进度控制：调查当日结束进行工作小结，现场结束及时进行数据录入。



3、数据处理

调查问卷数据录入：利用epidate建立数据模型，在建库的时候就设置一些逻辑判断，然后录入数据，同时所有数据进行双录入，进行一致性检验，质控。导出excel格式，备用。

标本数据录入：因为数据格式比较简单，所以我们就利用excel建立数据模板，录入数据，抽查进行质控，备用。

调查问卷的数据库与标本数据库关联：利用血清标本条形码是唯一的且两个数据库均有，进行数据库关联。

数据清洗：

(1) 数据筛选：在我们的调查问卷中，第一个问题被调查者调查时工作或者生活是否至少2周？如果是，我们经过筛选就纳入进来，作为我们的准备分析的数据。

(2) 数据缺失值处理：我们通常有四种方法：第一种是通过被调查者电话进行回访，补充缺失数据。第二种删除缺失值个案、但是这种方法会导致样本量减少，以及样本结果发生变化。第三种，保留缺失值个案，仅在相应的分析中进行必要的排除。第四种，是用统计值来代替缺失值，如平均值、或者前后平均值，这种方法适用于数值型数据。至于选择用平均值还是前后平均值，这个要根据具体变量的特点，不一定就是前后均值优于平均值，譬如年龄，可能用整体平均值更适合，如果用前后平均值可能代表性要差异性。

二、数据分析

利用不同的统计方法，从以下几个方面分析数据：

- 1、整体情况，如年龄、性别、学历等信息。
- 2、暴露史：接触病死家禽、接触病死野禽情况等。
- 3、发热情况
- 4、接种疫苗情况：流感疫苗、其他疫苗
- 5、感染情况（阳性率）
- 6、环境标本中检出H5N1病毒。养殖人员未检出H5N1病毒。
- 7、风险因素分析

三、得出结论

禽类暴露人群感染高致病性禽流感H5N1的感染率很低。从家鸭分离的禽流感病毒有可能传播给野禽，也很可能扩散到其他陆地家禽。

完成整个项目，对于数据分析有几点体会：

1、如何解决问题？保持勿骄勿躁、沉稳有序的心态，心急手稳，查阅资料，撰写研究方案。

2、数据分析作为整个项目的一部分工作，是服务于项目的。

3、数据质量是保障数据分析准确完美的关键基础。

对于我们数据分析师来说，数据质量是至关重要的，因为不同数据质量导致不同的结果。在调研过程中，加强培训、及时沟通、态度认真，是保障数据质量的关键。

4、数据分析方法的选择，基于我们的项目需求，根据项目需求，利用Excel制作简单图表更加快捷美观，利用SPSS进行统计分析。我认为，选择什么统计工具都没有关系，只要熟练运用就可以，不要在选择方法或者设定操作的时候出现错误就可以。数据分析过程中，除了本身计划的设计分析，有时也可以大开脑洞，也许会有意外收获。即使现在是大数据时代，但对于常用的统计方法来说，没有太大变化。原来我还拿过SAS和SPSS、Excel做过同一批数据，结果没有什么太大差别。

5、最后一点，不同领域方法是相通的，不要担心没有涉猎的领域，我们就不能做。但是，也一定要注意：在进行数据分析的时候，一定要详细了解本行业领域的背景资料，这样才会更好的出具数据分析报告。

另外，我个人认为，随着大数据行业的发展，数据量足够大的时候，我们可能不在考虑什么影响因素，可能更多的关注更多的是趋势和预测未来。

/ 广西尚辰合讯数据分析师事务所 /

文 / 广西尚辰合讯数据分析师事务所 编辑 / 协会会员处 李苗苗 日期 / 2019-12

广西尚辰合讯数据分析师事务所位于中国—东盟博览会永久举办地、国家“一带一路”有机衔接的重要门户城市-绿城南宁，事务所于2019年8月8日完成工商和税局手续注册成立。



2019年10月，广西尚辰合讯数据分析师事务所通过中国数据分析行业权威机构——中国商业联合会数据分析专业委员会的考核，成为中国商业联合会数据分析专业委员会的事务所会员单位，受行业协会监管。

目前事务所技术顾问由广西大学统计学博士后副教授担任。现在事务所已成功签约成为星环信息科技（上海）有限公司和成都四方伟业软件股份有限公司教育行业核心代理商，思迈特软件有限公司广西渠道技术服务中心。

今年事务所承接广西教育厅举办的2019年广西第一届大学生人工智能设计大赛大数据建模赛道，负责赛题设计，赛事宣传推广，各高职和本科院校参赛选手赛前培训，比赛评委安排。事务所计划通过大数据分析的比赛推广，汇聚和培养高校圈数据分析团队，解决岗前培训，人员考核和科研项目落地环节。

我事务所服务范围如下：

1. 智慧校园数据平台和数据可视化项目建设

根据国家对智慧校园的建设规范，对本科和高职院校做

智慧校园顶层设计规划。通过校内数据平台建设，消除由于历史原因，造成的信息孤岛问题。通过校内数据标准制定和元数据设计，形成本校数据资产积累。采用专业的数据分析模型和数据可视化工具，对校内数据进行教学质量保障数据分析和数据挖掘，促进学校教学质量提高。

2. 行业大数据分析和数据挖掘

通过专业的大数据采集、挖掘、分析、可视化平台技术，专业的数据分析师、架构师等技术团队，为包括金融、电信、交通、物流、政府、公共安全、媒体、电力、能源、零售、制造业、医疗、教育等多个行业提供大数据分析和数据挖掘服务。

3. 产学研生态合作

9月份开学，与广西机电职业技术学院签订校企合作协议，共建大数据工作室。开展大数据分析人才联合培养，实训室共建，联合科研等项目。

4. 大数据和人工智能专业建设

通过和星环信息科技（上海）有限公司合作，提供广西区内高职、本科全套大数据和人工智能专业建设方案。包括：大数据人工智能专用实验室建设、大数据人工智能课程及教学、资源建设、师资团队建设等。

欢迎数据分析行业全国CPDA学员、数据分析师事务所同仁，以及需要我所提供支持的广大企业来访洽谈。

联系方式

联系人：韦杰

电 话：18677066168

邮 箱：549327093@qq.com

地 址：广西南宁市青秀区双拥路42号富仕园1201B室





广州数据场科技有限公司
CPDA广州授权管理中心



广州数据场科技有限公司拥有超过十年市场运营、国际猎头、企业培训经验的专业团队，与众多中大型企业保持着良好的合作关系。

得益于中国商业联合会数据分析专业委员会的指定授权，开展CPDA数据分析师在广东地区的认证培训工作。肩负起为广东地区培养大数据人才供给的重任。大数据的热度和应用必将形成广州数据场科技有限公司独一无二的庞大社群资源。业务延伸范围必将逐步迅速加深至国内外大数据企业的专业实训落地、大数据职业猎头服务及大数据分析事务所集群！

广州地区科技互联网产业发达，在大数据产业方面，已经形成气候，对数据人才的需求迫切！缺口已达数十万，未来数据专业人才的薪资必将水涨船高，CPDA数据分析师课程内容将数据分析技术与企业运营决策实务结合起来，旨在培养大数据时代能够有效对数据进行综合应用的数据分析专业化、实用型人才，为国家大数据产业发展培养专业人才。

宗旨：用数据说话，做理性决策。

愿景：让数据分析改变每个人的未来，分析引领大数据落地，搭建大数据活动生态圈。



联系方式:王老师 020-39283117 / 13352892978
咨 询 QQ:2693634131 (微信同号)
电子邮件:2693634131@qq.com
培训网址:www.gz-cpda.com
办公地址:广州市天河区天河北路183号大都会广场.1401A