



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01期 总第41期 (季刊)
咨询热线: 400-050-6600
<http://www.chinacpda.org/>
投稿邮箱: xiehui@chinacpda.org

关注CPDA数据说

China Data Analysis

中国商业联合会数据分析专业委员会 主办

“数战数决”

——没有一个冬天不可逾越，没有一个春天不会来临！

1980年，未来学家阿尔文·托夫勒在《第三次浪潮》一书中，将大数据赞颂为“第三次浪潮的华彩乐章”，那时的他估计很难预测，40年后的今天，中国正在借助大数据的力量与新型冠状病毒肺炎展开一场无硝烟的战争。

2020年开年之际，新冠肺炎疫情汹涌而至，相比2003年的SARS，十七年后新一代信息技术迅猛发展，大数据触手可及、无处不在，实时、准确、全面地运用大数据分析提供强有力的决策支撑和科技支撑，为抗击疫情保驾护航。用数据说话、用数据决策、用数据管理、用数据创新，从来不是一句空话。

回顾人类发展史，重大疫情在对人类健康、生存构成威胁和挑战的同时，往往也蕴含着突破和进步的机遇。我们不难看到：出行数据可保障返程旅客安全；搜索数据可提供权威信息指导；电商数据可智能调控物资供应；基因数据可辅助疫苗研制工作……在当前的流动人员疫情监测、智能体温监测、返程高峰督导，追踪溯源、防控救治，疫情舆情传播、减少谣言，物资需求供给、资源调配等领域，大数据都有绝佳的应用场景，已在打赢疫情攻坚战中成为一枚利器！而随着数据价值的释放，及运用数据能力的爆发，全国各地区、各行业都紧密联系在数据“防护网”中，数据和对其适当的分析已成为征服疾病的重要力量，加强数据信息共享，通过多渠道、多种类、多维度数据的整合，提升大数据技术支撑疫情防控的能力水平已是不可或缺的组成部分。

数据有着自己的全生命周期，危机之中，数据时刻在生长，也随时在更新。如果说十七年前我们因缺少大数据思维，而忽视了危机中一手数据的积累，而面对眼前这场危机，我们不仅需要专业的医护工作者，也需要大量的数据科学家，更需要知晓大数据、善用大数据的管理者。不仅要与迅速扩散的病毒“赛跑”，对疫情的发展作出超前预判，为各级疫情防控指挥机构提供科学权威的决策依据；还需要有专门的数据团队，与疾控中心等部门合作，开展数据采集和分析建模，应对下一场“战争”，研究设计成为庇护人类的“第一道关口”！群众生命的代价，换来的不该只是感伤，而还应是比病毒传播变异更快的算法，以及能够见所未见、居安思危的危机预警模型。面对这场疫情风暴，各级地方政府迎来治理能力的大考，大数据在其中发挥出了关键作用，但仍有提升的空间。福祸相倚、危中有机，深刻总结经验教训，充分运用大数据技术实现政府决策科学化、社会治理精准化、公共服务高效化，必然将更好地保障人民群众的安全感。

除了生命安危，一场突如其来的疫情，也彻底改变了我们每个人的生活和工作方式，更改变了各行各业原本的运行轨迹……疫情期间，在全国数据分析师的教育领域中，协会积极发挥自身在大数据方面的资源、平台、技术和人才等优势，在短时间内迅速启动“CPDA智慧教室”，配合Datahoop专属实操、演练平台，开启浸入式直播教学模式，实现“停课不停学”，以信息化手段助力疫情期间网络教学工作。从而也更有效地促进培养数据分析行业人才，助力挖掘大数据潜在规律及价值，提升数据决策能力，在数据分析行业中发挥重要作用！

疫情猛于虎，一场没有硝烟的战争仍在持续。而大数据是一个有温度有力量的技术，正在“智慧”相助疫情防控，“数战数决”、速战速决！早日胜利！期待春暖花开，你我相遇！

没有一个冬天不可逾越，没有一个春天不会来临！

中国商业联合会数据分析专业委员会

邵东生





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

本期目录 CONTENTS

卷首语

- 01 “数战数决”——没有一个冬天不可逾越,没有一个春天不会来临!

协会动态

- 05 2019年中国商业联合会工作总结会议在京成功举行
05 数据分析行业关于抗击新型冠状病毒肺炎疫情倡议书
06 CPDA数据分析训为您带来不一样的课堂,一样的精彩
08 CPDA助力东航人才升级,精准教学获一致好评!
10 用大数据思维武装自己 | 数据分析线上沙龙活动

行业动态

- 17 顺丰2019生鲜大数据报告:揭秘中国式生鲜物流的“超能力”
20 大数据还原病毒扩散路径:哪些城市压力更大?
25 企业数据分析的10种思维和技巧

学数交流

- 27 从几个传染病模型定量地去理解当前的措施
34 遗忘在角落里的数据库-Access
36 城市垃圾污染也是新冠病毒传播途径
37 大数据在识别电信罪案中涉及的统计学原理浅析
39 运用数据分析评估营销效果
41 微信公众号文章量化分析

事务所专栏

- 43 基于新型冠状病毒肺炎疫情下的数据与模型思考
49 疫情下的《社区互联网应用》调查问卷分析报告



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

2020年03月出版 总第41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》行业特刊的支持! 投稿邮箱: xiehui@chinacpda.org

2019年中国商业联合会工作总结会议在京成功举行

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 / 2020-01

为了认真贯彻落实党的十九大和十九届二中、三中、四中全会和中央经济工作会议精神，深入学习贯彻习近平总书记系列重要讲话精神，牢固树立和贯彻落实创新、协调、绿色、开放、共享的新发展理念，中国商业联合会于2020年1月中旬在中国科技馆召开了2019年工作总结会议。



中商联党委书记、会长姜明就中商联党委2019年工作总结和2020年工作要点作了工作报告，对中商联党委2020年工作进行了部署；中商联党委副书记、常务副会长兼秘书长王民对中商联2019年工作进行了总结，提出了2020年工作思路和重点任务，同时提到数据分析专业委员会在全年取得的成绩并给予肯定；中商联党委副书记兼纪委书记周京英就中商联纪检监察工作进行了报告，会议由中商联副会长、党委常委傅龙成同志主持。

在《中国商业联合会分支机构2019年度考核结果的通报》中，数据分析专业委员会在2019年开展工作中遵守各项规章制度，服从管理，在业务范围内积极开展活动，工作业务内容和项目丰富，形成了品牌活动，在行业内具有较大影响力，获得“2019年度优秀分支机构”的荣誉！

数据分析行业关于抗击新型冠状病毒肺炎疫情倡议书

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 / 2020-02

数据分析行业会员单位、行业同仁们：

当前，新型冠状病毒肺炎疫情防控工作处于关键阶段，习近平总书记2月3日再次主持召开中央政治局常委会，对疫情防控工作作出一系列重要部署。

为进一步认真贯彻落实党中央、国务院一系列决策部署，切实加大规范有序参与疫情防控工作的工作力度。中国商业联合会数据分析专业委员会向全体会员单位及行业同仁发出倡议：积极行动起来，携手努力，为做好疫情防控工作贡献行业的智慧和力量。

一、坚定信心，坚决拥护决策部署。深刻学习领会习近平总书记“始终把人民群众生命安全和身体健康放在第一位，把疫情防控工作作为当前最重要的工作来抓”等一系列重要指示精神，始终坚信有党中央坚强正确的领导、有全国人民众志成城，我们一定能共克时艰，打赢这场抗击疫情的阻击战。团结带领工作人员坚决贯彻党中央、国务院关于疫情防控的决策

部署，坚守政治纪律和政治规矩，不信谣、不传谣、不造谣，全面贯彻“坚定信心、同舟共济、科学防治、精准施策”要求，依法有序参与疫情防控，做到帮忙不添乱。

二、落到实处，全力做好自身防护。增强自我防护意识，掌握自我防范本领。要积极配合所属地区做好防控工作，自觉遵守疫情防控的各项规定，工作人员出现疑似症状立即隔离，发现疫情主动向所居住街道相关部门报告。非特殊情况不进入湖北疫区，因支援需要前往疫区的人员要做好防护措施，返程后及时组织人员隔离、车辆装备消毒等。要自觉采取必要的防控措施，参照所属地区关于企业复工时间等稳妥安排恢复上班事宜，做好流动人员的筛查检测、防控情况统计上报工作，落实早发现、早报告、早隔离、早治疗等措施；在疫情得到有效控制之前，一般不得举办年会、展会、研讨会、论坛、讲座、培训等各类聚集性活动，采取网上办公、通讯联络等措施，减少人员流动，降低疫情发生风险。

三、勇于担当，主动履行社会责任。充分发挥自身优势、发挥行业自律作用，进一步规范行业行为，配合有关主管部门共同维护疫情期间市场秩序；充分发挥业务专长和行业示范引领作用，结合各行业疫情防控工作的需要，配合行业管理部门做好所在行业的疫情防控工作，为打赢疫情防控阻击战做出贡献。

四、正确科学的舆论宣导。全体同仁要及时关注政府机构发布的疫情动态，坚持正确的舆论导向，杜绝一切恶意谣言和不实流言，及时、准确地解读国家政策规定、科学防护知识和各级政府发布的防控措施。利用自身宣传渠道，构建有影响

力的舆论宣导体系，营造有效防控疫情的社会氛围。

各会员单位要及时总结疫情防控中涌现出的典型事例和先进人物，并将相关事迹报送协会和相关新闻媒体，协会将及时进行宣传报道。

五、行之有效的行业自救。鉴于可以预料的经济下行压力，会员单位间要携手互帮、融通发展，协会也将予以推动。

我们坚信在以习近平总书记为核心的党中央坚强领导下，全国人民团结一心、同舟共济，不断凝聚起众志成城抗疫情的强大力量，一定打赢这场疫情防控的阻击战！

CPDA数据分析师为您带来不一样的课堂，一样的精彩

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 / 2020-02

2020年，一场突如其来的疫情，彻底改变了我们每个人的生活和工作方式，更改变了各行各业原本的运行轨迹，与你我息息相关的教育行业就发生着这样的“变革——通过在线方式完成考、培、学成为当下探讨的热点。

2019年底，这一创新教学模式已顺利完成测试在北京、广州地区启用。2019年底，在中国商业联合会数据分析专业委员会的督导下，CPDA数据分析师培训引入了“智慧教室”线上教学新模式，配合Datahoop专属实操、演练平台，真正实现浸入式直播教学时代。此次疫情来临，智慧教室立刻能大显身手，解决广大学员的燃眉之急。进入智慧教室，宅在家一样能不断精进，高效实现隔离期间充电，复工之日逆袭！

走进CPDA智慧教室，最大的特点是我们摒弃了录屏直播和一些传统的老师“大头像”直播模式。

将整个教学场景完整的呈现在学员的电脑屏幕前，无论你在哪个城市都能和授课老师进行实时互动。真正做到教学范式、信息技术、教学空间深度融合，让每一节常态课都兼具“及时反馈、有效激励、灵活互动、讲义同步”的特点。



CPDA智慧教室全方位构建教学新生态

教学场景高保真还原,仿佛置身教室

线上开课是很容易的，但到底能不能上好课是另外一回事。“教育在线化并不是一个直播工具+老师就能搞定的，而是一个闭环。工欲善其事，必先利其器，CPDA数据分析师培养的核心要素是“学习效果”，这不是一个有和无的问题，而在于是不是“身临其境”——

有了这些“神器”加持。

一场在线教学的飓风行动即将开启↓↓↓



前后左右“四个方位”均配备多机位跟踪摄像机，可360度追踪并直播老师授课全过程。



身临其境的CPDA智慧教室PK传统在线课堂



4K超高清交互智能黑板融合多项黑科技

不仅能让学生更清晰地观看教学内容。
还可以随时切换讲师讲义。
调取学生答题结果进行分析。

优质教育资源共享,全国统一输出标准

据统计,单纯以录屏授课、或简单线上宣讲方式授课的在线课程,教学质量会衰减至少20%,学员在15分钟后会因为缺少互动,分神、浏览网页、接听电话等等。CPDA数据分析师科研团队和技术团队,在这个方面做出了相应的投入和研究。综合了17年培训经验及多地优秀学员反馈,通过科技的应用拆分“学”与“习”,通过“双师模式”,“主讲老师”负责学生课堂的“学”,“辅导老师”则负责学生课中与课后的“习”。

通过学习通教学软件+教学群+Datahoop专属分析平台的综合应用,轻松完成课前预习、远程上课、教学群督学、考前辅导、在线互动等所有环节,轻松考取证书。



方便、快捷的在线签到

因材施教+黑科技在线课堂,让学习不再枯燥

每个人的学习进度有差异,即存在个性化需求,因此每个人都拥有独一无二的学习路径。CPDA数据分析的学习过程

中同学们可以先进行自测,根据测试结果选修课程更有针对性的提高自己的短板。

每个人在课堂中的表现、状态、情绪等方面也不同,这就导致互动方式需要有变化。

智慧教室运用“黑科技”能够捕捉学员的每一个细微表情,迅速分析出本节课学生的理解情况,教师的教学水平,从而提醒授课老师,是否要调整课堂气氛、学员是否需要休息、是否要来个随机抽取提问等环节来提高互动,为学生们带来更高效的学习体验。



多种课堂互动方式

2019年CPDA智慧教室已经在北京、广州等地启用反响热烈,为了响应国家号召暂停线下授课,我们将陆续在全国多地开通智慧教室教学,让更多小伙伴体验沉浸式在线教学。



CPDA助力东航人才升级，精准教学获一致好评！

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 / 2020-02



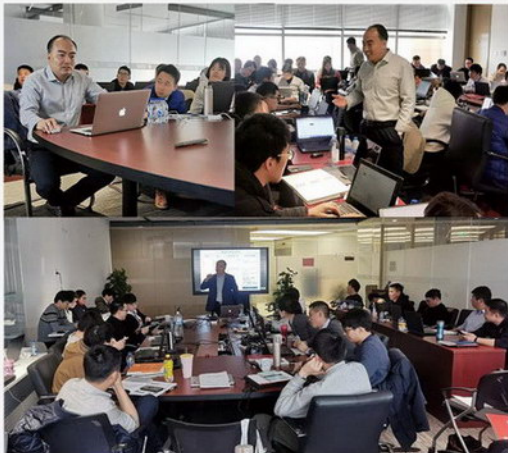
科技大繁荣的时代，大数据及大数据分析可以为航空公司提供研究未来旅行市场趋势的更多机会，在机务维修、收益管理、市场营销、人员配备和燃油管理等各个领域的业务中发挥作用。利用大数据尝试新想法、新办法，提高运营效率、提升消费者体验、开辟新的收入渠道、增加消费者粘度更是航空公司数据管理的终极秘诀。

在这样的大背景下，CPDA数据分析师为中国东方航空股份有限公司上海营业部数据洞察部的工作人员，就“CPDA数据分析系列课程”进行了专题培训。



中国东方航空股份有限公司（以下简称“东航”），是我国三大国有骨干航空运输集团之一，机队总体拥有600架飞机，排名全球第九。业务跨越177个国家、1062个目的地，年总运输量超过1亿人次，排名全球第七。

此次CPDA数据分析师课程的参训人员均为中国东航上海营业部数据洞察部的业务骨干，其中不乏博士、硕士学位等学历工作人员。



课堂讲解

学员们通过学习大数据思维、业务逻辑、实战案例、技术工具、分析框架及企业内部运营的数据分析课程体系，巩固了大数据分析相关技术及知识点，提升学习效果，培训反馈非常好。

CPDA数据分析师培训课程注重培养学员的实战经验，课程来源于众多企业经营数据管理与应用项目的提纯，结合数据分析理论的讲解与数据分析模型的实操，培养学员数据分析思维，开拓学员思考空间，助力学员挖掘数据潜在规律及价值，提升自身数据决策能力，并提供丰富的数据分析解决方案，引导学员结合自身所在行业与环境分享多样化数据分析解题思路。通过完整的培训体系培养学员的全局观、大局观，既可以自顶向下的探索数据背后蕴含的价值，又可以自底向上的去实现数据获取、数据挖掘、以及数据决策的全流程，以适应大数据时代的发展。

其次，本次CPDA的培训课程，是由有着10余年资深数据分析职场导师经验的王兴海老师授课。

最后，在课程内容上，CPDA的课程内容及优势正好与东航数据洞察部的需求相契合：

更加注重提升员工的数据洞察力；培养数据化思维能力、逻辑思维能力、商务决策能力；为学员增加数据分析算法在航空案例中的应用实战。



学员互动

本次课程，同学们都表示收获很大。最后的互动环节，大家也纷纷举手，希望王兴海老师给予数据分析及应用方面的更多指导。



课间交流



合影留念

未来，CPDA数据分析师还会为大家带来更多的数据分析领域的精品课程与实战案例经验，希望对大家有帮助。

用大数据思维武装自己 | 数据分析线上沙龙活动

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 / 2020-02

这场新冠肺炎疫情的战斗中，大数据发挥的作用有目共睹，无论是感染人群流动监测、智能体温监测、返程高峰督导、疫情走势分析，还是病毒溯源、防控救治、资源调配都发挥着决定性作用。大数据分析已在打赢疫情攻坚战中成为一枚利器！作为数据分析行业的从业人员，或者单纯对这个行业感兴趣的人群，我们在这段“被迫宅”的日子里可以利用时间，抓住机会，迅速充电，实现复工后的华丽晋升！

2月23日，云上江西携手中国商业联合会数据分析专业委员会，特邀西安交通大学管理学博士、中国商业联合会数据分析专业委员会研发师李妹老师，为江西数据分析爱好者带来《数据分析思维模型与案例讲解》的直播沙龙活动。旨在培养学员数据分析思维，开拓学员思考空间，助力学员挖掘数据潜在规律及价值，提升自身数据决策能力，通过用户画像、购物篮分析、疾病预测、房产价格预测、银行信贷等实操案例，引导学员结合自身所在行业与环境分享多样化数据分析解题思路。本次活动吸引了来自江西地区的新能源环保、航天航空、软件开发、互联网电商、通信科技等领域的500名数据分析爱好者参加。

此次线上沙龙活动不同于以往线下传统模式，而是通过CPDA智慧教室，在线上进行了一场体验感十足的“云直播”。在2个多小时的直播过程中，广大数据分析爱好者身临其境的参与到了此次活动，与老师在线上进行互动交流、深入探讨学习。让想了解数据分析的同学们，体验了一把“宅家沙龙”的畅快，现在就让我们一起来回顾吧！

线上沙龙回顾一：数据分析的前沿应用案例



直播开始，李妹老师通过新冠肺炎的疫情模型分析、facebook用户画像精准推送及婚姻离婚率预测等当下热点问题分享作为案例，将同学们带到了数据分析的场景之中，并让大家深刻理解数据分析在各行各业中的实际应用价值，大到政

策制定、小到个人决策，一切问题都可以用数据说话。

线上沙龙回顾二：学习数据分析的一些误区

紧接着，李妹老师分享了自己在数据分析工作中的实战经验结和小秘籍，为大家讲解了在学习数据分析过程中，遇到的一些错误和误区，让大家深刻理解了思维、方法与工具之间的优先级排序，正确、快速地建立数据分析知识体系，少走弯路。

线上沙龙回顾三：数据分析师知识体系构建

在2017年的世界物联网大会上，马云提到：未来的数据分析师将被人工智能取代。



数据分析师将要被人工智能替代了吗？

接下来，为大家正确地看待数据分析师这个职位。李妹老师以“数据分析师都将被人工智能取代？”的疑问抛出了一个值得大家深思的问题，当然也给出了答案——数据分析的核心竞争力应是思维方式，而不是重复性的抓取数据的工作。人工智能只能解决数据分析工作中最初级的问题，让数据分析师留出更多精力聚焦在分析本身。同时鼓励让大家成为能进行趋势预测、行业研究、评估的数据分析专业人才，最终为企业保驾护航。

线上沙龙回顾四：数据分析思维六步法及其案例解析

最后，李妹老师结合福建移动和香港赌马公司的真实案例，为我们讲解并验证了数据分析思维6步法的实现过程。让大家清晰完整地了解数据分析的思路，真正地通过数据分析解决工作、生活中遇到的各种问题。

近两个小时的精彩讲座分享，兼具理论与实践、讲授与互动，并穿插以李老师从业心得和经验分享，让所有收听参加此次线上沙龙活动的小伙伴收获颇丰，对于此次寓意深刻而又富有实战作用的课程，大家纷纷表示大有收获，并纷纷期待下期的线上沙龙活动。

习近平：要鼓励运用大数据、人工智能、云计算等数字技术，在疫情监测分析、病毒溯源、防控救治、资源调配等方面更好发挥支撑作用

来源 / 新华社 编辑 / 协会会员处 李苗苗 日期 / 2020-02



中共中央总书记、国家主席、中央军委主席、中央全面深化改革委员会主任习近平2月14日下午主持召开中央全面深化改革委员会第十二次会议并发表重要讲话。他强调，确保人民群众生命安全和身体健康，是我们党治国理政的一项重大任务。既要立足当前，科学精准打赢疫情防控阻击战，更要放眼长远，总结经验、吸取教训，针对这次疫情暴露出来的短板和不足，抓紧补短板、堵漏洞、强弱项，该坚持的坚持，该完善的完善，该建立的建立，该落实的落实，完善重大疫情防控体制机制，健全国家公共卫生应急管理体系。

中共中央政治局常委、中央全面深化改革委员会副主任李克强、王沪宁、韩正出席会议。

习近平在讲话中指出，这次抗击新冠肺炎疫情，是对国家治理体系和治理能力的一次大考。要研究和加强疫情防控工作，从体制机制上创新和完善重大疫情防控举措，健全国家公共卫生应急管理体系，提高应对突发重大公共卫生事件的能力水平。

习近平强调，要强化公共卫生法治保障，全面加强和完善公共卫生领域相关法律法规建设，认真评估传染病防治法、野生动物保护法等法律法规的修改完善。要从保护人民健康、保障国家安全、维护国家长治久安的高度，把生物安全纳入国家安全体系，系统规划国家生物安全风险防控和治理体系建设，全面提高国家生物安全治理能力。要尽快推动出台生物安全法，加快构建国家生物安全法律法规体系、制度保障体系。

习近平指出，要改革完善疾病预防控制体系，坚决贯彻预防为主的健康工作方针，坚持常备不懈，将预防关口前移，避免小病酿成大疫。要健全公共卫生服务体系，优化医疗卫生资源投入结构，加强农村、社区等基层防控能力建设，织密筑牢第一道防线。要加强公共卫生队伍建设，健全执业人员培养、准入、使用、待遇保障、考核评价和激励机制。要持续加强全科医生培养、分级诊疗等制度建设，推动公共卫生服务与医疗服务高效协同、无缝衔接，健全防治结

合、联防联控、群防群治工作机制。要强化风险意识，完善公共卫生重大风险研判、评估、决策、防控协同机制。

习近平强调，要改革完善重大疫情防控救治体系，健全重大疫情应急响应机制，建立集中统一高效的领导指挥体系，做到指令清晰、系统有序、条块畅达、执行有力，精准解决疫情第一线问题。要健全科学研究、疾病控制、临床治疗的有效协同机制，及时总结各地实践经验，形成制度化成果，完善突发重大疫情防控规范和应急救治管理办法。要平战结合、补齐短板，健全优化重大疫情救治体系，建立健全分级、分层、分流的传染病等重大疫情救治机制，支持一线临床技术创新，及时推广有效救治方案。要鼓励运用大数据、人工智能、云计算等数字技术，在疫情监测分析、病毒溯源、防控救治、资源调配等方面更好发挥支撑作用。

习近平指出，要健全重大疾病医疗保险和救助制度，完善应急医疗救助机制，在突发疫情等紧急情况时，确保医疗机构先救治、后收费，并完善医保异地即时结算制度。要探索建立特殊群体、特定疾病医药费豁免制度，有针对性免除医保支付目录、支付限额、用药量等限制性条款，减轻困难群众就医就诊后顾之忧。要统筹基本医疗保险基金和公共卫生服务资金使用，提高对基层医疗机构的支付比例，实现公共卫生服务和医疗服务有效衔接。

习近平强调，要健全统一的应急物资保障体系，把应急物资保障作为国家应急管理体系建设的重要内容，按照集中管理、统一调拨、平时服务、灾时应急、采储结合、节约高效的原则，尽快健全相关工作机制和应急预案。要优化重要应急物资产能保障和区域布局，做到关键时刻调得出、用得上。对短期可能出现的物资供应短缺，建立集中生产调度机制，统一组织原材料供应、安排定点生产、规范质量标准，确保应急物资保障有序有力。要健全国家储备体系，科学调整储备的品类、规模、结构，提升储备效能。要建立国家统一的应急物资采购供应体系，对应急救援物资实行集中管理、统一调拨、统一配送，推动应急物资供应保障网更加高效安全可控。

会议审议通过了《中央全面深化改革委员会2019年工作总结报告》、《中央全面深化改革委员会2020年工作报告》。

会议审议通过了《关于新时代加快完善社会主义市场经济体制的意见》、《企业职工基本养老保险全国统筹改革方案》、《赋予科研人员职务科技成果所有权或长期使用权试点实施方案》、《关于深化事业单位改革试点工作的指导意见》、《关于深入推进国家高端智库建设试点工作的意见》、《关于推动基础设施高质量发展的意见》、《关于进一步推进服务业改革开放发展的指导意见》。

会议指出，新时代加快完善社会主义市场经济体制，要全面贯彻新发展理念，充分发挥市场在资源配置中的决定性作用，更好发挥政府作用，深化市场化改革，扩大高水平开放，

按照系统集成、协同高效的要求纵深推进，在精准实施、精准落实上下功夫，在关键性基础性重大改革上突破创新。

会议强调，要坚持公平统一、权责一致、循序渐进，推动养老保险全国统筹制度更加公平更可持续，全面建成覆盖全民、城乡统筹、权责清晰、保障适度、可持续的多层次社会保障体系。要加强知识产权保护和产权激励，赋予科研人员职务科技成果所有权或长期使用权，健全决策机制，规范操作流程，探索形成赋权形式、成果评价、收益分配等方面制度。要坚持试点先行，坚持优化协同高效，改革事业单位机构设置和职能配置，探索实行政事权清单、机构编制职能规定、章程管理等制度。建设中国特色新型智库是党中央立定党和国家事业全局作出的重要部署，要精益求精、注重科学、讲求质量，切实提高服务决策的能力水平。

会议指出，基础设施是经济社会发展的重要支撑，要以整体优化、协同融合为导向，统筹存量和增量、传统和新型基础设施发展，打造集约高效、经济适用、智能绿色、安全可靠的现代化基础设施体系。要分类放宽服务业准入限制，构建监管体系，深化重点领域改革，健全风险防控机制，完善相关法律法规，提升供给质量和效率。

会议强调，2019年，面对错综复杂的国内外风险挑战，中央全面深化改革委员会坚持稳中求进工作总基调，把准脉搏、扭住重点、有的放矢，立足党和国家工作大局谋划推进改革，继续在重要领域和关键环节推出重大改革方案，完成深化党和国家机构改革，抓好改革部署落地，指导地方有序开展改革试点并及时总结推广，加强和改进改革宣传工作，一体推进落实党的十九届四中全会重要举措，全面深化改革迈出重要步伐。

2020年是全面建成小康社会和“十三五”规划收官之年。要坚持问题导向，继续推出一批重要改革方案。要坚持目标导向，推进改革系统集成、协同高效，牢牢抓住制度建设这条主线，体现针对性、操作性、实效性，把落实党中央要求、满足实践需要、符合基层期盼统一起来，把解决问题、务实管用、简便易行统一起来。要坚持结果导向，聚焦重点、紧盯实效，开展重要领域改革进展情况评估检查，克服形式主义、官僚主义，一个领域一个领域盯住抓落实。

中央全面深化改革委员会委员出席会议，中央和国家机关有关部门负责同志列席会议。

工信部总工程师张峰： 加快推进国家监管基础设施和运行监测大数据中心的建设

来源 / 中国信通院CAICT 编辑 / 协会会员处 李苗苗 日期 / 2020-02



要稳扎稳打，围绕国家监管及安全产业发展需要，加快推进国家监管基础设施和运行监测大数据中心的建设，以更实举措强化硬实力建设；

1月13日，中国信息通信研究院召开2020年工作会议，工业和信息化部党组成员、总工程师张峰出席会议并讲话。

张峰在讲话中指出，2019年，中国信息通信研究院认真贯彻落实党中央、国务院重大决策部署和部党组重点工作安排，紧紧围绕部中心工作，在院领导班子带领下，全体干部职工凝心聚力、锐意进取，各项工作扎实推进、亮点突出，智库作用持续彰显，平台优势更加突出，综合实力不断提升，为信息通信业高质量发展，网络强国和制造强国建设作出了新的贡献。

张峰强调，2020年是全面建成小康社会和“十三五”规划收官之年，中国信息通信研究院要围绕中心、服务大局，扎实推进各项工作，不断开创高质量发展新局面。

一是要提高站位，坚定不移推进和深化政治建设、廉政风险防控体系建设、党支部标准体系建设，以更严格要求推动党

的建设；

二是要善谋善为，结合国家、行业发展需求，着眼于推动高质量发展、“两个强国”建设等事关全局的重大问题，以更高标准打造智库品牌；

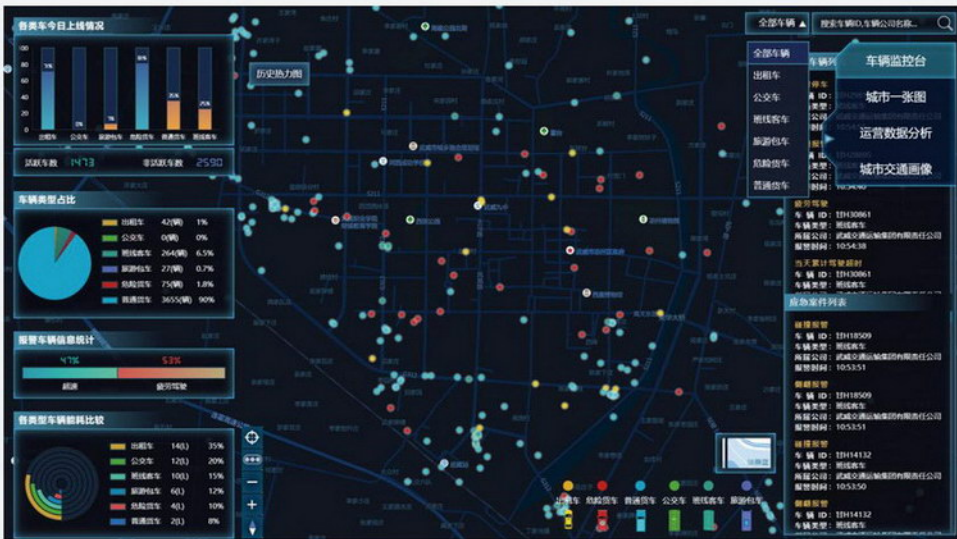
三是要开拓奋进，进一步完善科研布局，优化业务结构，拓展新空间，以更大力度提升平台实力；

四是要稳扎稳打，围绕国家监管及安全产业发展需要，加快推进国家监管基础设施和运行监测大数据中心的建设，以更实举措强化硬实力建设；

五是要固本强基，加强制度和文化建设，注重资源整合和人才培养，以更强保障厚植管理优势。

交通运输部印发 《推进综合交通运输大数据发展行动纲要(2020—2025年)》

来源 / 交通运输部 编辑 / 协会会员处 李苗苗 日期 / 2020-02



据交通运输部官网消息，为推进交通运输治理体系和治理能力现代化，提升综合交通运输服务水平，加快建设交通强国，交通运输部于近日印发《推进综合交通运输大数据发展行动纲要(2020—2025年)》。

根据文件，到2025年，力争实现以下目标：综合交通运输大数据标准体系更加完善，基础设施、运载工具等成规模、成体系的大数据集基本建成。政务大数据有效支撑综合交通运输体系建设，交通运输行业数字化水平显著提升。综合交通运输信息资源深入共享开放。

大数据在综合交通运输各业务领域应用更加广泛。大数据安全得到有力保障。符合新时代信息化发展规律的大数据体制机制取得突破。综合交通大数据中心体系基本构建，为加快建设交通强国，助力数字经济勃兴提供坚强支撑。

以下为全文内容：

交通运输部关于印发《推进综合交通运输大数据发展行动纲要(2020—2025年)》的通知

省、自治区、直辖市、新疆生产建设兵团及计划单列市

交通运输部(局、委)，部属各单位、部内各司局：

为贯彻落实习近平总书记关于网络强国的重要论述和国家大数据战略部署，推进交通运输治理体系和治理能力现代化，提升综合交通运输服务水平，加快建设交通强国，现将《推进综合交通运输大数据发展行动纲要(2020—2025年)》印发给你们，请结合实际认真贯彻落实。

交通运输部
2019年12月9日

推进综合交通运输大数据发展 行动纲要(2020—2025年)

为贯彻落实习近平总书记关于网络强国的重要论述和国家大数据战略部署，加快建设交通强国，依据《国务院关于印发促进大数据发展行动纲要的通知》(国发〔2015〕50号)、《交通强国建设纲要》等文件，特制定本纲要。

一、总体思路

以数据资源赋能交通发展为切入点，按照统筹协调、应用驱动、安全可控、多方参与的原则，聚焦基础支撑、共享开放、创新应用、安全保障、管理改革等重点环节，实施综合交通运输大数据发展“五大行动”，推动大数据与综合交通运输深度融合，有效构建综合交通大数据中心体系，为加快建设交通强国提供有力支撑。

二、主要目标

到2025年，力争实现以下目标：

综合交通运输大数据标准体系更加完善，基础设施、运载工具等成规模、成体系的大数据集基本建成。政务大数据有效支撑综合交通运输体系建设，交通运输行业数字化水平显著提升。综合交通运输信息资源深入共享开放。大数据在综合交通运输各业务领域应用更加广泛。大数据安全得到有力保障。符合新时代信息化发展规律的大数据体制机制取得突破。综合交通大数据中心体系基本构建，为加快建设交通强国，助力数字经济勃兴提供坚强支撑。

三、主要任务

（一）夯实大数据发展基础。

1.完善标准规范。重点研究制定综合交通运输信息资源分类分级、脱敏、溯源、标识等标准规范，完善信息资源目录编制、采集、传输、交换共享等标准规范。不断完善各类交通运输业务数据技术规范。引导相关社会组织和产业技术联盟制定满足市场和有关需要的团体标准。（部科技司牵头负责，国家铁路局有关司局、部内各司局、各省级交通运输主管部门按职责分工负责）

2.强化数据采集。完善修订有关交通基础设施工程建设规范，将采集设备、传输网络等信息基础设施纳入铁路、公路、港口、航道、机场、综合交通枢纽等有关交通基础设施工程建设内容，实现同步规划、同步设计、同步建设、同步运维。完善运载工具运行监测制度规范，实现对车辆、船舶、航空器等运载工具数据自动化采集。（部综合规划司、公路局、水运局、运输服务司、科技司、海事局、国家局有关司局、各省级交通运输主管部门按职责分工负责）

3.加强技术研发应用。推动各类交通基础设施、运载工具数字孪生技术研发，加快交通运输各领域建筑信息模型（BIM）技术创新，形成具有自主知识产权的应用产品。研究制定交通运输行业互联网协议第六版（IPv6）地址规划，推进第五代移动通信技术（5G）、卫星通信信息网络等在交通运输各领域的研发应用。开展综合交通运输体系下大数据关键技术研发应用。（部科技司牵头负责，国家铁路局有关司局、部内各司局、各省级交通运输主管部门按职责分工负责）

（二）深入推进大数据共享开放。

4.完善信息资源目录体系。完善政务信息资源目录，建立覆盖多种运输方式、多级行业管理的综合交通运输信息资源目

录体系，完善信息资源目录采集、更新和发布机制。开展信息资源目录合规性检测，形成完整规范的综合交通运输行业信息资源“总账本”。（部科技司牵头负责，国家局有关司局、部内各司局、各省级交通运输主管部门按职责分工负责）

5.全面构建政务大数据。深入推进国家综合交通运输信息平台建设。升级完善交通运输信息资源共享平台，建立综合交通运输政务信息资源共享交换机制，研究开展政务信息资源质量评估，汇聚形成全覆盖、高质量的综合交通运输政务大数据。建立完善与公安、自然资源、生态环境、水利、文化和旅游、卫生健康、应急、海关、市场监管、气象等部门的政务信息资源共享交换机制，有效支撑综合交通运输政务大数据跨部门应用。（部科技司牵头负责，国家局有关司局、部内各司局、各省级交通运输主管部门按职责分工负责）

6.推动行业数字化转型。以试点示范方式推动建立行业大数据平台。促进行业企业应用大数据、云计算等技术提高企业建设能力、运输效率和经营水平，鼓励各行业企业依法积极利用大数据开展融合应用，不断丰富行业大数据内容，推动行业企业数字化转型。（部科技司牵头负责，各省级交通运输主管部门按职责分工负责）

7.稳步开放公共信息资源。研究完善交通运输相关领域公共信息资源开放机制，建立公共信息资源开放清单及更新制度，建立统一的综合交通运输公共信息资源开放平台，优先开放与民生紧密相关、社会需求迫切、有利于激活市场活力的公共信息资源，推动公共信息资源有序开放。（部科技司牵头负责，部内各司局配合）

8.引导大数据开放创新。组织开展综合交通运输公共信息资源开放创新活动，支持各类主体开展大数据创新创业。推动政企数据融合创新，引导行业公共企事业单位依法开放自有公共信息资源，形成良好数据开放生态。（部科技司牵头负责，各省级交通运输主管部门按职责分工负责）

（三）全面推动大数据创新应用。

9.构建综合性大数据分析技术模型。综合运用跨行业、跨层级交通运输大数据集，研究建立具有较强应用价值的综合性、全局性大数据分析模型，形成一系列易理解、有结论、实用性和权威性较强的大数据分析成果，有效支撑综合交通运输决策管理与服务。（部科技司牵头负责，部内各司局配合，各省级交通运输主管部门按职责分工负责）

10.加强在服务国家战略中的应用。在服务“一带一路”建设、京津冀协同发展、长江经济带发展、粤港澳大湾区建设、长江三角洲区域一体化发展、黄河流域生态保护和高质量发展等重大国家战略中，充分利用大数据开展综合交通运输规划和实施情况评估，持续提升规划决策和评估工作的科学性。（部综合规划司牵头负责，国家局有关司局、部内各司局配合）

11.提升安全生产监测预警能力。推动基于大数据的综合交通运输安全生产全流程监管，实现重点营运车辆、超限车辆、重点船舶及水域、重点建设工程、客运码头、危险货物港

口装卸存储设施以及铁路运输、民航运行等重点领域安全生产风险的监测预警和重大风险的防范化解，利用大数据深化“平安交通”建设。（国家局有关司局、部内各司局、各省级交通运输主管部门按职责分工负责）

12. 推动应急管理综合应用。构建跨部门、跨运输方式应急管理大数据，有效掌握重点领域和区域综合交通运输运行动态，及时获取突发事件信息、应急资源和力量分布，有力支撑综合交通运输应急处置和调度指挥。（部应急办牵头负责，国家局有关司局、部内各司局、各省级交通运输主管部门按职责分工负责）

13. 加强信用监管。推动跨部门、跨行业、跨层级信用信息的归集和共享，利用大数据研究推进综合交通运输信用舆情监测、信用风险预警、信用指数分析等工作，推动信用信息在行政审批、行政检查、招投标等业务中深度应用，加快构建以信用为基础的综合交通运输新型监管机制。（部政策研究室牵头负责，国家局有关司局、部内各司局配合，各省级交通运输主管部门按职责分工负责）

14. 加快推动“互联网+监管”。优化综合执法业务流程，推动实现行政检查、行政处罚及行政强制等执法活动全过程数字化。加快建立交通运输“互联网+监管”系统，实现与国家“互联网+监管”系统对接，为实现综合监管、智慧监管提供支撑。（部法制司、科技司牵头负责，部内各司局配合，各省级交通运输主管部门按职责分工负责）

15. 深化政务服务“一网通办”。依托全国一体化在线政务服务平台，深入推进交通运输政务服务“一网通办”，推动政务服务事项动态更新，完善政务服务门户，推动电子证照信息共享，逐步实现统一身份认证、统一电子印章和统一电子归档。（部科技司、办公厅、法制司牵头负责，部内各司局配合，各省级交通运输主管部门按职责分工负责）

16. 促进出行服务创新应用。鼓励各类市场主体培育“出行即服务（MaaS）”新模式，以数据衔接出行需求与服务资源。促进交通旅游服务大数据创新应用。利用大数据分析评价道路客运、公共汽电车、出租汽车、汽车分时租赁等领域新老业态发展特征，推动新老业态动能转化和融合发展。（部运输服务司、综合规划司、科技司、各省级交通运输主管部门按职责分工负责）

17. 推动货运物流数字化发展。逐步完善国家交通运输物流公共信息平台，推动全国多式联运公共信息系统建设，促进多种运输方式间数据交换共享。研究制定货运物流单据电子化相关技术标准。完善全国快件数据监测体系，为全程可跟踪、可追溯的“一站式”快递服务提供数据支撑。鼓励网络平台道路货运、车货匹配、智能航运等“互联网+货物流”新模式发展。建立完善道路货运行业运行监测分析体系，充分利用大数据预测发展趋势，引导货运物流行业健康发展。（部综合规划司、运输服务司、水运局、国家局有关司局、各省级交通运输主管部门按职责分工负责）

（四）加强大数据安全保障。

18. 完善数据安全保障措施。推进交通运输领域数据分类分级管理，加强重要数据和个人信息安全保护，制定数据分级安全管理、数据脱敏等制度规范。推进重要信息系统密码技术应用和重要软硬件设备自主可控。（部办公厅、科技司、各省级交通运输主管部门按职责分工负责）

19. 保障国家关键数据安全。全面识别梳理交通运输领域国家关键数据资源，将重要数据保护纳入交通运输关键信息基础设施安全规划，推进国家关键数据资源全面实现异地容灾备份，推进去标识化、云安全防护、大数据平台安全等数据安全技术普及应用。（部科技司牵头负责，部内各司局、各省级交通运输主管部门按职责分工负责）

（五）完善大数据管理体系。

20. 推动管理体制变革。加强与机构编制部门沟通，争取基于现有内设机构，研究设立或明确综合交通运输大数据发展管理部门，负责统筹推进综合交通运输大数据发展。（部科技司、人事教育司牵头负责，各省级交通运输主管部门按职责分工负责）

21. 完善技术管理体系。研究明确统一的公益性综合交通运输大数据管理与应用机构，负责综合交通运输大数据技术统筹。（部科技司、人事教育司牵头负责，各省级交通运输主管部门按职责分工负责）

四、保障措施

（一）加强组织领导。

深入学习领会和贯彻落实习近平总书记关于网络强国的重要论述，将综合交通运输大数据发展作为“一把手”工程予以推进。增强各级领导干部获取数据、分析数据、运用数据的基本功，让广大领导干部“懂得大数据、用好大数据”。

（二）加强队伍建设。

引导建立综合交通运输大数据智库，创新行业大数据人才培养机制和激励机制，引进招聘高层次人才，培育层次多样、良性流动、激励有效的专业化大数据人才队伍。

（三）加强经费保障。

发挥财政资金引导作用，加大对政务大数据建设的支持力度，鼓励设立专项资金保障大数据发展。研究利用社会资本开展大数据创新应用，引导以市场化基金、政府和社会资本合作等模式推动行业大数据市场发展。

（四）加强效果评估。

推动将综合交通运输大数据工作绩效纳入部门和单位工作评估体系，对综合交通运输大数据工作情况开展评估，以评促建，推动综合交通运输大数据不断发展。

顺丰2019生鲜大数据报告：揭秘中国式生鲜物流的“超能力”

来源 / CPDA数据说 编辑 / 协会会员处 李苗苗 日期 / 2020-02

在过去的2019年,虽然很多人都在说“太南了”,甚至《新周刊》还把“南”字评选为年度汉字,但从现实数据上看,人们对于生活品质的追求并没有降低,在全国乃至全球原产地采购好东西的消费需求反而更容易得到满足。

SF EXPRESS
顺丰速运



2020年元旦刚过,顺丰发布了《2019年度中国居民生鲜消费升级报告》,基于顺丰生鲜寄递的上百亿条大数据,从一个个鲜活的侧面,对国人的生鲜消费做了年度盘点。生鲜消费历来被视为居民生活品质的“晴雨表”,“精致生活要鲜食”,生鲜消费需求的增加、尤其是原产地生鲜消费的增长,反应出人们对生活品质的重视程度越来越深。

“生鲜自由”背后的消费升级

前两年有一个词特别火,叫“车厘子自由”,但在2019年,越来越多的生鲜开始有了自己的“自由消费指数”。原产地、跨省乃至跨境的生鲜消费,背后折射出人们对生鲜品质的需求在进一步提升,促成处处不将就的“生活美学”,形成一幅幅精致生活图鉴。

一是“水果自由”指数提高。

报告显示,2019年常见水果自由指数大大提高,“樱桃自由指数上升8个百分点,杨梅自由指数上升22个百分点,枇杷自由指数上升50个百分点,桃自由指数上升70个百分点”。这主要得益于两个原因,一是生鲜运输能力提升,使得生鲜损耗率降低与自由流动能力增强,进而促进了市场的供需平衡;二是国内生产要素驱动,国产生鲜市场认可度增强,例如,根据顺丰生鲜大数据,“2019年取得樱桃发件量第一的地区是烟台,国外的车厘子自由正在被国内樱桃自由取代。”



二是节日消费高峰呈现。

报告数据显示,“春节期间牛羊肉和海鲜水产寄递总量占全年总量的1/5。”生鲜消费已成为中国人节日消费的重要组成部分。



三是热门产品的多样化趋向。

大数据显示,“2019年海鲜水产,寄递总件量占比较18年提升28%,大闸蟹增幅减缓”“大闸蟹C位被夺”,而赣南脐橙等热门水果涌现,这说明生鲜消费热门品类趋向多样化,生鲜市场更具活力。



四是养生等新理念凸显。

“2019年，从松茸的寄递到达量来看，北京购买力TOP1；其中8月份，松茸的寄递总量占全年的1/2”这显示，消费者的生鲜消费日趋品质化，不仅要吃的多样，更要吃的“健康”。



五是生鲜消费呈地区性集中。

一线城市消费能力突出，对“精致生活”要求更高。报告显示，“2019年顺丰收件城市top5分别是上海市、北京市、深圳市、广州市、杭州市。”

上海被冠以“更小资城市”称号，夺得精致生活排行榜第一，并连续三年成为大闸蟹收货top1城市；北京地区生鲜消费品类呈均衡态势，成为“更博爱的城市”；而深圳成为“更小清新的城市”，草莓与橙子消费量庞大，水果自由指数居高位；鲜花与牛肉丸成为广东人的最爱，成为“更感性的城市”；杭州作为新一线城市，不声不响却能进入前五榜单，成为“更低调的城市”。



中国式生鲜物流的“超能力”

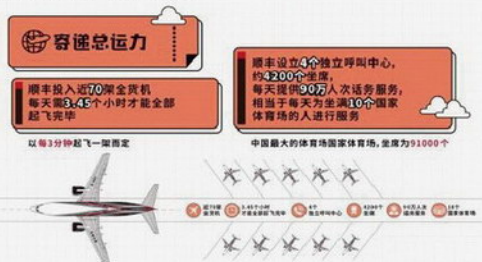
有一些国外媒体羡慕中国是“基建狂魔”，特斯拉工厂从动工到第一辆车下线只用了不到11个月。但让很多普通外国人羡慕的，还有“物流超能力”。

今天刚从湖里捞上来的大闸蟹，明天晚上就可能到了千里之外消费者的餐桌。报告显示，2019年顺丰寄递生鲜总重量超69.8万吨，总里程超1437.7亿公里，最远的一单是从大连到纽约。顺丰是如何做到的？

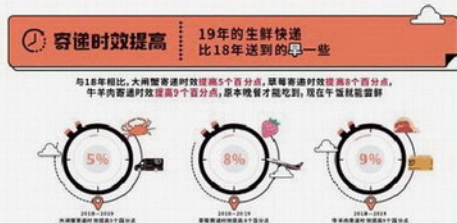


顺丰强大的运力是支撑基础。

顺丰投入近70架全货机，设立4个独立呼叫中心，4200个坐席每天提供90万人次话务服务，让日益增长的生鲜寄递需求妥善得到高质量满足。



在运送时效层面，顺丰大数据显示，与18年相比，各典型品类寄递时效均有大幅度提高。“大闸蟹时效提高5个百分点，草莓寄递时效提高8个百分点，牛羊肉寄递时效提高9个百分点。”这得益于运输技术的显著提升与发展，生鲜时效的提高大大降低了生鲜损耗率，让消费者能享受更新鲜的高品质食材。



技术黑科技是发展驱动力。

顺丰为不同的热门生鲜品类，量身定制了运输全流程黑科技，利用人工智能与自动化技术，提升食材采摘、包装等运输前期处理环节效率，赋能运输产业链，让食材更新鲜到家。例如，顺丰成立了全国首个松茸预处理中心，“无人

机”加入顺丰硬核运输阵营，200架无人机带领松茸出山，48小时内将甘孜松茸发往全国300多个城市，打造“会飞的松茸”。

又比如，顺丰还打造了全国首台捆蟹机，实现了物流捆扎从人工到自动的变革。设备拥有近10项国家专利知识产权，单只蟹捆扎效率提升100%、成本降低20%，实现了大闸蟹的捆扎美学。



从下至上的“生鲜革命”

品质生鲜寄递的发展，让千里之外的消费者以性价比更高的价格，在短时间内品尝到优良品质的生鲜。“一骑红尘妃子笑”已成为历史，物流与科技的发展为“寻常百姓家”提供了越来越多样化、品质化的消费选择。在这一过程中，顺丰品质生鲜寄递的发展，则是中国生鲜物流发展水平的典型代表。

生鲜寄递数据增长的背后，一方面是国民消费的进一步升级，另一方面是物流配送的运力大幅度提升。未来，消费者将有更自由、更多样化、更品质化的消费选择。而大众所追求的精致生活，也将由极速“鲜食”开始。

大数据助力物流精准配送

物流路线

物流企业运用大数据来分析商品的特性和规格、客户的不同需求等问题，从而用最快速度对这些影响配送计划的因素做出反映(比如选择哪种运输方案、哪种运输线路等)，制定最合理的配送线路。

而且企业还可以通过配送过程中实时产生的数据，快速地分析出配送路线的交通状况，精确分析配送整个过程的信息，使物流的配送管理智能化。

基于大数据实现车货高效匹配，不仅能减少空驶带来的损耗，还能减少污染，是一举多得的好事情！

物流选址

物流中心选址问题要求物流企业在充分考虑到自身的经营特点、商品特点和交通状况等因素的基础上，使配送成本和固定成本等之和达到最小。

通过运用大数据，物流运输效率将得到大幅提高，大数据为物流企业间搭建起沟通的桥梁，物流车辆行车路径也将被

最短化、最优化定制。

仓库储位

合理的安排商品储存位置对于仓库利用率和搬运分拣的效率有着极为重要的意义。对于商品数量多、出货频率快的物流中心，储位优化就意味着工作效率和效益。

哪些货物放在一起可以提高分拣率，哪些货物储存的时间较短，都可以通过大数据的关联模式法分析出商品数据间的相互关系来合理的安排仓库位置。

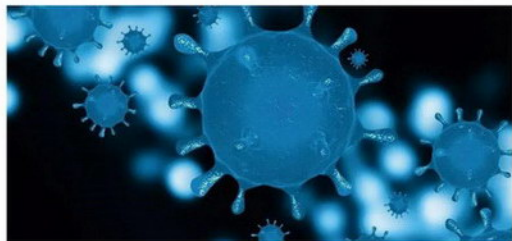


大数据时代的来临，不是技术的变革，首当其冲是思维的变革，随之而来的是商业模式的改变，所以我们要意识到，对物流系统的大数据管理、供应链管理是物流发展的必然趋势，是所有实业经济发展的必然趋势，是改变经济发展方式的杀手锏，通过信息技术、数据监测，实施商流、物流、信息流、资金流的一体化运作，使市场、行业、企业、个人联结在一起，才能实现智能化管理与智能化生活。

大数据还原病毒扩散路径：哪些城市压力更大？

来源 / CPDA数据说 编辑 / 协会会员处 李苗苗 日期 / 2020-02

导读：这场疫情究竟是怎么开始的？目前的疫情又发展到了什么阶段？在接下来的时间里，哪些城市将面临更大的压力？本文通过大数据进行详细分析。（资料来源：21世纪经济报道、21数据新闻实验室、澎湃、DT财经）



1月12日，一位常住湖北黄冈的居民坐上了前往广东珠海的高铁。到达的第二天，她开始咳嗽，以为是普通感冒，自己就先买了点药吃着，可是症状并没有得到缓解。

1月19日，她因咳嗽加重来到中山大学附属第五医院诊治。

三天后，晚上10点，她被确诊为新型冠状病毒感染肺炎患者。这天是1月21日。

她不知道的是，再过15天，黄冈会累计确诊1645例新型冠状病毒肺炎病例，位居全国第二。

1月18日至2月4日各地确诊病例数量



有类似情况的还有孝感、荆门和襄阳等湖北地级市。

甚至1000公里外的浙江温州、广东广州、深圳等城市，也因与武汉的密切人员往来，随着春运返乡的开始，慢慢成为了重点疫区。

而这一切究竟是怎么开始的？目前的疫情又发展到了什么阶段？在接下来的时间里，哪些城市将面临更大的压力？

本文除了结合案例数据详细还原病毒扩散的路径外，还将案例数据汇总在一起，拼贴出疫情的感染、死亡和治愈特征画像，并结合数据和专家观点观察疫情走向。

还原新冠病毒向全国扩散的路径

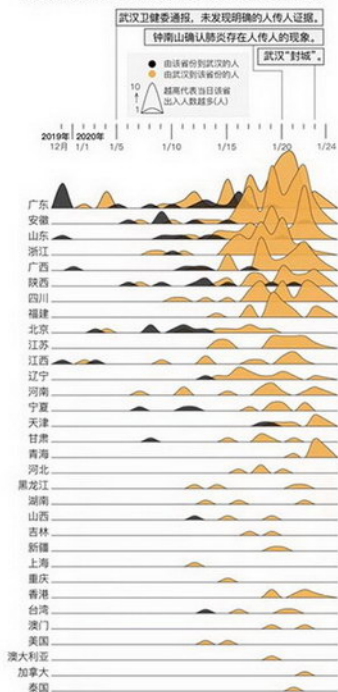
澎湃新闻从截至1月27日的各地卫健委公布确诊病例中，搜集并分析了763例确诊患者的详情，试着还原病毒扩散的路径。

值得指出的是，各地之间的个案数量差异主要是因为政府的信息公开程度不同，因此地区之间的数量不具有可比性。

368名后期确诊患者在武汉“封城”前一周离汉

1月23日早上10点，武汉关闭机场、火车站离汉通道，而这距离今年春运正式开始的1月10日已经过去了13天。

封城前多少后期确诊患者出入武汉



过去的这13天，正是每年人们固定返乡的日子。没有多少人被警示到，大家依旧像往年一样，拿起行李，探亲、出游、回家聚餐。

如上图所示，在506例有明确行程信息披露的病例里，368例在武汉“封城”前一周离汉。出入武汉的后期确诊患者人数高峰期在1月19日和20日，这两天一共有130名后来确诊的患者出城，占到了总数的27%。

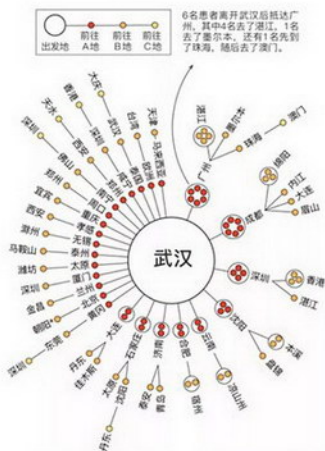
1月20日晚，钟南山在央视《新闻1+1》栏目中确认了病毒具有人传人的特征。从那天晚上，再到武汉23号封城，又过去了3天时间。21日，56名后来确诊患者出城；22日，62名；23日，36名。

47名患者离开武汉后曾去过不止一个城市

来自河北邯郸的一名患者，1月15日和父母一起到武汉旅游，18日移动到厦门，19日发病，22日抵达深圳就立即入院就诊。

和他一样，离开武汉可能只是很多人春节出行的第一步。在公开的506例确诊患者行程中，47名患者从武汉离开后，又经过了多个目的地，占到了总数的9%。

离开武汉后前往多个城市 后期确诊患者



此外，还有28例确诊患者案例显示，他们和武汉并无直接交际，仅有湖北其他省份的旅游史。更有甚者只因在旅途中途经武汉而患病。

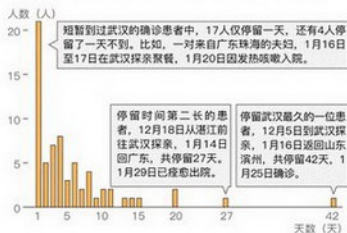
安徽合肥的叶某，1月9日前往湖南郴州出差，中途在武汉换乘，1月10日返回合肥。1月10日发病，1月23日确诊。

还有一名长居深圳的患者，1月19日从深圳出发前往天津，高铁在武汉停留了数分钟，1月22日到达吉林市，1月27日在当地确诊。

不过，停留时间的长短和后期是否患病确实没有直接关

系。如下图所示，67名短暂停留过武汉的患者中，17人仅停留一天，4人停留了一天不到。

短暂到过武汉的患者的停留时间



值得指出的是，上述数据也只反映了患者行程的一部分，只有少数患者的城市内部移动轨迹得以公示。

从出现症状到就诊入院最长间隔近一月

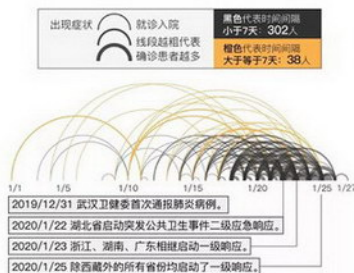
防控在于警示。“这次疫情其实各方面对我们信息的披露是不满意的”，武汉市市长周先旺27日接受央视访问时表示。

确诊患者从出现症状到就诊入院的时间间隔变化，侧面验证了这一现象。由于新冠肺炎症状和普通感冒很像，大多患者甚至医生一开始并不能察觉到患病的严重性。

陕西安康市的一名患者，长期在湖北十堰居住，他早在12月底就出现了症状，但直到1月21日才到白河县就诊。

山东成武县的患者田某，长期在湖北荆州工作。1月9日返回老家，1月20日出现症状，当日在村卫生室输液治疗，未见好转，22日前往成武县田集镇卫生院就诊，依旧未见好转，23日到成武县医院就诊，被隔离治疗。1月26日确诊。

确诊患者从出现症状到 就诊入院的时间间隔



如上图所示，以1月18日作为分界点，前后的时间间隔差距很大。特别是从1月22日起，当日出现症状即就诊的患者数量剧增。这天，湖北省启动了突发公共卫生事件二级应急响应。第二天，浙江、湖南、广东相继启动了一级响应。

家属和同事是人传人特征的主要受害者

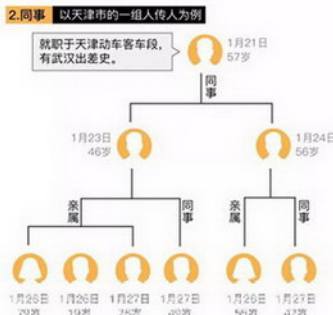
以上所述的时空特征，再加上人传人的特征，带来的后果就是大批的后期确诊患者早已到达全国各地，将病毒带给密切接触者，甚至引发聚集性传染。据新华社报道，截至2月4日中午，山东省发现聚集性疫情60起，其中49起为家庭聚集性疫情。

总体而言，家属和同事是最主要受害者，因为他们属于每天在熟人空间能遇见的人。

在151例人传人确诊患者中，共有64例属于这种情况。下图以山东和天津的两组传播为例。

熟人空间产生的互动

熟人空间产生的互动包括家属（邻里）和同事两类，是疫情人传人最常见的路径。在151例人传人病例中，共有64例属于这种传染路径。



公共空间也是病毒传播的重要途径。在151例人传人确诊患者中，共有20例属于这种情况。

他们之中，有的人是因为参加了同行者中有武汉籍患者的云南旅行团，有的人是因为和武汉籍患者同事聚餐，还有的人则是因为在公共交通工具上遇到了武汉旅客。

公共空间产生的互动

公共空间产生的互动包括外地出行、聚餐、公共交通三类。占比较少，在151例人传人病例中，共有20例属于这种传染路径。



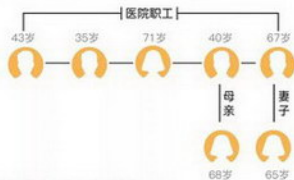
2. 聚会和公共交通



最后一种常见的人传人场景则是在医院。这种传染案例的详情公布得不多，截至1月27日，仅有以下来自江西新余市的5例。5名医院职工被感染，其中两名又分别将病毒传染给了自己的母亲和妻子。

医疗活动产生的互动

医院职工传染的详情记录较少，仅有以下来自江西新余市的5例，确诊日期在1月27日到28日之间。



死亡和治愈病例有何特征？

能给我们哪些启示？

真实病例表现表现出怎样的特征？

在此分享DT财经的一篇大数据报道，具体情况如下：

新冠病毒肺炎感染者的特征

对于感染者的特征研究，下面的分析直接引用了中国疾控中心主任高福、副主任冯子健于1月29日发表在《新英格兰医学杂志》上的论文《新型冠状病毒感染肺炎在中国武汉的初期传播动力学》。

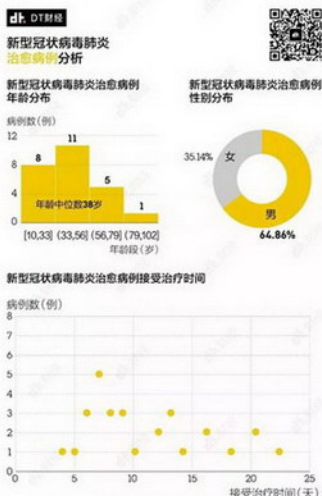
目前关于这篇论文的争议很大，暂且不表。但它确实是目前对感染者最为详尽的研究，样本包含了1月22日前武汉的425名新型冠状病毒肺炎感染者。

其中引起争议的部分是关于病例接触史的分析。

数据显示，2020年1月1日之前的感染病例中，虽然大部分（55%）去过华南海鲜市场，但亦有26%的病例既没有去过海鲜市场，也没有接触过有明显症状的人。

从这些较早感染者的整体情况来看，“中老年”和“男

龄、性别、接受治疗时间和地区分布来分析这些人的特质。

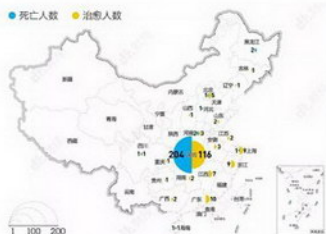


在年龄上，治愈出院的病例整体要比感染人群年轻——甚至年轻得多。在有记录年龄的31例治愈病例当中，年龄的中位数为38岁，过半病例没有超过40岁，九成没有超过60岁。年轻人在这场没有硝烟的战争当中，的确存在一定的优势。

性别上的分析结果也较符合市场预期。由于前期感染病例大多为男性，所以相对地，治愈出院的病例当中也有将近65%的人为男性。

从治疗时间来看，在治愈病例当中，接受治疗天数的中位数为9天，众数为7天，都比死亡病例的接受治疗时长略长几天。综合比较死亡病例和治愈病例，一定程度上可以认为，要打赢新型冠状病毒，很重要的得撑过接受治疗的第一周。

最后把观察点落在了地区分布上。除湖北以外，广东、浙江和上海是治愈人数最多的三大省份。其中广东和浙江是除湖北外疫情最为严重的地区，但到目前为止还没有死亡病例出现，治愈病例则在不断增加。



如果把观察样本扩大至治愈人数排在前10的省份，依次还能看到江西、北京、河南、安徽、湖南、山东、江苏等省份名。从实际案例来看，经济发达地区的应对水平仍然值得信任。

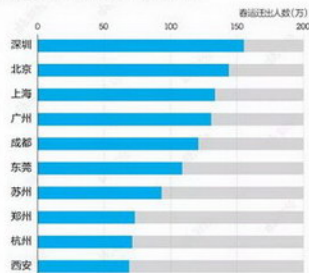
哪些城市压力更大？疫情什么时候可以控制下来？

截止2月6日24时，全国新增疑似病例4833例（5日为5328例），其中湖北新增2622例。

值得注意的是，全国非湖北地区新增疑似病例出现轻微上扬之势，这或许与陆续出现的返工潮有关。因此，疫情防控仍不可放松。



百度迁徙数据显示，在2020年除夕前这一周，迁出人口数量最多的城市主要集中在珠三角、长三角和成渝；其中深圳、北京、上海、广州、成都都是迁出人口数最多的五个城市，而东莞、苏州等制造业城市和郑州、杭州、西安等重要节点城市，也在春运前迁出了最多的人口。



从2月2日的少量返工人口流动来看，上述城市确实是最热门的目的地。



（图片说明：2020年2月2日全国热门迁入地/目的地，图片来源：百度迁徙）

在湖北正在尽全力实现“应收尽收、应治尽治”时，湖北外的城市们也在跟时间赛跑——尽快将疫情传播控制下来，以最小的风险去迎接返工潮的压力。

目前大家最关心，也是最关键的问题是，疫情什么时候可以控制下来？拐点如何才能到来？

7日下午，钟南山院士回应“#全国新增确诊病例连续两天明显下降#”时表示，这并不能证明拐点到来，他个人估计

这个拐点到来还得有几天。现在看起来早发现、早隔离的工作开始起效了。

中国医学科学院院长王辰2月5日对话白岩松时表示，拐点预判的难点在于：

- ①疫情的底数不明，判断根据不足；
- ②很多人还没隔离，社区、家庭传播相当严重；
- ③此次爆发的是新将人体作为宿主的病毒，适应宿主的过程会促使病毒变异，但变异过程未可预期；
- ④大批人员即将返程流动是个考验；
- ⑤或许还有天气变暖的因素。

世界卫生组织发言人贾萨瑞维奇6日接受《环球时报》记者采访时也表示，现阶段尚不能预测中国国内疫情的“拐点”何时到来，也无法预测新型冠状病毒肺炎疫情是否会在中国以外大规模爆发或演变成全球性流行病，因为新型冠状病毒仍有大量关键信息未能为世界科学界所掌握。

但他同时表示，只要各国采取强有力的措施，及早发现、隔离、治疗并追踪接触者，采取与风险相称的社会阻隔手段，阻断病毒蔓延依然是可能的。

美国加州大学洛杉矶分校（UCLA）公共卫生学院副院长、流行病学资深终身教授张作风认为，湖北确诊病例占全国

超70%，要重点做好湖北、以及与湖北邻接的河南、湖南等病人的诊断，把可疑病人尽快排查，诊断出来后，疫情拐点才会浮现。

当然，从乐观的方面来想，拐点预判不一定准确的点还在于，如果特效药等有效治疗手段突然出现，新冠病毒肺炎变得不再那么可怕，致命性大大降低，变得比流感还要寻常。那么，即使没有消灭它，我们也不用再如此严阵以待。

无论如何，要迎来拐点，现阶段的关键还是得我们共同发力。在这个关键时刻，希望不要过度悲观、盲目乐观。尽量少出门，不让自己感染的同时，把资源留给真正有需要的人。

最后，借用上海医疗救治组组长张文宏的一句话：“从现在开始，每一位都是战士。你闷两个礼拜就把病毒给闷死了。你在家不是在隔离，你是在战斗啊！”

企业数据分析的10种思维和技巧

来源 / CPDA数据说 编辑 / 协会会员处 李苗苗 日期 / 2020-02

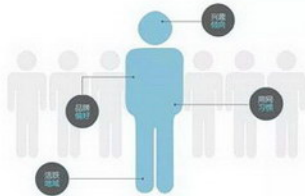


随着时代的发展，人们每天在互联网上产生大量的数据，对于企业来讲这些数据都是十分宝贵的资源。企业可通过数据挖掘进行战略调整以及营销部署，尤其是对于互联网公司而言，用户行为产生的数据就是企业最宝贵的资源。

一、企业中各项数据的挖掘

数据挖掘（Data mining），又译为资料探勘、数据挖

掘、数据采矿。它是数据库知识发现中的一个步骤。数据挖掘一般是指从大量的数据中自动搜索隐藏于其中的有着特殊关系性的信息的过程。



不过目前为止，在很多企业中都存在着很大的数据分析问题，如何进行数据分析？数据挖掘的结果要如何展示？企业中各个部门要如何才能最大化的利用数据分析结果？这些一直困扰着数据团队。

方法，多角度看问题，不要钻牛角尖，多学习别人的好方法，学会集思广益，发散性思维。比如说：你的学习能力和方法有效吗？

7、联合思维

很多销售数据，需要我们能站在当事人的角度去思考和分析，这样你才会理解人、事、物。关键在于多了解当事人的情况，学会换位思考。比如：你了解你周边的情况吗？你了解你周围的人吗？

8、离开思维

通过数据分析，你发现你处在一个不太有利的地位，那么，此时，你就要有离开思维去替你想办法，离开困境。关键是学会自我调节，自我放松。实际情况如：遇到难解的结，你怎么办？

9、接近思维

怎样达成目标，实现销售增长，这时候你需要接近思维来帮助你。关键是多接触你要解决的问题，花时间分析，你要的是方案，不是问题。实际情况如：你在做选择题还是问答题？责任点在哪？

10、理解层次

问题发现是第一步，要怎样分析问题，找到真正的原因，那么熟练的运用理解层次。关键是：你需要熟悉客观环

境，员工的能力、行为的规律、他需要什么？实际情况如：你能够分析到哪一步？

四、数据分析师需要掌握的 10 个技巧

- 会用一款或以上的数据分析工具
- 经常浏览数据统计的网站
- 在数据分析前先进行调研
- 在分析数据的时候用户体验的角度出发的么并不是以

公司利益为主

- 了解数据采集的方式以及数据内容和质量内容
- 熟悉各种样式低量和定位的不同，尤其是竞争的产品
- 做一个饥渴的探索者
- 在企业内部有效的沟通者
- 街头智慧
- 防御中带有进攻

对于数据分析师来说，只是关注数据的相关性，从相关性解释数据的因果性。如何将企业收集的杂乱数据进行分析处理，最终为其他部门提供一份清晰明朗的数据报告就显得格外重要。

从几个传染病模型定量地去理解当前的措施

作者 / CPDA数据分析师 广州 刘程浩 编辑 / 协会会员处 李苗苗 日期 / 2020-02

最近全国上下都在和NCP做斗争，这次也通过参与了“抗疫”的全民战争，使我懂得了很多过去没怎么重视，但是却非常重要的道理。同时也对目前国家的一些措施有更深刻的理解。

道理，一般分两种层面的理解，一种是定性的理解，简单说即是你知道做什么事是对的，从逻辑上理解为什么是对的；另一种是定量的理解，就是说你明白了做什么事是对的，也明白这件事做到什么程度，相应“做对”的效果有多大。

比方说这次全民对NCP这类抗呼吸道“传染病”，从定性的角度来看，我认为是包含着如下这些道理：

对抗“传”：就是切断和阻塞疾病的传播途径，例如“封城”、“社区交通管制”、“14天的隔离”、“戴口罩”，“勤洗手”……，这些是为了切断病毒携带者（包括潜伏期内）把疾病通过人体移动、飞沫、体液等载体“传”出去。

对抗“染”：譬如说这段时间宅在家也别只是吃喝追剧，要多做些室内健身运动；多喝一些有预防作用的“凉茶”；科研机构也紧锣密鼓地开展研发疫苗……这些措施都是为了增加人的抵抗力，降低万一被病毒接触造成的被感染可能。

对抗“病”：就是研发出治疗NCP的药物，无论是中药还是西药。总之就是尽快将病人治愈，或者降低被感染者的死亡率。

上面这些道理，有一些是从小到大爸妈耳边叮铃、老师嘴上唠叨，耳朵都听出茧了的。

我当然知道做这些措施是可以抵抗传染病，而且有些我也正在做。

但是有时候就是搞不明白，我认真去做和我应付去做这些事儿，区别有多大？

难道就只是影响自己健康而已，或者就只是影响一个社区？



我戴上普通的口罩，和没戴口罩相比，当然我知道被感染的风险降低了，但是这对我周围的人的影响到底有多大？

专家们说的 $R_0 > 1$ 和 < 1 背后有什么更深一层内容？

.....

等等很多疑问，在我这段时间宅在房间里认真看了几个传染病的模型后，很多都已经茅塞顿开，甚至感到惊叹！

原来从定量的层面去理解这些“道理”，能比定性的理解有更深刻的印象！

下面就和大家分享下我的心得：

一、首先看最简单的传染病模型

把这个模型看懂了，也就能从另一个角度了解为什么说一旦传染病爆发，感染人数是呈指数级增长了。

简单的指数级增长比较容易理解，比方说今天1个人得病，假设每次只传染1个人，那么初始第0天1个病人，第1天就2个人病人，第2天就4个病人、第3天就8个人，第n天就 2^n个人得病。

这个刚好是一个指数函数，底数为常数2，变化的是它的幂。

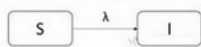
不过在传染病研究领域，指数级的增长是从另外一个视角来看

我们假设在某个时点某个区域内，不考虑出生和其他原因的死亡；也不考虑外部环境对传染病的干预；不考虑人口流动；潜伏期也可以传染；这种病无法治愈；时间的计量刻度是按照天。

初始化一些模型的变量

t时刻的易感染者（Susceptible）人数，用符号S(t)表示；

t时刻的被感染者（Infective）人数，用符号I(t)表示；
每个被感染者每天有有效接触且足以使人致病的人数为 λ ，这些 λ 均为健康的人，那么每天新感染的人数就是 $\lambda I(t)$



建立模型：

$$I(t + \Delta t) - I(t) = \lambda I(t) \Delta t \quad (1.1)$$

(1.1)式的模型左边和右边同时除以 Δt ，当 $\Delta t \rightarrow$ 无限小或0时，就会得到导数的定义式

$$\lim_{\Delta t \rightarrow 0} \frac{I(t + \Delta t) - I(t)}{\Delta t} = \frac{dI(t)}{dt} = \lambda I(t)$$

然后以此建立微分方程

$$\begin{cases} \frac{dI(t)}{dt} = \lambda I(t) \\ I(0) = I_0 \end{cases} \quad (1.2)$$

解这个微分方程。第一个式子容易看出是一阶线性微分方程，根据解这类微分方程的方法容易得到该方程的解：

$$I(t) = I_0 e^{\lambda t}$$

$I_0 e^{\lambda t}$ 也就是被感染者的数量随着时间推移的变化规律。

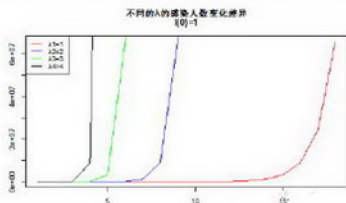
看到这个常数e了吗，以及e右上角的幂是不断增大的，也就是为什么被传染病领域说的感染的人数的增长是指数级了。

e的幂增长速度 and 直线比那是很快的，从这个最简单的传染病模型我们可以看出，只要随着时间的推移 $t \rightarrow \infty$ ，I(t)

$\rightarrow \infty$ ，也即是最终世界上全部健康人都会被传染。

被感染的人的数量是不是只根据一种指数级增长而变化呢，其实不是的。

解决传染病感染人数增长快慢还有一个的关键就是 λ ，也就是控制已感染者接触健康人的数量，让这个值越小越好。



上图不同的颜色曲线代表不同的 λ 值造成的感染人数变化。

可以看出， λ 值越大，传染人数增长的就越快！

假设最初只有1个人被感染，即 $I(0)=1$ 时，不做任何干预和控制的前提下

当 $\lambda=4$ ，也就是黑色的曲线到了第3天第4天后，几乎就直直的往上窜了，变化速度简直就是火箭；

而当 $\lambda=1$ ，到了第17天左右，感染人数才开始迅速上升

这2种增长速度的差异是相当惊人的，虽然人数只差了3个人。

当然极端情况下 $\lambda=0$ 是最好的，因为只要从最初时刻发现病人迅速隔离，那么就能够直接杜绝传染病的蔓延。

从传染病模型推导出来的不同指数级增长之间的差异如此的巨大，当时我是直接一个震撼：

这就是为什么要大家宅在家里别出去，背后数字上跃迁的可怕！

这也解释了武汉人民做这么大牺牲，通过“封城”来阻断疫情的发展的意义！

这也从定量解读的层面阐述了小区、村落要封闭，虽然在短期内对生活带来不便，但实际上对疫情控制有重要的意义！

但 $\lambda=0$ 是这个理想状态，且不适合于普通的社会环境。

因为那是在大家都对这个传染病有了解和认知的情况下。而且某些传染病的症状和普通可自愈（治愈）的疾病症状很相似，也就是人们一开始并没在意，因此就会有后续的传播。

不过对于已经收治的病人，本身传染病病房隔离措施比较好，治疗室里面假设仪器、医护人员、消杀措施配置都到位的情况下，就能实现上面的目的。可医疗资源是有限的，因此当一个区域的医疗资源比较紧张或者用罄后，就必须补充或支援。

这个时候我们就看到了全国各地的医疗队伍驰援武汉的场景，甚至看到了让人心潮澎湃和泪目的各个医疗专家队伍的请战书。



感兴趣的朋友可以接着往下看上面这个简单传染病模型推导的过程

$$\frac{dI(t)}{dt} = \lambda I(t)$$

$$\frac{1}{I(t)} dI(t) = \lambda dt$$

$$\int_0^t \frac{1}{I(t)} dI(t) = \int_0^t \lambda dt$$

$$\ln(I(t)) - \ln(I(0)) = \lambda t$$

$$e^{\ln(I(t)) - \ln(I(0))} = \frac{I(t)}{I(0)} = e^{\lambda t}$$

$$I(t) = I(0) e^{\lambda t} \text{ or}$$

$$I(t) = I_0 e^{\lambda t}$$

接下来再逐步把模型的应用具体化一些，就是观察某个区域的传染速度。

二、SI模型

这个模型看懂了，你就明白为啥戴口罩，哪怕带自制的口罩不只是对自己负责，也是对这个社会负责。

我们假设某个观察区域的总人数 N 不变，不考虑区域内出生和其他原因的死亡，也不考虑人口迁移。

增加需要考虑的是，病人每日接触到的人不全是健康的人，包括已经被感染的人群。这样的话后者是不会再考虑进这个系统；

另外考虑到感染需要一定的条件，因此不是每次病人的接触都会造成感染，存在一个感染的概率

模型的变量稍微调整如下：

t 时刻的总人口数保持不变 $=N$ ，用符号 $N(t)=N$ 表示；

t 时刻的易感染者（Susceptible）人数，用符号 $S(t)$ 表示；

t 时刻的已感染者（Infective）人数，用符号 $I(t)$ 表示；

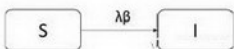
可知， t 时刻的 $N(t)=S(t)+I(t)$

β 为感染者接触到易感染者后，易感染者被感染的概率；

每个已感染者每天有效接触的人数为 λ ，但由于接触到的人可能是已经被感染过的了，因此要剔除掉这部分人。

假设t时刻平均易感染人数占比为 $S(t)/N(t)$ ，那么每天每个已感染者能够让 $\lambda[S(t)/N(t)]\beta$ 个健康者感染，那么每天新增的被感染人数为 $\lambda[S(t)/N(t)]\beta I(t)$ 。

建立模型：



建立模型：

$$I(t+\Delta t) - I(t) = \lambda \frac{S(t)}{N(t)} \beta I(t) \Delta t \quad (1.3)$$

然后以此建立微分方程

$$\begin{cases} \frac{dI(t)}{dt} = \lambda \frac{S(t)}{N(t)} \beta I(t) \\ N(t) = S(t) + I(t) \end{cases} \quad (1.4)$$

解这个微分方程，和上面的解法相比略有复杂，需要用到“常数变易法”。

首先，将第二个方程变形， $S(t) = N(t) - I(t)$ 代入到第一个方程得到

$$\frac{dI(t)}{dt} - \lambda \beta I(t) + \frac{\lambda \beta}{N(t)} I(t)^2 = 0 \quad (1.5)$$

(1.5)的解就是SI模型中受感染人数随时间变化的值，如下

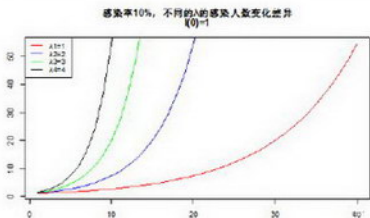
$$I(t) = \frac{NI_0}{I_0 + (N - I_0)e^{-\lambda \beta t}} \quad (1.6)$$

我们看式子(1.6)可以发现，感染率 β 成为影响传染病传播速度的又一个新增因子，由于它 <1 ，而且 λ 和它直接相乘，说明它会直接影响感染人数的增长。

那到底影响有多大呢，数据分析告诉你。

同样的，我们还是对比一下不同的传染人数和感染率下，被感染人数的变化情况

假设某个时点在一个50万人口的县城，由于普及了公共卫生医疗，条件比较好，某种疾病的感染率 β 为10%，我们分别对比一下不同的有效接触人数下感染人数增长的速度。



可以看到，由于感染率 β 值不高，传染病传播的传播速度受到了大幅削弱

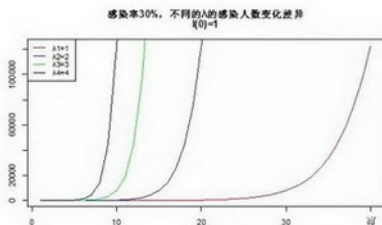
当 $\lambda=1$ ，到了第40天，受感染人数仅接近55人，占总人口比例很小；

而当 $\lambda=4$ ，要到第10天，感染人数也就达到上面这个水平！

这种虽然增长速度很快，但是受感染的力度被 β 大幅削弱了。

因此，和上面那个例子分析一对比，同样的天数受感染的人数下降的有多快！

那么假如对比另外一个县城，人口也是50万，假设卫生条件要差，感染率高达30%，我们再来分别对比一下不同的有效接触人数下感染人数增长的速度。



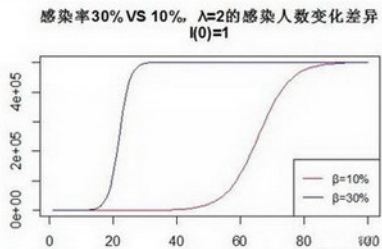
可以看到，由于感染率 β 值是原来的3倍，传染病传播的传播速度一下子就提升了

当 $\lambda=1$ ，到了第40天，受感染人数接近12.3万人，占总人口比例约25%；

而当 $\lambda=4$ ，要到第10天，感染人数也达到上面这个水平！

另外，我们还可以看到感染人数的值 $I(t)$ ，实际上是机器学习里面的Logistics模型的变形，而它是有峰值的。

如下图我们可以看出，假设 $\lambda=2$ ，如果大家都讲卫生，做好自己的防护，感染率从30%降到10%的话，感染的峰值出现的时间会从25天延长到90天。



现在我们再仔细想一下，为什么我们要出门戴口罩，不去和人肢体触碰，要勤洗手，就是要大幅的降低病毒的感染成功率，也就是把感染率给降下来！

那我认真去做了，对社会有多少贡献？

从这个计算中我们可以看出，如果戴口罩能将感染率能

降3倍,那么传染病造成的感染人数的增速,能够下降上千倍!是55人和12.3万人的区别啊!

并且,如果注意个人防护,戴口罩出门、勤洗手勤消毒,就很可能把疾病传染的速度给压制下去,那是25天出现峰值和3个月后才出现的峰值的区别啊!这段时间差内我们可以做很多事情和干预措施的。

这可真是事半功千倍!

换句话说,不必奢求N95,你哪怕只买到了普通的医用外科口罩,你对防止疾病传播的贡献,绝对超出你的想象!



有兴趣的小伙伴,接下来可以看看上面结论的简单推导过程

$$1.5 \text{式} \quad \frac{dI(t)}{dt} - \lambda\beta I(t) + \frac{\lambda\beta}{N(t)} I(t)^2 = 0 \text{的求解过程}$$

$$\text{考虑微分方程的一般形式: } \frac{dx}{dt} + ax + bx^2 = 0$$

方程两边同时乘以 $\frac{1}{x^2}$, 得

$$\left(\frac{1}{x^2}\right) \frac{dx}{dt} + a \frac{1}{x} + b = 0$$

$$-\frac{d\left(\frac{1}{x}\right)}{dt} + a \left(\frac{1}{x}\right) + b = 0$$

$$\frac{d\left(\frac{1}{x}\right)}{dt} - a \left(\frac{1}{x}\right) = b$$

可见,变形后的方程左边是一阶线性齐次微分方程,而方程右边 $\neq 0$,所以先计算左边=0的场景

$$\frac{d\left(\frac{1}{x}\right)}{dt} - a \left(\frac{1}{x}\right) = 0$$

令 $\frac{1}{x} = y$, 则有

$$\frac{dy}{dt} = ay, \text{ 移项变形得}$$

$$\frac{1}{y} dy = a dt$$

两边积分得

$$\ln(y) = at + C$$

$e^{\ln(y)} = e^{at+C}$, 故该微分方程的解

$$y = Ce^{at}$$

下面求C, 假设 $C=U(x)$, 将上面的结论回代得原微分方程形式

$$\frac{dy}{dt} - ay = b, \text{ 令 } y = u(x)e^{at}$$

$$\frac{d(u(x)e^{at})}{dt} - au(x)e^{at} = b, \text{ 整理得}$$

$$\frac{du(x)}{dx} e^{at} + au(x)e^{at} - au(x)e^{at} = b$$

$$\frac{du(x)}{dx} = be^{-at}$$

而前述 $y = u(x)e^{at}$ 即 $u(x) = ye^{-at}$, 因此上式

$$\frac{du(x)}{dx} = \frac{d(ye^{-at})}{dx} = be^{-at}, \text{ 两边积分得}$$

$$\int_0^t \frac{d(ye^{-at})}{dx} = \int_0^t be^{-at} dt$$

$ye^{-at} - y_0 = -\frac{b}{a}(e^{-at} - 1)$, 其中 $y = \frac{1}{x}$ 和 $y_0 = \frac{1}{x_0}$, 上式整理得

$$\frac{1}{x} e^{-at} - \frac{1}{x_0} = -\frac{b}{a}(e^{-at} - 1)$$

$$x = x(t) = \frac{e^{-at}}{\frac{1}{x_0} - \frac{b}{a}(e^{-at} - 1)}$$

我们把 $a = \lambda\beta$ 和 $b = \lambda\beta/N(t)$ 代入得到上式中, 就可以得到

$$I(t) = \frac{NI_0}{I_0 + (N - I_0)e^{-\lambda\beta t}}$$

不过,我们接下来再增加一些模型的复杂度,让它更接近些实际。

三、SIS和SIR模型

这两个模型看懂了,你就更能深刻地理解各地医护人员除了要驰援湖北、武汉外,为啥要用中国的基建狂魔的特有速度,在各地建设当地的“小汤山”和“方舱”。

另外,看懂了模型中的几个参数的组合(其实不难懂),你就明白专家们说的 $R_0 > 1$ 和 < 1 是咋回事儿了?

上面的模型中,我们假设某个观察区域的总人数 N 不变,不考虑区域内出生和其他原因的死亡,也不考虑人口迁移。考虑了不充分接触的场景,也考虑了感染率的场景, SIS 和 SIR 模型的区别,在于下面2点:

SIS 模型增加需要考虑的是,某些传染病病人是能够被治愈的。不过由于自身没有形成免疫(例如伤风、普通流感),康复的人还会有一定的几率重新被传染;

SIR 模型增加需要考虑的是,病人被治愈后,自身形成了免疫(例如麻疹),康复后便终身不再被传染;

因此模型的变量稍微调整如下:

t时刻的总人口数保持不变 N , 用符号 $N(t) = N$ 表示;

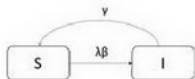
t时刻的易感染者(Susceptible)人数,用符号 $S(t)$ 表示;

t时刻的已感染者(Infected)人数,用符号 $I(t)$ 表示;

t时刻的康复者(Recovered)人数,用符号 $R(t)$ 表示;

可知,对于SIS模型而言, t时刻的 $N(t) = S(t) + I(t)$

对于SIR模型而言, t 时刻的 $N(t) = S(t) + I(t) + R(t)$
 β 为感染者接触到易感染者后, 易感染者被感染的概率;
 γ 为平均每日被感染者康复人数占总被感染人数的比率;
 λ 定义如前, 为每个已感染者每天有效接触的人数;
 现在分别讨论SIS和SIR模型
 先讨论SIS模型



建立模型:

$$I(t + \Delta t) - I(t) = \left(\lambda \frac{S(t)}{N(t)} \beta I(t) - \gamma I(t) \right) \Delta t \quad (1.7)$$

并在此上建立微分方程

$$\begin{cases} \frac{dI(t)}{dt} = \lambda \frac{S(t)}{N(t)} \beta I(t) - \gamma I(t) \\ I(0) = I_0 \end{cases} \quad (1.8)$$

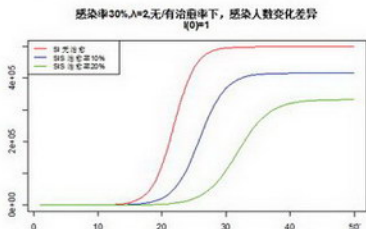
如果刚才读了SI模型的推导部分的小伙伴, 其实很容易就能依据之前的方法得到 (1.8) 的解, 这个解也就是SIS模型下被感染人数随着时间变化而变化的值

$$I(t) = \frac{N(\lambda\beta - \gamma)}{\lambda\beta} \left[\left(\frac{N(\lambda\beta - \gamma)}{I_0 \lambda\beta} - 1 \right) e^{-(\lambda\beta - \gamma)t} + 1 \right] \quad (1.9)$$

这个看上去很多蝌蚪在游的公式, 揭示了什么呢?

考虑了治愈和康复的情况, 治愈率哪怕再低都好 (假设只有10%), 不考虑愈后自身产生免疫, 得到治疗和得不到治疗之间的差异有多大?

我们假设还是那个50万人口的小县城, 最初感染人数还是1, 感染率还是30%, 没有封闭的措施, 每天已感染者有效接触人数 $\lambda=2$, 我们做一下对比。



我们看, 在前提假设不变的情况下:

如果没有得到治疗, 到了25天时感染人数出现峰值;

如果有得到治疗, 哪怕只有10%的治愈率, 感染人数的峰值出现的时间都晚出现5-6天, 不仅如此, 感染人数的峰值也相差接近1万人;

如果有得到治疗, 哪怕治愈率从10%只是上升到20%,

感染人数的峰值出现的时间就能晚出现15天左右, 不仅如此, 感染人数的峰值也相差接近2万人, 减少了约40%!

现在我们可以定量的了解, 我们国家发挥基建狂魔的优势, 迅速的在10天内建好“火神山”、“雷神山”, 以至于全国各地都在抓紧时间抢建自己的“小汤山”和“方舱”, 甚至征用党校和学校、大型场馆做“方舱”的意义有多大!

每天多治愈一例病人, 就意味着全体的传染的人数会有机会大幅的降低。

这就是和时间的赛跑!

另外, 在这个模型里的几个参数, 能够构建出我们经常看到文章里专家们说的 R_0 。

想要理解 R_0 , 就必须从这模型中的几个参数理解开始。

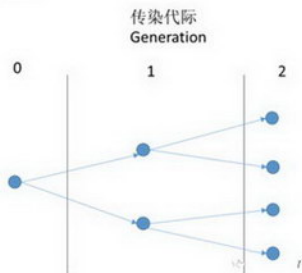
$I(t)$, 代表 t 时刻被感染者的总数, 这个已知;

λ , 每个已感染者平均每天接触到的能有效传染的人数。这个容易理解;

β , 已感染者平均能够成功传染给易感染者的概率, 这个容易理解, 感染率嘛;

γ , 平均每天被感染者康复人数占总患病人数的比率, 这个也容易理解, 治愈率嘛; 那么 $1/\gamma$ 就是平均的传染期 (这个需要好好消化一下)。

现在开始讲 R_0 (Reproductive Number), 它代表了一个被感染者在整个传染期内能够进行继发传染的人数。以下图为例, 假设 $\lambda=2$



上图就是一个继发传染的代数图, 表示了如果 $\lambda=2$ 的时候被感染的速度在第1代时是1传2, 在第二代的时候是2传4, 如此以往, 被感染的病人越来越多。这个时候我们说传染病会爆发。

在不同的模型里面, 由于参数的不同, R_0 的内涵是略有不同, 但可以作为一个阈值来判断在该模型下传染病是否会最终爆发或终止。

以SIS模型为例, $R_0 = \lambda\beta/\gamma$ 。

它的物理意义把公式变形一下就好理解了, $R_0 = \lambda\beta[1/\gamma]$, 就是说 R_0 代表平均每个被感染者在一个传染期内, 在感染率的约束下, 能够传染有效接触的人的数量。

换句话说这就是一个传染病“杀伤力”指标

从上面的公式中, 我们发现:

如果 γ 越大,也就是治愈率越高,那么一个传染期就会拉长, R_0 就会越小。

如果 λ 越小,也就是被感染者接触到健康人的数量越少,那么 R_0 也会越小。这个很自然,你连传染病传播的土壤都给破坏了,它还怎么爆发呢?

如果 β 越小,也就是感染率越低,那么 R_0 就会越小,这个就不在话下了。

因此,从 R_0 的内涵这一点上也说明了,武汉人民付出巨大牺牲采取的“封城”、全国人民“躺在家里不给国家添乱”、“出门戴口罩”、“勤洗手”、“10天建成一座‘火神山’”、“驰援武汉、驰援湖北”,等等措施都是有其重要的价值和意义的。

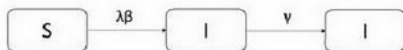
回到 R_0 本身我们再看:

如果 $R_0 > 1$,也就是说1个被感染者如果能够在在一个传染期内传染到至少1个人,那么逻辑上就知道整个群体被感染只是时间上的问题;

如果 $R_0 < 1$,也就是说1个被感染者连一个人都传染不了,那么传染病就会最终被终止;

不过, R_0 的参数中,各个值都是在不断的变化的,因此 R_0 本身也不是个固定的值。但是他可以作为一个阈值来监测,越早发现它比1大,越有利于控制传染病的后续“杀伤力”。

然后是SIR模型



建立模型

$$\begin{aligned} I(t+\Delta t) - I(t) &= \left[\lambda \frac{S(t)}{N(t)} \beta I(t) - \gamma I(t) \right] \Delta t \\ S(t+\Delta t) - S(t) &= -\lambda \frac{S(t)}{N(t)} \beta I(t) \Delta t \\ R(t+\Delta t) - R(t) &= \gamma I(t) \Delta t \end{aligned} \quad (1.10)$$

并在此之上建立微分方程

$$\begin{cases} \frac{dI(t)}{dt} = \lambda \frac{S(t)}{N(t)} \beta I(t) - \gamma I(t) \\ \frac{dS(t)}{dt} = -\lambda \frac{S(t)}{N(t)} \beta I(t) \\ \frac{dR(t)}{dt} = \gamma I(t) \\ I(0) = I_0, S(0) = S_0, R(0) = R_0 \end{cases} \quad (1.11)$$

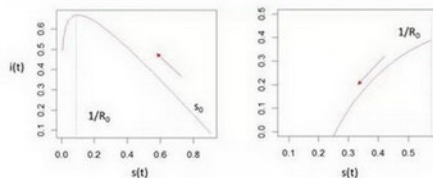
方程(1.11)是没有解析解的。因此得转到相平面来讨论这个解的性质。

为了能够在二维平面内显示相轨线的变化,我们先通过一个特定的 $S(0)/N$ 和 $I(0)/N$ 初始值,根据(1.11)式求出 $I(t)$ 和

$S(t)$ 的相轨线方程。

$$\text{令 } s(t) = S(t)/N, s_0 = S(0)/N \text{ 和 } i(t) = I(t)/N, i_0 = I(0)/N$$

做出相轨迹图如下



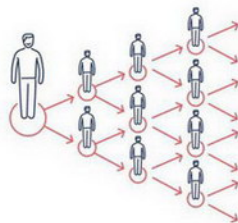
上图中的红色箭头是相轨迹中 $s(t)$ 变化方向。

上图(左)有一个显而易见的认知在生活中是成立的,也就是说当健康人数比率逐渐下降时,说明有传染病发生。当然了,被感染人数比率就自然会上升。

当感染人数比率达到峰值的时候,也就是当健康人数比率达到 $1/R_0$ 的时候。以上图左为例, $1/R_0=0.1$,也就是 $R_0=10$ (假设),那已经是一个相当高的恶性传染病水平了,或者说发生什么严重的事情了。

如果一开始 R_0 控制的不够好,例如 R_0 偏大,此时 $1/R_0$ 出现在 s_0 左侧足够远,那么必然会有传染病暴发的。像这次NCP来势汹汹,其实主要原因倒不是我们国家的公共卫生条件多不好,而是错过了黄金时期。到了后来才去大面积的去进行“刮骨疗伤”,去对 λ 、 β 、 γ 这些有关 R_0 如何降下来的措施下猛药。

而上图(右),一开始 $1/R_0$ 出现在最高值,于是迅速的往下滑,这就是因为缺少传播者或者传播的途径非常难,最终传染病终止了。



而缺少传播者,要么就是提前预警的早,做好了准备(如疫苗、公共卫生宣传和建设),也可以是中途加大了干预(有专门的药物,且很普及)、或遇到了对传播不力的环境因素(如气温升高...等);要么就是这种疾病传染力不强。不过这次来看NCP本来可以是服从第二种的轨迹线的,希望这次的教训之后,前事不忘后事之师,我们都能对自己的子孙后代们负责。

有兴趣的小伙伴,接下来可以看看SIR相轨迹方程的简单推导过程

由 (1.11) 进行变形

$$\begin{cases} \frac{dI(t)}{dt} = \lambda \frac{S(t)}{N(t)} \beta I(t) - \gamma I(t) \\ \frac{dS(t)}{dt} = -\lambda \frac{S(t)}{N(t)} \beta I(t) \end{cases}$$

消去 dt 运算, 即 $\frac{dI(t)}{dt} / \frac{dS(t)}{dt}$

$$\frac{dI(t)}{dS(t)} = \frac{\gamma}{\lambda \beta} \frac{N(t)}{S(t)} - 1, \text{ 由于 } N(t) \text{ 始终不变}$$

$$\frac{d\left(\frac{I(t)}{N(t)}\right)}{d\left(\frac{S(t)}{N(t)}\right)} = \frac{\gamma}{\lambda \beta} \frac{1}{s(t)} - 1$$

$$\frac{di(t)}{ds(t)} = \frac{\gamma}{\lambda \beta} \frac{1}{s(t)} - 1$$

$$\text{引入 } R_0 = \frac{\lambda \beta}{\gamma}$$

$$\frac{di(t)}{ds(t)} = \frac{1}{R_0} \frac{1}{s(t)} - 1, \text{ 两边积分得}$$

$$i(t) - i_0 = \frac{1}{R_0} \ln s(t) - s(t) - \left(\frac{1}{R_0} \ln s_0 - s_0 \right)$$

$$i(t) = i_0 + s_0 - s(t) + \frac{1}{R_0} \ln \frac{s(t)}{s_0}$$

当然了, 还有更多的影响感染人数的因子的模型, 本次

文章没有提到。

例如包含潜伏期的模型、考虑超级传染者的模型、考虑传染模式的模型……, 出于篇幅及知识储备所限, 本次我就不再引用了。

最后还是不忘强调下, 通过这次简单的学习, 我发现定量的去理解一个道理, 有时候确实比定性的去理解它要深刻的多。我也就更深刻理解到这次全民抗疫行动的很多措施的意义。



遗忘在角落里的数据库-Access

作者 / CPDA数据分析师 广州 史金乐 编辑 / 协会会员处 李苗苗 日期 / 2020-02

说起数据库, 大家第一反应肯定是Mysql, Oracle, Mssqlserver等, 这些都是比较常见的数据库, 其实在数据库里面还有一位被大家遗忘的小兄弟, 那就是Access数据库。

Access数据库和Word、Excel、PPT等是Microsoft Office全家桶的组件, 但是我们常用的是其他三个, Access这位小兄弟常常被我们丢到角落里, 甚至被遗忘。

那么问题来了: Access真的没什么用吗?

当然不是一无是处, 而且它还挺有用!

总体上来讲, Access算是数据库中的轻量级中的轻量级, 数据库的基本功能都具备, 界面对我们十分熟悉, 十分友善, 打开后感觉像是Excel。对! 你没听错, 界面和Excel比较相似, 毕竟是一个系列。

对于Excel, 我们常常面临的问题是:

- 1、数据量大了之后, 会容易卡死, 未响应;
- 2、一个sheet只能容纳104万行的数据, 超出容量就需要切成两个或是几个sheet来存储。

解决上述问题我们会将数据存放到Mysql数据库中, 但是又会遇到新的问题, 使用Mysql还要学习Sql语言以及数据库的相关知识, 需要投入时间和精力。这便让许多人望而却步。

如果是上述这种情况, 那么Access数据库就可以完美解决这个问题。首先免安装, Access是Office的组件之一, 如果安装完整版的Office, Access是默认安装的。其次, 这个数据库不需要学习Sql语言, 因为界面类似Excel, 可直接上手, 是不会Sql语言同学的福音。



接下来就从以下几个方面简单介绍Access的使用方法：

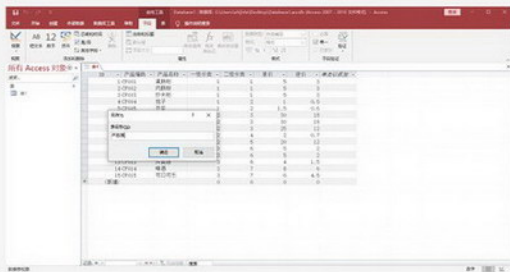
- 1、新建数据库、新建表
- 2、导入导出
- 3、增删查改
- 4、与power query结合使用
- 5、局限性

的数据进去即可，然后保存会出来一个让写表名的窗口，写上表名即可。第二种是导入创建新表，下面介绍。

新建数据库、表

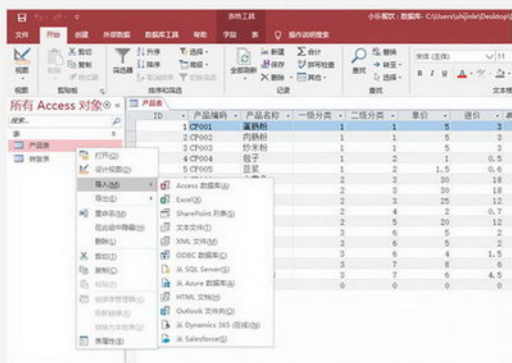


创建数据库比较简单，直接按照图表上面1、2、3创建一个数据库就好了，2那里要写好数据库名称和选择地址，好了之后直接点击创建即可。存放位置随便填写，方便找到即可，因为Access还可以像Excel一样可以随便移动，传递起来非常方便。



创建表的方式有两种，第一种就是直接复制几行带标题

导入导出



导入和导出操作类似，所以这里只介绍导入。如果是空白表导入，则是创建新表；如果是向非空的表导入，则是增加。导入时候根据导航一步一步操作（导航有具体的文字介绍），点击下一步即可，十分简单。

增删查改

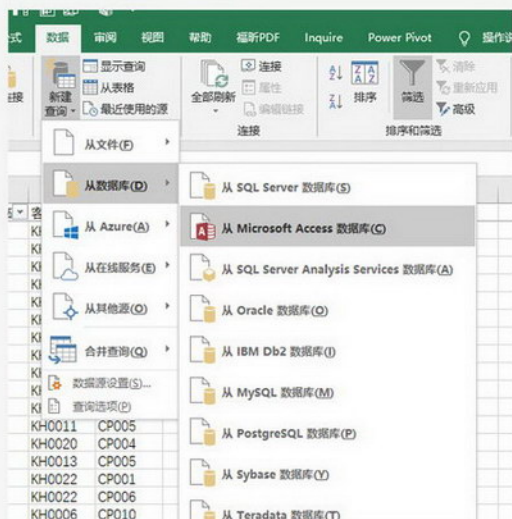
在Access中的“查”是利用筛选功能来实现的，“删”和“改”是在筛选到目标行的基础上来实现的。“增”的实现如上面导入的功能，在原数据的基础上增加行。

当然，如果会VB也可以使用VB来进行增删查改的操作。sql语句也是支持的，但是要结合其他编程语言来实现，比如说python，这部分就超出了本文的范围，不再做详细阐述。

与power query结合使用

之前工作中Excel建模的数据是将每月的源数据放到文件夹里面，用power query来加载文件夹里面的数据来进行加载的。使用power query来加载Access的方法与原来的类似，选

择加载Access即可，具体如下：



局限性

Access的优势也造成了他的局限性，Access可以满足平时Excel建模、存储数据的需求，但是大型、复杂的项目还是要选择传统的关系型数据库。Access数据库内代码编写只支

持VB，无法像传统关系型数据那样进行复杂的操作。Sql语言也支持，但是需要结合Python等编程语言来实现。

总体上来讲，Access的轻便快捷、易于上手，对于不会Sql、Excel满足不了需求的同学来说是个福音。但他的局限性也使他只能是个过度产品，个人建议，等学会了Sql，还是要使用Mysql、Mssqlserver等关系型数据库的。



城市垃圾污染也是新新冠病毒传播途径

作者 / CPDA数据分析师 武汉 李文强 编辑 / 协会会员处 李茜茜 日期 / 2020-02

由于武汉新冠肺炎的来袭凶猛，2020年1月23日武汉封城，随着全国各省市医疗救治队伍大力驰援武汉，使武汉的疫情得到控制，近10多天来，由13日新确诊的新冠肺炎3910人减少到次日的1923人，接着连续4天保持在1500-1700人之间，特别是近几天新确诊病人减少至314-615人之间。

尽管近些天的新确诊人数由4位数减少到三位数，但是，疫情还在继续，有的区域新确诊人数较突出，从武汉市通报的21日13个市辖区新确诊的新冠肺病人数分布情况看，全市有314人，按武汉市地理位置以长江（江南、江北）汉江流域带分块区看，长江以北（简称江北，含江汉、江岸、硚口等6个区）21日新确诊190人，占全市确诊人数的61%，江北6个区常住人口合计为506.8万人，该块区新确诊人数占常住人口

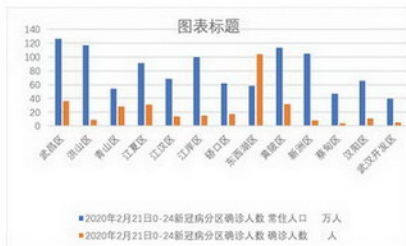
的37.5%，其中，东西湖区常住人口为58.48万人，新确诊的人数为104人，占常住人口的77.8%，并占全市新确诊人数的33%。

长江以南（简称江南，含武昌、洪山、青山、江夏区）21日新确诊104人，占全市确诊人数的33%，4个区的常住人口合计为388.9万人，该地区新确诊人数占常住人口数的26.7%。

汉江块区（简称汉阳，含蔡甸、汉阳、开发区）21日新确诊20人，占全市新确诊的6.4%，汉阳常住人口数合计为172.46万人，确诊人数占本区常住人口数的11.6%。

全市（江北，江南，汉阳三块区域）21日新确诊人数中，从江北，江南分别确诊人数占本区常住人口数的比率看，江北比江南高10.8个百分点，从绝对数看，江北比江南新确诊

的人数多86人。江北6个区中，东西湖区的确诊人数与江南所有区确诊的人数合计相等为104人，21日这天东西湖区的确诊人数占全市的三分之一，是这天的高发区。



武汉自从发生人感染新冠肺炎至今，已有2个多月了，除了人传人的主要途径，还有空间环境的传播是不可轻视的。距离存量消化已有10多天了，从武汉21日一天新确诊人数看，多数区域的这一天均少于40人，而江北东西湖区的新确诊人数在百位数，在全市13个行政区中居首位（上图表汉南区无确诊病人）。东西湖区地处武汉市的北边，属于城郊，当地有金银潭医院（也是这次新冠肺炎治疗的主要医院之一），还有几个殡葬单位和垃圾回收处理焚烧场，由于特殊时期，救治和病亡人员的频频发生，在空间环境和频繁的交通运输途中，免不了有病毒传播的机会存在。

通过如下几点看，也是武汉市新冠肺炎病毒传播或被感染的重要途径：

第一、全市有5个左右垃圾处理焚烧点，分布在远郊区，江北，江南和汉阳块区城郊，殡葬单位也是多个，并分布在各区市郊，东西湖区设点相对稍多。从这次新冠肺炎病毒扩散的情况看，有垃圾转运，也有因新冠肺炎病故者的后事处理均是由运输汽车辗转往返地奔跑，必然会有病毒洒在空间和物体上，甚至直接与路人擦肩而过，都有感染的机会。

第二、医院院区垃圾存放中转点是新冠肺炎病毒集中的毒窝，各医院均有大量医疗垃圾产生，特别是新冠病毒发生的前期，感染人数多，医院能力匮乏，别说严格收集垃圾，病人将医院拥挤的水泄不通，因此，医院除人传人的广泛环境，还有医疗废弃物的堆放处，即是医院的人群通道之一，路人走过，同样广泛传播和感染病毒。

第三、社会公共场所居民生活区域的垃圾集中地，均是集中新冠肺炎病毒传播的自然地。封城后的社会公共场所就是生活垃圾堆放处，武汉在家隔离的人员中有三类人：有健康（非疑似非确诊）的，疑似的和确诊的，三类人产生的生活垃圾均丢在一处，在扔垃圾时势必会有被感染的机会，再是一般生活垃圾均在社区居民点路边，小区居民进出路过垃圾处，过路的人同样易染上病毒。

第四、转运垃圾的车在每个居民住宅点转装过程和运输途中，沿路都是传播病毒的途径，因为垃圾有的是散洒的，一般垃圾袋密封不严，在将垃圾桶往车里倒时，垃圾被散落在垃圾车里，这些垃圾中必定有肺炎病毒，车走多远病毒便沿途随风飘扬。

以上的传播病毒途径形成了暴露空间的点与线的循环传播轨迹，因此，城市垃圾和病人运转过程的污染是不可忽视的新冠肺炎病毒传播途径。为了消除上述新冠肺炎病毒扩散的途径，建议首先加强对居民生活小区垃圾堆放处的消毒，以及对所有接送病人的医疗机构垃圾堆放中转点每天定期消毒。马上市气温回升，垃圾是苍蝇蚊虫最爱处，也会成为病毒传播的媒介；其次，环保部门应考虑改良生活垃圾袋，设计相对密封的、在转载过程中不至于散开的垃圾袋；第三，所有运转病人的车，垃圾车及运转过世新冠病人的车，每趟运行除车内消毒，车外也要加强消毒，因为车身外是暴露在外环境，污染空气，最好像洗车店的方式洗车消毒。

大数据在识别电信罪案中涉及的统计学原理浅析

作者 / CPDA数据分析师 上海 吴梁纯 编辑 / 协会会员处 李苗苗 日期 / 2020-02

通常而言，谈及统计学，人们认为就是将数据通过简单的样本计算和传统的数学模型将其中有用的信息提取出来。然而，放到现代的观念来说，这些认知偏颇狭隘了。

当今，大量的动态数组，千兆、亿兆的数据，在商业、制造业、环境科学、航天航空、数据网络等各个领域屡见不鲜。传统的统计学场景和现代统计学已经相去甚远。

更进一步，随着数据量的增加，计算机科学成为了比传统统计学更为适合处理大数据计算的方式，数据可视化的应用挑战也成为计算机科学发展的一个重要方面。

这篇文章，将浅谈统计学原理在数据信息提取、数据处理和现实应用场景中的研究意义。

1. 信息新定义



信息如何从数据中提取，取决于不同目标主体。有时，信息就是总结一下当前数据，不针对未来决策、过程、实验做推断。但即使是总结也不容易，如果数据量庞大则需要使用特殊函数来处理。更为隐晦的是，目标可能是推断未知参数或关联关系。例如，目标可能是理解一项未知或无法获取的事物，例如性能输出量，这类数据没有误差变量则无法衡量，

因此，难点就是从实验数据中提取关联行和参数来解释这类数据。或者，目标或许是预测当前抵押贷款未来可能会有拖欠情况的部分人员。预测未来价值，而不是推断隐藏的关联关系和参数。举例来说：某一项贷款已经下放，那么目标函数可能需要确定某部分数据应用用来预测还款是否会有拖欠，所以目前的贷款数据只是可用数据的训练集。

推断数据和预测数据都需要数据分析，而不仅仅是数据归集。

2. 基础统计概念

基础统计学模型很简单：

数据=主体数据+噪声数据

主体数据代表主要的数据形态，噪声数据表示围绕主要模型的变量，两者都具有高度复杂性。主体可能是参数类型的，这组参数可能是线性数据、非线性数据、复数、积性函数等。（例如：系数）。另外，主体数据也可以是非参数类型的，例如：阶梯函数、赋值函数或一连串解释型变量。噪声数据表示变量数据，会影响预测和评估的可靠程度。噪声数据是相对独立、标准、相关、持续期间偏差抽样，用于计算非随机样本或结构化数据。例如：如果噪声数据没有附加影响数据源，那更合

Data~ F θ

Mean (data) = g (θ)

适的模型就是：

F θ 是围绕在决策模型g (θ) 周边的分布函数，解释数据的冗余变量。F θ 也可以是著名的分布函数，例如伯努利分布，逻辑回归。其包括了计算非冗余样本，通过解析性变量和关联关系来预测持续期间样本变差等情况。主要参数 θ 有赖于解释型变量和预测型变量。数据归集的过程越复杂则，同济模型的复杂程度越高。

统计学理念认为，噪声数据建模与主要数据建模同样重

要。我们可以从噪声数据中获取可能偏差的预测情况、通过噪声属于的统计，我们可以知道这类属于可用于目标推断或问题预测。

3. 反诈骗行为统计模型

当今，各种各样的电信诈骗层出不穷，信用卡被盗都能引发周边一系列的电话欺诈。电信欺诈案中，作案人可以克隆各种电话号码。通过有线网络，黑客可以攻入大学的电信网络，把所有学生的电话信息都盗取后行骗。订阅欺诈案中，顾客被浏览购物信息窃取支付密码。

我们的目标在于尽快地获取每一次通话记录，并且记录诈骗电话的活跃和终止阶段的信息，更新采集样本。如果通过采集样本，我们可以精准预测出下一次的合规来电，通过已经捕获的来电号码对比预测出来的数据校验精准性。我们将采集对比的合规电话标记为0，而非0的预测来电则有可能为诈骗电话。

通过数据预处理的方法，我们可以定义出样本为0的数据为主体数据。而非0数据为噪声数据。通过逻辑回归等分布函数等噪声数据的处理，可以预测出在一周或者一日之内非0数据的可能来电情况。

但是实际情况远比样本计算复杂得多。我们很难通过简单的状态判断来决策主体变量和非主体变量的预测情况一定是符合现实场景的。

从统计学角度来说，我们将一通电话定义为一组随机向量 $X=(X_1, \dots, X_K)$, X_1 表示通话持续时长、 X_2 表示通话频次（每周一，每天一小时） X_3 表示通话率， X_4 表示号码归属地（例如：国家、地区、城市、行政区层级划分）。当所有可以采集的通话历史信息收集到后。一个合法的呼叫者 i 在通讯数据上会出现一个多元分布视图， y 轴为 C_i, n ， x 轴为 X_i, n . 诈骗分子有个 a 数据与多元分布矩阵 F 完全不同。

4. 结语

计算是处理海量数据分析的关键，统计学还有很多计算处理海量数据的方法要向计算机科学学习。与此同时，统计学也将新的要求提出给了计算机学科。例如计算机的数据挖掘需要提高。统计学原理是推动数据挖掘提取分析的关键原则。但是这不代表统计学具有数据探索意义。数据分块，尤其是当数据量巨大的时候，如何更好的利用数据，使数据更为有意义有作用是需要计算机科学采用更为有力的技术和模型构建方法的。

大量数据产生的更多问题远远不止分析能够解决，需要统计学和计算机科学双方一同发展，两者结合应用来处理。传统统计学与数学紧密相连，数学对于分析海量数据有重要作用。概率学则在每一步统计分析建模中起到关键作用。

总之，我们还有很多需要进步和研究的空间，更高效合理的结合统计学与计算机科学两者，将数据智能的应用场景结合到现实生活。

运用数据分析评估营销效果

作者 / CPDA数据分析师 广西 陈虹坚 编辑 / 协会会员处 李苗苗 日期 / 2020-02

一、分析背景

“得用户者得天下”，越来越多的企业将用户群体拓展及培育视为市场营销的首要目标。本文通过对某产品2019年市场数据进行分析，评估其营销团队一年以来对于该产品市场开发效果及群体拓展情况。

二、分析思路

(一) 产品所处阶段

该产品特性类似于快速消费品，所处区域市场常住人口5000万余人，2019年初产品单周期购买群体20万左右，市场规模不及竞品的1/4,仍有相当大的拓展空间。

(二) 分析目标

围绕营销目标：拓展产品购买群体，提升单周期销售额。对其投入的时间成本、市场机会成本、资金投入等进行营销效果分析。

(三) 数据选择

在支撑平台提取月度销售额、单周期销售额、购买群体记录、不同区间销售额及购买记录。

(四) 影响因素考虑

政策因素（关键）、资金情况、产品特性等。

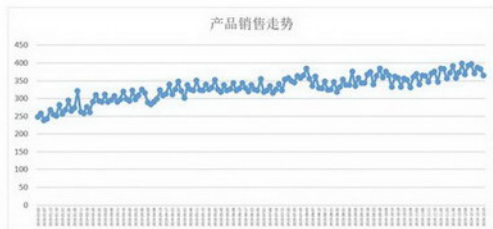
三、数据预处理

在产品支撑平台针对性选取数据字段，剔除无效的属性，进行数据清洗，观察字段集中趋势，均值、众数、中位数，确保数据的唯一性、连续性、一致性。

四、产品销售情况

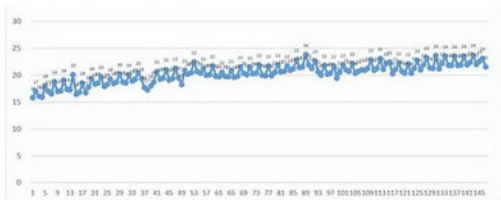
(一) 产品单期销售情况

对该产品2019年以来147个销售周期进行分析，单周期销售峰值接近400，最小值为239，平均值329；从单周期走势上看，在2019年4月前后出现一个折点，即1-4月的销售呈现缓慢上升趋势，4-11月呈平缓波动趋势，期间出现2个小高峰。从趋势上看，产品销量呈缓慢增长，末期较初期单期销售增长了约50。



(二) 产品单期购买群体变化走势

从该产品单期购买人群数量趋势上看，购买群体1-4月呈现缓慢上升趋势，4月起一直稳定在22-24左右，从趋势上看，并没有明显变化。



(三) 增长分析

分析单周期销售走势及群体变化走势不难看出，销售情况略有增长，但购买群体变化不大，需要进一步分析销售增量的来源。

根据业务特点，本案例中4月为年度该产品系列营销推广活动的开始点，11月为结束点，因此分别选取4月及11月的数据进行进一步分析，判断营销投入前后市场的变化，评估活动效果。

下表为该产品4月及11月市场销售情况对比表，从表中不难看出，该产品营销后相比营销前单周期增量有53，其中销售额区间在11-100的增量贡献较大。但销售额增加并不代表购买群体的增加，也有可能因营销活动刺激造成的客单价提升，需进一步对购买群体进行分析。

月份	月销售	2~10 销售额	11~100 销售额	101 以上销售额
4月	4067	1051	1725	1291
11月	4762	1252	2065	1445
增量	695	201	341	154
单周期增量	53	15	26	12

下表为4月，即活动一系列营销活动前及11月，活动后的购买群体对比，因业务特性需要剔除重复购买人群，因此取单周期购买群体增量进行比较。

由数据不难看出，增量贡献最大的11-100销售区间增加购买群体仅8000人左右，2-10销售区间增加购买群体20000人左右。



月份	2~10 群体数	11~100 群体数	101 以上群体数
4月	1700337	760587	46781
11月	1979518	864782	50841
增量	279181	104195	4060
单周期增量	21475	8015	312

五、影响因素分析

从产品单周期销售走势曲线上看，1-4月有一个缓慢上升的趋势，该产品2月进行了政策调整（国家层面），有利政策对市场有利向上得影响；该产品5-7月有一个缓慢上升的走势，7月下旬达到一个高点，后又下滑，10月迎来第二个高点，后又下滑，11月除缓慢上升后下旬下滑；整体呈现波浪形走势。

结合该产品一系列的营销推广活动时间线分析，5-7月开展第一阶段活动，营销宣传资金投入后，销售额增长，7月下旬达到高点，第一阶段活动结束后销售额回落；10月开展第二阶段活动，活动期间走势重复第一阶段，活动结束后销售额回落；11月因产品自身资金累计及分派，出现政策利好，销售额又缓慢上升，资金分派完毕后下旬回落。

综上，影响该产品销售走势的因素有：政策调整、营销

资金投入、宣传资金投入、产品资金累计及分派。

六、营销投入分析

该产品在2019年5月、7月、10月分别开展了营销活动，结合当地节庆开展品牌宣贯，总计花费300万元左右。

七、营销效果评估

通过上述数据的分析，该产品2019年以来通过各阶段营销资金的投入，排除4月前因政策影响的增长，经费投入前较投入后单周期销售额有53万元的增长，但通过对群体的分析得出，销售的增长并非来源于新群体的增加，而是来自营销经费刺激下原有群体客单价的提高，并且经费使用完毕后销售额回落，表明原有消费群体愈加精明，营销活动期间的经费支持下提高购买金额，活动结束后回到原有购买水平，但活动对新群体的吸引度不大，经费使用未达到理想的市场效果，仅能在短时间内通过增加投入拉高销售额，对于市场及群体的培育效果欠佳。

八、结论与建议

综上，该产品通过一年以来的资源及时间投入，营销效果未达预期，建议营销团队加强产品特性及用户群体研究与挖掘，评估营销投入产出、时机选择，设计更适合产品及区域市场的群体拓展方案。

微信公众号文章量化分析

作者 / CPDA数据分析师 苏州 刘雁南 编辑 / 协会会员处 李茜茜 日期 / 2020-02

1 分析背景

根据2016年1月中旬到7月中旬的公众号推文数据，对文章进行分析，以分析出文章质量的优劣，从而进行一步将文章特征细化，最终达到为公众号内容运营优化提供指导依据。

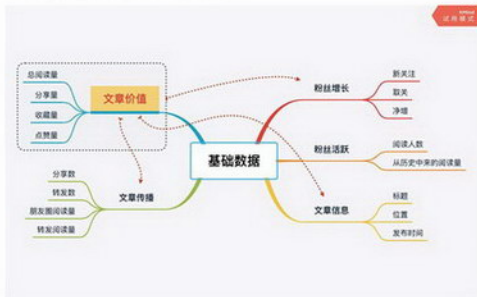
2 现有数据基本情况

2.1 现有数据共计202条，为2016年1月中旬至2016年7月中旬数据。

2.2 现有数据中包括：发文时间、文章标题、图文位置、送达人数、取关人数、新增关注、净增粉丝、分享人数、收藏人数、从公众号分享到朋友圈数量、朋友圈转发数量、朋友圈阅读量、好友转发量、公众号会话阅读人数、来自历史文章的阅读人数、图文阅读量、总阅读量。

2.3 现有数据基本完整，仅有两篇文章缺少标题。

2.4 由于本次数据分析的主要目标是对用户阅读行为进行分析，从而对文章画像，为内容运营优化提供指导，因此重点分析文章价值对粉丝增长、文章传播的影响，同时分析文章信息与文章价值之间的关联度。



3 文章价值分析

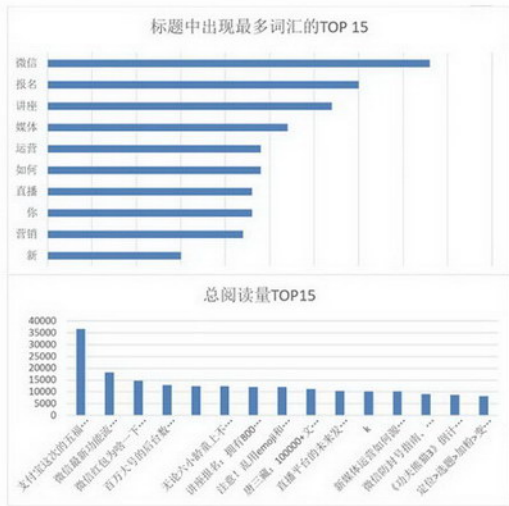
3.1 直接以文章标题进行分析

3.1.1 文章标题有缺失，根据总阅读量排名前15位中，有两篇文章缺少标题，需要补充。

3.1.2 文章标题中词频图谱，可以说明公众号文章标题中最常用的词汇。

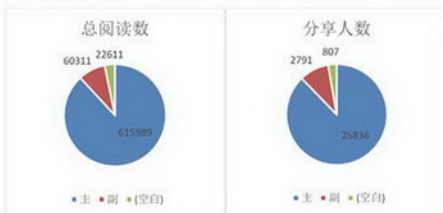


3.1.3 总阅读量高的文章，与词频对比，以确定文章标题中出现某些词汇对读者有直接吸引力。“微信”在所有文章标题中出现率较高，在202篇文章标题中出现43次，占21%；总阅读量排名靠前的文章中，“微信”在标题中出现的频率占30%，可以做出判断，该公众号读者对于与“微信”相关的内容较感兴趣。



3.2 文章信息对文章价值的影响

3.2.1 由现有数据可看出，在公众号主图文的位置文章的总阅读数远高于副图文的阅读数；主图文位置文章的分享人数也远高于副图文。



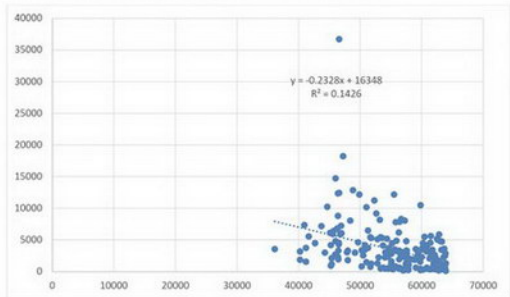
3.2.2 在内容运营时，可以参考此情况：将需要重点宣传和高质量的文章放在主图文位置，便于客户在第一时间掌握和第一时间分享，迅速增加阅读量、分享量，进一步增加粉丝量。

3.3 总阅读量的主要影响因素

3.3.1 首先考量送达人数与总阅读量的关系。

3.3.1.1 送达人数以公众号粉丝为基数，一般认为，送达人数越多，则阅读量相应增长。

3.3.1.2 根据现在数据分析，送达人数与阅读量非正相关，反而是送达人数增长时，会带来一定的阅读量下降，因此，送达人数不能作为内容运营的主要目标。



3.3.2 通过分析，把原始数据中的一些数据做合并处理。

3.3.2.1 关注净增粉丝数，以此确认文章的质量优秀，为公众号带来新的粉丝。

3.3.2.2 关注文章收藏人数，这能说明文章有质量，客户愿意收藏后仔细阅读或者是反复阅读。

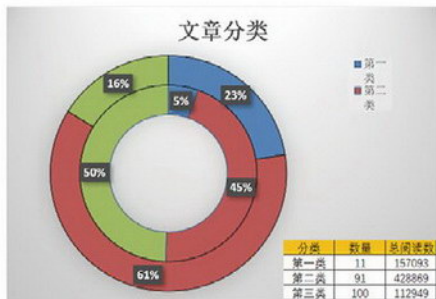
3.3.2.3 关注分享数，这里包括了从公众号直接分享、在朋友圈再次分享。这都能说明因为文章质量好，客户愿意分享并因此而引发讨论。

3.3.2.4 关注因分享而带来的阅读量，这里包括了在朋友圈的阅读人数、来自好友转发的阅读数。这能说明分享的客户群体本身认同文章内容，并因这些分享而带来新的粉丝、新的阅读量，从而能进行广泛传播。

3.3.2.5 关注自带流量，这里包括直接在公众号进行的阅读、来自历史文章的阅读、图文阅读。这是公众号一直在累积客户的结果，如果公众号持续产出优质的文章，阅读量会保持在一个恒定的较高的水平。

3.3.3 通过以上数据的分析，将文章分为三个类别，分好的三个类别用总阅读量进行检验，基本符合实际情况。

3.3.3.1 三个类别的文章中，按其数量和总阅读量分析：



第一类文章数量占总数据量的5%，提供了总阅读量的

23%，这类文章为最优质的文章，可以在后期重点关注，可以将文章的模式做出总结；

第二类文章数量占总数据量的45%，提供了总阅读量的61%，这类文章可以成为日常的主要推送类型，保持公众号文章的曝光度；

第三类文章数量占总数据量的50%，提供了总阅读量的16%，这一类文章需要进一步分析，从中总结出最不成功的模式，在未来的运营中要重点避免这类型的文章大量出现。

文章标题	总阅读量
微信可以向全世界打电话了	7311
微信红包为啥一下火了？今晚20点摇一摇还会怎么玩	14690
无论六小龄童上不上春晚，这家公司都赚了	12355
	12404
支付宝这次的五福红包有多尴尬？	36698
微信最新功能流出，预告了新媒体的未来	18225
百万大号的后台数据告诉你：爆文是怎样炼成的	12849
讲座报名：拥有800万小老婆的关爱八卦成长协会凭啥躺着把钱赚	12148
注意！乱用emoji和表情包可能被判刑	12141
多开微信会被封号	7788
直播平台的未来发展的7个趋势	10484

3.3.3.2 重点关注第一类文章，是占比小，但总阅读数高的类型。需要总结这一类文章的标题、内容、板式等，以此判断客户的关注重点。

4 总结

4.1 该公众号客户对于与“微信”相关的内容较感兴趣。

4.2 主图文的位置文章的总阅读量远高于副图文的阅读数，因此需要重点推广、宣传的内容应该放于主图文。

4.3 分类总结文章：

4.3.1 第一类文章数量占总数据量的5%，提供了总阅读量的23%，这类文章为最优质的文章，可以在后期重点关注，可以将文章的模式做出总结；

4.3.2 第二类文章数量占总数据量的45%，提供了总阅读量的61%，这类文章可以成为日常的主要推送类型，保持公众号文章的曝光度；

4.3.3 第三类文章数量占总数据量的50%，提供了总阅读量的16%，这一类文章需要进一步分析，从中总结出最不成功的模式，在未来的运营中要重点避免这类型的文章大量出现。

基于新型冠状病毒肺炎疫情下的数据与模型思考

作者 / 山东智谷数据分析师事务所 徐斐飞 编辑 / 协会会员处 李苗苗 日期 / 2020-02

基本概述

2019年12月以来,湖北省武汉市部分医院陆续发现了多例有华南海鲜市场暴露史的不明原因肺炎病例,后已证实为2019新型冠状病毒感染引起的急性呼吸道传染病,随着疫情蔓延,我国其他地区及境外多个国家地区也相继出现此类病例。2020年2月,国家卫健委1号公告将冠状病毒感染的肺炎纳入《中华人民共和国传染病防治法》规定的乙类传染病,并采取甲类传染病的预防控制措施。截至3月4日,国家卫生健康委员会(以下简称“国家卫健委”)官网已发布试行第七版的新型冠状病毒肺炎诊疗方案。随着疫情的不断蔓延,国家、省及各地区相应采取一系列防控措施,积极应对疫情发展。

在疫情全球肆虐之下,作为数据从业者,数据在哪些方面可以助力防控疫情,如何利用这些数据等问题,都值得我们去思考与研究。有专家学者表示,在疫情爆发前,数据就已经在发挥举足轻重的作用,数据可以清晰地描绘出人与动物接触的方式及频次,以及多地物种数量的分布情况等,基于这些数据,专家学者不难分析出病毒有多少几概率是来源于动物。在此前的SARS、埃博拉病毒等大范围爆发的流行病,科学家估算出,约四分之三的新病毒都是源于动物,多达80万种未知动物病毒可以传染给人类,这正是通过数据解析之后获得有价值的信息。在此基础上,整合已知病毒、动物数量、人口统计等,利用新技术来预测新病毒可能出现的区域,从而帮助政府和公共卫生机构提前采取措施,在疫情爆发前做好准备及预防性工作。

新冠肺炎疫情影响发生以来,各地政府也通过多种渠道发布疫情数据,对于满足公众知情权,加强自我防护起到了积极重要的作用,还有些数据被专业人员进一步利用,开发出许多受欢迎的应用服务。但是,各地发布数据的内容和方式各有不同,带来的阅读体验和应用价值也不相同。本文将从数据源头入手,以不同的角度探讨此次疫情的数据标准及质量实施情况,并结合已发布的数据,分析呈现疫情趋势,尝试探讨构建趋势拟合模型。

由于本人能力有限,以及时间上的仓促,恐难免有以偏概全之论,不足之处,望予以批评指正。

二、数据来源

国家卫健委官方网站、国家各地区卫生健康委员会官方网站、山东省公共数据开放网站、百度疫情实时大数据报告。

三、数据透析

基于政府发布的新型冠状病毒肺炎确诊病例、疑似病例、治愈出院情况、死亡病例等数据,从政府数据发布形式,提供格式等方面探讨数据标准、质量情况,并以公开发布数据呈现分析全国各地区、省(以山东省为例)、市(以泰安市为例)疫情发展趋势。

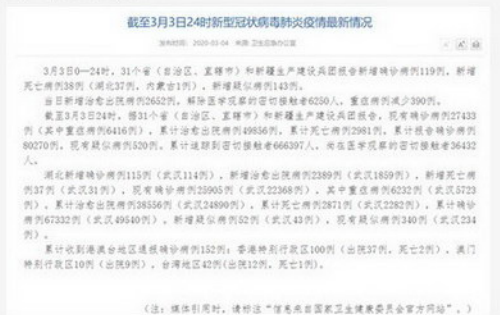
(一) 全国疫情数据

1. 数据采集



国家卫健委在其官方网站最显著位置增设疫情防控专题模块,涵盖疫情通报、防控动态、通知公告、医者风采、防控知识、你问我答、普法知识、新闻报道功能,其中本文全国疫情数据均来自疫情通报。

国家卫健委每日发布昨日数据,包括31个省(自治区、直辖市)和新疆生产建设兵团报告新增确诊病例、新增死亡病例、新增疑似病例、当日新增治愈出院病例、解除医学观察的密切接触者、累计治愈出院病例、累计死亡病例、累计报告确诊病例、现有疑似病例、累计追踪到密切接触者、尚在医学观察的密切接触者等统计数据,另外湖北省疫情数据再单独通报一次,并发布港澳台地区的通报确诊病例。疫情通报数据格式如下(文字+数字):



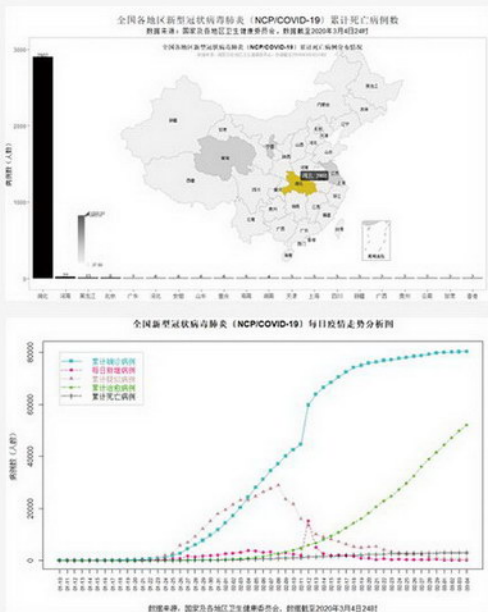
(注:媒体引用时,请标注“信息来自国家卫生健康委员会官方网站”。)

2020年1月10日起，每日通报疫情数据，1月10日至1月19日数据以武汉地区为主；2月12日，对湖北省以及湖北省以外其他省份的病例诊断标准进行了区分，湖北省增加了“临床诊断病例”分类，对疑似病例具有肺炎影像学特征者，确定为临床诊断病例，以便患者能及早按照确诊病例相关要求接受规范治疗，进一步提高救治成功率，因此会导致确诊病例数量突然大量增加；2月13日，湖北省因重复统计，核减108例；2月21日，湖北省将19日、20日核减的病例数重新加回到确诊病例，并对当日新增病例数进行了校正。疫情通报数据格式期间更改过一次，整体数据比较完善。

各省级疫情数据均来自各省卫健委网站疫情通报。各省卫健委疫情通报模块，大部分与上级单位一致，疫情前期每天通报昨天数据，中途更改为每半天通报一次，通报格式略有调整。无论是开始的“文字+数字”形式，还是现在“文字+数字+图片”形式，均不便于被开发利用，需要对这些数据先做一些技术处理，将数据从这些文字中、图片中提取出来，整理成结构化数据。

2. 数据呈现

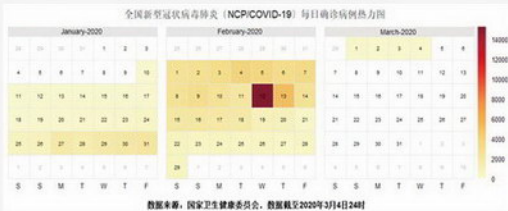
数据呈现主要从累计确诊病例、累计治愈病例、累计死亡病例、每日新增病例和累计疑似病例等五个方面展示全国及各地区疫情形势。



从上图，可以看出2月11日全国累计确诊病例持续增长，2月12日突然呈垂直增长，经查询材料发现，2月12日骤增的重要原因主要是，为做好新型冠状病毒肺炎患者早诊早治，落实好湖北病例应收尽收、应治尽治工作，按照《新型冠状病毒肺炎诊疗方案（试行第五版 修正版）》，对湖北省以及湖北省以外其他省份的病例诊断标准进行了区分。

湖北省增加了“临床诊断病例”分类，对疑似病例具有肺炎影像学特征者，确定为临床诊断病例，湖北省报告的13,332例临床诊断病例纳入确诊病例统计，以此来降低病死率，所以导致当天，数据异常增长。

累计疑似病例从2月8日起，持续下跌，每日增加确诊病例数自2月12日起，持续下跌，累计治愈病例数呈指数增长，可见疫情在全国上下齐心协力一致抗击疫情下，态势得到了良好的控制。



从每日确诊病例热力图中看出，2月12日颜色程度较深，表明这一天确诊病例较其他时间确诊人数众多。自2月下旬以来，颜色程度较浅，也印证了确诊病例每日持续下跌的事实。

(二) 省级疫情数据 (山东)

1. 数据采集



省卫健委与国家卫健委同样，在其官方网站最显著位置增设疫情防控专题模块，涵盖通知公告、防控动态、医者风采、健康知识、媒体视窗、一封家书等功能，其中本文省级疫情数据均来自通知公告。另外，省卫健委在官网首页增设了万众“疫”站 谨防严控和“疫”线视频两个专题模块。

省卫健委提供数据格式1月份与国家卫健委一致，“文字+数字”的形式对外公布，自1月30日12时~24时开始，通报形式变为“文字+数字+表格图片”，表格图片对于一般读者，便于阅读和理解，但是对于数据使用者来说，但还不好用，如要对这些数据进行分析，需要先将图片中的数据转化为机器可以读取和处理的结构化数据。数据表格在不同的时间发布，呈现为不连续的、碎片化，如果想要分析某段时期内的趋势，就要把之前发布的一条条通报搜寻出来，然后再把其中的数据整合起来。

为充分分析挖掘疫情数据，向社会提供可获取的、完整的、结构化的数据，用于疫情防控、帮助公众理解数据背后信息等应用的开发，省卫健委在官网通知公告中每日通报外，还通过省数据公共开放网，将全省及各地区疫情数据以开放数据为基本标准，向社会提供excel、csv、XML、json、API等形式的疫情数据服务，数据内容丰富，数据整体质量较高。

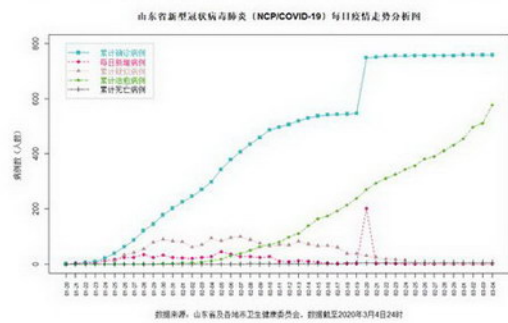
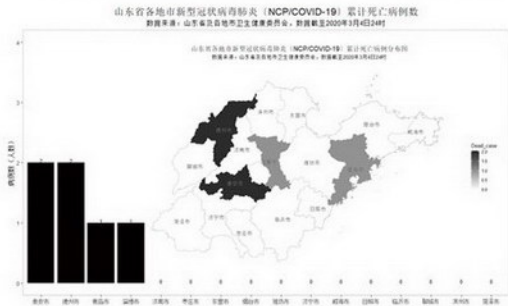


本部分利用的数据主要来源于省卫健委官方网站的公告，省公共数据开放网发布的疫情数据作为补充。

2. 数据呈现

省级数据呈现主要从累计确诊病例、累计治愈病例、累

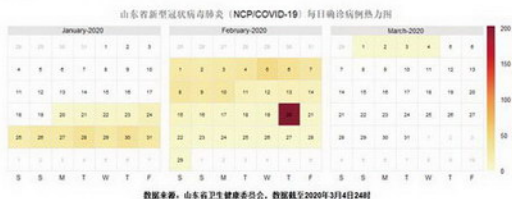
计死亡病例、每日新增病例和累计疑似病例等五个方面展示山东省及各地市疫情形势。



从上图，可以看出2月10日山东省累计确诊病例持续增长，2月11日至2月19日累计确诊病例增速放缓，但在2月20日，突然骤增，经查询材料发现，2月20日骤增的重要原因主要是，山东省任城监狱疫情感染事件造成的，当天仅任城监狱新增确诊200例，2月21日起累计确诊病例趋于平缓。

累计疑似病例从2月13日起，持续下跌，每日增加确诊病例数自2月10日起，持续下跌，仅因任城监狱感染事件，出现异常数据外，其余时间一直保持下跌态势。

累计治愈病例数持续增长，在省委省政府及各地市政府共同努力下，疫情发展态势得到了控制，为复工复产奠定了良好的基础。



从省每日确诊病例热力图中看出，2月20日颜色程度较深，表明这一天确诊病例较其他时间确诊人数众多。自2月下旬以来，颜色程度较浅，也印证了确诊病例每日持续下跌的事实。

(三) 市级疫情数据 (泰安)

1. 数据采集

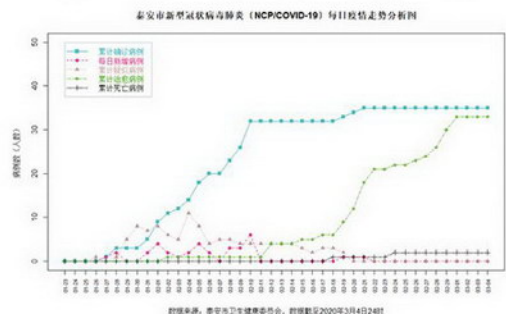
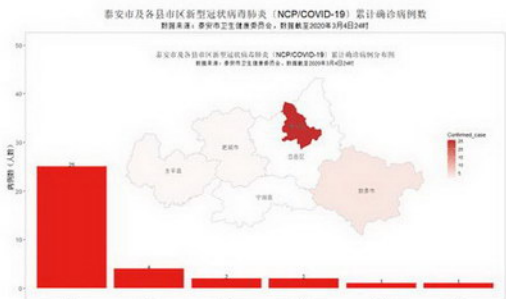


同样，在市级卫健委官网醒目的位置，设置有新型冠状病毒肺炎疫情防控专栏。每半天更新上半天的数据，数据格式到目前为止，主要是“文字+数字”一直未有更改，但数据更新时间，多次调整（由一天更新一次，变为两次，中途又出现一天更新一次），自2月9日到现在，数据更新时间（一天两次）保持一致。

本部分采用的数据均来自本级卫健委官网发布的疫情通告，在数据存在异议的情况下，结合上级部门的数据进行校正。

2. 数据呈现

市级数据呈现主要从累计确诊病例、累计治愈病例、累计死亡病例、每日新增病例和累计疑似病例等五个方面展示泰安市及各县市区疫情形势。



从上图，可以看出2月10日泰安市累计确诊病例持续增长，2月11日至2月18日累计确诊病例不再增加，2月19日至2月21日增长速度放缓，随后不再有病例增加。累计疑似病例从2月4日起，持续下跌，每日增加确诊病例数自2月10日起，下跌至0，除2月19日至2月21日期间每日新增1例确诊病例外，后期均无新确诊病例。累计治愈病例数一直呈增长态势。



从市级每日确诊病例热力图中看出，2月10日颜色程度较深，表明这一天确诊病例较其他时间确诊人数众多，2月1日、2月5日确诊人数次之。自2月下旬以来，颜色程度较浅，印证了确诊病例每日持续下跌的事实。疫情态势的良好控制，得益于市委市政府在疫情期间积极落实上级防控政策，因地制宜的采取一系列符合现实状况的防控措施。

四、疫情模型

通过构建统计学模型、数学模型，或者利用机器学习、深度学习方法拟合疫情发展趋势，利用历史数据对未来的确诊病例等疫情形势进行预测，比如说，逻辑斯蒂生长曲线拟合数据，预测未来几天可能的发展趋势；或者利用时间序列模型构建预测模型；也可用LSTM构建预测模型，一种特殊的RNN网络。以上方法，除生长曲线外，其他模型，需要大量数据做训练，就目前情况看，数据量并不大，即使构建出模型，参考价值并不大，并没有与业务做融合，只是以数据理解数据。

另外一个建模思路，可以从传统疾病传播模型（SIS、SIR、SEIR等），建立传染病模型，结合此次冠状病毒的传播特性，利用现有的样本数估计出一个大概的参数，建立适当的传染病数学模型，能较为精准的预估疫情的发展趋势，当然这是一个较为复杂且专业的问题。近日，由钟南山院士团队研究构建的「具有饱和发病率（其解释，任何传染病都具有饱和发病率，即不可能完全被消灭）SIQS 传染病模型」虽然被国外权威期刊退回，但研究成果还是符合国内疫情发展趋势。

据有关学者介绍，SIQS传染病模型实际上是在传统SEIR模型基础上，加上两个干预因素，即国家的强力干预和春节后的回程高峰，另外，2020年2月28日，钟南山院士团队发表了一篇名为《公共卫生干预下COVID-19流行趋势的 SEIR和AI预测修正》，将2020年1月23日前后的人口迁移数据及最新的新冠肺炎流行病学数据整合到SEIR模型中生成流行曲线，同时，团队还利用人工智能技术，以2003年SARS数据为基础进行训练，从而更好地预测新冠疫情。研究团队还使用长短期记忆模型，预测新增感染数随时间的变化。

对于基本训练数据集的处理，研究团队利用 2003年4-6月SARS的病例统计，纳入COVID-19流行病学参数。从钟南山院士团队的研究成果来看，假设是一支纯技术团队，是无法作出解释性强、可信度高的预测模型，所以说数据建模不仅仅依靠的是技术工具，更多的是业务理论背景，模型不应该是冰冷的技术实现，更应该是有温度、有内涵的业务与技术的融合。

因本人不具备传染病、医疗专业领域相关知识，从非专业角度，尝试利用Logistic生长曲线模拟泰安地区累计确诊病例数量，并试着简单叙述传统疾病传播模型-SEIR。

（一）Logistic生长曲线

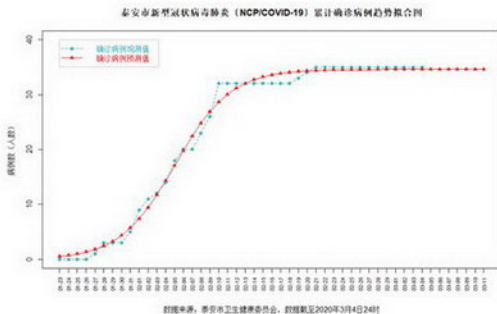
逻辑斯蒂曲线是由比利时数据学家首次发现的特殊曲线，后来，生物学家皮尔(R. Pearl)和L·J·Reed根据这一理论

研究人口增长规则，因此，逻辑斯蒂生长曲线也被称为生长曲线或者珍珠德曲线。逻辑斯蒂生长曲线一般形式如下：

$$Y_t = \frac{L}{1 + ae^{-bt}}$$

L,a,b均为未知参数，需要根据历史数据进行估计。生长曲线在现代商业、生产行业、生物科学等方面有着非常广泛的应用。

我们利用生长曲线模型，拟合泰安本地累计确诊病例数据，建立生长曲线模型。数据拟合如下图所示，蓝色部分显示的确诊病例观测值，红色部分显示的是确诊病例预测值，并计算出未来7天的确诊病例数据。



从上图预测值生成的曲线来看，生长曲线模型整体呈现“S”型，按照相关参考文献说明，生长曲线可以分为初期、中期和末期三个阶段：

在初期，虽然 t 处于增长阶段，但是 y 的增长较为缓慢，这时曲线呈现较为平缓的上升；在中期，随着 t 的增长，y 的增长速度逐渐加快，曲线呈现快速上升的态势；当达到拐点 (t, Y) 后，因函数饱和程度的增长达到末期，随着 t 的增长 y 的增长较为缓慢，增长速度趋近于 0，曲线呈水平状发展。

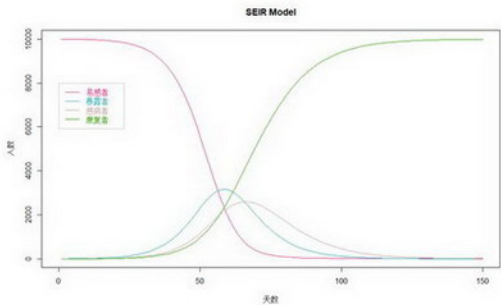
在了解模型特点后，假设外部因素干预事件发展，就会导致数据的突然增多或减少，会影响模型的预测精度。因此，logistic 增长模型只是对疾病进行预估，并不能准确判断，也并不是最佳模型。当然可以通过模型优化，来提高预测精度，有的文献提出可以根据华罗庚提出的 0.618 选优法，对得到的模型进行优化（计算该模型是否能得到预测值和测量值最小残差平方和）。这里我们就不再展开，可以后期进行探讨学习。

（二）疾病传播模型-SEIR

查阅相关文献后，发现常见的传染病模型按照传染病类型分为 SI、SIR、SIRS、SEIR 模型等，用于研究传染病的传播速度、空间范围、传播途径等问题，用来指导对传染病的预防和控制。模型中涉及 S、E、I、R、r、β、γ、α 参数：

S类：表示易感者 (Susceptible)，指未得病者，但缺乏免疫能力，与感染者接触后容易受到感染；E类：表示暴露者 (Exposed)，指接触过感染者，但暂无能力传染给其他人的人，对潜伏期长的传染病适用；I类：表示感染者 (Infectious)，指染上传染病的人，可以传播给 S 类成员，将其变为 E 类或 I 类成员；R类：表示康复者 (Recovered)，指被隔离或因病愈而具有免疫力的人。如免疫期有限，R类成员可以重新变为 S 类。

r：感染患者 (I) 每天接触的易感者数目； β ：传染系数，由疾病本身的传播能力，人群的防控能力决定； γ ：恢复系数，一般为病程的倒数，例如流感的病程5天的话，那么它的 γ 就是1/5； α ：潜伏者的发病概率，一般为潜伏期的倒数。



我们这里不再利用采集到的数据，模拟疫情发展形式，一方面原因是我们并不能较好的估计模型中涉及到各个参数，需要考虑的参数较多，另一方面数据并不能支撑其模型推导，特别是疫情的政府干预因素、社会舆情因素，对疫情发展趋势都会产生一定的影响，应将相关的因素考虑进去。

所以这个问题相对来说是比较复杂的过程，我们这里不再进行过多探讨。大家有兴趣的可以去查找相关文献材料，进行深入研究学习。

五、参考文献

[1] 百度百科. 新型冠状病毒肺炎[EB/OL].

<https://baike.baidu.com/item/%E6%96%B0%E5%9E%8B%E5%86%A0%E7%8A%B6%E7%97%85%E6%AF%92%E8%82%BA%E7%82%8E/24282529?fr=aladdin>.

[2] 刘俊囊. 疫情中数据的作用无法忽视，但也不能盲目信模型 | 专访前DHS生物监管总监[EB/OL].

<http://www.bigdatadigest.cn/#/articleDetail?articleId=591&type=0>

[3] 国家卫健委官网. 全力做好新型冠状病毒肺炎疫情防控工作[EB/OL].

http://www.nhc.gov.cn/xcs/xxgzbd/gzbd_index.shtml

[4] 山东省卫健委官网. 全力做好新型冠状病毒肺炎疫情防控工作[EB/OL].

<http://wsjkw.shandong.gov.cn/ztlz/rdzt/qlzhfkgz/index.html>

[5] 泰安市卫健委官网. 新型冠状病毒肺炎疫情防控 [EB/OL].

<http://wjw.taian.gov.cn/col/col119732/index.html>

[6] 百度. 新型冠状病毒肺炎疫情实时大数据报告[EB/OL].

https://voice.baidu.com/act/newpneumonia/newpneumonia/?from=osari_pc_1

[7] 山东省大数据局. 山东公共数据开放网[EB/OL].<http://data.sd.gov.cn/>

[8] 金涛. 逻辑斯蒂增长曲线模型的优选研究[J].1003-8965(2015)04-0130-01

[9] 雷锋网. 钟南山等人用 LSTM 力证“早发现早隔离”重要性[EB/OL].

https://www.sohu.com/a/377229230_114877

[10] CSDN昵称：土豆西瓜大芝麻. SEIR传染病模型[EB/OL].

<https://blog.csdn.net/jinking01/article/details/104145509>

[11] 郑磊. 哪些地方开放了真正“能用”和“好用”的疫情数据？这些数据能用来做什么？ [EB/OL].<https://mp.weixin.qq.com/s/2V6tF58gJkxA4-GLRbJ1mw>

=====

联系方式：

公司：山东智谷数据分析师事务所

地址：泰安市高新区南天门大街1110泰山科技创业城A区

公司电话：0538-8932988

公司邮箱：data@zzdz.com.cn

联系人：苏航

公司公众号：



疫情下的《社区互联网应用》调查问卷分析报告

作者 / 山东智谷数据分析师事务所 苏航 编辑 / 协会会员处 李苗苗 日期 / 2020-02

本次问卷调查的对象为社区居民，社区范围是山东省泰安市，共包括90个小区社区。其中泰安市泰山区41个，岱岳区28个，高新区13个，肥城市3个，新泰市2个，宁阳县2个，东平县1个。调查对象中，男士占58%，女士占42%。年龄构成上，30岁-50岁之间的居民占68%，30岁以下的居民占26%，50岁以上的居民占6%。全部居民中，76%是业主，20%是家属，剩下4%是租客。

针对社区是否具备面向居民的社区应用问题，80%居民反映无任何社区应用，拥有社区公众号的仅占4%。共有30%的居民反映所在社区智能化或信息化管理水平不好，他们中96%的是社区无任何信息化应用的；另有56%的居民反映社区智能化或信息化管理水平一般，他们中有76%是社区无任何信息化应用的。

居民回忆入住小区或者社区时完成过的登记方式时，47%的居民表示使用了纸质表格进行登记，23%的居民表示使用了信息系统进行登记，7%使用了其他方式登记，另有23%表示没有登记。可以看出，使用现代化信息系统进行登记的比例还是不高的，纸质登记表在疫情来临时，很难做到快速实时的数据分析统计。

关于通知下达

疫情期间，政府和社区管理者基本每日都会下达最新通报或防疫类通知，那么社区又是如何进行下达的呢？有64%的居民表示通过QQ群或者微信群接收社区通知，使用电话或者通知板的社区占到62%，另有6%的居民反映没有渠道接收通知，仅有1%表示使用社区应用接收通知。（本项为多选题，所以请勿简单累加百分比验证数字正确性。）

此次疫情下，居民纷纷响应号召，减少出门，那么社区通知板并非很有效的通知方式。电话通知的方式，因居民众多，也显得不太可行。社交软件群组通知的方式，您又会不会存在因开启了免打扰或没有被‘@’而错过的情况呢？所以笔者认为，社区应用中的虚拟通知板是比较有效的通知方式，然而在实际使用中并不多。

关于信息上报

疫情下社区管理者掌握居民每日健康状况显得尤为重要，那么现实情况下，居民们又是如何上报每日健康信息或其他特殊事件的？49%的居民说使用了QQ群或者微信群上报；17%使用过电话上报；27%反映没有渠道；14%使用了纸质表格；9%是其他渠道。（该项为多选题）

疫情下，使用纸质表格提报上交必然存在接触风险，并

不提倡；全员每日通过电话上报健康状况，对于社区管理者是一种负担，仅上报异常还能接受；那么将近一半的居民说使用社交工具的群组交流进行上报，且不说是否会涉及隐私泄露，试问，管理者又该如何用肉眼及时发现多数正常记录中的异常，又要花费多么的时间人工将一段段非结构化数据一条一条复制粘贴到其它表格中，进一步向上级汇报呢？所以，社区急需一套方便居民上报、数据自动汇集、智能筛查预警的管理软件。

关于物业报修

关于居民使用何种方式向物业进行报修的问题，70%的居民使用了电话，20%使用了QQ或微信。那居民对社区处理报修人员的响应速度、处理时间或处理结果，是否满意呢？15%的受调查居民表示不满意，62%的居民表示一般，表示很满意的仅占14%。

笔者认为，居民对物业报修的满意度不高，主要原因是没有建立规范有效的物业报修、任务派发、任务反馈、业主评价机制，导致物业管理者对物业维修工作人员缺乏有效的监督考核。所以，规范有效的社区工作管理机制和信息化工具是必须的。

关于投诉建议

业主和物业之间缺乏信任和沟通，是包括笔者在内的很多居民切身感受到的。那么大家的社区都具备何种意见反馈渠道呢？54%的居民使用投诉电话或投诉箱；27%的居民使用了社交聊天工具；仅有1%使用了社区应用；另有相当比例使用了其他方式。12%的居民表示目前的意见反馈渠道无效；67%的居民表示效果一般；仅有11%的居民表示效果很好。

投诉建议是业主和物业之间的敏感行为，投诉箱的方式可实现匿名投诉，社交工具和电话都不能保证匿名投诉，可能无法真实有效的反映情况。但是，通过软件实现匿名投诉是非常简单的，居民也不用特地跑到投诉箱前，一键提交即可。

关于沟通交流

我们所在的社区居民都很多，您是不是经常希望社区居民们就某件事发表意见，希望进行问卷调查，依靠集体的智慧和力量解决问题？78%的居民都表示需要，13%表示无所谓，仅有9%表示不需要。同样与沟通交流有关，63%的居民表示希望有自己社区的网络邻里圈，分享生活中的趣事等点点滴滴，10%表示不需要，其余27%表示无所谓。大家又是否希望社区能多组织社区活动，增进邻里交流？62%的居民表示希望，13%表示不需要，其余25%表示无所谓。

所以，现实情况中，居民还是存在很大的相互之间的沟通交流需求的，社区本该是一个其乐融融的大家庭，大家互帮互助，团结友爱。

关于物业缴费和政务便民

让数据多跑路，让居民少跑腿是智慧社区和人性化服务的口号，关于小区或者社区是否可以足不出户，完成水、电、气、暖、物业等各种生活缴费？52%的居民表示部分可以，11%的居民表示全部不可以，只有37%的居民表示全部可以。同样是手机办事，如果一部手机在手，就能完成政务办事大厅现场的各种办事流程，86%的居民表示会经常点击使用，14%居民表示不会使用。

疫情发生以后，关于各种物业缴费，无法线上完成的社区突显了管理弊端，在尽量减少接触的号召下，很多居民还要到物业办公室，手递水卡、电卡，扫码支付，增加了病毒传播风险。笔者所在的小区就存在这种情况。另外，相关政府部门前期也开发了各种政务应用，帮助群众线上办理业务，大大方便了隔离在家的社区居民。如果有一款包含了物业缴费和政务便民功能的社区应用，会为居民生活的提供很大的便利，疫情下也能保证身体健康安全。

关于社区周围商圈

疫情来临以后，不仅企业延期复工，个体商铺也纷纷停业，社区更是封闭管理，禁止随意出入。为了给社区居民提供生活必需品，社区周边超市与物业密切合作，建立起了物品安全输送渠道。但是，笔者所在小区的居民，都是通过社交软件群组，领导下单，中间夹杂着新进群组居民的各种提问，漏单错单时有发生。72%的居民都希望社区周边的店铺商圈出现在社区应用中，提供下单或订餐送餐便利。

将社区周围商铺加入社区应用，借助软件严密的下单、出单、支付逻辑，定能充分解决社区居民生活必需品等的采购需求。

关于智能硬件

很多新建社区中都会有监控、道闸、门禁等智能硬件，保障居民的日常安全。关于监控，有86%的居民希望能看到社区公共区域监控画面，提升居住安全感；有10%表示无所谓；仅有4%表示不希望。

最后，社区居民们是否希望有一款面向社区居民生活的应用投入使用，更好的实现信息的上传和下达？92%的居民都表示希望。尤其是在这个特殊时期，社区的智能化和信息化管理显得尤为重要。山东智谷数据分析师事务所的母公司，山东众志电子有限公司，为解决广大社区居民、物业和政府管理的迫切需求，快速研发了《泰山e情综合云管系统》，解决了社区管理的痛点难点，系统上线以来得到了山东省大数据局的高度重视，在省大数据局《关于发布疫情防控有关互联网平台产品清单的公告》460项产品中，列在第66位。

公司还将继续推出更加功能全面的智慧社区应用，欢迎详细咨询合作相关事宜。



联系方式:

公司：山东智谷数据分析师事务所
地址：泰安市高新区南天门大街1110泰山科技创业城A区
公司电话：0538-8932988
公司邮箱：data@zdz.com.cn
联系人：苏航



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