

运营商大数据助力复工复产的探索与应用

Exploration and Application of Big Data for Helping Production Resume

朱常波^{1,2,3},程新洲³,高 洁³(1. 中国联通智能城市研究院,北京 100048;2. 中讯邮电咨询设计院有限公司,北京 100048;3. 中国联通网络技术研究院,北京 100048)

Zhu Changbo^{1,2,3}, Cheng Xinzhou³, Gao Jie³(1. China Unicom Smart City Research Institute, Beijing 100048, China; 2. China Information Technology Designing & Consulting Institute Co., Ltd., Beijing 100048, China; 3. China Unicom Network Technology Research Institute, Beijing 100048, China)

摘 要:

随着全国疫情防控工作的有序开展,国内新型冠状病毒的疫情已经进入平稳阶段,然而,企业复工复产对保持经济社会稳定发展至关重要,迫在眉睫。因此,如何统筹疫情防控和复工复产,成为摆在各级政府面前的又一大挑战。国务院联防联控机制发文要求各地“分区分级精准防控,统筹疫情防控与经济社会秩序恢复”。在此背景下,中国联通充分发挥运营商大数据特有的多源、海量、融合的优势,从区域到岗复工率实时洞察、栅格化危险指数评估、区域人口流入风险分析、员工出行方式风险分析等4个方面,助力企业复工复产,为企业决策者制定科学政策提供支撑依据,逐步实现全面到岗复工。

Abstract:

With the national epidemic prevention and control work carried out in an orderly manner, the COVID-19 epidemic in China has entered a stable stage. However, the resumption of work and production by enterprises is crucial to maintaining the stable development of the economy and society. Therefore, how to co-ordinate epidemic prevention and control and resume production has become a major challenge for governments at all levels. The joint prevention and Control Mechanism of the State Council issued a document requesting all localities to "carry out precise prevention and control at different levels, coordinate the prevention and control of epidemic situations and restore economic and social order". In this context, China Unicom gives full play to the operator's unique advantages of multi-source, massive and integrated big data, from the four aspects of the real-time insight into the regional return to work rate, the evaluation of Rasterized Risk Index, the risk analysis of the regional population inflow and the risk analysis of the employee's travel mode, it helps the enterprises to return to work and provide the support basis for the decision-makers of enterprises, formulate scientific policies and means to gradually realize the full resumption of work.

Keywords:

Resume production; COVID-19; Telecom operator; Big data

引用格式:朱常波,程新洲,高洁. 运营商大数据助力复工复产的探索与应用[J]. 邮电设计技术,2020(5):1-6.

1 概述

在这场举全国之力抗击疫情的“战役”中,各级政府通过暂缓企业开工、学校开课等政策手段,有效控

制了疫情的蔓延和发展,当前已初步呈现疫情防控形势持续向好、生产生活秩序加快恢复的态势。然而,企业复工复产对保持经济社会稳定发展至关重要,迫在眉睫。因此,如何统筹疫情防控和复工复产,成为摆在各级政府面前的又一大挑战。2020年2月11日,国务院应对新型冠状病毒感染肺炎疫情联防联控机制举行“切实加强疫情科学防控,有序做好企业复工复产”工作发布会。要求各地区区分不同疫情状况,在继续加强

基金项目:工业和信息化部大数据产业发展试点示范项目(融合异构数据及深度学习的民生大数据创新应用试点示范)

收稿日期:2020-04-16

关键词:

复工复产;新型冠状病毒;运营商;大数据

doi: 10.12045/j.issn.1007-3043.2020.05.001

文章编号:1007-3043(2020)05-0001-06

中图分类号:TP391

文献标识码:A

开放科学(资源服务)标识码(OSID):



疫情科学防控的同时,及时协调解决企业的困难和问题,有序推动企业复工复产,尽早恢复正常生产。

中国联通作为拥有4亿用户的电信运营商,充分发挥全国31省一点集中的数据优势,面对疫情,快速响应,在保证用户隐私安全的基础上研究开发了一系列大数据疫情防控产品并面向全国部署,为企业复工复产洞察、城市人口动态透视等应用场景提供了数据支撑,确保疫情防控、复工复产的科学有效推进。

本文主要介绍了运营商大数据在疫情防控趋于稳定的情况下,企业复工复产、学校开学开学等场景的探索及应用。文章组织结构如下:第2章对运营商大数据的特征与分析架构进行了介绍;第3章阐述了运营商大数据在复工复产背景下的区域到岗率实时洞察与栅格化危险指数评估关注点、区域人口流入与员工出行方式风险点等4个方面的数据分析及研判;第4章对全文进行了总结。

2 运营商大数据技术架构

运营商大数据相较于其他行业,有着无法比拟的诸多优势。

首先,运营商拥有十亿级的出账用户数、PB级日均数据生成及采集量,数据维度全面包含了用户信息、地理位置、上网行为、历史轨迹、社交关系、支付能力等多元内容,可准确实时地反映用户全息行为状态。

另外,运营商拥有多张移动网络,全生命周期时间连续、空间完整,在覆盖时间、覆盖范围上遥遥领先于

其他行业。运营商大数据可以通过电话号码标识来整合打通各类数据,与GPS、统计数据等相比,可直接追溯到自然人。通过运营商大数据可以获取用户身份信息、时间信息、位置信息、终端信息、行为信息和感知体验,形成每个用户的动态标签库,从行为轨迹、业务偏好、时空分布、社交属性等维度刻画人们的日常工作与生活。

中国联通在大数据技术及场景化应用产品研发方面有着多年的技术积累与沉淀,在产品迭代更新上线时,更多的是基于原有的用户标签库、指纹数据库以及场景化分析模型的抽取和组合,快速形成针对不同场景应用的整体技术解决方案,并可以在很短的周期内完成相关产品的模块式开发及上线应用。运营商大数据在助力复工复产方面的应用模型如图1所示。

整体的技术方案,可以分为数据中台层和场景化应用层。

针对数据中台层的技术方案,主要包含:

a) 数据处理:运营商数据包含23类用户侧及网络侧数据,数据类型包括关系型数据、NoSQL数据等。同时,互联网及垂直行业包含大量的异构数据。数据存储层根据各类数据的采集方式、数据特点制定了一套规范化的数据管理存储体系,实现对以往分散数据的归集与规范化管理,存储能力达到PB级。

b) 模型搭建:针对数据存储层所存储的大量关系型数据、NoSQL数据、异构数据,为了后续进行准确、高效的数据分析,根据各个数据源的不同,研究数据解析

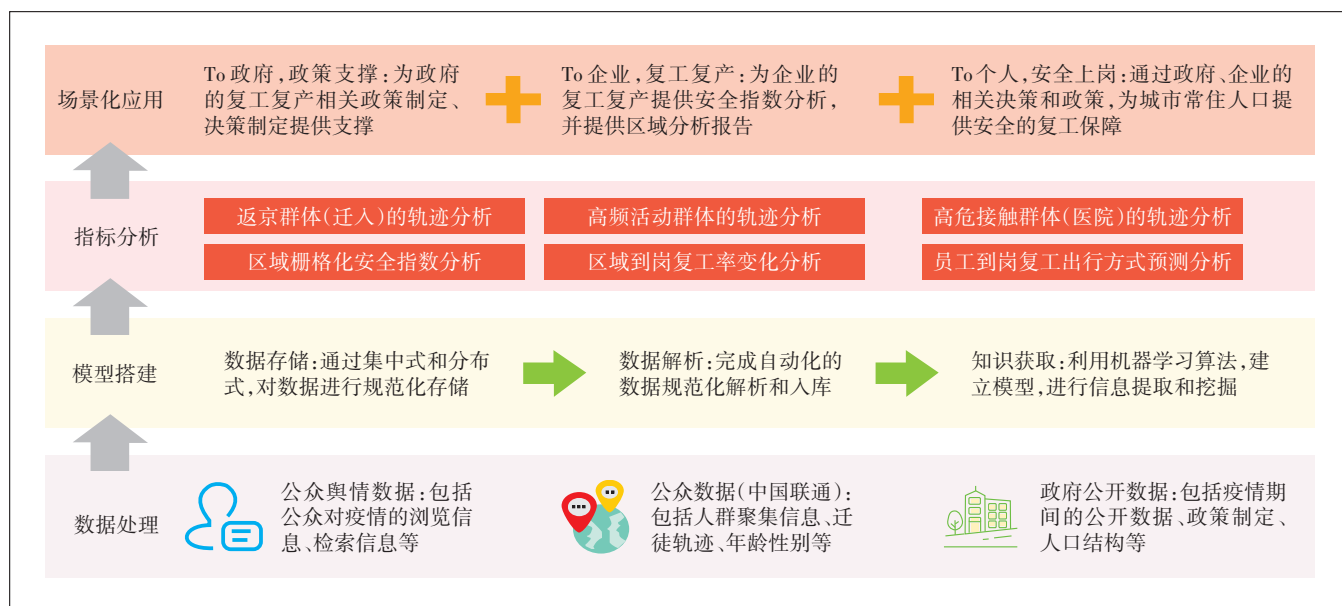


图1 大数据助力复工复产分析的技术方案

与入库方法,实现高效且精准的数据解析,并搭建支撑上层指标分析的机器学习模型。

c) 指标分析:通过搭建 Impala 集群、Vertica 集群,形成用户和业务为主体的6个基础模块分析。在基础模块之上,数据分析层将进行关联分析,构建用户属性标签库、用户位置档案库,形成用户、位置、驻留时长等分析指标的数据中间层。同时,数据分析层还将引入分布式计算、商业智能BI等技术,提升该平台的性能及功能。

场景化应用层主要包括场景化应用报告及平台服务。在指标分析层以上,搭建地图服务器、GIS 服务器、Web 服务器,进而实现数据分析层的用户、位置以及栅格化分析指标的实时查询及地理化呈现,以及针对不同应用场景生成相应的分析报告。

3 运营商大数据助力复工复产

首先,通过政府公开的媒体网站调研了北京市疫情防控的实时进展情况,截至2020年2月底,北京市疫情走势如图2所示。单日新增疑似、新增确诊的病例数均成下降趋势,且趋近于0。

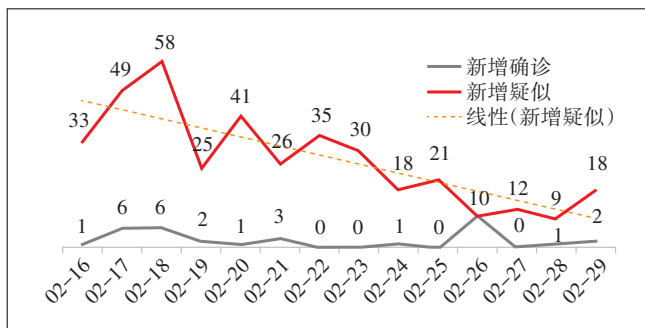


图2 北京单日新增病例数走势

同时,通过分析用户浏览微博、新闻等APP的业务使用行为,结合互联网公开数据接口,获取用户对疫情的关注程度。基于自然语言处理算法对用户微博、新闻留言等多元文本进行情感研究与分析,可进一步判定用户情绪指数(积极/消极,喜悦/愤怒/恐惧等),关注人群舆论导向。如图3和图4所示,可见公众对疫情的关注度在2020年2月逐步下降,同时公众情绪正在缓慢由消极向积极进行转变。

从数据分析结果看,截至2020年2月底的新冠疫情从医学防控和心理防控2方面整体处于积极向好的稳定局面,这主要得益于常驻人口的自我隔离和迁入人口的有效控制。

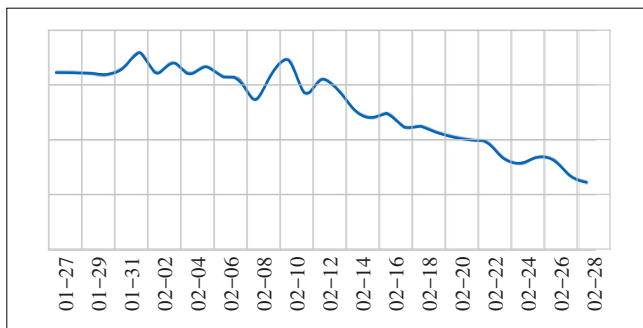


图3 公众疫情关注度变化趋势

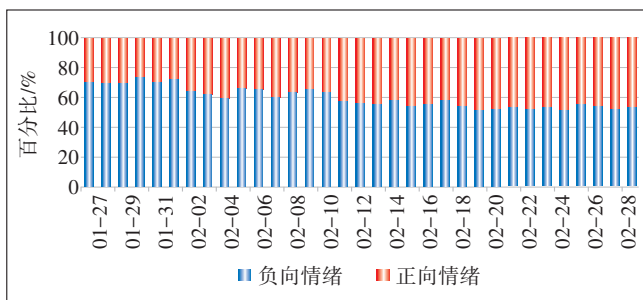


图4 公众心理情绪变化趋势

综上所述,针对大数据助力复工复产的分析应用场景,应重点关注复工复产带来的城市迁入人口变化以及常住人口的隔离变化2个方面。因此,本文从区域到岗复工率分析和区域栅格化安全指数2个关注点,以及人口迁入分析和员工出行方式预测2个风险点等4个方面进行分析及研判。

3.1 关注点1:区域到岗复工率分析

运营商数据具备分析用户群体时空信息的能力,通过位置信息、时间戳,可以获取用户的时空轨迹,从而掌握用户出行特征。通过对信令数据的实时分析和用户属性更新,可以逐步形成稳定的用户标签库数据。通过对某特定区域内用户属性分析以及用户数量变化分析,可以实时关注特定区域内的在岗员工到岗复工的变化情况,为企业决策者在制定复工复产相关政策、措施时提供数据分析依据。

通过用户标签库进行数据筛选分析,主语国际周边区域(2 km×2 km)约有8 000名在岗员工。通过实时样本数据观测,针对主语国际周边区域进行详细分析,如图5所示,从2020年2月25日—3月1日,周边区域到岗复工比例整体在40%左右,从2020年2月25日—2月28日(周二—周五)单日到岗复工人数从3 000人左右增长至3 700人左右,且持续呈上升趋势。随着国务院相关政策的发布,各企事业单位陆续进入逐步到岗复工的阶段。

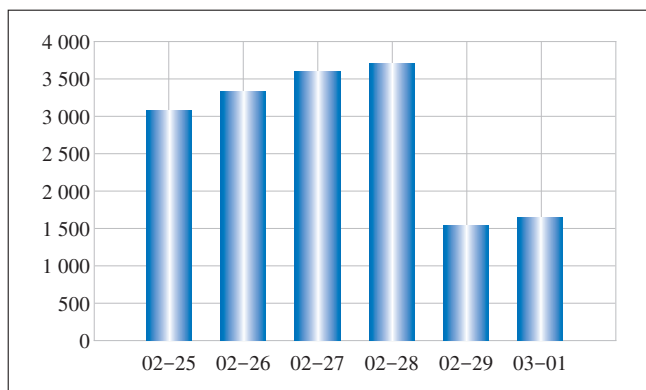


图5 区域到岗复工人数变化趋势

对周边区域人数比较集中的企业、商业楼宇进行个体分析,如图6所示,公安部第1研究所、国兴大厦、机械研究院等单位的到岗复工比例较高,分别为41.3%、32.2%、31.8%;从地图上看,如图7所示,其中红点表示人员相对集中的区域,包括公安部及机械研究院区域、百万庄区域、中国标协和云建大厦区域等。

3.2 关注点2:区域栅格化安全指数分析

基于运营商大数据,考虑人口密度、职业群体分布、自我隔离水平、用户出行计划和意图等因素,综合评估城市中各个栅格的危险指数及防疫安全等级(见图8),并通过可视化平台观测风险预警,可将疫情风险度量到栅格单位内,为复工复产、精准施策的防疫支撑工作提供参考依据。

从图8可以看出,主语周边区域平均安全指数3.75,超越市区54.31%的区域,处于中等水平。

同时,在SEIR传染病模型中,新增感染人数和潜伏者、传染者、感染者接触到的易感者人数、传染概率、潜伏者转化为感染者概率等因素正相关。因此,高频

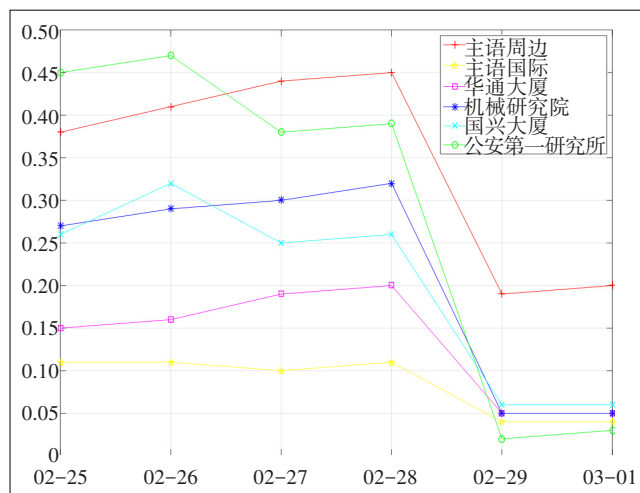


图6 区域到岗复工率分析

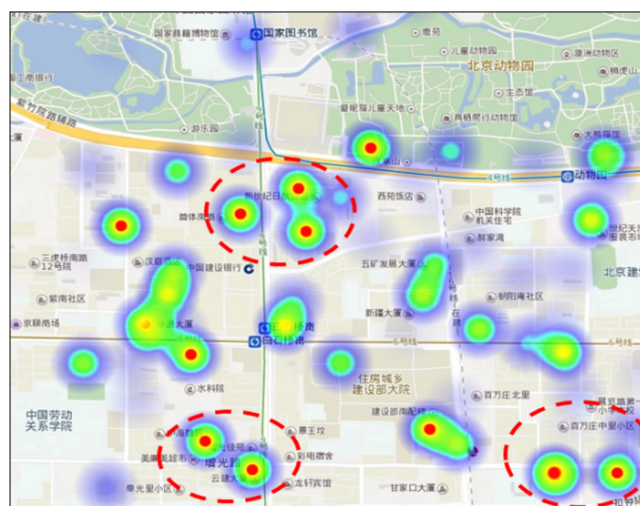


图7 区域到岗复工率分析

活动群体成为潜伏者的概率较大,而监测管理该群体,可以有效降低感染者接触到的易感者人数,对于病毒

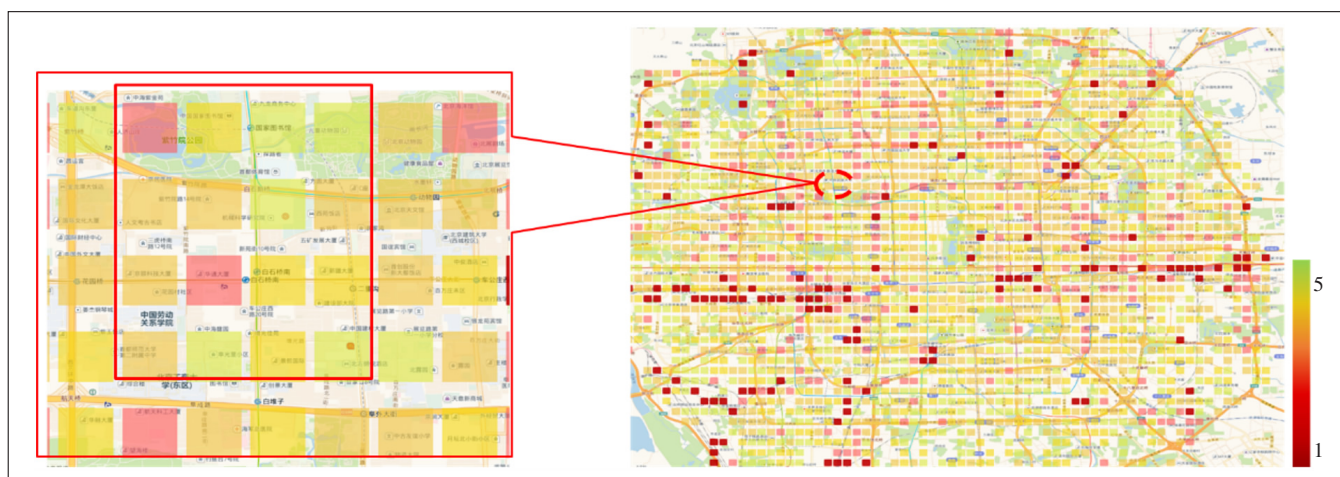


图8 栅格安全指数分布

防控有积极作用。

从数据分析结果看,主语周边区域,频繁进出大中型医院的群体人员,比例较低,基本稳定在0.3%左右,平均每天为600人。

3.3 风险点1:城市人口迁入风险分析

受疫情影响,全国大多数城市都推迟了复工日期,因此原本应该在农历初六、七出现的“回流”高峰得到了抑制,整个节后春运期间并未出现明显的集中返程高峰,交通表现相对比较平稳。对比北京市2019年和2020年同期人口迁出情况,如图9所示,截至2020年1月25日前,2020年人口迁出指数与同期相比并无较大差异,但从2020年人口迁入情况看,与同期形成较大的对比差异。截至2020年2月底,返京人员总量仍处

于较低的水平,统计来看,有约400万人员尚未返京。如图10所示,从2020年2月15日—2月28日的人口迁入变化分析,单日返京人口数量呈上升趋势,因此,随着逐步复工复产的趋势,北京仍面临大量人口流入压力。

3.4 风险点2:员工出行方式预测分析

全面到岗复工复产过程中,企业员工的出行方式,同样也是需要关注的重点之一,尤其是乘坐公共交通的群体。通过对用户属性标签库进行筛选分析,中讯院、网研院北京在岗员工约为1100人,通过与近期的实时数据关联出800人左右在京远程办公,主要居住在海淀区和丰台区(见图11)。

通过对员工居住地与主语国际直线距离的分析

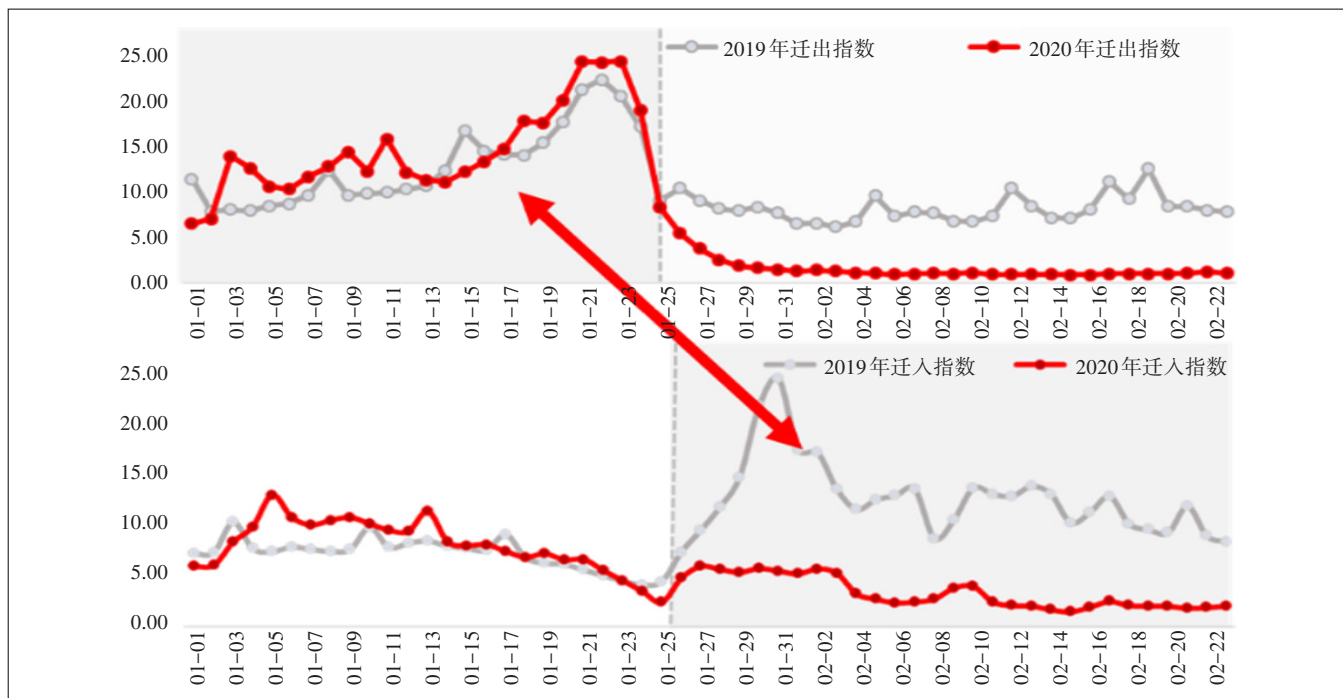


图9 北京市人口迁入迁出对比

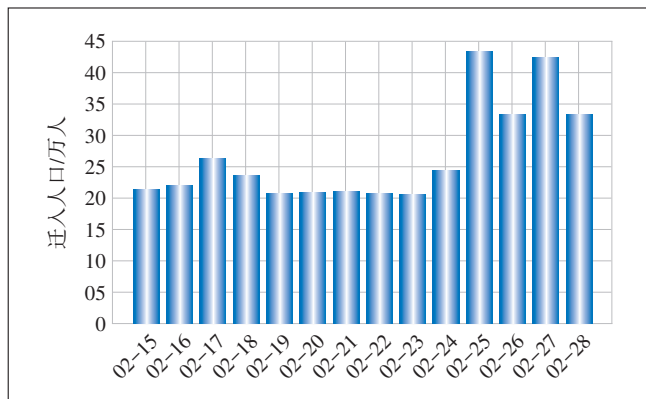


图10 北京市人口迁入实时变化趋势

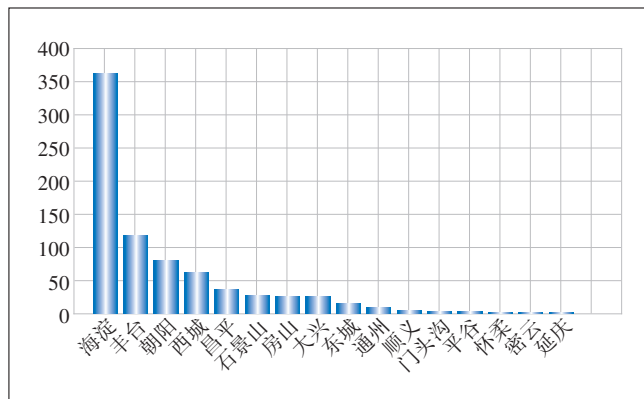


图11 北京地区居住行政区分布

(见图12),发现仅有11%员工居住在离主语2 km范围内,这部分员工可采用步行或单车方式出行上班。如图13所示,其余约有43%员工出行方式以地铁为主,约有46%员工出行方式为路面交通(如网约车、公交车、自驾),均存在一定接触风险。

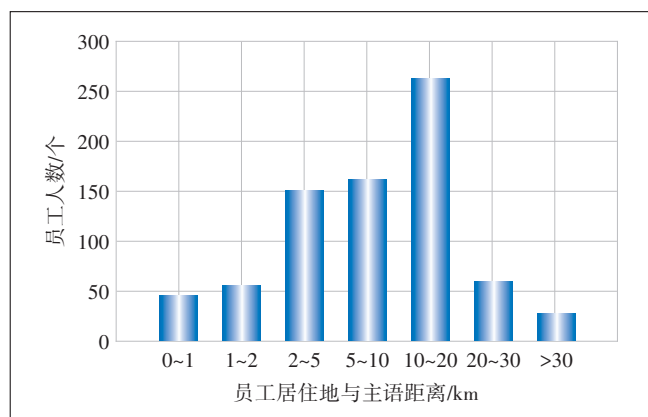


图12 员工居住地与主语距离分布

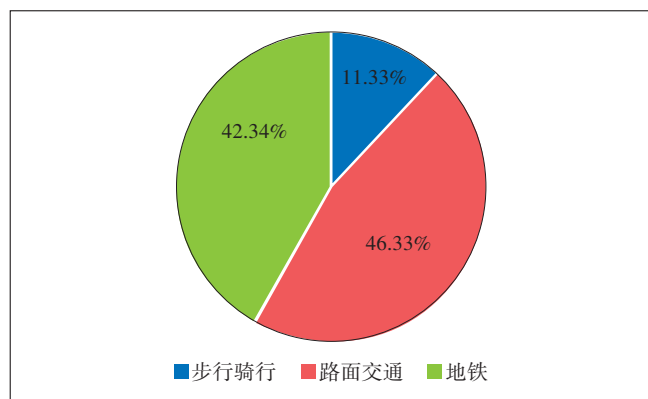


图13 员工出行方式预测

考虑到接触风险,北京公交集团为复工企事业单位推出定制化公交通勤班车的政策,企业可以参考员工住址标签数据,生成建议定制路线,为员工解决上下班的通勤风险问题。

4 总结

本文从区域到岗复工率实时洞察、栅格化危险指数评估、区域人口流入风险分析、员工出行方式风险分析等4个方面的数据分析及研判,实现基于运营商大数据的复工复产分析及应用。主要包括:

a) 区域到岗复工率分析,关注特定区域内的在岗员工到岗复工的变化情况,为企业决策者在制定复工复产相关政策、措施时提供数据分析依据。

b) 针对区域栅格化安全指数分析,并针对不同栅格评估其危险指数,为员工出行、复工复产提供参考依据。

c) 对城市人口迁入风险分析,对节后复工背景下可能产生的城市疫情传播风险进行评估。

d) 员工出行方式预测分析,预测分析需要乘坐公共交通的员工群体,为复工复产企业解决员工上下班的接触风险问题。

通过上述分析结果,可在疫情态势研判等方面为政府提供有力支撑,助力企业复工复产,为企业决策者制定科学政策及手段提供支撑依据,逐步实现全面到岗复工。

参考文献:

- [1] 新华社. 国务院联防联控机制印发《关于科学防治精准施策分区分级做好新冠肺炎疫情防控工作的指导意见》[EB/OL]. [2020-03-25]. http://www.xinhuanet.com/politics/2020-02/18/c_1125592416.htm.
- [2] 知乎. 在家宅着也能抵抗肺炎! 玩一玩SEIR传染病模型[EB/OL]. [2020-03-25]. <https://zhuanlan.zhihu.com/p/104268573?scene=1>.
- [3] 朱常波,程新洲,叶海纳. 5G+大数据赋能智慧城市[J]. 邮电设计技术,2019(9):1-4.
- [4] 郭中梅,朱常波,夏俊杰. “疫情大考”下的智慧城市未来发展思考[J]. 邮电设计技术,2020(2):5-8.
- [5] 程新洲,朱常波,晁昆. 掘金大数据-电信数据金矿详解、挖掘及应用[M]. 北京:机械工业出版社,2019:34-40.
- [6] 高洁,张涛. 一种基于LightGBM机器学习算法的用户年龄及性别预测方法[J]. 邮电设计技术,2019(9):36-39.
- [7] 韩玉辉,高洁,成晨. 电信大数据在交通行业的应用探讨[J]. 邮电设计技术,2018(10):15-19.
- [8] 吴钱锋. 疫情防控中大数据应用的浙江经验与建议[J]. 浙江经济,2020(3):2.
- [9] 俞柏锋. 科学战“疫”,看大数据如何发挥作用[J]. 宁波通讯,2020(5):38-39.
- [10] 常雯,祁昌涵. 基础电信企业大数据助力复工复产[N]. 人民邮电报,2020-02-25(4).
- [11] 董润莎,徐争莉,程新洲. 基于机器学习的用户离网预测研究[J]. 邮电设计技术,2018(10):23-26.

作者简介:

朱常波,毕业于华中科技大学,中国联通智能城市研究院院长、中讯邮电咨询设计院/中国联通网络技术研究院副总经理,正高级经济师,博士,荣获多项省部级科技进步奖、管理创新奖、优秀设计奖;程新洲,毕业于北京邮电大学,教授级高工,硕士,长期从事移动通信、大数据研究与应用工作;高洁,毕业于北京邮电大学,高级工程师,硕士,长期从事运营商大数据在各场景化应用的研究工作。