

Residential and Travel Characteristics Analysis Based on Mobile Phone Signaling Data

—A Case in Shenzhen City

Jiandong Qiu¹, Qingya Lin^{1*}, Qiang Li²

¹Shenzhen Urban Transport Planning Center Co., Ltd., Shenzhen Guangdong

²Traffic Information Engineering & Technology Research Center of Guangdong Province, Shenzhen Guangdong
Email: qiujd@sutpc.com, *1160193194@qq.com

Received: Aug. 20th, 2018; accepted: Sep. 5th, 2018; published: Sep. 12th, 2018

Abstract

There is no obvious tendency for the data of mobile phone signaling data to be objective, comprehensive and sampling. The data has strong spatiotemporal continuity and can be observed in the whole process of traffic travel, which is unmatched by any other data source. Using mobile phone signaling data to analyze the characteristics of urban traffic operation makes up for long cycle of the traditional traffic survey, heavy workload, small sample size and high cost. This paper studies the use of space-time information of mobile signaling data to visualize and analyze the characteristics of urban traffic travel. The analysis shows that there is still a strong job-commuting attraction in the original SAR and the SEZ; most people live in Dongguan, Huizhou, working in Shenzhen. Living and employment are basically at the junction of the city. Early peak of travelling shows a clear east-west, north-south formation. Analyzing the results provides a reliable basis for transport planning and operations.

Keywords

Cell Phone Signaling Data, Data Quality Inspection, Big Data Analysis, Visualize

基于手机信令数据的居住和出行特征分析

——以深圳市为例

丘建栋¹, 林青雅^{1*}, 李 强²

*通讯作者。

¹深圳市城市交通规划设计研究中心有限公司, 广东 深圳

²广东省交通信息工程技术研究中心, 广东 深圳

Email: qiujd@sutpc.com, 1160193194@qq.com

收稿日期: 2018年8月20日; 录用日期: 2018年9月5日; 发布日期: 2018年9月12日

摘要

手机信令数据样本量大、数据客观、全面、采样不会有很明显的倾向性,且数据具有较强的时空持续性,可以观测到交通出行整个过程,是任何其它数据源无法比拟的。用手机信令数据分析城市交通运行特征,弥补了传统的交通调查周期性长、工作量大、样本量少和花费高的特点。本文研究利用手机信令数据的时空信息,对城市交通出行特征进行可视化和科学分析,分析结果显示,原特区内与特区外仍有较强职住通勤吸引;住在东莞、惠州,职在深圳的人,居住和就业位置基本在城市间交界地带;早高峰出行形成明显东西向、南北向通道。分析所得结果可为交通规划和运营部门提供可靠的依据。

关键词

手机信令数据, 数据质量检查, 大数据分析应用, 可视化

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

目前,中国城市化进程仍处在快速演变阶段,城市建设步伐加快,城市用地规模不断加大,城市人口不断攀升,城市居民的时空出行范围进一步扩展,所面临的交通问题越来越严重[1]。作为城市居民生活与生产的两大载体,“居住”与“就业”是城市空间结构中的两个核心内生变量,它们彼此依存、互相影响,两者是否均衡发展是影响城市居民生活幸福指数的关键指标,制约城市可持续发展的重要因素[2]。

传统的数据获取方法局限性明显。作为新兴的研究方向——手机信令数据,一方面,手机用户群体数量和手机使用率的不断提高,保证了手机信令数据的样本数量和随机性特征[3];另一方面,无线网络覆盖区域的不断扩大,使得手机信令数据的时空信息具备时间连续性和空间广覆盖性等优点[4]。

本文基于手机数据,从多个层面对深圳市居民活动与空间环境间的复杂关系做更加精细与全面的描述与解析,重点关注城市居民的职住分布、城市空间的流动性与相互作用机制,以及特定场景、特定人群的空间活动特征,开展深莞惠区域城市群居民居住和就业空间关系及往来联系强度研究,对推动就业与居住均衡发展,把握城市发展规划,促进居住空间与产业空间拓展的联动与共赢显得非常必要。

2. 停留和出行地识别算法

2.1. 数据概述

数据来源于深圳市城市交通规划设计研究中心有限公司——未来交通实验室。一共有2017年11月12日到2017年11月25日(共14天)的数据,其中工作日有10天。数据空间范围为深圳、东莞和惠州这

三个城市的全域。本文主要分析深圳居民交通出行时空变化和深莞惠市域间的职住关系。手机信令数据包含的信息如下，见表 1 所示。

2.2. 人口类型划分

本次研究通过人口的停留日期、时间和位置总体上将人口划分为两大类型(见图 1)：常住人口(停留时间超过 7 小时/天的、天数大于等于 50%天)以及非常住人口。常住人口又分为常驻人员和固定设备，其中定位位置一直不变的识别为固定设备；非常住人口中所有出现的日期里，停留时间均少于 3 个小时的识别为过境人员，其他为流动人员。

2.3. 居住地和_work地识别

居住地的判别依据为 0~8 点、19~24 点停留时间最长的基站为居住基站。9~18 点停留时间最长的基站，且在基站周围 500 米范围内停留时间大于 4 小时、天数大于等于 50%天的为工作地基站。

2.4. 出行识别

首先，按 500 米的距离对个体出行的轨迹点进行空间聚合，将停留时间超过 30 分钟的点作为一次停留活动点，两个相邻活动点为一次出行，一天中的第一个活动点和最后一个活动点也作为一次出行，这样可得到用户一天的出行链轨迹及出行次数和出行时间。

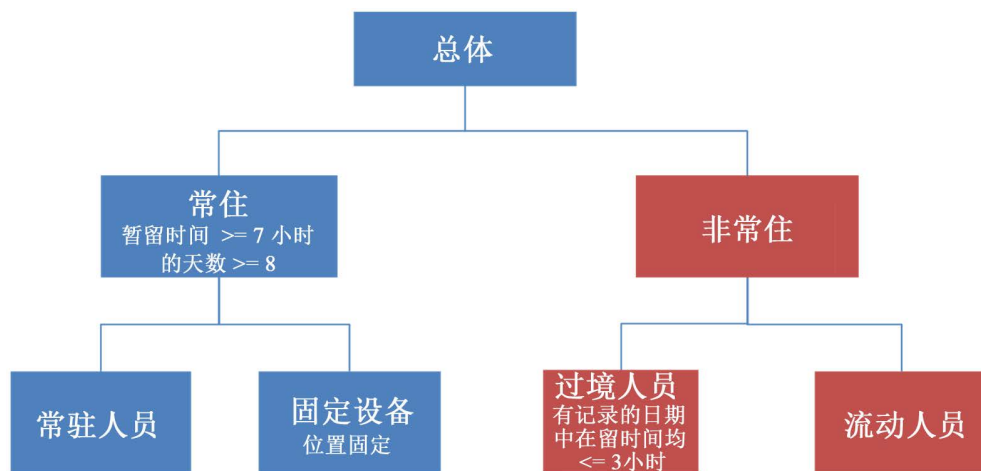


Figure 1. User type division
图 1. 人口类别划分

Table 1. Description of Mobile phone signaling data field
表 1. 手机信令数据字段说明

| 序号 | 字段名称 | 说明 |
|----|-------|----------------------------------|
| 1 | 用户 ID | F14BE1E7E64922F873B889ED15FD900A |
| 2 | 时间 | 2016-09-11 08:12:24 |
| 3 | 基站大区 | 9904 |
| 4 | 基站小区 | 76,060,311 |
| 5 | 经度 | 114.6627 |
| 6 | 纬度 | 23.6428 |

2.5. 数据分析流程

本文通过对比每天的手机信令数量，分析数据的接收质量情况，选取接收数据较稳定的日期用于后续的数据处理。数据分析流程见图 2，数据处理过程中剔除有杂质的数据，将 Python 语言编译的算法部署于 Spark 平台上，通过人口类别的划分、居住地、工作地和出行识别，计算得到相应的中间表，中间表存放于 postgis 数据库中，用于后续应用层做分析时调用计算，应用层通过 Arcgis 和 Transcad 将计算结果进行可视化。

3. 数据质量检查

现今，手机信令数据已经被广泛应用于城市的交通特征出行分析，但是，不同的人清洗数据的算法、阈值设置都不一样。为了保证算法的可靠性，需要对清洗的数据进行数据质量检验。主要验证手机基站手机数据是否缺失，居住地和在工作地识别、出行识别算法是否正确。大致的检验过程如下：用手机数据的人口数与已有的数据进行比较，判别居住地和在工作地识别算法的结果是否可靠；计算每小时的出行数量，从而找出数据收集比较稳定的日期作为后面出行特征分析；计算手机数据的早高峰 OD，用于与宏观模型的早高峰出行 OD 对比，检查出行识别算法的正确性。

3.1. 人口分布比较

考虑到手机数据中的人口不是全样数据，这里用各区人口数量占总人口比重进行比较。比较结果显示(图 3)，不管是与宏观模型数据还是统计年鉴数据比较，误差最大不超过 5%，误差最大的行政区为福田，可能原因是福田、罗湖区原本发展程度已达到饱和，且深圳原关外地区近年来建设用地不断扩大、交通路网不断完善、居民出行更加方便、配套设施逐渐齐全、居民的择居选择受就业岗位地理位置的制约力下降，人们很多会选择居住在原关外。

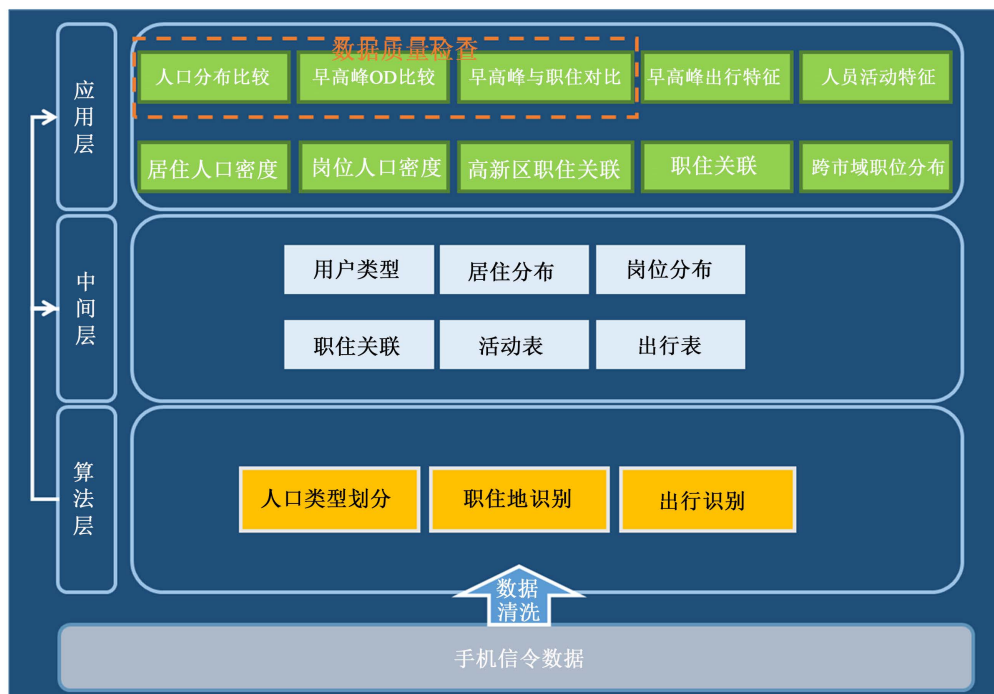


Figure 2. Process of data analysis
图 2. 数据分析流程

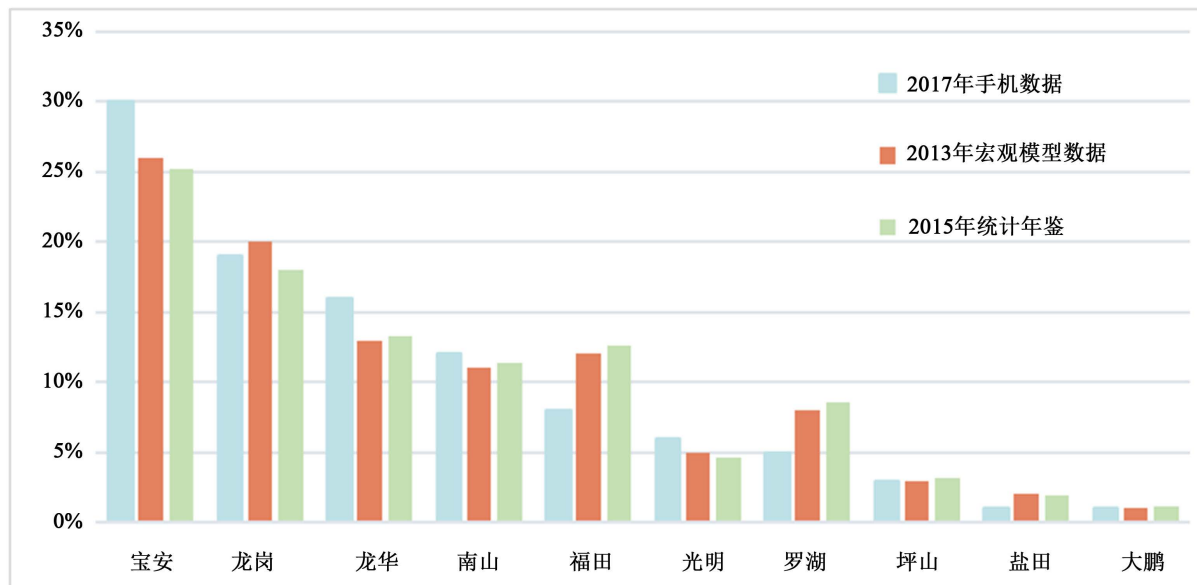


Figure 3. Population distribution of mobile data compare with macro-models and statistical yearbook
图 3. 手机数据与宏观模型和统计年鉴人口分布比较

3.2. 早高峰 OD 比较

与宏观模型(2014 年)比较结果如图 4 所示, 最大流量通道基本相似, 特别是南部几个大区之间的关联度, 连接宝安、南山、福田和罗湖的通道联系强度基本一致; 龙华到南山和福田、龙岗到罗湖和龙华的联系关系也一样紧密。

3.3. 早高峰 OD 与职住对比

职住期望线和早高峰出行对比见表 2 和图 5, 两者的行政区之间的联系非常相似, 宝安 - 南山的关联度最强, 然后是龙华 - 龙岗的联系度次之。

根据数据质量检查结果得出, 区级别的职住和早高峰的 OD 基本符合实际情况。但可以从中选择出符合分析要求的日期用于本文研究。

4. 大数据分析应用

本次分析分职住分布和出行特征分析两大模块。居住和就业是人们比较关心的两大民生问题, 基于手机信令数据和基站位置信息, 从不同层面上分析居住地就业岗位分布、区域的职住关联、兴趣点客流来源去向分布、居民的时空出行分布、TAZ 间 OD 客流空间分布以及客流来源空间分布。通过手机信令数据的采集、处理和分析, 可以获得城市的人口分布的动态信息, 可以为交通规划提供详细、实时的人口分布信息, 并可以避免因人工调查带来的各种弊端, 如成本较高、抽样样本有限等。

4.1. 职住分布

4.1.1. 居住和就业密度分布

居住和就业密度分布见图 6 中区层面上人口主要集中在福田中心区、粤海、龙华; 深圳市行政区居住人口占比中, 宝安所占比例最高, 为 30%。

岗位密集区主要是福田区的南园、华强北、园岭, 罗湖区的东门、南湖、桂圆以及南山区的粤海。深圳市行政区岗位占比中, 宝安所占比例最高, 为 33%。

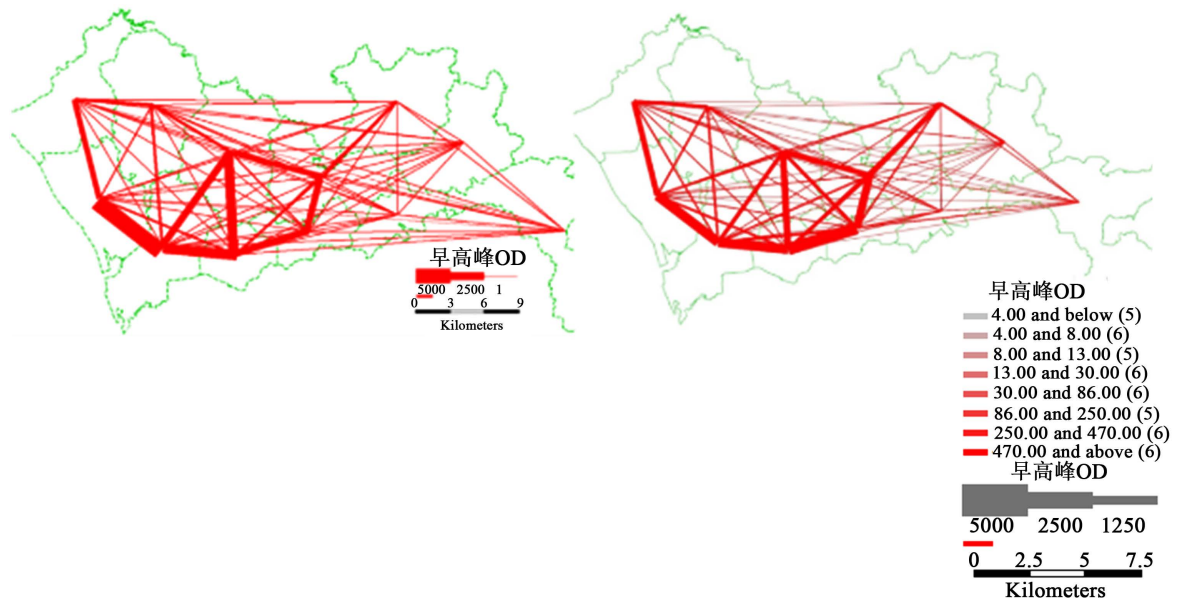


Figure 4. Comparison of mobile phone data with the early peak of macro model
图 4. 手机数据与宏观模型早高峰对比

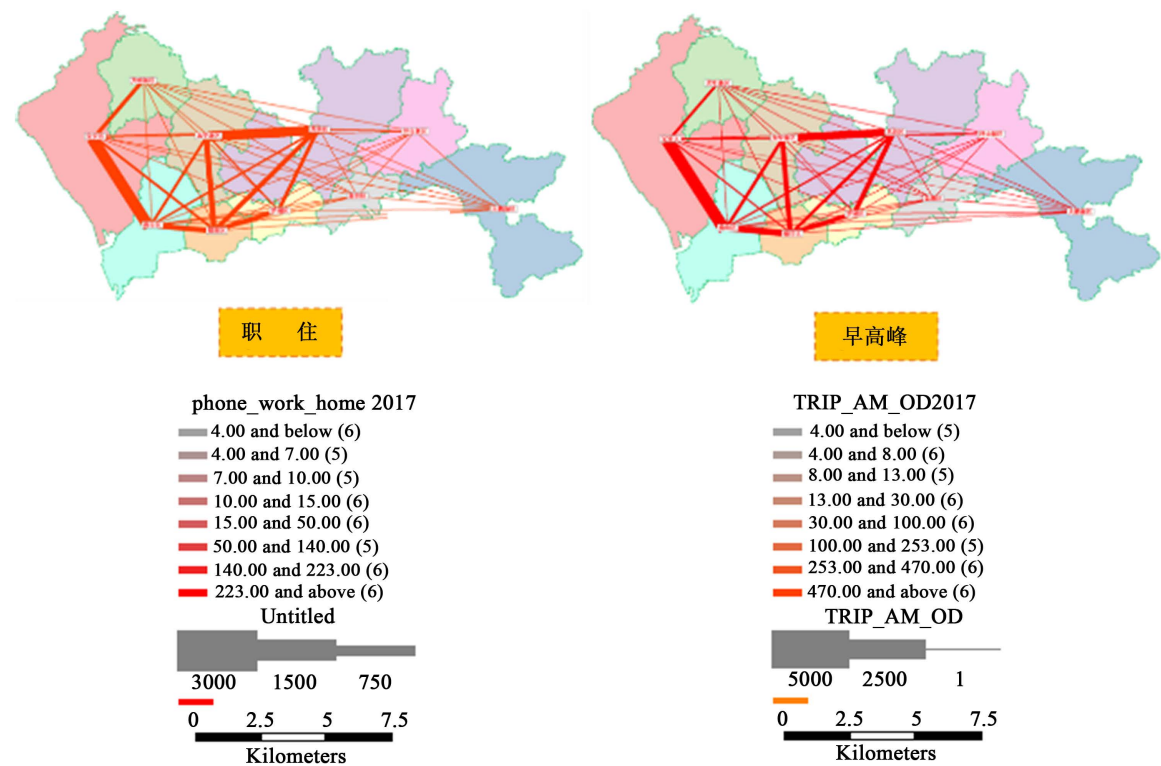


Figure 5. Morning peak OD (left) and accommodation (right) comparison
图 5. 早高峰 OD(左)和职住(右)对比

根据数据质量检查结果，手机数据的居住人口与统计年鉴数据对比，误差在可接受范围之内，所以行政区的人口比例数据是可靠的。从表 3 中可看出，宝安区无论是居住还是就业，人口占比都是最高，而大鹏新区占比最低。

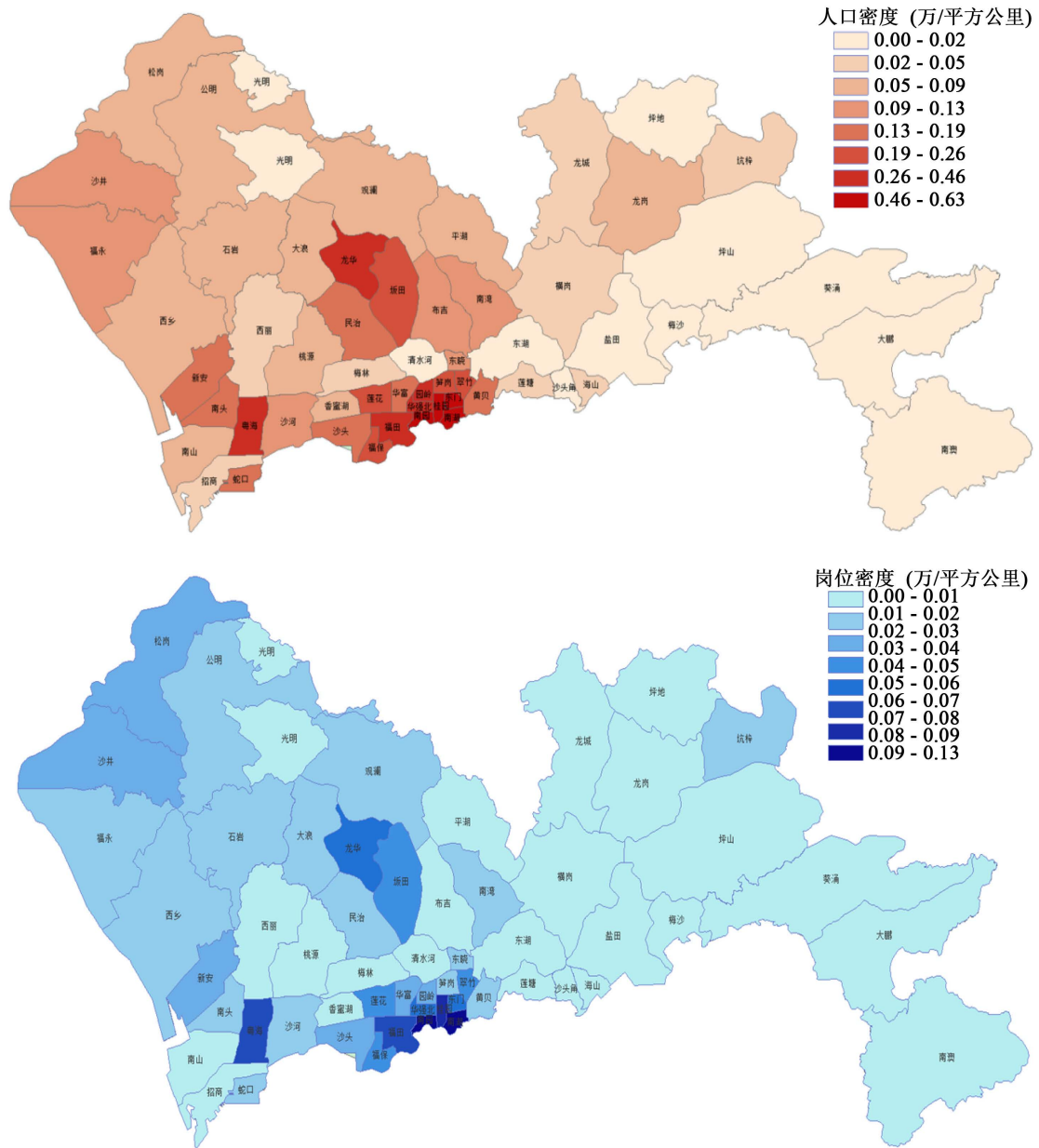


Figure 6. Residential (upper) and employment (lower) density distributions
图 6. 居住(上)和就业(下)密度分布

Table 2. The comparison between OD and accommodation in early peak of Nanshan district
表 2. 南山区早高峰 OD 与职住对比

| 行政区 | 工作地在南山区居住地比例 | 目的地在南山区出发地比例 |
|-----|--------------|--------------|
| 宝安 | 58% | 57% |
| 福田 | 14% | 17% |
| 龙华 | 13% | 14% |
| 龙岗 | 9% | 8% |
| 罗湖 | 4% | 3% |

Table 3. The proportion of administrative staff and residents
表 3. 行政区职住比例

| 行政区 | 居住人口比例 | 就业人口比例 |
|-----|--------|--------|
| 宝安 | 30% | 33% |
| 龙岗 | 19% | 17% |
| 龙华 | 16% | 15% |
| 南山 | 12% | 11% |
| 福田 | 8% | 8% |
| 光明 | 6% | 7% |
| 罗湖 | 5% | 4% |
| 坪山 | 3% | 3% |
| 盐田 | 1% | 1% |
| 大鹏 | 1% | 1% |

4.1.2. 职住关联

基于手机信令数据分析全市职住特性见图 7：宝安 - 南山，南山 - 福田，福田 - 罗湖，龙华 - 龙岗有大量的职住通勤，其中，南山 - 宝安、龙华 - 龙岗的职住关联关系最密切；原特区内与特区外仍有较强职住通勤吸引。在南山工作的人，有 84% 居住在南山，9% 居住在宝安。

4.1.3. 高新区职住关联

将 500 米栅格与基站关联，得到基站有栅格的关联关系，将就业基站与栅格 - 基站关联表匹配，按栅格 id 做人口汇总，求得每个栅格的就业人口，从而计算栅格等级的就业密度分布，选出 CBD 比较集中的区域，分析到主要 CBD 上班的人在哪里居住。本次分析选取高新园作为主要 CBD，分析结果见图 8 在高新园上班的人 72% 居住在南山，17% 居住在宝安，7% 居住在福田。

4.1.4. 跨市域职住分布

市域职住分布，将居住城市和就业城市分为两个图层，分别与基站空间关联，然后分别从职住表中提取职和住人口，同时展现职在深圳且住在东莞、惠州的这些人在这两地的职住分布；然后建立职住关联关系模型，可知道住在某个小区的人到另一小区上班占总通勤量的百分比。

深莞惠职住密度分布如图 9 所示，住在东莞、惠州，职在深圳的人，大多数居住和就业位置在城市间交界地带。东莞、惠州来深圳上班的人中，有 30% 的人住在东莞南部的长安镇，工作在宝安西北部的松岗镇；东莞、惠州来深圳上班的总人数中，5% 的人工作在南山、福田和罗湖。

4.2. 出行特征分析

早高峰出行特征

行政区之间早高峰出行特征分析结果如图 10 所示，宝安 - 南山之间出行量最大，龙华 - 龙岗次之；目的地在南山的出发地比例中，南山内部出行占 63%，宝安占 21%。

交通小区层面上，早高峰出行特征如图 11 所示，交通小区早高峰人口出行形成明显的东西向连接宝安、南山、福田、罗湖的通道和南北向连接龙华、龙岗、福田和罗湖的通道。

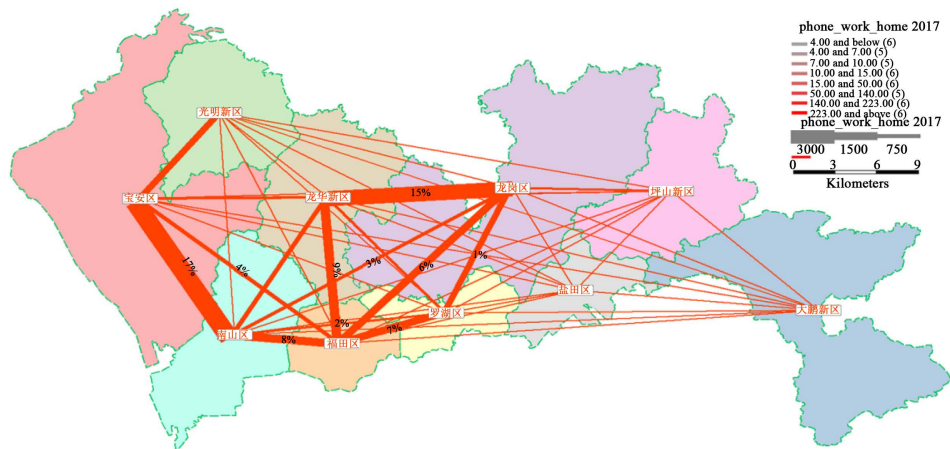


Figure 7. The relationship of district's Administrative and residential relations
图 7. 行政区职住关联

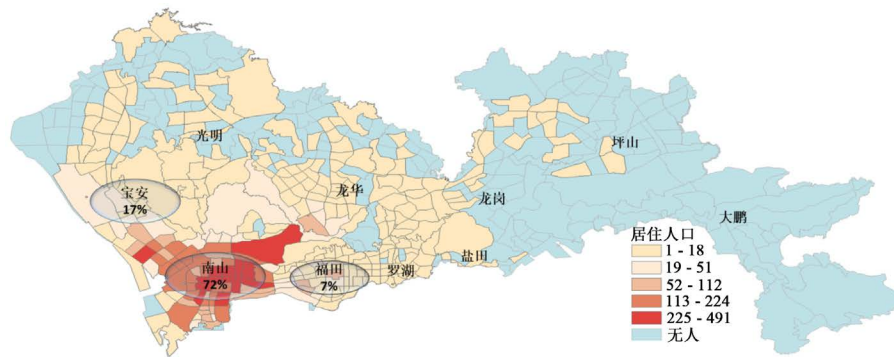


Figure 8. The proportion of people who work in Gaoxin Park
图 8. 到高新园上班的人居住地比例

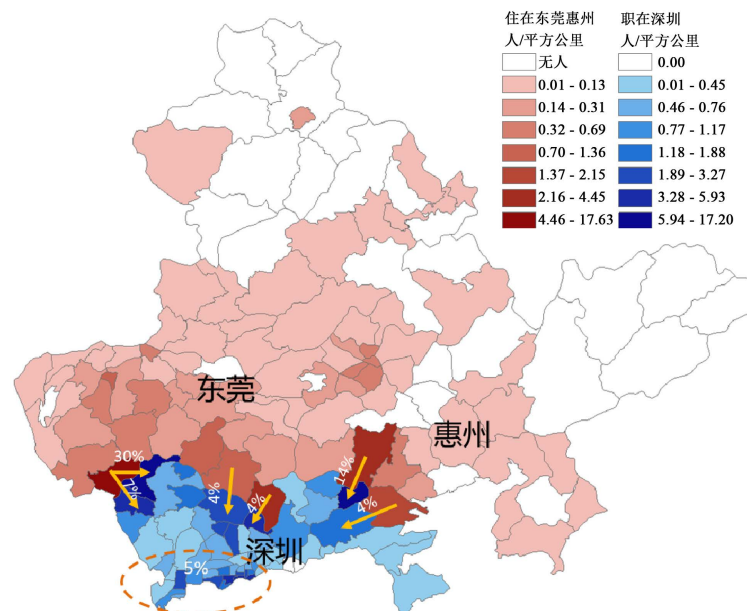


Figure 9. Shenzhen-Dongguan-Huizhou occupation and residence distribution
图 9. 深莞惠职住分布

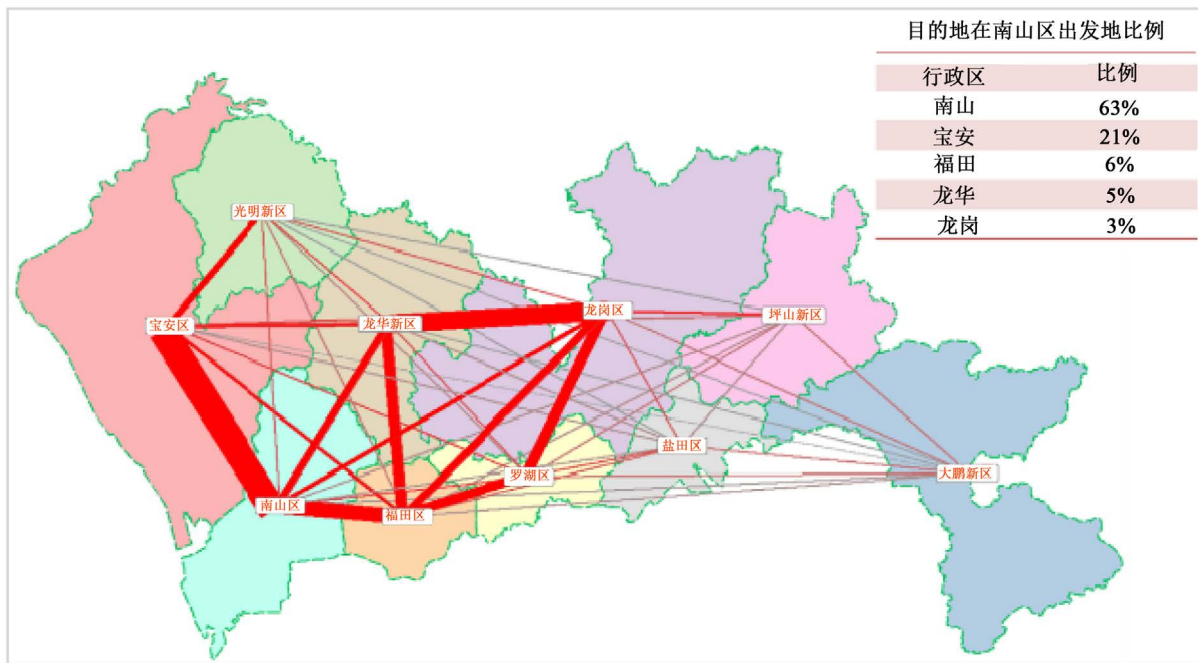


Figure 10. District early peak desire-line
图 10. 行政区早高峰 OD 期望线

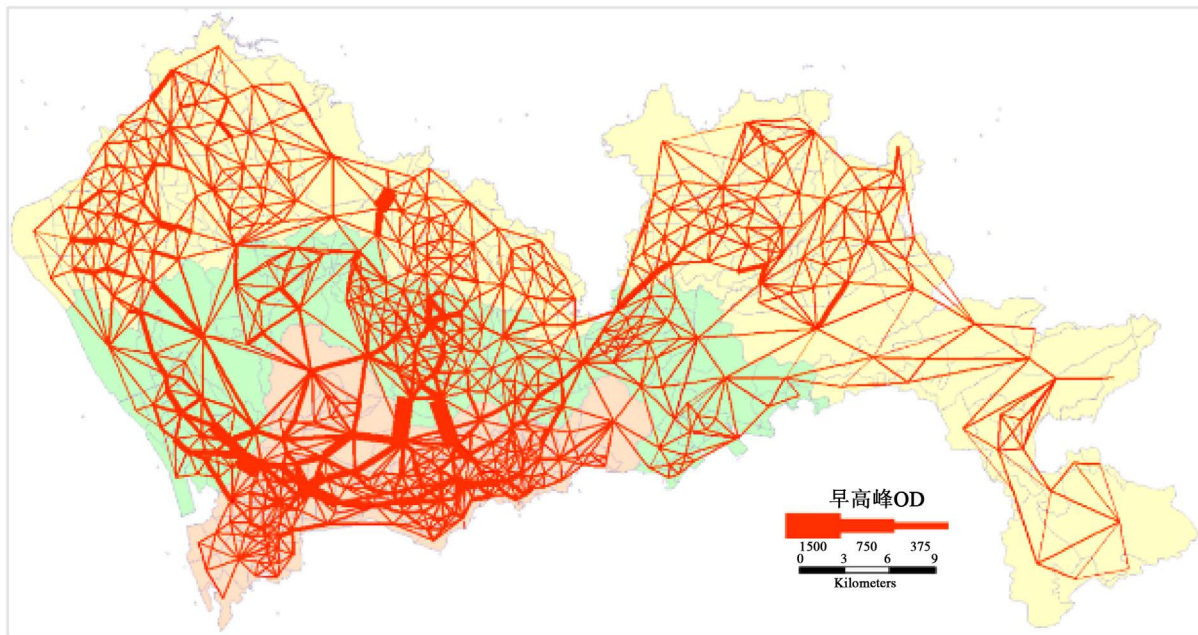


Figure 11. Early peak OD spider web map of traffic community
图 11. 交通小区早高峰 OD 蛛网图

5. 人口活动特征

人口活动特征如图 12 所示，早上十点人口主要分布在高新园、福田中心区、东门、国贸和坂田；晚上九点人口主要分布在粤海、福田中心区、东门、国贸、龙华和坂田。从人口时变图对比，清楚看出人口的时空分布特点，可从中区分出商务热点区和居住热点区。

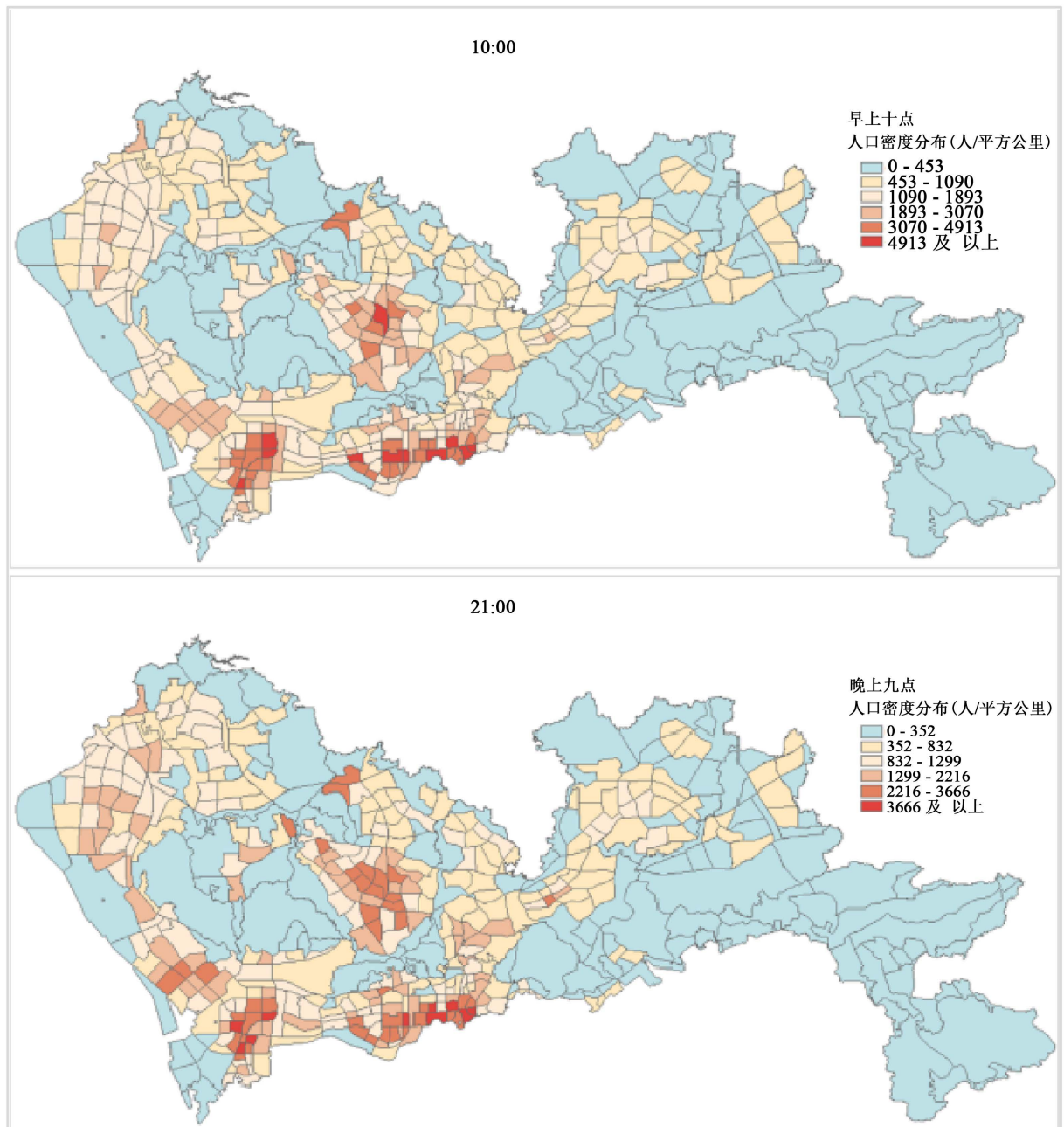


Figure 12. Population activity monitoring plan
图 12. 人口活动监测平面图

6. 总结和展望

原特区内与特区外仍有较强职住通勤吸引；在南山、福田和罗湖的 CBD 工作的人大部分居住在该行政区内；住在东莞、惠州，职在深圳的人，居住和就业位置基本在城市间交界地带；行政区间出行龙华 - 南山最密切；早高峰出行形成明显东西向、南北向通道。

地铁沿线基站尚未识别出来，无法准确知道轨道客流分布特点，下一步期望能区分出轨道基站，将手机数据与公交刷卡数据融合，进一步分析轨道客流分布和预测；采集更多的出行目的标签，分析不同出行目的的构成和人口分布特征。为交通规划建设提供更多有力的依据。

致 谢

感谢深圳市科技计划项目(项目编号 GGFW2016033017241891, 项目名称“深圳市交通大数据公共技术服务平台”)和深圳市战略性新兴产业发展专项资金 2017 年第一批扶持计划(项目名称: 深圳市交通碳排放工程实验室, 批复文号: 深发改[2017]550 号)的资助。

参考文献

- [1] Lu, M., Lai, C.F. and Ye, T.Z. (2017) Visual Analysis of Multiple Route Choices Based on General GPS Trajectories. *IEEE Transactions on Big Data*, 3, 234-247. <https://doi.org/10.1109/TBDATA.2017.2667700>
- [2] 陈佳, 胡波, 左小清, 乐阳. 利用手机定位数据的用户特征挖掘[J]. 武汉大学学报信息科学版, 2014, 39(6): 734-738.
- [3] 张啟梅, 廖玉梅, 任永成, 黄鹏. 基于大数据下的旅客流量分析[J]. 数据挖掘, 2017, 7(1): 26-36.
- [4] 黄涛, 周晨, 黄本雄, 涂来. 基于谱聚类的手机用户日出行移动行为分析[J]. 数据挖掘, 2012, 2(4): 38-42.
- [5] 张楠, 李路华, 周甜甜, 李扉. 小区开放对周边道路通行能力影响的研究[J]. 数据挖掘, 2017, 7(4): 83-92.
- [6] Liu, W., Zheng, Y., Chawla, S., Yuan, J. and Xing, X. (2011) Discovering Spatio-Temporal Causal Interactions in Traffic Data Streams. *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Diego, California, 21-24 August 2011, 1010-1018. <https://doi.org/10.1145/2020408.2020571>
- [7] Yuan, J., Zheng, Y. and Xie, X. (2012) Discovering Regions of Different Functions in a City Using Human Mobility and Pois. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Beijing, 12-16 August 2012, 186-194. <https://doi.org/10.1145/2339530.2339561>

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2163-145X, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: hjdm@hanspub.org