

探索运营商数据在精准扶贫和应急救援中的应用



□尤伟杰¹ 高 见^{2,3} 周 涛^{2,3}

[1. 中国移动通信集团四川有限公司 成都 610041; 2. 电子科技大学 成都 611731;
3. 成都新经济发展研究院 成都 610094]

[摘 要] 准确感知社会经济状态和及时识别应急事件对于智慧社会治理至关重要,然而传统的政府部门数据和统计分析方法无法满足准确性和实时性的需求。近年来,运营商积累的大规模数据,以低获取成本、实时更新和高时空分辨率等优势,为解决问题提供了新思路。回顾运营商数据结合深度挖掘分析算法在精准扶贫和应急救援中的具体应用基础上,并进一步讨论了运营商数据在定量评估扶贫和救灾效果、提高决策效率和治理能力等方面的应用前景。

[关键词] 大数据; 扶贫; 救灾; 社会治理; 计算社会学

[中图分类号] C931; X4

[文献标识码] A

[DOI] 10.14071/j.1008-8105(2018)-1020

Application of Carrier Data on Precise Poverty Alleviation and Emergency Management

YOU Wei-jie¹ GAO Jian^{2,3} ZHOU Tao^{2,3}

(1. China Mobile Group Sichuan Co., Ltd. Chengdu 610041 China;
2. University of Electronic Science and Technology of China Chengdu 611731 China;
3. Institution of New Economic Development Chengdu 610094 China)

Abstract Accurate perception of socioeconomic status and timely identification of emergencies are critical to smart social governance, however, traditional public sector data and statistical analysis methods cannot meet the accuracy and real-time requirements. Recently, large-scale data accumulated by the private sector, with many advantages including low acquisition cost, real-time updates and high spatio-temporal resolution, provide new directions for tackling the problem. This paper overviews the application of carrier data in combination with deep mining analysis algorithms on precise poverty alleviation and emergency management, and further discusses some prospects of applying carrier data to quantitatively evaluate the effect of poverty alleviation and disaster relief and to improve the decision-making efficiency and governance capability.

Key words big data; poverty alleviation; disaster relief; social governance; computational socioeconomics

[收稿日期] 2018-11-05

[基金项目] 国家自然科学基金 (61433014) .

[作者简介] 尤伟杰 (1977-) 男, 中国移动通信集团四川有限公司工程师; 高见 (1989-) 男, 电子科技大学计算机科学与工程学院博士研究生; 周涛 (1982-) 男, 博士, 电子科技大学互联网科学中心教授, 成都新经济发展研究院执行院长。

引言

党的十九大以来,民生工程被放到了极其重要的位置。其中,精准扶贫和应急救灾是民生工程中两个关键的组成部分。党的十九大报告中指出“坚决打赢脱贫攻坚战。让贫困人口和贫困地区同全国一道进入全面小康社会是我们党的庄严承诺。要动员全党全国全社会力量,坚持精准扶贫、精准脱贫……确保到2020年我国现行标准下农村贫困人口实现脱贫……”十九大后,国务院整合多部委原有功能后新设立了“中华人民共和国应急管理部”,习近平总书记指出:“防灾减灾救灾事关人民生命财产安全,事关社会和谐稳定,是衡量执政党领导力、检验政府执行力、评判国家动员力、体现民族凝聚力的一个重要方面”,充分体现了应急救灾工作的重要性。

然而,精准扶贫和应急救灾工作难度极大。在偏远乡村地区,采集信息的手段相对匮乏,信息传输条件相对较差,对于村民的收入情况评估往往只能依赖填报数据甚至几年一次的人口普查数据,存在时间滞后和数据造假的问题。农村发生自然灾害后,对于灾区情况和受灾民众情况的精确信息,往往会滞后数天才能被政府掌握^[1],这也给灾害援救和灾后援助工作带来了困难。Batty预测在21世纪末,世界上绝大多数人口(大于90%)将生活在城市,可以说整个地球都被城市化了^[2]。实际上,现在世界上已经有超过50%的人口生活在城市,中国的城镇化率也到了60%左右。在这种情况下,城市应急事件,包括恐怖袭击、骚乱暴乱、大规模示威游行等社会事件以及地震、暴雨、飓风等容易影响城市的自然灾害,也成为亟待解决的重大挑战。

精确感知社会经济状态,掌握贫困的空间分布,定位贫困人群,实时识别突发应急事件,及时掌握灾害地区受灾群众的情况等,都需要对人群广覆盖且实时性好的感知器。移动手机用户在2014年就超过了全球人口(发达地区人均手机数量超过1部),是目前可利用的最好的感知器之一(有类似作用的感知器还包括社交媒体和遥感卫星等)。因此,运营商拥有的移动智能手机数据完全有望在精准扶贫和应急救灾中发挥巨大作用。

事实上,运营商已经开始了一些有益的工作。以中国移动为例,其精准扶贫平台利用海量数据,努力尝试实现“六大精准”——扶持对象精准、项目安排精准、资金使用精准、措施到户精准、因

村派人精准、脱贫成效精准。该系统已在河南、湖南、重庆等7省市的55个市县落地,覆盖697万贫困人口,服务近40.8万扶贫干部。在应急救灾方面,中国移动汇聚整合国土资源、旅游、水务、气象、林业、海洋渔业等六个部门的数据,结合自身海量数据,建设综合防灾减灾救灾指挥中心监测系统平台,实现了山塘水库水位报警、旅游景点人数分析、船只位置监控、地震速报、火山安全监测等功能。

总体而言,运营商利用自身数据在精准扶贫和应急救灾中已经做出了重要的贡献。与此同时,我们也注意到,大部分平台和系统仅仅具有信息化和可视化的功能。但整体而言,智慧程度不高,数据利用还很不充分。本文将介绍国内外在利用手机数据进行贫困感知和分析应急突发应急事件的监测和分析方面的一些前沿进展,并进一步讨论我国可以在这方面开展的一些有益工作。

一、应用手机数据解决贫穷问题

贫困是21世纪全球发展所长期面临的难题,消除贫困是联合国“千年发展目标”的核心议程之一^[3]。为了实现消除贫困的目标,首先要精准和及时地感知社会经济状态,掌握贫困的空间分布,准确定位贫困人群。传统经济普查在贫困地区的操作更加费时、耗力,导致对贫困状况的感知非常滞后。随着手机在全球的普遍使用,近年来,基于手机数据的社会经济水平推断方法已经在感知贫困区域、识别贫困人群、制定精准扶贫策略等方面展现出巨大的应用价值和前景^[4]。

不同社会经济状况的个体在手机的拥有和使用上也千差万别,因此可以通过手机使用数据来推测个体的财富状况。Blumenstock等人分析了卢旺达的手机使用情况,发现拥有手机的人都相当富裕,而且主要是男性^[5]。Wesolowski等人分析了肯尼亚手机拥有和使用情况数据,发现在农村里贫困人口共享使用手机的情况非常普遍^[6]。Sundsoy等人基于大规模手机使用数据构建深度学习模型,能够以77%的准确性来区分社会经济水平高低的个体^[7]。手机同时也记录着与个人财务相关的数据,例如通话费用和所购买的通话时间。Gutierrez等人分析了科特迪瓦手机用户的话费充值金额变化情况,发现个人的相对收入水平与通话时间购买量和频次有显著关联性^[8]。

个体的社会经济状况与移动行为模式密切相

关,而恰恰手机数据在分析人类移动上有独特优势。Frias-Martinez等人分析了一个拉美国家的大规模手机轨迹数据,发现高收入群体在城市内的移动范围更大^[9]。特别地,Frias-Martinez等人发现社会经济水平与移动总距离和回转半径非常相关——他们基于此提出的模型能够以72%的准确性预测个体社会经济状况^[9]。类似地,基于2000万法国手机用户数据,Pappalardo等人分析人类移动模式与社会经济发展状况之间的关系,发现移动多样性比移动总量更能预测人均收入、初等教育率和失业率等社会经济指标^[10]。基于哥伦比亚首都150万手机用户数据,Florez等人构建了不同收入群体的通勤网络,发现通勤路线的多样性与收入水平正相关,贫困人群通勤距离更远、花费时间更多^[11]。

基于手机通讯数据可以构建社会网络,进而利用社会网络结构推断个体财富状况。Leo等人分析了墨西哥手机数据与银行信贷数据,发现社会网络中存在个人经济状况的同配关联现象,能够根据网络结构能推断个体经济阶层^[12]。类似地,Fixman等人发现手机通讯网络有很强的社会经济同质性,收入水平相近的个体通讯更密切,他们提出的贝叶斯方法能以71%的准确性预测个体收入水平^[13]。Luo等人分析了超过1亿用户的手机通讯数据,发现个体的经济状况与其在社会网络中的位置和影响力非常相关。综合考虑个体的年龄和网络位置,他们提出的复合指标与经济状况的相关性达到0.99^[14]。最近,Jahani等人基于手机数据构建自我中心网络(ego network)^[15],发现社会网络结构多样性与个体收入水平具有强相关性^[16]。

根据位置将手机通讯数据进一步聚合,还能预测区域的社会经济发展水平。基于25亿条科特迪瓦手机数据,Smith-Clarke等人构建区域通讯网络,发现通讯接收量大的区域社会经济水平高^[17]。Šćepanović等人利用时空移动模式来预测社会经济指标,发现回转半径概率分布的空间变化能判断区域财富状况^[18]。基于科特迪瓦手机数据,Smith等人发现区域内的通讯活跃性与贫困指数强负相关^[19]。Mao等人引入CallRank指数刻画区域相对重要性,发现CallRank指数与区域的基尼系数显著相关,电话拨出比率与区域的年收入和贫困率相关^[20]。Eagle等人分析了英国手机通讯网络,发现复合的社会网络多样性指标与区域经济水平的关联性达到0.78^[21]。Blumenstock等人将卢旺达手机通话详单与问卷数据结合,用856个用户的数据训练机器学习模型预测超过150万人的财富状况,从而绘制出高

分辨率的国家财富分布地图(见图1),得到的结果与政府普查数据相关性高达0.79^[22]。



图1 基于150万手机用户通话详单数据预测得到的卢旺达高分辨率财富状况地图

二、应用手机数据解决应急问题

通过挖掘和分析大规模移动手机数据,能为应急救援提供重要信息或者不易直接观察的深刻洞见。从实时发现突发事件和精确感知受灾群众状态这两个方面,具体介绍手机数据解决应急问题的主要进展。

手机被认为是“性价比最高”的社会探针,很多学者相信当人们遇到炸弹袭击、骚乱暴动、飓风地震等突发事件后,手机通讯的数量和模式都会发生变化。因此,可以用手机数据来实时发现可能的突发事件。Dobra等人基于卢旺达地区2005~2009年的通话数据,利用手机通话频率和手机用户移动频率两个主要维度,来实时检测是否发生了突发事件^[23]。他们将总通话量分成每天的数据单元,并将卢旺达领土切割成5 km²的网格(通过基站响应判断用户的移动),然后对比目标时间和目标区域的人群通话频率和移动频率是否异常偏高。如果两者都偏离常态,则被认为有突发事件出现。Gundogdu等人分析了科特迪瓦地区的手机通讯数据^[24]。尽管他们相信移动轨迹也有帮助,但他们认为通话量的数据更加重要。Gundogdu等人建立了一个更精细的马尔科夫模型下的泊松过程来刻画通话量的变化,并通过数据反向估计对应时间序列中存在突发事件因素的概率。对比以前的算法,Gundogdu等人的算法精度有了大幅度的提升,特别是召回率——他们能够成功检测19个标注的突发事件中的15个和11个非突发事件(节假日、演唱会等带来非常态

人群活动但又不属于突发应急的事件)中的8个。

如何分辨同样有大规模人群非常态活动的突发和非突发事件是一个有现实意义的问题。例如恐怖分子炸弹袭击区域的人群和明星演唱会附近的人群有什么不同吗? Gundogdu等人就承认这是一个棘手的问题。Bagrow等人对比了8个突发事件和8个非突发事件前后手机通讯模式的不同^[25]。他们发现突发事件之后手机通话量会立刻激增,几乎没有时延,到达峰值后会很快按指数下降到正常值,其明显偏离正常态的时间较短。反过来,节假日、重大演出等非突发事件手机通话量也会上升,但是上升比较缓慢且往往持续更长的时间。他们还注意到,突发事件中手机通话量的增加部分很大程度上来源于平时在这个时间段不怎么使用手机的人,因此这个特点可以被很好利用起来分辨一次通话量异常是

否对应为突发事件。当然,这需要远远多于群体分析的计算量。他们进一步细致地分析发现,突发事件发生后,接收到突发事件区域用户电话的用户有远远超出寻常的概率会在接下来的一段时间内打回电话。这个现象既无法在非突发事件中观察到,也没有办法用社会互惠性加以解释^[26]。

尽管很多人群的突然聚集并不是来源于突发事件,但是通过手机数据了解这种聚集也是很有帮助的。例如,景区、演唱会、球场、博览会等地人员聚集太多,有可能导致拥挤踩踏等恶性事件^[27]。又比如,突然聚集的人群可能为某些传染病突然大爆发提供了温床^[28]。最近, Finger等人就通过手机数据发现塞内加尔在朝圣的时候会出现超大规模人群聚集(见图2),而这种聚集很可能是导致2005年塞内加尔霍乱疫情突然激增的主要原因^[29]。

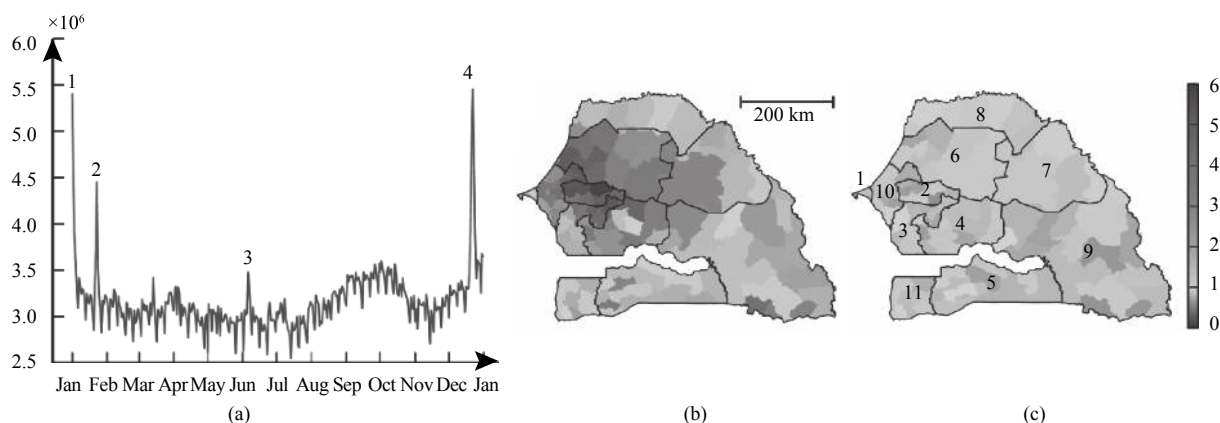


图2 塞内加尔朝圣时期的人群聚集。

注:图A是利用手机数据估计的离开家的用户数,其中1和4对应两个朝圣的节日;

图B中深色区域是朝圣时圣地人群数目;图C是全年平均人数

手机数据也可以用于及时了解灾害,特别是造成严重影响的自然灾害发生之后受灾地区民众的情况^[30]。吕欣等人研究了2010年海地地震前后190万手机用户的移动轨迹数据^[31]。他们发现震中地区太子港有23%的人口在灾后迁移并居住在其他城市,然后经过近一年的时间,这些人中的大部分逐渐回到了原居住地(见图3)。不仅是因灾迁移的人数,而且迁移人员的去向,都可以通过手机数据实时获得,而民政部门要获得精确的数据所需要的费用和时间都要高得多。类似的技术已经开始在中国应用。四川九寨沟地震区域刚好是旅游区,震后的一项主要工作是人员疏散和撤离景区。根据运营商数据显示,地震发生时九寨沟景区及周边区域人员数量在8万人左右,经过24小时的疏散,锐减到2万人左右,有6万人左右撤离灾区。

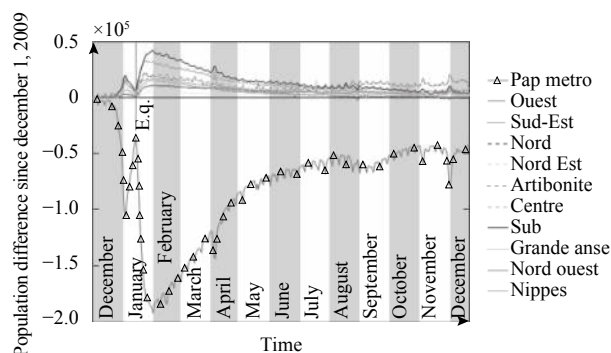


图3 海地地震后人口移动数据。

注:—△—线代表的是震中太子港。短时间内,大量人口离开,另外十个城市的人口都有一定程度的增长。随着地震发生时间推移,太子港人口总量逐渐恢复常态,其他城市人口增量也逐渐降低

吕欣等人研究了地震后海地的灾民并没有出现所谓的“混乱”,实际上灾民行为的可预测性在经

过短时间的降低后变得比平时还高。Kenett和Portugali^[32]认为这一发现实际上从某个侧面印证著名的“Hägerstrand理论”^[33]——人们绝大多数时间都生活在某种重复的常规状态中(受空间、经济、社会、心理等限制)。如果长时间脱离常规状态,人们会试图回到原来的常态或者重建一种新常态,然后继续过“有规律的生活”^[34]。

三、结论与讨论

本文对手机数据结合深度挖掘分析算法在精准扶贫和应急救援中的可能应用进行了回顾。总体而言,我们坚信运营商的数据(最好在进一步整合一些其他关键部门数据^[35])可以在这两个重大民生工程中发挥巨大作用。我们也注意到了运营商已经做出的一些有益尝试,但这种尝试只是应用了大量数据的简单统计分析,距离真正的智慧化以及给出深刻洞见以支撑高效正确决策,还有很长的距离——当然,这也是全球共同面对的困难和挑战!

在精准扶贫方面,运营商已经建立了比较好的信息化系统和业务平台,例如“渝扶通”扶贫大数据平台能够实现对贫困对象、扶贫干部、扶贫工作等基础数据的收集和更新,包括贫困程度、主要致贫原因、贫困户属性、家庭照片等。这类数据如果能够结合动态的移动手机数据,有望在定位贫困人群和量化扶贫前后效果方面做得更精确。进一步地,对一个城市数据之间关联的深度分析,也可以帮助在其他城市建立更准确地从移动手机数据到贫困程度之间的模型。

在应急事件发现方面,原来的专家规则,特别是仅仅关注重点区域的人员数量的阈值模型,很容易造成误报(报警但实际没有突发事件)和漏报(有突发事件但没有报警)。运营商和政府应该学习先进的分析算法,大幅度提高预测的精准度。在援灾救灾方面,也需要通过数据了解灾区人员流动和救灾人员及物资所在位置,以提高援灾救灾的效率。特别地,手机数据可以用于检测政策是否产生效果,譬如Morales和Pastor就注意到墨西哥洪灾预警的信号发出来之后,涉灾区域的群众根本不为所动(从手机通讯和移动模式上看不到异常),直到洪水到了,灾民才作出反应,结果造成重大损失^[36]。这种对政策效果的定量分析和评估对于提高我国政府决策和治理能力特别重要。

参考文献

[1] CHUA A Y, KAYNAK S, FOO S S. An analysis of the

delayed response to Hurricane Katrina through the lens of knowledge management[J]. *Journal of the American Society for Information Science and Technology*, 2007, 58(3): 391-403.

[2] BATTY M. When all the world's a city[J]. *Environment and Planning A*, 2011, 43(4): 765-772.

[3] United Nations. Millennium Development Goals [EB/OL]. (2008-05-20). <http://www.un.org/millenniumgoals>.

[4] 高见, 周涛. 大数据揭示经济发展状况[J]. *电子科技大学学报*, 2016, 45(4): 625-633.

[5] BLUMENSTOCK J E, GILLICK D, EAGLE N. Who's calling? Demographics of mobile phone use in Rwanda[J]. *Transportation*, 2010, 32: 2-5.

[6] WESOŁOWSKI A, EAGLE N, NOOR A M, et al. Heterogeneous mobile phone ownership and usage patterns in Kenya[J]. *PLoS ONE*, 2012, 7: e35319.

[7] SUNDSØY P, BJELLAND J, REME B, et al. Deep learning applied to mobile phone data for individual income classification[C]//2016 International Conference on Artificial Intelligence: Technologies and Applications. Paris, France: Atlantis Press, 2016: 96-99.

[8] GUTIERREZ T, KRINGS G, BLONDEL V D. Evaluating socio-economic state of a country analyzing airtime credit and mobile phone datasets[EB/OL]. (2013-09-17). <http://arxiv.org/abs/1309.4496>.

[9] FRIAS-MARTINEZ V, VIRSEDA-JEREZ J, FRIAS-MARTINEZ E. On the relation between socio-economic status and physical mobility[J]. *Information Technology for Development*, 2012, 18: 91-106.

[10] PAPPALARDO L, PEDRESCHI D, SMOREDA Z, et al. Using big data to study the link between human mobility and socio-economic development[C]//2015 IEEE International Conference on Big Data. New York, NY, USA: IEEE Press, 2015: 871-878.

[11] FLOREZ M, JIANG S, LI R, et al. Measuring the impact of economic well-being in commuting networks—A case study of Bogota, Colombia[C]//Proceedings of the Transportation Research Board 96th Annual Meeting. Washington, DC, USA: Transportation Research Board, 2018: 1-19.

[12] LEO Y, FLEURY E, ALVAREZ-HAMELIN J I, et al. Socioeconomic correlations and stratification in social-communication networks[J]. *Journal of The Royal Society Interface*, 2016, 13(125): 20160598.

[13] FIXMAN M, BERENSTEIN A, BREA J, et al. A Bayesian approach to income inference in a communication network[C]//2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. New York, NY, USA: IEEE Press, 2016: 579-582.

[14] LUO S, MORONE F, SARRAUTE C, et al. Inferring personal economic status from social network location[J].

Nature Communications, 2017, 8: 15227.

[15] WANG Q, GAO J, ZHOU T, et al. Critical size of ego communication networks[J]. *EPL (Europhysics Letters)*, 2016, 114(5): 58004.

[16] JAHANI E, SAINT-JACQUES G, SUNDSØY P, et al. Differential network effects on economic outcomes: A structural perspective[C]//*Social Informatics*. Cham: Springer International Publishing, 2017: 41-50.

[17] SMITH-CLARKE C, MASHHADI A, CAPRA L. Poverty on the cheap: Estimating poverty maps using aggregated mobile communication networks[C]//*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM Press, 2014: 511-520.

[18] ŠĆEPANOVIĆ S, MISHKOVSKI I, HUI P, et al. Mobile phone call data as a regional socio-economic proxy indicator[J]. *PLoS ONE*, 2015, 10: e0124160.

[19] SMITH C, MASHHADI A, CAPRA L. Ubiquitous sensing for mapping poverty in developing countries[C]//*Proceedings of the Third Conference on the Analysis of Mobile Phone Datasets, Data for Development Challenge*. Cambridge, MA, USA: MIT Media Lab, 2013: 1-7.

[20] MAO H, SHUAI X, AHN Y Y, et al. Quantifying socio-economic indicators in developing countries from mobile phone communication data: applications to Côte d'Ivoire[J]. *EPJ Data Science*, 2015, 4: 15.

[21] EAGLE N, MACY M, CLAXTON R. Network diversity and economic development[J]. *Science*, 2010, 328 (5981): 1029-1031.

[22] BLUMENSTOCK J, CADAMURO G, ON R. Predicting poverty and wealth from mobile phone metadata[J]. *Science*, 2015, 350(6264): 1073-1076.

[23] DOBRA A, WILLIAMS N E, EAGLE N. Spatiotemporal detection of unusual human population behavior using mobile phone data[J]. *PLoS ONE*, 2015, 10(3): e0120449.

[24] GUNDOGDU D, INCEL O D, SALAH A A, et al. Countrywide arrhythmia: emergency event detection using mobile phone data[J]. *EPJ Data Science*, 2016, 5(1): 25.

[25] BAGROW J P, WANG D, BARABÁSI A-L.

Collective response of human populations to large-scale emergencies[J]. *PLoS ONE*, 2011, 6(3): e17680.

[26] GAO L, SONG C, GAO Z, et al. Quantifying information flow during emergencies[J]. *Scientific Reports*, 2014, 4: 3997.

[27] 周涛, 韩筱璞, 闫小勇, 等. 人类行为时空特性的统计力学[J]. *电子科技大学学报*, 2013, 42(4): 481-540.

[28] WESOŁOWSKI A, EAGLE N, TATEM A J, et al. Quantifying the impact of human mobility on malaria[J]. *Science*, 2012, 338(6104): 267-270.

[29] FINGER F, GENOLET T, MARI L, et al. Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks[J]. *Proceedings of the National Academy of Sciences, USA*, 2016, 113(23): 6421-6426.

[30] BENGTSSON L, LU X, THORSON A, et al. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti[J]. *PLoS Medicine*, 2011, 8(8): e1001083.

[31] LU X, BENGTSSON L, HOLME P. Predictability of population displacement after the 2010 Haiti earthquake[J]. *Proceedings of the National Academy of Sciences, USA*, 2012, 109(29): 11576-11581.

[32] KENETT D Y, PORTUGALI J. Population movement under extreme events[J]. *Proceedings of the National Academy of Sciences, USA*, 2012, 109(29): 11472-11473.

[33] HÄGERSTRAND T. What about people in regional science?[J]. *Papers in Regional Science*, 1970, 24(1): 7-24.

[34] CAO Y, GAO J, LIAN D, et al. Orderliness predicts academic performance: Behavioural analysis on campus lifestyle[J]. *Journal of The Royal Society Interface*, 2018, 15(146): 20180210.

[35] EINAV L, LEVIN J. Economics in the age of big data[J]. *Science*, 2014, 346(6210): 1243089.

[36] MORALES A, PASTOR D. Studying human behavior through the lens of mobile phones during floods[C]//*Proceedings of the Third International Conference on the Analysis of Mobile Data*. Cambridge, MA, USA: MIT Media Lab, 2015: 69-71.

编辑 何婧