

文章编号:1000-5641(2015)05-0162-10

# 基于手机大数据的城市人口流动分析系统

包 婷, 章志刚, 金澈清

(华东师范大学 数据科学与工程研究院 上海市高可信计算重点实验室,上海 200062)

**摘要:** 分析城市人口流动行为有助于合理分配社会资源,有效应对交通压力、维护社会公共治安等。传统的人工分析方法,如问卷调查、座谈访问等,成本高昂且低效率。智能手机的不断发展与普及在为人们日常生活带来极大便利的同时,所产生的用户移动轨迹数据为有效分析城市人口流动行为提供了可能。然而,海量、低质的轨迹数据给查询分析工作带来了诸多挑战。文中提出了一个分布式人口流动分析框架,采用多节点处理任务,从而提升了算法的执行能力和可扩展性。利用手机运营商提供的手机轨迹数据,分析城市人口流动情况,建立了多个模型,包括进出城市的人口流动行为分析模型、市内各区县间的人口流动行为分析模型、居民工作地/居住地人口分析模型。与传统方法相比,本方案的成本更低,效率更高,覆盖人群更广。

**关键词:** 人口流动; 轨迹数据; 分布式框架

中图分类号: TP309 文献标识码: A DOI:10.3969/j.issn.1000-5641.2015.05.014

## An urban population flow analysis system based on mobile big data

BAO Ting, ZHANG Zhi-gang, JIN Che-qing

(Shanghai Key Laboratory of Trustworthy Computing, Institute for Data Science and Engineering,  
East China Normal University, Shanghai 200062, China)

**Abstract:** Analysis on urban population flow can help to make rational distribution of social resources, cope with traffic pressure and maintain public order, etc. The traditional manual analysis methods, such as questionnaire and interview, can not deal with this task efficiently. The continuous development and prevalence of smart phones bring great convenience to people's daily life and users' trajectory data generated by the connection between smart phones and base stations, which makes it possible to implement this task. However, trajectory data is massive and has low quality, which brings great challenge to related work. We propose a distributed framework for population flow analysis by using multiple computing nodes, thus greatly enhancing efficiency and scalability. In this paper, we use the massive trajectory data to analyze the behavior of urban population flow. We model flowing behavior among cities and among inner-city districts, and decide the work place and living place of each person. Compared with the traditional methods, our method is cheaper and more efficient.

**Key words:** population flow; trajectory data; distributed framework

收稿日期:2015-06

基金项目:国家重点基础研究发展计划(973)(2012CB316203);国家自然科学基金(61370101);上海市教委科研创新重点项目(14ZZ045)

第一作者:包婷,女,硕士研究生,研究方向为LBS和大数据处理技术。

通信作者:金澈清,男,博士生导师,研究方向为LBS和大数据处理技术. E-mail: cqjin@sei.ecnu.edu.cn.

## 0 引言

信息技术的高速发展加速了城市化进程。在此过程中,城市人口的剧增也加大了城市管理难度,例如交通压力、就业压力等。由于地区间经济发展不均衡,城市内部各区域的功能分工各有不同,导致城市内部人口会大量流动。受限于地理和社交等因素,人们的行为往往呈现出规律性<sup>[1-3]</sup>,就是人们在工作地和居住地的周期性位置变迁<sup>[4]</sup>。通过分析人口流动行为以及居民工作地和居住地等信息,有助于优化社会资源分配,应对交通压力、维护社会公共治安。

长期以来,人口流动行为研究往往采用如现场观察调查、问卷调查、座谈访谈等人工手段,成本高昂且效率不高。随着智能手机的不断发展与普及,海量的手机轨迹数据为研究城市人口流动行为提供了一种新方法。在各大城市中,各个移动通讯运营商均布置了多个基站。当用户接听/拨打电话、收发短信或者使用数据通讯服务时,就会生成基站连接记录,产生海量的手机数据。手机轨迹数据不仅数量庞大,而且质量低下。基站类型多样,包括微站、宏站、直放站和射频拉远站,覆盖范围从几百米到几千米不等。而定位精度很大程度上依赖于基站的分布密度及其覆盖范围的大小,在不同区域,基站的分布密度差异显著。例如市中心区域的基站密度远高于郊区的基站密度。此外,基站跳变也会极大地影响手机轨迹数据的质量;换言之,如果用户所处位置恰巧处于多个基站的服务范围之内,当用户稍微移动位置甚至固定在某个地方时,手机也会在多个基站间切换连接,而我们使用手机与基站连接日志记录来判定用户的移动轨迹,这就导致难以真正判定用户的真实位置。

尽管手机定位数据是离散和稀疏的,但利用手机数据仍然可以对人们的行为进行高精度的预测<sup>[5]</sup>。该结论为利用手机数据研究城市人口流动提供了理论前提。现有工作大都是针对集中式环境,无法直接应用于海量数据环境。为此,本文提出了一种基于 Map/Reduce 的分布式框架来对城市人口流动行为进行研究分析,具有较好的执行效率和可扩展性。本文利用运营商提供的海量手机轨迹数据,对手机用户在城市的流动行为进行分析和挖掘,同时对数据进行了模糊化处理以满足用户的隐私保护需求,并建立了多个模型,包括进出城市的人口流动行为分析模型、市内各区县间的人口流动行为分析模型,特别地,对区县间流动行为建立了居民工作地居住地流动行为分析模型。这些模型为更好地了解用户特征,分析城市人口流动提供了可能。

## 1 相关工作

近年来,已有不少工作针对手机轨迹数据研究用户的行为模式。文献[7]将 OD(Origin-Destination,起止)矩阵作为输入,从手机轨迹数据中提取用户起止点信息。文献[8]将移动手机流量关联到交通流量,设计 GSM(Global System for Mobile Communication)网络模拟器来模拟从电话网络中提取出的网络数据,将数据处理后转化成 OD 矩阵,从而判定出移动行为的起始地点。文献[9]将基站的连接记录与交通流量相结合,建立 OD 矩阵,从而进一步分析用户轨迹。文献[10]从手机数据中提取用户每天位置轨迹并转化为活动序列,将序列进行分类得到用户活动的转移模式。文献[11]利用近百万条手机数据提取用户行为模式并分析不同用户工作地间的相关性。文献[12]利用聚类、回归的方法分析匿名化的手机数据,根据用户稀疏的位置信息发现有意义的重要位置如工作地、居住地。海量的手机数据为分析人

们行为提供了可能,文献[13]利用手机轨迹数据挖掘用户异常聚集活动,如异常的社会活动的发现。文献[14]将手机数据与推荐系统相结合,挖掘用户行为模式并向用户推荐感兴趣的社会活动。

文献[15]利用实时采集的移动手机数据分析城市交通状况、预测行人活动序列。社会经济水平可以反映出人们住房、教育、健康以及其他基础服务情况,文献[16]利用手机数据聚类分析后的信息来确定社会经济水平,并利用 SVM 和随机森林模型来预测社会经济水平。有很多研究利用手机数据对交通流量进行评估,但这些研究往往忽视了每辆车可能有多个手机的情况,文献[17]利用聚类的方法判定同一辆车中是否有多台手机,从而利用手机数据确定车速、车辆密度等,并对高速公路交通流量作出更精确评估。

由于手机轨迹数据数量庞大且质量低下,同时为了不泄露用户隐私,移动运营商往往将用户手机轨迹数据进行模糊处理,这些给相关研究带来了很大的挑战。本文针对城市人口流入流出行为展开研究,利用手机轨迹数据发现用户行为模式,并挖掘用户工作地、居住地信息,为今后的研究提供了一种新思路。

## 2 系统框架结构

本节介绍系统的框架结构,如图 1 所示。

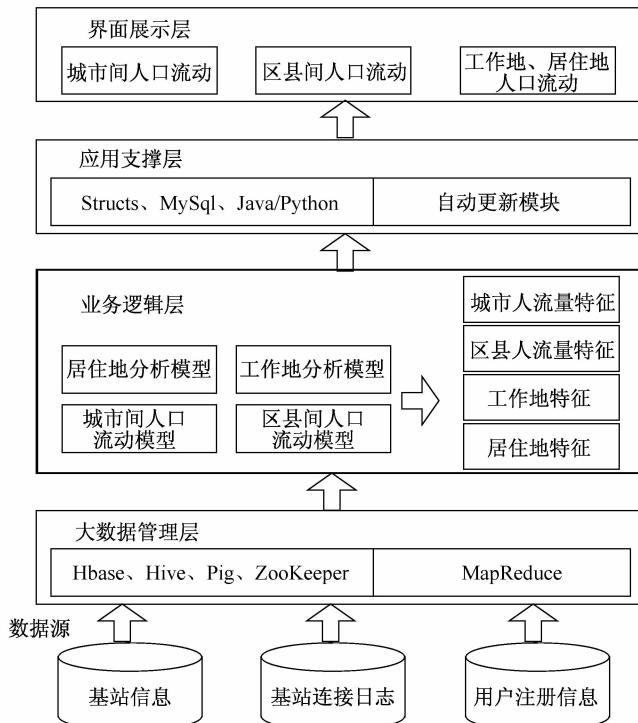


图 1 系统框架结构

Fig. 1 System architecture

构建人口流动分析平台需要使用以下数据:基站连接日志数据、基站信息数据、用户注册信息。基站连接日志数据描述了用户每一次手机连接基站的情况,包括:手机设备号、连接起始时间、连接基站号、连接扇区号、手机开关机状态、加载时间等字段。基站数据描述了基

站的基础信息,包括:基站 ID、地理位置、GPS 坐标、所在行政区等。用户注册信息包括:用户身份证号码、姓名、手机号、性别、出生年月等。

为了保护用户隐私和移动运营商的隐私,在发布使用数据时,本文对这两部分数据进行了隐私保护。对于用户注册信息,隐匿了身份证号码和姓名,且对其手机号和基站连接日志中的手机号使用了一致的加密手段。为了保护移动运营商的基站信息,对基站位置在不影响功能分析的基础上做了位置修正。

系统架构在设计上采用分布式、分层结构,包括大数据管理层、业务逻辑层、应用支撑层、界面展示层 4 层结构。

大数据管理层使用 HBase、Hive、Pig、ZooKeeper 来管理基站连接日志。Hadoop 集群上使用 MapReduce 框架来执行任务,数据库使用 HBase,数据仓库使用 Hive,并使用 Pig 语言来简化 Hadoop 工作任务,使用 Zookeeper 进行集群内的协作服务。

业务逻辑层利用数据管理层对手机数据分析处理后输出的数据,建立分析模型,包括城市间人口流动模型、区县间人口流动模型、居住地分析模型、工作地分析模型。针对这四个模型,对数据进行分析处理,训练出相应的特征,这些特征可用来表示城市人口在城市间、各区县间、居住地/工作地这三个层次的流动情况。

后台开发框架使用集成的 Struts。Struts 采用 Java Servlet/JSP 技术,开发 Web 应用程序的开放源码的框架。数据库主要用来存储后台分析好的结果,本项目采用 MySQL 数据库,MySQL 是开源数据库且体积小、速度快、适用于快速部署。后台处理程序使用 Java/Python 语言编写,负责处理前端发过来的请求,并从大数据平台获取分析结果,存放到数据库中。系统通过创建脚本文件并将文件加入到任务计划中,实现周期性更新数据。

界面展示层用来与用户进行交互,并展示系统分析结果。网页效果设计采用的主要 Flash 技术,使用 Flash 技术可以创作出可改变尺寸的导航界面以及其他奇特的效果。本项目采用 Flash 技术的主要原因是可以自定义开发,开发周期短,图形和动画效果丰富,并且 Flash 使用向量运算的方式,产生的文件占用存储空间较小。系统使用 JSON + XML 技术来获取数据库中的数据。

### 3 大数据管理

大数据管理使用 Hadoop 这一开源平台来实现。如图 2<sup>[18]</sup> 所示,该平台集成了 HBase、Hive、Pig、Zookeeper 等实用工具,方便了用户对数据的管理和操作。HBase 是 Hadoop 的数据库,能够对大数据提供随机、实时的读写访问功能,是一个高可靠、高性能、面向列、可伸缩的分布式存储系统。HBase 存储的数据从逻辑上来看就像一张很大的表,并且它的数据列可以根据需要动态地增加。Hive 是一个基于 Hadoop 文件系统之上的数据仓库架构。它为数据仓库的管理提供了许多功能:数据 ETL(抽取、转换和加载)工具、数据存储管理和大型数据集的查询和分析能力。同时,Hive 定义了类 SQL 的语言——Hive QL。Hive QL 允许用户进行和 SQL 相似操作,还允许开发人员方便地使用 map 和 reduce 操作,这对 Map/Reduce 框架是一个强有力的支持。Hive 本身建立在 Hadoop 的体系架构上,可将外部命令解析成一个 Map/Reduce 可执行计划。Pig 为大型数据集的处理提供了更高层次的抽象,它提供了一套强大的数据变换操作,这些操作整体上描述了一组数据流到另一组数据流的转换,而这些转换操作被转换成一系列的 Map/Reduce 作业,这样以来使得程序员仅仅需要编

写简单的脚本代码,就能轻松处理 TB 级的数据集<sup>[18]</sup>. 此外,Pig 和 Hive 还为 HBase 提供了高层语言支持,使得在 HBase 上进行数据统计处理变得非常简单.

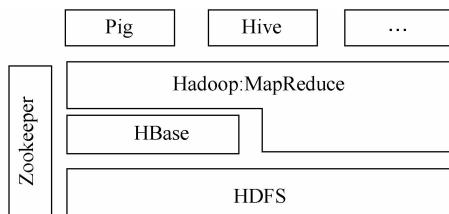


图 2 大数据管理的技术框架

Fig. 2 Technical architecture of data management

为方便各种分析任务对用户信息和基站信息的实时查询,本系统将用户的注册信息和基站的信息存放在 HBase 中. 同时,将用户连接基站数据存放在 HDFS 上,并导入到 Hive 所构建的数据仓库中进行管理. 为了满足各种任务分析需求,本系统提供了 3 种数据操作方式:首先,对于简单的数据查询使用 Hive QL 命令来进行操作. 比如查询指定用户某天连接过哪些基站,使用 Hive QL 编写一句查询语句就能完成任务;其次,对于批处理任务,本系统通过 Pig 脚本程序实现. 比如,由于用户连接基站数据每天会批量更新,如果想知道用户每天都出现在哪些区县. 在实现该任务时涉及到两个数据的连接操作,这时使用 Pig 脚本程序能够方便的完成上述分析任务;最后,对于复杂分析任务,通过编写 Map/Reduce 程序对存放在 HDFS 的数据进行操作,比如需要从用户连接基站历史记录中分析出用户的居住地时,Hive QL 和 Pig 程序无法满足需求,这时就需要用户自己编写 Map/Reduce 程序完成分析.

## 4 模型分析

本节介绍本系统所涉及到的各个分析模型. 人口流动的分析涉及范围很广,包括城市的流入流出分析、城市内部各区县间人口流动分析和用户居住地/工作地分析. 系统分别建立了三个相应的数据分析模型.

### 4.1 城市流入流出模型分析

本文提出了一种利用手机轨迹数据监测人口流动的方法,处理框架如图 3 所示. 首先对基站连接数据进行预处理,以降低手机基站数据低质问题的影响,然后分析进出城市的行为模式,利用分析分类模型判定用户轨迹是否进出城市.

#### (1) 数据预处理

该过程是为了减小基站定位不准以及信号跳变造成的影响,本系统分析手机基站数据特点,挖掘出用户的重要停留区域,由此降低用户在该区域范围内的信号定位不准以及信号频繁跳变问题的影响. 将这些停留区域按时序串联以构成用户的活动轨迹. 数据预处理部分包括三步:①去除异常点与建立移动轨迹;②挖掘用户活动轨迹;③建立枢纽区域和边境区域.

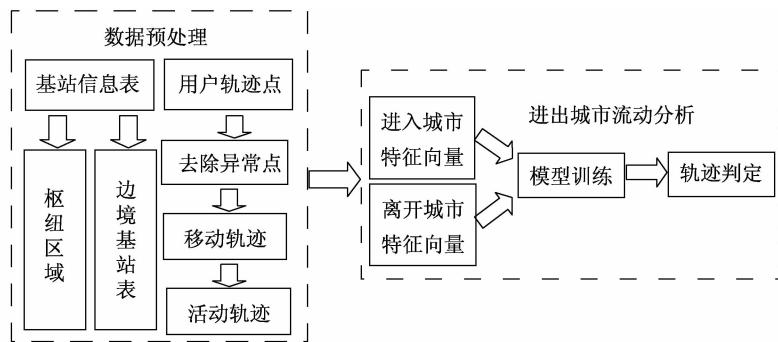


图3 进出城市人口流动处理框架

Fig. 3 Processing framework of population flow among cities

## (2) 进出城市流动分析

用户行为轨迹复杂,本文通过分析用户进出城市的轨迹行为特点,与非进入非离开城市行为进行对照,挖掘轨迹特征.具体特征如下:

- 1) 信号消失时长 相邻两停留点之间的时间间隔.
- 2) 枢纽区域出现概率 用户进入或离开城市时,在某交通枢纽处(如火车站、机场等)出现的可能性.
- 3) 枢纽区域停留指数 用户在某枢纽区的停留程度.
- 4) 是否在边境区域出现 用户的手机信号消失或出现时,是否与城市边境处的基站进行连接交互.
- 5) 与居住地和工作地的平均距离 手机信号消失或出现时,用户与其居住地和工作地之间的平均距离.

用户进出城市的行为必定发生在某一信号消失时段的前后,因此本文针对信号消失时段的前后,分别得到用户的离开行为特征向量和进入行为特征向量,然后利用标注数据训练分类模型(如决策树、逻辑回归等),之后运用分类模型的训练结果进行轨迹行为判定,最终判定用户在某时刻是否进入或离开了城市<sup>[19]</sup>.

## 4.2 城市各区县间人口流动模型分析

区县间人口分析框架如图4所示.首先同样需要进行数据预处理,处理方法与城市间人口流动数据预处理方式一致,然后通过分析各区县间人口流动判定用户在区县间流动情况.

针对某一用户的行为轨迹,数据预处理阶段可得到该用户的多个重要停留区,用户在这些区域中有较大的可能性进行活动,停留的时间较长.本文将这些重要的停留区域用圆表示,这些圆的圆心所在的区县即为用户的停留区县,将处于同一区县并且时间上相隔小于一定阈值的停留区域进行状态合并,即可得到该用户的区县停留状态,该状态信息中包括区县ID和停留的起止时间.

由于用户去往目的区县的过程中可能会经过某些“经过区”,例如某用户从普陀出发,去往闵行上班,途中经过长宁区和徐汇区,这两个区即为“经过区”.在分析用户在区县间流入流出行为时,需过滤“经过区”,以挖掘用户行走路线的真正意图.上述过程所得的停留区县即为用户真正的活动区县,“经过区”不构成停留区县状态,因此时间上相邻的两个停留区县间的状态转移伴随着用户的一次离开区县和进入另一区县的行为.最后通过汇总全体数据

集中用户所有停留区县间的状态转移情况,即可得到在各个时间段内的不同区县间的人口流动情况<sup>[19]</sup>.

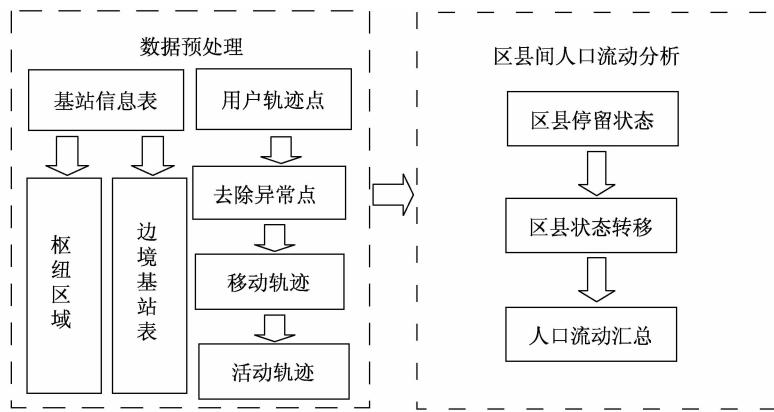


图 4 区县间人口流动处理框架

Fig. 4 Processing framework of population flow among districts

#### 4.3 居民工作地、居住地模型分析

为了发现用户工作地居住地信息,本文提出了一个基于 Map/Reduce 的框架,该处理框架主要包含 4 个步骤:①过滤各个用户的轨迹记录;②找出包含工作地、居住地的候选区域;③调用传统聚类算法对候选区域进行聚类操作;④从聚类结果中发现工作地、居住地信息。同时,本文在该框架中分别嵌入了两种不同的发现策略,即:GPMA 和 SPMA。这个分布式挖掘框架主要思路是:首先,通过 Map/Reduce 编程模型将同一个用户在总时间内的连接记录,合并到同一个计算节点。然后,选择合适的位置范围化方法,针对用户原始连接基站的记录使用状态生成算法生成状态序列。接着,根据停留时间和停留次数,删除那些不满足给定阈值的状态,从剩下的状态所对应的区域中找出那些可能包含工作地、居住地位置的候选区域。最后,对找出的候选区域继续聚类,从聚类结构中分析出用户的工作地、居住地信息。

针对基于网格范围和基于基站覆盖范围的两种区域范围化方法,利用提出的分布式挖掘框架,设计了两种并行挖掘算法:GPMA 算法和 SPMA 算法。GPMA 算法首先将整个区域进行栅格化,然后将用户连接基站的情况映射为在各个网格内的停留状态,停留状态包含停留的网格号、起始停留时间、结束停留时间。由于精度和基站跳变的原因,则认为该用户停留在某网格中时,实际所处位置也有可能是该网格的邻居网格。SPMA 算法利用基站覆盖范围来表示用户所在区域范围,它将用户连接基站的情况转换为连接各个基站的序列。GPMA 算法思想的想法比较简单直观,但该方法将同一网格中的基站等同看待,这会加大所得工作地、居住地位置的偏差。而 SPMA 算法克服了这一问题,它将每一个基站单独看待,每一个基站可以有不同的覆盖范围。当用户连接到某基站时,则用户可以确定用户在该基站的覆盖范围内,同时由于基站切换原因,用户也可能在该基站邻居的覆盖范围内<sup>[20]</sup>。

#### 5 界面展示

图 5 是城市各区县间人口流动展示界面。为了提高用户与系统的交互性,用户可直接点击地图上的相应区县直接选择,查看其它各区县流入到该区县或者该区县流出到其他区县

的人口流量。默认展示的是当天的人口流量。系统后端进行数据处理后将结果保存在 MySQL 数据库中,并编写脚本文件实现系统数据的周期性更新。前端采用 XML + JSON 技术获取所需数据,考虑到系统的多模块性以及用户所选时间段的多样性,采用分模块加载数据,提高系统的响应速度,提供良好的用户体验。

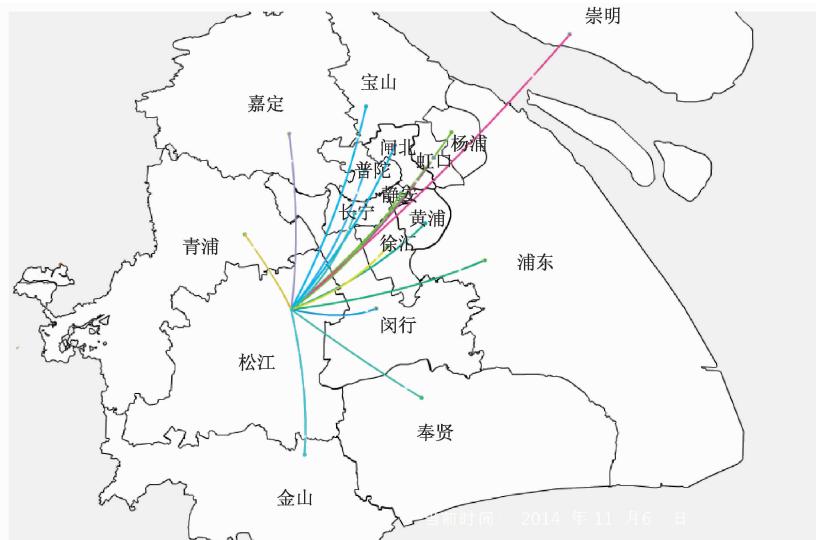


图 5 人口流动界面对比图

Fig. 5 Contrast figure of interface of population flowing

图 6 是系统的数据对比图模块,分别从上海市、各区县的角度来动态展示不同时期的人口流动对比情况。系统提供了月与月间、周末与工作日间的对比,多样化的展示人口流动的变化情况。系统设计了简单查询与复合查询两种查询功能。简单查询提供按照时间的查询,复合查询提供区县、时间的组合查询。用户可根据自身需求采用相应查询方法,获取所需数据。

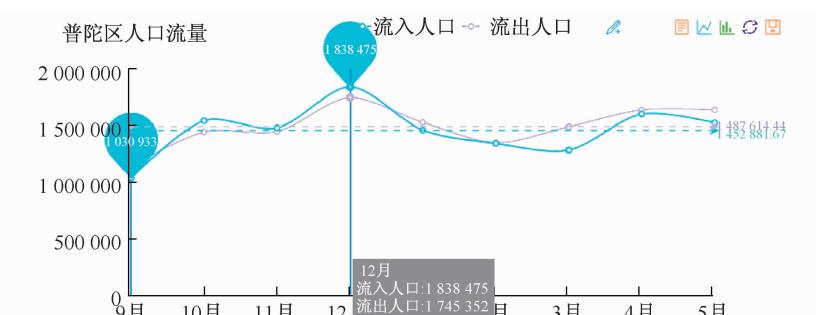


图 6 数据对比图

Fig. 6 Contrast figure of data

## 6 总 结

本文利用大数据平台分析用户的手机轨迹数据,挖掘用户的行为模式,设计了基于传统数据仓库、Hadoop 集群和 MySQL 数据库的上海市人口流动分析平台,包括三个分析模块,分别是城市间人口流动分析模块、区县间人口流动分析模块、工作地/居住地流动分析模块。在分析城市间人口流动时,提出了在分布式框架下的基于轨迹行为特征的判定算法;在分析区县间人口流动时,对用户在区县间流动行为加以分析;对工作地、居住地进行分析时,介绍了两种挖掘重要位置信息的算法:GPMA 和 SPMA。本文为有效、准确分析城市人口行为提供了有力支持。

### [参 考 文 献]

- [1] GONZALEZ M C, HIDALGO C A, BARABASI A L. Understanding individual human mobility patterns[J]. Nature, 2008, 453(7196): 779-782.
- [2] SONG C, QU Z, BLUMM N, et al. Limits of predictability in human mobility[J]. Science, 2010, 327(5968): 1018-1021.
- [3] SONG C, KOREN T, WANG P, et al. Modelling the scaling properties of human mobility[J]. Nature Physics, 2010, 6(10): 818-823.
- [4] LI Z, DING B, HAN J, et al. Mining periodic behaviors for moving objects[C]// Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2010: 1099-1108.
- [5] 陈佳,胡波,左小清,等.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.
- [6] ASHBROOK D, STARNER T. Using GPS to learn significant locations and predict movement across multiple users[J]. Personal and Ubiquitous Computing, 2003, 7(5): 275-286.
- [7] WHITE J, WELLS I. Extracting origin destination information from mobile phone data[C]// 11th International Conference on Road Transport Information and Control, 2002: 30-34.
- [8] CACERES N, WIDEBERG J P, BENITEZ F G. Deriving origin destination data from a mobile phone network[J]. Intelligent Transport Systems, IET, 2007, 1(1): 15-26.
- [9] IQBAL M S, CHOUDHURY C F, WANG P, et al. Development of origin-destination matrices using mobile phone call data[J]. Transportation Research Part C Emerging Technologies, 2014, 40(1): 63-74.
- [10] LIU F, JANSENS D, CUI J X, et al. Building a validation measure for activity-based transportation models based on mobile phone data[J]. Expert Systems with Applications, 2014, 41(14): 6174-6189.
- [11] PHITHAKKITNUKORN S, HORANONT T, LORENZO G D, et al. Activity-aware map: identifying human daily activity pattern using mobile phone data[C]// Proceedings of the First international conference on Human behavior understanding. Springer-Verlag, 2010: 14-25.
- [12] ISAACMAN S, BECKER R, CACERES R, et al. Identifying Important Places in People's Lives from Cellular Network Data[J]. Lecture Notes in Computer Science, 2011, 6696: 133-151.
- [13] TRAAG V A, BROWET A, CALABRESE F, et al. Social Event Detection in Massive Mobile Phone Data Using Probabilistic Location Inference[C]// Proceedings of the Third IEEE International Conference on Social Computing, 2011: 9-11.
- [14] QUERCIA D, LATHIA N, CALABRESE F, et al. Recommending social events from mobile phone location data [C]// Proceedings of the 10th International Conference on Data Mining (ICDM), 2010: 971-976.
- [15] CALABRESE F, COLONNA M, LOVISOLI P, et al. Real-Time Urban Monitoring Using Cell Phones: A Case

- Study In Rome[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(1): 141-151.
- [16] SOTO V, FRIAS-MARTINEZ V, VIRSEDA J, et al. Prediction of Socioeconomic Levels Using Cell Phone Records[J]. Lecture Notes in Computer Science, 2011, 6787: 377-388.
- [17] HONGYAN G, FASHENG L. Estimating freeway traffic measures from mobile phone location data[J]. European Journal of Operational Research, 2013, 229(1): 252-260.
- [18] 陆嘉恒. Hadoop 实战[M]. 第 2 版. 北京: 机械工业出版社, 2012: 85-329.
- [19] 孔扬鑫. 手机轨迹数据的人口流动分析[R]. 上海: 华东师范大学软件工程学院, 2015.
- [20] 章志刚. 面向海量手机轨迹数据的重要位置发现[R]. 上海: 华东师范大学软件工程学院, 2015.

(责任编辑 李万会)

(上接第 161 页)

- [12] PAILLIER P. PUBLIC-key cryptosystems based on composite degree residuosity classes[C]//Advances in cryptology-EUROCRYPT99. Berlin Heidelberg: Springer, 1999: 223-238.
- [13] ZHU H, MENG X, KOLLIOS G. Privacy Preserving Similarity Evaluation of Time Series Data[C]//EDBT, 2014: 499-510.
- [14] INAN A, KANTARCI OGLU M, BERTINO E, et al. A hybrid approach to private record linkage[C]//Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on. IEEE, 2008: 496-505.
- [15] HU H, XU J, REN C, et al. Processing private queries over untrusted data cloud through privacy homomorphism [C]//Data Engineering (ICDE), 2011 IEEE 27th International Conference on. IEEE, 2011: 601-612.
- [16] YAO A C C. How to generate and exchange secrets[C]//Foundations of Computer Science, 1986. 27th Annual Symposium on. IEEE, 1986: 162-167.
- [17] LINDELL Y, PINKAS B. A proof of security of Yao's protocol for two-party computation[J]. Journal of Cryptology, 2009, 22(2): 161-188.
- [18] KOLESNIKOV V, SADEGHI A R, SCHNEIDER T. Improved garbled circuit building blocks and applications to auctions and computing minima[M]//Cryptology and Network Security. Springer Berlin Heidelberg, 2009: 1-20.
- [19] KOLESNIKOV V, SCHNEIDER T. Improved garbled circuit: Free XOR gates and applications[M]//Automata, Languages and Programming. Berlin Heidelberg: Springer, 2008: 486-498.

(责任编辑 李万会)