计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2018,27(5):73−79 [doi: 10.15888/j.cnki.csa.006329]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于手机大数据的动态人口感知①

杨皓斐, 曹　仲, 李付琛

(北京交通大学 计算机与信息技术学院 交通数据分析与挖掘北京市重点实验室, 北京 100044)

摘　要: 随着通信市场迅速发展和手机的高普及率, 利用手机数据研究人类活动和城市规划发展成为可能. 本文主要工作是构建了大数据实时处理分析平台, 并基于此平台提出了一种利用手机数据感知城市人口分布的方法. 通过实验表明, 基于手机数据的动态人口感知能够反映实际城市人口分布, 对于城市交通监管、公共资源配置优化等方面具有重要意义.

关键词: 手机数据; 大数据; 人口感知; 城市规划; 位置服务

引用格式: 杨皓斐,曹仲,李付琛.基于手机大数据的动态人口感知.计算机系统应用,2018,27(5):73–79. http://www.c-s-a.org.cn/1003-3254/6329.html

Dynamic Population Perception Based on Mobile Phone Big Data

YANG Hao-Fei, CAO Zhong, LI Fu-Chen

(Beijing Key Lab of Traffic Data Analysis and Mining, School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

Abstract: With the rapid development of the communication market and the high penetration rate of mobile phones, it is possible to use mobile phone data to study the development of human activities and urban planning. The main work of this study is to build a big data real-time processing analysis platform, and based on this platform proposes a method to understand the urban population distribution by using of mobile data. Experiments show that dynamic population perception based on mobile phone signaling data can reflect the actual urban population distribution, which is of great significance to urban traffic regulation and public resource allocation optimization.

Key words: signaling data; big data; population perception; urban planning; location services

随着我国城市化进程的不断推进, 早期的城市规划不合理导致产业分布不均、经济发展不平衡的问题愈加凸显. 分析区域的人口数量以及时空特性[1,2], 对于城市发展政策的制定[3]、发展规划的调整具有着重要的意义. 过去几十年间, 国内外在相关方面都开展了统计分析工作, 并取得了一定的成果. 数据统计方式经历了从最初依靠人力的人口普查、到依靠红外、摇杆技术等技术测量方式, 再到近几年采用跨学科综合地理信息系统建模[4]的方式, 不断提高数据的精准度. 然而, 这些方法普遍存在着测量方式复杂, 数据获取难度大、更新过程慢、时效性低等缺点.

城市人口感知最早可以追溯到上世纪 30 年代, 沃斯在《作为一种生活方式的城市主义》中首次提到了昼夜人口数量区分[5]. 90 年代初期, 人口感知取得了迅速的发展. 在国外, Tobler W 等[6]提出了地理信息栅格化的方式, 将地理空间划分为等大小的栅格, 提高了人口分布的精度, 但是同时也削弱了与地理空间语意的结合. Linard C 等人[7]提出了一种融合人口普查数据和卫星数据的感知方法, 利用非洲地区的数据得到验证, 并提高了非洲地区人口感知分辨率. 在人口分辨率的基础上, 实现了人口分布中心性和可达性的分析.

Gaughan AE 等[8]在 Linard C 的基础上, 结合了土

① 基金项目: 教育部-中移动科研基金 (MCM20150513)
收稿时间: 2017-07-28; 修改时间: 2017-09-15; 采用时间: 2017-09-18; csa 在线出版时间: 2018-04-23

地利用率数据, 将分辨率提高至 100 m, 但是该方法采用的遥感数据和普查数据存在数据获取相对困难, 时效性低的特点. 在国内, 龙瀛等[9]以北京市一周的公交刷卡数据为基础, 分析了北京的职住关系和通勤出行, 识别出 20 万以上的人口出行, 但也只占全数据样本比例的 2.8%, 不能全面反映北京的人口情况. 郑宇等[10]采用用户历史移动数据和 POI 数据发觉城市功能区、寻找城市人口火山与黑洞. 总体而言, 上述方法都是基于抽样样本数据获取方式, 存在样本集合无法反应全集的问题, 而且呈现的结果时效性低.

近年来, 通信市场的迅速发展, 手机的高覆盖率, 为城市人口分布感知提供了一种新的途径. 手机用户在主叫、被叫、收发短信、开关机以及位置更新时, 运营商均可以记录包含着时空信息的数据内容. 这些手机数据存在着海量、实时的特性, 并且具有数据采集方便、数据样本覆盖全的特点, 使其可以广泛的应用在城市空间结构、建设环境评估、职住关系分类、交通通勤行为等众多研究领域, 同时也为动态人口感知提供了可能.

基于上述内容, 本文以北京市连续五天早 5 点到晚上 12 点采集的某运营商数据为基础, 结合 GIS 地理信息系统, 采用分布式消息中间件 Kafka 作为数据缓冲平台、Spark Streaming 作为实时处理方案及 HBase 作为数据存储平台, 以小时作为处理时间粒度单位, 尝试感知北京市动态人口分布时空特性, 为城市交通和城市规划的后续研究提供参考借鉴.

1 问题定义

当移动终端在蜂窝网络中发生基站小区切换时, 基站会记录用户位置更新相应的交互信息日志 (称为位置更新数据), 并汇总到运营商数据中心. 基于位置更新数据, 建立本文的数据模型.

定义 1. 原始定位点 OPP (Original Positioning Point). 表示位置更新数据经过降噪格式化处理之后的有效数据单元, 包含了三个基本信息: OPP=(ID, P, T), 其中 ID 表示用户唯一性标识, P=(LAC, CELLID) 分别为基站大区和小区号, T 为交互发生的有效时间戳.

如图 1 所示, 用户在单位时间内共发生 8 次基站小区切换, 形成 8 条有效记录, 经过转换形成 8 个 OPP 单元, 每一个单元均包含了位置小区以及时间信息.

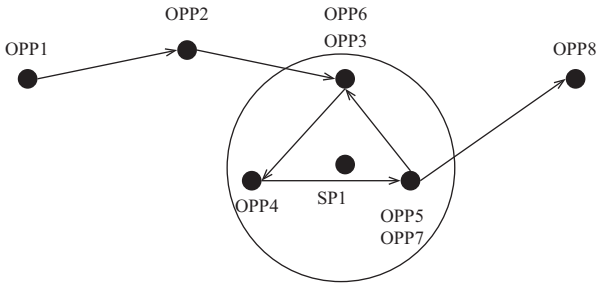

图 1 用户位置序列

定义 2. 移动轨迹 MT (Movement Trajectory). 一条移动轨迹由同一用户在单位时间内的原始定位点 OPP 按照时间排序组成, MT={OPP_i}.

在图 1 所示情况中, OPP 位置按照位置更新的有效时间作为主键进行升序排序, MT={OPP_1, OPP_2, OPP_3, OPP_4, OPP_5, OPP_6, OPP_7, OPP_8}.

定义 3. 用户驻留点 SP (Stay Point). 用户驻留点表示将用户的移动轨迹按照时间和空间两个维度聚类之后产生的位置信息点, SP=(P, T), P=(LAC, CELLID) 为聚类之后形成的新的聚类中心对应的基站小区编号, T 为聚类之后形成的新的有效时间信息点.

如图 1 所示, 用户的 8 个 OPP 中, 有五个点满足聚类条件, 形成了新的聚类中心 SP1.

定义 4. 用户位置空间 UGA (Geometry Area). 用户位置空间表示将基站按照泰森多边形算法[11]划分形成基站覆盖面, 之后将用户的驻留位置点信息 SP 根据基站小区编号位置映射, 形成用户位置空间. UGA=(ID, Geometry), 其中 ID 标识唯一用户, Geometry 为包含基站形成泰森多边形区域的地理语义信息.

2 大数据平台

大数据平台主要为本应用提供海量数据的采集、计算、存储、对外服务等基础支撑性功能. 在数据采集方面需要提供全面的数据清洗能力, 实现数据的无缝抽取、转换和加载. 在数据分析方面需要能够具备海量数据的实时处理分析能力和高度的可扩展性. 在数据存储方面需要具备低成本、高扩展、及时响应的存储性要求. 对外服务方面需要具有全面的应用服务接口, 对外提供交互友好的可视化效果.

2.1 技术架构

大数据分析处理平台采用分层架构, 利用当前主流的 Hadoop 生态圈产品. 平台技术架构如图 2 所示.

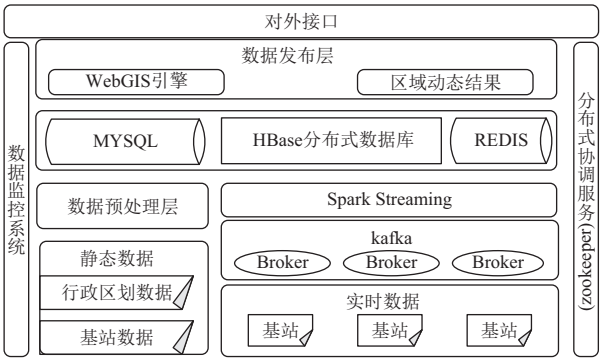

图 2 分布式处理平台架构

2.2 数据采集层

数据采用层要能够接入多种异构数据源, 包括了静态数据文件以及实时流数据. 静态数据文件如行政区划和基站等地理信息数据采用自定义工具实现数据的存储转换. 实时流数据在不同的时间段面对的数据量大不同, 为保证后台实时处理的效率及性能的动态收缩, 采用 Kafka 作为分布式消息队列, 减轻后后台业务处理压力.

2.3 数据存储层

传统关系型数据库对于小量结构化数据有着高速查询、分析处理的能力, 但是无法满足海量异构数据的存储分析性能要求. HBase 分布式数据库诞生之初就是为了解决海量结构化和非结构化数据的存储以及分析需求. 结合关系型数据库 MySQL 和分布式数据库 HBase 构建数据存储层. 将静态、变更周期长的数据存储在传统关系型数据库中, 而将海量实时流数据存储在分布式数据库中.

2.4 数据处理层

数据处理层需要具备海量数据实时处理分析的能力, 并在数据爆发时具有高可伸缩性. 采用构建在 Spark 基础之上的 Spark Streaming 作为分布式实时处理架构, 其基于内存的高速执行引擎具有使其能够达到秒级响应并具备高效的容错性.

2.5 数据监控层

数据监控层主要用于监控集群的运行状态以及对数据的变更做出及时的响应. 一方面采用开源 Zookeeper 组件保存部分应用状态, 如 HBase 元信息和 Kafka 偏移量控制; 另一方面采用自定义监控工具实现对静态数据变更的监控及响应, 当基础数据如基站数据或者行政区划数据发生变更时, 及时更新数据库中的状态.

2.6 对外服务层

对外服务层采用 B/S 架构, 提供数据结构调用, 并采用 HTML5+JaSvaScript 实现可视化效果, 后台采用 Tomcat 作为 Web 应用服务发布层, 并利用 GeoServer 发布应用.

3 研究方法

3.1 数据预处理

用户位置更新数据存在海量、实时等大数据的优点, 但是由于人为因素或者其他客观因素的存在, 原始数据需要经过一定的数据清洗优化措施才能供分析使用.

首先, 存在基站异常数据, 用户连接基站信息缺失以及连接基站小区不存在等情况.

其次, 存在非真实用户数据, 针对数据的统计分析工作发现, 存在用户位置更新频率异常高, 超出人类动力学[12]活动范围, 针对此类用户, 通过设定阈值方式, 剔除不合理用户数据.

最后, 存在切换点异常的用户, 这类用户通过建立用户的个体轨迹方式, 剔除异常值.

针对以上异常数据内容, 提出一种基于时间和空间两个维度的数据预处理方案以减少这样的低质量数据, 消除异常数据所带来的影响.

数据预处理整体算法描述如算法 1.

算法1. 位置更新数据预处理

输入: 原始记录, 时间阈值, 距离阈值

输出: 用户移动轨迹

1) 将从消息源读到的数据格式化为OPP对象, 删除不包含OPP基本信息的记录.

2) 将OPP记录按照用户ID为主键进行合并, 再按照时间为主键进行二次排序, 形成用户的基站小区轨迹序列MT.

3) 从用户移动轨迹MT的第二个点开始遍历, 计算当前点和前一个的时间差, 如果时间差小于时间阈值参数, 计算两点之间的欧式距离, 如果欧式距离大于距离阈值, 移除当前点, 开始下一个点的计算.

4) 结束用户轨迹的遍历之后, 返回每个用户的有效轨迹序列.

3.2 用户驻留点模型

经过预处理之后的用户移动轨迹反应的是用户在单位时间内空间位置移动序列, 但是这样的移动序列并没有和地理空间相结合起来. 考虑时间和空间两个因素, 采用密度聚类 DBSCAN[13]方式, 结合地理语义信息, 生成用户在单位时间内的驻留点模型, 其中聚类需要满足以下时间参数 T 和距离参数 D:

$$\begin{aligned} &OPP_i.T - OPP_{i-1}.T \le T \\ &Dis(OPP_i.P - OPP_{i-1}.P) \le D \end{aligned} \quad (1)$$

$Dis(OPP_i.P\text{-}OPP_{i-1}.P)$ 表示相邻 OPP 点在道路网中采用最短路径搜索算法 pgRouting[14]搜索出来的道路长度之和, 将满足条件的点进行聚类, 形成新的聚类位置中心为聚类范围点的中心位置、聚类时间中心为聚类点的时间和. 不满足聚类条件的点生成单独的驻留点. 整体算法描述如算法 2.

算法2. 用户驻留点聚类

输入: 用户移动轨迹, 时间阈值, 距离阈值

输出: 用户单位时间内的驻留小区

1) 读取用户移动轨迹并将其转换为聚类算法的标准时空格式List.

2) 从List的第一个点开始, 形成第一个聚类中心, 并加入用户驻留点序列, 聚类中心时间点为第一个点的值, 遍历第二个点, 如果满足聚类条件, 更新新的聚类位置中心为两个点相对位置中心, 聚类时间点为两个点相对分钟数之和. 如果第二个点不满足聚类条件, 则将第二个点表示成另一个聚类中心, 开始第三个点的遍历, 直到List循环结束.

3) 将用户的驻留点序列按照时间长短排序, 选择时间最长的为用户当前时间范围内的驻留点, 更新用户驻留点为覆盖用户当前驻留位置的基站小区编号.

4) 返回用户的当前时间单位内的驻留小区编号.

3.3 用户位置空间模型

用户位置空间模型将用户个体映射到定义区域网格中, 实现不同标准的人口数据在地理尺度上的收缩. 当前的研究方法主要有: Meng 等[15]提出的线性面积权重法, 是一种使用较为广泛的社会经济数据空间化处理方法, 在假设同种类型的人均面积权重的前提下, 根据目标区内各个源区所占面积的百分比确定目标区的属性值.

$$P_j = \sum_{i=1} n\left(\frac{S^*_{ji}}{S^*_i}P_i\right)P_i ij \quad (2)$$

其中 P_j 表示网格 j 中的人口数, P_i 表示第 i 类用地内总人口数, S_{ji}^*表示网格 j 的中 i 的面积权重, S_i^*表示用地的总面积权重. 但是这种方法要综合考虑地理空间各种属性数据, 获取数据难度较大. Deville 等[16]采用了非线性方程 $\rho_c=\alpha(\sigma_c)^\beta$ 表征手机活跃度和人口密度之间的关系. 其中, ρ_c 表示小区人口密度, σ_c 表示小区手机用户密度. 并且, 通过实验表明非线性方程有很好的拟合效果.

综合考虑上述两种模型, 考虑某运营商在通信市场上的无差别市场占有率, 提出本文人口分布感知方法, 采用移动端区域面积模型反应实际人口指标. 公式如下:

$$N_i = \alpha\left(\frac{N_{ic}}{A_i}\right)^\beta A_i \quad (3)$$

N_i 代表当前基站实际用户数量, N_{ic} 代表第 i 个基站的手机活跃用户数. 对公式 (3) 进行变换, 得到 $N_i=\alpha(N_{ic})^\beta(A_i)^{1-\beta}$, 然后对全北京市基站人口数据做累加运算, 得全北京的城区的用户数据:

$$Y = \sum_{i=1}^{n} N_i = \alpha\sum_{i=1}^{n}(N_{ic})^\beta(N_{ic})^{1-\beta} \quad (4)$$

其中 Y 代表全北京移动用户数. 由于$\sum N_i$ 已知, 为求解参数 α、β, 将模型可以转化为求解函数 $f(\alpha, \beta)$ 最小化问题.

$$f(\alpha,\beta) = \sum_{i=1}^{n} N_i - \alpha\sum_{i=1}^{n}(N_{ic})^\beta(A_{ic})^{1-\beta} \quad (5)$$

考虑 $f(\alpha, \beta)$>=0, 改进公式:

$$f(\alpha,\beta) = \left(\sum_{i=1}^{n} N_i - \alpha\sum_{i=1}^{n}(N_{ic})^\beta(A_{ic})^{1-\beta}\right)^2 (f(\alpha,\beta) \ge 0) \quad (6)$$

即求 $f(\alpha, \beta)$ 的最小化问题.

采用梯度下降算法求解最优值, 对五天以小时为单位数据集分别求解 α、β 最优值, 结果如图 3 所示, 图 2 中较高的折线表示 α 从早上 7 点到晚上 21 点的变化曲线. 较低的折线表示 β 随时间变化曲线.

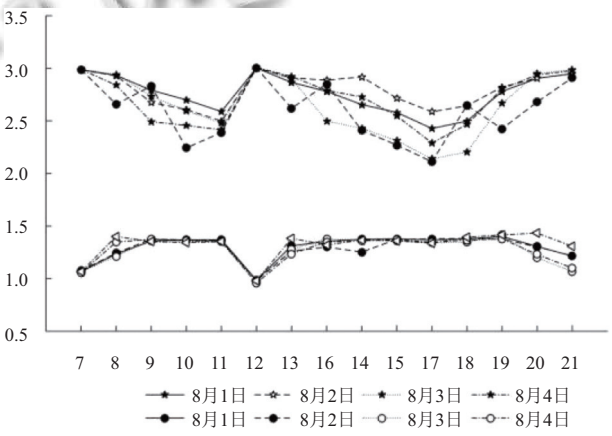

图 3 α、β 随时间变化曲线

从图中可以看出, α、β 参数随时间基本保持稳定, 对 α、β 分别求 5 天不同时刻的几何平均值, 得 α、β 值如表 1 所示.

表 1 α、β 随时间变化值

整点时刻	α	β
7	2.987	1.066
8	2.860	1.287
9	2.705	1.366
10	2.520	1.361
11	2.475	1.358
12	3.005	0.977
13	2.843	1.293
14	2.761	1.341
15	2.628	1.344
16	2.485	1.369
17	2.311	1.354
18	2.492	1.373
19	2.700	1.397
20	2.876	1.294
21	2.954	1.181

以 12 点为例, 取 α=3.005, β=0.977 将结果带入公式, 求得基站 i 对应的真实用户数 $N_i=3.055(N_{ic})^{0.977}(A_i)^{0.003}$.

3.4 实时处理

以小时作为数据处理时间单位, 用户在单位时间内的移动序列形成了用户的移动轨迹. 不同用户的移动轨迹在时空维度上都呈现出异构性, 同一用户不同时间段的移动轨迹、活动频率也不相同. 为分析用户在单位时间内的活动轨迹, 由于在时间维度上不同时间段内用户发生位置更新的频率不一致, 采用 Kafka 作为消息缓冲中间件; 面对海量的数据集, 为提高算法的横向可扩展性以及提高时间效率, 采用 Spark Streaming 作为分布式处理平台, 并以 HBase 作为海量数据存储平台, 最后的分析结果存入传统关系型数据库实现数据可视化.

从 Kafka 中读取数据的周期为 1 小时, 然后在 map 阶段按照读取到的用户 ID 为主键进行元组映射; 以用户 ID 为 key 进行 Reduce, 将同一个用户的移动轨迹序列在同一个计算节点上进行分析, 并行处理分析用户数据, 在同一用户的数据传输到同一个计算节点之后, 首先进行数据预处理, 再分析用户的驻留点模型, 最后分析用户的停留基站小区, 用于空间模型分析.

算法描述如算法 3.

算法3. 实时处理算法

输入: 原始位置更新数据

输出: 基站位置小区用户数

1) 将从通信网络中采集原始位置更新数据格式化存储到Kafka中.
2) 按照读取周期, 从Kafka中读取原始数据, 按照用户ID为主键进行分区.
3) 调用数据预处理算法, 生成用户的有效活动轨迹.
4) 调用驻留点算法将用户的活动轨迹数据进行聚类分析, 生成用户单位时间内的有效驻留地.
5) 按照位置空间模型将用户驻留地结果进行转换, 按照驻留小区为主键进行统计. 再按照时间伸缩比例进行人口数据伸缩.

4 实验结果分析及可视化

4.1 数据集描述

本文中使用的实验数据来自某运营商提供的北京 2016 年 8 月 1 日到 5 日连续 5 天从凌晨 5 点到晚上 12 点的包含了用户通话位置切换和正常位置更新信息两类日志数据. 如图 4 所示, 当用户发生位置区切换时, 与移动端进行连接的通信基站和通信建立连接时刻将会被记录下来. 记录信息如表 2 所示, 每条数据包含了用户身份标示、记录时间、连接基站等信息.

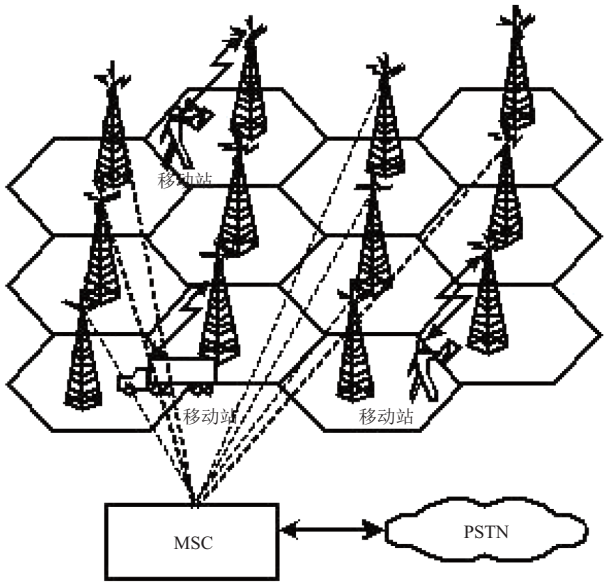

图 4 蜂窝小区

表 2 位置更新记录属性

序号	属性英文名称	属性描述	字段类型
01	time	事件发生时刻	Datetime
02	Id	用户唯一性表示	String
05	Dlac	发生基站切换的目的位置区	Int
06	Dci	发生基站切换的目的小区	Int

同期, 运营商提供了切换记录对应的基站位置小区数据, 如表 3 所示, 其中包含了基站的位置区、位置小区、经纬度坐标等基本信息. 如图 5 所示, 一个基站

有一般有 1~3 个扇区和多个位置小区. 对基站按照经纬度进行地理位置的合并, 形成有效基站数据.

表 3 基站位置小区属性

序号	属性英文名称	属性描述	字段类型
01	Lac	位置区	Int
02	Ci	位置小区	Int
03	Lon	基站所在地理位置的经度	Double
04	Lat	基站所在地理位置的维度	Double

验证原始数据集来自由北京统计局发布的《北京统计年鉴 2016》中第六次人口普查数据中的本地户籍常驻人口和常驻外来人口数据以及北京市旅游发展委员会公布的北京旅游统计数据.

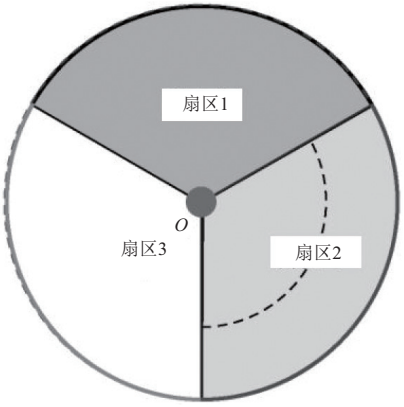

图 5 基站信息

4.2 结果分析

本文利用第 2 节提到的分布式算法对数据进行了分析处理, 结果如图 6 所示, 原始统计数据表示将每天的位置更新数据进行统计之后得到了人口数量, 分析结果数据表示将最初的人口数据按照计算法进行处理分析之后得到的每天的人口总数, 验证数据表示将北京的原始验证数据乘以某运营商的市场占比后得到的人口数据. 从实验结果中得出, 人口感知的误差率保持在 30%~10% 之间, 平均误差率为 21.5%.

4.3 数据可视化

数据可视化使用 OpenStreets Map 作为后台 GIS 地理信息系统基础服务平台. OpenStreets Map 是一款由网络大众共同打造的免费开源、可编辑的地图服务, 它利用公众集体的力量和无偿的贡献来改善地图相关的地理数据. 采用 GeoServer 作为 Web 应用服务器, GeoServer 是采用 J2EE 规范编写的用来发布地图数据的服务器, 支持多种数据源, 用户能够简单便捷的发布地图数据, 并可以对数据进行删除、更新, 添加等操作. 最后采用 Html+JavaScript 组合进行前端可视化, 其中采用 OpenLayers 作为地图加载引擎, OpenLayers 是专门为 WebGIS 开发提供的开源 JavaScript 类库.

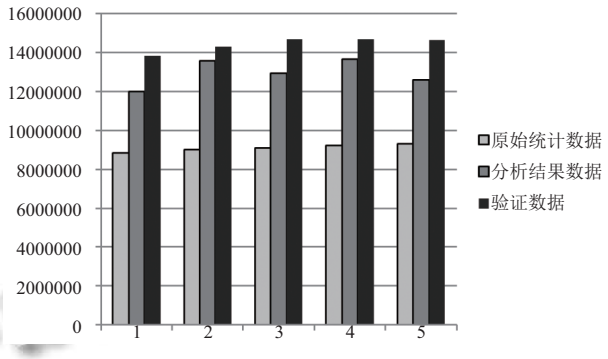

图 6 实验结果对比

采用 B/S 架构进行数据可视化展示, 前后台交互采用 SpringMVC 技术, 前台通过参数向后台发送数据请求, 后台进行相应的查询分析之后将数据返回给前台, 然后 OpenLayer 进行数据渲染.

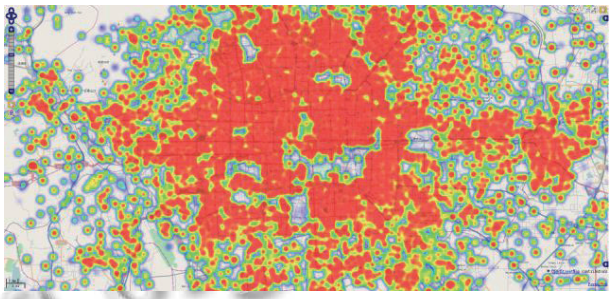

图 7 可视化展示

5 总结

本文从城市感知出发, 利用白天连续时段的位置更新信令数据, 分别从时间维度和地理空间维度量化分析了人口数据. 同时, 为减小个体数据低质量的影响, 对个体位置信息进行了分析挖掘, 反映真实用户信息, 以降低高频用户造成的数据偏差.

实验结果表明, 在以小时数据量为单位的情况下, 该平台在面对海量数据, 依旧能够保证性能与分析结果的准确性. 本文提出的大数据城市人口感知分析方法, 在动态人口感知中, 取得了一定的成果. 但是, 实验数据缺少了夜晚信息数据, 没有反映全天 24 h 的城市人口分布, 处理单位维持在小时粒度级别, 未来可以考虑获取全天数据和进一步提高数据的分辨率.

参考文献

1 Wang L, Yang YZ, Feng ZM, *et al*. Prediction of China's population in 2020 and 2030 on county scale. Geographical Research, 2014, 33(2): 310–322.

2 Pan Q, Jin XB, Zhou YK. Population change and spatiotemporal distribution of China in recent 300 years. Geographical Research, 2013, 32(7): 1291–1302.

3 Checchi F, Stewart BT, Palmer JJ, *et al*. Validity and feasibility of a satellite imagery-based method for rapid estimation of displaced populations. International Journal of Health Geographics, 2013, (12): 4.

4 卓莉, 黄信锐, 陶海燕, 等. 基于多智能体模型与建筑物信息的高空间分辨率人口分布模拟. 地理研究, 2014, 33(3): 520–531. [doi: 10.11821/dlyj201403011]

5 Wirth L. Urbanism as a way of life. American Journal of Sociology, 1938, 44(1): 1–24. [doi: 10.1086/217913]

6 Tobler W, Deichmann U, Gottsegen J, *et al*. World population in a grid of spherical quadrilaterals. International Journal of Population Geography, 1997, 3(3): 203–225. [doi: 10.1002/(ISSN)1099-1220]

7 Linard C, Gilbert M, Snow RW, *et al*. Population distribution, settlement patterns and accessibility across Africa in 2010. PLoS One, 2012, 7(2): e31743. [doi: 10.1371/journal.pone.0031743]

8 Gaughan AE, Stevens FR, Linard C, *et al*. High resolution population distribution maps for Southeast Asia in 2010 and 2015. PLoS One, 2013, 8(2): e55882. [doi: 10.1371/journal.pone.0055882]

9 龙瀛, 张宇, 崔承印. 利用公交刷卡数据分析北京职住关系和通勤出行. 地理学报, 2012, 67(10): 1339–1352. [doi: 10.11821/xb201210005]

10 Hong L, Zheng Y, Yung D, *et al*. Detecting urban black holes based on human mobility data. Proceedings of the 23rd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. Seattle, WA, USA. 2015. 35.

11 Fu TL, Yin XT, Zhang Y. Voronoi algorithm model and the realization of its program. Computer Simulation, 2006, 23(10): 89–91, 128.

12 李楠楠, 周涛, 张宁. 人类动力学基本概念与实证分析. 复杂系统与复杂性科学, 2008, 5(2): 15–24.

13 Ester M, Kriegel HP, Sander J, *et al*. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining. Portland, OR, USA. 1996. 226–231.

14 Zhang LJ, He XH. Route search base on pgRouting. In: Wu YW, ed. Software Engineering and Knowledge Engineering: Theory and Practice. Berlin Heidelberg: Springer, 2012. 1003–1007.

15 Meng B, Wang JF. A review on the methodology of scaling with geo-data. Acta Geographica Sinica, 2005, 60(2): 277–288.

16 Deville P, Linard C, Martin S, *et al*. Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences of the United States of America, 2014, 111(45): 15888–15893. [doi: 10.1073/pnas.1408439111]