

引用格式: 邬群勇, 苏克云, 邹智杰. 基于 MapReduce 的海量公交乘客 OD 并行推算方法[J]. 地球信息科学学报, 2018, 20(5): 647-655. [Wu Q Y, Su K Y, Zou Z J. A mapreduce-based method for parallel calculation of bus passengers origin and destination from massive transit data [J]. Journal of Geo-information Science, 2018, 20(5): 647-655.] DOI: 10.12082/dqxxkx.2018.170374

基于 MapReduce 的海量公交乘客 OD 并行推算方法

邬群勇, 苏克云, 邹智杰

1. 福州大学地理空间信息技术国家地方联合工程研究中心, 福州 350002; 2. 空间数据挖掘与信息共享教育部重点实验室, 福州 350002

A MapReduce-based Method for Parallel Calculation of Bus Passenger Origin and Destination from Massive Transit Data

WU Qunyong*, SU Keyun, ZOU Zhijie

1. National & Local Joint Engineering Research Center of Geo-spatial Information Technology, Fuzhou University, Fuzhou 350002, China;
2. Key Laboratory of Spatial Data Mining & Information Sharing of MOE, Fuzhou 350002, China

Abstract: Bus passengers' origin and destinations (OD) can truly reflect travel characteristics and demands of residents, which is an important basic data for bus system evaluation, scheduling and route optimization, with significantly practical value in urban planning. Existing OD estimation methods are mostly applied to a small amount of bus data, which cannot directly and rapidly calculate mass transit passenger OD. In order to solve these problems, a parallel method for calculation of massive transit passengers' origin and destinations based on MapReduce is investigated. Firstly, database migration tool was applied to transfer massive bus data stored in relational database to HBase. Secondly, MapReduce parallel computing framework was introduced to divide the IC card data into multiple Map tasks in the light of region numbers in HBase to calculate origins. The origins are grouped and stored into HDFS by user in the Reduce function. Thirdly, the destinations are estimated by origins in parallel which are divided into multiple Map tasks according to block numbers stored in HDFS. According to the travel record of each passenger, destinations can be accurately calculated by the means of public transit chain method and history similarity. In the end, taking IC card data and GPS bus data in Xiamen from June 13 to 26, 2015 as the example, which has 295 bus lines, 16 879 661 bus records, and 14 410 058 complete OD pairs which accounted for 78.9% of IC card data. Comparing with the traditional method, the computational efficiency has substantially improved. The results illustrate that the parallel method can not only calculate bus passenger OD accurately, but also has higher computational efficiency.

Key words: massive transit data; public transit origin and destination; MapReduce; public transit trip chain; travel rule

*Corresponding author: WU Qunyong, E-mail: qywu@fzu.edu.cn

收稿日期 2017-08-10; 修回日期: 2018-03-07.

基金项目: 国家自然科学基金项目(41471333); 中央引导地方科技发展专项项目(2017L3012)。[**Foundation items:** National Natural Science Foundation of China, No.41471333; The Central Guided Local Development of Science and Technology Project, No.2017L3012.]

*通讯作者 邬群勇(1973-), 男, 山东诸城人, 博士, 研究员, 研究方向为时空大数据分析、地理信息服务。E-mail: qywu@fzu.edu.cn

摘要 公交乘客出行OD能够反映居民出行特征和出行需求,是进行公交系统评价、调度和线路优化的重要基础数据,对城市规划具有重要的实用价值。现有公交OD推算方法多适用于少量公交数据,无法直接快速地推算海量公交乘客出行OD,因此本文提出了一种基于MapReduce的海量公交乘客OD并行推算方法。首先将公交数据从关系型数据库迁移至HBase数据库;接着利用MapReduce并行计算框架,根据HBase中IC卡数据的Region数量分成多个map任务,每个map任务中Map函数计算上车站点,Reduce函数将上车站点以用户为单位进行归并输出到HDFS;然后在上车记录数据的基础上,根据HDFS存储的块数量分成多个map任务,针对每个乘客的出行记录,综合考虑出行链方法和历史相似出行行为规律实现对公交乘客下车站点较为精确的推算。最后以厦门2015年6月13日至26日的IC卡数据和公交车辆GPS数据进行实例分析,共计算出295条公交线路,16 879 661条上车记录,14 410 058条完整OD记录,占IC卡数据的78.9%,计算效率相比传统方法有较大幅度提升。结果表明:该方法不仅可以较为准确地推算公交乘客上下车站点,而且计算效率较高。

关键词 海量公交数据;公交OD;MapReduce;公交出行链;出行规律

1 引言

随着智能公交系统的发展与普及,智能公交系统收费、车辆监控等系统中积累了海量的IC卡数据和公交车辆GPS数据,蕴藏着真实、全面的公交客流信息^[1-2]。如何利用大数据技术和数据挖掘技术从这些海量数据中快速获取真实全面的公交客流OD(Origin-Destination)信息,发现其中隐含的公共交通模式及规则,获得高层的、潜在的规律是智能公交系统建设的重点,也是研究的热点问题。

目前我国的公交收费大都是一票制的,即只需上车刷卡,下车无需刷卡。在缺乏准确定位信息的情况下,通过对IC卡数据的刷卡时间间隔进行聚类,将聚类结果结合公交车辆调度信息和线路站点信息来推算公交乘客的上车站点^[3-6]。随着车辆GPS定位技术的发展与普及,公交车辆实时的位置信息以及其他状态信息更容易获得,研究者更青睐于基于公交车辆GPS数据和IC卡数据来推算上车站点,利用乘客在公交线间的换乘信息以及地铁和公交线间的换乘信息来推算公交乘客的上车站点^[7-8]。而更多的是基于IC卡刷卡时间和公交车辆到离站时间,建立一定的匹配约束规则来推算上车站点^[9-14]。对于下车站点的推算,根据公交乘客的出行特征主要有3种推算方法:基于公交站点的站点吸引权法、基于单个乘客的连续出行链方法和基于多源数据通勤出行方法。基于公交站点的站点吸引权法是指同时考虑居民公交出行距离近似符合泊松分布规律、受站点周边用地性质影响以及该站点上下车人数吸引强度3个影响因子来确定乘客出行在每个站点的下车概率,推算乘客的下车站点^[15-16]。基于单个乘客的出行链方法是在连续公交出行假设的基础上推算乘客的下车站点^[17-21],但连续公交出行只是公交出行一部分,有的学者在出行链

的基础上结合站点吸引权法来推算下车站点^[22]。基于多源数据通勤出行方法是在多种数据源的条件下追踪乘客的出行及换乘记录推算下车站点^[23-24]。

上述研究方法大多是利用一条或者几条公交线路的刷卡数据来推算乘客的上下车点。随着大数据和并行技术的发展,为海量公交数据OD推算提供了一种新的解决方案。本文提出一种基于MapReduce的海量公交乘客OD并行推算方法,这种方法运用大数据领域目前比较成熟的MapReduce并行计算框架、HBase非关系型数据库,融合连续出行链方法和居民出行特征较为准确地推算乘客的上下车站点,真实全面地反映居民公交出行的客流信息,并以厦门市的公交数据验证了所提出方法的实用性。

2 数据预处理与分区存储

本文数据主要有厦门IC卡数据、公交车辆GPS数据、公交车辆基本信息数据和厦门市公交线路站点数据。公交IC卡原始数据包括运营公司、数据库时间(刷卡记录存储到服务器时间)、IC卡编号、刷卡时间、刷卡金额、卡类型、线路编号、IC卡设备编号、公交车辆编号等字段。原始公交车辆GPS数据有22个字段,不仅记录车辆的实时位置信息,还记录车辆到达与离开站点的相关信息。在进行公交OD推算之前需要对原始数据进行预处理,删除冗余字段,保留所需字段信息,并转换成所需的数据格式。本文研究选取IC卡数据的IC卡编号、刷卡时间、线路编号、公交车辆编号4个字段属性,选取公交车辆GPS数据的GPS设备编号、GPS时间、纬度、经度、行驶方向、进出站标志、站点编号、线路编号8个字段属性。

本文使用HBase数据库来存储公交数据,公交

IC卡数据以IC卡编号和刷卡时间为组合行键,分别将IC卡编号、刷卡时间、线路编号、公交车车辆编号添加到指定列簇,如表1所示。公交车车辆GPS数据以GPS设备编号和GPS时间为组合行键,分别将GPS设备编号、线路编号、行驶方向、GPS时间、进出站标志、站点编号、经度、纬度添加到指定的列簇,如表2所示。HBase数据库根据IC卡数据表中行键的值进行分区,每个行键区间构成一个分区(Region),共有52个Region,每个Region表示一定数量IC卡号的刷卡记录。公交车车辆GPS数据表和IC卡数据表一样按照行键值被分成184个Region,每个Region表示一定数量公交车车辆的GPS数据。

表1 IC卡数据结构

Tab. 1 Data structure of IC card

行键	列簇
IC卡编号:刷卡时间	IC卡编号 刷卡时间 线路编号 公交车车辆编号

表2 公交车车辆GPS数据结构

Tab. 2 Data structure of bus GPS

行键	列簇
GPS设备编号:GPS时间	GPS设备编号 线路编号 行驶方向 GPS时间 进出站标志 站点编号 经度 纬度

公交线路站点数据包括公交线路信息及其所属站点的经纬度信息,本文将公交线路站点数据以公交线路编号为键,以该条线路所属的所有站点信息为值,存储到HDFS。公交车车辆信息包括车辆编号与GPS设备编号的对应关系,以公交车车辆编号为键,以对应的GPS设备编号为值存储到HDFS。

3 公交乘客上车站点并行推算方法

通常来讲,公交乘客是在公交车车辆进站停稳后依次刷卡上车,车辆在乘客全部上车后启动离开公交站台,公交乘客的刷卡时间应位于公交车车辆的进出站时间之内。但是实际的公交乘车情况比较复杂,尤其是上下班高峰期,出现多辆公交车排队进站时就上下客的情况,导致部分乘客的刷卡时间早于车辆进站时间。也可能由于车内拥挤,出现一些乘客先上车在车辆出站后刷卡的现象,导致一部分乘客的刷卡时间晚于车辆出站时间。为了提高推算公交乘客上车站点的准确率,为后续推算下车站点提供保障,本文采用基于相邻进站时间的时间匹配方法来推算公交乘客的上车站点。基本思想是:公交乘客的刷卡时间如果位于公交车车辆相邻进站

时间之内,计算乘客刷卡时间与出站时间差的绝对值和刷卡时间与后进站时间差的绝对值,比较这二者时间差,比较结果将对应着不同的上车站点,刷卡时间和车辆进出站时间关系如图1所示。

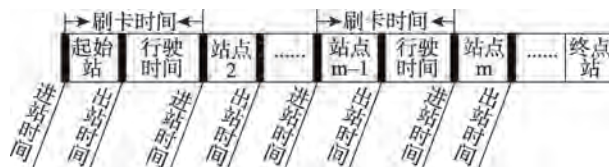


图1 乘客刷卡时间与车辆进出站时间关系示意图

Fig. 1 The relationship between boarding time and arrival-departure time

MapReduce模型的核心是Map函数和Reduce函数,Map函数主要是读入一组键值对数据,经过一系列变换处理,然后映射到一组新的键值对并输出;Reduce函数主要任务是归并,将Map函数输出的一系列具有相同键的键值对以某种方式组合起来,输出处理后的键值对数据。基于HBase数据库进行MapReduce并行计算时,一个map任务对应一个Region,根据IC卡数据的Region数量将IC卡数据划分成多个map任务并行计算上车站点。在Map函数中实现上车站点的推算,Reduce函数负责实现整理、归并Map函数的推算结果并输出到HDFS存储,公交乘客上车站点并行推算方法如图2所示。Map函数中推算上车站点的具体流程如图3所示,图中T为乘客该次乘车的刷卡时间, a 和 b 为时间阈值, $T-a < T < T+b$ 表示刷卡时间前后一定的时间范围。 S 表示从车辆GPS数据表中获取同一辆车基于刷卡时间一定时间范围内的车辆进出站记录。 Δt_1 表示刷卡时间与出站时间的时间差绝对值, Δt_2 表示刷卡时间与第二次进站时间的时间差绝对值。

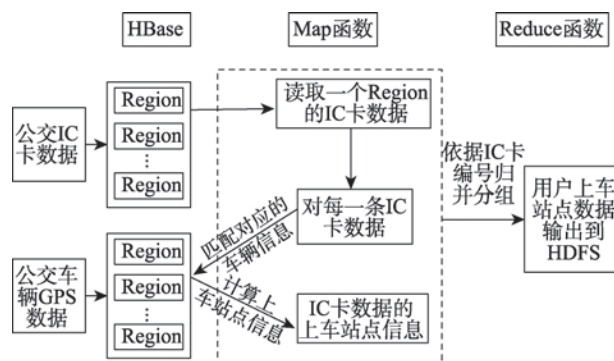


图2 上车站点并行推算流程

Fig. 2 The parallel estimating flow of bus passengers boarding station

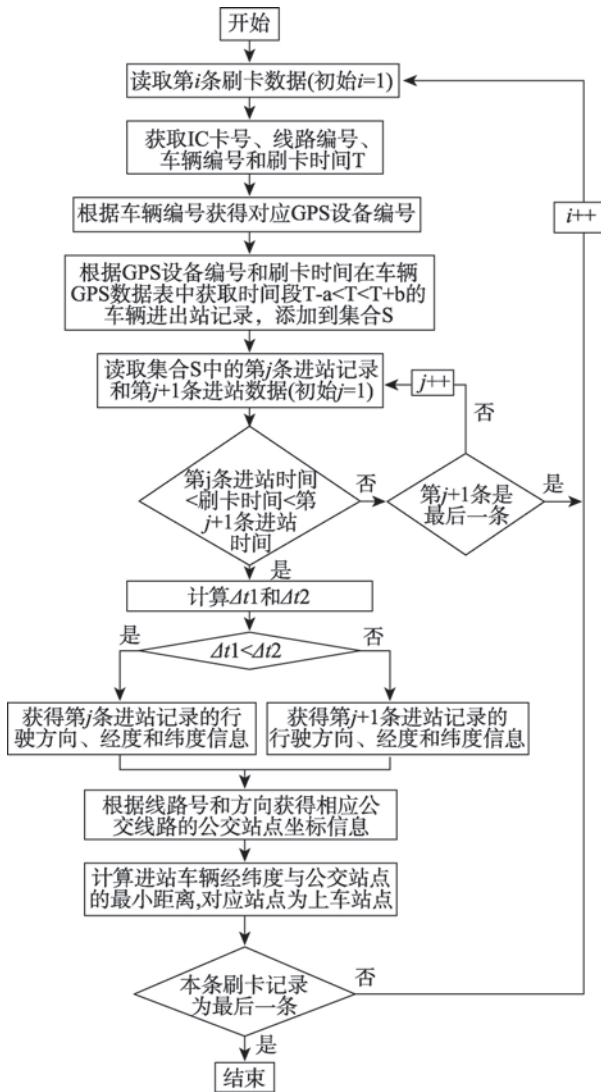


图3 Map函数中推算上车站点流程

Fig. 3 The estimating flow in Map function of bus passengers boarding station

针对Map函数中输出的上车记录,在Reduce函数中以IC卡号为键,以该用户按时间排序的所有上车记录为值,归并输出到HDFS,作为后续下车站点推算的数据基础。

4 公交乘客下车站点并行推算

本文乘客下车站点推算方法只针对一票制的公交刷卡数据,未考虑其它交通(如BRT、地铁等)刷卡记录。在通常的公交出行中,普遍存在一种代刷卡的现象,即一张IC卡在同一辆公交车出现连续两次或者多次的刷卡记录(连续定义为同一张IC卡的相邻刷卡时间小于一定的时间阈值)。本文假设

这些代刷记录都具有相同的出行目的地,即具有相同的下车站点。居民的公交出行行为根据出行链方法可以分为连续性公交出行链和非连续性公交出行链,如图4所示。图4(a)表示连续性公交出行链,有3个假设条件:①连续两次公交出行之间不使用其他交通工具。②下一次公交出行的起点是上一次公交出行的终点或者终点可接受步行换乘距离之内的站点。③一天中的最后一次出行终点为当天的第一次出行起点或者第二天首次出行的起点。而非连续性公交出行链如图4(b)所示,其中实线为公交出行,虚线为其他方式出行。从图4可以看出,非连续性公交出行包含以下3种情形:①用户乘坐公交连续出行,但是最后一次公交出行并未返回当天的起点;②用户在换乘时使用其他方式,并且最后一次出行并未返回当天的起点;③用户在换乘时使用其他方式,但最后一次出行返回当天的起点。所有的非连续性公交出行均由这3种基本情形自由组合而成。

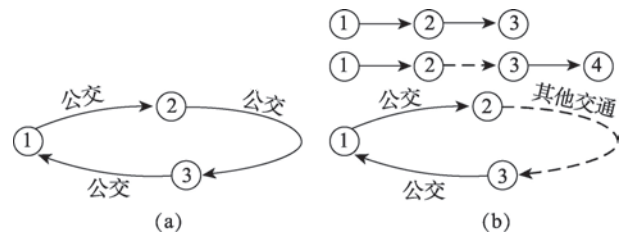


图4 公交出行链示意图

Fig. 4 Public transit trip chain

在获取乘客上车站点的基础上,本文基于连续性公交出行链方法和历史相似出行行为规律并行推算公交乘客下车站点。乘客的上车站点数据存储在HDFS,被分割成多个Block存储于不同节点上,基于HDFS的MapReduce并行计算时一个map任务对应一个Block数据。本文根据Block数量将乘客上车站点数据分成多个map任务并行计算下车站点,在Map函数中实现公交乘客下车站点的推算,而Reduce函数负责归并每个乘客完整的OD记录并存储到HDFS,具体推算过程如图5所示。

4.1 基于连续性出行链的下车站点并行推算方法

在Map函数中基于连续性出行链推算下车站点的流程如图6所示,主要有以下4步:

(1)读取一个IC卡用户的所有上车记录。如果该用户只有一条上车记录,则将该用户添加到 S_a , (S_a 是未能按照连续性出行链推算下车站点的上车

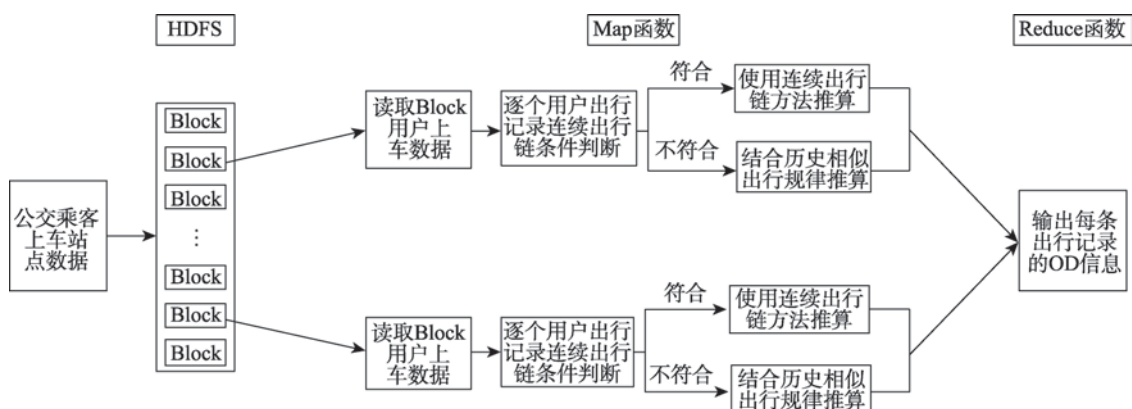


图5 公交乘客下车站点并行推算过程

Fig. 5 The parallel estimating flow of bus passengers destination station

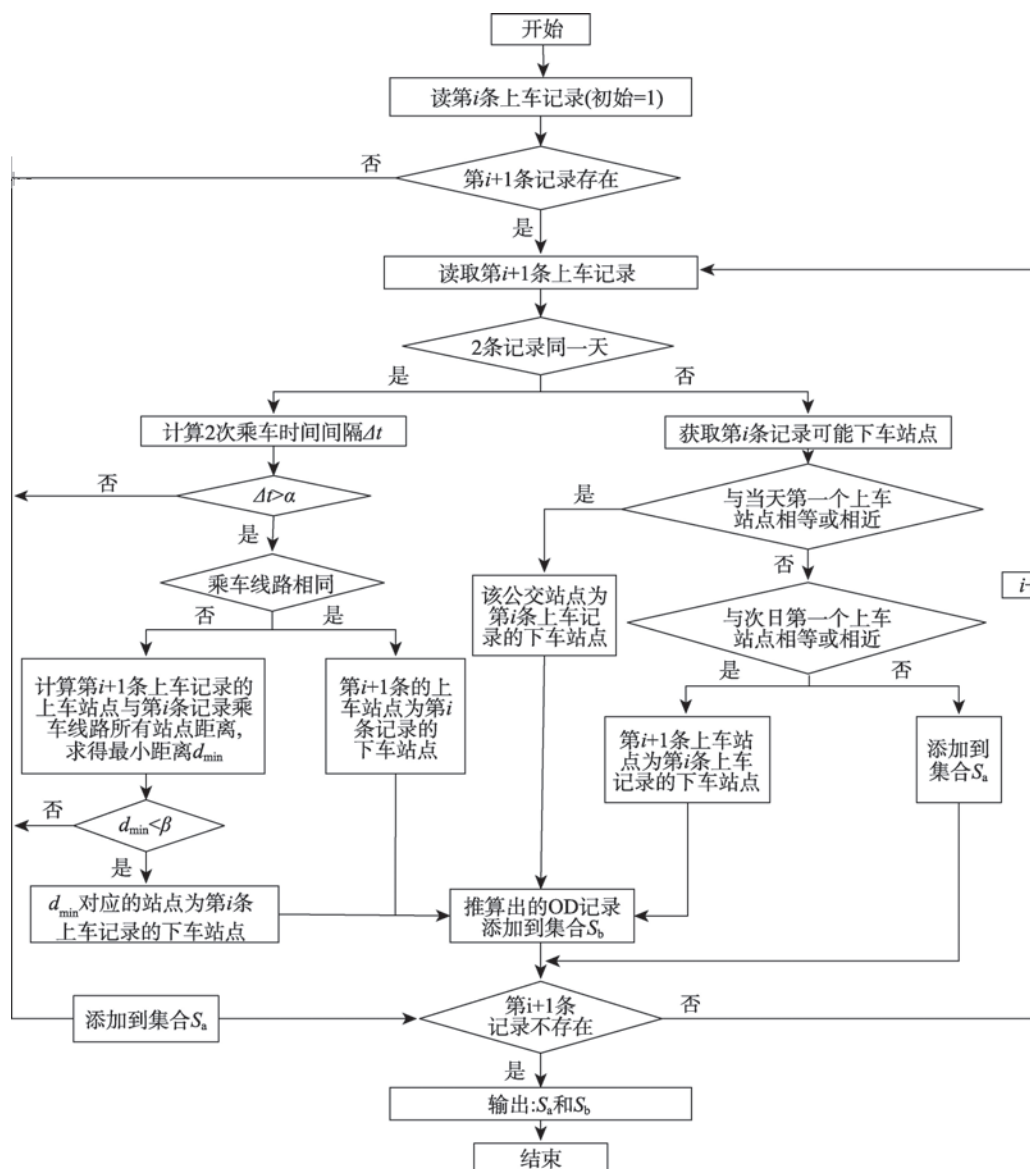


图6 Map函数中基于连续性公交出行链下车站点推算流程

Fig. 6 The destination station inference flow for continuous public transit trip chain in Map function

记录集合)。如果该用户有多条上车记录,则读取第 i 条和第 $i+1$ 条上车记录(初始 $i=1$)。

(2)比较这2条上车记录的日期,如果在同一天,计算2次乘车的时间差的绝对值 Δt 并与时间阈值 α 比较,若 $\Delta t < \alpha$,则可认为是代刷情况,将第 i 条上车记录添加至 S_a ,否则计算两条记录的乘车线路编号,若乘车线路编号相等,则符合连续性公交出行链,将第 $i+1$ 条上车记录的上车站点视为第 i 条上车记录的下车站点,将第 i 条OD记录添加到 S_b (S_b 是基于连续出行链方法推算出完整OD记录的集合);若不相等,则计算第 $i+1$ 条记录的上车站点与第 i 条上车记录所有可能下车站点的距离,获得其最小距离 d_{\min} ,如果 $d_{\min} < \beta$ (β 为距离阈值,代表换乘中可接受的最大步行距离),则符合连续性出行链(换乘), d_{\min} 所对应的公交站点为第 i 条上车记录的下车站点,将第 i 条OD记录添加到 S_b ,否则将第 i 条上车记录添加到 S_a 。

(3)若第 i 条上车记录和第 $i+1$ 条上车记录不在同一天时间内,则第 i 条为前一天的最后一条乘车记录,而第 $i+1$ 条为次日第一条乘车记录(上车记录已按刷卡时间排序)。根据第 i 条上车记录的线路编号、线路方向、站点编号信息获取该次乘车的可能下车站点。将这些可能的下车站点与当日第一次乘车的上车站点进行比较,如果2个公交站点相等或者相近(相等:站点名称一样,相近:2个站点距离 $< \beta$),则该公交站点为第 i 条上车记录的下车站点,并将第 i 条OD记录添加到 S_b 。否则将这些可能的下车站点与次日第一条乘车记录(即第 $i+1$ 条记录)的上车站点进行比较,如果2个站点相等或相近,则该站点为第 i 条上车记录的下车站点,并将第 i 条OD记录添加到集合 S_b ,都不符合条件,将第 i 条上车记录添加到集合 S_a 。

(4)输出集合 S_a 和集合 S_b 。

4.2 非连续性公交出行链下车站点推算

在上节中 S_a 是不符合连续性公交出行链推算下车站点的上车记录集合,针对这一部分用户上车记录,本文根据居民工作日和非工作日的出行特征,结合历史出行站点频次来推算下车站点。这部分的下车推算同样是在上节的Map函数中执行,具体推算流程如图7所示。从图7中可以看出对非连续性公交出行推算下车站点主要包含以下4步:

(1)输入上节的集合 S_a 和集合 S_b ,读取 S_a 的第 k

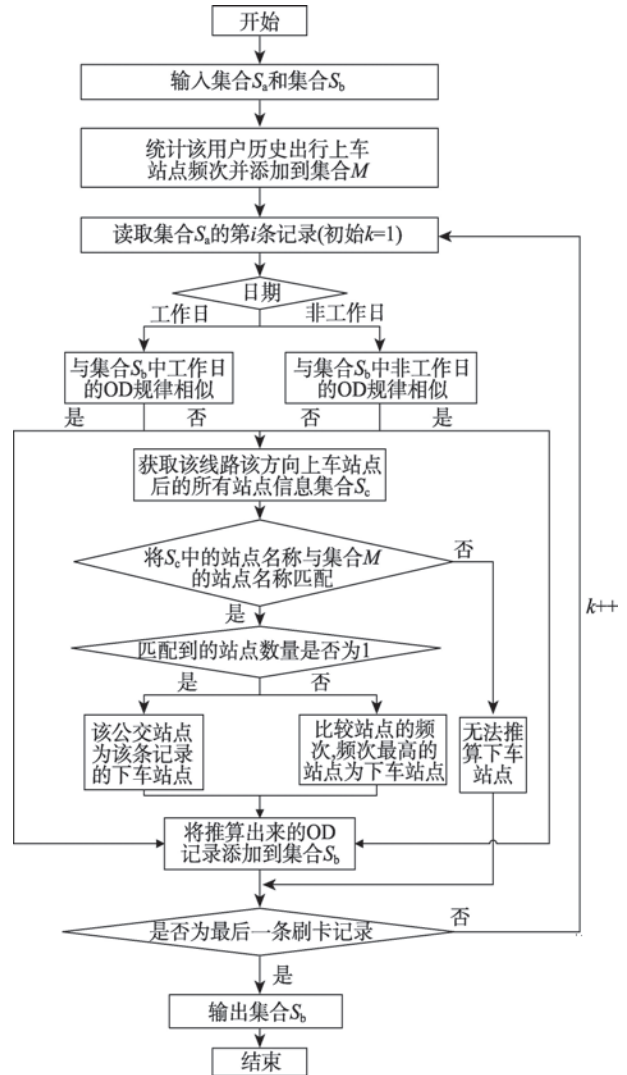


图7 Map函数中非连续性公交出行链下车站点推算流程

Fig. 7 The destination station inference flow for discontinuous public transit trip chain in Map function

条记录(初始 $k=1$),判断第 k 条记录的刷卡时间是否属于工作日。若是,则在集合 S_b 中寻找相似的工作日D点作为该条上车记录的D点,将OD记录添加到 S_b ;如果未找到相应的D点则将该条上车记录添加到集合 S_a 。

(2)若刷卡时间属于非工作日,则在集合 S_b 中寻找相似非工作日的D点作为该条上车记录的D点,将OD记录添加到 S_b ;如果未找到相应的D点则将该条上车记录添加到集合 S_a 。

(3)统计该用户历史出行的上车站点频次,按频次排序并存入一个Map集合 M 。读取集合 S_a 的第 j 条上车记录(初始 $j=1$),根据线路编号、行驶方向、站点编号获取该次乘车所有可能的下车站点,添加到集合 S_c 。将 S_c 的站点名称和集合 M 中的站

点名称进行匹配,若匹配到的站点数为1,则该站点就是第 j 条上车记录的下车站点,将第 j 条OD记录添加到 S_b ;若匹配的站点数大于1,则比较匹配结果中每个站点的频次,频次最高的站点为第 j 条记录的下车站点,将第 j 条OD记录添加到 S_b 。

(4)若第 j 条上车记录未匹配到相应的D点,则该条记录无法推算到下车站点,若该条记录不是该用户最后一条上车记录,则读取下一条上车记录,否则就结束该用户的下车站点推算工作,输出集合 S_b 。

5 实例分析

本文以厦门市2015年6月13日至26日的公交数据为例进行研究,原始数据包括:3337辆公交车,39 306 315条车辆GPS记录,295条公交线路,21 112 499条IC卡刷卡记录。经筛选和数据预处理,得到18 268 031条有效IC卡刷卡记录,IC卡刷卡数据和公交车辆GPS数据如表3、4所示。

本实验平台为Hadoop平台,该平台具有4个节点,每个节点为两核处理器,4 G内存和500 G磁盘空间,通过实现MapReduce并行推算厦门公交乘客OD点。按照调查及经验值,在上车站点的推算中,

表3 厦门市公交IC卡数据示例

Tab. 3 Sample of IC card data in Xiamen

IC卡编号	刷卡时间	线路编号	公交车辆编号
5238111601	2015-06-13 00:01:42	000099	8986
0878416539	2015-06-12 18:18:52	000129	9318

a 取值3 min, b 取值7 min,获取刷卡时间前3 min,后7 min的所有车辆进出站记录。推算出1 280 607个IC卡用户,16 879 661条上车记录,占IC卡刷卡记录的92.4%,有7.6%的IC卡记录因为进出站时间的丢失和异常未能成功推算出上车站点。按照图6和图7所示的下车站点并行推算流程进行下车站点推算,其中刷卡时间间隔阈值 α 取值1 min,是为了区分代刷情况, β 取值500 m,是由站点间距及乘客换乘步行最大接受范围确定的。共推算出1 085 853个IC卡用户,14 410 058条完整OD记录,占上车记录的85.3%,IC卡刷卡记录的78.9%。其余的14.7%的上车记录未能推算出下车站点原因有以下几点:①用户一天只有一次公交出行并且总体公交出行记录少。②用户一天公交出行两次,但第二次是公交换乘,无法推算第二次公交出行的下车站点。③部分用户两周时间内只有一次公交刷卡记录。表5为某个乘客OD推算结果,从中可以看出本文

表4 厦门市公交车辆GPS数据示例

Tab. 4 Sample of GPS bus data in Xiamen

GPS设备编号	GPS时间	纬度/ $^{\circ}$	经度/ $^{\circ}$	行驶方向	进出站标志	站点编号	线路编号
550001313902	2015-06-12 18:18:52	24.489145	118.072485	4	1	12	122

表5 OD推算结果

Tab. 5 The result of origin and destination inference

IC卡编号	刷卡时间	线路编号	线路方向	上车站点	下车站点	推算依据
4078181794	2015-06-13 16:07:31	24	4	中医院	岳阳小区	历史站点频次
4078181794	2015-06-15 07:24:23	31	4	岳阳小区	江头市场	连续出行链
4078181794	2015-06-15 16:23:44	31	5	江头市场	岳阳小区	连续出行链
4078181794	2015-06-17 07:43:44	859	5	岳阳小区	莲花路口东	相似出行规律
4078181794	2015-06-17 07:43:46	859	5	岳阳小区	莲花路口东	连续出行链
4078181794	2015-06-17 15:39:32	42	4	莲花路口东	市行政中心	相似出行规律
4078181794	2015-06-17 15:39:33	42	4	莲花路口东	市行政中心	连续出行链
4078181794	2015-06-17 17:00:12	18	4	市行政中心	枋湖车站	相似出行规律
4078181794	2015-06-17 17:00:14	18	4	市行政中心	枋湖车站	连续出行链
4078181794	2015-06-17 17:11:25	45	5	枋湖车站	岳阳小区	相似出行规律
4078181794	2015-06-17 17:11:26	45	5	枋湖车站	岳阳小区	连续出行链
4078181794	2015-06-18 09:04:16	15	4	岳阳小区	叉车厂	连续出行链
4078181794	2015-06-18 09:25:52	86	4	禾山路	未推算出	无

所提出的方法能够较为准确的推算公交乘客的上下车站点。

为更好地体现本文所提方法能够快速准确提取海量公交乘客OD的优势,将本文推算公交OD的方法与传统推算公交OD方法分别从上车站点推算和下车站点推算2个方面进行对比分析。

目前基于单一公交数据源推算公交乘客上车站点均是通过IC卡数据和公交车辆GPS数据相匹配计算,与本文上车站点计算方法一致。基于这同一方法,本文从传统的基于关系型数据库(如Oracle)计算、基于HBase数据的非并行计算和基于HBase数据库的MapReduce并行计算3个方面进行计算效率的对比分析,其结果如图8所示。从图8中可以看出,在公交乘客上车站点计算效率方面,基于HBase的MapReduce并行计算要高于其余2种计算方法,并且数据量越大,并行计算的优势越明显。

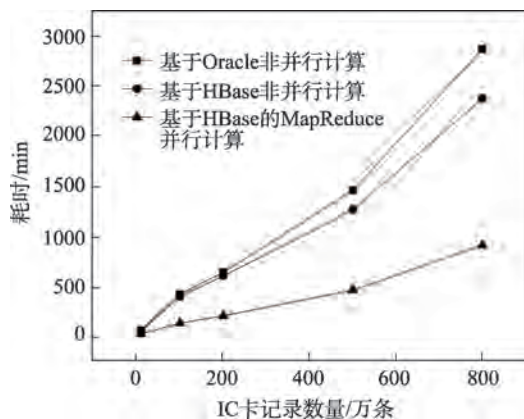


图8 上车站点计算效率对比图

Fig. 8 The comparison of computation efficiency of calculating boarding station

在下车站点推算方面,目前常见的方法是连续出行链方法、站点吸引权重方法以及这两种方法的结合使用。站点吸引权重法是根据站点概率来推算乘客下车站点,可信度不高,因而本文将与传统的连续出行链方法进行对比分析。使用传统的连续性公交出行链方法推算本文公交数据的下车站点,得到完整的公交OD记录11 380 983条,占IC卡数据总量的62.3%,低于本文公交乘客下车站点的提取率78.9%。从而可以看出本文使用连续出行链结合乘客历史出行规律及其出行特征方法推算公交乘客下车站点的提取率要高于传统基于连续出行链的方法。

6 结语

准确快速推算出海量公交乘客OD是及时获取全面真实城市公交客流信息的重要前提。面对海量的公交数据,本文提出一种基于MapReduce并行推算海量公交乘客OD的方法。该方法充分利用MapReduce并行计算的优势和乘客历史出行规律相似性的特点,准确快速地提取海量公交乘客OD信息,其优点在于:①结合乘客历史出行记录的站点出现频次推算下车站点,与以往研究相比,具有更高的公交乘客OD提取率;②面对海量的公交数据,可以快速计算出乘客OD信息,计算效率高,及时反映出全面真实的城市公交OD客流信息,对公交线路的优化和城市规划具有重要的实用价值。同时本文对公交数据的挖掘还有进一步待提高,如对于未能推算出OD点的数据,可以考虑结合其他数据源(地铁、出租)推算乘客出行OD点,获得更加完善的公交客流OD信息,对于此类数据还有待于进一步深入研究。

参考文献(References):

- [1] 孙慈嘉,李嘉伟,凌兴宏.基于云计算的公交OD矩阵构建方法[J].江苏大学学报(自然科学版),2016,37(4):456-461. [Sun C J, Li J W, Ling X H. Estimation of bus origin-destination matrix based on cloud computing[J]. Journal of Jiangsu University (Natural Science), 2016,37(4): 456-461.]
- [2] 陈锋,刘剑锋.基于IC卡数据的公交客流特征分析——以北京市为例[J].城市交通,2016,14(1):51-58,64. [Chen F, Liu J F. Characteristics of bus passenger flow based on IC card data: A case study in Beijing[J]. Journal of Urban Transportation, 2016,14(1):51-58,64.]
- [3] 戴霄.基于公交IC信息的公交数据分析方法研究[D].南京:东南大学,2006. [Dai X. Approach on the information analysis of urban public traffic base on the data of bus intelligent card[D]. Nanjing: Southeast University, 2006.]
- [4] 于勇,邓天民,肖裕民.一种新的公交乘客上车站点确定方法[J].重庆交通大学学报(自然科学版),2009,28(1):121-125. [Yu Y, Deng T M, Xiao Y M. A novel method of confirming the boarding station of bus holders[J]. Journal of Chongqing Jiaotong University (Natural Science), 2009,28(1):121-125.]
- [5] 尹长勇,陈艳艳,陈绍辉.基于聚类分析方法的公交站点客流匹配方法研究[J].交通信息与安全,2010,28(3):21-24. [Yi C Y, Chen Y Y, Chen S H. Bus station passenger matching method based on cluster analysis method[J].

- Journal of Traffic Information and Safety, 2010,28(3):21-24.]
- [6] 侯艳,何民,张生斌.基于公交IC卡刷卡记录的居民出行OD推算方法研究[J].交通信息与安全,2012,30(6):109-114. [Hou Y, He M, Zhang S B. Origin-destination matrix estimation method based on bus smart card records [J]. Journal of Transport Information and Safety, 2012,30(6):109-114.]
- [7] 张颂,陈学武,陈峥嵘.基于公交IC卡数据的公交站点OD矩阵推导方法[J].武汉理工大学学报(交通科学与工程版),2014,38(2):333-337. [Zhang S, Chen X W, Chen Z R. A method of deriving bus stops O-D matrix based on bus IC card data[J]. Journal of Wuhan University Technology (Transportation Science & Engineering), 2014,38(2):333-337.]
- [8] 宋晓晴,方志祥,尹凌,等.基于IC卡综合换乘信息的公交乘客上车站点推算[J].地球信息科学学报,2016,18(8):1060-1068. [Song X Q, Fang Z X, Yin L, et al. A method of deriving the boarding station information of bus passengers based on comprehensive transfer information mined from IC card data[J]. Journal of Geo-information Science, 2016,18(8):1060-1068.]
- [9] 靳佳.基于IC卡的北京市公交出行特征分析[D].北京:首都师范大学,2013. [Jin J. Analysis of bus trip characteristics in Beijing based on IC card[D]. Beijing: Capital Normal University, 2013.]
- [10] Hazelton M L. Statistical inference for time varying origin-destination matrices[J]. Transportation Research Part B Methodological, 2008,42(6):542-552.
- [11] Alsger A A, Mesbah M, Ferreira L, et al. Use of smart card fare data to estimate public transport origin - destination matrix[J]. Transportation Research Record Journal of the Transportation Research Board, 2015,2535:88-96.
- [12] Munizaga M A, Palma C. Estimation of a disaggregate multimodal public transport Origin-Destination matrix from passive smart card data from Santiago, Chile[J]. Transportation Research Part C, 2012,24(9):9-18.
- [13] Alsger A, Tavassoli A, Mesbah M, et al. Evaluation of effects from sample-size origin-destination estimation using smart card fare data[J]. Journal of Transportation Engineering Part A-Systems, 2017,143(4):1-10.
- [14] 马晓磊,刘从从,刘剑锋,等.基于公交IC卡数据的上车站点推算研究[J].交通运输系统工程与信息,2015,15(4):78-84. [Ma X L, Liu C C, Liu J F, et al. Boarding stop inference based on transit IC card data[J]. Journal of Transportation Systems Engineering and Information Technology, 2015,15(4):78-84.]
- [15] 吴祥国.基于公交IC卡和GPS数据的居民公交出行OD矩阵推导与应用[D].济南:山东大学,2011. [Wu G X. Urban public transportation trip OD matrix inference and application based on bus IC card data and GPS data[D]. Jinan: Shandong University, 2011.]
- [16] 胡郁葱,梁杰荣,梁枫明.基于IC卡数据挖掘获取公交OD矩阵的方法[J].交通信息与安全,2012,30(4):66-70. [Hu Y C, Liang J R, Liang F M. A way to get bus regional OD matrix based on mining IC card information[J]. Journal of Transport Information and Safety, 2012,30(4):66-70.]
- [17] 李海波,陈学武,陈峥嵘.基于公交IC卡和AVL数据的客流OD推导方法[J].交通信息与安全,2015,33(6):33-39, 95. [Li H B, Chen X W, Chen Z R. A method for estimating origin-destination matrix of public transit based on smart card and AVL data[J]. Journal of Transport Information and Safety, 2015,33(6):33-39,95.]
- [18] Munizaga M, Devillaine F, Navarrete C, et al. Validating travel behavior estimated from smartcard data[J]. Transportation Research Part C Emerging Technologies, 2014, 44(4):70-79.
- [19] Nunes A A, Dias T G, Cunha J F E. Passenger journey destination estimation from automated fare collection system data using spatial validation[J]. IEEE Transactions on Intelligent Transportation Systems, 2015,17(1):133-142.
- [20] 杨万波,王昊,叶晓飞,等.基于GPS和IC卡数据的公交出行OD推算方法[J].重庆交通大学学报(自然科学版),2015,34(3):117-121. [Yang W B, Wang H, Ye X F, et al. OD matrix inference for urban public transportation trip based on GPS and IC card data[J]. Journal of Chongqing Jiaotong University (Natural Science), 2015,34(3):117-121.]
- [21] 宋竹,秦志光,徐进,等.基于公交IC卡数据的大规模OD矩阵推导算法研究[J].计算机应用研究,2016(7):2007-2013. [Song Z, Qin Z G, Xu J, et al. Research on large-scale OD matrix estimation method based on bus IC card data[J]. Application Research of Computers, 2016(7):2007-2013.]
- [22] 胡继华,邓俊,黄泽.结合出行链的公交IC卡乘客下车站点判断概率模型[J].交通运输系统工程与信息,2014,14(2):62-67. [Hu J H, Deng J, Huang Z. Trip-chain based probability model for identifying alighting stations of smart card passengers[J]. Journal of Transportation Systems Engineering and Information Technology, 2014,14(2):62-67.]
- [23] 胡继华,高立晓,梁嘉贤.基于交通大数据的公交线路OD矩阵推断方法[J].科学技术与工程,2017,17(11):309-314. [Hu J H, Gao L X, Liang J X. An inference method of public transit OD matrix based on traffic big data[J]. Science Technology and Engineering, 2017,17(11):309-314.]
- [24] Zhao J, Qu Q, Zhang F, et al. Spatio-temporal analysis of passenger travel patterns in massive smart card data[J]. IEEE Transactions on Intelligent Transportation Systems, 2017,99:1-12.