

引文格式：

熊文, 周钱梅, 杨昆, 等. 基于时空相似性的大规模轨迹数据融合技术 [J]. 集成技术, 2019, 8(5) : 26-33.

Xiong W, Zhou QM, Yang K, et al. Spatial-temporal similarity-based data fusion for large-scale trajectories in metro system [J]. Journal of Integration Technology, 2019, 8(5) : 26-33.

基于时空相似性的大规模轨迹数据融合技术

熊 文¹ 周钱梅¹ 杨 昆¹ 代 浩² 孙 黎³

¹(云南师范大学信息学院 昆明 650500)

²(深圳北斗应用技术研究院 深圳 518055)

³(中国科学院深圳先进技术研究院 深圳 518055)

摘 要 如何利用大数据技术来支撑地铁路网规划、运营调度、应急管理和公共服务是当前学术界和工业界的研究热点。该文使用集成电路(Integrated Circuit, IC)卡交易记录和手机 WiFi 信号记录两种不同的数据,提出一种基于时空相似性的设备关联方法来重现乘客的完整轨迹。通过计算 IC 卡和手机两种不同设备历史轨迹的时空相似性,来关联同一乘客的 IC 卡和对应的手机。基于这种关联可以融合粗粒度的 IC 卡轨迹和细粒度的手机站内轨迹,进而重现乘客在地铁网络里的完整轨迹。实验通过对深圳地铁连续两个月智能 IC 卡刷卡数据和 WiFi 信号数据进行分析,同时利用最长公共子序列方法,在 Spark 集群计算了 728 万张 IC 卡轨迹数据和 4 010 万个移动设备轨迹数据的时空相似性。实验结果显示,该方法可以重现 20.3 万乘客的完整轨迹,足以用来支撑地铁清分和智慧警务等应用。

关键词 数据融合; 时空相似性; 轨道交通; 最长公共子序列

中图分类号 TG 156 文献标志码 A doi: 10.12146/j.issn.2095-3135.20190729001

Spatial-Temporal Similarity-Based Data Fusion for Large-Scale Trajectories in Metro System

XIONG Wen¹ ZHOU Qianmei¹ YANG Kun¹ DAI Hao² SUN Li³

¹(School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China)

²(Shenzhen Institute of Beidou Applied Technology, Shenzhen 518055, China)

³(Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

Abstract As the metro system becoming more and more important. How to utilize big data technology to support operational and management tasks is a hot topic in academia and industry communities. These tasks include metro network developing, service scheduling, risk response management, and public services. To address

收稿日期: 2019-07-29 修回日期: 2019-08-13

基金项目: 国家自然科学基金项目(61862066)

作者简介: 熊文, 博士, 研究方向为大数据系统基准测试和性能优化、时空数据存储和索引等; 周钱梅, 学士, 研究方向为地理信息系统; 杨昆(通讯作者), 教授, 博士研究生导师, 研究方向为空间信息技术, E-mail: kunyang@163.com; 代浩, 硕士, 研究方向为大规模数据处理; 孙黎, 博士研究生, 研究方向为数据融合。

these issues, we propose a data fusion-based approach on two sources to rebuild a passenger's full trip. The key idea is that we leverage the WiFi signal data and the smart card data together. We first calculate the spatio-temporal similarity between of smart card's trajectories and mobile device's trajectories. Then, we associate a passenger's smart card and the corresponding mobile device via their similarity. Finally, we combine the in-station trajectory hidden in the WiFi signal record and the coarse-grained trip presented by smart card record. We validate our approach on an extremely large dataset in the Chinese city Shenzhen. We calculate the similarity of trajectories generated by 7.28 million of smart cards and trajectories generated by 40.1 million of mobile devices in a Spark cluster. Experimental results show that this approach can rebuild 203 000 passenger trajectories. These results are enough to support many important applications in metro system.

Keywords data fusion; spatial-temporal similarity; metro transportation system; longest common subsequence (LCS)

1 引言

现代轨道交通由于其便捷、高效、准时和舒适等特点, 已经成为城市居民市内出行的首选交通工具。以深圳地铁为例, 单日内地铁出行规模最高达到了 500 万人次, 承担了整个城市超过三分之一的交通流量。随着城市空间的不断扩展, 地铁网络的规模与复杂性也在不断地增长。深刻理解数以百万计乘客的出行模式及其时空特征, 是地铁路网远景规划、运营调度、应急管理和智慧警务的重要基础。

当前, 大多数关于乘客出行时空特征的研究都集中在自动售检票 (Automated Fare Collection, AFC) 采集的集成电路 (Integrated Circuit, IC) 卡刷卡信息上^[1-4]。这些研究都属于采用单源数据的单视角研究方法, 尽管这些研究成果能很好地满足了部分应用场景 (如乘客在站内的平均行走时间、在站台平均等待时间和到达时间评估等), 但 AFC 的固有特性导致这些方法存在局限性。具体来讲, IC 卡的读卡器安装在地铁站内的闸机处, 读卡器只记录了乘客经过进站闸机和离开出站闸机的时空信息, 然而, 这些信息只反映了乘客的部分轨迹。至于乘客何时进入地铁站、如何换乘及在地铁内的轨迹, 只能依靠

基于概率的模型进行推断。因此, 仅仅以 IC 卡交易数据为对象的单视角研究方法无法重现乘客在地铁网络中的完整路径。

地铁站内近期广泛部署的无线 Mesh 网络为我们提供了一个潜在的机会。由于地铁站内手机信号较弱, 乘客无法使用无线互联网。为了提供更好的上网体验, 地铁运营商在地铁站内部署大量的无线访问点 (Access Point, AP) 感应器为乘客提供无线网络 (Wireless Fidelity, WiFi) 上网服务。当乘客进入地铁站, 在走道、大厅和站台内使用手机 WiFi 连接 AP 时, 每个 AP 都记录了对应手机的介质访问控制 (Multiple Access Channel, MAC) 地址和信号数据, 此时乘客在地铁站内的轨迹就隐藏在 WiFi 信号数据中。因此, 如果能够关联乘客的 IC 卡和对应的手机, 就有可能重现乘客的完整轨迹。但是, IC 卡和 MAC 数据完全不同的时空特征给设备的关联和轨迹的融合带来了巨大的挑战。一方面, IC 卡对应的轨迹在时间上相对稀疏而在空间上高度精确, 每次行程只产生两条记录, 并且可以确定刷卡时乘客就在对应的闸机处; 另一方面, MAC 对应的轨迹在时间上密集, 但在空间上不精确, MAC 在单个地铁站内可能产生几十甚至上百条信号数据, 无法确定 WiFi 信号产生时乘客的具

体位置。除此以外,乘客在每次行程中不一定都会开启手机 WiFi 功能。

本文提出一种基于多源异构数据的多视角研究方法。其核心是利用不同设备的时空相似性来进行设备关联并重现乘客的完整轨迹。首先,利用最长公共子序列(Longest Common Subsequence, LCS)计算 IC 卡和手机两种不同设备历史轨迹的时空相似性;然后,以相似性为基础,本研究关联了乘客 IC 卡对应的手机(移动设备);最后,基于这种关联可以融合粗粒度的 IC 卡轨迹和细粒度的手机站内轨迹,进一步重现了乘客在地铁里的完整轨迹。最终,在两个超大规模数据集上(深圳地铁连续两个月的 IC 卡交易记录和 WiFi 信号记录)验证了该方法的有效性。具体来讲,本研究的贡献有以下几点:

(1) 本文首次提出了利用多源异构数据来重现地铁乘客的完全轨迹。尽管任何单个数据源都不足以重现乘客的完整轨迹,但两种数据源可以互为补充,进而可以重现乘客的完整轨迹。

(2) 本文使用了 LCS 来计算 IC 卡轨迹和 MAC 轨迹的时空相似性,并以此为基础来关联乘客对应的 IC 卡和手机(或移动设备)。

(3) 本文在 Spark^[5] 集群计算了 728 万条 IC

卡轨迹数据和 4 010 万条 MAC 轨迹数据的时空相似性。实验结果显示,该方法可以重现 20.3 万乘客的完整轨迹。其中,这些轨迹足以用来支撑地铁清分和智慧警务等应用。

2 研究背景和动机

在本节中,首先介绍了支撑深圳地铁的一个大数据平台;其次,介绍了本研究动机-重现乘客的完整轨迹;最后,介绍了本文所用到的两个大规模数据集。

2.1 研究背景

目前深圳地铁包含 11 条线路和 224 个地铁站点,为了高效管理大规模地铁线路和提供良好的乘车体验,地铁运营商部署了多套传感网络。最初部署了 AFC 系统用于地铁票务,接着部署了安检系统和视频监控系统,后来又部署了 AP 感应器为乘客提供 WiFi 上网服务。这些传感器网络每天产生大量的数据。例如,每天产生的 IC 卡刷卡数据超过 1 000 万条;每天产生的 WiFi 信号数据超过 10 亿条。为了存储和分析这些数据,深圳地铁及其信息服务提供商构建了一个大数据平台(如图 1 所示)。大数据平台分为 3 个层次:最底层为

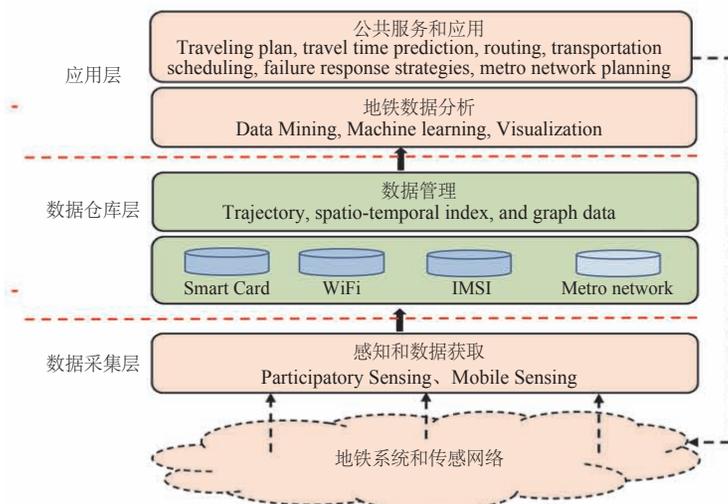


图 1 智能轨道交通大数据平台

Fig. 1 Big data platform for smart metro system

数据采集层, 主要由多个不同传感网络对应的终端进行数据采集; 中间层为数据仓库层, 以多源异构数据为基础构建不同的主题库; 最上层为应用层, 主要负责提供公共服务和应用等功能。该大数据平台主要支撑两方面的应用: 一方面, 用于地铁管理, 如实时客流量估计、地铁服务调度、应急管理和地铁远期规划; 另一面, 为乘客提供公共服务, 如出行安排、换乘推荐和到达时间评估。

2.2 研究动机-重现乘客的完整轨迹

深入理解数以百万计乘客的出行时空特征, 是地铁管理和公共服务的基础。当前, 大多数相关研究都以单视角的方式集中在 IC 卡交易数据。但是, IC 数据的特性导致了单视角的局限性。具体来讲, AFC 系统中读卡器安装在每个站点的闸机口, 读卡器只记录了乘客进入或离开闸机的时间。IC 对应的轨迹数据并没有完整描述乘客的行程。例如, 乘客何时进入地铁站? 从进站口行走到闸机的时间有多长? 在站内的逗留时间多长? 乘客在中途是否发生了换乘(若是, 在哪一站换乘)? 如何准确地回答这些问题是地铁管理、公共服务、智慧警务和地铁清分等重要应用的基础。因此, 在大规模地铁网络环境下重现乘客的完整轨迹显得尤为重要。

部署在地铁站内的 AP 为我们提供了一个潜在可能。不管乘客是进站、换乘还是出站, 开启 WiFi 功能的手机均有可能连接在站内广泛部署的 AP 感应器, 而乘客的站内轨迹就隐藏在这些 AP 感应器的 WiFi 信号数据中。尽管 IC 卡数据和 WiFi 数据都不足以单独重现一个乘客的完

整轨迹, 但两者在时空特征上却互为补充使乘客轨迹重现成为可能。这些都促使我们尝试通过数据融合的方法, 关联乘客的 IC 卡和手机(移动设备), 来重现乘客在地铁站内的完整轨迹。

2.3 数据集介绍

作为深圳地铁的信息服务提供商, 本文团队参与了深圳地铁大数据平台的解决方案建设, 且相应数据全部保存在该平台中。本研究每次分析的数据全部来自深圳地铁 11 条线路和 224 个地铁站点。数据产生的时间为 2018 年 9 月 1 日至 2018 年 10 月 31 日, 连续 8 个星期的时间。如表 1 所示, 728 万张 IC 卡共产生 2.49 亿条刷卡交易记录。刷卡数据中包含了智能卡号、刷卡时间、地铁站和进出站标志等。4 010 万个不同 MAC 地址共计产生 47.36 亿条 WiFi 信号数据。信号数据中包含了移动设备 MAC 地址、信号采集时间、地铁站和 AP 编号等。

需要特别指出的是, MAC 地址的个数远超 IC 卡个数。由于部分品牌手机操作系统, 如华为和苹果会随机产生伪 MAC 地址, 导致数据中存在大量虚假 MAC。后续在第 3 节会介绍如何通过技术手段来过滤这些假 MAC 地址。另外, 本文使用了足够的技术手段来保护乘客的隐私和数据安全, 技术人员无法通过轨迹数据推断某个具体乘客的行程。

3 问题定义和研究方法

为了重现乘客在地铁内的完整轨迹, 首先将

表 1 数据集和格式介绍

Table 1 Dataset summary and record format

IC 卡		MAC 地址	
数据名称	数值或描述	数据名称	数值或描述
智能卡个数	7 284 394	移动设备数	40 701 710
智能卡记录数	249 417 856	信号记录条数	4 736 953 391
刷卡记录的格式	卡号、时间戳和地铁站编号	信号记录格式	物理地址、时间戳和地铁站编号

IC 卡刷卡记录和 WiFi 信号记录分别转化为智能卡 ID 对应的轨迹和 MAC 地址对应的轨迹；然后，以 LCS 对两种不同设备的轨迹进行了两两相似性计算；最后，基于相似性计算结果将智能卡所属乘客对应的手机(移动设备)进行关联。具体流程(图 2)主要包括：(1)原始数据的预处理和轨迹数据的生成；(2)轨迹数据的两两相似性计算；(3)公交智能卡和手机设备的关联。

3.1 数据预处理

IC 卡交易记录的预处理相对简单。正常情况下，乘客的每次行程会有两条刷卡记录，分别对应着乘客起点站名称与拍卡时间、终点站名称与拍卡时间。尽管有部分刷卡数据缺失(小于万分之一的行程会缺失入站信息或出站信息)，但不影响最终的实验结果。

WiFi 信号数据的处理相对复杂，因为部分手机操作系统出于隐私保护的原因会产生伪 MAC 地址。这些伪 MAC 地址会导致部分设备的轨迹出现异常：一方面，干扰了轨迹相似度计算的精度；另一方面，浪费了很多计算资源。本文采用了 Zheng^[6]介绍的一系列方法，如卡尔曼和粒子滤波器对 MAC 轨迹数据进行了过滤。详细的过滤方法将在后续的研究中进行介绍。

过滤以后的数据，按照 IC 卡卡号(MAC 地址)和时间排序，可以得到不同设备的完整轨迹数据集。一条个体轨迹通常表示为有带地理坐标和时间标记的点序列，如公式(1)所示。

$$Tr=\{P_1, P_2, \dots, P_n\} \quad (1)$$

其中， P_i 表示个体的第 i 个位置记录点，如公式(2)所示。

$$P_i=(t, x, y) \quad (2)$$

其中， t 表示数据采集的时间戳； x 、 y 分别对应地铁站的经度和纬度。

3.2 相似性计算

Zheng^[6]和 Yuan 等^[7]总结了一系列轨迹相似性计算的方法，如欧式距离、动态时间规整、杰卡德相似性(Jaccard Similarity Coefficient)和 LCS 等。考虑到本研究中时空轨迹的特征，如时序数据非等长、数据缺失等原因，最终选择 LCS 作为轨迹数据的相似性计算方法。

最长公共子序列，其定义为已知一个序列 $X=\{x_1, x_2, \dots, x_m\}$ 和另一个序列 $Y=\{y_1, y_2, \dots, y_n\}$ ，存在一个序列 $Z=\{z_1, z_2, \dots, z_c\}$ 。其中， $c \leq m$ 且 $c \leq n$ ； $Z=\{z_1, z_2, \dots, z_c\}$ 既是前者 X 的子序列，也是后者 Y 的子序列，且 Z 是所有符合条件的子序列中最长的序列。用 $C[m, n]$ 来记录序列 X 和 Y 最长公共子序列的长度，得到递归关系：

$$C[i, j]= \begin{cases} 0, & i=0 \text{ 或 } j=0 \\ C[i-1, j-1]+1, & i, j>0 \text{ 且 } X_i=Y_j \\ \max(C[i, j-1], C[i-1, j]), & i, j>0 \text{ 且 } X_i \neq Y_j \end{cases} \quad (3)$$

其中， $0 \leq i \leq m$ 。

在计算出两个序列 X 和 Y 的公共子序列长度之后，使用公式(4)来量化两个时序数据的相似性。

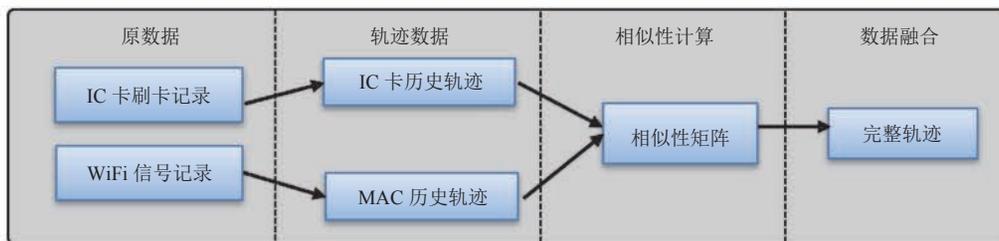


图 2 数据处理流程

Fig. 2 Workflow of data analysis

$$Sim(X, Y) = C[m, n] / k, \quad k = \min(m, n) \quad (4)$$

其中, $Sim(X, Y)$ 代表两个 X 和 Y 的相似性; k 为序列 X 和 Y 中长度的较小值。

在本文中, X 和 Y 可分别对应 IC 卡轨迹和 MAC 轨迹, $C[m, n]$ 代表两种不同的设备在相同的地铁站出现的次数。两条不同轨迹的相似性系数 $Sim(X, Y)$ 越大, 意味着 X 和 Y 对应的智能卡和移动设备越有可能属于同一位乘客。

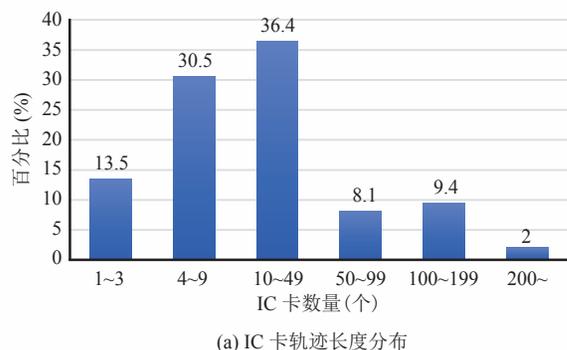
我们在一个 Spark 集群上实现了轨迹数据的两两相似性计算。这是一个典型的向量的两两相似性计算问题, 我们采用了倒排索引、前缀过滤、小表(IC 卡轨迹)广播和大表(MAC 轨迹)分区的优化方法。具体优化方法会在后续研究中详细描述。

4 实验结果分析

实验使用的 Spark 集群包含 10 个计算节点, 数据集包含 728 万张 IC 卡和 4 010 万个移动设备连续 8 周产生的轨迹数据。本节将分别介绍轨迹数据的特征和设备关联的准确度。

4.1 轨迹数据特征

通过数据分析发现: (1) 每张 IC 卡平均产生 35 条记录, 每个 MAC 地址平均产生 118 个记录, MAC 产生的轨迹点是 IC 的 3.2 倍; (2) 单张 IC 卡刷卡次数最大达到 1 347 次, 刷卡记录达到



(a) IC 卡轨迹长度分布

500 次以上的智能卡有 907 张。

如图 3 所示, 轨迹长度小于 10 的 IC 卡数量占总数量的 44%; 轨迹长度小于 10 的 MAC 个数占总数量的 68.3%。尽管 MAC 对应的轨迹普遍长度较短, 但单个 MAC 产生的信号记录最大已超过 120 万次(1 211 563), 并且产生 1 347 次以上记录的 MAC 地址超过 19 万个(192 131)。最终导致了 MAC 地址产生的记录数比 IC 卡产生的记录数密度大。

4.2 设备关联精度

我们采集了 140 组数据来验证该方法。其中每组数据包含了一位乘客的真实 IC 卡卡号和对应的手机 MAC。

通过分别计算以上每个 IC 卡历史轨迹和所有 MAC 历史轨迹的相似性, 并将 MAC 与该 IC 卡的相似性(公式(4))由大到小进行排列。从图 4 可以看出: (1) 14.3% IC 卡与相似性最大的 MAC

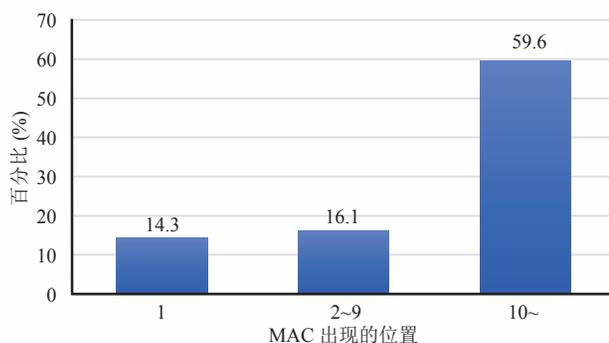
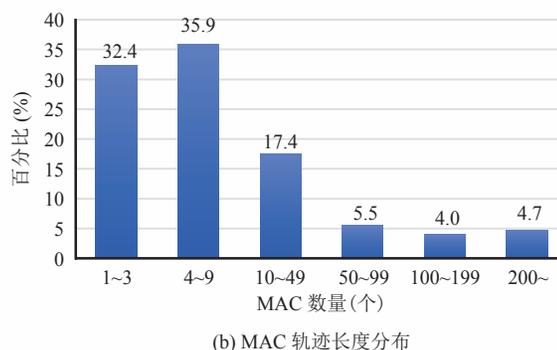


图 4 IC 卡对应的真实 MAC 在相似度排序中的位置

Fig. 4 The corresponding MAC's offset



(b) MAC 轨迹长度分布

图 3 IC 卡和 MAC 轨迹长度分布

Fig. 3 Length distribution of trajectories

属于同一名乘客；(2) 16.1% 的 IC 卡实际对应的 MAC 出现在第 2 至第 9 位间；(3) 59.6% 的 IC 卡对应的 MAC 出现在第 10 位以外。另外，当 IC 卡长度达到 49、并且 IC 卡轨迹长度与 MAC 轨迹长度比值大于 0.25 时，设备关联的准确率最高。

根据统计数据，IC 卡轨迹长度超过 49 的卡数占整个 728 万卡数的 19.5%。根据公式(5)可以初步推断，本文方法可以重现 20.3 万乘客的历史轨迹。

$$N = SmartCards \times LenThreshold \times ModelAccuracy \quad (5)$$

其中， N 为能够重现完全轨迹的乘客数量； $SmartCards$ 为 IC 卡的总数(728 万)； $LenThreshold$ 为轨迹长度大于 49 的 IC 占总数的百分比(19.5%)； $ModelAccuracy$ 为设备关联的准确率(14.3%)。

4.3 重现乘客的完整轨迹

这里将以一个例子来说明如何通过设备关联重现乘客轨迹。其中，选择深圳通卡号(6***373**)对应的 MAC 为(**-72**-3D**-51)进行乘客轨迹重现。如表 2 所示，分别显示了两种不同设备某次行程的轨迹。

从表 2 可以看出，该乘客从地铁塘朗站入口到入站闸机行走时间是 1 分 2 秒，在前海湾发生换乘并且换乘的时间开销是 2 分 26 秒，从福永站出站闸机到出站口的行走时间是 1 分 18 秒。实验结果验证了本文的预期：(1) 如果仅仅只有 IC 卡刷卡数据，无法得知该乘客何时进入地铁

站，及从入站口到达闸机的行走时间，也无法确定该乘客在哪一站发生了换乘和换乘所需的时间；(2) 如果只观察移动设备 MAC 对应的轨迹，无法确定该乘客行程的起点站和终点站。由此可见，两种设备的关联互为补充，可以重现乘客完整详细的乘车轨迹。

5 讨论与分析

探究城市空间中的群体活动分布及其时空变化能够帮助数据驱动的城市规划与城市治理。而基于大数据的时空群体活动研究是当前时空大数据研究的一个热点^[1-4]。轨道交通作为城市的重要基础设施和城市时空数据的重要来源，吸引了越来越多政府部门、企业和学者的关注。深刻理解数以百万计的轨道交通乘客是地铁远期规划、服务调度、应急管理、智慧警务和公共服务等应用的基础。

蒲一超等^[3]利用换乘站点的 AP 感应器采集乘客的 WiFi 信号用于改进杭州地铁的地铁清算算法和模型。但该方法需要大规模的人力物力到地铁内采集数据，并且这些轨迹的规模不足千个，在大规模地铁网络条件下不足以反映乘客出行特征的全局视图。本文方法利用现有的数据采集系统，不需要人工参与数据采集，并且可以甄别几十万个不同轨迹。

Sun 等^[4]通过对智能卡刷卡数据的分析和挖

表 2 两种不同设备的时空轨迹

Table 2 Two trajectories of smart card and mobile device

时间	IC 卡	MAC	备注
2018-09-05 20:12:44	/	塘朗站第一次出现	起点站
2018-09-05 20:13:46	塘朗站拍卡进站	/	/
2018-09-05 20:44:12	/	前海湾第一次出现	换乘站
2018-09-05 20:46:38	/	前海湾最后一次出现	换乘站
2018-09-05 21:06:52	福永站拍卡出站	/	/
2018-09-05 21:08:10	/	福永站最后一次出现	终点站

注：“/”为无相应数据记录

掘来了解轨道交通乘客的出行模式和时空特征, 用于支撑新加坡轨道交通系统的应急管理。Zhao 等^[8]分析了深圳地铁乘客出行需求的时空特征, 用于支撑地铁系统车辆调度等应用。Zhang 等^[9]和 Lee 等^[10]分别用基于概率的不同模型方法对乘客在地铁内的行程进行了详细的分析, 并提出了乘客旅行时间评估的方法。Zhao 等^[11]通过对智能卡历史刷卡数据的分析来为地铁乘客提供推荐换乘的服务。这些基于概率的模型较好地解决了部分应用, 但从单源数据出发单视角的研究方法具有局限性。例如, 这些研究只能从概率的角度来描述乘客出行的时空特征, 而不能从个体的角度来呈现乘客在地铁系统内的完整轨迹。本文方法从多视角的方式利用两种不同的数据源, 通过数据融合的方法来重现乘客在地铁系统内的完整轨迹。

6 总结和展望

在本文中, 我们尝试以数据融合的方式来重现乘客在地铁内的完整行程。首先, 利用 LCS 计算 IC 卡和移动设备两种不同时空特征轨迹的相似性; 然后, 在相似性基础之上将 IC 卡对应的属于同一位乘客的手机 MAC 地址进行关联; 最后, 利用这种关联融合了 IC 卡粗粒度的轨迹和 MAC 地址对应的细粒度的站内轨迹, 为重现乘客在地铁内的完整轨迹提供了基础。最终, 在一个超大规模数据集上实施了该方法: 在一个 10 节点的 Spark 集群完成了 728 万个 IC 卡轨迹和 4 010 万个 MAC 地址轨迹的相似性计算, 并验证了数据融合的效果。实验结果显示, 该方法能够重现 20.3 万乘客的完整轨迹。实验结果表明, 尽管单个数据源无法达成目标, 但两种数据互为补充可以重现乘客的轨迹。其中, 20.3 万乘客的完整轨迹足以支撑地铁清分、公共服务等重要应用。

目前, 系统是以线下方式利用历史轨迹数据完成的相似性计算和数据融合。下一步工作计划

实现在线增量计算, 能够实时计算乘客的轨迹相似性和重建完整轨迹, 支撑更多的地铁公共服务, 如换乘推荐、到达时间评估和站点流量预测等。另外, 还将通过算法改进来提高设备的关联精确度和数据融合的效率。

参考文献

- [1] 曹劲舟, 涂伟, 李清泉, 等. 基于大规模手机定位数据的群体活动时空特征分析 [J]. 地球信息科学学报, 2017, 19(4): 467-474.
- [2] 郑宇. 城市计算概述 [J]. 武汉大学学报(信息科学版), 2015, 40(1): 1-13.
- [3] 蒲一超, 尹梅枝, 朱琼斯, 等. 基于手机数据的地铁线网清分模型验证与优化 [J]. 都市快轨交通, 2017, 30(4): 102-109.
- [4] Sun LJ, Lee DH, Erath A, et al. Using smart card data to extract passenger's spatio-temporal density and train's trajectory of MRT system [C] // ACM Sigkdd International Workshop on Urban Computing ACM Sigkdd International Workshop on Urban Computing, 2012.
- [5] Zaharia M, Chowdhury M, Franklin MJ, et al. Spark: cluster computing with working sets [C] // Usenix Conference on Hot Topics in Cloud Computing, 2010.
- [6] Zheng Y. Trajectory data mining: an overview [J]. ACM Transactions on Intelligent Systems and Technology, 2015, 6(3): 1-41.
- [7] Yuan G, Sun PH, Zhao J, et al. A review of moving object trajectory clustering algorithms [J]. Artificial Intelligence Review, 2017, 47(1): 123-144.
- [8] Zhao JJ, 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, 18(11): 3135-3146.
- [9] Zhang F, Zhao JJ, Tian C, et al. Spatiotemporal segmentation of metro trips using smart card data [J]. IEEE Transactions on Vehicular Technology, 2016, 65(3): 1137-1149.
- [10] Lee HJ, Zhang DS, He T, et al. MetroTime: travel time decomposition under stochastic time table for metro networks [C] // 2017 IEEE International Conference on Smart Computing (SMARTCOMP), 2017.
- [11] Zhao JJ, Zhang F, Tu L, et al. Estimation of passenger route choice pattern using smart card data for complex metro systems [J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 18(4): 790-801.