

DOI:10.13203/j.whugis20160303



文章编号:1671-8860(2018)08-1152-07

轨迹数据的时间采样间隔对停留识别和出行网络构建的影响

赵志远¹ 尹凌² 方志祥^{1,3} 萧世伦⁴ 杨喜平⁵

1 武汉大学测绘遥感信息工程国家重点实验室,湖北 武汉, 430079

2 中国科学院深圳先进技术研究院,广东 深圳, 518055

3 时空数据智能获取技术与应用教育部工程研究中心,湖北 武汉, 430079

4 美国田纳西大学地理系,田纳西州 诺克斯维尔, 37996-0925

5 陕西师范大学地理科学与旅游学院,陕西 西安, 710119

摘要:个体轨迹数据已经广泛用于人群活动的研究中。在静止的局部空间开展的活动是个体日常生活的基本元素,在轨迹数据中对应停留部分。因此学者常从轨迹数据中识别停留来研究个体活动信息。然而,轨迹数据的时间采样间隔会对停留识别带来影响。针对该问题,首先提出了一个框架,量化不同持续时间长度的活动在不同时间采样间隔的轨迹数据中被识别为停留的概率。其次,考虑到个体出行网络依赖于停留识别结果,基于该框架,研究分析了时间采样间隔对出行网络分析结果的影响。最后,利用该框架分别对深圳市居民出行调查数据和手机轨迹数据进行了分析。研究表明,在面向人群活动的研究和应用中,该框架能支持时间采样间隔的选择决策和面向活动类型的研究结果评价。

关键词:轨迹数据;时间采样间隔;人类活动;停留识别;手机定位数据

中图分类号:P208

文献标志码:A

当前,中国快速的城市化进程伴随着诸多城市问题,理解城市居民的需求成为解决这些问题的前提。随着信息技术的进步,个体的位置追踪变得便捷、廉价,而人群级别的位置追踪也成为可能^[1]。位置追踪数据为了解个体行为习惯和人群对城市的需求提供了新的途径^[2]。基于活动的模型(activity-based model)是研究人群时空行为的新模型之一,该模型认为,活动尤其是在静止的局部空间开展的活动(也有在移动环境下进行的,如地铁上读书,如无特别说明,本研究的活动指在静止的局部空间开展的活动)是人们日常生活的基本要素,出行则满足了活动的空间位置转移需求^[3-4]。人们的现实生活轨迹数据中有相应的映射特征,活动表现为停留,而出行表现为停留之间的移动。因此,个体轨迹的时空采样点呈现出不同的集聚特征,文献[5]提出了停留与移动模型(stop and move of trajectory, SMoT)对轨迹数

据进行建模。停留的识别成为利用轨迹数据进行面向活动研究的基础^[6-8],进一步的研究,如表示个体日常活动的空间结构的出行网络,则依赖于停留识别的结果。

目前用于人群时空行为研究的轨迹数据主要有GPS轨迹数据^[7]、带地理位置的社交媒体数据^[9]、公交智能卡刷卡数据^[10-11]等。在时间采样间隔上,它们有各自的时间采样间隔特征^[12-13]。在时空分析领域,空间和时间维度上均存在着聚合效应(aggregation effect),即观测的分辨率变得粗略时,分析的结果会随之变化^[14-15]。学者们对时间采样间隔分析结果的影响进行了讨论,进一步阐明了聚合效应的存在^[16-17],但轨迹数据中聚合效应对停留识别的影响尚未有系统讨论。

针对这一问题,本文给出了活动停留识别的量化分析框架,并以出行网络为例研究了该影响将为进一步研究带来的变化。此问题的研究有助

收稿日期:2017-02-21

项目资助:国家自然科学基金(41231171, 41371420, 41301440);广东省自然科学基金(2014A030313684);深圳市基础研究项目(JCYJ20140610151856728)。

第一作者:赵志远,博士,现主要从事时空 GIS 理论与方法研究。zhaozhiyuan@whu.edu.cn

通讯作者:方志祥,博士,教授。zxfang@whu.edu.cn

于更好地理解各个轨迹数据的可用性和适用性,避免数据的误用,以降低研究结论的偏差。

1 活动与轨迹中的停留识别

1.1 轨迹中的停留

个体的日常活动需要在特定的时间与空间中开展,时空轨迹中的停留对应个体的活动。停留在轨迹上表现为持续的时间过程中,个体的空间位置未发生移动或在局部范围内移动^[5]。停留的位置与持续时间对应了现实世界中活动发生的地点与时间信息。一个上班族工作日的活动可以表示为图 1(a)所示:早上从家中(位置 A)出发到单位(位置 B)上班,工作期间需要到位置 C 去处理业务,而下班后,到位置 D 吃晚饭然后回到家中(位置 A)。

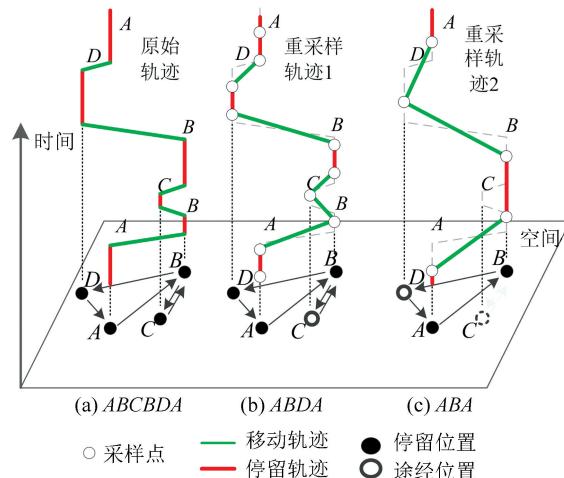


图 1 轨迹的时间采样间隔与停留

Fig.1 Temporal Sampling Intervals and Stays in a Trajectory

当对这个轨迹用不同时间采样间隔来收集位置信息时(图 1(b)和图 1(c)),不同持续时长的活动受到的影响存在差异。例如持续时间较短的活动 C 在图 1(b)中仅有一个轨迹点,随着时间采样间隔的增加,在图 1(c)中被完全忽略了。然而活动 A 和 B 持续时间相对较长,即使时间采样间隔增加,在图 1(b)和图 1(c)中也能得以识别出来。

1.2 时间采样间隔与活动的停留识别类型

采样点是组成轨迹数据的原始记录数据,表示的是个体在某一个时刻下的空间位置信息。活动的停留识别类型是指个体活动在轨迹数据中被识别为停留的类型。一个活动在轨迹数据中能否被识别为停留取决于在活动的持续时间内相同位置采样点的数量。活动是在特定的空间范围内发生的,当活动持续时间内有两个及以上相同位置采样点时,该活动在轨迹数据中体现为一个停留;而当持续时间内没有上述采样点时,该活动在轨迹数据中被忽略,无法识别为停留;当持续时间内有且只有一个采样点时,则无法直接满足停留识别的条件,但有可能与其他采样点形成停留。相应类别和规则见表 1。

2 停留识别概率框架

2.1 采样点数量问题建模

采样点的位置受时间采样间隔的直接影响,时间间隔直接反映了数据的时间分辨率,参考讨论可塑面积单元问题时用到规则的空间网格^[14, 18],以及讨论时间采样间隔影响时用到规则的采样间隔^[15],本研究采取等时间间隔的采样方式。

表 1 活动的停留识别类型与条件

Tab.1 Types and Conditions of the Stay-Identification of Activities

类别	描述	持续时间内采样点数量	示例
I	识别为停留	至少有两个相同位置采样点	图 1(a)~1(c)中的活动 A 和 B
II	潜在识别为停留	有且只有一个采样点	图 1(b)中的活动 C 和图 1(c)中的活动 D
III	无法识别为停留	无采样点	图 1(c)中的活动 C

在等间隔采样方式下,由于每个个体对应的采样起始时间和各个活动的时间是随机的,因此,活动持续时间内的采样点个数取决于时间采样间隔 ΔT 和活动持续时间 D 之间的关系。这个问题可以建模成数轴上均匀分布的采样点落入一个区间范围内点数量的问题:在一维数轴上分布着多个点位,点的绝对位置是随机的,相邻点之间的距离 t 是固定的,计算长度为 d 的区间中至少具有 N 个点的概率 p 。

该问题可以分为 3 种情况考虑。由于相邻点的间隔是固定的,其他点的分布方案可以由一个标记采样点来决定,该标记采样点的取值范围为长度为 t 的一个区间。①当 $d < (N-1)t$ 时,显然不论标记采样点的位置如何,都不会有超过 N 个点位于 d 区间中, p 为 0;②当 $d \geq Nt$ 时,不论标记采样点的位置如何,至少会有 N 个点位于 d 区间中, p 为 1;③当 $(N-1)t \leq d < Nt$ 时,有 N 个点位于 d 区间中。

选定一个采样点为标记采样点,其取值范围为时间采样间隔 t ,图 2 表示标记采样点在不同位置情况下区间 d 中采样点个数的变化情况。其中, ϵ 为区间长度 d 与 $(N-1)t$ 的差值。

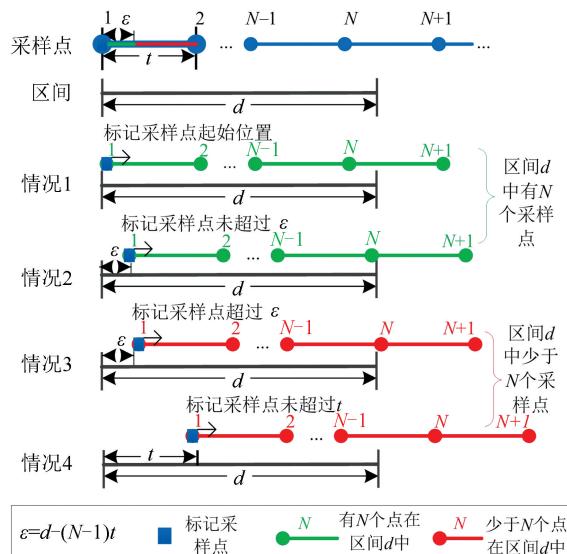


图 2 不同位置的标记采样点对应区间 d 中采样点的个数

Fig.2 Number of Sample-Points in Range d with Different Positions of the Marked Point

1)图 2 中的情况 1 变为情况 2 时,即当标记采样点的位置从起始时间到 ϵ 的范围内,在区间 d 中存在 N 个采样点;

2)图 2 中的情况 3 变为情况 4 时,即当标记采样点的位置超过 ϵ 而未超过 t 时,区间 d 中仅有 $N-1$ 个采样点;

3)当标记采样点超过 t 时,将重复此过程。

在此种情况下,至少有个 N 个采样点的概率为 ϵ 与 t 的比值。因此,区间长度为 d 的范围内,至少有 N 个间隔为 t 的采样点落入其中,其概率表示为 $P(d, t, N)$:

$$P(d, t, N) = \begin{cases} 0, & d < (N-1)t \\ \epsilon/t, & (N-1)t \leq d < Nt, \\ \epsilon = d - (N-1)t \\ 1, & d \geq Nt \end{cases} \quad (1)$$

2.2 不同停留识别类型概率

若轨迹数据的时间采样间隔为 ΔT ,持续时间为 D 的活动分别被识别为不同停留类型的概率可根据式(1)计算,其中,活动的停留识别类型 I 表示至少有 2 个采样点落入活动持续时间中($N=2$),该类型的概率 P_1 为:

$$P_1(D, \Delta T) = P(D, \Delta T, 2) \quad (2)$$

而活动的停留识别类型 II 表示有且只有 1

个采样点落入活动持续时间中,该类型的概率 P_{II} 可以通过区间中“至少 2 个采样点”($N=2$)的概率与“至少有 1 个采样点”($N=1$)的概率差值计算:

$$P_{II}(D, \Delta T) = P(D, \Delta T, 1) - P(D, \Delta T, 2) \quad (3)$$

停留识别类型 III 表示没有采样点落入活动持续时间中,是区间中“至少有 1 个采样点”($N=1$)的逆事件,该类型概率 P_{III} 为:

$$P_{III}(D, \Delta T) = 1 - P(D, \Delta T, 1) \quad (4)$$

在基于轨迹数据停留识别的研究中,相应的结果会因为停留识别的结果变化而变化,出行网络便是其中一种。

2.3 停留识别变化与出行网络的“退化”

出行网络是个体活动链在空间维度上的投影。活动链是由不同的活动组成的,在特定的时间采样间隔下,对应的每一个活动在轨迹数据中表现为 3 种停留识别类型,相应的出行网络也随之变化。个体的出行网络在较粗略时间采样间隔的轨迹中,会有不同的更为简单的结果,类似于“退化”的过程。如图 1 所示,原始的轨迹中有 6 个停留,如果对每个空间位置用字母编码,其出行网络为 $ABCBDA$;如果相应的每一个活动对应的停留识别类型分别是 I-II-II-I-I-I,则得到的出行网络为 $ABDA$;而如果对应的停留识别类型分别为 I-II-III-II-II-I 时,得到的出行网络为 ABA 。因此,出行网络的进化取决于每个停留类别的类型,定义算子 op 来描述每个停留对应的类别序列:

$$op = [\tau_1 \ \tau_2 \dots \tau_N], \tau_i \in \{I, II, III\} \quad (5)$$

式中, N 为原始活动的数量;算子的每一个要素 τ 表示对应活动的停留的识别类型,有 3 种情况:
①当要素为类别 I 时,对应的活动会识别为停留;
②而当要素为类别 III 时,对应的停留会被忽略;
③当要素对应的类别为 II 时,需要进一步考察其前后停留的类别。如图 1(c)中所示,在位置 C 的停留是类别 III,但被忽略停留两端邻接停留的类别如果为 II,而且两个停留的空间位置一致,此两个停留会合并为一个停留而保留下。将最终识别为停留的部分构建“退化”的出行网络,每一个算子对应一个出行网络。

算子中元素的类型决定了 3 种概率计算方式,对应概率 P_τ 根据 τ 的取值,分别对应 P_1 、 P_{II} 、 P_{III} ,即式(2)~(4)。相应地,算子的概率 op 为:

$$P_{op} = \prod_{i=1}^N P_{\tau}(\tau_i) \quad (6)$$

考虑到算子之间是独立的, 将具有相同出行网络结果的算子概率相加, 获得该结构的“退化”概率。据此, 对给定时间采样间隔下的活动模式, 可以计算其在粗略的采样间隔下可能的“退化”网络以及相应的概率情况。

3 时间采样间隔对停留识别率和出行网络影响实例分析

3.1 各类型活动停留识别率

为系统地评价不同类型活动在不同采样间隔下的停留识别情况, 本文利用居民出行调查数据反映出的不同类型活动的持续时间特点, 来评价不同采样间隔下不同类型活动的停留识别情况。

3.1.1 深圳市居民出行调查

在本研究中, 活动的停留识别率定义为该类型活动识别为 I 类停留的概率。据此, 对 2010 年深圳市居民出行调查的数据进行分析。深圳市出行调查数据中共包含约 19 万个调查对象, 调查对象在问卷中填写了一个工作日中的每一次出行的目的、开始时间和结束时间, 包含 9 种类型, 不同类型的活动都有自身的活动持续时间长度特征。工作、住家等活动持续时间较长, 而购物、娱乐等持续时间相对较短。对每一种类型活动持续时间的统计信息见表 2。

表 2 各类型活动持续时间的四分位数统计信息/min

Tab.2 Duration Time Statistics for Different Types of Activities/min

类型	平均值	中位数	第一分位数	第三分位数
工作	487	550	255	610
上学	406	450	230	570
住家	716	800	650	860
购物	110	50	30	90
娱乐	126	67	33	120
访友	253	160	90	300
看病	206	130	75	230
接送人	107	10	2	25
其他	336	175	60	595

3.1.2 不同类型活动停留识别率变化

对每一类的每一个活动, 在不同的采样间隔下, 根据其活动持续时间, 按照式(3)计算其分别在不同的时间间隔下被识别的概率, 结果如图 3 所示。

为了说明不同活动类型的停留识别率的分异性, 需要选择适当的统计检验方法和统计量^[19]。

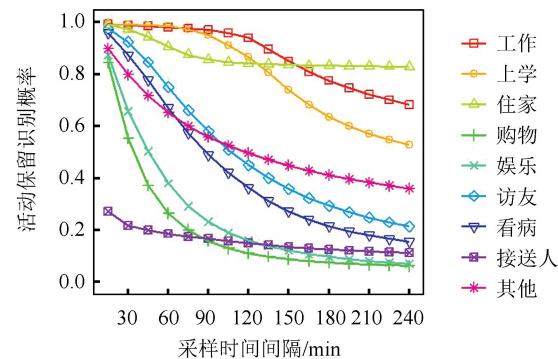


图 3 各时间采样间隔下不同类型活动被识别为停留的概率

Fig.3 Probabilities of Stay-Identification for Different Types of Activities by Different Temporal Sampling Intervals

空间分层异质性检验可以利用 Q-statistic 对空间、时间和属性等分层性质的分异性进行统计检验^[20]。利用此方法对图 3 中各活动类别的停留识别率的区别进行检验。结果表明, 活动类型的识别率存在显著的分异性, 分异性水平 q 的数值在 0.22~0.36 之间, 说明活动之间的分异程度不高。这主要是因为, 虽然部分类型的活动识别率对时间采样间隔的变化有迥异的反应(如工作与接送人), 但部分活动类型在特定的采样间隔下反应近似(如购物与娱乐)。

具体而言, 随着时间采样间隔的增加, 各类活动的停留识别率呈现下降的趋势。持续时间较长的住家、工作和上学等下降程度相对较小, 而购物、娱乐、接送人等持续时间较短的活动下降程度相对较大。在具体的识别概率数值上, 在 15 min 时, 除了接送人活动在 30% 以下, 多数类型的活动均具有较高的停留识别率, 在 85% 以上。而在 1 h 的情况下, 购物、娱乐活动的识别率显著下降, 不足 40%。访友、看病和其他类型活动也受到显著的影响, 但此两种类型的停留识别率尚在 65% 以上。而当时间采样间隔达到 2 h 时, 只有住家、工作和上学的活动识别率在 80% 以上, 剩余类型活动识别率均不足 50%。

图 3 中不同类型活动在不同采样间隔下的停留识别率在实际应用中有重要意义。一方面, 若轨迹数据的收集已完成, 在开展支持面向不同活动类型研究时, 给定结果的允许偏差范围, 可以判定数据的规格是否满足相应的研究要求。而在评价了时间采样间隔对不同类型活动停留识别结果的影响基础上, 可以进一步纠正偏差, 提高结论的可信度。另一方面, 当研究目的先行确定, 为此制

定轨迹数据采集的规则时,该结果能够根据需要帮助确定时间采样间隔的范围,例如若研究看病活动,采样间隔不宜大于30 min。数据的采集与管理成本以及后期的数据处理与时间采样间隔有关,采样越密,成本越高。参照本文的分析结果能够在满足应用需求的前提下,确定最优的时间采样间隔,从而降低成本。

3.2 出行网络分析

出行网络在人群分类和基于活动的交通需求分析模型中具有广泛的应用基础。本文以此为例来分析时间采样间隔对建立在停留识别基础上研究的影响。大规模手机定位数据在大样本量和低采集成本方面的优势,使其被用来研究城市人群的出行网络特征^[21]。为了说明时间采样间隔对出行结构的影响,本研究利用手机定位轨迹数据提取的出行网络来进行分析。

3.2.1 手机定位数据

实验数据为深圳市某通讯公司的手机定位数据,包括一个工作日中140多万个匿名用户的轨迹数据,当用户接打电话、收发短信以及连接移动互联网时会产生位置记录。该数据中共有约3 400个基站,基站间平均距离约为550 m。为了表明不同时间采样间隔下的出行网络状况,本研究用3条规则筛选一个高频采样的数据集:①1 d的记录个数不小于288条,以保证平均间隔不大于5 min;②记录的时间跨度不小于16 h,以保证尽量全地覆盖日常活动;③最大采样间隔不大于1 h,以保证记录在时间上的均匀分布。根据以上规则,本文共获得具有高频采样的用户19 370个,用户的平均采样间隔为2 min。为了降低手机数据中“乒乓”效应的影响,用80 km/h的速度阈值对该数据集进行处理。然后,对此数据集使用文献[21]中的方法利用时间阈值和空间阈值来识别停留。参数设定上考虑到基站的平均距离,设定距离阈值为500 m,停留持续时长考虑到人们日常活动的实际情况,将时间阈值设定为10 min。在此基础上,以500 m为阈值、以停留中心点位置为参照进行聚类,空间上邻近的停留认为是发生在同一个位置上。据此可以得到每一个个体的出行网络情况,然后根据本研究的分析框架推演不同时间采样间隔下出行网络的变化。

3.2.2 出行网络变化

根据高频采样数据的出行结果分别分析了此数据在0.25 h、0.5 h、1 h、2 h和4 h的时间采样间隔下的情形。为了说明相应的变化,本研究选用高频采样中出行网络最多的10种作为统计样

例^[21],各个采样间隔下不同出行结果的比例如图4所示。

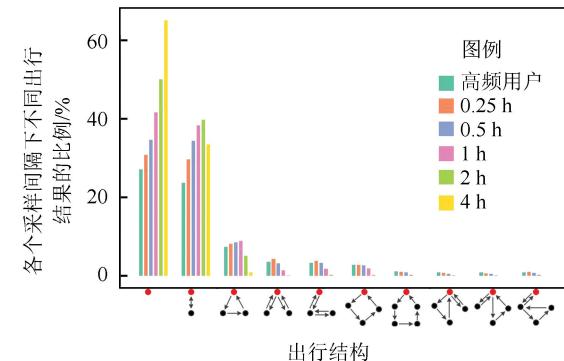


图4 前10种出行结构在典型的时间采样间隔下的比例
Fig.4 Proportions of Top-10 Motifs in Typical Temporal Sampling Intervals

图4中的结果表明,随着时间采样间隔的增加,轨迹逐渐简化,前3种简单的出行网络比例增加,出行网络呈现出简单化的趋势。具体而言,高频数据能够捕捉更为全面详尽的活动,前3种简单出行网络在人群所占的整体比例相对较低;当时间采样间隔为0.25 h和0.5 h时,前两种出行网络有显著增加,其他出行网络中与高频采样数据有较好的一致性,比例损失不大;时间采样间隔为1 h时,前两种出行网络的比例继续增加,而剩余部分相比高频采样数据的结果,呈现出显著减少的现象;而当时间采样间隔大于等于2 h时,出行网络主要集中于前3种,尤其是前两种网络。

从手机数据中分析出的出行网络和比例依赖于时间采样间隔。而在基于活动的交通需求分析中,这些出行网络类型和比例可以作为微观模拟步骤的输入信息,因此时间采样间隔会对模拟结果产生影响。图4的结果表明,当时间采样间隔不大于0.5 h时,相比于高频采样的结果,变化相对较小,而当时间采样间隔大于1 h的情况下,大量的出行网络将被忽略。因此,在基于轨迹数据提取出行网络的研究中,要慎重对待时间采样间隔的影响。

4 结语

针对轨迹数据中时间采样间隔对停留识别以及出行网络的影响,本研究根据采样点的时空特征与活动持续时长的关系,提出了一个概率框架来量化在特定时间采样间隔的轨迹数据中不同持续时间的活动被识别为停留的概率,以及进一步对出行网络分析的影响;基于该分析框架,以居民

时间采样间隔提供决策参考;利用手机定位数据分析了时间采样间隔对出行网络的影响,结果表明,为了保证人群出行网络的完整性,轨迹数据的时间采样间隔最好不大于 0.5 h。

本研究的结论表明,在利用轨迹数据研究人群活动时,时间采样间隔会对相关结果产生显著影响。这提醒学者需要根据研究目的来评价既有数据的可用性和适用性或去选择合适的参数收集轨迹数据,尤其是在各种观测数据大量涌现的当下,更加需要预防个体轨迹数据的误用。后续的研究工作包括:①时间采样间隔的影响在空间上是具有异质性的,不同的城市或同一城市不同的区域均各有特点,进一步探讨其差异可以帮助我们加深对时间采样间隔影响的理解,从而更好地降低其负面影响;②时间采样间隔的影响对不同的人群也有异质性,对其进行进一步探讨可以帮助细分不同人群的移动模式对该影响的反应,提高研究结论对精细化个体应用的可信度。

参 考 文 献

- [1] Zheng Yu. Introduction to Urban Computing [J]. *Geomatics and Information Science of Wuhan University*, 2015, 40(1): 1-13 (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)
- [2] Liu Yu. Revisiting Several Basic Geographical Concepts: A Social Sensing Perspective [J]. *Acta Geographica Sinica*, 2016, 71(4): 566-577 (刘瑜. 社会感知视角下的若干人文地理学基本问题再思考 [J]. 地理学报, 2016, 71(4): 566-577)
- [3] Timmermans H, Arentze T, Joh C-H. Analysing Space-Time Behaviour: New Approaches to Old Problems [J]. *Progress in Human Geography*, 2002, 26(2): 175-190
- [4] Davidson W, Donnelly R, Vovsha P, et al. Synthesis of First Practices and Operational Research Approaches in Activity-Based Travel Demand Modeling [J]. *Transportation Research Part A: Policy and Practice*, 2007, 41(5): 464-488
- [5] Alvares L O, Bogorny V, Kuijpers B, et al. A Model for Enriching Trajectories with Semantic Geographical Information [C]. The 15th Annual ACM International Symposium on Advances in Geographic Information Systems, New York, 2007
- [6] Palma A T, Bogorny V, Kuijpers B, et al. A Clustering-Based Approach for Discovering Interesting Places in Trajectories [C]. The 2008 ACM Symposium on Applied Computing, New York, 2008
- [7] Zheng Y, Li Q, Chen Y, et al. Understanding Mobility Based on GPS Data [C]. The 10th International Conference on Ubiquitous Computing, New York, 2008
- [8] Xu Jinlei, Fang Zhixiang, Shaw Shihlung, et al. The Spatio-Temporal Heterogeneity Analysis of Massive Urban Mobile Phone Users' Stay Behavior: A Case Study of Shenzhen City [J]. *Journal of Geo-information Science*, 2015, 17(2): 197-205 (徐金垒, 方志祥, 萧世伦, 等. 城市海量手机用户停留时空分异分析——以深圳市为例 [J]. 地球信息科学学报, 2015, 17(2): 197-205)
- [9] Zhang T, Cui P, Song C, et al. A Multiscale Survival Process for Modeling Human Activity Patterns [J]. *PLoS ONE*, 2016, 11(3): e0151473
- [10] Hou Yan, He Min, Zhang Shengbin. Origin-destination Matrix Estimation Method Based on Bus Smart Card Records [J]. *Journal of Transport Information and Safety*, 2013, 30(6): 109-114 (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, 2013, 30(6): 109-114)
- [11] Blondel V D, Decuyper A, Krings G. A Survey of Results on Mobile Phone Datasets Analysis [J]. *EPJ Data Science*, 2015, 10(4): 1-55
- [12] Wang De, Wang Can, Xie Dongcan, et al. Comparison of Retail Trade Areas of Retail Centers with Different Hierarchical Levels: A Case Study of East Nanjing Road, Wujiaochang, Anshan Road in Shanghai [J]. *Urban Planning Forum*, 2015(3): 50-60 (王德, 王灿, 谢栋灿, 等. 基于手机信令数据的上海市不同等级商业中心商圈的比较——以南京东路、五角场、鞍山路为例 [J]. 城市规划学刊, 2015(3): 50-60)
- [13] Xu Ning, Yin Ling, Hu Jinxing. Identifying Home-Work Locations from Short-Term, Large-Scale, and Regularly Sampled Mobile Phone Tracking Data [J]. *Geomatics and Information Science of Wuhan University*, 2014, 39(6): 750-756 (许宁, 尹凌, 胡金星. 从大规模短期规则采样的手机定位数据中识别居民居住地 [J]. 武汉大学学报·信息科学版, 2014, 39(6): 750-756)
- [14] Openshaw S. The Modifiable Areal Unit Problem [M]. Norwich, England: Geo Books, Regency House, 1984
- [15] Cheng T, Adepeju M. Modifiable Temporal Unit Problem (MTUP) and its Effect on Space-Time Cluster Detection [J]. *PLoS ONE*, 2014, 9(6): e100465
- [16] Tang Luliang, Kan Zihan, Huang Fangzhen, et al. Travel Time Detection at Intersection from Taxi's

- Trace Data [J]. *Geomatics and Information Science of Wuhan University*, 2016, 41(1): 136-142(唐炉亮, 阚子涵, 黄方贞, 等. 利用低频时空 GPS 轨迹进行交叉口通行时间探测 [J]. 武汉大学学报·信息科学版, 2016, 41(1): 136-142)
- [17] Ranacher P, Brunauer R, van Der Spek S, et al. What is an Appropriate Temporal Sampling Rate to Record Floating Car Data with a GPS? [J]. *ISPRS International Journal of Geo-Information*, 2016, 5(1): 1-17
- [18] Chen Jiangping, Zhang Yao, Yu Yuanjian. Effect of MAUP in Spatial Autocorrelation [J]. *Acta Geographica Sinica*, 2012, 66(12): 1 597-1 606(陈江平, 张瑶, 余远剑. 空间自相关的可塑性面积单元问题效应 [J]. 地理学报, 2012, 66(12): 1 597-1 606)
- [19] Wang J F, Stein A, Gao B B, et al. A Review of Spatial Sampling [J]. *Spatial Statistics*, 2012, 2: 1-14
- [20] Wang J F, Zhang T L, Fu B J. A Measure of Spatial Stratified Heterogeneity [J]. *Ecological Indicators*, 2016, 67: 250-256
- [21] Jiang S, Fiore G A, Yang Y, et al. A Review of Urban Computing for Mobile Phone Traces: Current Methods, Challenges and Opportunities [C]. The 2nd ACM SIGKDD International Workshop on Urban Computing, New York, 2013

Impacts of Temporal Sampling Intervals on Stay Detection and Movement Network Construction in Trajectory Data

ZHAO Zhiyuan¹ YIN Ling² FANG Zhixiang^{1,3} SHAW Shihlung⁴ YANG Xiping⁵

¹ State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

² Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

³ Engineering Research Center for Spatiotemporal Data Smart Acquisition and Application, Ministry of Education of China, Wuhan 430079, China

⁴ Department of Geography, University of Tennessee, Knoxville, Tennessee 37996-0925, USA

⁵ School of Geography and Tourism, Shaanxi Normal University, Xi'an 710119, China

Abstract: Trajectory data have been extensively used in human mobility studies. Activities, especially conducted on a static and local space are basic elements of people's daily life, and they are represented as stays in trajectories. Hence detecting stays from trajectories has become a base for many activity-oriented studies. The temporal sampling interval (TSI) of trajectory data can impact the result of stay detection. However such impacts have not been systematically studied yet. This study proposes a probability-based framework, which aims to quantify the probability of an activity that with a specific duration time can be detected as a stay with different TSIs. Moreover, this framework can support further analysis on the evolution of daily movement network with different TSIs. We demonstrate the impacts of TSIs on stay detection and movement networks construction by using a trip survey dataset and a mobile phone location dataset of Shenzhen, China respectively. This study provides both methodological and empirical guidance on the decision-making of a TSI selection as well as the estimation of the results of activity-oriented studies.

Key words: trajectory data; temporal sampling interval; human activities; stay detection; mobile phone location data

First author: ZHAO Zhiyuan, PhD, specializes in space-time GIS. E-mail: zhaozhiyuan@whu.edu.cn

Corresponding author: FANG Zhixiang, PhD, professor. E-mail: zxfang@whu.edu.cn

Foundation support: The National Natural Science Foundation of China, Nos. 41231171, 41371420, 41301440; the Natural Science Foundation of Guangdong Province, No. 2014A030313684; Basic Research Project of Shenzhen City, No. JCYJ20140610151856728.