

DOI:10.13203/j.whugis20160203



文章编号:1671-8860(2017)01-0063-07

城市手机用户移动轨迹时空熵特征分析

康朝贵¹ 刘 瑜² 邬 伦²

1 武汉大学遥感信息工程学院,湖北 武汉,430079

2 北京大学地球与空间科学学院,北京,100871

摘 要:利用手机话单数据分析城市个体居民移动活动的时空熵特征,一方面探讨了从原始话单记录中进行出行识别的必要性,另一方面提出了一种考虑空间邻近性的轨迹近似熵特征分析方法。其中,出行识别可以克服手机定位数据采样频率较低的缺陷;近似熵分析方法具有强空间鲁棒性,可以减少因手机定位数据空间精度较低带来的影响。实证结果表明,城市居民出行活动既具有强烈的目的地选择倾向,同时也具有强烈的移动路径选择偏好。

关键词:手机话单;个体轨迹;活动提取;出行模式;时空熵

中图法分类号:P208

文献标志码:A

根据大数据研究个体或群体行为活动中蕴含的空间认知规律及空间行为和交互模式,建立以人为本的地理信息服务,进而支持个体或群体时空行为决策,已经成为地理信息科学研究的前沿问题^[1]。此类研究也将为联系“计量”和“行为”两个地理学思想流派以及重新审视人地关系提供一个全新的视角,从而推动理论地理学的进一步发展。研究成果表明,居民移动模式分析能够为城市规划和管理^[2]、交通监控与预测^[3]、信息与疾病传播^[4]、旅游监测与分析^[5]等众多领域的研究提供工作基础与方法指导。

在刻画人类空间运动模式方面,停留时间、出行距离和轨迹熵是 3 个关键的统计特征^[6]。其中,停留时间分布对人类个体的空间扩散速度有决定性影响^[7];出行距离分布直接反映城市的经济性和效率,并显著影响信息、疾病等的传播^[8];轨迹熵则集中反映了个体居民移动活动的主观意愿和惯性模式,既能揭示个体居民对于城市空间利用的不均衡性,也能突出居民移动活动内在的重复性、规律性和可预测性^[9]。

近年来,文献^[8]开创性地使用手机通话数据来分析海量个体的移动活动规律,极大地促进了对人类空间运动模式认知的深度和广度。相对传统行为学研究常采用的调查方法而言,手机数据具有覆盖人群广、可持续时间长且代价成本低等

优势。但是,手机数据采集过程一般缺少严格的控制,重构得到的人类移动轨迹往往具有较大的时空不确定性^[10],只能作为个体居民移动活动规律的近似反映。尽管如此,现有研究大多追求用户样本覆盖的最大化,而缺少对手机轨迹时空不完整性导致的移动行为规律认知偏差的探讨。文献^[10]指出了从手机轨迹中进行活动提取的意义及必要性。但是,如何针对手机移动轨迹时空特征的内在缺陷,提供系统性、更加可靠的人类移动活动规律分析方法和框架仍有待进一步探索研究。

本文利用手机话单数据分析城市个体居民移动活动的时空熵特征。首先,探讨了从原始话单记录中进行出行识别的必要性;然后,提出了一种考虑空间邻近性的轨迹熵特征分析方法。其中,出行识别可以克服手机定位数据采样频率较低的缺陷;近似熵分析方法则具有强空间鲁棒性,可以减少手机定位数据空间精度较低带来的影响。通过在哈尔滨市开展实证分析发现,城市居民出行活动既具有强烈的目的地选择倾向,同时也具有强烈的移动路径选择偏好。

1 研究区与数据

本文分析的手机话单记录包含哈尔滨市某主

收稿日期:2016-05-02

项目资助:国家自然科学基金(41601484);测绘遥感信息工程国家重点实验室开放研究基金(15S01)。

第一作者:康朝贵,博士,助理教授,主要从事城市信息学和轨迹计算的理论与方法研究。cgkang@whu.edu.cn

要移动运营商全部用户历时一个月(自7月1日至7月31日,共31 d)的通话记录,用户总数为262万,累计通话次数超过1.5亿次,通话时长超过2.6亿min,空间位置唯一的有效基站557个。该数据集由被动手机定位方式产生,只记录有用户拨打或接听电话时的接入基站空间位置,不包含用户收发短信以及网络接入等活动信息。由于用户具体空间位置的确定基于基站定位原理,手机定位的精度取决于基站的分布模式。一般情况下,研究区内相邻基站间的距离为300~500m(见图1中城区部分)。

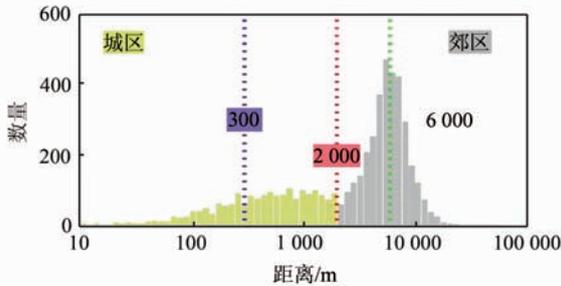


图1 研究区相邻手机基站距离分布

Fig.1 Distances Between Adjacent Mobile Station Antennas

研究区内用户的通话活动具有明显的时间节律:(1)以d为单位,表现为工作日通话活动(次数和总时长)强于非工作日通话活动;(2)以h为单位,无论是工作日还是非工作日,都呈现出明显的“双峰”特点,早晚(9:00和17:00左右)各有一个峰值,午间(12:00左右)则存在一个明显的波谷。此外,通过对比通话次数和通话时长的时间变化趋势发现,晚18:00~20:00之间用户每次通话的平均时长明显增加。具体表现为:白天(6:00~17:00)平均每次通话的时长基本稳定在1~1.5min,18:00之后开始迅速上升,在22:00左右达到峰值(4min),之后再次下降。上述特征表明研究区内用户每天的通话活动及其变化模式非常稳定,不同天之间的差异很小,说明用户使用手机在时间上比较规律。

表1统计了研究期内所有手机用户的通话次数分布。其中,约58.32%的用户平均每天通话不超过1次,22.95%的用户平均每天通话2~3次,15.64%的用户平均每天通话4~10次,只有少数用户(3.09%)平均每天通话10次以上。上述结果表明,用户的手机使用习惯和频率具有非常显著的差异。

表1 手机用户通话次数统计

Tab.1 Number of Calls by Each Mobile User

月通话次数	用户数	比例/%
(0,31)	1 527 555	58.32
(31,93)	601 085	22.95
(93,310)	409 739	15.64
(310,8 189)	80 919	3.09

2 活动点与出行提取

以单个手机用户为对象,根据其不同时刻接入手机服务基站的记录,可以将这些位置按照时间顺序串联生成相应的移动轨迹。考虑到个体移动性分析主要关注日常生活中有意义的出行,必须对生成的轨迹进行概括合并,识别出有意义的活动地点和对应的出行。为识别手机轨迹中包含的个体出行起始点信息,本文采用常用的时空点聚类方法^[11]:(1)识别出用户移动轨迹中的锚点;(2)定义用户在不同锚点之间的连续运动为一次出行。

假设某用户第*i*次通话时接入的手机基站位置为 X_i ,总共*n*次通话构成一个时空序列 $X = \{X_1, X_2, \dots, X_n\}$ (称为初始轨迹)。在该序列中,如果连续*s*次通话所接入基站 X_k, \dots, X_{k+s} ($1 < k < k+s < n$) 在空间上相距小于给定的距离阈值 Δd ,即 $\forall i \in [k+1, k+s], \|X_k - X_i\| < \Delta d$,则这些点将被合并成一个新的点 Y_j ,其坐标为相应点

集的中心点 $Y_j = \frac{1}{s+1} \sum_{i=k}^{k+s} X_i$ 。在生成的新轨迹

序列 $Y = \{Y_1, Y_2, \dots, Y_m\}$ 中,如果该用户在新位置 Y_j 通话次数多于1次且停留时间超过给定的阈值 Δt ,则该点被标记为候选出行起始点;否则,将该位置从 Y 中剔除。经过上述处理,可生成候选起始点序列 Y (称为候选轨迹)。最后,按照同序列 X 相同的处理方法,对序列 Y 中的点进行合并和剔除处理,最终生成由*v*个有效的出行起始点组成的轨迹序列 $Z = \{Z_1, Z_2, \dots, Z_v\}$ (称为活动轨迹)。其中,用户在 Z_k 和 Z_{k+1} 之间的移动即代表一次出行。活动轨迹对用户访问空间邻近基站的通话记录进行了合并处理,可以在一定程度上降低手机定位精度和采样频率较低对个体用户轨迹熵计算的影响。

考虑到用户使用手机在时间上具有高随机性,本文舍弃了平均每天通话次数较少的用户记录,只选取平均每天6:00~22:00之间通话次数不少于10次(即保证选取通话活跃的用户),且平均每天通话时段数大于10个(即保证相对均匀的时间采样)的74 026位用户的16 266 590条通话

记录进行出行活动提取分析。通过设置 $\Delta d = 2 \text{ km}$ 和 $\Delta t = 30 \text{ min}$, 最终共提取出 74 015 位用户

在 31 d 内产生的 1 181 944 次出行。其总体统计特征如图 2 所示。

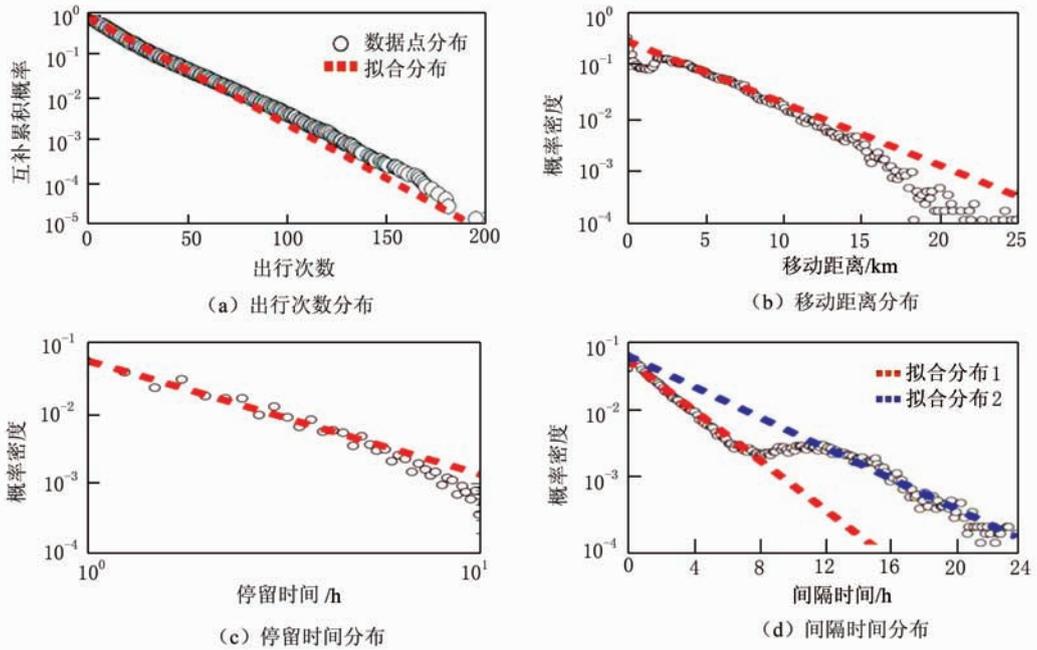


图 2 手机用户出行统计特征分布

Fig. 2 Characteristics of Mobile Users' Travel Activities

1) 用户的出行次数(见图 2(a))呈明显的指数分布($y = a \cdot \exp(-bx)$, $b \approx 16$), 绝大部分用户在整个研究期内出行的次数在 50 次以内, 且多数用户平均每天的出行次数为 2~3 次, 说明城市居民日常出行活动的目的地往往比较有限, 具有很强的可预测性。

2) 出行距离分布(见图 2(b))具有“偏峰”分布的模式, 具体体现为: 在 2~3 km 出行距离处存在一个明显的峰值; 3 km 以内的出行概率随距离的增加而增大; 3 km 以上的出行发生的概率随出行距离的增大而呈指数形式减少($y = a \cdot \exp(-bx)$, $b \approx 0.28$)。可见, 在群体层面上, 居民日常出行的距离往往较短, 只在少数情况下才进行长距离的移动。而且, 在 2~3 km 处出现的峰值反映了手机定位方式空间精度较低产生的影响。

3) 出行活动停留时间的统计分布(见图 2(c))近似服从幂律分布($y = \lambda x^{-\beta}$, $\beta = 1.6$)。这种现象暗示, 群体(每次)出行的目的具有多样性, 个体在日常生活中不同目的的出行次数符合某种(具有快速衰减性质的)特殊分布。其中, 绝大多数种类出行目的的出行活动产生的停留时间较少而固定, 极少数类型的移动活动(如居家、工作等)产生的停留时间较长。

4) 出行活动的间隔时间分布(见图 2(d))具有明显的分段模式, 近似服从分段式的指数分布。

具体可以划分为两个不同的模式: 区间 1(0~8 h)服从指数分布($y = a \cdot \exp(-bx)$, $b \approx 0.55$); 区间 2(12~24 h)服从指数分布($y = a \cdot \exp(-bx)$, $b \approx 0.32$)。此外, 8~12 h 可以被视为过渡区间。以上结果说明大多数活动间隔时间都较短, 同时反映了作息规律对活动间隔的限制。在现实生活中, 个体在睡眠之前的某个时刻常常会因使用电话而被记录一次位置; 然后, 在夜间个体因睡眠而停止活动, 其位置不会被记录; 次日清晨, 个体睡醒之后恢复活动, 不久后某个时刻在工作地(或其他地点)因通话暴露下一次位置。上述过程将产生一次合理的间隔时间, 其时长是睡眠时间、睡眠之外部分停留时间和移动耗时的总和, 其取值范围一般介于 8~12 h 之间。这一现象会导致群体层面间隔时间在 8 h 之后先出现一个明显的上升趋势, 之后恢复下降的趋势, 从而形成了两段式的指数分布模式。

3 轨迹熵分析

熵能够反映个体移动活动的(不)规则程度^[12]。如果个体在固定地点之间的移动占主导地位, 相应的轨迹熵就小。对居民移动轨迹熵进行分析, 可以帮助理解个体移动活动的选择倾向以及可预测程度。因此, 本文分析和讨论了初始熵、空间

熵、条件熵和近似熵等4种不同的轨迹熵形式。

3.1 初始熵

在本文中,初始熵^[9]被定义为用户通话记录初始轨迹 X 的不规则程度度量指标。如 §1 节所述,研究区手机用户总数为 262 万,累计通话次数超过 1.5 亿次,并且可以按照通话次数多少将用户分为 4 种类型(见表 1)。对任一手机用户而言,其初始时间熵为 $S^T = - \sum p(t) \log_2 p(t)$ (其中, $p(t)$ 表示该用户在第 t 个时段发生(至少一次)通话活动的概率, $t \in [1, 24]$)。通过与对应的随机时间熵 $S^{T_R} = \log_2 N^T$ (其中, N^T 为给定用户的总通话时间段数)比较发现(见图 3): (1) 手机用户一天内可能发生通话活动的时间段分布比较广,大多介于 8~18 个时段之间; (2) 通话越活跃(平均每天通话次数越多),通话发生时间的覆盖

范围越广,不确定性越大。

相应地,按照通话活动发生的空间位置,计算每位手机用户的初始空间熵 $S^L = - \sum p(l) \cdot \log_2 p(l)$ (其中, $p(l)$ 表示该用户在第 l 个基站通话的概率, $l \in [1, 557]$)。通过与对应的随机空间熵 $S^{L_R} = \log_2 N^L$ (其中, N^L 为给定用户访问的基站总数)比较发现(见图 4): (1) 用户的通话位置比较固定,经常访问的基站数非常有限(一般少于 10 个); (2) 每天通话次数越多,通话活动可以捕获的用户位置信息越完整。结合时间熵分布,说明用户在少数固定地点发生通话活动的概率比较高,但是通话发生的具体时间具有较大的随机性。这一结论暗示,日常生活中人们习惯在少数固定地点停留比较长的时间^[13],该位置被手机通话记录捕获的机会更大。

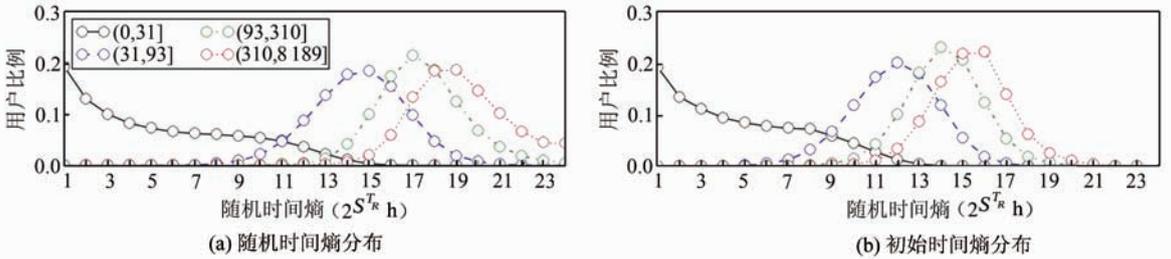


图 3 手机用户初始轨迹时间熵分布

Fig. 3 Temporal Entropy of Mobile Users' Trajectories

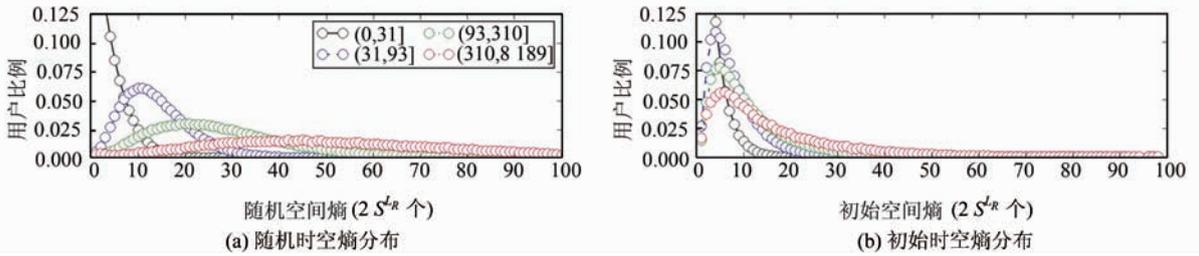


图 4 手机用户初始轨迹空间熵分布

Fig. 4 Spatial Entropy of Mobile Users' Trajectories

3.2 空间熵

初始熵直接依赖于用户在不同位置进行通话活动的频率,但是无法反映用户在不同地点的活动时间差异。因此,本文进一步分析了手机用户的空间熵指标。

给定某个体的连续活动轨迹 $Z: \{\langle x_1, y_1, \Delta t_1 \rangle, \dots, \langle x_n, y_n, \Delta t_n \rangle\}$, 其中, $\langle x_i, y_i \rangle$ 为第 i 个访问地点的空间坐标, Δt_i 为用户在该地点停留的时间。轨迹中的点是按访问时间顺序排列的,相同地点的出现次数反映了个体访问该地点的次数。假设轨迹中相异的地点数为 m , 其空间坐标为 $\{(\bar{x}_1, \bar{y}_1), \dots, (\bar{x}_m, \bar{y}_m)\}$, 定义用户在各个位置

i 出现的时间概率 $p(i)$ 为:

$$p(i) = \frac{\Delta T(i)}{\Delta T} \quad (1)$$

其中,个体在 m 个访问地点的总停留时间 $\Delta T = \sum_{j=1}^m \Delta t_j$, 在地点 i 的停留时间 $\Delta T(i) = \sum_{x_j=x_i, y_j=y_i} \Delta t_j$ 。本文定义个体的空间熵为:

$$E^L = - \sum_{i=1}^m p(i) \log_2 p(i) \quad (2)$$

对 §2 节中提取的 7.4 万手机用户的活动轨迹进行分析发现(见图 5(a)), 其空间熵符合正态分布 $f(x) \sim N(\mu, \sigma^2)$ 。其中,均值和均方根的估

计分别为 $\mu = 1.8$ 和 $\sigma = 2.0$, 拟合优度为 $R^2 = 0.9228$ 。显然, 拟合得到的正态分布的均值非常小, 近似等于 2, 说明多数个体居民的日常移动活动发生的地点都比较少, 平均意义上约为 4 (即 2^2) 个, 即表明个体的活动地点具有明显的规律性, 主要集中在少数几个固定的空间位置。文献[14]通过分析个体日常移动活动的模体结构发现大多数个体的移动活动发生在少数 (一般不超过 5 个) 的固定位置之间, 与本文的发现具有较高的一致性。

值得指出的是, 当计算个体在给定位置的访问概率时, 空间熵严格要求用户访问给定的精确坐标位置点。但是, 现实中个体对空间的利用是基于地点而非精确坐标产生的。因此, 空间熵会对个体移动活动的认知带来一定程度的偏差。本文采用的手机话单数据是由基站定位产生的, 其空间位置实际是基站服务范围的一种概括, 从这一意义上讲, 该数据集适用于个体轨迹熵分析。但是, 由于空间基站的分布与人口密度密切相关, 在人口稠密的城市区域, 许多基站间的间隔距离很小, 因此会出现用户访问同一 (或邻近) 区域但是被不同基站捕获位置而增加空间熵的情况。另一个值得指出的因素是手机基站的服务方式。受限于单个移动基站的服务人数上限, 为了保证服务信号的稳定性, 在人口稠密的区域, 多个基站的空间服务范围一般存在较大的重叠。因此, 即使用户在相同的 (精确) 位置重复出现, 其也可能被不同的移动基站标定不同的空间坐标。解决上述两个问题的一般做法是对空间进行重采样或对移动基站位置做合并处理。本文在 § 2 节中已经对空间邻近的通话记录做了合并处理, 一定程度降低了上述问题对个体用户空间熵计算的影响。

3.3 条件熵

条件熵是对用户轨迹熵度量更有效的指标, 同时具有更直观的实际意义。对于某个移动个体而言, 假定其当前时刻 (或历史) 访问地点序列 X' 已知, 那么其下一位置 Y' 的可选择地点的多样性程度就可以用条件熵来度量:

$$\begin{aligned} E(Y' | X') &= - \sum p(x) E(Y' | X' = x) \\ &= \sum p(x) \sum p(y | x) \log_2 p(y | x) \end{aligned} \quad (3)$$

根据熵的链式规则^[12], 个体连续访问 n 个特定地点的联合熵为:

$$E(Z_1, \dots, Z_n) = \sum_{i=1}^n E(Z_i | Z_{i-1}, \dots, Z_1) \quad (4)$$

因此, 可以直接根据个体的前 $n-1$ 个访问地点历史序列 (称为阶数), 计算个体第 n 个可能出现位置 Z_n 的条件熵:

$$E(Z_n | Z_{n-1}, \dots, Z_1) = E(Z_1, \dots, Z_n) - E(Z_1, \dots, Z_{n-1}) \quad (5)$$

显然, 同空间熵一样, 条件熵对用户的访问位置的空间坐标具有严格的要求。如果直接对个体的轨迹坐标序列进行计算, 将可能高估个体移动的条件熵 (即判定个体的移动路径更多样化)。

3.4 近似熵

为了克服因定位精度导致熵值估计过高的问题, 本文采用近似熵^[15]来度量根据个体历史访问轨迹预测其下一位置的难易程度。通过严格的理论推导, 本文将证明近似熵是条件熵的一种变形。而且, 作为条件熵的一种估计, 近似熵对于空间位置坐标具有更好的鲁棒性。

给定个体的时空活动轨迹序列 $Z = \{Z_1, Z_2, \dots, Z_{N-1}, Z_N\}$ (N 为序列总长度), 可以从中提取出 $(N-m+1)$ 个长度为 m 的连续子活动轨迹序列 $L(i) = \{Z_i, Z_{i+1}, \dots, Z_{i+m-1}\}$ ($i \in [1, N-m+1]$)。进而, 得到一个新的轨迹子序列集合 $L: \{L(1), \dots, L(N-m+1)\}$ 。对于任意两个长度为 m 的子序列 $L(i)$ 和 $L(j)$, 如果两者对应顺序 k 上的访问地点间的距离 $d(k) = \|Z_{i+k}, Z_{j+k}\|$ (其中, $k \in [0, m-1]$) 都小于或等于给定的距离阈值 r , 则标定子序列 $L(i)$ 和 $L(j)$ 相同。

基于上述原则, 可以计算个体轨迹序列 L 中与子序列 $L(i)$ 相同的所有子序列 (包括 $L(i)$ 本身) 出现的概率 $C_i^m(r)$:

$$C_i^m(r) = \frac{n(\{L(k) | d(L(k), L(i)) \leq r\})}{N-m+1} \quad (6)$$

其中, $n(\ast)$ 为集合中元素 (即子序列) 的个数; $d(L(k), L(i))$ 为子序列 k 和 i 间的距离。显然, 当距离阈值 r 趋近于 0 时, 只有完全重合的两个子序列 $L(i)$ 和 $L(j)$ 才会被判定为相同的序列。在这种情况下, $C_i^m(0)$ 等价于个体按子序列 $L(i)$ 进行空间移动的联合概率 $p(Z_i, \dots, Z_{i+m-1})$ 。而且, 在得到的联合概率集合 $\{C_1^m, C_2^m, \dots, C_{N-m+1}^m\}$ 中与子序列 $L(i)$ 相对应的联合概率 C_i^m 将重复出现 $p(Z_i, \dots, Z_{i+m-1}) \times (N-m+1)$ 次。在此基础上, 可以进一步得到以下关系:

$$\begin{aligned} \Phi^m(r) &= \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \log_2 C_i^m(r) = \\ &= \frac{1}{N-m+1} \log_2 \prod_{i=1}^{N-m+1} C_i^m(r) = \end{aligned}$$

$$\frac{\log_2 \prod_{Z_i, \dots, Z_{i+m-1}} p(Z_i, \dots, Z_{i+m-1})^{p(Z_i, \dots, Z_{i+m-1}) \times (N-m-1)}}{N-m+1} =$$

$$\frac{\sum_{Z_i, \dots, Z_{i+m-1}} \{p(Z_i, \dots, Z_{i+m-1})(N-m+1) \log_2 p(Z_i, \dots, Z_{i+m-1})\}}{N-m+1} =$$

$$\sum_{Z_i, \dots, Z_{i+m-1}} p(Z_i, \dots, Z_{i+m-1}) \log_2 p(Z_i, \dots, Z_{i+m-1}) = -E(Z_1, Z_2, \dots, Z_m) \quad (7)$$

根据文献[15]提出的近似熵定义,可最终得到近似熵与条件熵之间的关系:

$$E^C = \Phi^m(r) - \Phi^{m+1}(r) =$$

$$-E(Z_1, Z_2, \dots, Z_m) +$$

$$E(Z_1, Z_2, \dots, Z_m, Z_{m+1}) =$$

$$E(Z_{m+1} | Z_m, \dots, Z_1) \quad (8)$$

该关系说明近似熵是 m 阶条件熵的一种估计。当距离阈值 $r=0$ 时,近似熵等价于条件熵;当距离阈值 $r>0$ 时,其会对邻近的地点做相似性判定,因此使条件熵的估计受精确地理位置的限制减少,从而具有更好的空间鲁棒性,可以进一步降低因手机定位精度较低导致的认知偏差。

考虑到条件熵具有随着给定历史(位置)序列长度增加逐渐递减的性质,本文仅分析了手机用户的一阶近似熵(其中, $m=1$, 且 $r=3$ km)。事实上,在取一阶的情况下,用户的近似熵取值已非常低(见图 5(b)),一般都在 3 以内,显示出极强的可预测性。对所有用户而言,近似熵具有明显的正态分布特点(其中,均值和均方根的估计分别为 $\mu=0.98$ 和 $\sigma=0.64$,拟合优度为 $R^2=0.9724$),暗示居民在进行移动的过程中,目的地的选取具有强烈的倾向。近似熵反映了个体在给定当前位置条件下,进行一次移动的潜在目的地的多样程度。从统计上讲,我们观察到平均意义上个体进行一次出行决策时可能选择的目的地数目约为 2(即 2^μ)个,这一数值与文献[9]的观测具有高度一致性。

结合空间熵和近似熵分析结果,可以得出以下结论:群体水平上,城市居民的日常固定活动地点数一般较少,统计上平均值为 4 个左右;而且,在这些固定地点之间移动时目的地选择具有倾向性,一般存在两条频繁的移动路径。这些特点说明个体的移动行为是高度规律的,移动活动的模式容易多次重复出现,因此可预测性较高。

4 结 语

城市居民日常活动具有非弹性活动与弹性活动的结构性特征。其中,日常活动模式常以非弹

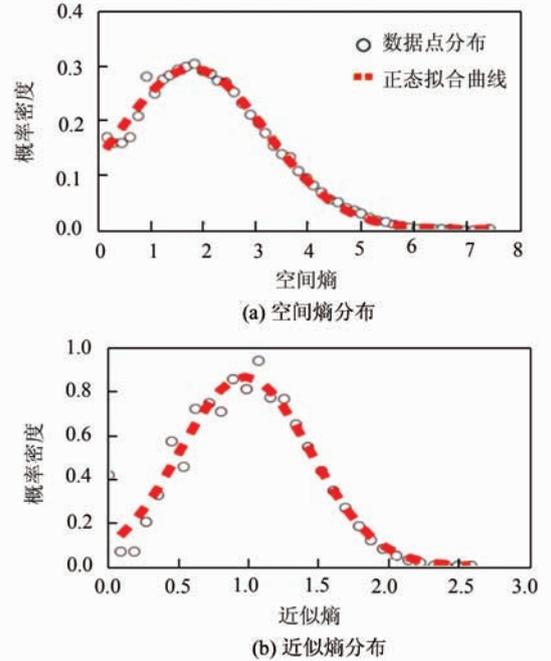


图 5 手机用户活动轨迹熵分布

Fig. 5 Entropy of Mobile Users' Travel Activities

性活动为主要框架,这类活动在活动安排中居于主要位置,而弹性活动因时空间上的灵活性和可调整性而相对地处于次要位置。因此,个体居民的日常移动活动所受的时空间制约非常显著,表现为存在少数几个固定的活动地点以及频繁的位置访问序列模式^[16],从而大大降低了个体移动活动预测的难度。

轨迹时空熵分析作为人群活跃度异质性的一种反映,可以为传统的人文地理学(尤其是时间地理学)研究人类出行活动提供新的量化分析工具。例如,在现实应用中,时空熵分析方法有助于揭示个体居民移动活动的主观意愿和惯性模式。通过分析居民移动活动内在的重复性、规律性和可预测性,可以为交通需求预测、位置精准推荐等应用系统提供必要的辅助信息,提高系统的有效性。此外,通过分析居民的频繁活动场所(作为主观“出行-活动”意愿的一种体现),也能帮助揭示个体居民对于城市空间利用的不均衡性,为研究“职-住分离”、“规划瑕疵”等城市管理问题提供量化分析方法。简而言之,轨迹时空熵分析正逐

渐成为分析、理解海量居民出行规律的一般性方法。

但是,以手机话单为代表的新数据被广泛用于分析城市个体居民移动活动的时空特征容易产生具有偏差的发现与结论。通过探讨从原始话单记录中进行出行识别的方法及其必要性,本文提供了一种克服手机定位数据采样频率较低缺陷的途径。更重要的是,本文提出了一种考虑空间邻近性的轨迹熵特征分析方法。该方法则具有强空间鲁棒性,可以减少手机定位数据空间精度较低带来的影响。以上探索有助于更加准确地发现城市居民的出行活动规律,进而更加客观地揭示城市内部的人群活动动态。

参 考 文 献

- [1] Liu Y, Liu X, Gao S, et al. Social Sensing: A New Approach to Understanding Our Socio-Economic Environments[J]. *Annals of the Association of American Geographers*, 2015, 105(3):512-530
- [2] Ratti C, Frenchman D, Pulselli R M, et al. Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis[J]. *Environment and Planning B-Planning & Design*, 2006, 33:727-748
- [3] Liu L, Andris C, Ratti C. Uncovering Cabdrivers' Behavior Patterns from Their Digital Traces [J]. *Computers, Environment and Urban Systems*, 2010, 34(6):541-548
- [4] Wesolowski A, Eagle N, Tatem A J, et al. Quantifying the Impact of Human Mobility on Malaria[J]. *Science*, 2012, 338(6 104):267-270
- [5] Ahas R, Aasa A, Roose A, et al. Evaluating Passive Mobile Positioning Data for Tourism Surveys: An Estonian Case Study[J]. *Tourism Management*, 2008, 29(3):469-486
- [6] Zhou Tao, Han Xiaopu, Yan Xiaoyong, et al. Statistical Mechanics on Temporal and Spatial Activities of Human[J]. *Journal of University of Electronic Science and Technology of China*, 2013, 42:481-540(周涛,韩筱璞,闫小勇,等. 人类行为时空特性的统计力学[J]. 电子科技大学学报, 2013, 42:481-540)
- [7] Brockmann D, Hufnagel L, Geisel T. The Scaling Laws of Human Travel[J]. *Nature*, 2006, 439(7 075):462-465
- [8] González M C, Hidalgo C A, Barabási A L. Understanding Individual Human Mobility Patterns [J]. *Nature*, 2008, 453:779-782
- [9] Song C, Qu Z, Blumm N, et al. Limits of Predictability in Human Mobility[J]. *Science*, 2010, 327(5 968):1 018-1 021
- [10] 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
- [11] Calabrese F, Lorenzo G D, Liu L, et al. Estimating Origin-Destination Flows Using Mobile Phone Location Data[J]. *IEEE Pervasive Computing*, 2011, 10(4):36-44
- [12] Cover T M, Thomas J A. Elements of Information Theory[M]. New York: Wiley, 1991
- [13] Sevtsuk A, Ratti C. Does Urban Mobility have a Daily Routine? Learning from the Aggregate Data of Mobile Phone Networks [J]. *Journal of Urban Technology*, 2010, 17(1):41-60
- [14] Schneider C M, Belik V, Couronné T, et al. Unravelling Daily Human Mobility Motifs[J]. *Journal of the Royal Society Interface*, 2013, DOI: 10.1098/rsif.2013.0246
- [15] Pincus S M. Approximate Entropy as a Measure of System Complexity[J]. *The National Academy of Sciences*, 1991, 88(6):2 297-2 301
- [16] Eagle N, Pentland A S. Eigenbehaviors: Identifying Structure in Routine[J]. *Behavioral Ecology and Sociobiology*, 2009, 63(7):1 057-1 066

An Analysis of Entropy of Human Mobility from Mobile Phone Data

KANG Chaogui¹ LIU Yu² WU Lun²

1 School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

2 School of Earth and Space Sciences, Peking University, Beijing 100871, China

Abstract: Human mobility patterns have been intensively investigated by scientists from computational social science, statistical physics, and complexity science. In last decade, mobile phone data have pro-

(下转第 129 页)

spheres intersecting segments were efficiently located using the AABB-tree algorithm. Then, initialized by the sphere center, initial segment-surface intersections approximated within bounding spheres were parallel-computed by iterative segment-sphere intersection. Finally, the surface was meshed by Delaunay refinement, unambiguously. This approach can reconstruct surfaces with a good aspect ratio in comparison with the matching-cube algorithm. Experiments show that the new algorithm can efficiently, robustly, and accurately reconstruct surfaces from point clouds with high noise, but the time cost and memory consumption rapidly increases when executing precise models.

Key words: Delaunay refinement; sphere fitting; noisy point clouds; surface reconstruction

First author: LI Guojun, master, specializes in surface reconstruction and time-frequency measurement. E-mail:1010551750@qq.com

.....
(上接第 69 页)

vided an unprecedented tool for capturing individual travel activities in space and time. However, the temporal sparsity and spatial imprecision of this data imposes significant biases upon the derived mobility patterns. Two efficient techniques to cope with this issue are proposed. First, we implement an activity-location and travel-OD identification method to reconstruct reliable trajectories from call detailed records for mobile users. Second, we introduce the approximate entropy, which is superior to conditional entropy, for quantifying the regularity of individuals' consecutively visited locations. With a case study in Harbin, the proposed approaches enable us to uncover meaningful patterns of urban mobility in terms of frequently and consecutively visited locations.

Key words: call detailed records; digital traces; displacement identification; mobility patterns; trajectory entropy

First author: KANG Chaogui, PhD, assistant professor, specializes in the theories and methods of urban informatics and trajectory computing. E-mail: cgkang@whu.edu.cn

Foundation support: The National Natural Science Foundation of China, No. 41601484; the Open Research Fund Program of Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, No. 15S01.