

基于手机信令轨迹点识别的职住地综合决策算法*

陈略, 熊宸, 蔡铭

中山大学智能工程学院 / 广东省智能交通系统重点实验室, 广东 广州 510006

摘要: 由于手机信令存在稀疏性和空间不确定性等特点, 给职住地识别带来一定难度; 同时, 当前的手机信令职住地识别算法对于影响识别结果的关键要素, 如时间规则和空间聚合距离等的设置存在依据不足且有效性未知等问题。针对以上难点, 提出了一种基于手机信令轨迹点识别的职住地综合决策算法。通过轨迹点识别消除了手机信令的空间不确定性, 有效界定了停留区域的时空边界。在停留点识别的基础上, 提取各时段中停留点的分布特征, 计算时段信息熵, 从而度量时段属于职住时段的可能性, 对时段赋权值。通过对任意语义停留点到达和离去时间轴离散化, 组合语义停留点的停留时段, 并通过时段权值反映停留时段特征。以语义停留点的停留时段和停留时长为特征构建职住地综合决策矩阵, 识别可能度最大的语义停留点作为职住地, 并采用带有移动停留标签的信令数据验证算法的有效性。

关键词: 手机信令; 轨迹点识别; 信息熵; 职住地; 决策矩阵

中图分类号: U491.2⁶² **文献标志码:** A **文章编号:** 2097-0137 (2022) 02-0106-11

A comprehensive decision-making algorithm of residence and workplace based on the identification of cellular signaling track points

CHEN Lue, XIONG Chen, CAI Ming

School of Intelligent Systems Engineering, Sun Yat-sen University / Guangdong Key Laboratory of Intelligent Transportation Systems, Guangzhou 510006, China

Abstract: Due to the characteristics of cellular signaling, such as sparsity and spatial uncertainty, it is difficult to identify the location of workplace and residence. At the same time, the current cellular signaling residence and workplace identification algorithm is an insufficient basis and unknown effectiveness for the key elements affecting the identification results, such as time rules and spatial aggregation distance. In view of the above difficulties, this paper proposes a comprehensive decision-making algorithm based on cellular signaling track point identification, which eliminates the spatial uncertainty of mobile signaling and effectively defines the space-time boundary of the stay area. Based on the recognition of stay points, the distribution characteristics of stay points in each period are extracted, and the information entropy of each period is calculated to measure the possibility that the period belongs to the residence and workplace period, and the weighted value of the period is given in this paper. By discretizing the arrival and departure time axis of any semantic stay point, the stay time slot of semantic stay point is combined, and the characteristics of stay time slot are reflected by the time slot weight. Based on the characteristics of the stay time slot and stay time, a comprehensive decision matrix of residence and workplace is constructed to identify the most likely semantic stay point as the residence and

* 收稿日期: 2020-03-27 录用日期: 2020-04-14 网络首发日期: 2021-05-21

基金项目: 国家重点研发计划 (2018YFB1601001); 高校基本科研业务费 (19lgpy290)

作者简介: 陈略 (1994年生), 女; 研究方向: 数据挖掘与分析; E-mail: 394371296@qq.com

通信作者: 蔡铭 (1977年生), 男; 研究方向: 数据挖掘与分析; E-mail: caiming@mail.sysu.edu.cn

workplace, and the cellular signaling with stay tag is used to verify the effectiveness of the algorithm.

Key words: cellular signaling; identification of cellular signaling track points; information entropy; residence and workplace; decision matrix

近年来手机信令由于其覆盖范围广、采样率高等特点被广泛应用于城市职住空间组织结构、居民通勤特征等研究。张天然利用历史每日20:00-次日8:00的夜间手机数据训练识别,以出现概率最高且比例超过60%的区域作为用户的居住地。同理,利用历史工作日9:00-18:00的手机数据训练识别工作地^[1]。丁亮等选取早晚特殊时间点,若某一手机用户每天至少有3个时间点在同一基站或附近1000 m内的基站,就将该基站识别为该日该用户的工作地或居住地^[2]。许宁等根据深圳市企事业单位工作时段和深圳市居民生活习惯,设定工作时段为09:00-18:00,夜间睡眠时段为00:00-06:00,筛选用户在工作(居住)时段累计停留时间超过阈值的基站作为该用户的工作(居住)地所对应的基站^[3]。职住地识别方法虽然已有众多成果发表,但大多文献更关注职住地识别结果的应用,极少关注职住地识别算法本身的可靠性。

由于手机信令的固有特性以及当前职住地识别算法存在的问题,在实际应用中仍有诸多问题亟待解决。本研究通过阐述手机信令特性,剖析目前职住地识别算法中的关键因素,并提出了解决这些问题的职住地识别算法。

1 问题描述

手机信令数据是一种典型时间序列数据,由基站的经纬度和时间戳构成。手机信令数据影响职住地识别的两个重要特性为稀疏性和空间不确定性。

由于手机信令产生方式的无规律性,导致其采样频率非常不规则。与一般时空轨迹数据不同,手机信令包括主动和被动两种产生方式。主动产生为数据流量上网、发短信、拨号等行为触发的基站响应;被动产生包括收短信、接电话以及心跳数据等行为触发的基站响应。因此手机数据的时间密度因用户而异,并且非常不规则。通常在某一小段时间内,手机数据密集出现,在较长且不规则的时间后,手机数据再次密集发生^[4]。手机数据的稀疏特性使得轨迹点时间间隔较大,CDR数据的时间间隔可以达到数小时之久^[5],而信令

数据的时间间隔通常也超过1 h^[6],长时间间隔的轨迹无法分析用户在间隔时段的定位,因此在通常职住地识别算法中,如果以基站经纬度位置点为停留单元,通过信令逐行相减的方法得到每一个用户在每一个位置的停留时间^[7]。当相邻的长时间间隔轨迹点位置不相同,将该时段纳入基站停留时间,可能会导致增加虚假的停留点或者基站停留时长被夸大。

手机信令中大量存在的乒乓和漂移数据导致手机信令数据存在着较大的空间不确定性。文献指出,乒乓切换和漂移数据约占数据总量的30%^[8]。乒乓切换和漂移数据是由手机信令的定位原理和基站工作机制所产生的。用户定位以其所在基站覆盖范围内的基站位置所表示,而我国目前城市内基站布设普遍比较密集,基站的覆盖范围相互重叠区域大。处于多个基站覆盖范围内的地点地理位置受基站信号强度的变化而不断切换基站,因而产生乒乓切换现象。一般来说基站有负荷优化调节机制,当邻近基站的用户负荷过大,会自动切换更远但用户负荷更少的基站,从而产生了信号漂移^[4]。乒乓切换和数据漂移并非用户真实的移动轨迹,这类震荡的轨迹点具有紧密的时空关系,可视为一个整体,表示同一个语义地点,若把乒乓切换和漂移数据当作普通轨迹数据处理,则会多出许多虚假的语义地点,真正的语义地点停留时间则会被削弱。

从手机信令广泛存在的稀疏性和空间不确定性可以看出,以基站为对象单元,研究职住地识别算法不合理,存在较大的误差。

1.1 时间规则

职住地识别算法涉及的重要因素包括时间规则和空间聚合距离^[9]两个部分。职住地识别方法的时间规则可以归纳为:

1.1.1 时间阈值法 划分居住时间段,例如20:00至次日8:00。通常将居民每天夜间时间段内停留时间超过阈值时间且一个月内出现次数大于阈值次数的停留点,作为居民的居住地^[7]。划分日间时间段,以同样的方法识别工作地^[10]。

时间阈值法的职住地识别率完全由划分时间段、停留时间阈值和天数阈值的经验值决定,具

有较强的不确定性。

1.1.2 累计停留时间法 划分居住时间段和工作时间段,选取时间段内停留时间最长的停留点作为居住地和工作地^[11-13]。累计停留时间的变形算法将全天划分为多个时窗,依次计算轨迹与每个时窗的叠合度,将轨迹点停留时间区间转换为连续的多个时窗,通过多日轨迹叠加识别每个时窗的多日稳定点。合并距离小于阈值的多日稳定点的时窗,选取日间和夜间时间段内停留时间最长的多日稳定点作为工作地和居住地^[14]。累计停留时间法的识别结果与划分时间段直接相关。

1.1.3 信息熵法 将观测时间段内的各停留点停留时长与相对观测时长的比例作为一种信息输入,通过信息熵反映划分时段内用户活动的强度^[7],信息熵越小,活动强度越小,居住或工作停留的确定性越强。

计算时间段内各停留点 U_1, U_2, \dots, U_n 的停留时间 T_i 相对总观测时长 T 的比例 p_i 、信息熵 $H(U)$ 。若信息熵小于设定信息熵阈值,则将其中停留时间概率最大的停留点识别为居住地和工作地。即

$$p_i = T_i / (\sum_{j=1}^n T_j), \quad H(U) = -\sum_{i=1}^n p_i \log_2 p_i. \quad (1)$$

信息熵法可以算是累积停留时间法的变形,但信息熵的阈值不容易确定,且信息熵与划分时段相关。

通过时间规则可知,识别结果与划分时间段紧密相关,但时间段划分由人为决定,没有度量的指标,划分的有效性未知,且忽视了个体不同的作息规律。

1.2 空间聚合距离

由于移动通信网络空间不确定性的特点,用户定位的基站并不一定所处实际位置最邻近的基站。因此需要对一定空间距离的基站进行聚合,并视为同一位置点。若以基站作为研究单元,不设置空间聚合距离,则基站间的震荡切换会被识别为不同的停留位置,实际停留位置的停留时间变短,实际停留位置将被切分成多个停留位置,停留时间可能无法达到时间规则的要求;将基站按一定的算法空间聚合,若聚合距离过大会将出行的轨迹点也纳入停留位置,导致真正的停留位置时长变长,从而导致职住地识别错误,同时职住地理位置准确度也会受到影响,因此选择空间聚类算法正确界定停留区域的边界对职住地的识别至关重要。倾向于职住结果应用的职住地识别算法中对空间测度的描述一般直接采用基站为研

究单元,如以基站经纬度位置点为单元,通过信令逐行相减的方法得到每一个用户在每一个位置的停留时间^[7],但这种操作会导致职住地识别算法的有效性降低。以距离邻近的基站空间聚合的停留点为研究单元,空间聚类算法一般定义一定的距离阈值作为合并的依据,如文献^[14-15]。但距离阈值的大小对空间测度影响有待研究。对于停留区域边界的界定通过停留时间影响时间规则进而影响算法对职住地的识别。总结上述职住地识别算法可以发现:

(1) 时间规则对时段划分和停留时长阈值等参数的选择依据不足,而这些参数会对职住地识别的有效性产生严重影响。

(2) 以基站为停留研究单元,不对基站进行空间聚合,忽视了手机信令的稀疏性和空间不确定性,直接削弱实际停留位置的停留时间;以基站聚类的停留区域为停留单元,空间算法对停留边界识别的准确度会影响时间规则内停留时长的度量,以及职住地的判断和职住位置的准确度。

(3) 职住地识别是一个综合性决策问题,需同时考虑停留时段、停留时长、停留天数等特性,但目前算法先对职住时段划分,再统计划分时段内的停留时间,缺乏对停留时段和停留时长等特征的综合分析。

(4) 目前职住地识别算法只能定性地获得职住地的地理位置而无法获得其职住规律特性,如职住的时段和时长。

因此,本研究提出了一种基于停留点语义的职住地决策方法。

2 算法流程

本研究提出的基于手机信令轨迹点识别的职住地综合决策算法流程,如图1所示。其中,联结同化流程和时空密度聚类算法已在文献^[16]中详细介绍,本文仅着重于轨迹点识别中的空间簇联合以及职住地判断中的相关算法。

2.1 轨迹簇空间联合识别停留区域

时空联结同化流程已经将具有停留特点的时间紧密性的网格位置联结成簇,但由于受手机信令的基站定位和调节机制影响,同一语义的活动地点的空间位置通常不止一个基站,由于乒乓切换和数据漂移发生的无规律性,同一语义地点的不同基站之间存在时间连续的但不发生切换的情况,此时同一语义地点的不同基站表现为两个不

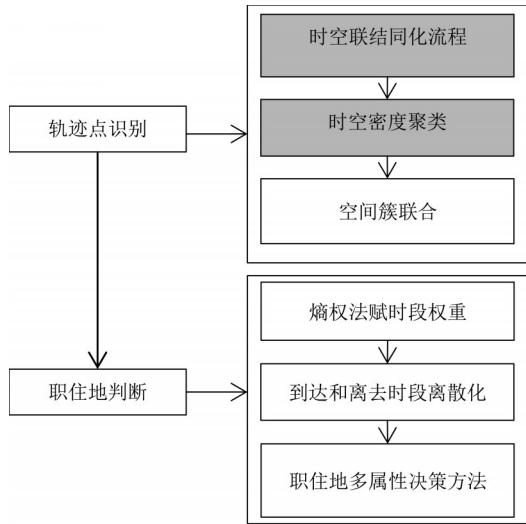


图 1 算法流程

Fig. 1 Algorithm process

同的轨迹簇; 同时由于手机信令采样间隔不均等特点, 若同一语义地点的不同基站坐标之间时间间隔较大, 缺乏时空紧密性, 同一语义地点会被分成不同的轨迹簇。

以上情况均会切割真正的停留点, 削弱停留时长, 对职住地的时间规则识别造成影响。因此, 空间簇联合目的是通过簇的长时间的定位规律识别区分地点语义, 进一步区分停留区域的边界, 以弥补手机信令的采样间隔造成的空间不确定性, 并确定时间不紧密的簇之间的空间关系。

考虑到有意义的活动语义地点是经常发生且具有规律性的, 可以通过多次的时空紧密的轨迹簇之间位置点的相互交集, 将具有位置点交集的簇联合为同一语义地点。多天出现的语义地点定义为固定语义地点, 表示经常性停留或经过的语义地点。通过识别固定语义点可以更加锚固工作地、居住地以及餐厅等经常性停留的活动地点, 使得职住地的识别更加准确。

固定语义地点通过多天的交集轨迹簇重叠空间联合确定。将含有相同网格位置标号的轨迹簇联合成为同一语义地点。轨迹簇空间联合识别语义地点如表 1 所示, 以簇 6 为例, 簇 6 是一个典型的震荡的时空联结同化簇, 簇中含有的位置点 (2037, 1094) 和 (2041, 1093) 分别用红框和蓝框表示, 簇 2、3、5 中均含有簇 6 中的位置点, 因此可以将原本缺乏时间紧密性的簇 1、2、3, 以及簇 5、6 空间联结起来, 识别为同一语义 c3。

语义地点的识别结果如图 2 所示。同样颜色的位置点代表同一语义点。从图中可以看出, 即使

表 1 轨迹簇空间联合识别语义地点
Table 1 Spatial association of trace clusters to identify semantic locations

时间戳	簇标号	网格标号	语义地点
2020-01-10 09: 49: 20	0	(2043, 1085)	c8
2020-01-10 10: 39: 43	1	(2041, 1093)	c3
2020-01-10 11: 05: 33	1	(2041, 1093)	c3
2020-01-10 11: 05: 38	2	(2037, 1094)	c3
2020-01-10 11: 43: 38	2	(2037, 1094)	c3
2020-01-10 11: 43: 50	3	(2041, 1093)	c3
2020-01-10 12: 13: 29	3	(2041, 1093)	c3
2020-01-10 12: 56: 09	4	(2052, 1086)	c5
2020-01-10 13: 19: 19	4	(2052, 1086)	c5
2020-01-10 13: 19: 24	5	(2041, 1093)	c3
2020-01-10 14: 27: 50	5	(2041, 1093)	c3
2020-01-10 14: 27: 55	6	(2037, 1094)	c3
2020-01-10 14: 43: 10	6	(2037, 1094)	c3
2020-01-10 14: 43: 15	6	(2041, 1093)	c3
2020-01-10 14: 45: 00	6	(2041, 1093)	c3
2020-01-10 14: 45: 05	6	(2037, 1094)	c3
2020-01-10 14: 45: 15	6	(2037, 1094)	c3
2020-01-10 14: 45: 20	6	(2041, 1096)	c3
2020-01-10 14: 45: 25	6	(2041, 1096)	c3
2020-01-10 14: 45: 30	6	(2037, 1094)	c3
2020-01-10 14: 47: 10	6	(2037, 1094)	c3
2020-01-10 14: 47: 15	6	(2041, 1093)	c3
2020-01-10 14: 47: 30	6	(2041, 1093)	c3
2020-01-10 14: 47: 36	6	(2037, 1094)	c3
2020-01-10 18: 31: 46	6	(2037, 1094)	c3
2020-01-10 18: 31: 51	7	(2047, 1083)	c7
2020-01-10 18: 32: 06	7	(2046, 1082)	c7
.....

语义地点之间距离并不远, 但语义的区分度却很明显, 没有出现大片轨迹点连成同一语义地点的情况。其中, 从上到下圆圈中的语义地点依次为工作地、居住地、餐厅, 与用户标签地点一致, 可见规律停留语义地点的网格位置之间常有切换, 可通过含有相同位置的轨迹簇联合进一步锚固停留语义点。而且通过多天轨迹簇没有交集可以区分距离靠近的不同语义点。

对比簇空间联合前后停留点识别精度, 如表 2 所示。簇空间联合后的空间精度和时间精度均高于簇空间联合前, 说明簇空间联合识别语义地点



图2 语义点识别结果

Fig.2 Semantic point recognition results

能够使得停留区域更锚固, 时间识别精度更高, 且簇联合固定语义能够有效地识别停留区域边界。

2.2 职住地判断

2.2.1 熵权法赋时段权重

由于职住地的停留时段比其他活动的停留时段有更强的规律性, 表现

在职住地停留时段与其他活动时段相比停留次数更多且停留语义地点比较单一和固定, 因此可以计算时段语义熵来表征处于职住时段的可能性。某时段内停留次数越多, 停留语义地点越固定, 时段熵越小, 表明该时段为职住时段的可能性越大; 反之, 熵越大, 表明该时段的停留地点的偶然性越大, 即该时段为职住时段的可能性越小。

对时间轴各时间段内停留的各语义进行统计, 某时间段 i 的语义熵由该时段数据集内被访问的次数 n_i 和语义停留区域的分布构成。 p_{ij} 为语义 j 的停留次数在该时段 i 访问次数中出现的概率。 k 为时段 i 内出现的语义停留区域的个数, 时段熵为

$$H(i) = -\frac{1}{\ln n_i} \sum_j^k p_{ij} \ln p_{ij}. \quad (2)$$

可以看出, 相同语义停留区域的分布下, 被访问次数 n_i 大的时段信息熵更小; 相同的访问次数 n_i 下, 语义停留区域分布确定性更强的时段信息熵更小。以表3为例, 某时段语义停留区域 A、B、C 各访问3次与 A、B、C 各访问2次的分布相同, 但被访问次数多 (A、B、C 各访问3次) 的时段语义熵更小; 当时段被访问总次数 n_i 相同时, 如 A、B、C 各访问3次与 A 访问8次、B 访问1次相比, 语义停留区域分布更均匀的信息熵更大, 时段不确定性更强。

根据时段语义熵对时段赋权值。时段语义熵

表2 簇空间联合前后停留点识别精度的对比

Table 2 Comparison of recognition accuracy of retention points before and after cluster space combination

参数	簇空间联合前	簇空间联合后
标记的停留区域个数	26	26
算法检测停留区域个数	34	26
相符的停留区域个数	25	25
停留区域准确率 P	0.735	0.962
停留区域召回率 R	0.962	0.962
停留区域值	0.833	0.962
标记的停留时间	6 days 08: 54: 15	6 days 08: 54: 15
算法检测的停留时间	5 days 19: 24: 30	6 days 01: 21: 08
相符的停留时间	5 days 19: 24: 30	6 days 01: 21: 08
停留时间准确率 P	1.0	0.966
停留时间召回率 R	0.912	0.951
停留时间值	0.954	0.975

表3 时段熵示例
Table 3 Period entropy example

某时段 <i>i</i> 的语义停留区域分布	时段语义熵 <i>H(i)</i>
A (2)、B (2)、C (2)	0.613
A (3)、B (3)、C (3)	0.50
A (8)、B (1)	0.159

更小的, 权值更大, 有

$$w_i = \frac{1 - H(i)}{\sum_{i=1}^k (1 - H(i))} \quad (3)$$

2.2.2 到达和离去时段离散化 为了提取用户出行、停留的时段和时长等特征, 根据语义停留区域的到达、离去时间分布对连续时间轴进行离散化, 合理划分时间段, 使得区间既能有效区分不同时段分布特征又不过度分散独立。本研究使用基于数据分布特征信息熵聚类的MDLP算法^[17-18]对数据进行聚类并划分时间轴, 以避免划分后破坏数据的分布特征。MDLP算法如表4所示。MDLP离散化以任意语义点Cm到达(离去)时间段的次数为特征, 依据信息熵聚类时间段。具体步骤如下:

(1) 先将时间轴划分为*k*个基本时间段 D_i ($1 \leq i \leq k$), 计算任意语义Cm停留区域在多天数据时间轴上的总信息熵 $E(Cm)$, 如式(4)所示。

$$E(Cm) = -\sum_{i=1}^k p(D_i) \ln p(D_i), \quad (4)$$

其中 $p(D_i) = |D_i|/N$ 表示语义停留区域Cm在时间段 D_i 中的概率, D_i 表示语义Cm在时段 D_i 中到达(或离去)时刻的次数, $N = \sum_{i=1}^k |D_i|$ 表示语义Cm到达(或离去)时刻在整个时间轴上的次数。同时计算当前时间轴序列包含的基本时间段数*k*(表4中算法1的3~5)。

(2) 遍历每两个时间段 D_i 和 D_{i+1} 之间的候选划分 (D_i, D_{i+1}) , 计算划分后左右两部分的总加权信息熵 $E(D_i, D_{i+1})$, 如式(5)所示。

$$E(D_i, D_{i+1}) = \frac{\sum_{j=1}^i |D_j|}{N} E(Cm^L) + \frac{\sum_{j=i+1}^k |D_j|}{N} E(Cm^R). \quad (5)$$

选出划分后总加权信息熵 $E(D_i, D_{i+1})$ 最小的候选划分 (D_i, D_{i+1}) 。式(5)中, $E(D_i, D_{i+1})$ 计算方法为对切分点 (D_i, D_{i+1}) 左右两部分加权信息熵, 左边权重为语义Cm在切点左边时段 D_j ($1 \leq j \leq i$)

表4 信息熵MDLP离散化具体算法伪代码
Table 4 Information entropy MDLP discretization algorithm code

MDLP_Cm	
输入:	$D = \{D_1, D_2, \dots, D_i, \dots, D_k\}$ //时间轴基本时间段的数 据集
输出:	$cutD = \{(D_{i_0}, D_{i_0+1}), (D_{i_1}, D_{i_1+1}), \dots, (D_{i_n}, D_{i_n+1})\}$ ($1 \leq i_0 \leq i_1 \leq i_n < k$) //时间轴的切点集合
1	$cutD = []$
2	{
3	$E(Cm) = -\sum_{i=1}^k p(D_i) \ln p(D_i)$ //计算时间轴序列的总体 信息熵 $E(Cm)$
4	$N = \sum_{i=1}^k D_i $ //计算时间轴序列中包含的到达(或离去) 总次数 <i>N</i>
5	$k = \text{len}(D)$ //计算时间轴序列包含的基本时间段数 <i>k</i>
6	$ED_lis = []$ //记录遍历切点的划分信息熵
7	for i in range ($k-1$): //遍历候选时间轴切点
8	计算切分后的左右两部分加权信息熵 $E(D_i, D_{i+1})$
9	$ED_lis.append(E(D_i, D_{i+1}))$
10	$(D_i, D_{i+1}) \leftarrow \min(ED_lis)$ //选出切分后信息熵 最小的切点为最优切点 (D_i, D_{i+1})
11	$k_l, k_r \leftarrow (D_i, D_{i+1})$ //分别计算时间轴序列切分后的左 右两部分的基本时间段数
12	$\text{Gain}(D_i, D_{i+1}) = E(Cm) - E(D_i, D_{i+1})$ //计算切分前后 信息增益 $\text{Gain}(D_i, D_{i+1})$
13	if $\text{Gain}(D_i, D_{i+1})$ satisfies ():
14	$cutD.append((D_i, D_{i+1}))$
15	MDLP ($\{D_1, D_2, \dots, D_i\}$)
16	MDLP ($\{D_{i+1}, D_{i+2}, \dots, D_k\}$)
17	}

中的累计到达(或离去)时刻的次数 $\sum_{j=1}^i |D_j|$ 与语义Cm到达(或离去)时刻在整个时间轴上的次数*N*的比值 $\sum_{j=1}^i |D_j|/N$, 右边权重为语义Cm在时段 D_j ($i+1 \leq j \leq k$)中的累计到达(或离去)时刻的次数 $\sum_{j=i+1}^k |D_j|$ 与语义Cm到达(或离去)时刻在整个

时间轴上的次数 N 的比值 $\sum_{j=1}^i |D_j|/N$ (表 4 中算法 1 的 8~11)。

(3) 计算最小的候选划分 (D_i, D_{i+1}) 的信息增益 $\text{Gain}(D_i, D_{i+1})$ 如式 (6)。

$$\text{Gain}(D_i, D_{i+1}) = E(\text{Cm}) - E(D_i, D_{i+1}). \quad (6)$$

检验信息增益是否满足切分准则式 (7), 若满足, 则对切分后的左右时间轴 Cm^L 和 Cm^R 递归 MDLP 切分时间轴; 否则停止切分 (算法 1 的 12~16)。

$$\text{Gain}(D_i, D_{i+1}) > \frac{\ln(N-1) + \ln(3^k - 2)}{N} - \frac{kE(\text{Cm}) - k_l E(\text{Cm}^L) - k_r E(\text{Cm}^R)}{N}. \quad (7)$$

2.2.3 职住地多属性决策方法 到达和离去时间轴离散化得到切点的集合 $\text{cut}D = \{(D_{i_0}, D_{i_0+1}), (D_{i_1}, D_{i_1+1}), \dots, (D_{i_n}, D_{i_n+1})\}$, 根据切点划分时间轴, 得到对应的时间轴切片的时间段集合

$$T\text{set} = \{T_{1,i_0}, T_{i_0+1,i_1}, T_{i_1+1,i_2}, \dots, T_{i_{n-1}+1,i_n}, T_{i_n+1,k}\}, \quad (8)$$

其中 $T_{x,y}$ 为基础时段 D_x 到 D_y 的时段集合,

$$\begin{aligned} T_{1,i_0} &= \{D_0, D_1, \dots, D_{i_0}\}, \\ T_{i_0+1,i_1} &= \{D_{i_0+1}, D_{i_0+2}, \dots, D_{i_1}\}, \\ T_{i_{n-1}+1,i_n} &= \{D_{i_{n-1}+1}, D_{i_{n-1}+2}, \dots, D_{i_n}\}, \\ T_{i_n+1,k} &= \{D_{i_n+1}, D_{i_n+2}, \dots, D_k\}. \end{aligned}$$

以此类推。基于到达和离去的时间段可以组合成语义 Cm 停留区域各停留时间段的区间数, 假设到达时间段为 T_{a_l, a_u} , 离去时间段为 T_{b_l, b_u} , 则组成的停留时间段区间 $l_i (1 \leq i \leq s)$ 为 $[T_{a_l, b_l}, T_{a_u, b_u}]$, s 为语义 Cm 停留区域所属的停留时间段个数。根据时段语义熵的时段权值将停留时间段的区间数转化为时间段权值区间数 $W_l = [W_{T_{a_l, b_l}}, W_{T_{a_u, b_u}}]$, 其中

$$W_{T_{a_l, b_l}} = \sum_{j=a_l}^{b_l} w_j, \quad W_{T_{a_u, b_u}} = \sum_{j=a_u}^{b_u} w_j,$$

通过时间段权值区间数可以表示停留时段属于职住时间段的可能性值。统计多天语义 Cm 停留所属的各个停留区间 l_i 以及次数 n_i , 即: $N = \sum_{i=1}^s n_i$ 。则, 语义 Cm 停留时段的属性值为多天停留时段属性值之和, 即

$$W(\text{Cm}) = \sum_{l=1}^s n_l W_l = [\sum_{l=1}^s n_l W_{T_{a_l, b_l}}, \sum_{l=1}^s n_l W_{T_{a_u, b_u}}].$$

统计语义 Cm 停留时段区间 l_i 中的每次停留时长最大最小值, 得到停留时长 st_l 区间, 语义 Cm 的停留时长区

间为多天停留时段区间 l_i 的停留次数 n_i 与停留时长 st_l 的乘积之和, 即 $\text{ST}(\text{Cm}) = \sum_{l=1}^s n_l st_l$ 。

正如前文所述, 职住地的判断是一个区间数多属性权重未知的决策问题。信息熵越大, 信息的不确定性越大, 信息的可用性越低; 在综合决策系统中, 所有方案在属性 G_j 下差异越大, 则对方案的排序作用越大, 属性的权重应赋越大值, 因此可以通过信息熵确定属性权重。

相较于属性值为实数的决策问题, 同一属性下方案选择的差异为各方案区间数上限 a_{ij}^+ 之间差异和下限 a_{ij}^- 之间差异之和, 因此属性为区间数的权重赋值可以看做分别对区间上限实数和下限实数赋权值, 即分别得到决策矩阵上限属性权重 $w^+ = (w_1^+, w_2^+, \dots, w_m^+)$ 和 下限属性权重 $w^- = (w_1^-, w_2^-, \dots, w_m^-)$ 。定义属性的综合权重 w 为上下限权重的线性加权, 即: $w = \alpha w^+ + \beta w^- (\alpha + \beta = 1)$ 。其中, α 和 β 为对属性上下限权重的偏好^[19]。基于综合权重计算各方案的综合属性值 $Z_i = \sum_j a_{ij} w_j, (i = 1, 2, \dots, n)$ 。各方案的综合属性值 Z_i 两两比较得到可能度矩阵 $P = (p_{ij})_{n \times n}$, 然后使用文献^[11]中给出的可能度矩阵排序公式对方案进行最终的排序。

基于信息熵权法的区间数多属性决策问题的决策过程如下:

(1) 把决策矩阵 $A = (a_{ij})_{m \times n}$ 分割成上下限决策矩阵 $A^+ = (a_{ij}^+)_{m \times n}$ 和 $A^- = (a_{ij}^-)_{m \times n}$, 利用式 (9) 和 (10) 分别对上下限决策矩阵规范化^[20]得到 $R^+ = (r_{ij}^+)_{m \times n}$ 和 $R^- = (r_{ij}^-)_{m \times n}$ 。其中

$$r_{ij}^- = a_{ij}^- / \sqrt{\sum_{i=1}^n (a_{ij}^-)^2}, \quad (9)$$

$$r_{ij}^+ = a_{ij}^+ / \sqrt{\sum_{i=1}^n (a_{ij}^+)^2}. \quad (10)$$

(2) 对上下限规范化决策矩阵 $R^+ = (r_{ij}^+)_{m \times n}$ 和 $R^- = (r_{ij}^-)_{m \times n}$ 列归一化得到 $\dot{R}^+ = (\dot{r}_{ij}^+)_{m \times n}$ 和 $\dot{R}^- = (\dot{r}_{ij}^-)_{m \times n}$ 。其中

$$\dot{r}_{ij}^+ = r_{ij}^+ / \sum_{i=1}^n r_{ij}^+, \quad (11)$$

$$\dot{r}_{ij}^- = r_{ij}^- / \sum_{i=1}^n r_{ij}^-. \quad (12)$$

(3) 对上下限归一化矩阵 $\dot{R}^+ = (\dot{r}_{ij}^+)_{m \times n}$ 和 $\dot{R}^- = (\dot{r}_{ij}^-)_{m \times n}$ 计算各属性 G_j 下所有方案的信息熵

$$E_j^+ = \frac{1}{\ln n} \sum_{i=1}^n \dot{r}_{ij}^+ \ln \dot{r}_{ij}^+, \quad (13)$$

$$E_j^- = \frac{1}{\ln n} \sum_{i=1}^n \dot{r}_{ij}^- \ln \dot{r}_{ij}^-. \quad (14)$$

(4) 通过各属性 G_j 的信息熵值, 分别对上下限归一化矩阵 $\dot{R}^+ = (r_{ij}^+)_{m \times n}$ 和 $\dot{R}^- = (r_{ij}^-)_{m \times n}$ 计算各属性 G_j 的权重, 得到上下限各属性的权重向量 $w^+ = (w_1^+, w_2^+, \dots, w_m^+)$ 和 $w^- = (w_1^-, w_2^-, \dots, w_m^-)$ 。其中

$$w_j^+ = \frac{1 - E_j^+}{\sum_{j=1}^m (1 - E_j^+)}, \quad (15)$$

$$w_j^- = \frac{1 - E_j^-}{\sum_{j=1}^m (1 - E_j^-)}. \quad (16)$$

(5) 通过定义上下限各属性的综合权重 $w = \alpha w^+ + \beta w^- (\alpha + \beta = 1)$, 计算各方案 i 的属性综合值

$$Z_i = \sum_j a_{ij} w_j. \quad (17)$$

(6) 通过两两方案的可能度比较, 组成可能度矩阵 $P = (p_{ij})_{n \times n}$ 。而, 当 a 和 b 至少有一个为区间数 (实数退化为上下限相同的区间数), 设 $a = [a^-, a^+]$, $b = [b^-, b^+]$, $l_a = a^+ - a^-$, $l_b = b^+ - b^-$, 则区间数 a 大于区间数 b 的可能度为

$$p(a \geq b) = \frac{\max\{0, l_a + l_b - \max\{0, b^+ - a^-\}\}}{l_a + l_b}. \quad (18)$$

(7) 由排序向量 $v = \frac{1}{n(n-1)} (\sum_{j=1}^n p_{ij} + \frac{n}{2} - 1)^{[20]}$ 对可能度矩阵 $P = (p_{ij})_{n \times n}$ 进行计算, 得到排序向量 $v = (v_1, v_2, \dots, v_n)^T$ 。通过分量的大小对方案进行排序, 确定最优方案。

3 职住地识别算例

本研究所采用自行开发的 app 采集志愿者的一周的信令数据, 数据除时间戳和基站经纬度外还带有移动或停留标签, 以验证算法的识别效果。

3.1 利用熵权法赋时段权重

用户的时段活跃度, 如图 3 所示。用户夜间 00:00-8:00 时段活跃度最低, 日间下午时段活跃度也较低, 结果显示通过信息熵来度量时段有效性是合理的。

3.2 职住地多属性决策方法

信息熵 MDLP 离散化能根据离去分布特征把连续的时间轴划分为具有特征规律的时间段, 有利于离去规律的提取。通过离散化语义停留点到达

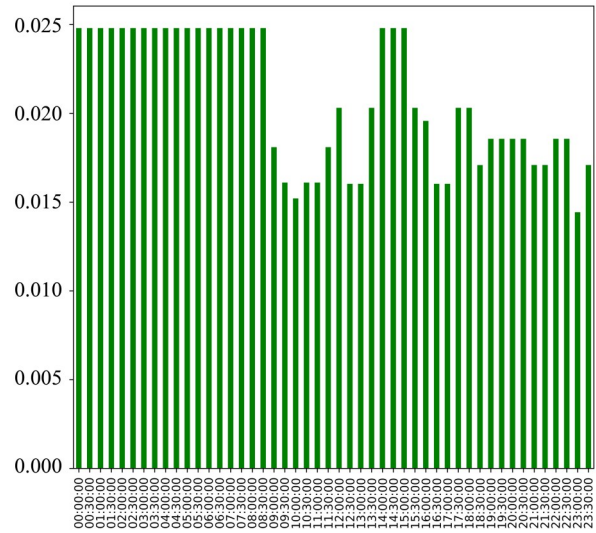


图 3 时段权值示例

Fig. 3 Example of period weight

时间段和离去时段, 组合成停留时段区间数, 并计算停留时段区间数的时段权重区间数。以停留时段权重区间数和停留时长区间数为属性, 经职住地识别问题转变为停留时段和停留时长的多属性区间数决策问题, 得到停留时段和停留时长的区间数决策矩阵, 如表 5 所示。

表 5 职住地决策矩阵 A

语义停留点 Cm	停留时段 W(Cm)	停留时长 ST(Cm)
C2	[1.098 0, 1.098 0]	[0 days 01:06:03, 0 days 01:06:03]
C6	[2.187 7, 2.187 7]	[0 days 01:17:27, 0 days 01:17:27]
C5	[1.098 0, 1.098 0]	[0 days 00:23:10, 0 days 00:23:10]
C3	[2.399 4, 5.365 0]	[2 days 04:08:42, 3 days 19:35:52]
C8	[3.439 1, 5.988 6]	[2 days 15:58:40, 2 days 22:51:46]
C10	[1.090 2, 1.090 2]	[0 days 00:30:20, 0 days 00:30:20]
C7	[0.248 7, 0.248 7]	[0 days 03:38:50, 0 days 03:38:50]

对决策矩阵 A 进行规范化, 停留时段和停留时长均为效益型属性, 得到规范化矩阵 R, 如表 6 所示。决策矩阵 R⁻、R⁺ 列归一化, 得到矩阵 \dot{R}^- 、 \dot{R}^+ , 如表 7、8 所示。

根据归一化的上下限矩阵计算各属性的熵和权重, 如表 9、10 所示。设区间数上、下限的偏好相同, 即 $\alpha = \beta = 0.5$ 可得各属性的综合权重 $w = (0.2065, 0.7935)$ 。利用综合权重计算各方案的综合属性值 Z_i 分别为

$$Z_1 = (0.0184, 0.0267), Z_2 = (0.0325, 0.0474),$$

$$Z_3 = (0.0151, 0.0221), Z_4 = (0.3791, 0.4940),$$

$$Z_5 = (0.4045, 0.4740), Z_6 = (0.0156, 0.0227),$$

$$Z_7 = (0.0201, 0.0280).$$

表 6 职住地规范化矩阵 R Table 6 Normalization matrix R of residence and workplace

语义停留点 C_m	停留时段 $W(C_m)$	停留时长 $ST(C_m)$
C2	[0.2657, 0.3229]	[5.0749, 5.9542]
C6	[0.5294, 0.6434]	[5.9508, 6.9819]
C5	[0.2657, 0.3229]	[1.7800, 2.088]
C3	[0.5806, 1.5779]	[240.3912, 495.4382]
C8	[0.8322, 1.7613]	[294.9410, 383.2858]
C10	[0.2638, 0.3206]	[2.3306, 2.7345]
C7	[0.0602, 0.0731]	[16.8139, 19.7273]

表 7 职住地归一化矩阵 \hat{R}^- Table 7 Residence and workplace normalization matrix \hat{R}^-

语义停留点 C_m	停留时段 $W(C_m)$	停留时长 $ST(C_m)$
C2	0.0950	0.0089
C6	0.1892	0.0105
C5	0.0950	0.0031
C3	0.2075	0.4238
C8	0.2975	0.5199
C10	0.0943	0.0041
C7	0.0215	0.0296

表 8 职住地归一化矩阵 \hat{R}^+ Table 8 Residence and workplace normalization matrix \hat{R}^+

语义停留点 C_m	停留时段 $W(C_m)$	停留时长 $ST(C_m)$
C2	0.0643	0.0065
C6	0.1281	0.0076
C5	0.0643	0.0023
C3	0.3142	0.5407
C8	0.3507	0.4183
C10	0.0638	0.0030
C7	0.0146	0.0215

表 9 矩阵 R^- 各属性的信息熵和权重Table 9 Entropy and weight of matrix R^- attributes

信息熵和属性权重	停留时段	停留时长
E_j	0.9016	0.4825
w_j	0.1597	0.8403

表 10 矩阵 R^+ 各属性的信息熵和权重Table 10 Entropy and weight of matrix R^+ attributes

信息熵和属性权重	停留时段	停留时长
E_j	0.8143	0.4526
w_j	0.2533	0.7467

$Z_i (i = 0, 1, 2, 3, 4, 5, 6)$ 两两比较的可能度矩阵 $P = (p_{ij})_{7 \times 7}$, 如表 11 所示。

表 11 各语义综合属性值两两比较的可能度矩阵 P Table 11 The possibility matrix P for the comparison of the comprehensive attribute values of each scheme

语义 C_m	C2	C6	C5	C3	C8	C10	C7
C2	0.5	0.0	0.760	0.0	0.0	0.722	0.410
C6	1.0	0.5	1.0	0.0	0.0	1.0	1.0
C5	0.240	0.0	0.5	0.0	0.0	0.461	0.135
C3	1.0	1.0	1.0	0.5	0.486	1.0	1.0
C8	1.0	1.0	1.0	0.514	0.5	1.0	1.0
C10	0.278	0.0	0.539	0.0	0.0	0.5	0.175
C7	0.590	0.0	0.865	0.0	0.0	0.825	0.5

由排序向量 v 及矩阵 P 中的可能度, 得到可能度矩阵 P 的排序向量, 通过分量的大小对方案进行排序, 确定最优方案。

由于居住时段比工作时段的活动更少, 稳定性更强, 因此在排序的前两个语义地点中选择停留时段属性值较大的作为居住地, 停留时段属性值小的作为工作地。第 i_0 语义与第 i_1 语义停留时段属性值大小以决策矩阵停留时段区间可能度来度量。若 $p(a_{i_0j} \geq a_{i_1j}) > 0.5$, 则 i_0 为居住地, i_1 为工作地, 否则 i_1 为居住地, i_0 为工作地。在例子中, C3 和 C8 的决策矩阵停留时段区间分别为 $a_{41} = [2.3994, 5.3650]$ 、 $a_{51} = [3.4391, 5.9886]$, 可能度 $p(a_{41} \geq a_{51}) = 0.3492$, 小于 0.5, 因此 C8 为居住地, C3 为居住地, 识别结果与轨迹标签吻合。

3.3 多属性决策算法与时间阈值法的比较

为了比较多属性决策算法与传统的职住地算法的优劣, 对采集的标签数据采用两种算法进行职住地识别。由于时间阈值法是传统职住地算法的基础算法, 其余算法大多是其的延伸, 因此选取时间阈值法作为常规算法。

对数据量、划分的时间段、停留时间阈值和天数阈值, 采用全样本量一周的数据, 根据经验设定不同的划分时间段 slot、停留时间阈值 T 和天数阈值 θ 等参数, 比较不同参数下时间阈值法的识别结果, 如表 12、13 所示。统一设置居住地天数

表12 各参数组合的时间阈值法居住地识别结果

Table 12 Time threshold method of residence identification results of each parameter combination

居住地参数	满足参数阈值的停留点居住天数	居住地
slot = [day20:00 - day ₊₁ 08:00], T = 6 h	{C8: 5, C9: 1}	C8
slot = [day22:00 - day ₊₁ 06:00], T = 5 h	{C8: 5, C9: 1}	C8
slot = [day22:00 - day ₊₁ 08:00], T = 5 h	{C8: 5, C9: 1}	C8
slot = [day22:00 - day ₊₁ 10:00], T = 6 h	{C8: 5, C9: 1}	C8
slot = [day00:00 - day06:00], T = 4 h	{C8: 6}	C8
slot = [day00:00 - day08:00], T = 4 h	{C8: 6}	C8

表13 各参数组合的时间阈值法工作地识别结果

Table 13 Time threshold method of residence identification results of each parameter combination

工作地参数 (T = 6h)	满足参数阈值的停留点工作天数	工作地
slot = [day 07: 00 - day 17: 00]	{C3: 2}	无法识别
slot = [day 08: 00 - day 18: 00]	{C3: 4}	C3
slot = [day 09: 00 - day 17: 00]	{C3: 2}	无法识别
slot = [day 07: 30 - day 17: 30]	{C3: 2}	无法识别
slot = [day 08: 30 - day 17: 30]	{C3: 2}	无法识别
slot = [day 09: 30 - day 17: 30]	{C3: 2}	无法识别

阈值为4, 工作地天数阈值为3, 表示在划分时间段 slot 内, 停留时间阈值大于 T 的停留点出现天数大于4天的停留点为居住地, 类似地得到工作地。

由表中可以看出, 常规职住地识别算法受划分的时间段 slot、停留时间阈值 T 和天数阈值 θ 等参数直接影响较大, 由于时间段的经验划分无法挖掘用户的作息规律, 对于日间活跃时段无法有效识别用户工作地, 因此常规职住地算法具有很

强的局限性。

本研究提出的多属性决策算法不依赖于时段的划分, 以及一系列阈值参数的选取, 可以根据用户日常出行作息规律习惯自动挖掘用户职住的规律, 从而识别职住地。多属性决策算法只与用户数据样本量有关。因此对不同样本量分别为3天、4天、5天、6天和一周全样本量的识别结果进行对比, 如表14所示。

表14 各语义综合属性值两两比较的可能度矩阵 P

Table 14 Possibility matrix for pairwise comparison of each semantic comprehensive attribute value

样本日期	停留点序列	可能度序列	居住地	工作地
2020. 01. 10-2020. 01. 12	[C3,C8,C7,C6,C5]	[0. 3,0. 25,0. 2,0. 15,0. 1]	C3	C8
2020. 01. 11-2020. 01. 13	[C8,C3,C6]	[0. 58,0. 42,0. 25]	C8	C3
2020. 01. 12-2020. 01. 14	[C8,C3,C10,C6]	[0. 375,0. 292,0. 168,0. 165]	C8	C3
2020. 01. 13-2020. 01. 15	[C8,C3,C10,C6]	[0. 375,0. 292,0. 168,0. 165]	C8	C3
2020. 01. 14-2020. 01. 16	[C3,C8,C2,C10,C6]	[0. 28,0. 27,0. 177,0. 138,0. 135]	C8	C3
2020. 01. 10-2020. 01. 13	[C3,C8,C6,C5,C7]	[0. 276,0. 274,0. 2,0. 139,0. 111]	C8	C3
2020. 01. 11-2020. 01. 14	[C8,C3,C6,C10]	[0. 349,0. 318,0. 208,0. 125]	C8	C3
2020. 01. 12-2020. 01. 15	[C8,C3,C10,C6]	[0. 375,0. 292,0. 168,0. 165]	C8	C3
2020. 01. 13-2020. 01. 16	[C8,C3,C2,C10,C6]	[0. 287,0. 263,0. 167,0. 142,0. 141]	C8	C3
2020. 01. 10-2020. 01. 14	[C8,C3,C6,C10,C5,C7]	[0. 242,0. 225,0. 183,0. 125,0. 122,0. 103]	C8	C3
2020. 01. 11-2020. 01. 15	[C8,C3,C6,C10]	[0. 35,0. 317,0. 208,0. 125]	C8	C3
2020. 01. 12-2020. 01. 16	[C8,C3,C2,C10,C6]	[0. 289,0. 261,0. 163,0. 144,0. 143]	C8	C3
2020. 01. 10-2020. 01. 15	[C8,C3,C6,C7,C10,C5]	[0. 235,0. 232,0. 183,0. 126,0. 114,0. 11]	C8	C3
2020. 01. 11-2020. 01. 16	[C3,C8,C6,C2,C10]	[0. 277,0. 273,0. 2,0. 149,0. 101]	C8	C3
2020. 01. 10-2020. 01. 16	[C8,C3,C6,C7,C2,C10,C5]	[0. 203,0. 202,0. 167,0. 126,0. 117,0. 095,0. 09]	C8	C3

根据表14可知, 除了2020年1月10日至2020年1月12日的样本由于规律性差别较大, 导致职住地识别反向外, 其余15个样本的职住地均识别正确, 说明职住地多属性决策算法能够较为全面地考虑停留时间段和停留时长等职住属性, 自动挖掘

职住的规律特征, 同时具有较强的稳定性, 且在样本量较小的情况下仍然有较强的职住识别能力。

4 结论

本研究总结了目前职住地算法中存在的时间

规则中时段划分和时间阈值设置依据不足、空间聚合距离大小的设置方法有效性未知、缺乏对多种职住属性的综合考虑以及只能定性识别职住地理位置而无法挖掘用户作息规律等突出问题。并基于轨迹点识别,提出了轨迹簇空间联合流程,通过实验证明该流程能够提升停留区域识别的时空精度,锚固停留区域时段和时长的准确度。在准确划分轨迹停留区域时空界限的前提下,通过时段停留次数和停留点分布特性计算时段熵,度

量时段属于职住时段的可能性,对时段赋权值。通过对任意语义停留点到达和离去时段的离散化,以停留时段和停留时长为特征构建职住地综合决策矩阵,识别出最可能的职住地语义。通过与传统的职住地识别算法对比,可知多属性决策方法的识别准确率优于传统职住地算法,在样本较小的情况下仍具有较强的稳定性,可以直接应用于现实数据。

参考文献:

- [1] 张天然. 基于手机信令数据的上海市域职住空间分析[J]. 城市交通, 2016(1):15-23.
- [2] 丁亮,钮心毅,宋小冬. 上海中心城就业中心体系测度—基于手机信令数据的研究[J]. 地理学报, 2016, 71(3):484-499.
- [3] 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职住地[J]. 武汉大学学报(信息科学版), 2014, 39(6):750-756.
- [4] 吴子啸. 基于手机数据的出行链推演算法[J]. 城市交通, 2019(3):11-18.
- [5] GONZÁLEZ M C, HIDALGO C A, BARABÁSI A L. Understanding individual human mobility patterns [J]. Nature, 2008,453:779-782.
- [6] CALABRESE F, COLONNA M, LOVISOLO P, et al. Real-Time urban monitoring using cell phones: a case study in rome [J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(1):141-151.
- [7] 宋少飞,李玮峰,杨东援. 基于移动通信数据的居民居住地识别方法研究[J]. 综合运输, 2015, 37(12):72-76.
- [8] LEE J K, HOU J C. Modeling steady-state and transient behaviors of user mobility: formulation, analysis, and application [C]//Conti M, et al. Proceedings of the 7th ACM International Symposium on Mobile Ad Hoc Networking & Computing. Florence: ACM, 2006:85-96.
- [9] 钮心毅,谢琛. 手机信令数据识别职住地的时空因素及其影响[J]. 城市交通, 2019, 17(3):19-29.
- [10] 简志春. 基于多源时空数据的深圳市居民职住分布和通勤特征研究[D]. 武汉:武汉大学, 2018.
- [11] 许宁. 基于手机定位数据的居民职住地分布特征研究[D]. 长沙:中南大学, 2014.
- [12] 许洪波,许金辉. 基于手机信令的空间位置数据计算职住地[J]. 北京测绘, 2016(6):69-71+90.
- [13] 唐小勇,周涛,陆百川,等. 一种基于手机信令的通勤OD训练方法[J]. 交通运输系统工程与信息, 2016, 16(5):64-70.
- [14] 王艳涛,魏海平,何源浩,等. 基于位置轨迹挖掘的城市居民职住地识别方法研究[J]. 测绘与空间地理信息, 2017, 40(2):113-116.
- [15] 陈略,熊宸,蔡铭. 一种基于手机信令的时空密度轨迹点识别算法[J]. 计算机工程, 2021, 47(3):83-93.
- [16] FAYYAD U M. Multi-Interval discretization of continuous-valued attributes for classification learning [C]// Proc of the Int'l Joint Conf on Artificial Intelligence. 1993: 1022-1027.
- [17] 吴瑕,唐祖锴,祝园园,等. 近似到达时间约束下的语义轨迹频繁模式挖掘[J]. 软件学报, 2018, 29(10):3184-3204.
- [18] 张金波,赵攀. 基于信息熵的区间数多属性决策方法[J]. 赤峰学院学报(自然科学版), 2019, 35(3):4-6.
- [19] 樊治平,宫贤斌,张全. 区间数多属性决策中决策矩阵的规范化方法[J]. 东北大学学报, 1999(3):104-107.
- [20] 徐泽水. 模糊互补判断矩阵排序的一种算法[J]. 系统工程学报, 2001(4):311-314.

(责任编辑 王海蓉)