工程与管理科学
Engineering and Management Science
第 8 卷◆第 9 期◆版本 1.0◆2026 年
文章类型：论文|刊号（ISSN）：2705-0637(P) / 2705-0645(O)

基于手机 GPS 定位数据的个体出行目的识别方法研究

王亮[1] 杨茜[1] 陈升升[2] 姚振兴[2*]
1.中国电建集团西北勘测设计研究院有限公司 陕西西安 710065；2. 长安大学运输工程学院 陕西西安 710061

DOI：10.32629/ems.v8i9.22154

[摘 要] 出行目的识别是理解居民出行行为规律、支撑交通规划与管理决策的关键环节。传统基于问卷调查的出行目的获取方法成本高、时效性差，难以满足精细化交通治理需求。本文利用手机 GPS 定位数据高精度、连续采样的优势，提出一种融合多源特征的多层证据链出行目的识别方法。首先，采用改进的时空密度聚类算法（ST-DBSCAN）识别个体出行停留点，提取出行端点及停留时长；其次，从时空特征、用地属性、POI 语义、历史规律四个维度构建证据体系，分别量化各证据因子与出行目的的条件依赖关系；最后，基于贝叶斯网络实现多源证据的概率融合，输出最大后验概率对应的出行目的。实验结果表明，本文方法在典型出行目的识别中准确率达到 87.5%，为城市交通需求分析与治理决策提供了有效支撑。

[关键词] 手机 GPS；出行目的；ST-DBSCAN；多源证据；贝叶斯网络

1 引言

出行目的是交通需求预测和居民活动分析的重要变量。问卷与电话调查能够获取活动语义，但存在样本有限、回忆偏差和更新周期长等问题。智能手机 GPS 可连续记录时间、位置和速度，在开阔环境下具有较高定位精度，为自动提取出行端点和活动特征提供了条件；但城市峡谷、室内遮挡和设备差异会造成漂移或缺测，且轨迹本身不含出行目的，仍需融合外部语义信息进行推断。

既有研究经历了规则、统计学习和深度学习的发展。Wolf 等利用 GPS 与 GIS 信息推断目的[1]，Shen 等引入活动持续时间和出行链信息[2]；Liu 等利用机器学习赋予位置记录活动语义[3]。随后，人工神经网络、随机森林提高了分类能力，注意力模型开始利用出行链上下文[4-5]。但复杂模型依赖较多标注样本，且难以解释时空、用地和 POI 证据冲突时的判别过程。

本文构建“轨迹清洗-端点识别-证据提取-概率融合”流程：以数据驱动方式标定 ST-DBSCAN 参数，将时空行为、用地、POI 和历史规律纳入统一概率框架，并通过参与者分组验证和证据消融检验模型。研究重点是提高小样本条件下的可解释性，而非以更复杂模型替代数据质量控制。

2 出行目的识别方法

2.1 基于时空聚类的出行端点识别

传统 DBSCAN 算法仅考虑空间邻近性，无法区分不同时段在同一空间区域的重复停留。本文采用 ST-DBSCAN（Spatial-Temporal DBSCAN）算法，在空间邻域基础上引入时间邻域约束。相关定义如下：

E 邻域：给定对象 O，其 E 邻域为满足空间距离 $\leqslant Eps$ 且时间差 $\leqslant \Delta T$ 的所有对象的集合；核心对象：若对象 O 的 E 邻域内包含的对象数量 $\geqslant MinPts$，则 O 为核心对象；密度可达：若存在一个核心对象链 $p_1, p_2, \cdots, p_n$，使得 p_{n+1} 在 p_i 的 E 邻域内，则称 p_n 从 p_1 密度可达。

对清洗轨迹执行 ST-DBSCAN，每个稳定聚类簇作为候选停留点，中心位置取簇内坐标的时间加权中心，并记录到达、离开和持续时间。相邻两个停留点之间的轨迹段构成一次出行，其目的端点进入后续语义识别。短时交通控制停留由持续时间和前后运动状态共同排除。

2.2 多维证据特征提取

单一空间信息难以区分混合功能区内的活动，本文构建以下四类证据：

1）时空行为证据：到达时段、停留时长、工作日属性和到访频率。工作日白天长时且重复停留通常更支持上班目的；2）用地属性证据：提取居住、商业、办公、教育、医疗和娱乐等功能；混合用地采用面积占比形成概率向量，不作唯一硬匹配；3）POI 语义证据：统计停留点 100、200 和 500 m 缓冲区内 8 类 POI，采用 TF-IDF 降低高频通用类别的权重，并保留多尺度结果。4）历史规律证据：依据至少 7 d 轨迹统

Copyright © This work is licensed under a Commons Attibution-Non Commercial 4.0 International License.

计相同地点、日类型和时段的重复到访。该证据只描述访问规律，不使用测试样本的真实目的标签。

2.3 基于贝叶斯网络的出行目的识别

模型以出行目的 C 为目标变量，以 E_{time}，E_{dur}，E_{land}，E_{poi}，E_{hist} 为观测变量。四类证据与五个节点并不矛盾：时空行为被拆分为到达时段和停留时长两个变量，其余三类各对应一个节点。给定目的后的条件独立是假设性近似，建模前通过合并重复 POI 类别和归一化降低证据冗余。

目的集合 $C = c_1, c_2, \cdots, c_8$，依次表示上班、上学、回家、餐饮、购物、娱乐、接送和就医；证据向量 $E=\{e_{time}, e_{dur}, e_{land}, e_{poi}, e_{hist}\}$。其中“上班”包含与工作直接相关的公务活动。给定证据 E 时，目的 c_i 的后验概率为：

$$P(c_i|E) = \frac{P(E|c_i)P(c_i)}{P(E)} = \frac{P(c_i)\prod_{j=1}^{n} P(e_j|c_i)}{P(E)} \quad (3\text{-}1)$$

其中，$P(c_i)$为出行目的先验概率，可根据城市居民出行调查统计结果或历史信令数据统计获得；$P(E|c_i)$为条件概率，表示在出行目的为c_i的条件下证据 e_j出现的概率，通过训练样本的频率统计或参数估计得到；$P(E)$为归一化因子。

计算八类目的的后验概率，并选择最大值对应类别：

$$\hat{c} = \arg\max_{c_i \in C} P(c_i|E) \quad (3\text{-}2)$$

置信度阈值在训练折内标定，本文取 θ =0.65。低于阈值的样本标记为“不确定”，但评价总体准确率时仍计入样本总数，以免通过剔除难样本虚高性能。采用按参与者分组的五折交叉验证，每折约 48 名参与者用于训练、12 名用于测试，同一人的轨迹不跨训练集和测试集。训练阶段估计目的先验、条件概率和核密度参数；测试阶段独立提取证据、计算后验概率并与日志标签比较。该划分可降低同一人固定地点和规律性特征造成的信息泄漏。

3 数据采集与预处理

3.1 研究区域与数据采集

研究区域为西安市主城区。招募 60 名志愿者采集 2023 年 5-11 月出行数据，职业包括职员、教师、学生、个体经营者和医疗人员等，最终获得 1112 次有效出行，涵盖上班、上学、回家、餐饮、购物、娱乐、就医和接送 8 类目的。

数据由自主研发的“易行”手机应用采集，目标采样频率为 1 Hz，字段包括匿名编号、时间戳、经纬度、速度、卫星数和水平精度因子。志愿者同步填写电子日志，记录起止时间、目的地和活动类型，作为端点及目的标签真值。研究仅使用匿名化后的必要字段。

3.2 数据预处理

原始 GPS 数据依次进行格式、定位质量和空间范围检查：删除重复、缺失及研究区外记录；结合卫星数、水平精度、相邻点速度与加速度识别异常跳点，避免仅以固定速度阈值误删真实快速出行。对不超过 60 s 的短时缺测采用线性插值，长时间缺测则切断轨迹，不在地下或室内区段强行补点。清洗后按匿名编号和时间重构轨迹，并计算有效采样间隔。每条轨迹平均约含 300 个有效点；后续聚类以实际有效采样率标定最小点数，减少设备休眠和缺测对停留判定的影响。

4 实验与结果分析

4.1 出行端点识别结果

在 1112 个日志标注的目的端点上，ST-DBSCAN 正确检出 1043 个，检出率为 93.8%。上班、回家和餐饮端点检出率超过 95%；接送活动因停留通常仅 10-20 min，漏检较多。该指标反映对真实端点的检出能力，不等同于同时考虑误检的分类准确率；端点时间偏差总体控制在 3 min 以内。

4.2 多维证据有效性分析结果

为验证各维度证据对出行目的识别的贡献，设计了证据逐项增加的控制实验。实验采用相同的贝叶斯网络结构和训练/测试划分，依次加入不同证据组合，结果如表 4-1 所示。

表 4-1　不同证据组合的识别准确率

证据组合	平均准确率
仅 POI 语义	68.3%
POI 语义 + 时空特征	78.6%
POI 语义 + 时空特征 + 用地属性	83.5%
全证据（四类）	87.5%

单证据（仅 POI）准确率仅为 68.3%，说明仅依靠周边 POI 类型易受功能混合和偶然访问干扰。加入时空特征后提升 10.3 个百分点，再加入用地属性后提升 4.9 个百分点，最后加入历史规律后提升 4.0 个百分点。四类证据融合相较单证据提升了 19.2 个百分点，证明多源证据融合的必要性和有效性。其中，时空特征和历史规律对规律性强的出行目的（上

Copyright © This work is licensed under a Commons Attibution-Non Commercial 4.0 International License.

班、回家）贡献尤为显著，用地属性和 POI 语义对非通勤目的（购物、就医）贡献更大。

4.3 出行目的识别总体结果

基于全证据的贝叶斯网络模型对 8 类出行目的的整体识别准确率为 87.5%（五折交叉验证的平均值）。各类别的详细结果如表 4-2 所示。

表 4-2　出行目的识别准确率

出行目的	测试样本数	正确识别数	准确率
上班	68	62	91.2%
上学	10	9	90.0%
回家	70	65	92.9%
餐饮	36	32	88.9%
购物	14	12	85.7%
娱乐	7	6	85.7%
接送学	15	12	80.0%
就医	3	3	100%
总计	223	201	90.1%

注：表中为一折结果（223 个样本，准确率 90.1%），不代表五折总体水平；就医等小样本类别不宜单独外推。

从表 4-3 中可以看出，回家和上班识别率分别为 92.9%和 91.2%，得益于稳定的时段和重复到访规律。餐饮、购物分别为88.9%和85.7%,POI 共址及时长重叠是主要混淆来源；接送为 80.0%，易受短暂停留影响。就医仅 3 个样本，100%结果不足以证明该类别具有稳定泛化能力。

5 结论

本文提出融合 GPS 时空行为、用地、POI 和历史规律的个体出行目的识别方法，主要结论如下：

（1）通过训练折内的数据驱动标定确定 ST-DBSCAN 参数，目的端点检出率为 93.8%，为后续语义识别提供基础。

（2）四类证据在贝叶斯框架中形成互补，完整模型较仅采用 POI 提高 19.2 个百分点,能够缓解功能混合造成的目的混淆。

（3）按参与者分组的五折验证得到平均准确率 87.5%；回家和上班识别较稳定，接送及小样本类别仍受停留时间和样本规模限制。

该方法适用于 GPS 辅助出行调查和群体交通需求分析。受 60 名志愿者、单城市和类别不均衡限制，结果不宜直接外推。后续应扩大跨区域样本，补充宏平均精确率、召回率和 F1 值，并比较不同阈值下的精度与覆盖率。

6 致谢

本研究受中国电建集团西北勘测设计研究院有限公司项目（20250128KJWWCC）资助。

[参考文献]

[1]Wolf J, Guensler R, Bachman W. Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data[J]. Transportation Research Record, 2001, 1768(1): 125-134.

[2]Shen L, Stopher P R. A process for trip purpose imputation from Global Positioning System data[J]. Transportation Research Part C: Emerging Technologies, 2013, 36: 261-267.

[3]Liu F, Janssens D, Wets G, et al. Annotating mobile phone location data with activity purposes using machine learning algorithms[J]. Expert Systems with Applications, 2013, 40(8): 3299-3311.

[4]Xiao G, Juan Z, Zhang C. Detecting trip purposes from smartphone-based travel surveys with artificial neural networks and particle swarm optimization[J]. Transportation Research Part C: Emerging Technologies, 2016, 71: 447-463.

[5]Lyu S, Han T, Li P, et al. A dual-flow attentive network with feature crossing for chained trip purpose inference[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(1): 631-644.

*通讯作者：姚振兴（1989—），男，浙江金华人，副教授，博士生导师，工学博士。E-mail: yaotraffic@chd.edu.cn。

Copyright © This work is licensed under a Commons Attibution-Non Commercial 4.0 International License.