

Research on the Recognition Method of Travel Mode Based on Mobile Phone Signaling Data

Shejun Deng^{1,2}, Yucheng Tang², Dongmei Liu³, Shijun Yu²

¹Opening Project of Key Laboratory of Intelligent Transportation Systems Technologies, Ministry of Communications, Beijing

²College of Civil Science and Engineering, Yangzhou University, Yangzhou Jiangsu

³ITS Center, Research Institute of Highway Ministry of Transport, Beijing
Email: yzrx6@163.com

Received: Dec. 25th, 2019; accepted: Jan. 8th, 2020; published: Jan. 15th, 2020

Abstract

In order to study the application of mobile signaling data in the field of identifying user's travel mode, and improve the accuracy of travel mode identification, firstly, we introduce the principle of signaling data generation, the selection and principle of clustering algorithm. Then, according to the difference between the aboveground and underground base stations, the identification of travel mode can be divided into two directions: the identification of ground travel mode, and the identification of subway. Then, the signaling data is combined with the navigation track data and the base station database data to establish the model for the ground traffic and the subway travel, so as to determine the user's travel mode. Finally, the model is validated based on the signaling data of volunteers provided by operators. The result shows that the accuracy of the algorithm is 86.12%, which effectively improves the accuracy of the algorithm.

Keywords

Mobile Signaling Data, Navigation Data, Recognition of Travel Method

基于手机信令数据的出行方式识别方法研究

邓社军^{1,2}, 唐玉成², 刘冬梅³, 于世军²

¹智能交通技术交通运输行业重点实验室, 北京

²扬州大学建筑科学与工程学院, 江苏 扬州

³交通运输部公路科学研究院ITS中心, 北京

Email: yzrx6@163.com

摘要

为了研究手机信令数据在识别用户出行方式领域的应用，进一步提高出行方式识别的准确度。首先，介绍了手机信令数据产生的原理、聚类算法选取的原则。然后，根据地上地下基站的不同将出行方式的识别分为两个方向：地面出行方式、地铁出行的识别。接着，针对地面交通、地铁出行将信令数据分别与导航轨迹数据、地铁基站库数据进行结合建立了模型，以此判定用户的出行方式。最后，基于某运行商提供的志愿者的信令数据对模型进行验证。研究结果表明，该识别算法正确率达86.12%，有效的提升了出行方式识别算法的精度。

关键词

手机信令数据，导航数据，出行方式识别

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

基于手机信令数据的居民出行信息采集技术作为一种新兴的调查技术已广泛应用于交通调查中。相较于传统的调查方法，其具有成本低、周期短、覆盖面广等优点，能更加全面高效、持续的获取居民的出行方式，可以为城市规划、交通组织管理提供决策依据[1] [2]。

目前国内外关于手机信令的研究主要集中在出行轨迹、交通流等方面[3] [4] [5] [6]，而针对出行方式识别的研究成果尚不多见。国际上，相关学者最先通过对平均速度、最大速度、出行时长等设定不同的阈值实现出行方式的识别[7] [8] [9]；Anderson 等[10]利用速度、信号强度波动与服务小区数量之间的关系，采用隐马尔可夫模型 HMM (Hidden Markov Model)对交通方式进行识别；Wang 等[11]使用 K-means 聚类方法对信令数据的轨迹进行分组，然后通过时间匹配从而判定出行方式；Yu Ning [12]等利用手机信令数据与 GPS 数据、基站数据、加速度数据等构建了交通方式的识别系统；Hong 等[13]，Reddy 等[14]将地理信息系统与数据结合，利用平均速度、最大速度等特征将数据分为许多不同速度区间，依据速度大小，结合地理信息，利用决策树识别交通方式。

国内对于手机信令数据的研究起步较晚，研究主要集中在 2008 年之后。张博[15]考虑出行距离、时间、平均速度三个指标结合出行轨迹提出了出行方式模糊识别模型；但模型选取的指标受天气、实际路况等因素的影响较大。部分学者利用 GPS 定位数据采用神经网络、支持向量机、多层感知器神经网络模型进行出行方式的识别，达到了良好的精度[16] [17] [18]。赖见辉[19]使用手机信令数据和轨道交通线路信息识别地铁出行轨迹。

纵观国内外文献，以往研究主要通过 GPS 定位数据进行出行方式的识别，而手机信令数据的研究则主要集中在 OD 的获取等方面。为进一步提高居民出行方式判定的准确率，本文主要利用手机信令数据同时结合地图导航数据、地铁基站数据库，采用 DBSCAN 算法，构建出行方式的识别模型，最后通过实际案例对模型的有效性进行了验证。

2. 基本原理

手机信令数据产生原理

手机处于待机状态时通过基站与无线网络进行信息交互，与此同时无线网络记录手机所处的位置区信息。

由于基站分布在道路附近，因此信令数据记录的坐标散落在真实出行轨迹附近，将信令数据与道路的轨迹点进行聚类分析，大量的信令数据将与实际出行路径的轨迹点聚类可以获得很多簇。基于此原理，同时考虑到出行方式与出行路径相互联系又相互制约，将出行方式与路径通过导航地图进行关联。使用百度地图应用程序接口请求同一个 OD 下不同交通方式的各种导航轨迹数据，然后将信令数据中的位置数据与导航轨迹数据进行聚类分析，通过聚类簇中的信令数据的个数计算各交通方式下导航地图规划路径的匹配度。结合出行时间相似度判定地面出行过程中最有可能的出行方式。

3. 识别算法

3.1. 识别算法分类

基站包括位于地面的普通基站和位于地铁内的微蜂窝基站。为满足各自通信质量的需求，地面移动通信网络与地铁移动通信网络存在以下不同点：首先是基站布设形式方面，地铁基站按照地铁线路走线呈点式分布，地面基站的布设则较为复杂；然后是基站覆盖范围方面，地铁站点间重复覆盖度较小，相邻站点间的基站不会出现信号交叉干扰，而地面基站信号干扰的现象则较为普遍。地铁移动通信网络与地面移动通信网络在基站布设、信号覆盖、服务需求等方面的差异性使得用户进入和离开地铁站点时都会发生一次地面基站与地铁基站的切换，并记录在信令数据中。这一特征可以为判定用户是否采用地铁出行方式，提供了技术条件。结合地面基站与地铁基站的特点，本文将出行方式的识别方法分为地面与地铁两个方面进行研究。如图 1 所示。

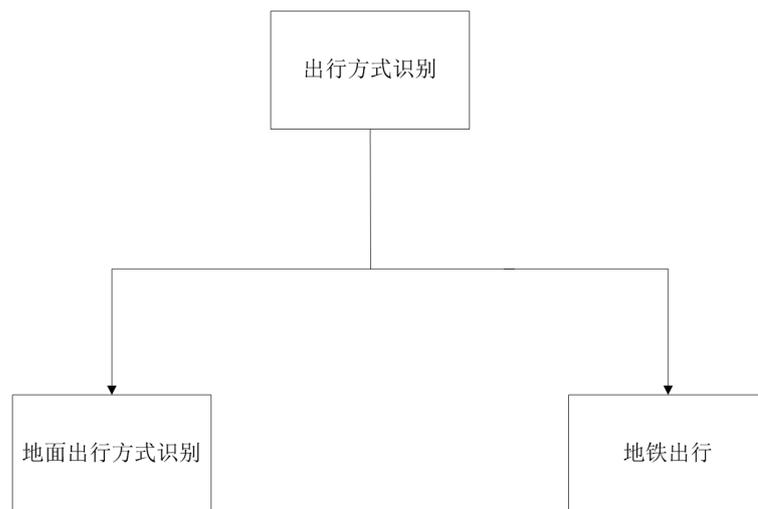


Figure 1. Division of mode
图 1. 方式划分

3.1.1. 地面出行方式识别

地面交通出行方式识别算法流程如图 2 所示。具体步骤如下：

- 1) 获取用户一次有效出行段的信令数据集 A，包含此次出行轨迹点的经纬度坐标、时间戳等信息。

2) 由于出发时间对出行方式以及出行路径的选择影响较大。本文通过对出发时间进行情景划分, 在相似情景下利用导航地图的应用程序接口进行各交通方式的路径规划, 减少这一因素对算法识别准确率的影响。情景划分见表 1。

Table 1. Division of cases

表 1. 情景划分

	工作日	周末	节假日
高峰时段	情景 1	情景 2	情景 3
平峰时段	情景 4	情景 5	情景 6

3) 使用百度地图应用程序接口请求从起点到终点各种交通方式(骑行、步行、小汽车(不区分私家车和出租车)、公交车)在相似情景下的导航数据。定义获得的导航数据集合为 $H = \{h_1, h_2, \dots, h_i, \dots, h_n\}$, h_i 表示第 i 个导航线路; 其中导航线路数据包含从起点到终点所需要经过的道路沿线的经纬度坐标点 $L = \{l_1, l_2, l_3, \dots, l_i, \dots, l_n\}$, l_i 表示第 i 个导航路线 h_i 对应的经纬度坐标 $l_i = (l_x, l_y)$, 距离长度 $S = \{s_1, s_2, s_3, \dots, s_i, \dots, s_n\}$, S 表示距离集合, s_i 表示第 i 个导航路线 h_i 对应的才能够起点到终点的距离长度、交通方式 $M = \{m_1, m_2, m_3, \dots, m_i, \dots, m_n\}$, M 表示交通方式的集合, m_i 表示 h_i 对应的交通方式, 以及所用时长 $T = \{t_1, t_2, t_3, \dots, t_i, \dots, t_n\}$, T 表示所用时长的集合, t_i 表示 h_i 对应的所用时长。

4) 设一次有效出行段的信令数据的坐标集合为 $A = \{a_1, a_2, a_3, \dots, a_i, \dots, a_n\}$, A 表示信令数据的坐标集合, a_i 表示出行段中第 i 个基站的经纬度坐标。将信令数据的经纬度坐标 a_i 与导航路线的经纬度坐标 l_i 进行聚类分析。设定好参数, 聚类得到 k 个坐标点的簇 $B = \{b_1, b_2, b_3, \dots, b_i, \dots, b_k\}$, B 为簇的集合, b_i 为第 i 个簇。

5) 分析并获取 B 中含有导航轨迹坐标 l_i 的簇的集合 BL , 统计集合 BL 各簇中含有基站坐标 a_i 的个数记为 N_{li} ; 信令数据坐标集合 A 中坐标点的个数为 N 。将不同出行方式下的各种导航轨迹与信令数据进行聚类分析, 获得各出行方式下导航轨迹数据与信令轨迹数据的轨迹匹配度 $R = \{r_{l1}, r_{l2}, \dots, r_{li}, \dots, r_{ln}\}$; 轨迹匹配度计算公式如下:

$$r_{li} = \frac{N_{li}}{N} \tag{1}$$

r_{li} : 轨迹匹配度;

N_{li} : 簇中信令数据位置数据的个数;

N : 信令数据中位置数据的个数。

6) 对于某些路段可采用多种交通方式, 所以存在交通方式不同但导航路线相同的情况。对出行实际用时和导航所用时长进行相关性分析。若两者相关性越大则该出行方式下的该路径是真实情况的概率就越大。时间相似度的计算公式如下:

$$t_{li} = \begin{cases} 1 - \frac{|t_i - T_0|}{t_i} & t_i > \frac{T_0}{2} \\ 0 & t_i \leq \frac{T_0}{2} \end{cases} \tag{2}$$

t_{li} : 时间相似度;

t_i : h_i 对应的所用时长(s);

T_0 : 用户出行的实际用时(s)。

7) 综合考虑轨迹匹配度与时间相似度, 设定相应的权重, 最后计算得出所有出行方式下的出行路径的概率集合 $P = \{P_1, P_2, \dots\}$ 。查询 P 中的最大值 P_{max} , 判定 P_{max} 对应的出行方式 m_i 为实际出行方式。计算公式如下:

$$P_i = K \times r_{ii} + (1 - K) \times t_{ii} \quad (3)$$

P_i : 一种交通方式 m_i 下一条出行路径 h_i 的概率;

r_{ii} : 轨迹匹配度;

t_{ii} : 时间相似度。

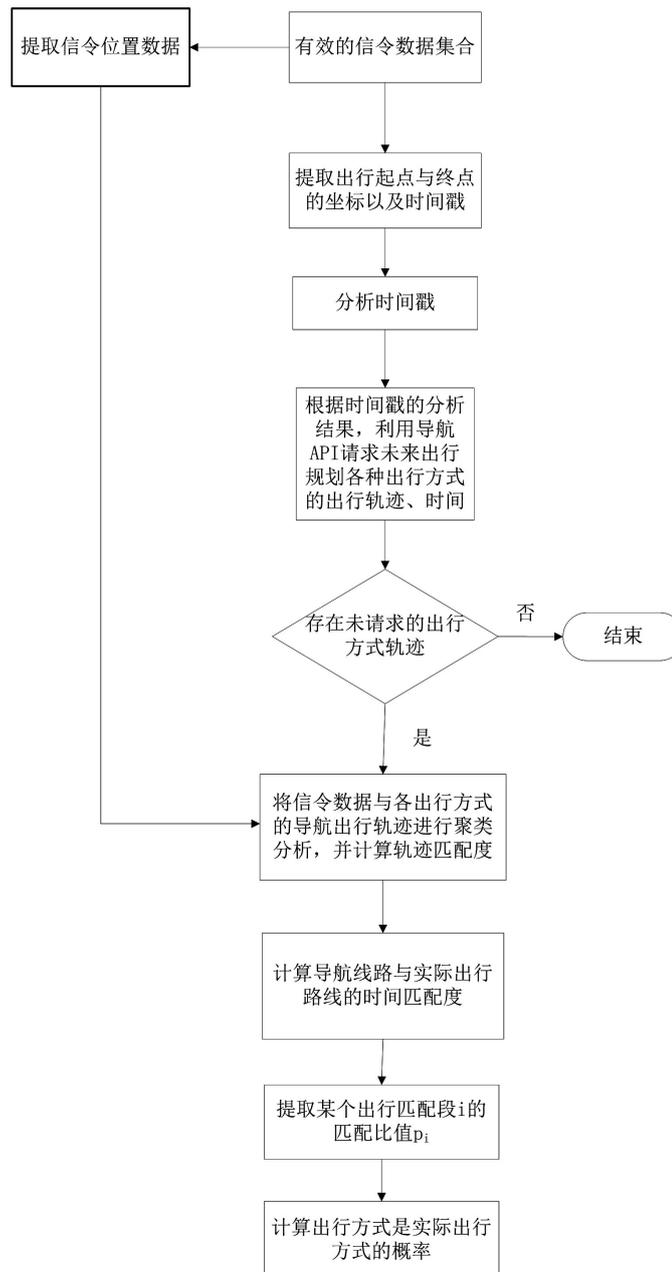


Figure 2. Identification process of ground transportation travel mode
图 2. 地面交通出行方式识别流程

3.1.2. 地铁出行方式识别

将信令数据中的位置数据在地铁基站数据库中进行遍历查询，若信令数据中出现与地铁基站数据相同的情况则判断该出行段的出行方式为地铁出行。

3.2. 评价指标

运用本文提出的识别算法，对出行段进行分割、相似性度量以及获取的出行轨迹点，选用查准率、召回率作为指标进行评价。

1)查准率(The precision, P)，指判定正确的样本的个数占判定结果总数的比例。

$$P = \frac{TP}{TP + FP} \tag{4}$$

2)召回率(The recall, R)，指判定正确的样本的个数占方式下实际样本个数的比例。

$$R = \frac{TP}{TP + FN} \tag{5}$$

在公式(4)~(5)中 TP 、 FP 、 FN 所代表的含义如下。

TP : 判定正确的样本的个数;

FP : 误判为该出行方式的样本的个数;

FN : 被误判为其他出行方式的样本的个数。

4. 实例分析

为对算法识别的准确性进行检验，本文采用某运营商提供的志愿者单向加密后的数据进行研究，部分数据如表 2 所示。

Table 2. Signaling data

表 2. 信令数据

uid	stime	etime	lon	lat
-9223*****	2017/1/2 7:19:31	2017/1/2 7:19:50	116.877865	39.954074
-9223*****	2017/1/2 7:20:07	2017/1/2 7:20:12	116.876572	39.95519

根据信令数据提取出行段的信息，见表 3。

Table 3. Signaling data information extraction

表 3. 信令数据信息提取

判断指标	提取信息
是否包含地铁出行	否
所属情景	情景 1
出行起点时间	7:19:31
出行终点时间	7:43:16
出行时长	1425s
出行起点坐标	116.877865, 39.954074
出行终点坐标	116.803362, 39.971522

使用百度地图应用程序接口请求从起点到终点各种交通出行方式在情景 1 下的路径规划，提取规划路径的坐标、时间。

根据基站小区的范围以及基站与道路之间的映射关系，令 DBSCAN 算法中的 $Eps = 300\text{ m}$ ， $MinPts = 2$ ；轨迹匹配度 $K = 0.7$ ，时间匹配度 0.3。各出行方式下的概率见表 4。

Table 4. Travel mode and path probability

表 4. 出行方式及路径概率

	公交		小汽车			骑行	步行
	线路 1	线路 2	线路 A	线路 B	线路 C	线路	线路
轨迹匹配度	68.13%	26.23%	93.41%	79.56%	86.39%	28.54%	84.02%
时间匹配度	32.21%	14.6%	67.25%	60.79%	51.58%	0%	0%
概率	57.35%	22.74%	85.56%	73.93%	75.95%	19.98%	58.81%

由此可知该用户此次从起点到终点采用小汽车的出行方式沿线路 A 出行。

利用出行方式识别模型，对信令数据样本进行测试。表 5 是样本交通出行方式识别结果。

Table 5. Comparison between recognition results and actual results

表 5. 识别结果与实际对比

		识别结果					查全率
		步行	骑行	公交车	小汽车	总和	
真实 轨迹 数据	步行	121	12	4	5	142	85.2%
	骑行	9	109	3	9	130	83.8%
	公交车	6	7	131	8	152	86.2%
	小汽车	5	10	7	168	190	88.4%
	总和	138	137	149	190	614	/
查准率		87.7%	80.0%	87.9%	88.4%	/	/
正确率		86.12%					

614 个出行段样本中，步行的样本个数为 142、骑行的样本个数为 130、公交车的样本个数为 152、小汽车的样本个数为 190。各出行方式的查准率、召回率均达到 80.0% 以上，整体识别的正确率也达到了 86.12%。

5. 结果和分析

为了对算法的有效性与合理性进行判断，选取了查准率、召回率、正确率对模型的精度进行评价。由表 5 可知，各出行方式识别的查准率、召回率均达到了 80% 以上，整体识别正确率达到了 86.12%。目前，部分学者利用手机信令数据进行出行方式识别方面的研究，文献[15]构建隶属度函数模型整体准确率为 68%。相比之下，本文提出的识别算法能更为精确的识别用户的出行方式。这是因为传统的隶属度函数模型考虑的影响因素过于单一。本文基于手机信令数据、结合导航数据和地铁基站数据有效的区分了地面与地铁的出行，降低了各出行方式之间的相互干扰。

6. 结语

针对目前基于手机信令数据的出行方式识别方法研究较少且正确率较低现状，结合手机信令数据、

地图导航数据、以及地铁基站数据,提出了出行方式与路径的识别算法。在基于志愿者的 614 个出行段信令数据样本的基础上,利用该识别模型得到的结果整体的正确率达到了 86.12%,高于现有的识别方法。此外本文针对地铁出行进行了分析,能够有效降低模型对地铁出行的误判率。但是,受样本数据的限制无法对地铁出行进行验证且导航数据也仅考虑了高峰、平峰时段对于导航数据的影响忽略了道路设施的因素,需要在以后的研究中进行一定的修正。

基金项目

智能交通技术交通运输行业重点实验室开放课题资助(F262019016);教育部人文社会科学研究规划基金(19YJAZH011);江苏省交通厅运输厅科技项目(KY2018049)。

参考文献

- [1] 唐小勇,周涛,陆百川. 基于手机信令的大范围人流移动分析[J]. 重庆交通大学学报(自然科学版), 2017, 36(1): 82-87.
- [2] Yang, F., Yao, Z. and Jin, P.J. (2015) Multi-Mode Trip Information Recognition Based on Wavelet Transform Modulus Maximum Algorithm by Using GPS and Acceleration Data. *Transportation Research Board 94th Annual Meeting*, Washington DC, 11-15 January 2015, 90-98.
- [3] Asakura, Y. and Iryo, T. (2007) Analysis of Tourist Behaviour Based on the Tracking Data Collected Using a Mobile Communication Instrument. *Transportation Research Part A: Policy and Practice*, **41**, 684-690. <https://doi.org/10.1016/j.tra.2006.07.003>
- [4] Caceres, N., Wideberg, J.P. and Benitez, F.G. (2007) Deriving Origin Destination Data from a Mobile Phone Network. *IET Intelligent Transport Systems*, **1**, 15-26. <https://doi.org/10.1049/iet-its:20060020>
- [5] Vajakas, T., Vajakas, J. and Lillemets, R. (2015) Trajectory Reconstruction from Mobile Positioning Data Using Cell-to-Cell Travel Time Information. *International Journal of Geographical Information Science*, **29**, 1941-1954. <https://doi.org/10.1080/13658816.2015.1049540>
- [6] 李祖芬,于雷,高永,等. 基于手机信令定位数据的居民出行时空分布特征提取方法[J]. 交通运输研究, 2016, 2(1): 51-57.
- [7] Stopher, P., Fitzgerald, C. and Zhang, J. (2008) Search for a Global Positioning System Device to Measure Person Travel. *Transportation Research Part C Emerging Technologies*, **16**, 350-369. <https://doi.org/10.1016/j.trc.2007.10.002>
- [8] Bohte, W. and Maat, K. (2009) Deriving and Validating Trip Purposes and Travel Modes for Multi-Day GPS-Based Travel Surveys: A Large-Scale Application in the Netherlands. *Transportation Research Part C*, **17**, 285-297. <https://doi.org/10.1016/j.trc.2008.11.004>
- [9] Chen, C., Gong, H., Lawson, C., et al. (2010) Evaluating the Feasibility of a Passive Travel Survey Collection in a Complex Urban Environment: Lessons Learned from the New York City Case Study. *Transportation Research Part A: Policy and Practice*, **44**, 830-840. <https://doi.org/10.1016/j.tra.2010.08.004>
- [10] Anderson, I. and Muller, H. (2008) Exploring GSM Data in Pervasive Environments. *International Journal of Pervasive Computing and Communications*, **4**, 8-25. <https://doi.org/10.1108/17427370810873075>
- [11] Wang, H., Calabrese, F., Di Lorenzo, G., et al. (2010) Transportation Mode Inference from Anonymized and Aggregated Mobile Phone Call Detail Records. *13th International IEEE Conference on Intelligent Transportation Systems*, Funchal, 19-22 September 2010, 318-323. <https://doi.org/10.1109/ITSC.2010.5625188>
- [12] Ning, Y. and Liu, X.X. (2015) Traffic Pattern Recognition System Design and Development Based on Smart Phones. *International Conference on Intelligent Computation Technology & Automation*, Changsha, 25-26 October 2014, 154-157. <https://doi.org/10.1109/ICICTA.2014.45>
- [13] Hong, S. and Vonderohe, A. (2014) Uncertainty and Sensitivity Assessments of GPS and GIS Integrated Applications for Transportation. *Sensors*, **14**, 2683-2702. <https://doi.org/10.3390/s140202683>
- [14] Reddy, S., Mun, M., Burke, J., et al. (2010) Using Mobile Phones to Determine Transportation Modes. *ACM Transactions on Sensor Networks*, **6**, 13. <https://doi.org/10.1145/1689239.1689243>
- [15] 张博. 基于手机网络定位的 OD 调查的出行方式划分研究[D]: [硕士学位论文]. 北京: 北京交通大学, 2010.
- [16] 闫彭. 基于 AGPS 手机的交通方式识别研究[D]: [硕士学位论文]. 北京: 北京交通大学, 2012.

- [17] 张治华. 基于 GPS 轨迹的出行信息提取研究[D]: [博士学位论文]. 上海: 华东师范大学, 2010.
- [18] 汪磊, 左忠义, 傅军豪. 基于 SVM 的出行方式特征分析和识别研究[J]. 交通运输系统工程与信息, 2014, 14(3): 70-75.
- [19] 赖见辉, 陈艳艳, 钟园, 等. 基于手机定位信息的地铁乘客出行路径辨识方法[J]. 计算机应用, 2013, 33(2): 583-586.