

移动端个体出行链数据自采设计及出行特征选择

王文静^{1,2}, 陈艳艳¹, 刘冬梅², 赵晓斐³

(1.北京工业大学, 北京 100124; 2.交通运输部公路科学研究院, 北京 100088; 3.清华大学土木工程系, 地球空间信息研究所, 北京 100084)

摘要:手机信令数据和GPS数据等移动互联网位置数据的应用,为获取完整的个体出行链提供了可能。基于移动互联网位置数据提取的出行特征及建立的出行分析模型,与真实的出行情况是否匹配,需要被测试。因此,有必要通过移动互联网位置数据采集和交通调查两种方式,获取同一个体完整的出行链数据,用交通调查数据来标定出行分析结果。设计一种基于移动端的个体出行链数据自采方案,同时采集个体的GPS数据、基站位置数据以及出行链记录数据。在出行特征分析的基础上,应用主成分分析和决策树两种方法,筛选可用于交通方式识别的关键特征。结果表明,速度最大值、速度均值、加速度标准差、速度标准差和方位角变化标准差可显著区分交通方式。

关键词: 交通工程; 出行数据调查; 出行链; 出行方式; 出行特征

Data Mining of Individual Travel Chain Based on Mobile Phone and Travel Characteristics

Wang Wenjing^{1,2}, Chen Yanyan¹, Liu Dongmei², Zhao Xiaofei³

(1.Beijing University of Technology, Beijing 100124, China; 2.Research Institute of Highway, Ministry of Transport, Beijing 100088, China; 3.Institute of Geomatics, Department of Civil Engineering, Tsinghua University, Beijing 100084, China)

Abstract: The application of mobile internet location data, such as cell phone GSM data and GPS data, makes it possible to obtain a complete individual travel chain. Whether the travel characteristics extracted from mobile internet location data and the established travel analysis model match the real-world situation needs to be tested. Therefore, it is necessary to obtain the complete travel chain data of the same individual through two ways of mobile internet location data collection and traditional travel survey, and use travel survey data to calibrate travel analysis results. This paper designs an individual travel chain data acquisition scheme based on cell phone, and collects three types of data simultaneously, including individual GPS data, base station location data and travel chain. Based on the analysis of travel characteristics, two methods of principal Component Analysis (PCA) and decision tree-based methods are applied to screen key characteristics that can be used for travel mode identification. The results show that the maximum speed, average speed, standard deviation of acceleration, standard deviation of speed and standard deviation of azimuth change can distinguish travel modes significantly.

Keywords: transportation engineering; travel data survey; travel chain; travel modes; travel characteristics

收稿日期: 2019-08-06

基金项目: 国家重点研发计划(2017YFC0840200)课题六“海量个体出行数据驱动的城市综合客运协同服务评估: 技术与应用”(2017YFC0840206)

作者简介: 王文静(1987—), 女, 山西河津人, 在读博士研究生, 高级工程师, 主要研究方向: 智能交通、交通数据分析。E-mail: wenjingwang@emails.bjut.edu.cn

通信作者: 陈艳艳(1970—), 女, 河南郑州人, 博士, 教授, 主要研究方向: 智能交通、交通大数据挖掘等。E-mail: cdyan@bjut.edu.cn

1 研究概述

由公共汽车、出租汽车、城市轨道交通、非机动车等构成的城市客运交通网络,

涵盖人们出行的主要过程及方式^[1-2]。城市交通管理部门可以获取多种不同方式的出行数据,并基于这些数据的分析和优化为公众提供出行服务。然而,这些过程仍然存在无法

有效感知的环节。例如，选择公共交通出行的乘客在使用公共交通之外的环节中，出行方式及出行时间无法获知；公共汽车和轨道交通换乘过程中的步行换乘时间和站台等候时间无法监测。有效的衔接换乘以及出行全过程的便利程度，对人们选择不同的交通方式有很大影响^[3-5]。这些环节和过程的完整感知，需要获取个体完整的出行链数据。

出行链又称为基于活动的出行模式研究。活动是指在一个连续的时段内为达到某种目的而完成的事件。出行目的通常有上班、上学、就医、购物、用餐、接送人等。个体出行链^[6-7]描述个体完整的出行，用以分析出行者在何时、何地、花了多长时间，用什么交通方式及其活动目的。出行链可包含以下信息：出行OD、交通方式^[8]、活动类型^[9]，以及与这些信息对应的位置和时间。

手机信令数据和GPS数据等移动互联网位置数据的采集，为获取完整的出行链提供了可能^[10]。由于移动互联网位置数据和交通行业的出行数据无法进行直接匹配，基于移动互联网位置数据建立的出行特征提取模型和交通方式识别模型，能否体现出行者真实的交通方式和出行状态值得研究。

因此，从移动互联网位置数据和交通调查数据两个层面获取同一个体的完整的出行链数据，用交通调查数据来标定移动互联网

位置数据分析的交通方式和交通状态是非常必要的。在此背景下，本文设计了基于移动端的个体出行链数据采集方案，同时采集个体的出行调查数据和移动互联网位置数据，获取个体完整的出行链数据，为建立基于个体出行链的交通方式识别模型提供数据支持。

在基于出行特征提取的交通方式识别方面，国内外学者已做了一些研究。数据源从交通调查逐渐扩展到便携式的手机端，数据量也随之增大。从表1中可以看出，数据类别以GPS数据和加速度数据^[11-12]为主，辅以不同的出行调查数据^[13]；基于数据提取了不同的交通特征^[14]，如平均速度、最大速度、平均加速度和最大加速度等作为模型输入；采用随机森林、支持向量机、贝叶斯等算法，识别志愿者或者个体的交通方式数据。然而，依然有问题值得探讨。一是在数据采集层面，需要采集或记录哪些数据来描述出行链；二是哪些出行特征可以更显著地区分交通方式。

因此，本文设计了基于移动端的出行链数据采集方案，采集个体的GPS数据、基站位置数据和出行链记录数据，以识别个体出行链为目的，将各类特征进行汇总对比；进而分析出行特征的重要程度及其对交通方式识别的影响，为大规模移动互联网位置数据的应用及个体出行链模型建立提供基础。

表1 出行特征提取及出行方式识别研究现状

Tab.1 Current research on travel characteristics extraction and travel mode recognition

文献来源	主题	数据量	数据类别	模型输入	结果输出	算法
文献[11]	基于GPS的多日出行调查的出行目的和交通方式；在荷兰的一个应用	1 104名志愿者携带GPS设备一周	GPS, GIS, 个体特征数据	距离, 速度, 最大速度, 与POI的距离	交通方式(步行、自行车、公共交通、汽车、火车)和出行目的	基于规则的算法
文献[12]	使用GPS设备从旅行调查中提取出行行程和交通方式；以哥本哈根地区为例	183人携带GPS设备3~5天	GPS数据	与轨道交通车站的距离, 平均速度和最大速度, 加速度, 公共汽车线路	交通方式(火车、步行、自行车、汽车、公共交通)	结合模糊逻辑和基于GIS的原始GPS数据处理算法
文献[13]	将随机森林方法应用于模型的交通方式选择行为	使用1 435户家庭共2 991人7 276次出行进行分析	南京2013年出行调查数据	家庭特征, 个体属性, 建筑环境, 出行时间和出行目的	交通方式(步行、自行车、电动自行车、公共交通、机动车)	随机森林
文献[14]	使用GPS和加速度数据识别交通方式	80 670个数据记录	GPS, 加速度参数, 出行活动记录	仅用GPS数据, 仅用加速度计数据, 加速度计和GPS数据组合	交通方式(步行、跑步、自行车、摩托车、公共交通、汽车、有轨电车和火车)	贝叶斯信念网络模型
文献[15]	使用GPS识别交通方式	900名受访者100多万份出行记录	GPS, GIS, 出行日记	平均速度, 最大速度, 75分位加速度, 出行时间, 速度标准差	交通方式(步行、自行车、机动车)	多项式逻辑模型
文献[16]	使用多个移动窗口从GPS数据中识别活动和出行段	100个模拟活动约23.5万个标记数据点	GPS数据	不同时间窗口下的平均速度, 最大速度, 平均加速度, 最大加速度	交通状态(移动、停留)	朴素贝叶斯, 支持向量机, 随机森林

2 研究方法

2.1 基于移动端的出行链数据自采方案设计

基于移动端的出行数据采集,是通过移动端应用程序(APP)结合手工记录,完成个体出行过程记录,即个体的完整的出行链记录,包含大量的时间、空间、交通方式和活动类型信息。具体信息包括:1)个体GPS数据;2)个体的基站位置数据(与信令数据类似,但当前数据来源于移动端应用程序采集的出行过程中的基站位置数据,下阶段可探讨运营商的信令数据分析);3)出行全过程的出行链记录;4)对出行各环节的评价信息。出行数据调查方案设计思路如图1所示,包括数据需求、数据调查与采集、数据分析三部分。

1) 数据需求。主要采集两类数据:可反映完整出行信息的自动采集数据,如个体GPS数据和用户手动记录的真实的出行数据。

2) 数据调查与采集。数据调查要求包括:采集的数据需涵盖3类日常活动出行链,强制性(上班和上学等通勤出行)、非强制性(社交、休闲娱乐和自由活动等生活出行)和在家活动;交通方式涵盖公共汽车、轨道交通、小汽车、班车、自行车、步行等;出行目的涵盖上下班、上下学等通勤类以及购物、休闲娱乐等生活类;出行者类别涵盖通勤者、学生、老年人等。在调查组织与实施阶段,征集志愿者,安装移动端数据

采集软件后,收集数据并校验。

3) 数据分析。对采集的出行数据进行分析,计算单个位置点数据及出行段数据的出行特征,基于出行方式划分结果对出行特征的重要度进行排序。

2.2 出行特征提取及分析

出行特征提取的目的,一是区分交通状态,判断个体的状态是出行还是停留;二是区分交通方式。当前,出行特征提取考虑的因素^[17]主要有速度、缺失值、方向变化特征、与路网的偏离距离等。将出行特征作为输入,可建立交通方式识别模型。这些特征值的重要程度如何,可以多大程度表征不同的交通方式?本文将提取出行链分段后的多种特征值进行分析。

图2a为一条出行链的轨迹示意图,可转化为交通状态和交通方式的出行链表示(见图2b):交通状态包括出发(起点)、等车及到达(终点);交通方式为步行+公共汽车+步行。

为了区分不同的交通方式,需要计算与交通方式相对应的出行段的特征,包括距离、方向、速度、加速度等。对每一个经纬度点,可计算其与前一位置点的移动距离、当前点的方向分量、速度方向分量和加速度方向分量。计算公式如下:

$$D_i = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}, \quad (1)$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad (2)$$

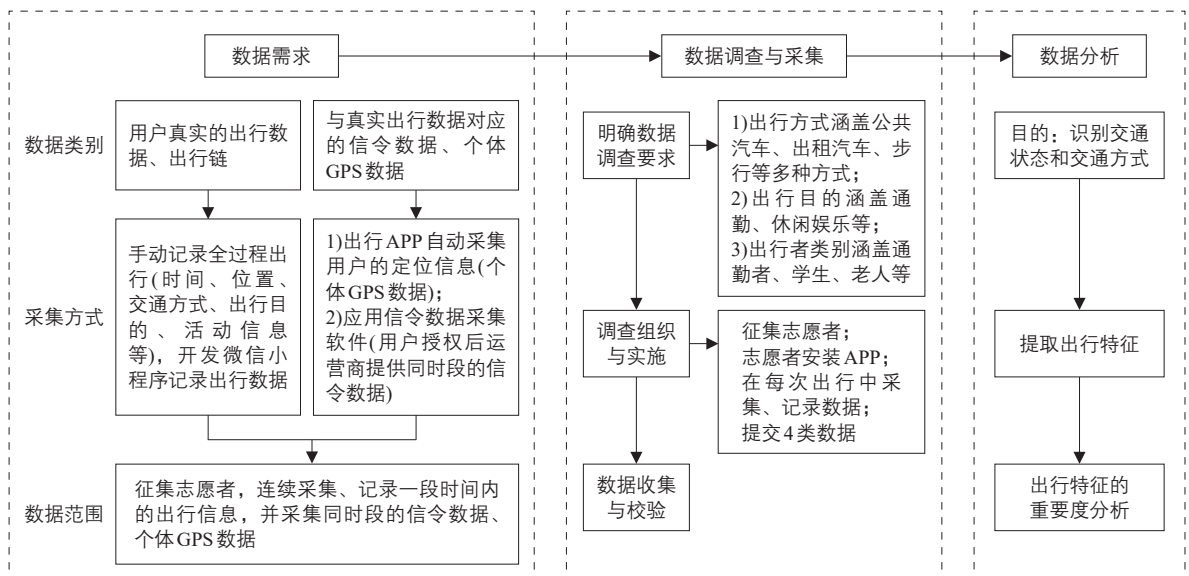


图1 出行数据调查方案设计思路

Fig.1 Travel survey scheme design

$$\begin{bmatrix} v_{x,i} \\ v_{y,i} \end{bmatrix} = \frac{1}{T} \left(\begin{bmatrix} x_i \\ y_i \end{bmatrix} - \begin{bmatrix} x_{i-1} \\ y_{i-1} \end{bmatrix} \right), \quad (3)$$

$$\begin{bmatrix} a_{x,i} \\ a_{y,i} \end{bmatrix} = \frac{1}{T} \left(\begin{bmatrix} v_{x,i} \\ v_{y,i} \end{bmatrix} - \begin{bmatrix} v_{x,i-1} \\ v_{y,i-1} \end{bmatrix} \right), \quad (4)$$

式中： D_i 为从前一位置点到当前点的移动距离/m； x_i 和 y_i 为当前位置点的坐标； x_{i-1} 和 y_{i-1} 为前一位置点的坐标；在一段时间 T 内有 n 个位置记录， \bar{x} 和 \bar{y} 为方向分量， $v_{x,i}$ 和 $v_{y,i}$ 为速度分量/($m \cdot s^{-1}$)， $a_{x,i}$ 和 $a_{y,i}$ 为加速度分量/($m \cdot s^{-2}$)。

在单个轨迹点的特征计算基础上，结合文献学习^[11-16]，对各出行段从速度、加速度、方向变化、距离、状态等方面进行特征分析，形成出行段的数据特征集(见表2)。

2.3 出行特征选择模型

出行特征选择的依据是各出行特征对交通方式的解释程度，自变量为出行特征，因变量为交通方式。分别应用主成分分析(Principal Component Analysis, PCA)和决策树建立各出行特征与交通方式之间的关系模型，根据模型分析结果评判各出行特征的重要程度。

其中，PCA将已有的自变量转换为一系列名为“主成分”的新变量，即捕捉可以最大程度解释数据集波动的一些属性。PCA的数学定义是^[18]：一个正交化线性变换，把数据变换到一个新的坐标系统中，使得这一数

据的任何投影的第一大方差在第一个坐标(第一主成分)上，第二大方差在第二个坐标(第二主成分)上，以此类推。公式如下：

$$C_n = b_{n1}(x_1) + b_{n2}(x_2) + \dots + b_{np}(x_p), \quad (5)$$

式中： C_n 为每组特征组合在主成分 n 上的得分； b_{np} 为观察变量即特征 p 的回归系数(或权重)； x_p 为当前组合中第 p 个特征上的得分。根据主成分的排名，筛选前三个主成分中的各特征组合，即可得到重要程度相对较高的特征。

决策树又称分类树，是一种基于规则的工具。通过系统地考虑所有属性对最终归类的影响，把数据分成不同类别^[19]。在使用训



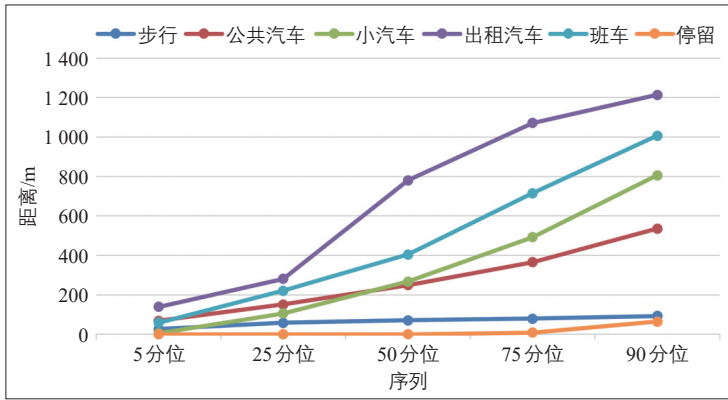
图2 出行轨迹及出行链示例

Fig.2 Example of travel path and travel chain

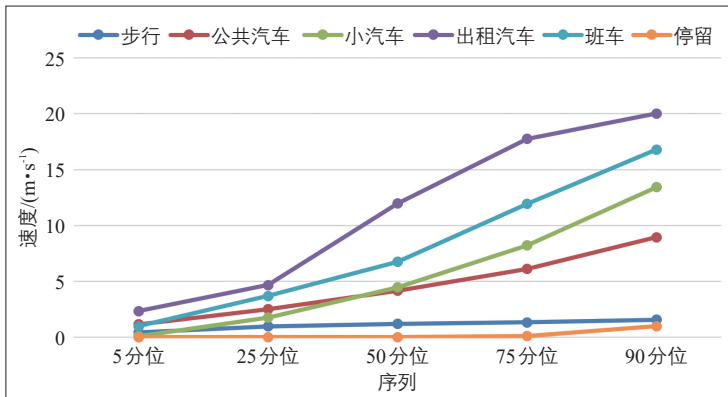
表2 出行段的数据特征集

Tab.2 Data characteristics of travel segment set

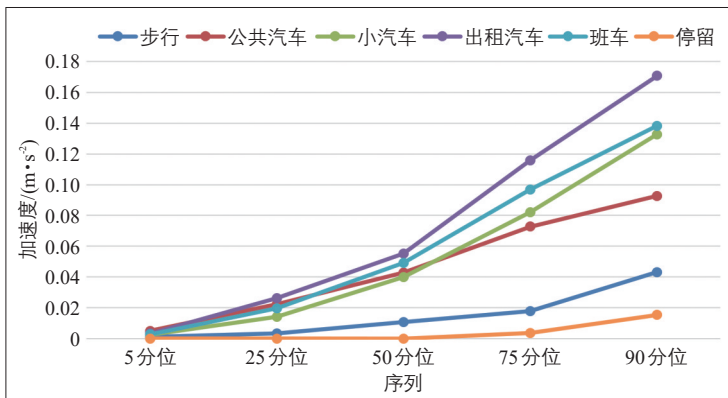
特征分类	特征名称	物理意义	计算方法
速度类	平均速度	整个出行段的平均速度、速度最大值、速度变化情况	为出行段内各轨迹点的速度升序排序后，取排位在95%，85%，50%，5%位置的速度值
	速度中位数		
	速度95分位数		
	速度标准差		
加速度类	平均加速度	整个出行段的平均加速度、加速度最大值、加速度变化情况	为出行段内各轨迹点的加速度升序排序后，取排位在95%，85%，50%，5%位置的加速度值
	加速度中位数		
	加速度95分位数		
	加速度标准差		
时间类	总出行时间	发生时刻用来区分出行是工作日还是周末，以及是否早晚高峰；前停驻时间指出行前在路边等待的时间，公共汽车有前停驻时间，私人汽车几乎没有	根据各位置点的时间和状态进行计算
	出行发生的时刻 前停驻时间		
方向类	平均方向变化 方向变化的95分位数 方向变化的标准差	方位角的变化情况	统计各点的方位角分量
距离类	出行距离 直线距离 非直线系数	出行距离和直线距离的比值一般被表征为非直线系数，不同交通方式的非直线系数存在一定差异	出行段各段的距离相加为出行距离
停留类	一个出行段的停留次数 停留段的距离 停留时长	不同交通方式的可能停留次数不同	统计停留段的相关特征



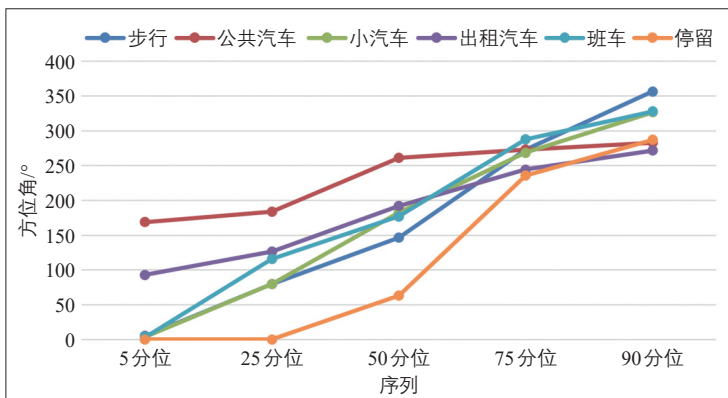
a 距离



b 速度



c 加速度



d 方位角

图3 不同交通方式的距离、速度、加速度和方位角的分布统计
Fig.3 Distribution of travel distance, velocity, acceleration, and azimuth by different travel modes

练集构建模型的同时，完成特征选择。决策树的主要算法是C4.5^[20]，使用决策树进行决策的过程就是从根节点开始，测试待分类项中相应的特征属性，并按照其值选择输出分支，直到到达叶子节点，将叶子节点存放的类别作为决策结果。

3 数据采集情况说明

在出行链数据采集测试阶段，共征集30名志愿者，均为通勤用户，于2018年4—6月收集到167条完整的出行链记录。后期数据大规模采集阶段将覆盖通勤者、老年人和学生等人群的数据，并对不同人群的出行链进行分析。根据手动记录的活动及停留状态对数据分段后，共得到1125条出行段数据，每条数据对应一种交通方式。

由于轨道交通记录因信号问题缺失，得到步行、公共汽车、小汽车、出租汽车和班车5种交通方式的出行数据和停留状态的数据，对每种交通方式的出行距离、速度、加速度和方位角的分布情况进行统计(见图3)。初步推断，速度均值可对步行、公共汽车和其他机动车进行区分；距离、速度和加速度的高分位数指标比低分位数指标更容易区分不同交通方式。

4 数据结果分析

每条出行数据有7个特征值，1个标签列(交通方式)。各特征值的统计分析数据如表3所示。特征“距离”的均值和中位数相差较大，极差(最大值和最小值之差)很大，说明该特征可能存在离群点；各特征的众数为0，即出现次数最多的数据是0，这与停留状态有关。特征“距离”和“一段时间的最高速度”方差较大，说明各数据点该项特征距离中心点的分布更分散。

接下来，应用PCA及基于决策树的特征选择筛选关键特征。

4.1 基于PCA的特征选择

将PCA应用于数据集，通过计算，前三个主成分解释近86%的方差。从前三个主成分的特征向量中找出原来的变量，可知速度最大值、速度均值、速度标准差、加速度标准差、方位角变化标准差、方位角变化均

值、出行开始时间对结果的影响较大(见表4)。

一步提高交通方式分类的准确率。

4.2 基于决策树的特征选择

将决策树应用于数据集，位于树形结构的上层节点的一些属性即为相对重要的属性。

从图4中可以看出，关键特征包括：速度最大值(85分位速度)、速度均值、加速度标准差、速度标准差、方位角变化标准差。继续变化决策树的深度、最小增益以及最小叶节点数等参数，得到的关键特征无明显变化。

将数据集划分为70%的训练集和30%的测试集，应用决策树进行交通方式分类分析，交通方式的平均识别率为80%。其中步行、公共汽车、小汽车(包含出租汽车和班车)的识别率依次为82%，80%，78%。结果表明，所选特征可以用于交通方式分类。在后续的研究中，将探索建立不同的模型，进

4.3 小结

对出行特征进行统计，并应用PCA和基于决策树的特征选择，对关键特征进行排序。如表5所示，决策树得到的关键特征包括，速度最大值(85分位速度)、速度均值、加速度标准差、速度标准差、方位角变化标准差，同时也是PCA得到的排名靠前的关键特征。此外，应用PCA可获得更多的关键指标，如速度中位数、加速度中位数、95分位加速度、方位角变化均值、出行开始时间等。因此，可判定速度最大值(85分位速度)、速度均值、加速度标准差、速度标准差、方位角变化标准差等能更显著地区分交通方式。在需要探索算法精度或增加特征时，可考虑PCA中排名相对靠前的其他指标。

应用PCA和基于决策树的特征选择结

表3 出行数据集与统计分析

Tab.3 Travel location data sets and statistical analysis

项目	距离/m	速度/ (m·s ⁻¹)	方位角/ °	加速度/ (m·s ⁻²)	速度均值/ (m·s ⁻¹)	速度差	一段时间的 最高速度/(m·s ⁻¹)	
出行数据 集ID	1	71.50	1.19	266.11	0.02	1.13	0.06	1.70
	2	80.91	1.35	273.13	0.00	1.24	0.11	1.70
	85.01	1.42	269.44	0.00	1.33	0.08	1.70
	1 025	64.16	1.07	202.94	0.02	0.44	0.63	0.71
均值	295.93	4.66	164.23	0.04	4.67	-0.01	6.88	
中位数	145.83	2.43	176.96	0.02	3.28	0.00	5.50	
众数	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
极差	14 392.79	28.70	359.93	0.33	27.71	34.60	28.70	
标准差	554.04	5.74	115.19	0.05	5.23	3.77	6.76	
方差	306 956.30	32.90	13 269.70	0.00	27.37	14.22	45.71	

表4 出行特征前三种主成分的属性组成

Tab.4 The attributes composition of the first three principal components in travel characteristics

属性	PC1	属性	PC2	属性	PC3
速度最大值	0.261	加速度标准差	0.364	方位角变化标准差	0.596
速度均值	0.258	方位角变化标准差	0.318	方位角变化均值	0.469
速度标准差	0.257	加速度最大值	0.301	出行开始时间	0.397
速度中位数	0.254	方位角变化均值	0.271	出行时间	0.294
加速度中位数	0.233	加速度均值	0.270	速度标准差	0.125
加速度最大值	0.214	非零加速度均值	0.157	速度最大值	0.124
加速度最小值	0.204	加速度中位数	0.120	速度均值	0.086
加速度均值	0.201	出行开始时间	0.075	速度中位数	0.076

果，以一条下班回家的典型的公共汽车出行链为例，对出行链进行划分，横坐标为时间，纵坐标为当前时刻对应的各出行特征值。如图5所示，速度、加速度、方位角和速度最大值等出行特征，可在一定程度上区分不同出行段的交通状态及交通方式。

5 结语

现有的交通方式识别研究多侧重于建立

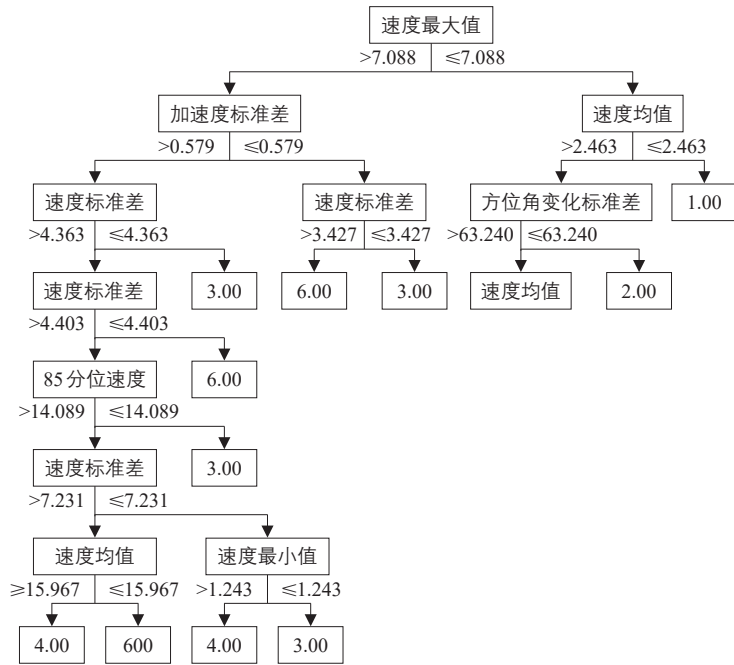


图4 基于决策树的特征选择

Fig.4 Feature selection based on decision tree

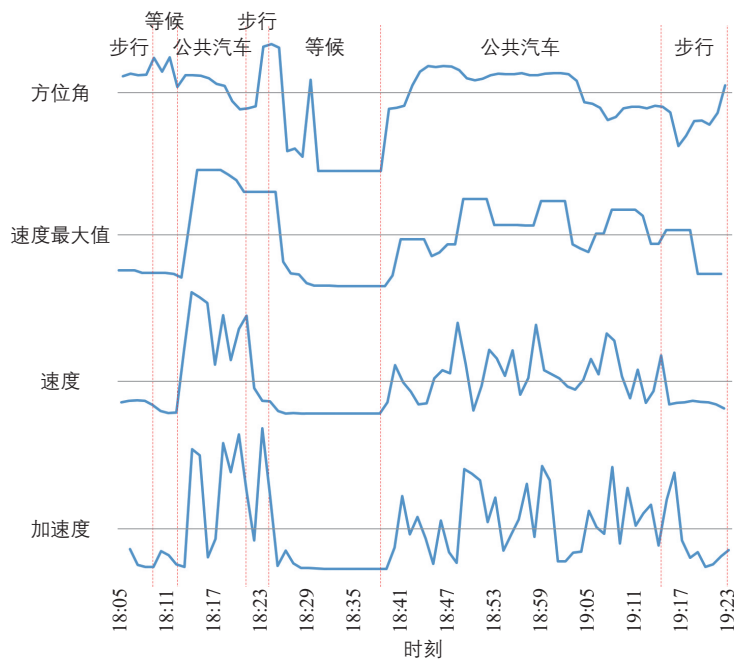


图5 典型的公共汽车出行链出行特征示例

Fig.5 A typical example of travel characteristics of bus travel chain

特征、训练模型并分析精度，但是不同出行特征对模型的影响情况，特征选取和设置是否可以有效区分交通状态和交通方式，仍值得探讨。本文从真实的出行状态和交通方式数据分析出发，进行了出行特征的多维度探索。通过出行数据采集对移动互联网位置数据的交通属性进行真值标定，为模型校验提供数据支持。对可识别交通方式的特征进行数据探索，建立出行状态和出行方式识别的特征集，并使用PCA和基于决策树的特征选择对特征进行分析。结果表明，速度最大值、速度均值、加速度标准差、速度标准差、方位角变化标准差能更显著地区分交通方式。

本文的创新点在于：1)设计了一种出行调查方法，基于移动端手机应用，同时采集手机GPS、基站位置数据和出行调查三类数据，拓展了可采集的数据源，为后续基于GPS数据或手机信令数据分别建立出行链模型提供数据基础；2)对比了可提取的各类出行特征，筛选出关键特征，为基于移动互联网位置数据的出行链建模提供有效的解释变量。

下一步研究的重点在于：一方面，以更轻量化的方式采集大规模的真实出行数据，在特征分析的基础上，将用户进行分类，针对不同群体的用户建立出行特征，并将特征数据应用于多种交通状态和交通方式判别模型，提升算法对出行链标定的准确度，在可获取的较大范围的居民出行数据上验证并应用；另一方面，探索自动判别方法代替实际出行调查的可行性，例如，与地图服务商合

表5 PCA与决策树对特征的选择排序

Tab.5 Feature selection and ranking by PCA and decision tree

排序	PCA	基于决策树的特征选择
1	速度最大值	速度最大值
2	速度均值	速度均值
3	速度标准差	加速度标准差
4	速度中位数	速度标准差
5	加速度中位数	方位角变化标准差
6	加速度标准差	
7	方位角变化标准差	
8	95分位加速度	
9	方位角变化均值	
10	出行开始时间	

作, 将用户的出行路径查询数据、出行常访地数据和轨迹点数据进行匹配, 获取较大范围的真实出行数据及其交通方式。

参考文献:

References:

[1] Akin D. Model for Estimating Increased Ridership Caused by Integration of Two Urban Transit Modes: Case Study of Metro and Bus-Minibus Transit Systems, Istanbul, Turkey[J]. Transportation Research Record: Journal of the Transportation Research Board, 2007 (1986): 162-171.

[2] Sun Yang, Sun Xiaonian, Li Baoqing, et al. Joint Optimization of a Rail Transit Route and Bus Routes in a Transit Corridor[J]. Procedia - Social and Behavioral Sciences, 2013, 96: 1218-1226.

[3] Guo Zhan, Wilson N. Assessing the Cost of Transfer Inconvenience in Public Transport Systems: A Case Study of the London Underground[J]. Transportation Research Part A: Policy and Practice, 2011, 45(2): 91-104.

[4] Liu Tao, Ceder A. Communication-Based Cooperative Control Strategy for Public Transport Transfer Synchronization[J]. Transportation Research Record: Journal of the Transportation Research Board, 2016(2541): 27-37.

[5] Ceder A, Chowdhury S, Taghipouran N, et al. Modelling Public-Transport Users' Behaviour at Connection Point[J]. Transport Policy, 2013, 27: 112-122.

[6] Ye Xin, Ram M, Gottardi G. An Exploration of the Relationship Between Mode Choice and Complexity of Trip Chaining Patterns[J]. Transportation Research Part B: Methodological, 2007, 41(1): 96-113.

[7] 褚浩然, 郑猛, 杨晓光, 等. 出行链特征指标的提出及应用研究[J]. 城市交通, 2006, 4(2): 64-66.

Chu Haoran, Zheng Meng, Yang Xiaoguang, et al. A Study on Trip-Chain Indices and Their Application[J]. Urban Transport of China, 2006, 4(2): 64-66.

[8] 刘侃, 张蕊, 杨静, 等. 流动人口出行特征及出行方式选择模型: 以北京市短期旅居流动人口为例[J]. 城市交通, 2018, 16(3): 87-92.

Liu Kan, Zhang Rui, Yang Jing, et al. Travel Characteristics and Mode Choice Model of

Floating Population: Take Short-Term Floating Population in Beijing as an Example[J]. Urban Transport of China, 2018, 16(3): 87-92.

[9] 刘冬梅, 荣建, 张劲泉, 等. 面向公交全出行链的短时服务评价模型研究[J]. 公路交通科技, 2018, 35(8): 123-132.

Liu Dongmei, Rong Jian, Zhang Jinquan, et al. Study on Public Transport Trip Chain Oriented Short-Time Service Evaluation Model [J]. Journal of Highway and Transportation Research and Development, 2018, 35(8): 123-132.

[10] 吴子啸. 基于手机数据的出行链推算算法[J]. 城市交通, 2019, 17(3): 11-16.

Wu Zixiao. Travel Chain Estimation Based on Cell Phone Data[J]. Urban Transport of China, 2019, 17(3): 11-16.

[11] Bohte W, Maat K. Deriving and Validating Trip Purposes and Travel Modes for Multi-Day GPS-Based Travel Surveys: A Large-Scale Application in the Netherlands[J]. Transportation Research Part C: Emerging Technologies, 2009, 17(3): 285-297.

[12] Rasmussen T, Ingvarsson J, Halldórsdóttir K, et al. Improved Methods to Deduct Trip Legs and Mode from Travel Surveys Using Wearable GPS Devices: A Case Study from the Greater Copenhagen Area[J]. Computers, Environment and Urban Systems, 2015, 54: 301-313.

[13] Cheng Long, Chen Xuewu, De Vos J, et al. Applying a Random Forest Method Approach to Model Travel Mode Choice Behavior[J]. Travel Behaviour and Society, 2019, 14: 1-10.

[14] Feng Tao, Timmermans H. Transportation Mode Recognition Using GPS and Accelerometer Data[J]. Transportation Research Part C: Emerging Technologies, 2013, 37: 118-130.

[15] Zong Fang, Yuan Yixin, Liu Jianfeng, et al. Identifying Travel Mode with GPS Data Using Support Vector Machines and Genetic Algorithm[J]. Transportation Planning and Technology, 2017, 40(2): 242-255.

[16] van Dijk J. Identifying Activity-Travel Points from GPS-Data with Multiple Moving Windows[J]. Computers, Environment and Urban Systems, 2018, 70: 84-101.