

基于自动售检票数据的城市轨道交通通勤客流辨识

邹庆茹¹,赵 鹏¹,姚向明¹,汪 波²

(1. 北京交通大学 交通运输学院,北京 100044; 2.北京市交通委员会,北京 100161)

摘 要:城市轨道交通具有明显的通勤客流主体特征,把握通勤客流出行规律对运营管理具有重要意义.本文以海量自动售检票数据潜在包含的时空关系为基础,从时间、空间、个体属性、出行规律四方面构建基于规则的出行目的辨识算法,重点针对通勤(上下班及上下学)客流进行辨识.以 2014 年北京市轨道交通售检票数据进行实证分析,结果显示:该方法能有效辨识上班、上学、下班回家、放学回家、其他回家和其他 6 类客流,其中上班占比 26.77%,上学占 0.44%,回家占 44.49%(包含下班回家、放学回家及其他回家),其他占 28.30%.结合 2014 年北京市公共交通出行调查结果,验证了辨识结果的准确性.该研究扩展了售检票数据应用范围,为精细化客流特征研究提供了一种低成本、高效的分析方法.

关键词:城市轨道交通;通勤客流辨识;数据挖掘;规则辨识法;自动售检票数据

中图分类号:U239.5 **文献标志码:**A

Commuters identification for urban rail transit using automatic fare collection data

ZOU Qingru¹, ZHAO Peng¹, YAO Xiangming¹, WANG Bo²

(1.School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China;

2.Beijing Municipal Commission of Transport, Beijing 100161, China)

Abstract: The passenger flow of Urban Rail Transit (URT) mainly composes of commuters, and it is of great significance to understand the law of travel characteristics of commuters for the operation and management. Based on the potential spatial-temporal relationship within Automatic Fare Collection (AFC) records, a rule-based identification method is proposed to infer trip purpose for URT travelers, which is constructed from the view of time, space, personal properties and regular travel behaviors. An empirical transit network from Beijing in China is applied to verify the efficiency of the proposed method, results show that six types of trips (home-work, work-home, home-school, school-home, others-home and other) can be efficiently identified, where the work trips cover 26.77 %, the school trips cover 0.44 %, home trips cover 44.49 % and other trips cover 28.30 %. Compared to the public transportation survey results in Beijing of 2014, the

收稿日期:2017-06-01
基金项目:国家自然科学基金(51478036,71701011);中央高校基本科研业务费专项资金资助(2017RC032);中国国家留学基金委(CSC)资助(201707090039)
Foundation items: National Nature Science Foundation of China (51478036, 71701011); Fundamental Research Funds for the Central Universities(2017RC032); China Scholarship Council Funds (201707090039)
第一作者:邹庆茹(1987—),女,黑龙江齐齐哈尔人,博士生.研究方向为轨道交通规划与管理.email:15114194@bjtu.edu.cn.
通信作者:赵鹏(1967—),男,内蒙古通辽人,教授,博士,博士生导师.email:pzhao@bjtu.edu.cn.
引用格式:邹庆茹,赵鹏,姚向明,等.基于自动售检票数据的城市轨道交通通勤客流辨识[J].北京交通大学学报,2018,42(3):44—52.
ZOU Qingru, ZHAO Peng, YAO Xiangming, et al. Commuters identification for urban rail transit using automatic fare collection data[J]. Journal of Beijing Jiaotong University, 2018, 42(3): 44—52. (in Chinese)

identification results are verified to be reasonable and acceptable. This study improves the value of AFC data, and provides a new method for deeply analyzing travel demand.

Keywords: urban rail transit; commuters identification; data mining; rule-based method; automatic fare collection data

自动售检票系统(AFC)不仅提高了乘客支付便捷性,而且为交通规划及运营管理提供了海量基础数据.我国城市公交一卡通(IC卡)均实行匿名制,虽然AFC系统能记录乘客每一次出行的时间及地点,但无法获取乘客个体属性(如职业、年龄、家庭住址)及出行属性(如出行目的),导致难以深度挖掘乘客出行特征.自动售检票数据在统计粒度、准确性、覆盖范围方面具有良好优势,能作为交通规划及管理的辅助数据源,对其深度挖掘具有重要实际意义.

近年来,基于AFC数据挖掘的交通规划及运营管理已成为国际研究热点.Pelletier综述了AFC数据在交通领域的前期研究成果^[1],分为3个层次:1)战略层,包括线网规划^[2]、出行行为^[3]以及需求预测^[4-5]等;2)战术层,包括运营计划调整^[6]、出行模式^[7]等;3)运营管理层,包括运营状态评价、服务水平评估^[8]等.国内研究主要集中于票款清分^[9-10]、客流分配^[11]、需求预测^[12]等方面,相比国外存在较大差距.前期主要通过交通调查(出行面板调查)来获取出行目的信息,该方法实施成本高、采集样本小.随着智能检测设备、全球定位系统(GPS)等先进设备在交通领域的应用,越来越多的自动采集信息得以获取,如GPS、自动车辆定位(Automatic Vehicle Location, AVL)、自动客流统计(Automatic Passenger Count, APC)、AFC、手机信令数据等.部分学者尝试利用这类新型数据对居民出行目的进行辨识,从而更好地理解需求特征.前期辨识方法主要分为两类:1)基于规则的辨识法,从出发时间、活动持续时间、土地利用性质、兴趣点(Point of Information, POI)等构建特定规则来辨识出行目的,如文献[13-14]利用GPS数据来辨识居民出行目的;2)机器学习法,如文献[15]利用马尔科夫模型对AFC数据进行目的辨识,文献[16]利用随机森林法对GPS数据进行出行目的辨识.对比而言,基于规则的辨识法含义明确、计算简便;而机器学习法在理论上较为完善,但计算过程复杂,难以适用于海量实际数据.考虑到方法对大规模轨道交通系统的适用性,本研究构建方法属于前者.不同采集数据在时间、空间粒度上存在较大差异,如GPS、手机信令数据较为精细,AFC数据较为粗略,因此,需结合特定数据类

型及对应居民的出行特征来构建恰当的辨识方法.

本文作者以城市轨道交通乘客为研究对象,利用AFC数据潜在包含的时空关联关系构建基于规则的乘客出行目的辨识方法.由于AFC数据包含的乘客出行信息十分有限,本研究充分结合轨道交通乘客出行时空稳定性强及通勤客流主体特征,重点针对工作日通勤客流(即出行目的为上班、上学、下班回家、放学回家)进行辨识.该研究为乘客出行行为、需求特征研究提供了一种低成本、高效率的分析方法.

1 问题描述

1.1 辨识过程及辨识要素

假设一天内某居民仅采用公共交通(包含公交和地铁)出行,将其IC卡对应的全部AFC记录按照时间顺序串联起来,即可得到该居民完整的出行序列(类似出行链).若出行序列中相邻两次出行不属于换乘时,则认为一次活动(如上班、购物)产生,出行目的辨识即对该活动类型进行判别.

以简单示例来阐述出行目的辨识过程.假设某乘客出行次序为序列1,见图1,其首次出行的出发时间位于早高峰,末次出行的出发时间位于晚高峰,两次出行间隔时间大于8 h,且首次出行的终点站与末次出行的起点站较为接近,那么,可初步判断该乘客为通勤乘客,首次出行目的为上班,末次出行目的为下班回家.

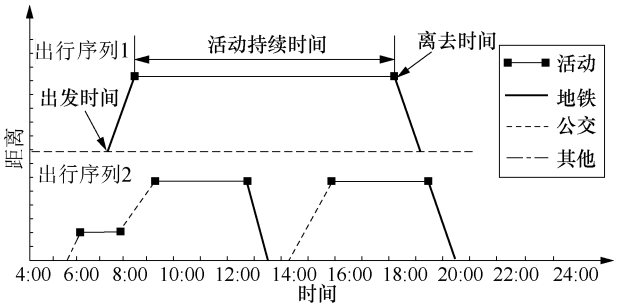


图 1 居民出行次序示意图

Fig.1 Schematic diagram of locators trip sequences

出行目的辨识过程即利用出行记录包含的时空关系(如出发时间、活动持续时间)来间接推断出行活动类型.出行活动影响因素可概括为内在因素和外在因素,内在因素包括个体属性(如职业、收入)、

家庭属性、出行习惯等;外在因素包括城市形态(购物场所、工作场所、家庭住址)、政策制度(上下班时间)等.结合 AFC 数据可获取的直接信息和潜在信息,将出行目的辨识要素分为以下四方面:

1)时间:时间是出行活动辨识的关键要素,包括出发时间、活动持续时间、离去时间;一般通勤乘客出发时间位于早高峰,活动持续时间较长,而娱乐、购物等活动的出发时间较为随机,活动持续时间较短.

2)空间:空间因素决定了出行活动的起讫位置(OD),例如工作地、家庭住址、学校位置直接决定了上班、上学出行的起讫点.

3)个体属性:个体属性是影响出行活动的重要因素,然而 AFC 记录不包括个体属性信息.以乘客所持 IC 卡类型来表征个体属性;IC 卡类型包括:储值卡(成人卡)、单程卡、学生卡、福利卡(老年卡等)、车站工作卡、员工卡 6 类.

4)出行规律:一定时期内居民在出行方式选择、出行 OD、出行频率等方面具有规律性,如通勤乘客工作日出行频率高、出行 OD 较为固定,根据出行规律可有效区分常乘客和普通乘客.

AFC 数据仅包含居民出行的部分信息(不包含常规公交、出租车等出行信息),难以构建居民的完整出行链,使得辨识过程面临诸多难点.涉及关键问题包括:1)出行连续性判断,如何确定两次相邻出行属于连续出行,即两次相邻出行间仅包含一次活动,以图 1 中的出行序列 2 为例,两次轨道交通出行不属于连续出行;2)辨识规则构建,如何构建恰当规则来辨识活动类型,某些活动(如购物、聚餐、访友等)在时空上无明显特征,如何有效辨识;3)单次出行辨识,若一天内居民仅利用轨道交通出行 1 次,信息严重缺失条件下如何进行辨识.

1.2 轨道交通乘客出行特征

基于 AFC 数据统计结果对轨道交通乘客出行特征进行分析,从而寻找解决关键问题的突破口.以 2014 年 11 月北京市轨道交通连续 5 个工作日 AFC 记录为基础数据,首先从乘客出行次数、相邻出行时间间隔、空间距离等方面分析.考虑到单程卡无法对应固定乘客以及车站工作卡的特殊性,分析时不包含单程卡和车站工作卡对应的 AFC 记录.据统计,一日内单程票对应 AFC 记录占比 13.71%,车站工作卡对应 AFC 记录占比 1.18%,该部分记录一般不对应通勤乘客出行,因此剔除该部分记录对通勤客流辨识结果不会造成影响.图 2 给出了不同出行次数对应的乘客数占比情况,图 3 给出了不同出行

次数对应的出行量占比情况.

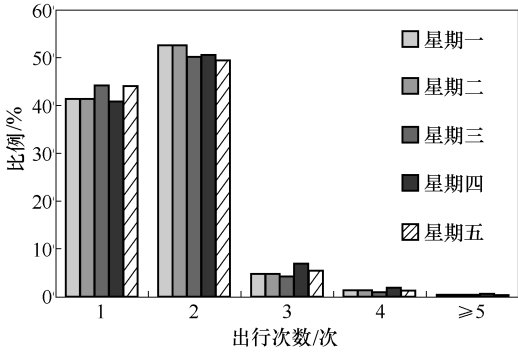


图 2 乘客出行次数统计

Fig.2 Count of passenger's travel times

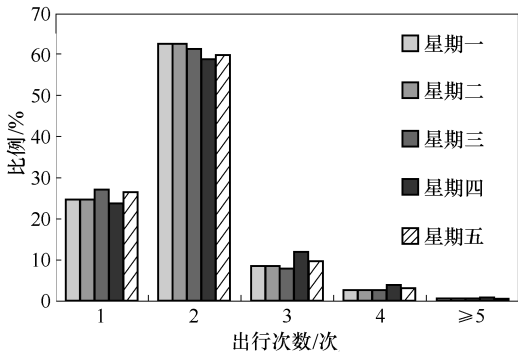


图 3 不同出行次数对应出行量统计

Fig. 3 Trip counts with different travel times

由图 2 可以看出:一天内居民利用轨道交通出行次数集中在 1 和 2 次,且 2 次出行乘客的比例大于 1 次;以工作日平均来看,1 次出行占约 42%,2 次出行占约 51%,超过 2 次出行的乘客数仅占 7%.因此,出行目的辨识主要针对 1 次和 2 次出行乘客.出行量为乘客数与出行次数的乘积,在此需要说明的是,出行量与官方发布的客运量存在差异,客运量考虑了换乘因素,客运量为出行量与日均换乘系数(一般为 1.5~2.0)的乘积.由图 3 可以看出:2 次出行对应的出行量占比最大,工作日均值达 62%,不难推测该部分客流主要为通勤乘客(出行过程为早晚往返出行);1 次出行对应的出行量占比约为 25%,也占据一定比例.

下面对关键问题 1)分析.文献[4]在需求预测研究中提出重要假设:乘客下一次出行起始站紧邻前一次出行的终点站,调查显示超过 90%的乘客满足该假设.在此,对北京市轨道交通乘客进行实证分析.由于工作日出行具有很强相似性,仅以周三为特征日进行分析.定义相邻出行站间距为前一次出行终点站和后一次出行起点站间距离,统计时以相邻两次出行为基本单位,对相邻出行站间距进行频次统计,见图 4. 由图 4 可以看出:1)大部分轨道交通

乘客出行满足下一次出行起始站等于上一次出行终点站,该比例约为 70%; 2) 假设乘客可接受步行距离范围为 1 km (平均站距), 那么约 87% 的乘客出行满足该假设. 实际中, 同一 OD 出行乘客在车站选择时具有多种方案, 其起点或终点距多个车站空间距离相近. 考虑到该因素, 出行连续性判别时设定距离阈值(1 km)来优化判定过程, 使其更加符合实际情况. 综上分析, 验证了该假设对北京市轨道交通乘客的合理性, 在进行出行活动辨识时可通过相邻出行站间距来判断相邻出行是否为连续出行.

构建恰当规则进行出行目的辨识是另一关键问题, 在此对时间要素中的活动持续时间进行分析. 定义相邻出行时间间隔为下一次出行出发时间与上一次出行到达时间之间的差值, 该时间与活动持续时间基本相同(忽略活动地点与下车站间时间影响). 假定相邻出行站间距小于等于 1 km 时为连续出行, 以 0.5 h 为时间粒度对连续出行间隔时间统计, 见图 5. 可以看出: 1) 时间间隔曲线存在明显峰值, 最高峰值位于 9.5~10.0 h 范围, 不难推测处于该范围的乘客为通勤乘客, 其活动持续时间与工作时间保持一致; 2) 时间间隔超过 5 h 的出行比例高达 63%; 3) 在 1.0~1.5 h 范围内出现了另一较小峰值, 该部分主要对应日常生活类出行乘客. 结合时间间隔分布特征及居民日常活动规律, 可认为活动持续时间能够表征活动类型.

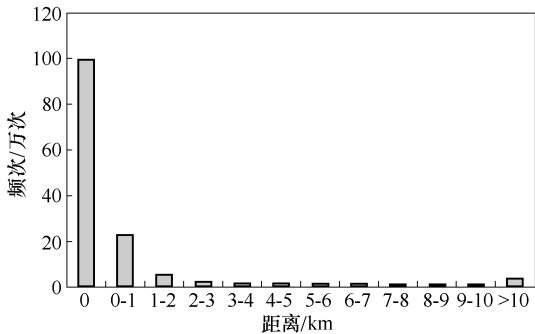


图 4 相邻出行间站间距分布

Fig.4 Distances distribution within two continuous trips

下面对乘客出行规律进一步分析. 定义工作日出行频率和最大 OD 出行频率来表征乘客出行时空稳定性. 工作日出行频率指工作日出行天数与工作日之比

$$\lambda = n / 5 \tag{1}$$

式中: λ 表示工作日出行频率; n 为工作日出行天数.

最大 OD 指一定时期内乘客出行最多的 OD, 最大 OD 出行频率表示一定时期内乘客最大 OD 出

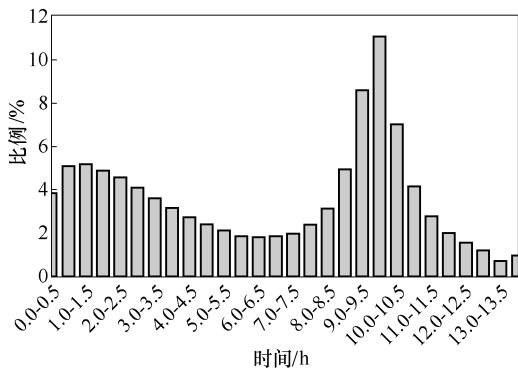


图 5 活动持续时间分布

Fig.5 Activity time distribution

行次数与总出行次数之比

$$\alpha_{\max} = \max\{n_{od}\} / N \tag{2}$$

式中: α_{\max} 表示最大 OD 出行频率; n_{od} 表示某特定 OD 对应的出行次数; N 为总出行次数. α_{\max} 越大则乘客出行 OD 越集中, $\alpha_{\max} \in (0, 1]$. 考虑到通勤乘客往返出行特点, 将对称 OD 视为同一 OD, 从而更好地刻画乘客出行的空间稳定性. 若某乘客每日往返家庭和工作地之间, 其最大 OD 出行频率 $\alpha_{\max} = 1$. 另外, 需要注意的是乘客在同一 OD 间出行(如家—工作地), 可能具有多种车站选择方案, 虽然出行车站不同但实为同一 OD, 为此, 设定距离阈值(1 km)来判别此类相近 OD, 将其视为同一 OD.

后续进行出行目的辨识时, 工作日出行频率及最大 OD 出行频率将作为关键参数, 为得到更为稳定的参数取值, 分析时段范围扩展至 2014 年 11 月一整月, 分别对每周 AFC 数据进行统计. 图 6 给出工作日出行频率所对应乘客量占比情况. 可以看出: 1) 工作日出行频率为 0.2 和 1.0 时达到峰值(对应出行 1 日和 5 日), 出行 1 日的乘客占比约为 30%, 该部分乘客较少利用轨道交通出行, 而出行 5 日的乘客为常乘客, 可初步判断为通勤乘客; 2) 对于其他乘客则难以判断乘客类型, 当工作日出行频率越高, 其属于通勤乘客的可能性越大.

图 7 给出最大 OD 出行频率所对应乘客量占比情况. 图 7 看出: 1) 连续四周统计得到的 OD 出行频率分布趋势保持一致, 最大 OD 出行频率在 0.40~0.50 以及 0.70~0.90 范围内达到峰值, 而在 0.60 附近明显降低, 可认为 0.60 为区分两类不同乘客在出行空间稳定上的分界点; 2) 若将所有乘客的最大 OD 出行频率按照升序进行排序, 50% 分位处对应的频率为 0.58, 70% 分位处对应频率为 0.72 (取值为四周均值).

1.3 出行活动类型分析

出行活动一般分为上班、上学、购物、娱乐、业

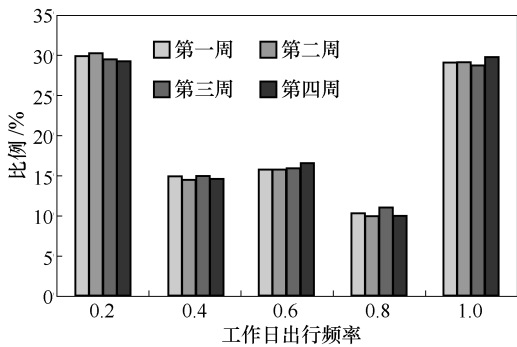


图 6 工作日出行频率统计

Fig.6 Travel frequency of weekdays

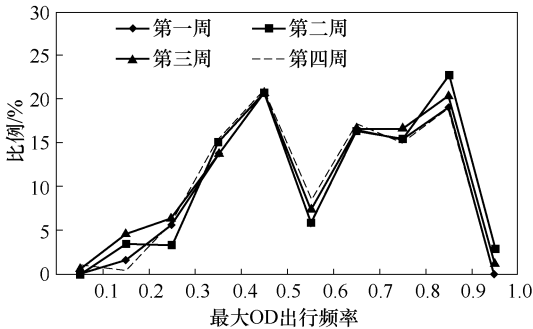


图 7 最大 OD 出行频率统计

Fig.7 Travel frequency of maximum OD

务、接送人、回家、其他 8 类.部分出行活动在时空间上无明显特征,如业务、购物和娱乐,导致难以通过自动采集数据自身包含的出行时空关系进行辨识.既有研究常见的解决办法是结合出行特征对出行目的予以简化,如文献[18]对上班、上学、其他 3 类出行活动进行辨识.考虑到轨道交通客流主体为通勤乘客,其他出行类型占总出行量比重较小.因此,本研究重点仅针对通勤客流(上班、上学、下班回家、放学回家)进行辨识.结合辨识要素对不同出行活动类型的特征进行分析,见表 1.

表 1 不同类型出行活动特征

Tab.1 Characteristics of different travel activities

类型	时间	空间	卡类型	出行规律
上班	出发时间一般为上午;活动持续时间较长;一般为首次出行	出行 OD 较为固定	储值卡;员工卡	时空稳定性强
上学	出发时间为上午;活动持续时间较长	出行 OD 较为固定	学生卡	时空稳定性强
回家	出发时间一般为下午;一般为末次出行	目的地为家	不限	其中通勤乘客、学生时空稳定性强
其他	不限	出行 OD 随机	不限	时空稳定性弱

2 基础数据及预处理

不同城市轨道交通系统可能采用不同类型的

AFC 系统以及票价体系,导致 AFC 记录包含的出行信息略有差异.本研究仅涉及乘客出行基本信息,所有类型 AFC 记录均包含.AFC 记录基础数据结构如表 2 所示.

表 2 AFC 记录基础数据结构

Tab.2 Basic data structure of AFC records

字段名称	数据类型	含义
CARD_CODE	字符型	IC 卡编号
CARD_TYPE	整型	IC 卡类型
TRIP_START_TIME	日期时间型	进站刷卡时间
START_LOCATION	字符型	进站车站编码
TRIP_END_TIME	日期时间型	出站刷卡时间
END_LOCATION	字符型	出站车站编码

数据预处理目的在于提高基础数据准确性及有效性.本研究中数据预处理包含两方面:1)有效性筛选,数据传输错误以及设备故障等会造成部分信息缺失或错误,以及部分信息对本研究无实际意义,剔除该类数据;2)数据转换,城市轨道交通系统每天将产生数百万甚至上千万条 AFC 交易记录,预处理能有效提高算法对海量数据的处理能力.

有效性判断主要包括:1)卡类型有效性判断,单程卡以及车站功能卡所对应 AFC 记录为无效记录;2)出行时间有效性判断,出站时间须晚于进站时间,行程时间处于合理范围,即行程时间不大于路网最远 OD 间行程耗时(本文设为 3 h),同时不小于基本行程时间(本文设为 10 min);3)出行位置有效性判断,乘客起始站与终点站不能为同一车站;4)完整性判断,AFC 记录包含完整的进出站时间及位置信息.

数据转换主要是获取同一 IC 卡的全部出行记录,以卡号为标识构建新的数据结构.定义 r_m^k 为卡号 m 对应乘客的第 k 次出行记录, $r_m^k = \langle p_m, t_m^k, o_m^k, t_m^{k'}, d_m^k \rangle$, 其中 p_m 为卡类型, t_m^k 和 $t_m^{k'}$ 分别表示进站时间和出站时间, o_m^k 和 d_m^k 分别表示起始站和终点站名称.提取某 IC 卡一天内全部出行记录,按进站时间排序,得到日出行序列.预处理后数据结构如表 3 所示.

表 3 预处理后数据结构示例

Tab.3 Examples of data structure after processing

卡号	日期	出行次数	卡类型	出行记录集合
40729285	2014-11-01	2	储值卡	(r_m^1, r_m^2)
40729323	2014-11-01	1	储值卡	(r_m^1)
40730187	2014-11-01	n	学生卡	$(r_m^1, r_m^2, \dots, r_m^n)$
...

3 出行目的辨识算法

结合 AFC 数据可获取的直接和间接出行信息,

选取相关指标来表征辨识要素,见表 4.其中,最大出行 OD、工作日出行频率、最大 OD 出行频率需利用多日(至少连续 5 个工作日)AFC 数据统计获取,家庭所在区域采用与“家”紧邻的车站表示,表示“家”处于该车站吸引范围内,一般“家”所紧邻的车站为乘客使用最多的车站,采用前期研究^[18]所构建方法进行辨识.前述分析可知,一天内居民使用轨道交通出行次数集中在 2 次和 1 次,为此,从多次和单次出行两方面对出行目的辨识规则进行构建.

表 4 出行活动辨识指标

Tab.4 Identification indexes for travel activities

辨识要素	表征指标
时间	出发时间、活动持续时间、 出行次序(首次或末次出行)
空间	最大出行 OD、家庭住址(所在区域)
个体属性	IC 卡类型(储值卡、学生卡、福利卡、员工卡)
出行规律	工作日出行频率、最大 OD 出行频率

3.1 多次出行辨识规则

为保证相邻出行间仅包含一次活动,即相邻两

表 5 多次连续出行辨识规则

Tab.5 Identification rules for continuous trips

出行目的	时间	空间	出行规律	个体属性
上班	活动持续时间>5 h;首次出行	OD=最大出行 OD	$\lambda \geq 0.4; \alpha_{\max} \geq 0.58$	储值卡、员工卡
上学	活动持续时间>5 h;首次出行	OD=最大出行 OD	$\lambda \geq 0.4; \alpha_{\max} \geq 0.58$	学生卡
下班回家	末次出行	D=“家”;OD=最大出行 OD	$\lambda \geq 0.4; \alpha_{\max} \geq 0.58$	储值卡、员工卡
放学回家	末次出行	D=“家”;OD=最大出行 OD	$\lambda \geq 0.4; \alpha_{\max} \geq 0.58$	学生卡
其他回家	末次出行	D=“家”	不限	不限
其他	不限	不限	不限	不限

3.2 单次出行辨识规则

部分研究中将单次出行记录剔除^[17],然而前述分析(图 2 及图 3)可知,单次出行所对应出行量占全日出行量比值约为 25%,占据较大比例,为此制定相应规则来对单次出行乘客的出行目的进行辨

表 6 单次出行目的辨识规则

Tab.6 Identification rules for one time traveling

出行目的	时间	空间	出行规律	个体属性
上班	出发时间<12:00	OD=最大出行 OD	$\lambda \geq 0.6; \alpha_{\max} \geq 0.72$	储值卡、员工卡
上学	出发时间<12:00	OD=最大出行 OD	$\lambda \geq 0.6; \alpha_{\max} \geq 0.72$	学生卡
下班回家	出发时间 ≥ 12:00	D=“家”;OD=最大出行 OD	$\lambda \geq 0.6; \alpha_{\max} \geq 0.72$	储值卡、员工卡
放学回家	出发时间 ≥ 12:00	D=“家”;OD=最大出行 OD	$\lambda \geq 0.6; \alpha_{\max} \geq 0.72$	学生卡
其他回家	出发时间 ≥ 12:00	D=“家”	不限	不限
其他	不限	不限	不限	不限

首先,以 12:00 为分界点来对出发时间进行限制,假定出发时间小于 12:00 的出行为首次出行,对应出行目的可能为上班、上学、其他,晚于 12:00 的出行为末次出行,对应出行目的可能为回家或其他.其次,在出行规律指标方面设置更为严格的限制,工作日出行频率 $\lambda \geq 0.6$,最大 OD 出行频率采用 70%

次出行为连续出行,以相邻出行站间距为约束条件进行判断

$$l(D_m^k,O_m^{k+1}) \leq \Delta l \tag{3}$$

式中: $l(D_m^k,O_m^{k+1})$ 表示乘客 m 第 k 次出行终点站与第 $k+1$ 次出行起点站间距离, $m;\Delta l$ 为距离阈值,结合乘客可接受步行距离以及实际出行记录统计结果(图 4),本文 $\Delta l=1\ 000\text{ m}$.

当相邻出行站间距不满足距离约束时,则认为不止一次出行活动产生,此时将两次出行分别视为单次出行(见 3.2 节)进行辨识.表 5 给出了相邻连续出行的辨识规则.

考虑到连续出行乘客能够利用活动持续时间这一关键要素来表征活动类型,因此,出行规律指标仅设置基本的判定条件,对于出行较为稳定的通勤乘客(上班、上学)设 $\lambda \geq 0.4, \alpha_{\max} \geq 0.58$,即乘客工作日出行天数需大于等于 2 天,最大 OD 出行频率大于等于 0.58(50%分位点).

识.由于单次出行无法获取出行活动的时间信息(如活动持续时间、首末次出行),给辨识带来极大困难.辨识规则的构建思路是利用更为严格的出行规律指标以及时空间指标来间接判断,见表 6.

分位处的出行频率表示($\alpha_{\max} \geq 0.72$).结合图 2 和图 6 看出,一天内约 51%的乘客为 2 次出行,以及 30%的乘客工作日出行天数为 5 日,因此,可初步判断约 30%的乘客为通勤乘客,故取 70%分位的最大 OD 出行频率作为出行规律指标的取值.该方法能够对出行较为稳定的乘客进行辨识,但对某些出行

则无法辨识,例如上午回家乘客、下午上班乘客(夜班)进行辨识,考虑到该部分乘客占比很小,对辨识结果不会产生显著影响.

3.3 出行目的辨识算法

图 8 给出了基于规则的出行目的辨识算法.结合前述辨识规则即可对乘客出行目的予以辨识,其中,对于多次出行条件下的非连续出行,虽采用单次出行规则辨识,但其出行次序(首末次)可直接获取,无须通过出行时间来确定.

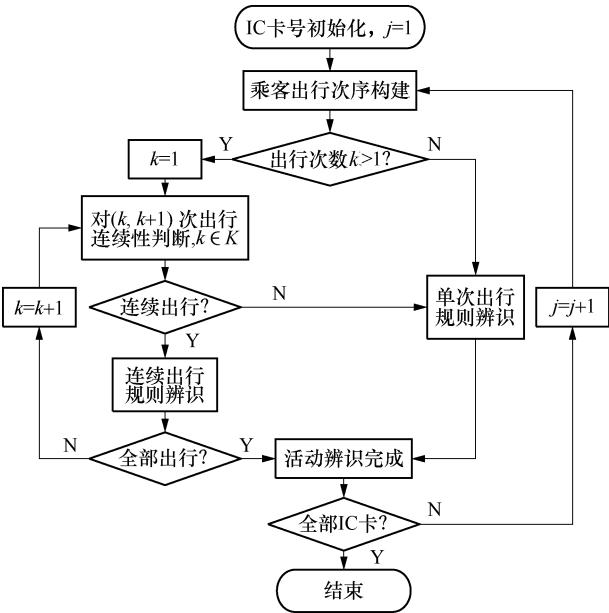


图 8 基于规则的出行目的辨识算法

Fig.8 Rule-based trip purpose identification algorithm

4 实证分析

4.1 辨识结果分析

以 2014 年 11 月北京市轨道交通为对象进行实证分析,路网共包含 17 条线路,281 座车站.客流数据为连续 5 日 AFC 记录.通过连续 5 日客流数据确定每一位乘客的工作日出行频率、最大 OD 出行频率、最大出行 OD,并对其家庭住址所在区域(紧邻车站)进行辨识.利用所构建算法对星期三全日内 AFC 记录进行出行目的的辨识,其中有效 AFC 记录约为 370 万条,IC 卡使用数约为 226 万张,辨识结果如表 7 所示.可以看出:1)上班和下班回家出行占全日出行量的 53.03%,与轨道交通通勤客流主体特征一致;2)上学和放学回家出行量仅占 0.99%,表明学生利用轨道交通上下学较少,更多的选择其他出行方式;3)其他出行(如购物、娱乐等)在全日内占据一定比例,达 28.30%.

根据 2014 年北京市居民公共交通出行目的调查,上下班出行占总出行比例为 50.70%,上下学为

6.4%,公务外出为 1.4%,生活类(购物等)为 22.50%,其他回家为 19.00%^[19].虽然调查对象不局限于轨道交通,但对辨识结果仍具有一定说服力.可以看出:通勤出行(包含上下班和上下学)是公共交通出行中的主要部分.辨识结果显示轨道交通上下班出行占 53.03%,调查显示上下班占 50.70%,两者较为接近.主要差异在于轨道交通上下学出行仅占 0.99%;调查显示利用公共交通上下学占 6.40%.上下学一般为短距离出行,更多选择常规公交、私家车等出行,实际中较少采用轨道交通出行,因此,辨识结果与实际出行情况一致.对比辨识结果与调查结果,表明辨识结果与实际具有较高吻合度.

表 7 乘客出行目的辨识结果

Tab.7 Identification results of traveler purpose

出行目的	出行量/(人次)	百分比/%
上班	992 348	26.77
上学	16 425	0.44
下班回家	973 597	26.26
放学回家	20 540	0.55
其他回家	655 281	17.68
其他出行	1049 049	28.30

4.2 不同出行目的乘客特征

出行目的的辨识意义在于能够对不同出行活动乘客的出行行为进行挖掘,进而制定针对性的运营管理与客运营销策略.下面对不同出行目的的乘客的出行时空特征进行简要分析.

图 9 给出了不同出行目的的乘客的行程时间分布.可以看出:1)从总体来看不同出行目的的乘客的行程时间分布总体趋势一致,约在 30~50 min 出现高峰,行程时间大于 80 min 的出行仅占约 10%;2)对比不同出行目的的出行,其他出行对应的行程时间分布最为平缓,而上班出行对应的行程时间最为集中,表明工作出行平均花费的时间要少于其他出行.

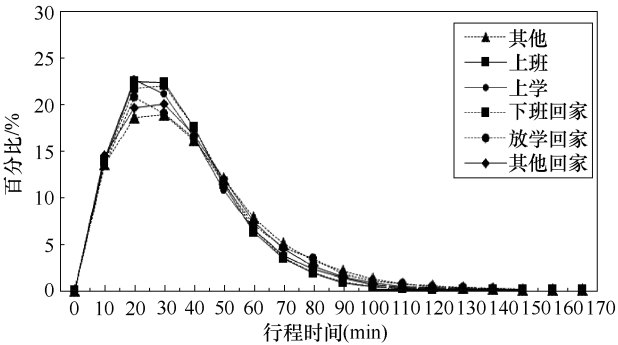


图 9 不同出行目的的乘客行程时间分布

Fig.9 Travel cost time distribution for different trip purposes

图 10 给出了不同出行目的的乘客的出发时间分

布.可以看出:1)上班和下班乘客在出发时间分布上存在明显峰值,集中于高峰时段(07:00~09:00 和 17:00~19:00),而其他出行乘客的出发时间分布较为均匀;2)对比上班和下班乘客的分布曲线,上班乘客的出发时间分布更为集中、且峰值更高,这是造成早高峰拥挤强于晚高峰拥挤的主要原因;3)早高峰时段中上班乘客约占总出行量的 75%,仍有约 25% 的乘客为非通勤乘客,晚高峰时段该比例更高,因此,制定相关政策来分散高峰时段其他出行目的的乘客是缓解拥挤的可行措施,存在较大的实施空间.

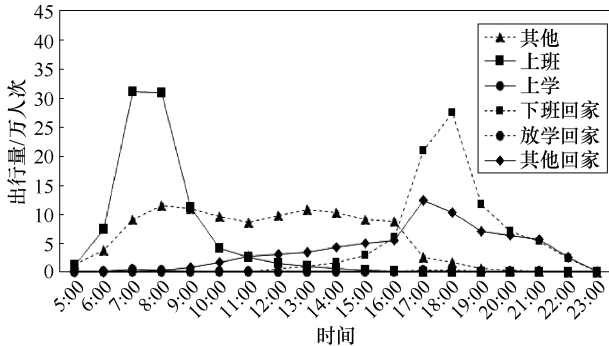


图 10 不同出行目的乘客出发时间分布

Fig.10 Departure time distribution for different trip purposes

5 结论

1)利用轨道交通海量 AFC 数据自身包含的出行时空关联关系,构建基于规则的乘客出行目的辨识算法,为精细化客流特征及出行行为分析提供了方法,扩充了 AFC 数据的应用范围和价值.

2)以 2014 年 11 月北京市轨道交通连续 5 个工作日 AFC 数据为对象进行实证分析,结果显示所构建方法能够有效对上班、上学、下班回家、放学回家、其他回家和其他 6 类出行进行辨识,结合北京市公共交通调查结果,验证了结果的准确性与合理性.

3)AFC 数据及其他自动采集数据能够为交通规划、运营、管理提供充足的基础数据,充分挖掘该类数据潜在价值具有重要实际意义,后期将结合辨识结果对不同出行目的乘客的特征进行深度挖掘.

参考文献 (References):

- [1] PELLETIER M P, TRÉPANIÉ M, MORENCY C. Smart card data use in public transit: a literature review [J]. *Transportation Research Part C: Emerging Technologies*, 2011, 19(4): 557—568.
- [2] PARK J Y, KIM D J, LIM Y. The potential of using the smart card data to define the use of public transit in Seoul [J]. *Transportation Research Record: Journal of*

- the Transportation Research Board*, 2008 (2063): 3—9.
- [3] MORENCY C, TRÉPANIÉ M, AGARD B. Measuring transit use variability with smart-card data [J]. *Transport Policy*, 2007, 14(3): 193—203.
- [4] BARRY J J, NEWHOUSER R, RAHBEE A. Origin and destination estimation in New York City with automated fare system data [J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2002(1817): 183—187.
- [5] ZHAO J, RAHBEE A, WILSON N H M. Estimating a rail passenger trip origin-destination matrix using automatic data collection systems [J]. *Computer-Aided Civil and Infrastructure Engineering*, 2007, 22(5): 376—387.
- [6] UTSUNOMIYA M, ATTANUCCI J, WILSON N H. Potential uses of transit smart card registration and transaction data to improve transit planning [J]. *Journal of the Transportation Research Board*, 2006 (171): 119—126.
- [7] LEE S, HICKMAN M. Travel pattern analysis using smart card data of regular users [C]// *Transportation Research Board 90th Annual Meeting*, 2011: 4258.
- [8] REDDY A, LU A, KUMAR S, et al. Entry-only automated fare-collection system data used to infer ridership, rider destinations, unlinked trips, and passenger miles [J]. *Journal of the Transportation Research Board*, 2009 (2110): 128—136.
- [9] 于江波, 王浙君. 基于出行时间匹配的 ACC 清分模型的研究 [J]. *城市轨道交通研究*, 2013, 16(1): 43—45.
YU Jiangbo, WANG Zhejun. On ACC clearing model based on travel matching time [J]. *Urban Mass Transit*, 2013, 16(1): 43—45. (in Chinese)
- [10] 黄荣, 蒋玉琨, 刘智丽, 等. 基于共线运营的城市轨道交通线网票款清分算法及实施 [J]. *交通运输系统工程与信息*, 2010, 10(2): 130—135.
HUANG Rong, JIANG Yukun, LIU Zhili, et al. Algorithm and implementation of urban rail transit network based on joint operation [J]. *Journal of Transportation Systems Engineering and Information Technology*, 2010, 10(2): 130—135. (in Chinese)
- [11] 孙延硕, 徐瑞华. 基于 AFC 数据的城市轨道交通乘客出行路径推算 [J]. *交通与运输(学术版)*, 2011(2): 85—90.
SUN Yanshuo, XU Ruihua. Estimation of rail transit passenger route using automated fare collected data [J]. *Traffic and Transportation (Academic)*, 2011(2): 85—90. (in Chinese)
- [12] 蔡昌俊, 姚恩建, 张永生, 等. 基于 AFC 数据的城轨站间客流量分布预测 [J]. *中国铁道科学*, 2015, 36(1):

126—132.

CAI Changyong, YAO Enjian, ZHANG Yongsheng, et al. Prediction of passenger flow distribution at urban rail stations based on AFC data[J]. China Railway Science, 2015, 36(1): 126—132. (in Chinese)

[13] 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: Journal of the Transportation Research Board, 2001, 1768(1): 125—134.

[14] STOPHER P, FITZGERALD C, ZHANG J. Search for a global positioning system device to measure person travel [J]. Transportation Research Part C: Emerging Technologies, 2008, 16(3): 350—369.

[15] HAN G, SOHN K. Activity imputation for trip-chains elicited from smart-card data using a continuous hidden Markov model [J]. Transportation Research Part B: Methodological, 2016, 83: 121—135.

[16] MONTINI L, RIESER-SCHÜSSLER N, HORNI A, et al. Trip purpose identification from GPS tracks[J]. Transportation Research Record: Journal of the Transportation Research Board, 2014(2405): 16—23.

[17] LEE S G, HICKMAN M. Trip purpose inference using automated fare collection data [J]. Public Transport, 2014, 6(1/2): 1—20.

[18] 姚向明, 赵鹏, 韩宝明, 等. 基于售检票数据挖掘的轨道交通乘客居住区辨识[J]. 交通运输系统工程与信息, 2016, 16(5): 233—240.

YAO Xiangming, ZHAO Peng, HAN Baoming, et al. Home district identification for urban rail transit travelers by mining automatic fare collection data[J]. Journal of Transportation Systems Engineering and Information Technology, 2016, 16(5): 233—240. (in Chinese)

[19] 北京市交通发展研究中心. 2015 年北京交通发展年度报告[R]. 北京: 北京市交通发展研究中心, 2015.

Beijing Transportation Research Center. Beijing transport annual report in 2015 [R]. Beijing: Beijing Transportation Research Center, 2015. (in Chinese)

(上接第 43 页)

GUO Zhan, SHANG Xiaolei, LI Hai. Model of rail transit safety evaluation system based on AHP [J]. Journal of China Railway Sciences, 2011, 32(3): 123—125. (in Chinese)

[12] 冯运卿, 李雪梅, 李学伟. 基于熵权法与灰色关联分析的铁路安全综合评价[J]. 安全与环境学报, 2014, 14(2): 38—49.

FENG Yunqing, LI Xuemei, LI Xuewei. Comprehensive evaluation of railway safety based on entropy weight method and grey relation analysis [J]. Journal of China Railway Sciences, 2014, 14(2): 38—49. (in Chinese)

[13] 沈鸿飞, 贾利民. 区域路网交通安全风险分析及管控策略研究[J]. 交通运输系统工程与信息, 2011, 11(5): 187—192.

SHEN Hongfei, JIA Limin. Regional road network traffic safety risk analysis and control strategies [J]. Journal of Transportation System Engineering and Information Technology, 2011, 11(5): 187—192. (in Chinese)

[14] 于少伟. 基于区间数的模糊隶属函数构建[J]. 山东大学学报(工学版), 2010, 40(6): 32—35.

YU Shaowei. Construction of a fuzzy membership function based on interval number [J]. Journal of Shandong University (Engineering Science), 2010, 40(6): 32—35. (in Chinese)

[15] 孙莹, 鲍新中. 一种基于方差最大化的组合赋权评价方法及其应用[J]. 中国管理科学, 2011, 19(6): 141—148.

SUN Ying, BAO Xinzhong. A method of combination weighting evaluation based on variance maximization and its application [J]. Chinese Journal of Management Science, 2011, 19(6): 141—148. (in Chinese)