

基于北京公交刷卡数据和兴趣点的功能区识别

IDENTIFYING URBAN FUNCTIONAL ZONES USING BUS SMART CARD DATA AND POINTS OF INTEREST IN BEIJING

韩昊英 于翔 龙瀛
HAN Haoying; YU Xiang; LONG Ying

【摘要】城市在其发展过程中逐渐形成居住区、工业区和商业区等不同的功能区。识别这些功能区并理解其分布特征，对于把握城市结构以及制定和使用科学合理的规划具有重要作用。本研究基于2008年4月北京市连续一周的7797余万条公交IC卡刷卡数据，将其转换为每个公交站台流量的二维时间序列数据，结合居民日常出行行为研究，利用数据挖掘技术，构建了基于公交刷卡数据和兴趣点的城市功能区识别模型，并将识别结果在交通分析小区尺度上汇总。研究结果显示，利用城市功能区识别模型，通过冗余数据的筛选和特征的创建实现对数据的有效降维，并选用期望最大化算法对处理后的数据进行聚类分析，结合居民日常出行相关特征和兴趣点分布数据对聚类结果进行诠释，可以快速有效地识别出与北京市土地利用现状地图具有一定匹配度的北京市各功能区。本研究的方法可以辅助规划人员和公众有效识别和理解复杂的城市空间结构，对城市地理及规划研究具有重要的理论和实践价值。

【关键词】公交IC卡刷卡数据；兴趣点；出行行为；功能区识别；北京

ABSTRACT: Cities form various functional zones including residential zone, industrial zone, commercial zone, etc, during their development process. It is important for urban planners to identify different functional zones and understand their spatial distribution characteristics in order to better comprehend city structure and formulate and use rational urban plans. In this research, we used 77,976,010 bus smart card data (SCD) records of Beijing City in one week in April 2008 and converted them into two-dimensional time series data of each bus platform. Then, by applying the data mining techniques in combination with citizens' daily travel behavior, we established the

DZoF (discovering zones of different functions) model based on SCD (smart card data) and POIs (points of interest), and pooled the results at the TAZ (traffic analysis zone) level. The results suggest that DZoF model and cluster analysis based on dimensionality reduction and EM (expectation-maximization) algorithm can identify functional zones that well match the actual land uses in Beijing. The methodology in the present research can help urban planners and the public understand the complex urban spatial structure and contribute to the academia of urban geography and urban planning.
KEYWORDS: bus smart card data (SCD); POIs; travel behavior; identification of functional zones; Beijing

城市是人类社会发展到一定阶段的产物，是一个开放的复杂系统。城市中存在着多种结构，包括空间结构、社会结构、文化结构、经济结构、人口结构和功能结构等。其中，城市空间结构的研究是以往城市地理学及城市规划学研究城市空间的核心内容之一^[1]。不论从物质层面还是社会层面，城市都是一个复杂的空间系统，而基于微观对象的研究是理解这一复杂系统运作规律的重要途径^[2, 3]。传统的对于城市要素、组织、结构等方面的研究由于受到数据条件的限制，往往局限在大尺度上(如乡镇或交通分析小区)。而随着大数据时代的到来，遍布的个人贡献数据为描述和理解城市空间结构提供了新的渠道^[4]。

各研究领域所关注的大数据，主要是由移动通信(GSM)、全球定位系统(GPS)、社会化网络(SNS)和无线宽带热点(Wi-Fi)等基于位置服务(location based services, LBS)技术所提供的公交智能卡刷卡记录、航班记录、银行卡记录、微博记录和手机通话记录等。这些数据可以形成城市居民的出行日志^[5]，并被用来对城市活动进行

实时监控^[6],分析城市活动的强度和时空分布特征^[7],研究城市中人类活动模型(human mobility pattern)^[8]。

人类活动(human mobility)与城市空间结构之间有着密切的关系^[9,10]。现有研究中,学者主要通过城市空间结构辅助研究人类活动,分析人们出行特征,探讨城市空间结构对人的出行影响^[11]。例如,通过城市土地利用结构来研究城市通勤模型^[12],分析空间结构对于居民通勤行为的影响^[13,14]。与此相反,对于如何利用已有的城市中人类活动数据,来进行城市空间结构的有关研究则鲜有学者涉及,但是这却又是非常重要的。因为伴随着城市的发展,城市的土地利用与空间结构正快速地发生变化,城市由过去的单一中心模式向多中心模式发展^[15,16],即时且明确的城市功能区块划分能够启发城市规划者对城市未来的规划,并对以往的用地规划进行验证。而传统的对于城市土地利用及空间结构的研究往往是基于遥感数据^[17,18],这些数据价格昂贵而且不能及时更新,不能满足城市规划者和学者研究与应用的需要。所以,利用LBS技术提供的海量数据,分析人类活动,进而开展城市空间结构的研究,将逐渐成为未来城市研究的热点。

现有城市空间结构研究中,移动通讯(GSM)和全球定位系统(GPS)数据应用最为广泛。例如,研究者利用杭州市出租车行驶的GPS信息,分析了出租车乘客上下车特征与城市区域社会功能的关系^[19]。近些年来,作为一种大规模的具有地理标识和时间标签的公交IC卡刷卡数据(SCD),也逐渐被用于城市研究中。约和黄(Joh and Hwang)利用首尔大都市区1000万条的SCD数据,分析了公交卡持卡人的出行轨迹与市区土地利用的特征^[20];龙瀛等人利用北京市公交IC卡刷卡数据进行了北京职住关系及通勤流向分析^[21]。

本研究基于不同的城市功能区对应不同的用地类型与设施,这些不同的用地类型对应了不同的广义上的人类活动(如上班、上学、购物等),而不同的人类活动又对应着不同的交通时空特征。基于这一理论前提,本研究利用SCD和兴趣点的原始数据,建立城市功能区识别(discovering zones of different functions, DZoF)模型。在此模型中,构建公交站台(platform)流量数据模型,利用基于模型的期望最大化(expectation-maximization, EM)算法,对北京市8691个公交站台进行聚类分析,并基于传统的居民出行行为研究、居民的普遍认知、兴趣点数据模型建立SCD数据挖掘的模式识别规则,对聚类所得的簇(cluster)进行功能诠释。根据DZoF模型,本研究最终确定了北京市各个公交站台的功能,并在交

通分析小区(traffic analysis zone, TAZ)尺度上进行汇总,实现不同区域功能的识别。为了验证DZoF模型识别结果的有效性,研究结果还与北京市城市总体规划(2004-2020)的用地现状图及该地区谷歌地图进行对比分析。

1 研究区概况及数据

1.1 研究区概况

本研究的研究区域为北京市域,总面积为16410km²,常住人口2069.3万。北京市拥有四通八达的现代化、立体交通网络。2008年,全市公共电汽车线路共有648条,运营里程1.7万km,运营车辆2.06万辆;轨道交通线路为8条,运营里程达到200km;出租车运营车辆6.6万辆。

1.2 数据

本研究的数据主要为北京市2008年4月连续一周(4月7日至4月13日)的公交IC卡刷卡数据(该数据不包括轨道交通刷卡数据),其中共涉及超过600条公交线路(上下行计算共计1287条,其中一票制线路566条,分段计价线路721条),约3.7万个公交站点(stop)和8691个公交站台(platform)。图1为北京市公交站台(platform)的分布图。同时,为了研究的方便,北京市被划分1118个交通分析小区(traffic analysis zone, TAZ),如图2所示。

本研究所使用的公交IC卡刷卡数据中的每条

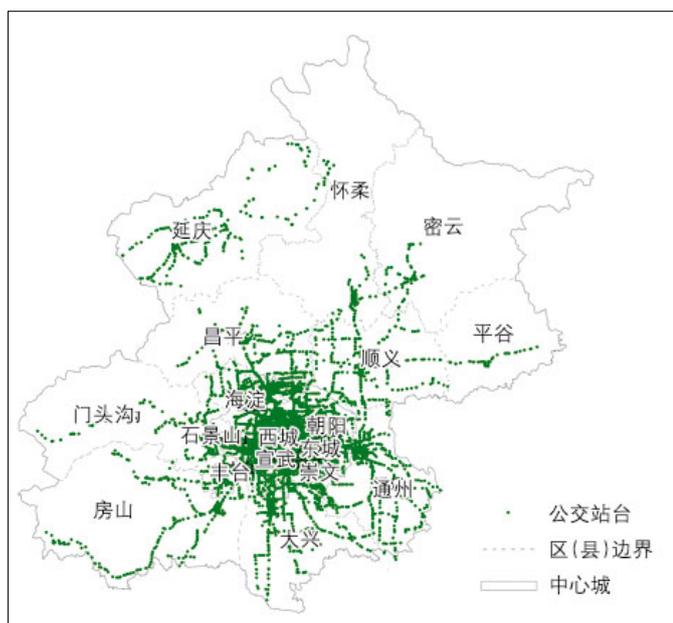


图1 北京市公交站台分布

Fig.1 Distribution of bus stations in Beijing Metropolitan Area (BMA)



图2 北京市交通分析小区分布
Fig.2 Traffic analysis zone of Beijing Metropolitan Area (BMA)

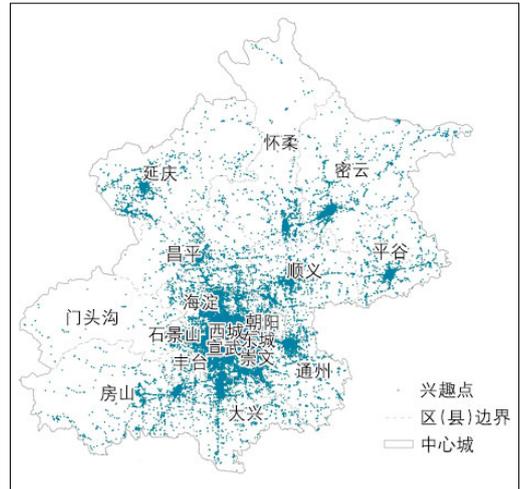


图3 2010年北京市兴趣点空间分布
Fig.3 Spatial distribution of POIs in Beijing Metropolitan Area (BMA) in 2010

记录主要包含以下基本信息：持卡人的基本信息、卡类型、每次刷卡的时间和地点等。该研究数据共包含77976010次乘客刷卡记录，共对应着北京市8549072张一卡通。鉴于非技术原因，本文所使用的公交IC卡刷卡数据不包括祥龙公司的运输线路数据和城市轨道交通数据。北京市公交线路按照计价方式主要分为两种：分段计价线路与一票制线路，本研究为了统计各公交站台上下车流量，所用数据主要为分段计价刷卡数据，共37649207条。

本研究中所使用的兴趣点数据(POI)为北京市2010年兴趣点数据，共有113810条，来自新浪微博地理服务平台，如图3所示。不同类别的兴趣点数量如图4所示。

2 研究方法

本研究以LBS技术所提供的城市公交IC卡刷

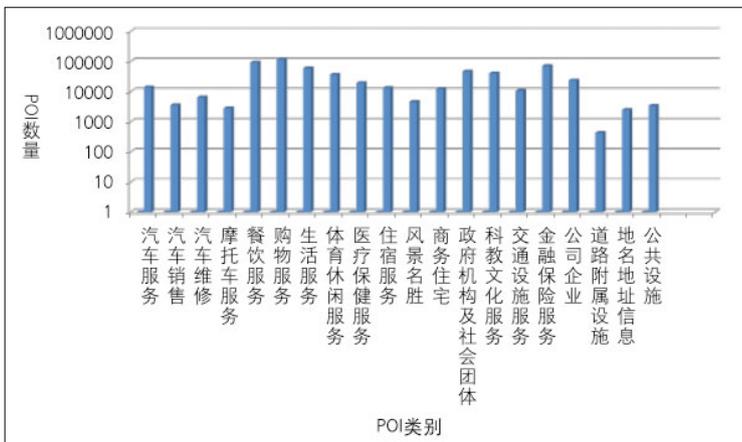


图4 2010年北京市兴趣点分类统计
Fig.4 Classification of POIs in Beijing Metropolitan Area in 2010

卡数据和兴趣点数据为基础，利用数据库系统对这些数据进行预处理，使之成为能被数据挖掘工具所直接使用的数据；之后构建城市功能区识别(discovery zones of different functions, DZoF)模型，通过聚类分析和模式识别实现对城市功能区的识别，本研究的整体技术路线如图5所示。

2.1 公交站台的聚类

2.1.1 公交IC卡刷卡数据预处理

原始数据是按照公交线路进行的流量统计，存在多条公交线路经过同一个站台(platform)的问题，利用SQL Sever软件，实现对相同公交站台的不同公交线路统计值的加和处理，统计得到每个公交站台的流量，其中x为Platform ID(x=1, 2, ..., 8691), y为日期(y=7, 8, ..., 13), z为时间(z=0, 1, ..., 23)。

2.1.2 PF(Platform flows)数据模型

针对每一个公交站台(platform)，构建一个上车流量统计向量(Inflows vector), $\langle X_{7,0}, X_{7,1}, \dots, X_{i,j}, \dots, X_{13,23} \rangle$ ，其中 $X_{i,j}$ 为该站点(platform)在2008年4月i日，第j个小时内的上车人数(i=7, 8, ..., 13; j=0, 1, ..., 23)。同时，构建一个下车流量统计向量(outflows vector), $\langle Y_{7,0}, Y_{7,1}, \dots, Y_{i,j}, \dots, Y_{13,23} \rangle$ ，其中 $Y_{i,j}$ 为该站点(platform)在2008年4月i日，第j个小时内的下车人数(i=7, 8, ..., 13; j=0, 1, ..., 23)。

本研究数据通过转换为二维时间序列数据，并进行维度归约，构造函数，利用不同时间上下车人数的比值作为公交站台流量相似性比较的指标数据：

$$Z_{ij} = \frac{X_{i,j}}{Y_{i,j}} \quad (1)$$

因此，针对每一个公交站台(platform)构建PF模型，即 $\langle Z_{7,0}, Z_{7,1}, \dots, Z_{i,j}, \dots, Z_{13,23} \rangle$ ，其中 $Z_{i,j}$ 为该站点(platform)在第*i*天，第*j*个小时内的上下车人数之比($i=7, 8, \dots, 13$ ； $j=0, 1, \dots, 23$)。

2.1.3 数据降维

根据北京市一星期内24小时平均公交上下车流量数据统计，发现出行时间集中在5:00-23:00时间段。对原数据进行冗余特征的约简，PF数据由168维降为126维。

接下来，对Platform ID为1934的公交站台一星期内24小时上车流量数据进行统计，通过对统计结果的观察，发现工作日内的流量数据具有较强的一致性，而周末两天的流量数据则具有较强的一致性。

进而，对工作日和周末的公交站台的流量数据进行了相关性分析，结果显示，不同工作日内相同时间段的流量数据在0.01(双侧)水平上呈显著相关，据此对一周工作日内和周末两天的数据分别计算其算数平均数，原数据维度就由126维降至36维，在保留原数据特征的基础上，避免了维度灾难。

2.1.4 聚类算法

以往城市问题的研究中常常选择K-Means作为聚类的首选算法，该算法广泛应用于将数据集分解为几个不同的簇，性能优越^[22]。但是当所要处理的数据维数增多时，其计算效率就会下降，且不会取得较好的结果^[23]。由于本研究用数据的高维特征，因此并没有采用同类研究中常用的K-means聚类算法作为公交站台的聚类算法，而是选用在高维数据聚类中表现较好的EM算法作为聚类算法。

该算法的基本过程大致可归结为：首先，对参数做初始猜想；然后，迭代地改进这些估计。算法中对参数的估计是利用最大似然法进行的。

一维高斯分布所产生点的概率密度为：

$$\text{prob}(x|\Theta) = \prod \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \quad (2)$$

如果 μ 和 σ 的值是未知的，就需要一个过程来估计他们，即选择使得上式最大化的 μ 和 σ 。这种估计模型参数的方法在统计学上称作最大似然估计法。

2.2 城市功能区的识别

2.2.1 公交服务区POIs数据的采集及预处理

本研究以500m为半径界定公交站服务区空间尺度^[24]，建立以公交站为中心，以500m为半径的城市公交站服务区。针对每个公交站台，统计不同类别的POIs数据个数，表示为，其

i 为公交站台的ID号($i=1, 2, \dots, 8691$)， j 为POIs一级编码($j=1, 2, \dots, 20$)。2010年北京市的POIs数据中，餐饮服务POIs共有90819个，购物服务POIs共有116499个，而风景名胜POIs只有4575个。在统计分析的过程中，这种POIs数量级的不同会影响对于功能的识别。因此，需要对原始POIs数据按照如下公式进行Z-Score标准化。

$$x_{ij}^* = \begin{cases} \frac{x_{ij} - \bar{x}_j}{S_j} & (S_j \neq 0) \\ 0 & (S_j = 0) \end{cases}$$

$$(i=1,2,\dots,m; j=1,2, \dots,n) \quad (3)$$

x_{ij}^* 为标准值， x_{ij} 为原始值， S_j 为标准差， \bar{x}_j 为均值， m 和 n 为数据矩阵行数、列数。

2.2.2 兴趣点(POIs)数据模型

对于每一个公交站台，构建一个POIs数据的特征向量，FD(frequency density)数据模型，表示为 $\langle fd_1, fd_2, \dots, fd_{20} \rangle$ 。

其中 fd_i 表示在公交站台服务区*r*内的第*i*类兴趣点的频数密度(frequency density, FD)，即：

$$fd_i = \frac{\text{标准化后该公交站台服务区内第}i\text{类POIs的个数}}{\text{公交站台服务区}r\text{的面积}} \quad (4)$$

同样，对于每一个公交站台，构建了另外一

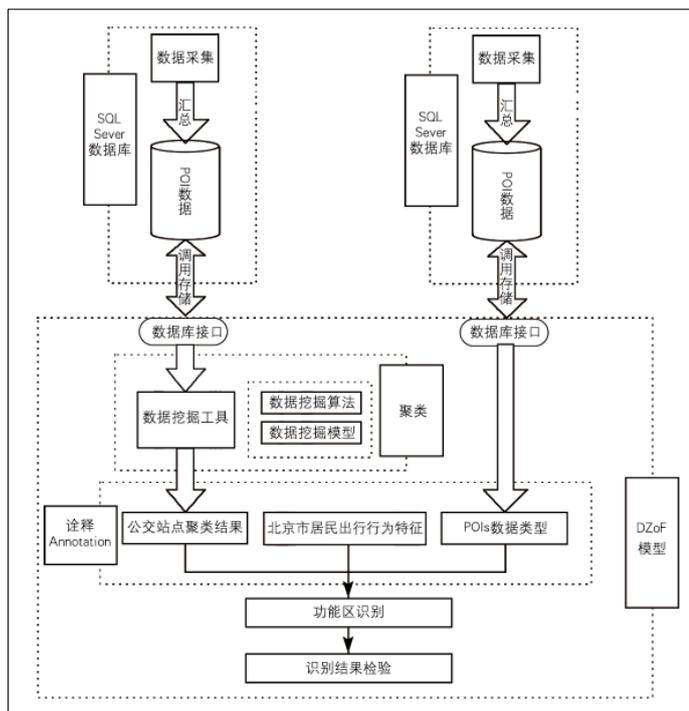


图5 技术路线
 Fig.5 Technical route

个POIs数据的特征向量, CR (category ratio)数据模型, 表示为 $\langle cr_1, cr_2, \dots, cr_{20} \rangle$ 。

其中 cr_i 表示第*i*类兴趣点在该区域中所有兴趣点的百分比, 即:

$$cr_i = \frac{\text{标准化后该公交站台服务区内第}i\text{类POIs的个数}}{\text{标准化后该公交站台服务区内POIs的总数}} \quad (5)$$

2.2.3 城市居民出行行为特征

城市居民出行行为的研究主要包括以下内容: 出行量、交通出行结构、交通出行的目的、交通出行空间分布以及出行影响因素等^[25], 本研究主要采用文献法, 重点关注出行结构和出行目的, 以及出行行为的时空特征。

2011年底, 北京市六环内日均出行总量为2873万人次, 其中公共汽(电)车日均出行量为811万, 轨道交通日均出行量为395万人次, 小汽车日均出行量为948万人次, 出租车日均出行量为199万人次, 自行车日均出行量为432万人次。公共汽(电)车依然是居民日常出行的主要交通工具。

居民每日出行时间分布整体呈现“两强两弱”的特征, 其中“两强”分别为7:00-8:00的早高峰和17:00-18:00的晚高峰, 而“两弱”则为11:00-12:00和14:00-15:00两个时间段。在白天时段, 出行的最低谷往往出现在中午12:30左右。

基本出行(上学、上班出行)是居民日常出行的主要目的。同时, 居民出行目的在工作日和休息日出现明显差异, 工作日居民出行目的上班、回家、上学排在前三位, 而休息日购物、文化娱乐、回家则排在前三^[26]。

除此之外, 基于现有居民出行调查数据, 部分机构与学者对资料进行了多因素综合分析。北京大学柴彦威教授运用时间地理学的基本理论, 从城市居民活动的时间结构入手, 调查居民连续时间内所有活动, 形成活动日志, 通过属性分析和行为特征分析, 描述居民日常出行的时空特征^[27, 28]。2007年长沙市规划管理局、长沙市统计局、国家统计局长沙城市调查队联合组织对长沙2000户居民家庭进行了入户调查, 研究报告显示中小学生学习日出行主要集中在6:00-7:00, 16:00-17:00两个时间段, 6:00-7:00这一时段出发地通常为居住地, 目的地则为学校, 而16:00-17:00这一时间段正好与之相反。针对以消费娱乐为目的的出行行为, 柴彦威等以深圳居民消费行为问卷调查为基础, 总结出了该城市居民日常消费购物行为的频率特征、购物时段特征以及购物行为的空间特征, 得出城市居民工作日购物时间主要集中在下午下班之后的时间段, 休息日购物主要集中在下午和晚上的时间段^[29]。

这些研究为人类活动的时间与空间之间架起了桥梁, 使得可以通过研究人类活动时间规律来预测人类活动的空间规律^[30]。

2.2.4 城市功能的识别

对于城市功能区的分类标准有很多^[31], 本研究主要根据区域所承载的社会经济功能进行功能区划分, 主要包括居住区、风景名胜区、商业娱乐区、公共管理及科教文化区等, 并根据人类活动的密集程度将部分功能区细化为成熟区与未成熟区。

就具体识别方法而言, 本研究主要利用居民出行调查所获得的数据中得到的居民在出行时间和出行目的相关性特征以及POIs数据模型等, 对公交站台聚类得到的结果进行功能识别判断。

功能区的识别主要是以下三种方法的综合应用:

(1) 通过计算得出每个簇(cluster)的FD(frequency density)模型, 并对该数据进行排序(得到内部排名); 其次计算得出每个簇(cluster)中的CR(category ratio)模型, 进行排序(得到外部排名)。

(2) 根据每个簇(cluster)的出行时间流量特征进行判断。

(3) 居民的普遍认知。人们通常了解一些知名区域的功能, 比如故宫、中关村、北海公园等。

3 结果

3.1 公交站台聚类结果及TAZ尺度汇总

利用EM算法对公交站点依据流量数据进行聚类, 分别得到6个不同的簇(每一个公交站点唯一地属于一个簇, C0~C5)。利用公交站台和交通分析小区的空间从属关系, 对每一个交通分析小区进行统计, 分别统计各小区中从属于不同簇别的公交站台数。

现实情况下, 商住、产住等混合用地是广泛存在的, 土地区块在功能上往往呈现多元化, 为了方便表征及可视化, 本文选取分布最多的簇别作为该交通分析小区的类别, 将聚类结果在交通分析小区的尺度上汇总(Sparse为未分类区域), 如图6所示。

3.2 功能识别

3.2.1 POIs模型建立

依照3.1节公交站台聚类结果, 对各聚类所得簇(C0~C5)分别建立POIs数据模型, 计算各功能区的频数密度值和类别比排名(rank of category ratio, RCR)值, 如表1所示。

3.2.2 居民出行时间流量特征

由EM算法聚类得出的簇(C0~C5)在一周内

的工作日、休息日出行时间流量特征(上下车人数),如图7~图10所示。

3.2.3 识别结果

针对EM算法聚类结果进行功能识别:

(1)成熟居住区(C0)。

该区域内住宅兴趣点分布较广,商务住宅比例较高(FD值为-0.073),且为居民生活服务的医疗卫生服务、住宿服务和教育服务等兴趣点配套均衡,是典型的居住区兴趣点分布。

同时通过一周流量数据分析,可以发现该区域工作日最大的出发流量高峰是在清晨(7 00-8 00点的上班时段),返程流量高峰出现在傍晚(17 00-19 00点的下班时段),是典型的居住区出行模型。

(2)待开发区(C1)。

该区域兴趣点主要为摩托车、汽车服务,分布较多的汽车4S店、摩托车销售、汽车摩托车维修点等,周边基础设施建设尚不健全。

(3)风景名胜(C2)。

该区域分布比例最高的兴趣点是风景名胜点,在该类别相较于其他区域具有较高FD值(FD值为0.042)。同时为游客服务的餐饮服务(1/20)、住宿服务(8/20)等在外排名(RCR)中也较靠前,而且该区域工作日、休息日出行流量

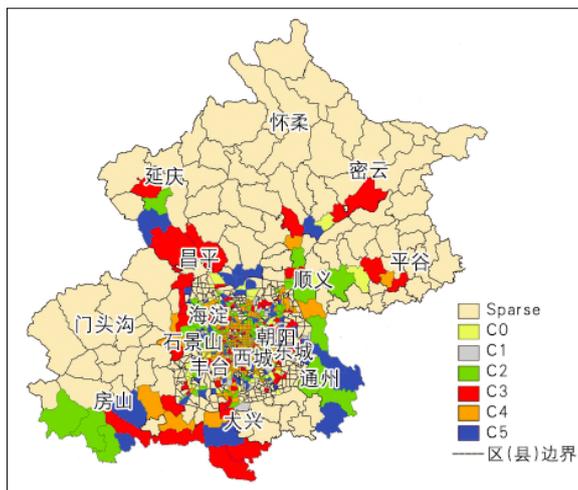


图6 北京市功能区域
Fig.6 Functional zones of Beijing

差距不大,每天不同时段出行相对平均,休息日出行流量高于工作日流量。

(4)商业娱乐区(C3)。

从该区域兴趣点数据分布特征可以看出:餐饮服务、购物服务、生活服务的FD值较高,在所有簇中分别排名第2、第1和第1。与此同时,餐饮购物信息点在区域内部所有信息点中占比例较高(CR值较大)。例如,餐饮服务类的簇内CR值

表1 EM聚类所得功能区的兴趣点特征值(FD:频数密度,RCR:类别比排名)

Tab.1 Overall POIs feature vector and ranking of functional zones formed by EM (FD: Frequency Density, RCR: Rank of Category Ratio)

兴趣点分类	C0		C1		C2		C3		C4		C5	
	FD	RCR										
汽车服务	-0.077	7	-0.025	6	0.073	9	0.03	19	0.021	18	-0.02	6
汽车销售	-0.075	6	0.034	2	-0.006	19	0.089	14	0.073	13	-0.063	12
汽车维修	-0.005	3	0.032	3	0	18	0.119	9	0.084	12	-0.012	4
摩托车服务	0.063	1	0.006	5	0.057	13	0.085	15	0.041	16	0.117	1
餐饮服务	-0.186	18	-0.109	13	0.142	1	0.149	7	0.205	5	-0.095	18
购物服务	-0.173	16	-0.141	16	0.039	15	0.214	3	0.107	11	-0.051	10
生活服务	-0.156	13	-0.157	18	0.099	5	0.216	2	0.114	10	-0.026	7
体育休闲服务	-0.16	14	-0.114	14	0.124	2	0.095	10	0.307	1	-0.057	11
医疗保健服务	-0.106	9	0.013	4	0.06	12	0.187	5	0.056	15	-0.004	3
住宿服务	-0.183	17	-0.14	15	0.075	8	0.187	4	0.18	6	-0.034	8
风景名胜	-0.129	12	-0.076	11	0.042	14	-0.033	20	0.167	8	-0.075	15
商务住宅	-0.073	5	-0.094	12	0.072	10	0.152	6	0.07	14	-0.018	5
政府机构及社会团体	-0.124	11	-0.18	20	0.082	6	0.135	8	0.224	2	-0.11	20
科教文化服务	-0.202	19	-0.173	19	0.068	11	0.067	18	0.22	3	-0.095	17
交通设施服务	-0.173	15	-0.076	10	0.111	4	0.089	13	0.17	7	-0.066	13
金融保险服务	-0.214	20	-0.057	8	0.114	3	0.094	11	0.216	4	-0.105	19
公司企业	-0.12	10	-0.144	17	0.075	7	0.069	17	0.117	9	0.017	2
道路附属设施	-0.008	4	-0.039	7	0.031	16	0.092	12	-0.036	20	-0.068	14
地名地址信息	0.015	2	0.039	1	0.016	17	0.084	16	0.021	17	-0.079	16
公共设施	-0.102	8	-0.068	9	-0.048	20	0.214	1	-0.036	19	-0.044	9

排名第7, 购物服务类的簇内CR排名第3。同时, 通过流量特征图, 可以看出该区域工作日下午时段(17:00-19:00)会出现下车流量高峰, 说明很多居民在该区域消费购物, 以及参加休闲娱乐活动。

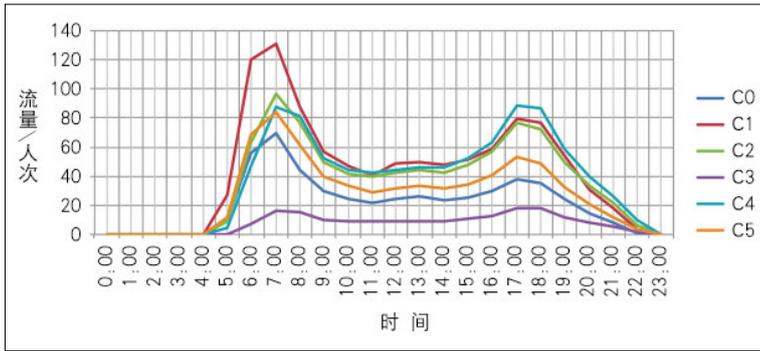


图7 EM聚类所得各簇的工作日上车流量
Fig.7 The inflows on work days of clusters formed by EM

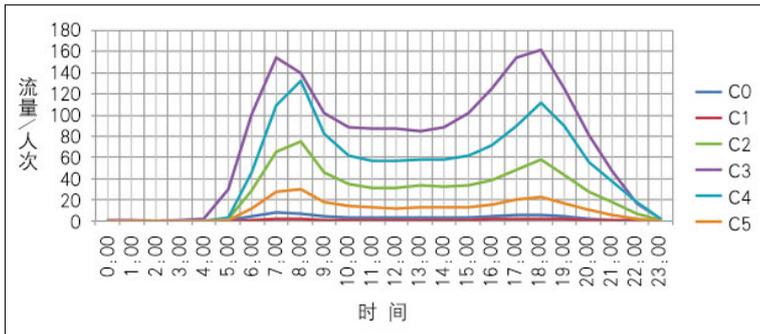


图8 EM聚类所得各簇的工作日下午车流量
Fig.8 The outflows on work days of clusters formed by EM

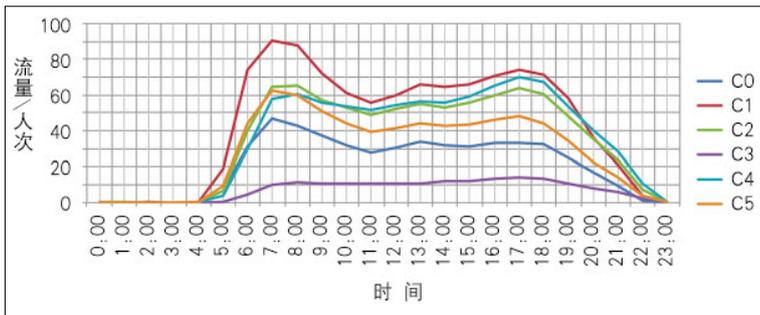


图9 EM聚类所得功能区的休息日上车流量
Fig.9 The inflows on weekends of functional zones formed by EM

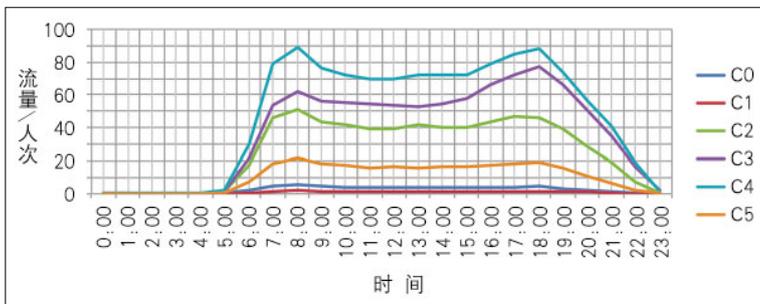


图10 EM聚类所得功能区的休息日下午车流量
Fig.10 The outflows on weekends of functional zones formed by EM

(5)公共管理及科教文化区(C4)。

该区域分布比例最高的兴趣点是政府机构及社会团体, 相较于其他区域具有较高FD值(FD值为0.22), 该类型兴趣点占该区域中兴趣点数的9.7%, RCR值排名为第二。并且该区域内科教文化兴趣点较多。同时交通服务设施、体育休闲服务、住宿服务等在外部排名中也较靠前。

(6)新兴居住区(C5)。

该区域的兴趣点数据结构和C0类似, 按照各类兴趣点数量占该区域内中兴趣点总数的比例进行排名, 住宅类位列第5, 同时该区域内医疗保健第3、生活服务类第7, 是典型的居住区兴趣点分布结构。

另一方面, 从工作日、休息日站点流量数据可以看出, 该区域工作日最大的出发流量高峰是在清晨(7:00-8:00点的上班时段), 返程流量高峰出现在傍晚(17:00-19:00点的下班时段), 是典型的居住区出行模型。但是该区域流量相对于C0区域较少, 日流量为C0的1/4左右, 这说明该区域人流量不大, 尚处在发展阶段。

(7)未分类区域(Sparse)。

由于山地、森林、河流等原因, 部分区域无公交流量数据, 本文将该类区域归为一类。

根据功能区识别结果, 对各功能区内的交通分析小区数目及面积进行了统计, 如表2所示。

3.3 识别结果的检验

为了检验城市功能区识别模型识别结果的准确性, 本研究将实验得到的北京市不同功能区图与北京市城市总体规划(2004-2020)中的用地现状图以及谷歌地图进行对比。此外, 我们还将研究结果与详细的北京市各交通分析小区土地利用数据, 进行分析对比(原始数据包含各个交通分析小区内的各类型用地的面积), 以检验识别的总体准确率。根据公共用地面积(具体包含公共设用地和市政用地)占该交通分析小区总面积的比例大小, 对1118个交通分析小区由大到小进行排序, 选取前50个公共交通分析小区, 除去无公交IC卡刷卡信息的分析小区, 共有44个分析小区。其中有28个被模型识别为公共管理及科教文化区, 准确率达63.6%。采用同样的方法, 对居住用地进行分析对比, 准确率为58.06%, 对比结果如表3所示。

综合考虑北京市商住、产住高度混合的用地现状与研究对比分析, DZof模型对于北京市主要的功能区能有效地加以识别, 具有一定的准确度。

4 讨论和结论

基于能够获取的城市公交刷卡数据和兴趣点

数据,本文通过构建DZoF模型,进行了北京市城市功能区的识别,共得到6个类别的功能区,分别为公共管理及科教文化区、风景名胜区、商业娱乐区、成熟居住区、新兴居住区和尚未分类区域。本研究基于北京市的公交IC卡刷卡数据和兴趣点数据,通过数据采集、数据预处理、聚类和模式诠释等一系列工作对北京市城市功能区进行了识别,并对识别结果进行了检验。其一,针对北京市部分具有代表性的用地区块,将最终的识别结果与北京市城市总体规划(2004-2020)中的用地现状图以及谷歌地图进行对比,具有较好的吻合度。其二,与2005年北京市居民出行调查所涵盖的北京市18个区县共计1118个交通分析小区的基础地理数据进行对比,其中公共管理及科教文化区和居住区识别的准确率分别为63.60%与58.06%,与研究初期70%的准确率预期尚有差距,这可能由以下几个方面的因素造成:(1)现实情况中,北京市商住和产住等混合用地的程度较高,特别是主城区土地利用结构复杂,识别难度大。(2)由于客观原因,未能获得与分段制公交线路相似的能够完整记录乘客出行时间空间信息的城市轨道交通数据,造成了对部分区域的识别误差。(3)为了研究观测方便,本研究将城市功能识别的结果在交通分析小区的尺度上进行汇总,部分交通分析小区空间尺度较大,精细化程度不够,从而在汇总过程中产生误差,对研究结果的准确性造成影响。

总体而言,研究结果显示DZoF模型对于北京市城市功能区特征具有一定的识别能力,更重要的是验证了基于大规模的人类活动数据,利用计算机数据挖掘分析技术并结合传统调查统计而开展城市空间结构研究的可行性。

本研究的结果能够更好地帮助人们轻松地理解一个复杂的城市的空间功能结构,辅助城市规划者基于人类活动和兴趣点开展不同城市功能区的规划,对城市规划具有指导和参照价值,也可以为房地产开发的选址等具体实践提供重要的决策支持。同时,研究中所使用的方法可以为未来城市空间结构的研究提供借鉴和参考。

本研究仍有很多不足之处,需要在未来的研究中加以改进:(1)本研究中忽略了一票制公交车乘客刷卡数据,未来可以通过构建新的数据模型,有效地利用一票制公交车乘客刷卡数据,提高识别结果的准确性。(2)公共交通数据不足,在未来研究中,将进一步通过增加轨道交通和出租车数据,完善人类活动信息。(3)现实情况中,商住、产住等混合用地是广泛存在的。本研究将实验结果在TAZ尺度上汇总时,忽略了混合用地的因素,选取比率最大的功能作为该区的功能表征。在未来的研究中,也可考虑加入混合用

表2 各功能区信息统计

Tab.2 Information of each functional zone

功能区	TAZ个数	面积(km ²)
未分类区域(Sparse)	357	11061.130
成熟居住区(C0)	63	326.292
待开发区(C1)	25	82.106
风景名胜区(C2)	155	973.087
商业娱乐区(C3)	129	2068.882
公共管理及科教文化区(C4)	267	918.993
新兴居住区(C5)	122	974.291

表3 识别结果与交通分析小区用地情况对比分析

Tab.3 Comparative analysis between identification results and land use of TAZ

分区	分项	有效对照数据	识别结果	准确率
公共管理及科教文化区	交通分析小区数目	44	28	63.60%
	总面积(m ²)	59596348	26824488	-
居住区	交通分析小区数目	31	18	58.06%
	面积(m ²)	49319636	15423667	-

地分类。

注释(Notes)

数据来自北京市2012年统计年鉴, <http://www.bjstats.gov.cn>。

数据来自北京市公交网统计数据, <http://www.bjbus.com/>。

该数字为所有公交线路站点数的综合,而不是站台数量,站点为某一公交线路某站台的名称。交通分析小区(TAZ)是为了整体而全面地了解城市内交通流状况,综合考虑土地形态、道路分布、行政区划以及抽样调查的方便程度而人为划分的区域,其是基础数据搜集、整理和汇总的单位,也是交通分析的基础。本研究TAZ与北京市2005年居民出行调查所划分的TAZ一致。

新浪开放平台,2012年4月新浪微博LBS平台正式开放,第三方开发者可免费接入新浪位置服务。新浪微博LBS平台最具特色的是基于用户及基于POI(具体位置点)的接口两个功能,基于用户的相关接口,使用户能获取单个人的时间线动态;POI接口是基于某个具体位置的接口。<http://open.weibo.com/>。

数据来源于北京市交通发展研究中心2012年8月公布的《2012北京市交通发展年度报告》。未分类区域形成的原因主要为该区域内公交站台流量数据稀疏或为零。

数据来源于北京市规划设计研究院。

数据来源于2005年北京市居民出行调查。

参考文献(References)

- 胡俊. 中国城市:模式与演进[M]. 北京:中国建筑工业出版社, 1995.
HU Jun. Chinese Cities: Patterns and Evolution[M]. Beijing: China Architecture & Building Press, 1995.
- 赖世刚, 韩昊英. 复杂:城市规划的新观点[M]. 北

京:中国建筑工业出版社, 2009: 9, 10-13.

LAI Shigang, HAN Haoying. Complexity: A New Perspective of Urban Planning[M]. Beijing: China Architecture & Building Press, 2009: 9, 10-13.

- 3 BATTY M. Cities as Complex Systems: Scaling, Interactions, Networks, Dynamics and Urban Morphologies[R]. Berlin: Springer, 2009.
- 4 BATTY M. Invisible Cities[J]. Environment and Planning B: Planning and Design, 1990(17):127-130.
- 5 NEUHAUS F. "Urban Diary-a Tracking Project" UCL Working Paper Series[R]. Paper 151. Available on line: <http://discovery.ucl.ac.uk/19245/>.
- 6 CALABRESE F, RATTI C. Real Time Rome[J]. Networks and Communication Studies, 2006, 20:247-258.
- 7 RATTI C, PULSELLI R M, WILLIAM S, et al. Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis[J]. Environment and Planning B: Planning and Design 2006, 33(5):727-748.
- 8 GONZALEZ M C, HIDALGO C A, BARABASI A L. Understanding Individual Human Mobility Patterns[J]. Nature, 2008, 453:779-782.
- 9 GOODCHILD M F, JANELLE D. The City Around the Clock: Space-time Patterns of Urban Ecological Structure[J]. Environment and Planning A, 1984, 16:807-820.
- 10 GOODCHILD M F, KLINKENBERG B, JANELLE D G A. Factorial Model of Aggregate Patio-Temporal Behavior: Application to the Diurnal Cycle[J]. Geographical Analysis, 1993, 5:277-294.
- 11 JIANG B, YIN J, ZHAO S. Characterizing Human Mobility Patterns in a Large Street Network[J]. Physical Review E, 2009, 80(2):1136-1146.
- 12 HAMILTON B W. Wasteful Commuting[J]. The Journal of Political Economy, 1982, 90(5):1035-1053.
- 13 刘志林, 王茂军. 北京市职住空间错位对居民通勤行为的影响分析: 基于就业可达性与通勤时间的讨论[J]. 地理学报, 2011, 66(4):457-467.
- 14 LIU Zhilin, WANG Maojun. Job Accessibility and Its Impacts on Commuting Time of Urban Residents in Beijing: From a Spatial Mismatch Perspective[J]. Acta Geographica Sinica, 2011, 66(4):457-467.
- 15 WANG D, CHAI Y. The Jobs-Housing Relationship and Commuting in Beijing, China: The Legacy of Danwei[J]. Journal of Transport Geography, 2009, 17: 30-38.
- 16 ANAS A, ARNOTT R, SMALL K A. Urban Spatial Structure[J]. Journal of Economic Literature, 1998, 36(3):1426-1464.
- 17 MCMILLEN D P, MCDONALD J F A. Nonparametric Analysis of Employment Density in a Polycentric City[J]. Journal of Regional Science, 1997, 37(4):591-612.
- 18 LU D, WENG Q. Urban Land-Use and Land-Cover Mapping Using the Full Spectral Information of Landsat ETM+ Data in Indianapolis, Indiana[J]. Photogrammetric Engineering & Remote Sensing, 2005, 71(11):1275-1284.
- 19 XIAO J, SHEN Y, GE J, et al. Evaluating Urban Expansion and Land Use Change in Shijiazhuang, China, by Using GIS and Remote Sensing[J]. Landscape and Urban Planning, 2006, 75(1-2):69-80.
- 20 QI G, LI X, LI S, et al. Measuring Social Functions of City Regions from Large-Scale Taxi Behaviors[C]. The 9th IEEE International Conference on Pervasive Computing and Communications (PerCom '11). Seattle, USA: March 21-25, 2011:384-388.
- 21 CHANG H J, HWANG C. A Time-Geographic Analysis of Trip Trajectories and Land Use Characteristics in Seoul Metropolitan Area by Using Multidimensional Sequence Alignment and Spatial Analysis[C]//2010 AAG Annual Meeting, Washington, DC:2010.
- 22 龙瀛, 张宇, 崔承印. 利用公交刷卡数据分析北京职住关系和通勤出行[J]. 地理学报, 2012, 67(10):1-12.
- 23 LONG Ying, ZHANG Yu, CUI Chengyin. Identifying Commuting Pattern of Beijing Using Bus Smart Card Data[J]. Acta Geographica Sinica, 2012, 67(10):1-12.
- 24 WU X, KUMAR V, ROSS Quinlan J, et al. Top 10 Algorithms in Data Mining[J]. Knowledge and Information System, 2008, 14(1):1-37.
- 25 HUANG Z. Extensions to the K-Means Algorithm for Cluster Large Data Sets with Categorical Values[J]. Data Mining and Knowledge Discovery, 1998, 2(3):283-304.
- 26 黄丽彬. 大城市的轨道交通站点对地区发展的影响评价研究——以上海徐家汇轨道枢纽为例[D]. 上海: 同济大学, 2006.
- 27 HUANG Libin. Research on the Influence of Rail Transit Stations on Regional Development in Large Cities, a Case of Xujiahui Shanghai Railway Junction[D]. Shanghai: Tongji University, 2006.
- 28 张文尝, 王成金, 马清裕. 中国城市居民出行的时空特征及影响因素研究[J]. 地理科学, 2007, 27(6):737-742.
- 29 ZHANG Wenchang, WANG Chengjin, MA Qingyu. Spatial-Temporal Characteristics of Urban Resident Trips and Influence Factors in China[J]. Scientia Geographica Sinica, 2007, 27(6):737-742.
- 30 刘东梅, 齐彬岩, 张可, 等. 北京市居民出行时间小样本调查分析[J]. 交通运输系统工程与信息, 2009, 2(9): 24-26.
- 31 LIU Dongmei, QI Tongyan, ZHANG Ke, et al. Beijing Residents' Travel Time Survey in Small Samples[J]. Journal of Transportation Systems Engineering and Information Technology, 2009, 2(9):24-26.
- 32 丁威, 杨晓光, 伍速锋. 基于活动的居民出行行为研究综述[J]. 人文地理, 2008, 23(3):85-91.
- 33 DING Wei, YANG Xiaoguang, WU Sufeng. A Review of Activity-Based Travel Behavior Research[J]. Human Geography, 2008, 23(3):85-91.
- 34 李铮嵘. 中国城市居民交通出行行为的时间地理学研究[D]. 北京: 北京大学, 2001:1-2.
- 35 LI Zhengrong. A Study on the Time Geography of the Urban Residents' Travel Behavior in China[D]. Beijing: Beijing University, 2001:1-2.
- 36 柴彦威, 翁桂兰, 龚华. 深圳居民购物消费行为的时空间特征[J]. 人文地理, 2004, 6(19):80-84.
- 37 CHAI Yanwei, WENG Guilan, GONG Hua. A Study on Temporal and Spatial Characteristics of Shopping Behavior of Shenzhen Residents[J]. Human Geography, 2004, 6(19):80-84.
- 38 黄潇婷, 柴彦威. 面向LBS使用者的时间地理学研究评介[J]. 地理科学进展, 2009, 28(6):962-969.
- 39 HUANG Xiaoting, CHAI Yanwei. The Introduction and Comment on the Study of User-Centered Time Geography for LBS[J]. Human Geography, 2009, 28(6):962-969.
- 40 颜芳芳. 城市功能区发展模式研究[J]. 经济研究导刊, 2010(12):134-136.
- 41 YAN Fangfang. A Study on the Development Model of Urban Functional Area[J]. Economic Research Guide, 2010(12):134-136.