

基于公交刷卡数据和兴趣点数据的城市街区功能类型识别研究

——以北京市朝阳区为例

钟少颖 岳未祯 张 耘

摘要 作者利用公交刷卡数据 (Smart Card Data, SCD) 和兴趣点 (Point of Interest, POI) 数据, 借助 Oracle 数据库和地理信息系统对城市街区层面的功能属性识别进行了研究。其中 SCD 用于刻画街区客流特征的功能属性, POI 数据用于刻画街区设施功能属性。作者首先利用公交站点客流出行规律的功能属性对城市街区进行初步的分类。在此基础上, 利用 POI 数据的设施功能属性字段进行蚁群聚类, 对城市功能进行进一步细分和识别。最后利用遥感影像分析对识别结果进行检验。检验结果表明, 综合利用公交刷卡数据和兴趣点数据能较为准确刻画城市街区的功能特征。

关键词 城市功能识别 公交刷卡数据 兴趣点 街区

【中图分类号】F294 【文献标识码】A 【文章编号】2095 - 851X (2016) 03 - 0067 - 19

一、引言

城市功能区划分是城市管理和城市规划的基础。以往对城市功能区的划分多基于土地利用类型, 但是单纯考虑土地利用类型而不考虑土地承载的人口和设施的密度及功能属性难以准确反映城市区块功能。在土地城镇化向人的城镇化转型的过程中, 城市功能区划分尤其要关注功能区内人的因素。传统的统计数据由于统计口径等因素,

【基金项目】国家社会科学基金青年项目“高速铁路建设对中国区域经济一体化影响的研究”(批准号: 15CJL086)。

【作者简介】钟少颖 (1984 -), 中国科学院科技战略咨询研究院助理研究员, 邮政编码: 100190; 岳未祯 (1989 -), 中国科学院科技战略咨询研究院博士研究生; 张耘 (1955 -), 北京城市学院城市管理所研究员, 本文通讯作者, 邮政编码: 100184。

致谢: 感谢评审专家匿名评审, 当然文责自负。

难以准确刻画街区层面的客流和设施特征。在大数据时代,由基于位置服务(Location Based Services, LBS)技术所提供的公交智能卡刷卡记录、微博等社交媒体使用记录和手机通话记录等数据可以用来客观刻画城市客流活动空间特征,分析城市活动的强度和时空分布特征(秦萧等,2013;龙瀛等,2014;杨振山等,2015)。利用LBS技术提供的海量数据进行城市空间研究已成为城市研究的热点(Fernandes et al., 2014; Chatterjee and Das, 2015)。目前,国内也有众多学者通过城市居民活动形成的大数据来研究城市空间的规律特征,如新浪微博数据(甄峰等,2012;王波等,2013)、大众点评网数据(秦萧等,2014)、签到数据(王波等,2015)、公交刷卡数据(Smart Card Data, SCD)(龙瀛等,2012)、出租车GPS定位数据(Pan et al., 2013)等。

本研究利用SCD和兴趣点(Point of Interest, POI)数据对城市街区层面的功能属性进行划分,既考虑了客流功能属性特征,又考虑了物化的设施功能属性特征。研究以北京市朝阳区为例,首先探讨了朝阳区内部不同街区公交站点的客量密度及其随时间变化的特征,通过客流量变化特征与街区功能的相关性,构建公交站点的分类指标,对朝阳区的公交站点进行聚类分析。然后根据公交站点与街区的隶属关系,实现街区功能的初步划分。在此基础之上,利用街区内POI的功能属性特征和分布情况,对街区功能做进一步细分。

二、研究范围和数据说明

(一) 研究范围

城市功能识别首先要明确功能识别的地理单元,本研究选取街区作为功能识别的基本单元。街区作为构成城市的基本空间单元,是人们认知城市和构建城市意象的基本要素之一,也是城市生活的基本社会单元(李炬,2013)。街区的边界可以是城市道路、河流、绿化带、围墙等要素,即可以明显地界定一个片区的范围,具有较强的向心性、独立性和完整性。

本研究利用等级为三和四的城市道路将朝阳区划分为178个街区(见图1)。在178个街区中,位于三环以内的共有27个街区,位于三环到四环之间共有43个街区,位于四环到五环之间的共有76个街区,位于五环以外的有32个街区。

(二) 数据处理

1. 公交站点和线路统计

具体到朝阳区,共涉及公交站点1116个,其空间位置分布如图2所示,其中五环内公交站点数量约占总量的60%,涉及公交线路495条。

2. 公交刷卡数据的决策树匹配算法和数据清洗

本研究主要采用北京市2014年4月11日到17日一周时间的公交一卡通刷卡数据(不包括轨道交通刷卡数据),共计74 152 897条。北京市公交刷卡升级改革以

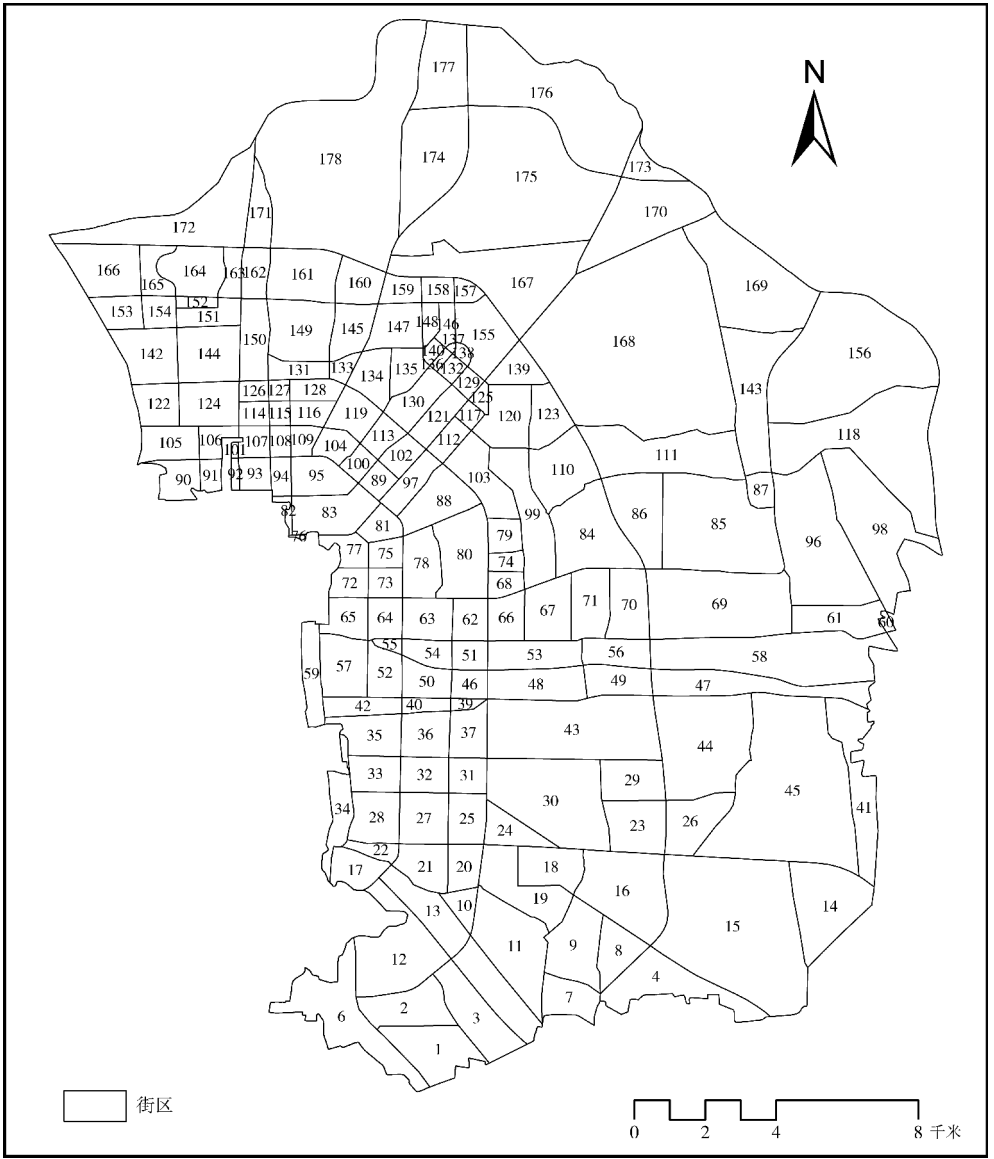


图1 朝阳区街区划分示意图

前，公交线路分为一票制线路和分段计价线路两种。其中，一票制线路只要求上车刷卡一次，下车无须刷卡，也就无法获取下车站点的信息。从朝阳区来看，此类线路为196条，占有公交线路的39.59%。如果不考虑这部分数据将造成很大的信息损失，使研究结果产生偏误。为解决这一问题，本文构建了一个一票制线路上车和下车点匹配的算法（如图3所示）。其基本假设是认为持卡人在晚高峰（17:00~20:00）的上车点（工作或其他目的地）为其在早高峰（6:00~9:00）时期的下车

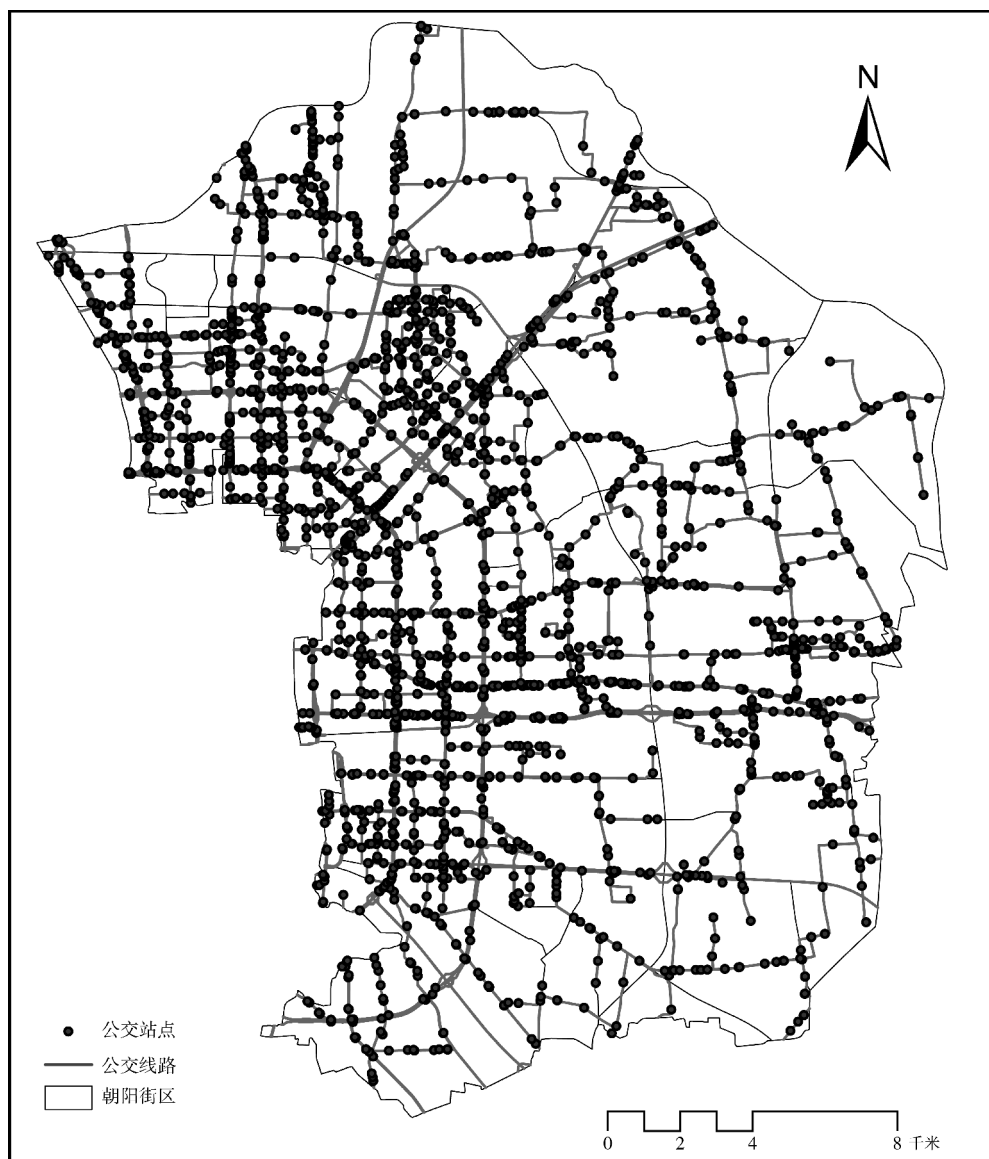


图2 朝阳区公交站点分布图

点（工作或其他目的地）。由于工作日出行和双休日出行的目的和规律不同，因此分别选择不同的匹配标准。工作日出行的主要目的是上下班通勤，其出行的起讫点、时间、线路等特征较少发生变化，是一种非弹性的出行；而双休日出行主要目的是购物、休闲和娱乐，其出行频率和出行的时空分布均显示出较强的随机性，是一种弹性出行。

经过数据的匹配和清洗后形成的有效乘车人数情况见表1。剔除掉的冗余数据占

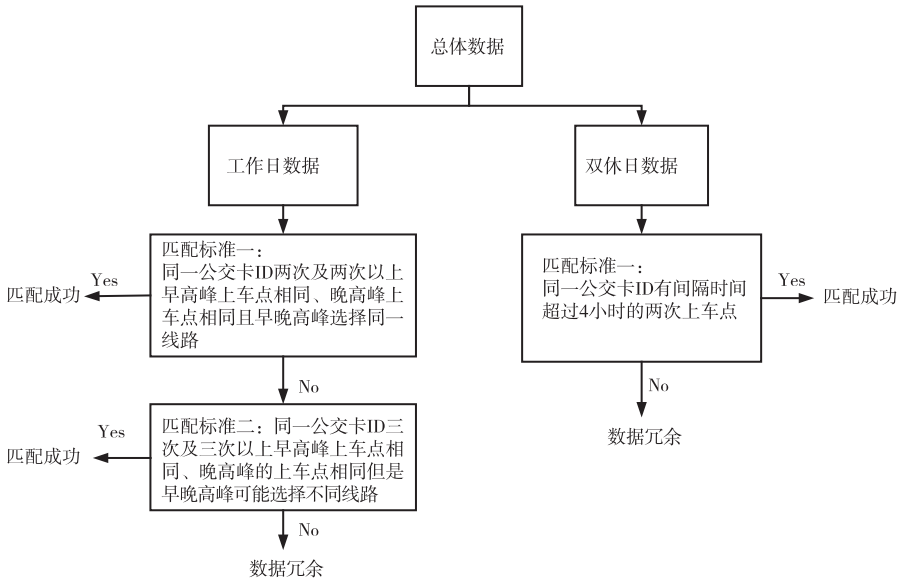


图3 公交刷卡数据匹配和清洗的算法思路

样本总量的 18.30%，匹配上的数据量占样本总量的 21.25%。根据不同时段乘车人数统计显示每天 5:00 至 22:59 时间段之外的乘车人数仅占当天总人数的 0.50% 左右。因此，本研究在分析公交站点的客流特征时，仅研究 5:00 至 22:59 之间的客流特征。

表1 数据匹配和清洗后不同时段乘车人数统计

时间段	2014/4/14(工作日)	2014/4/12(休息日)
0:00 ~ 4:59	16 768	1966
5:00 ~ 22:59	7 984 597	6 803 344
23:00 ~ 0:00	16 135	6652

3. POI 数据

POI 数据泛指一切可以被抽象为点的地理实体。POI 通常是在地图上标注的一个带位置的综合信息，包括名称、类别、经度、纬度等，涵盖了政府机构、学校、商业、金融、餐饮、娱乐、景点、加油站等基础信息。目前，越来越多的电子地图服务把目标集中在提供内容丰富且与人们生活密切相关的 POI 上，从而使 POI 数据具有多样化和精细化特征。

本研究采用 Esri 中国（北京）有限公司给出的地理信息公共服务平台 POI 的行业分类，并根据研究的特点进行归并，最后得到餐饮、购物、住宿、出行、文

体娱乐、金融服务、生活服务、汽车服务、教育、医疗、房产、旅游、企事业单位、行政机构和公共服务设施等 15 种行业分类，共计 88 236 个 POI，如图 4 和表 2 所示。

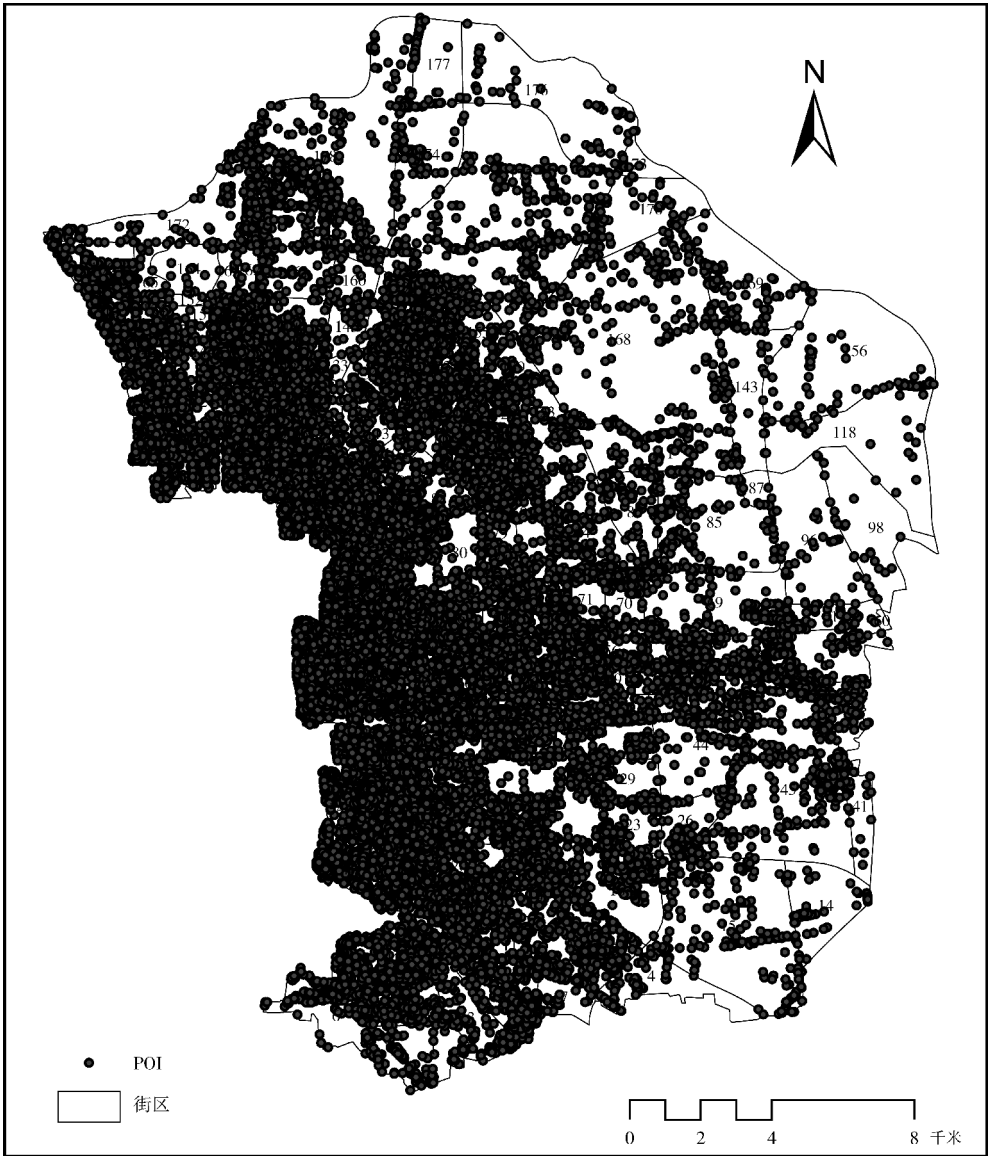


图 4 朝阳区 POI 分布图

数据来源：Esri 地理信息公共服务平台。

表 2 朝阳区各行业 POI 数量

单位：个

类型	数量	类型	数量
住宿	1455	教育	4021
行政机构	3216	金融服务	2584
医疗	2017	出行	4661
文体娱乐	2723	购物	14 509
生活服务	9986	企事业单位	16 003
房产	7500	公共服务设施	3145
汽车服务	2735	旅游	436
餐饮	13 245		

三、公交站点客流特征的功能属性分析

首先利用 Oracle 数据库软件构建一个公交刷卡数据库，再利用 SQL 操作对朝阳区公交刷卡数据进行总体分析，研究公交站点客流密度的时间特征，提出表征客流特点的参量，构建公交站点客流特征蚁群聚类模型对站点和站点所属的街区进行初步的类型划分。在此基础之上，利用 POI 及基础地理信息数据，采用 ArcGIS 的空间分析模块，提出与街区功能相关的 POI 类别，构建 POI 数据模型，同样利用蚁群聚类算法对街区功能进行进一步细分，完成对街区功能类别的识别。最后，利用遥感影像分析与实地调研相结合的方法对功能识别的结果进行检验。具体的思路和方法见图 5。

如前所述，居民在工作日出行和双休日出行的目的和规律不同。工作日出行以上班通勤为主，时间上具有高度集中性，早晚高峰相对集中，早通勤的起讫地点分别为居住地和工作地，晚通勤则相反。双休日通勤以购物、休闲娱乐为主，时间上相对分散，目的地相对集中。这种出行特点使得公交站点的客流在时间维度上具有不同的特征，也就可以从公交站点的客流特征来分析其所在街区的功能属性。本研究选择平房、望京、金台夕照、三里屯四个典型区域客流量较大的公交站点统计了工作日和休息日各时段（以小时为粒度）客流量均值来分析不同功能区的客流特征。

由图 6 可知，金台夕照工作日的出行客流量总的来说大于休息日出行客流量，同时形成了如下特征：（1）工作日晚高峰特征明显，也形成了早高峰，但早高峰客流量远小于晚高峰；（2）休息日客流量明显小于工作日。

图 7 描述了平房附近公交站点工作日和休息日各时段上车客流情况。从图中可见，总的来说形成了如下特征：（1）不论是工作日还是休息日，早高峰特征明显，也形成了晚高峰，但晚高峰客流量远小于早高峰；（2）休息日早晚高峰客流量明显小于工作日。

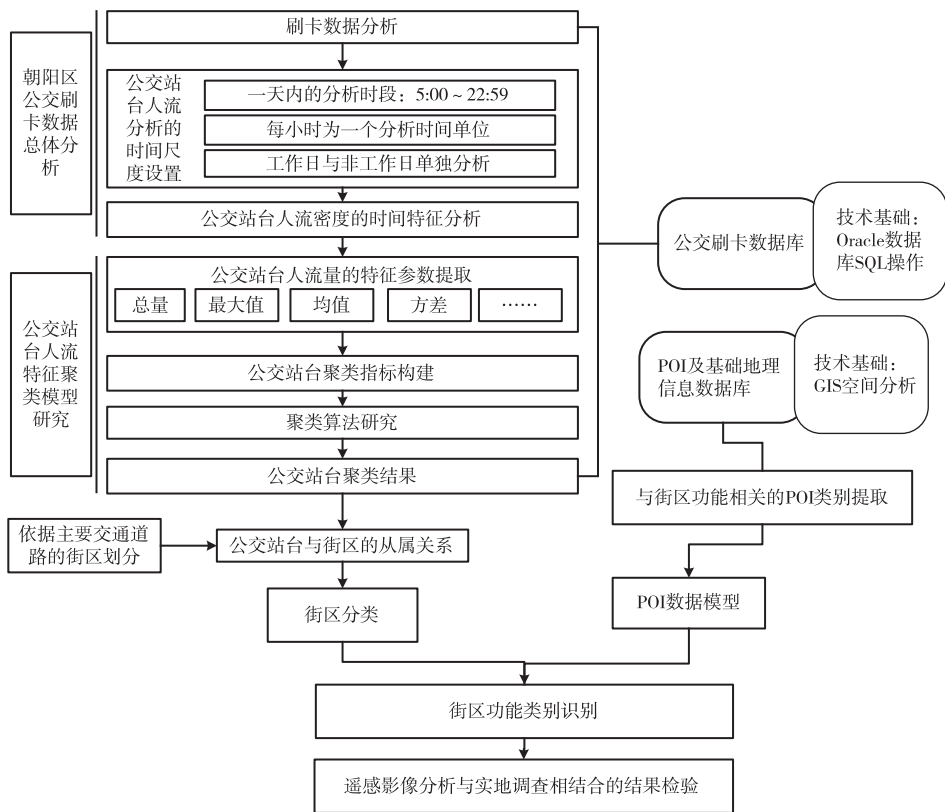


图5 基于SCD和POI数据的城市功能分区基本思路

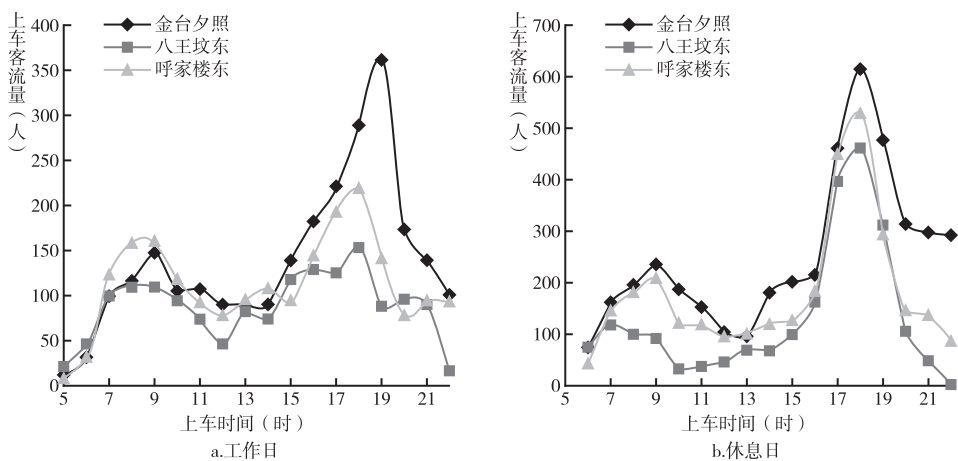


图6 金台夕照等公交站点各时段上车客流情况

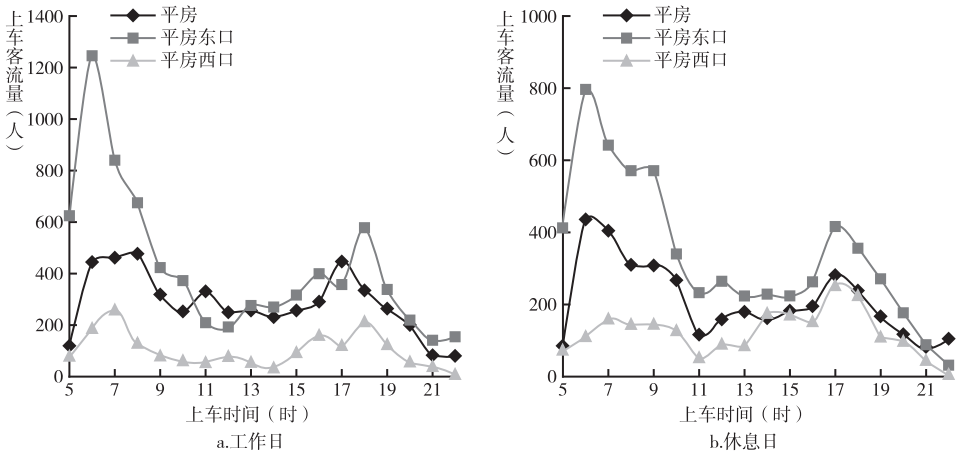


图7 平房附近公交站点各时段上车客流情况

图8描述了望京附近公交站点周末和工作日各时段上车客流情况。由图8可见，总的来说形成了如下特征：(1) 工作日早晚高峰特征都比较明显，形成了双峰分布，双休日的早晚高峰不是很明显，客流比较稳定；(2) 周末客流量明显小于工作日。

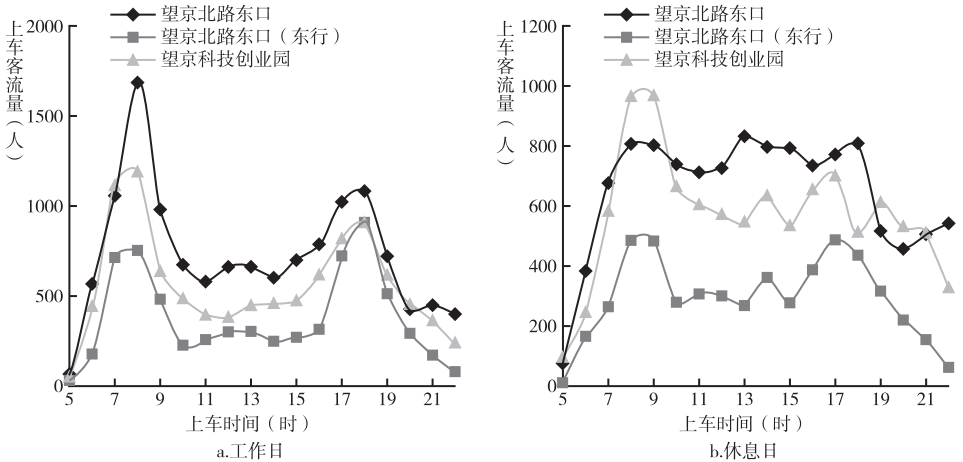


图8 望京附近公交站点各时段上车客流情况

图9描述了三里屯附近公交站点工作日和休息日各时段下车客流情况。从图中可见，总的来说形成了如下特征：(1) 工作日晚高峰特征都比较明显，晚高峰持续时间较长，形成了早高峰但是早高峰不明显；(2) 休息日客流量大于工作日，形成了比较明显的晚高峰，且晚高峰的持续时间较长。

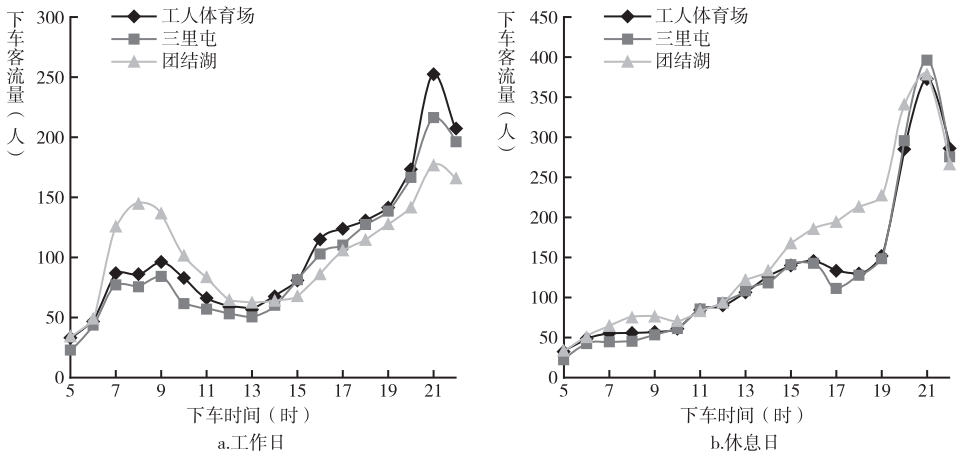


图9 三里屯附近公交站点各时段下车客流情况

四、基于客流变化特征的城市街区功能初步划分

上文对于不同站点客流随时间变化的特征进行了分析, 这些特征与站点周边街区的功能密切相关。本节将通过对客流特征参数与街区功能的关系分析, 提取街区功能聚类的指标, 采用聚类算法, 对街区功能进行初步的聚类分析。

(一) 聚类指标选择

工作日早高峰客流量体现了街区的居住功能, 而晚高峰客流量体现了街区的就业功能, 两者之比体现了街区的住宅功能与就业功能的比例。因此, 本文将(工作日)早高峰客流总量与晚高峰客流总量的比值(记为 index1)作为功能分区的指标之一。

本文采用的另一个指标是工作日晚高峰客流总量与周末同时段的晚高峰客流总量的比值(记为 index2)。周末晚高峰客流总量体现了街区的购物休闲功能, index2 可以表示街区就业功能与购物休闲功能的比例。需要说明的是, 根据对刷卡数据的分时段分析, 本文所指的早高峰时段为 6:00~9:00, 晚高峰时间段为 17:00~20:00。

(二) 蚁群聚类算法的流程和模型

蚁群聚类算法是基于蚁穴清理行为设计的聚类算法, 是一种结合群体智能的方法(Hou et al., 2014)。蚁群聚类算法具有很好的鲁棒性和适应性, 适合动态交错问题的求解, 多应用于多目标规划、旅行商问题、调度问题等优化问题的研究(Guo et al., 2015), 同样也适用于本文研究的问题。

本文采用的蚁群聚类算法基于蚂蚁堆积尸体行为, 将蚂蚁堆形成原理抽象为如下算法: 首先, 让蚂蚁分布在聚类空间上, 并让它们开始随机移动; 其次, 在移动的过程中, 蚂蚁需要就是否搬运或扔下蚂蚁尸体做出决策。在聚类过程中, 需聚类的对象

分属不同的类别，如何让蚂蚁在搬运物体的过程中实现同类物体的聚集，这就需要在对象属性空间引入一个相似度函数，以代替物体的个数，决定蚂蚁拾起或放下物体的概率。

将需聚类的对象集合定义为 $O = \{O_1, O_2, \dots, O_n\}$ ， $\{A_1, A_2, \dots, A_m\}$ 是对象集合 O 中的 m 个分类属性，则对象 O_i 的属性 A_j 可记为 A_{ij} 。对象 O_i 的相似度函数 $f(O_i)$ 的计算公式如下：

$$f(O_i) = \frac{1}{n} \sum_{O_j \in L(O_i, r)} d(O_i, O_j) \quad (1)$$

其中， $L(O_i, r)$ 表示对象 O_i 的半径为 r 的邻域， n 为邻域中对象的个数， $d(O_i, O_j)$ 表示对象 O_i 和 O_j 之间的相似度。在本研究中对象之间的相似度用对象属性之间的欧氏距离来计算：

$$d(O_i, O_j) = \sqrt{\sum_{k=1}^m (A_{ik} - A_{jk})^2} \quad (2)$$

(三) 聚类结果与分析

根据上述指标与方法，朝阳区内所有街区被分为四类，各类的聚类中心如表 3 所示。

表 3 基于 SCD 的街区聚类中心

	CLASS1	CLASS2	CLASS3	CLASS4
index1	2.341	0.761	1.260	1.230
index2	1.271	1.542	1.112	1.411

CLASS1 中居住功能远大于就业功能，而就业功能大于购物休闲功能，因此将其判定为居住区；CLASS2 中就业功能大于居住和购物休闲功能，因此将其判定为就业区；类似地，将 CLASS3 判定为购物休闲区，将 CLASS4 判定为混合区（图 10）。

上文已经利用公交站点客流特征的功能属性对公交站点所在的街区进行了初步的功能识别，形成了四类街区。这个划分充分考虑了街区客流的特征因素，但是仍过于粗略。

五、基于 POI 的街区功能聚类研究

为了对街区功能进行进一步的细分识别，本文在基于 SCD 的分类结果上，采用 POI 数据的功能属性对就业、购物休闲和混合街区的功能进行进一步细分。

(一) POI 数据的类型划分

POI 的分类是其应用的基础（张玲，2012），本研究根据 POI 所代表的地理实体

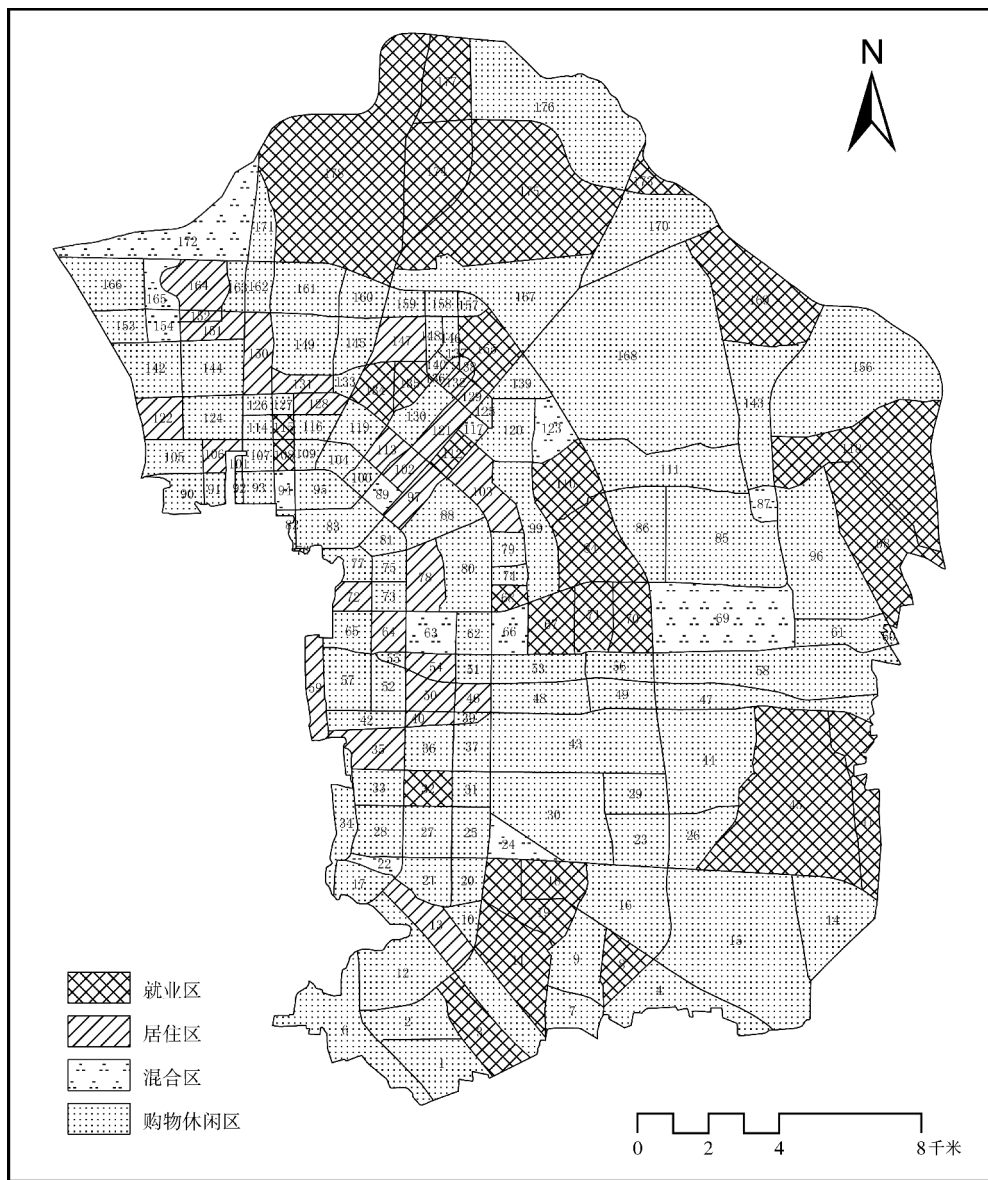


图 10 四大类街区分类结果

对街区功能影响的差异性对其进行分类。不同类型的 POI 对街区功能影响的方式不同，在计算其空间分布时所采用的方法也不同。本文将其分为密度型、区位型和距离型三类。

1. 密度型 POI。此类 POI 所代表的服务设施的密度与街区的主要功能相关，其密度大小因街区功能结构的差异而不同。以餐饮服务为例，餐饮服务设施密度较大的街

区主要功能为娱乐、购物的可能性较大。在分析它们的空间布局时，需计算某一类型设施在街区的密度。

2. 距离型 POI。此类 POI 所代表的服务设施的服务范围影响街区的功能，如幼儿园、小学、中学，距离这三类学校越远的街区以住宅为主要功能的可能性越小，相反距离这三类学校越近的街区以住宅为主要功能的可能性越大。该类 POI 在本研究中被称为距离型 POI。

3. 区位型 POI。此类 POI 所代表的服务设施占地面积较大，直接决定了街区的部分功能类型，例如风景名胜、高等院校等。在分析区位型 POI 的空间布局时，仅需判定街区内是否含有该类 POI。

(二) 聚类指标选择

上文根据 SCD 将街区分为了居住区、就业区、购物休闲区和混合区四类，但就业区还存在商务办公与产业园的区别，购物休闲类型中也存在购物功能与休闲功能的区别。就公交客流而言，商务办公区与产业园区、购物区与休闲区分别具有相似的特征，但商务办公与购物主要位于城市核心区域，产业园和休闲区则一般位于城市边缘区域。此外，混合区也可以根据 POI 的分布进一步区分各功能的强弱。基于此，本文选择对街区功能属性有较大影响的 POI 数据类型作为街区功能进一步划分的分类指标（见表 4）。

表 4 根据 POI 数据的街区分类指标

编号	指标名称	数据来源和处理说明	指标类型
11	街区常住人口密度(人/km ²)	先按街道计算平均人口密度,再根据街区计算街区常住人口密度,常住人口数据由北京市朝阳区统计局提供。	密度型
12	街区商务写字楼密度(栋/km ²)	利用 GIS 对 POI 数据进行提取计算	
13	街区产业园区密度(家/km ²)	利用 GIS 对 POI 数据进行提取计算	
14	街区购物中心密度(家/km ²)	利用 GIS 对 POI 数据进行提取计算	
15	外国使领馆和地市级及以上行政机构密度(家/km ²)	利用 GIS 对 POI 数据进行提取计算	
16	娱乐设施和综合性体育中心的密度(家/km ²)	利用 GIS 对 POI 数据进行提取计算	
17	到最近小学的距离的均值(km)	利用 GIS 对 POI 数据进行提取计算	距离型
18	到最近中学的距离的均值(km)	利用 GIS 对 POI 数据进行提取计算	
19	到最近医院的距离的均值(km)	利用 GIS 对 POI 数据进行提取计算	
110	是否含国家级公园和风景名胜	利用 GIS 对 POI 数据进行提取计算	区位型
111	是否含有 985 或 211 高校	利用 GIS 对 POI 数据进行提取计算	

(三) 数据计算和标准化处理

三类 POI 对应的指标分别采用以下方法进行计算：对于密度型指标，利用

ArcGIS 的 Zonal Statistic 工具统计街区内的 POI 个数，再除以对应街区的面积；对于区位型指标，利用 ArcGIS 的 Spatial Join 工具识别所在街区是否含有相应类型 POI，如有则赋值；对于距离型指标，采用 ArcGIS 的 Euclidean Distance 工具计算研究区域内各街区空间几何中心到其最近的 POI 的距离，然后将 POI 空间分布的信息转化为街区的属性信息，采用 Zonal Statistic 工具计算街区内各空间位置到最近 POI 的距离的均值。

由于各指标量纲不同，因此在进行聚类分析之前需对数据进行标准化处理。本文采用最大最小值标准化方法（离差标准化方法），结果值映射到区间 [0, 1]。转换函数如下：

$$X_{\text{new}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \quad (3)$$

其中， X_{max} 为样本数据的最大值， X_{min} 为样本数据的最小值。

(四) 街区功能识别结果

为对聚类结果进行分析、对街区各类别进行判别与理解，首先对各类别的聚类中心进行分析，聚类中心见表 5。为便于分析，增强各指标间的可比性，表中聚类中心的值为标准化后的值。

表 5 就业类街区 POI 聚类中心

	CLASS1	CLASS2	CLASS3
I2	0.018	0.066	0.427
I3	0.698	0.052	0.114

表 5 中 I2 是商务办公楼密度，I3 为产业园密度。从表中可见 CLASS1 的商务办公功能较弱，产业园的比重较大；CLASS2 的商务办公和产业园比重均较小；CLASS3 的商务办公功能较强，产业园比重较小。因此可判定 CLASS1 为产业园区、CLASS3 为商务办公区。

表 6 中 I4 代表了购物功能，I6、I10 代表了休闲功能。从表中可以判定 CLASS1 的购物、休闲功能都比较强；CLASS2 休闲功能比较强，而购物功能比较弱；CLASS3 的购物功能比较强，而休闲功能比较弱。

表 6 购物休闲类街区 POI 聚类中心

	CLASS1	CLASS2	CLASS3
I4	0.172	0.151	0.042
I6	0.171	0.055	0.033
I10	1.000	1.000	0.000

根据表 7 可判定 CLASS1 为住宅、商务办公、行政的混合街区；CLASS2 为住宅、产业园的混合街区；CLASS3 为住宅、购物的混合街区；CLASS4 主要是公共服务区；CLASS5 主要是住宅、高校的混合街区；CLASS6 主要是住宅和待开发的混合街区。

表 7 混合类街区 POI 聚类中心

	CLASS1	CLASS2	CLASS3	CLASS4	CLASS5	CLASS6
I1	0.256	0.334	0.558	0.182	0.213	0.354
I2	0.343	0.125	0.174	.006	0.034	0.516
I3	0.000	0.113	0.073	0.126	0.000	0.047
I4	0.058	0.089	0.210	0.035	0.055	0.085
I5	1.000	0.005	0.018	0.000	0.449	0.117
I6	0.175	0.050	0.094	0.009	1.000	0.367
I9	0.130	0.275	0.113	0.396	0.082	0.080
I10	0.000	0.000	1.000	0.000	0.000	0.000
I11	0.000	0.000	0.000	1.000	1.000	0.000

综上所述，本文在四大类的基础上，将所有的街区进一步划分为 12 个小类，分别为：住宅区（居住区），商务办公区，公共服务区，购物区，休闲娱乐区，购物休闲娱乐混合区，产业园区和商务办公混合区，住宅、商务办公和行政混合区，住宅、产业园混合区，住宅和购物混合区，住宅、高校（公共服务）混合区，住宅和待开发地区混合区。

由识别结果（图 11）可以看出，朝阳区的居住功能主要集中在东三环和东五环之间的环状区域，就业和购物休闲等其他功能则高度集中在东四环以内、奥运功能区以及以望京为核心的电子城功能区，而东五环以外主要为待开发区域。总体而言，朝阳区的功能区之间分割比较明显，就业和居住分布都比较集中。

六、城市街区功能识别结果检验

为了检验上述方法对街区功能识别的准确性，本文选择编号分别为 49、66、154 的三个街区，利用遥感影像信息与功能识别的结果进行比对（表 8）。结果表明，基于 SCD 和 POI 数据的功能识别结果基本符合实际情况。应该说，这一功能识别方法具有准确度高、成本低廉、计算方法简单的优势，对基于土地利用类型进行功能分区的传统方法有一定的改进。

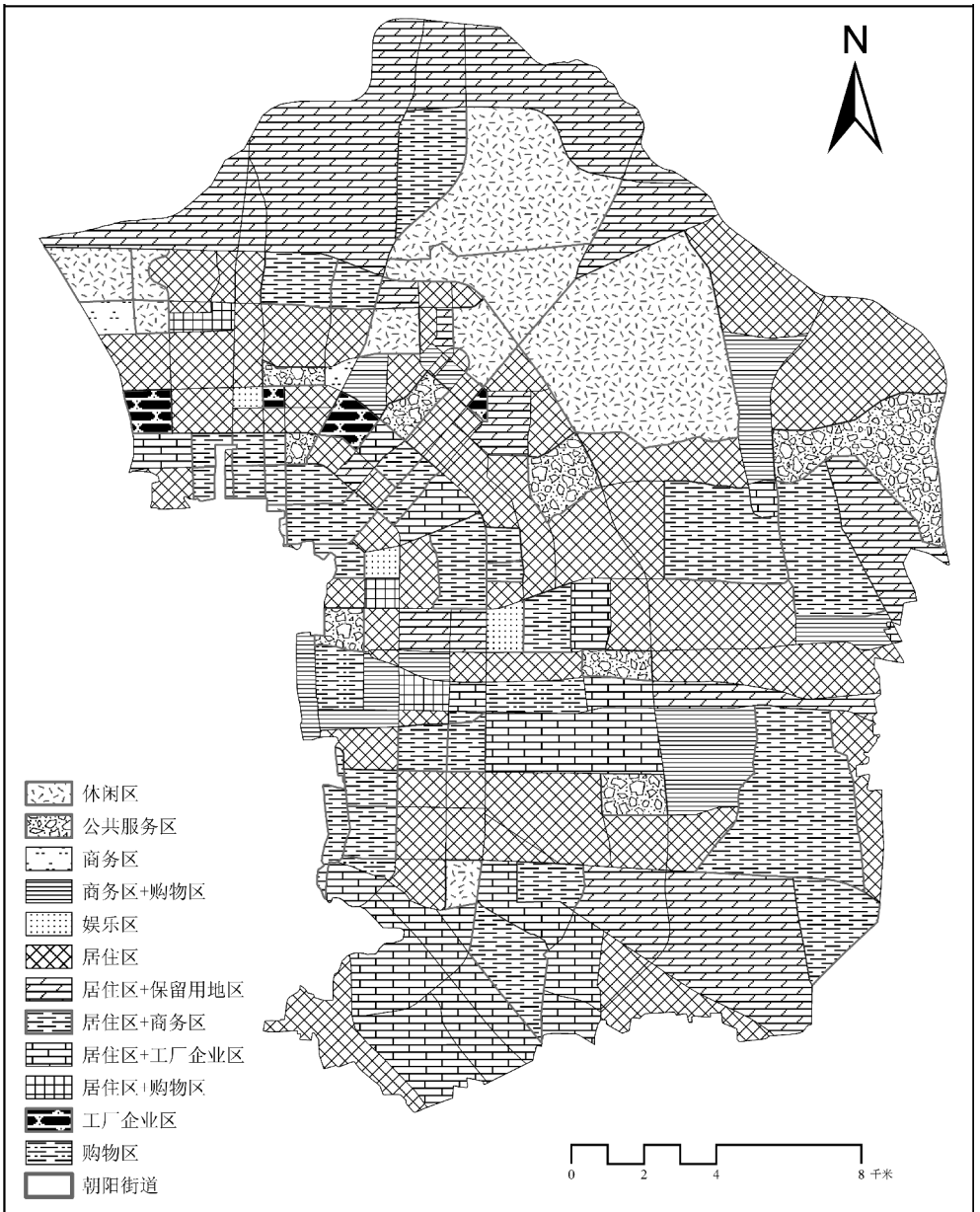





图 11 朝阳区街区功能识别分布图

表 8 基于遥感影像数据的城市街区功能识别结果检验

街区编号	街区范围	街区功能识别结果	街区遥感影像图	检验结果分析
49	朝阳路、东三环、东四环和建国路形成的区域	商务区		区域内有中央电视台等众多商务写字楼,住宅小区很少,是典型的商务就业区
66	石佛营路、姚家园路、青年路和朝阳北路形成的区域	居住区		区域内有国美第一城、华纺易城、珠江罗马家园、炫特区等社区,是典型的居住区
154	东五环、京密路、广泽路、阜通西大街、望京路形成的区域	混合区		区域内既有摩托罗拉大厦,中轻大厦等商务写字楼,又有保利中央公园、融科橄榄城等小区,是居住和商务就业的混合区

七、结论和展望

本文利用北京市 2014 年 4 月 11 日~17 日连续一个星期的 SCD 和朝阳区 POI 数据,构建了一个城市街区功能识别方法,得到 12 个类别的功能街区。利用遥感影像数据对街区功能识别结果进行检验,结果表明这一方法方法具有较好的准确性。

这一方法具有如下优势:第一,综合利用 SCD 的客流功能属性特征和 POI 数据的物化功能属性,规避了传统的通过土地利用类型来进行城市功能分区方法不能体现人口密度属性和客流功能特征属性的问题;第二,相对于土地利用数据,SCD 和 POI

数据具有处理简单且成本较低的特点,能够更方便地推广应用。

当然,由于数据的可获取性,本文只考虑了公交这一出行方式。事实上,2013年北京公交出行的比例仅为28.60%。数据取样方面存在的偏差可能在一定程度上影响结论的准确性。未来在数据允许的情况下可以将轨道交通和其他出行方式的数据纳入考虑,使结论更为精确。

参考文献

李炬(2013):《西安市明城区街区形态的类型化基础研究》,西安:西安建筑科技大学硕士学位论文。

龙瀛、茅明睿、毛其智等(2014):《大数据时代的精细化城市模拟:方法、数据和案例》,《人文地理》第3期,第7~13页。

龙瀛、张宇、崔承印(2012):《利用公交刷卡数据分析北京职住关系和通勤出行》,《地理学报》第10期,第1339~1352页。

秦萧、甄峰、熊丽芳等(2013):《大数据时代城市空间行为研究方法》,《地理科学进展》第9期,第1352~1361页。

秦萧、甄峰、朱寿佳等(2014):《基于网络口碑度的南京城区餐饮业空间分布格局研究——以大众点评网为例》,《地理科学》第7期,第810~817页。

王波、甄峰、席广亮等(2013):《基于微博用户关系的网络信息地理研究:以新浪微博为例》,《地理研究》第2期,第380~391页。

王波、甄峰、张浩(2015):《基于签到数据的城市活动时空间动态变化及区划研究》,《地理科学》第2期,第151~160页。

杨振山、龙瀛、Nicolas DOUAY(2015):《大数据对人文—经济地理学研究的促进与局限》,《地理科学进展》第4期,第410~417页。

张玲(2012):《POI的分类标准研究》,《测绘通报》第10期,第82~84页。

甄峰、王波、陈映雪等(2012):《基于网络社会空间的中国城市网络特征:以新浪微博为例》,《地理学报》第8期,第1031~1043页。

Chatterjee, S. and S. Das (2015), "Ant Colony Optimization Based Enhanced Dynamic Source Routing Algorithm for Mobile Ad-Hoc Network", *Information Sciences*, 295 (4), pp. 67 - 90.

Fernandes, C. M., A. M. Mora and J. J. Merelo, et al. (2014), "KANTS: A Stigmergic Ant Algorithm for Cluster Analysis and Swarm Art", *IEEE Transactions on Cybernetics*, 44 (6), pp. 843 - 856.

Guo, Y., Y. Li and Z. Shao (2015), "An Ant Colony-Based Text Clustering System with Cognitive Situation Dimensions", *International Journal of Computational Intelligence Systems*, 8 (1), pp. 138 - 157.

Hou, J., W. Mi and L. Li (2014), "Spatial Quality Evaluation for Drinking Water Based on GIS and Ant Colony Clustering Algorithm", *Journal of Central South University*, 21 (3), pp. 1051 - 1057.

Pan, G., G. Qi and Z. Wu, et al. (2013), "Land-Use Classification Using Taxi GPS Traces", *IEEE Transactions on Intelligent Transportation Systems (T-ITS)*, 14 (1), pp. 113 - 123.

Analysis of Function Identification of Urban Blocks Based on SCD and POI Data

——A Case Study of Chaoyang District

ZHONG Shao-ying¹, YUE Wei-zhen¹, ZHANG Yun²

(1. Institute of Science and Development, Chinese Academy of Science, Beijing 100190, China;

2. City Development Institute, Beijing City University, Beijing 100184, China)

Abstract: Using the SCD (Smart Card Data) and POI (Point of Interest) data, the authors identify the functional attribute of urban blocks. The SCD is used to describe the characteristics of pedestrian flow, while the POI data is used to describe the functional properties of the block facilities. First of all, this study makes a preliminary classification of urban blocks by using the functional attributes of public transit stations. On this basis, using the POI data and ant colony clustering algorithm, the authors carry on the further subdivision of urban blocks function. Finally, the recognition results are examined by remote sensing image analysis. The examined results indicate that the complex utilization of the SCD and POI data can precisely describe the functions of urban blocks.

Key Words: urban function identification; SCD; POI; urban blocks

责任编辑：丛晓男