

基于人群出行行为轨迹的城市功能区识别

凌 鹏¹, 诸彤宇¹, 周 轶², 吴爱枝², 张 鹏²

(1.北京航空航天大学 软件开发环境国家重点实验室,北京 100191; 2.北京市安全生产科学技术研究院,北京 101101)

摘 要:城市功能区识别对于城市规划和管理具有重要的支撑作用,目前大部分研究主要依赖于影像和兴趣点(POI)分布数据进行识别,但多将区域内不同出行行为的人群混杂在一起,没有考虑不同群体对区域产生的不同影响。结合物以类聚、人以群分的思想构建城市功能区识别模型UFAl,通过学习不同功能区人群出行活动的特征识别相应功能区。基于大样本粗粒度的匿名轨迹数据,刻画并提取个体出行特征,依据个体的出行特征划分人群类型。在此基础上,构建并训练多任务深度学习模型,实现城市功能区识别。选取北京市2 000万匿名用户10个月的手机信令数据作为人群出行轨迹数据,使用UFAl模型进行计算,并与决策树、随机森林、集成学习梯度提升决策树等7种传统分类模型进行对比。实验结果表明,UFAl模型的F1值达到0.95,与对比模型相比提升了0.10~0.29,具有更好的识别性能。

关键词:城市功能区;时空数据;行为轨迹;城市感知;深度学习

开放科学(资源服务)标志码(OSID):



中文引用格式:凌鹏,诸彤宇,周轶,等.基于人群出行行为轨迹的城市功能区识别[J].计算机工程,2022,48(7):36-41.

英文引用格式:LING P, ZHU T Y, ZHOU Y, et al. Urban functional areas identification based on crowd travel behavior trajectory[J]. Computer Engineering, 2022, 48(7): 36-41.

Urban Functional Areas Identification Based on Crowd Travel Behavior Trajectory

LING Peng¹, ZHU Tongyu¹, ZHOU Yi², WU Aizhi², ZHANG Peng²

(1.State Key Laboratory of Software Development Environment, Beihang University, Beijing 100191, China;

2.Beijing Academy of Safety Science and Technology, Beijing 101101, China)

[Abstract] Urban functional areas identification plays an important supporting role in urban planning and management. Most studies mainly rely on images and Point of Interest(POI) distribution data for identification, but mostly mix people with different travel behaviors in the region, without considering the different effects of different groups on the region. Basis that birds of a feather flock together and people flock together, an urban functional areas identification model, UFAl, is constructed to identify the corresponding functional areas by learning the different characteristics of people's travel activities in different functional areas. Based on large sample, coarse-grained anonymous trajectory data, individual travel characteristics are characterized and extracted, and population types are divided according to individual travel characteristics. A multitask deep learning model is constructed and trained to identify urban functional areas. The 10-month mobile signaling data of twenty million anonymous users in Beijing were selected as the crowd travel trajectory data, calculated by the UFAl model, and compared with seven traditional classification models, such as the decision tree, random forest, and integrated learning gradient boosting classifier. The results show that the F1 value of the UFAl model reaches 0.95, which is improved by 0.10~0.29, compared with the comparison model, and has better recognition performance.

[Key words] urban functional area; spatio-temporal data; behavior trajectory; urban perception; deep learning

DOI:10. 19678/j. issn. 1000-3428. 0061688

0 概述

目前,全球一半以上的人口居住在城市中,而且城市人口规模仍在不断扩大,到2030年,全球城市

人口数量估计会增长到50亿^[1]。因此,对如此庞大的城市人口和有限的城市地区进行管理非常重要。

城市功能区识别是进行合理城市规划和管理的关键一环。功能区是一个文化概念,描述了某个地

基金项目:北京市科技计划(Z181100009018010)。

作者简介:凌 鹏(1997—),男,硕士研究生,主研方向为时空数据处理;诸彤宇,副教授;周 轶、吴爱枝,高级工程师;张 鹏,工程师。

收稿日期:2021-05-19 **修回日期:**2021-06-19 **E-mail:** opbird@buaa.edu.com

区的人类活动,能够反映城市的复杂空间分布和社会经济功能^[2]。每个功能区在空间上都由不同的地理对象聚合,并从土地用途上进行语义抽象^[3]。功能区是城市规划的基本单位^[4-5],对城市交通、资源管理、工厂选址等都有重要影响^[6-7]。传统的城市功能区识别研究主要采用调研统计的方法,通过对人们出行日志的记录和问卷调查等方式进行。这种方法具有很大的局限性,不仅需要大量的人力、物力、财力,效率低下,而且调查者容易将主观意见和看法带入研究结果^[8]。

遥感影像数据可以很好地捕获陆地表面的自然外观,因此,大量研究者用遥感影像数据来进行功能区识别^[9-11]。但仅使用遥感图像进行区域功能识别存在不足:一方面,遥感数据只能反映地表的自然属性,而城市功能区具有社会经济属性,并由相关的人类活动决定;另一方面,遥感数据的获取需要巨大的耗费,且高密度城市中众多高层建筑存在阴影,这对遥感图像处理带来了巨大挑战。

公共交通工具(公共汽车、地铁、出租车等)能够产生大量与人移动强相关的位置数据用于功能区识别^[12-13]。QIAN等^[14]提出了一个集成模型,先基于出租车轨迹中的上下车点,使用K-Means和k最近邻算法提取区域社交属性,再基于决策树算法融合遥感数据得到功能区分。但公交数据会忽视行人对于城市功能区的影响,也会忽略路网未覆盖的区域。

兴趣点(Points of Interest, POI)数据作为城市设施的代表,被广泛应用于城市功能区提取。YUAN等^[15]使用基于主题的推断模型来推断每个区域的功能,该模型将区域视为文档,将功能视为主题,将POI的类别视为元数据,将人类流动模式作为词语。POI数据以建筑物的功能属性出发,能够覆盖所有区域,但并未考虑人的社会活动属性,同时数据更新成本较大。

手机呼叫详细记录(Call Detail Records, CDR)数据间接记录了人的活动时间、空间信息,同时隐含了人的社会活动属性^[16-17],数据获取成本较低,且支

持实时更新。江贵林等^[18]使用高斯混合模型,基于CDR数据的简单统计量设计多特征加权判决的功能区识别算法。TU等^[19]先基于手机信令数据推断人的职住位置,再基于隐马尔可夫模型由社交媒体签到数据获取人的活动的知识,从而推断城市功能。JIA等^[20]自定义融合规则,同时使用遥感影像数据与CDR数据。但目前基于手机数据的大部分研究都只用了简单的统计量,如区域内不同时间段的通话量等。

现有基于出行信息的功能区识别研究大部分停留在简单的统计,且多数将区域内不同出行行为的人群混杂在一起,并没有考虑不同群体对区域产生的不同影响。笔者通过研究相关数据发现:不同功能区的人群出行活动具有各自的特征,而通过学习这些特征可以识别出相应的功能区。本文基于人群出行行为轨迹构建城市功能区识别模型UFAI。利用粗粒度的匿名个体移动位置数据提取隐含的个体出行特征,并通过将这些出行特征与所在局部区域相关联,对该区域的群体出行特征进行分类。在此基础上,通过训练功能区的多分类深度学习模型,完成对功能区的识别。

1 UFAI模型

1.1 模型框架与符号定义

UFAI模型框架如图1所示,主要分为3个模块:

1)个体出行特征提取模块。对原始数据进行数据预处理后,使用ST-DBSCAN算法^[21]识别用户停留点,然后构建出行链,提取出用户的职住位置以及隐含的个体出行特征。

2)区域内人群出行行为刻画模块。以区域为研究对象,基于个体出行特征划分区域内的不同群体,提取出各类人群的停留特征和居民的活动特征。

3)功能区分类模块。首先人工标注高置信度样本,然后以时间周期为单位切分数据,从而扩充样本,构建并训练深度学习模型。

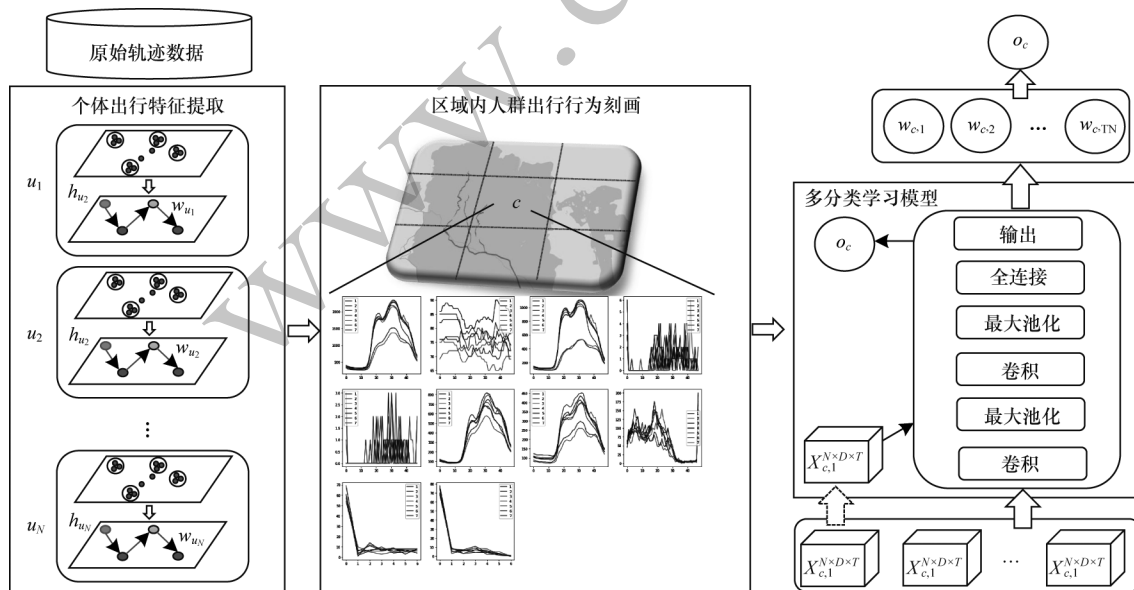


图1 UFAI模型框架

Fig.1 UFAI model framework

本文所使用符号定义如表1所示。

表1 相关符号定义

Table 1 Definition of related symbols

| 符号 | 定义 |
|------------|--|
| U | 全部用户的集合 |
| C | 全部区域的集合 |
| u_{id} | 用户的唯一标识 |
| δ_L | 人群停留时长长短的区分阈值 |
| δ_D | 本地常驻人群和非常驻人群的区分阈值 |
| τ | 时间采样粒度,时间片 t 代表的时间范围是 $[\tau \times t, \tau \times (t+1)]$ |

1.2 个体出行特征提取

一个用户 u_{id} 对应的全部原始轨迹记录集为 R_{id} , 如式(1)所示:

$$R_{id} = \{r_{id}^1, r_{id}^2, \dots, r_{id}^n\} \quad (1)$$

其中: r_{id}^i 是 R_{id} 的第 i 条记录; r_{id}^i 是 $\langle loc, ts, te \rangle$ 三元组, loc 代表记录所在位置, ts 、 te 代表记录开始时刻和记录结束时刻。

1.2.1 停留点识别

通过对聚类算法的研究可知,密度聚类算法 DBSCAN 比较适用于停留点提取,但传统的 DBSCAN 算法只考虑了空间这一单一维度,并不适用于处理具有多个维度的数据,因此,本文引入面向高维数据的 ST-DBSCAN 算法。ST-DBSCAN 算法相对 DBSCAN 主要有两点改进:时间阈值的加入与选择样本邻域方法的改变。时间阈值限制了簇集中样本点的最短时间跨度值,而选择样本邻域的方法由随机扩展方式改为按时序扩展方式,能够保证一个簇集中的样本点在时间上是连续的。

从手机 CDR 数据中获取用户的停留点集合,如式(2)和式(3)所示:

$$P_{id} = \{p_{id}^1, p_{id}^2, \dots, p_{id}^n\} \quad (2)$$

$$P_{id} = \text{stdbscan}(R_{id}, \varepsilon_S, \varepsilon_T) \quad (3)$$

其中: P_{id} 是一个用户对应的所有停留点的集合,集合的第 i 个停留点为 p_{id}^i , p_{id}^i 是 $\langle ts, te, loc \rangle$ 三元组; stdbscan 代表 ST-DBSCAN 聚类算法; ε_S 、 ε_T 分别是 ST-DBSCAN 聚类算法中的空间阈值和时间阈值。

1.2.2 职住位置提取

基于停留点集合 P_{id} 可以提取 u_{id} 在各个位置对应的所有停留时段。计算各个位置在居家时间段的累计停留时长,并取累计停留时长最长的停留点作为居住地 h_{id} 。同时,计算各个位置在工作时间段的累计停留时长,并取累计停留时长最长的停留点作为工作地 w_{id} 。上述计算过程如式(4)和式(5)所示:

$$h_{id} = \underset{p_{id}^i, loc \in C}{\operatorname{argmax}} \left\{ \sum_{p_{id}^i \in P_{id}} f_T[(p_{id}^i.ts, p_{id}^i.te) \cap d_H] \right\} \quad (4)$$

$$w_{id} = \underset{p_{id}^i, loc \in C}{\operatorname{argmax}} \left\{ \sum_{p_{id}^i \in P_{id}} f_T[(p_{id}^i.ts, p_{id}^i.te) \cap d_W] \right\} \quad (5)$$

其中: d_H 和 d_W 分别为居家时间段和工作时间段; f_T 是一个函数,用于计算任意两个时间段交集的时间长度。

1.3 区域内人群出行行为刻画

1.3.1 区域内的人群分类

以往针对区域内人群出行行为的研究,都是以区域内的所有人群为研究对象,但单个区域内的人员成分复杂,不同人群的出行行为也互不相同。因此,本文先将单个区域 c 内的人群分为 8 个研究类别,如表 2 所示。

表2 区域内人群研究类别定义

Table 2 Categories definition of population research in the region

| 符号 | 定义 |
|---------|------------------------|
| G_1^c | 经过区域 c 的所有人群 |
| G_2^c | 区域 c 内的居民 |
| G_3^c | 区域 c 内的职工 |
| G_4^c | 区域 c 内的职住一体人群 |
| G_5^c | 在区域 c 内偶尔经过的本地常驻人群 |
| G_6^c | 在区域 c 内偶尔经过的外地人群 |
| G_7^c | 在区域 c 内停留时长较长的本地常驻人群 |
| G_8^c | 在区域 c 内停留时长较长的外地人群 |

对于用户 u_{id} , 职住位置为 h_{id} 和 w_{id} , 则其在区域 c 内的停留时长 l_{id}^c 计算如式(6)所示:

$$l_{id}^c = \sum (p_{id}^i.te - r_{id}^i.ts), \forall p_{id}^i \in P_{id} \wedge (p_{id}^i.loc \in c) \quad (6)$$

用户在整个研究范围内的停留天数如式(7)所示:

$$d_{id} = |\{p_{id}^i.ts.date | \forall p_{id}^i \in P_{id}\}| \quad (7)$$

其中: $date$ 代表所取停留时段对应的日期。由此可得各类人群职住位置的计算公式如式(8)~式(15)所示:

$$G_1^c = \{u_{id} \in U | p_{id}^i.loc \in c, \forall p_{id}^i \in P_{id}\} \quad (8)$$

$$G_2^c = \{u_{id} \in U | h_{id} \in c\} \quad (9)$$

$$G_3^c = \{u_{id} \in U | w_{id} \in c\} \quad (10)$$

$$G_4^c = \{u_{id} \in U | h_{id} \in c \wedge w_{id} \in c\} \quad (11)$$

$$G_5^c = \{u_{id} \in U | (l_{id}^c < \delta_L) \wedge (d_{id} > \delta_D)\} \quad (12)$$

$$G_6^c = \{u_{id} \in U | (l_{id}^c < \delta_L) \wedge (d_{id} < \delta_D)\} \quad (13)$$

$$G_7^c = \{u_{id} \in U | (l_{id}^c > \delta_L) \wedge (d_{id} > \delta_D)\} \quad (14)$$

$$G_8^c = \{u_{id} \in U | (l_{id}^c > \delta_L) \wedge (d_{id} < \delta_D)\} \quad (15)$$

1.3.2 区域内各类人群的停留特征

区域内不同人群的出行行为在相同时间范围内具有不同的人数变化趋势,因此,先以 τ 为采样粒度,将一天划分为间隔相等的时间片序列,可以计算得到区域一天各时间片的各类人群的人数,最后将连续多天的结果进行拼接,用以刻画区域内各类人群的出行行为。

对于区域 c , 在第 k 个时间周期的第 d 天, 时间片 t 对应 g 类人群的数量如式(16)所示:

$$N_{c,k}^{g,d,t} = |\{u_{id} \in G_g^c | (p_{id}^i.loc \in c) \wedge (p_{id}^i.ts < \tau \times (t+1)) \wedge (p_{id}^i.te > \tau \times t), \forall p_{id}^i \in P_{id}\}| \quad (16)$$

由此,可得到各个区域的人群出行行为特征矩阵 N 。

1.3.3 区域内居民的活动特征刻画

对于区域 c 内的居民 u_{id} , 可以计算出时间周期 k 的第 d 天的最大活动半径 $a_{c,k}^{id,d}$ 和停留时长加权活动半径 $w_{c,k}^{id,d}$, 由于计算出的半径是连续量, 因此还需要进行离散化, 如式 (17) 和式 (18) 所示:

$$a_{c,k}^{id,d} = f_D(\max(\text{dis}(p_{id}^i, \text{loc}, h_{id})))$$

$$\text{s.t.}(\forall p_{id}^i \in P_{id}) \wedge (h_{id} \in c) \quad (17)$$

$$w_{c,k}^{id,d} = f_D\left(\frac{\sum(\text{dis}(p_{id}^i, \text{loc}, h_{id}) \times (p_{id}^i \cdot \text{te} - p_{id}^i \cdot \text{ts}))}{\sum(p_{id}^i \cdot \text{te} - p_{id}^i \cdot \text{ts})}\right)$$

$$\text{s.t.}(\forall p_{id}^i \in P_{id}) \wedge (h_{id} \in c) \quad (18)$$

其中: dis 为计算两点间的直线距离的函数; f_D 为分段离散函数。

f_D 将输入的任意原始连续值 x 映射为 t 段离散值中的一个离散类别 r :

$$f_D(x) = r, \exists r \in \mathbb{Z} \wedge (r \leq T) \wedge (\kappa_{r,1} \leq x < \kappa_{r,2}) \quad (19)$$

其中: $\kappa_{r,1}$ 和 $\kappa_{r,2}$ 分别代表离散类别 r 的范围上界和下界, 离散后的类别数与一天时间片个数相同。

对于区域 c , 在时间周期 k 的第 d 天, 半径离散类别 r 对应的值如式 (20) 和式 (21) 所示:

$$A_{c,k}^{r,d} = |\{u_{id} \in U \mid (h_{id} \in c) \wedge (ra_{c,k}^{id,d} == r)\}| \quad (20)$$

$$W_{c,k}^{r,d} = |\{u_{id} \in U \mid (h_{id} \in c) \wedge (rd_{c,k}^{id,d} == r)\}| \quad (21)$$

由此, 可得到各个区域的居民出行行为特征矩阵 A 和 W 。

1.4 功能区分类模型

区域人群出行行为刻画模型基于粗粒度轨迹数据, 从出行行为角度出发计算得到区域的人群出行行为特征和居民出行行为特征。

功能区分类模型先拼接两类特征矩阵, 得到模型的输入特征矩阵 X 。由经验可知, 一周为一个人类活动的周期, 因此, 本文选取一周七天为区域的一个时间周期, 构建区域 c 第 k 个时间周期输入特征的方式如式 (22) 所示:

$$X_{c,k}^{10 \times d \times t} = \{N_{c,k}^{8 \times d \times t}, A_{c,k}^{1 \times d \times t}, W_{c,k}^{1 \times d \times t}\} \quad (22)$$

其中: t 代表一天的时间片个数; d 代表一个时间周期的天数; $N_{c,k}^{8 \times d \times t}$ 代表区域内的 8 类人群的停留特征; $A_{c,k}^{1 \times d \times t}$ 和 $W_{c,k}^{1 \times d \times t}$ 代表区域内的居民活动特征; $X_{c,k}^{10 \times d \times t}$ 为拼接后的输入特征。当存在多个时间周期时, 一个区域 c 会包含多个 $X_{c,k}^{10 \times d \times t}$ 。

1.4.1 样本集扩充

结合实际经验和数据分析, 先人工标记部分功能区, 作为训练和测试数据。但由于功能区数量本身有限, 导致有标签的功能区数量有限, 因此需要扩充样本。

由于数据在同一个区域的不同时间周期上遵循同一个概率分布, 而且时间周期是可以不断扩展的, 因此本文将一个区域对应的多个时间周期数据切分为多个训练样本, 从而扩充样本集。

1.4.2 模型训练与分类结果

对于任意区域 c , 时间周期 k 对应的特征矩阵 $X_{c,k}^{10 \times d \times t}$ 是一个三维张量, 经分析发现其时空分布均匀, 类似图像数据, 因此, 本文采用类似卷积神经网络的结构构建多分类模型。

对于输入 $x_i^0 \in X$, 经过卷积后输出 y_j^1 :

$$y_j^1 = \sigma\left(\sum_{i \in M_j} x_i^0 \cdot k_{i,j}^1 + b_j^1\right) \quad (23)$$

其中: l 代表层数; k 为卷积核; M_j 表示选择的输入特征图的集合。每个输出特征图会给一个额外的偏置 b , 然后进行最大池化:

$$x_j^1 = \max_{i \in N \times N} (y_i^0 \cdot u(n, n)) \quad (24)$$

其中: $u(n, n)$ 为输入窗口函数。在此基础上, 进行第 2 次卷积和池化:

$$y_j^2 = \sigma\left(\sum_{i \in M_j} x_i^1 \cdot k_{i,j}^2 + b_j^2\right) \quad (25)$$

$$x_j^2 = \max_{i \in N \times N} (y_i^1 \cdot u(n, n)) \quad (26)$$

最后, 将 x_j^2 ($j = 1, 2, \dots, 10$) 顺序展开成向量, 并有序连接为一个长向量 x_{fold} , 作为全连接层的输入, 由此得到单个周期的分类结果 $y_{c,k}$:

$$y_{c,k} = W x_{\text{fold}} \quad (27)$$

其中: W 为全连接层的权重矩阵。

对于同一个区域, 当输入多个时间周期时, 得到的分类结果可能会有所不同。出现这种情况的原因可能是该区域的功能分类复杂, 也可能是数据采集的问题。因此, 本文选取所有时间周期对应的分类结果中出现次数最多的结果, 作为区域的功能区分类的最终结果。在第 k 个时间周期, 区域 c 的输入特征矩阵为 $X_{c,k}$, 模型的分类结果为 $y_{c,k}$, 获取区域的最终结果 o_c 的过程如式 (28) 所示:

$$o_c = f_M(y_{c,1}, y_{c,2}, \dots, y_{c,K}) \quad (28)$$

其中: f_M 用于返回一个序列的众数。

2 实验与结果分析

2.1 实验数据集

本文实验所用的数据集为某移动运营商提供的北京市 2019 年 1 月—2019 年 10 月的手机信令数据, 覆盖北京市范围内 2 000 万匿名手机用户, 数据采样频率为半小时, 主要字段说明如表 3 所示。

表 3 数据字段说明

Table 3 Data field description

| 字段名 | 字段说明 |
|-------|---------------------------------------|
| uid | 用户 id |
| stime | 事件开始时间 |
| etime | 事件结束时间 |
| loc | 事件发生位置, 基于基站生成误差 300 m~500 m 的粗粒度位置坐标 |

经过数据去噪后,本文选取数据质量好、无重大节日影响且具有典型日常生活特征的数据作为UFAI模型的输入。

2.2 实验设置

本文以北京市作为研究区域,以250 m×250 m为精度将其划分为18 106个网格区域。通过参考《北京市土地利用总体规划(2006—2020年)》《北京市主体功能区规划》以及网络地图,选取250个区域作为训练集,100个区域作为测试集,经过样本扩充,训练集个数最终为1 250个,功能区类别有居住区、工作区、其他。同时,本文的多分类模型采用交叉熵损失函数,如式(29)所示:

$$\text{loss}(x, \text{class}) = -\log_a \left(\frac{\exp(x[\text{class}])}{\sum_j \exp(x[j])} \right) \quad (29)$$

2.3 实验结果分析

采用本文模型进行训练,实验结果如表4所示。

表4 各功能区分类结果

| Table 4 Classification results of each functional area | | | |
|--|------|------|------|
| 类别 | 准确率 | 召回率 | F1值 |
| 居住区 | 0.95 | 1.00 | 0.97 |
| 工作区 | 0.97 | 0.91 | 0.94 |
| 其他 | 0.93 | 0.95 | 0.94 |

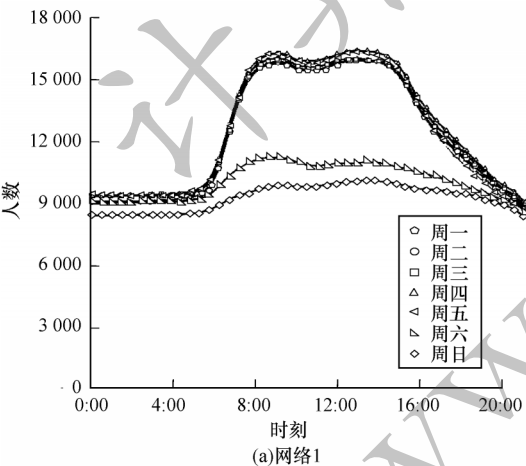


图2为某局部区域对应的功能区分类结果,从中随机选取多个网格进行数据分析,如图3所示,其中横坐标为一天48个时间片,纵坐标为各时间片对应人数,不同线条代表一周七天。可以看出:网格①平时人数多于周末人数,且每天人数趋势符合“凸”形;网格②平时人数少于周末,且每天人数区域符合“凹”形。以上数据表现符合日常生活经验。

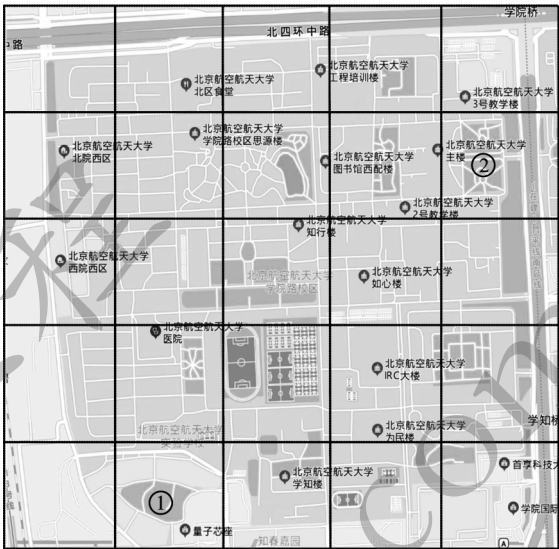


图2 某局部区域功能区分类结果
Fig.2 Classification result of functional area in a local area

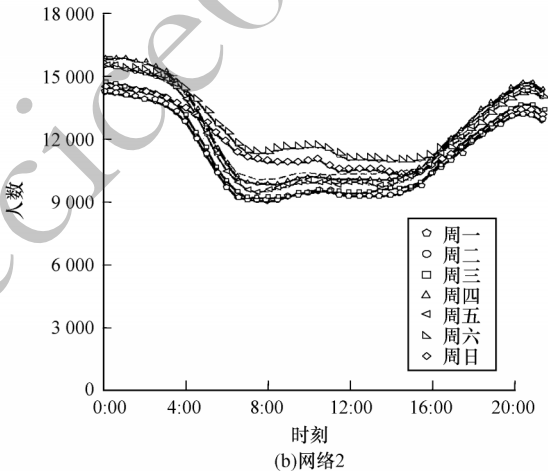


图3 一周七天网格内总人数随时间变化曲线

Fig.3 Curve of the total number of people over time in the daily grid in a week

实验使用以下对比模型:
1)决策树分类模型(Decision Tree)^[22]。
2)随机森林模型(Random Forest)^[23]。
3)核函数选择线性函数(SVM)^[24]。
4)多项式朴素贝叶斯算法(Multinomial NB)^[25]。
5)K最近邻分类算法(KNN)^[26]。
6)逻辑回归算法(Logistic Regression)^[27]。
7)集成学习梯度提升决策树分类模型(Gradient Boosting Classifier)^[28]。
对比结果如表5所示,与其他模型相比,本文UFAI模型具有最好的性能,F1值达到0.95。

表5 对比实验结果

| Table 5 Comparison experimental results | | | |
|---|------|------|------|
| 模型 | 准确率 | 召回率 | F1值 |
| Decision Tree | 0.67 | 0.66 | 0.66 |
| Random Forest | 0.83 | 0.85 | 0.85 |
| SVM | 0.82 | 0.82 | 0.82 |
| Multinomial NB | 0.75 | 0.75 | 0.72 |
| KNN | 0.79 | 0.81 | 0.80 |
| Logistic Regression | 0.79 | 0.79 | 0.78 |
| Gradient Boosting Classifier | 0.81 | 0.84 | 0.82 |
| UFAI | 0.93 | 0.92 | 0.95 |

3 结束语

本文基于粗粒度匿名个体移动位置数据构建城市功能区识别模型UFAI。挖掘个体出行特征并与局部区域结合, 将人群划分为不同类别。在此基础上, 通过刻画各个区域内不同人群的出行活动, 构建功能区多分类深度学习模型, 同时将多个时间周期数据划分为多个训练样本, 从而扩充样本集。实验结果表明, UFAI模型识别准确率达到0.93, 相比于决策树、随机森林等分类模型准确率更高。下一步将结合更多识别功能区的人群出行活动特征, 扩大本文模型的识别范围。

参考文献

- [1] TU W, HU Z W, LI L F, et al. Portraying urban functional zones by coupling remote sensing imagery and human sensing data[J]. *Remote Sensing*, 2018, 10(1): 141.
- [2] BAO H Q, MING D P, GUO Y, et al. DFCNN-based semantic recognition of urban functional zones by integrating remote sensing data and POI data[J]. *Remote Sensing*, 2020, 12(7): 1088.
- [3] ZHANG X Y, DU S H, WANG Q. Hierarchical semantic cognition for urban functional zones with VHR satellite images and POI data[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2017, 132: 170-184.
- [4] MATSUOKA R H, KAPLAN R. People needs in the urban landscape: analysis of landscape and urban planning contributions[J]. *Landscape and Urban Planning*, 2008, 84(1): 7-19.
- [5] MONTANGES A P, MOSER G, TAUBENBÖCK H, et al. Classification of urban structural types with multisource data and structured models[C]//*Proceedings of JURSE'15*. Washington D. C., USA: IEEE Press, 2015: 1-4.
- [6] HEIDEN U, HELDENS W, ROESSNER S, et al. Urban structure type characterization using hyperspectral remote sensing and height information[J]. *Landscape and Urban Planning*, 2012, 105(4): 361-375.
- [7] FAN J, TAO A J, REN Q. On the historical background, scientific intentions, goal orientation, and policy framework of major function-oriented zone planning in China[J]. *Journal of Resources and Ecology*, 2010, 1(4): 289-299.
- [8] 陈占龙, 周路林, 禹文豪, 等. 顾及兴趣点潜在上下文关系的城市功能区识别[J]. *测绘学报*, 2020, 49(7): 907-920.
- [9] CHEN Z L, ZHOU L L, YU W H, et al. Identification of the urban functional regions considering the potential context of interest points[J]. *Acta Geodaetica et Cartographica Sinica*, 2020, 49(7): 907-920. (in Chinese)
- [10] CAO R, TU W, YANG C X, et al. Deep learning-based remote and social sensing data fusion for urban region function recognition[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 163: 82-97.
- [11] XIA J M, DING Y, TAN L. Urban remote sensing scene recognition based on lightweight convolution neural network[J]. *IEEE Access*, 2021, 9: 26377-26387.
- [12] DU S J, DU S H, LIU B, et al. Large-scale urban functional zone mapping by integrating remote sensing images and open social data[J]. *GIScience & Remote Sensing*, 2020, 57(3): 411-430.
- [13] JIANG Z, EVANS M, OLIVER D, et al. Identifying K primary corridors from urban bicycle GPS trajectories on a road network[J]. *Information Systems*, 2016, 57: 142-159.
- [14] ZHANG F S, JIN B H, WANG Z Y, et al. On geocasting over urban bus-based networks by mining trajectories[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2016, 17(6): 1734-1747.
- [15] QIAN Z, LIU X T, TAO F, et al. Identification of urban functional areas by coupling satellite images and taxi GPS trajectories[J]. *Remote Sensing*, 2020, 12(15): 2449.
- [16] YUAN J, ZHENG Y, XIE X. Discovering regions of different functions in a city using human mobility and POIs [C]//*Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, USA: ACM Press, 2012: 186-194.
- [17] GONZÁLEZ M C, HIDALGO C A, BARABÁSI A L. Understanding individual human mobility patterns[J]. *Nature*, 2008, 453(7196): 779-782.
- [18] PEI T, SOBOLEVSKY S, RATTI C, et al. A new insight into land use classification based on aggregated mobile phone data [J]. *International Journal of Geographical Information Science*, 2014, 28(9): 1988-2007.
- [19] 江贵林, 胡访宇, 石立兴. 基于呼叫详细记录数据的城市功能区识别[J]. *计算机应用*, 2016, 36(7): 2046-2050.
- [20] JIANG G L, HU F Y, SHI L X. Urban functional area identification based on call detail record data[J]. *Journal of Computer Applications*, 2016, 36(7): 2046-2050. (in Chinese)
- [21] TU W, CAO J Z, YUE Y, et al. Coupling mobile phone and social media data: a new approach to understanding urban functions and diurnal patterns[J]. *International Journal of Geographical Information Science*, 2017, 31(12): 2331-2358.
- [22] JIA Y X, GE Y, LING F, et al. Urban land use mapping by combining remote sensing imagery and mobile phone positioning data[J]. *Remote Sensing*, 2018, 10(3): 446.
- [23] BIRANT D, KUT A. ST-DBSCAN: an algorithm for clustering spatial-temporal data[J]. *Data & Knowledge Engineering*, 2007, 60(1): 208-221.
- [24] LEO B. Random forests[J]. *Machine Learning*, 2001, 45(1): 5-32.
- [25] PRAAGMAN J. Classification and regression trees; Leo BREIMAN, Jerome H. FRIEDMAN, Richard A. OLSHEN and Charles J. STONE The Wadsworth Statistics/Probability Series, Wadsworth, Belmont, 1984, x + 358 pages[J]. *European Journal of Operational Research*, 1985, 19(1): 144.
- [26] ELOMAA T, MANNILA H, TOIVONEN H. Machine Learning [C]//*Proceedings of the 13th European Conference on Machine Learning*. Berlin, Heidelberg: Springer, 2002: 1-5.
- [27] KIBRIYA A M, FRANK E, PFAHRINGER B, et al. Multinomial Naive Bayes for text categorization revisited [M]//WEBB G I, YU X. AI 2004: advances in artificial intelligence. Berlin, Germany: Springer, 2004: 488-499.
- [28] ZHANG M L, ZHOU Z H. ML-KNN: a lazy learning approach to multi-label learning[J]. *Pattern Recognition*, 2007, 40(7): 2038-2048.
- [29] HOSMER D W, LEMESHOW S, STURDIVANT R X. Applied logistic regression[M]. [S. l.]: Wiley, 2013.
- [30] MASON L, BAXTER J, BARTLETT P, et al. Boosting algorithms as gradient descent in function space[EB/OL]. [2021-05-10]. https://www.researchgate.net/profile/Marcus-Frean/publication/2750572_Boosting_Algorithms_as_Gradient_Descent_in_Function_Space/links/09e4150617c0848d46000000/Boosting-Algorithms-as-Gradient-Desc-ent-in-Function-Space.pdf?origin=publication_detail.