

基于时空数据的城市人流移动模式挖掘



孙天旭¹ 赵蕴龙^{1,2} 练作为¹ 孙毅¹ 蔡月啸¹

¹ 南京航空航天大学计算机科学与技术学院 南京 211106

² 软件新技术与产业化协同创新中心 南京 210023

(sun29@nuaa.edu.cn)

摘要 随着城市的快速发展,城市中人流的管理与移动模式挖掘变得越发重要。同时,随着以群智感知为代表的各种感知技术的发展,提出了智慧城市概念,智慧城市中的大量感知数据为人流的分析提供了可能性。在智慧城市中,时空数据是最为常见的一种数据。本文基于城市中的时空数据,首先提出一种建模方法,将不同种类的时空数据表示为人流模型;然后基于聚类的思想,通过改进传统的基于密度的聚类算法来对人流的移动模式进行挖掘,提出一种人流的移动模式聚类算法:时空密度聚类(Spatio-Temporal Density-Based Spatial Clustering of Applications with Noise, ST-DBSCAN);接着设计了一个移动模式的交通应用场景,并提出对移动模式的评价方法;最后在中国某城市的真实数据集上进行实验与分析,结果表明本文得到的移动模式结果在统一交通服务的场景下可节省25%的交通成本,验证了本文所提移动模式的有效性。

关键词: 智慧城市;时空数据;数据挖掘;移动模式;城市人流

中图分类号 TP311.13

Mobility Pattern Mining for People Flow Based on Spatio-Temporal Data

SUN Tian-xu¹, ZHAO Yun-long^{1,2}, LIAN Zuo-wei¹, SUN Yi¹ and CAI Yue-xiao¹

¹ College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

² Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing 210023, China

Abstract With the accelerating urbanization of many countries, managing people flow and mining mobility patterns become more and more important. Simultaneously, with the development of information technology, especially mobile crowd sensing, the concept of smart city is proposed by many scholars, sensing data in smart cities also provides the possibility for analysis of people flow. In smart city, spatio-temporal data is the most common data. Based on the spatio-temporal data, this paper proposes a modeling method to represent different kinds of spatio-temporal data as people flow model. Then, based on the thinking of clustering, this paper mines mobility pattern from people flow by an improved density-based clustering algorithm, designs a transportation application in smart city, and proposes a method for evaluating the effectiveness of mobility pattern. Finally, experimenting on a real dataset of a city in China and analyzing the results. The results show that the mobility pattern obtained in this paper can reduce costs by 25% in the transportation application of smart city, verifying the effectiveness of the mobility pattern.

Keywords Smart city, Spatio-Temporal data, Data mining, Mobility pattern, People flow

1 引言

现今,随着中国城市化的快速发展,城市人口正急速增加,快速增长的人口给城市的公共资源如公共交通等方面带来了巨大的压力。同时,由于密集人流的不确定性,极易发生交通事故、踩踏等严重事件,影响了城市居民的安全与生活质量。因此城市人流的管理与移动模式挖掘是城市发展的重要环节。随着感知技术特别是群智感知的发展,许多学者提出了智慧城市的概念。城市中各种感知设备所收集到的感知数据被更多地用于城市人流分析,如手机信令数据、交通视频监控、GPS数据、WiFi、居民智能交通卡数据等^[1-2]。这些数据

一般统称为时空数据(Spatio-Temporal Data, ST Data),被应用于智能交通、通信系统与环境科学等多个领域。与一般的文本或图像数据不同,时空数据拥有空间特征与时间特征,其中空间特征包括距离与地理层次等属性,时间特征包括时序、趋势与周期性等属性。由于所含的信息丰富,时空数据为分析城市人流的分布和移动模式提供了可能性。现今智慧城市中的许多应用,如城市的功能区域规划^[3]、定制公交的路线制定^[4]、区域紧急事件的预警与疏散方案规划^[5]等都依赖于对时空数据的挖掘结果。

本文的主要工作如下:

(1)提出一种基于时空数据的人流建模方法,将多种类的

到稿日期:2020-01-02 返修日期:2020-05-16 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国防基础科研计划资助(JCKY2016605B006);江苏省“六大人才高峰”高层次人才(XYDXXJS-031)

This work was supported by the National Defense Basic Scientific Research Program of China (JCKY2016605B006) and Six Talent Peaks Project in Jiangsu Province (XYDXXJS-031).

通讯作者:赵蕴龙(zhaoyunlong@nuaa.edu.cn)

时空数据表示为统一的数学模型,从而为进一步的分析或计算提供基础。

(2)提出了一种基于聚类思想的人流移动模式挖掘方法。本文首先提出一种时空数据的建模方法,然后基于聚类的思想进行人流移动模式的挖掘。具体方式为:将移动模式挖掘问题转换为一个对六维对象聚类的问题,改进传统密度聚类的方法,进而提出一种时空域上的聚类算法来解决问题。

(3)设计了一种城市人流移动模式的应用场景,并在真实数据集上应用本文提出的算法,结合实际地理信息对结果进行分析与评价。实验表明,本文方法得到的结果符合城市人流量,证明了该方法的有效性。

2 相关工作

城市的人流时空分布规律研究可以感知人流随时间的变化趋势,为城市的功能区域规划、定制公交的路线制定、区域紧急事件的预警与疏散方案规划提供了重要的数据支撑,同时,人流移动规律研究作为轨迹数据挖掘的重要组成部分也引起了国内外学术界以及工业界的关注。

在时空数据的来源方面,蓝牙、射频技术、4G LTE等现代通信技术产生的大量数据在人流时空分布研究中扮演着越来越重要的角色^[6]。例如,运营商拥有的手机信令数据就是分析人流时空分布和移动模式的创新性工具之一^[7]。在此基础上,还有学者结合了多源数据来进行研究,如 Itoh 等将智能交通卡数据和社交媒体数据结合,从多个视角探索地铁交通中乘客行为的变化情况及异常情景^[1],这项研究中结合的社交媒体数据可以更好地反映乘客的个体行为特征。同时,许多学者在不同粒度的时空数据上进行了不同的研究,例如 Demissie 等应用信令数据对塞内加尔国家整体的交通分布模式进行建模^[8],同时他们在另一项研究中对交通需求进行了估计^[9],从而对发展较为落后的塞内加尔提出了交通规划建议。而在更细粒度的场景下,Fei 等对公交线路的运行信息进行聚类^[10],划分了几种公交线路运行时的状态,从而根据车辆状态估计城市的交通状况。在人流移动模式分析方面,Zheng 系统地阐述了在轨迹数据挖掘领域得到人流移动模式的主要研究方法和技术^[11]。Liu 等基于城市居民的出行链^[12],通过探究家庭住址和日常移动行为来构建人流移动模式。相似地,Yabe 等结合出行链与灾害数据^[13],分析了不同灾害模式对人流移动模式的影响。Do 等利用起始地-目的地矩阵^[14],使用决策树分类算法对手机数据进行处理来挖掘出行模式。在更高层面的应用中,Yan 等利用人流移动模式提出了一个定制公交线路的规划方法^[4],提高了公交系统的收益;Li 等则根据人流移动模式选取一组广告牌位置,使获得的收益最大化^[15]。

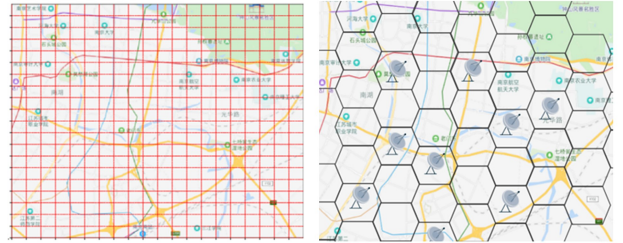
3 基于时空数据的人流建模

城市内有多种感知应用,每个应用所采用的感知设备与方法不尽相同。事实上,各种时空数据在时间和空间两个尺度上的主要属性的含义是一致的,所以本文提出一个统一的模型将不同种类的时空数据表示为统一的人流。

3.1 位置信息网格化方法

目前有许多不同思想、不同粒度下的位置定义方法,其中最常用到的即为经纬度。但考虑到有相当一部分的时空数

据,例如手机信令数据、智能卡刷卡数据所得到的位置信息依赖于基站或公交站点等固定区域位置,本文在位置定义方面采用网格化的方式。网格化即将一片大小为 $a \times b$ km 的城市区域划分为 N 个网格。当网格足够小且以正方形表示时,网格位置信息足以保留细粒度的经纬度信息;当网格尺寸较大或以特殊方式划分时,一些依赖特殊位置信息的数据也便于进行表示。图 1 给出了两个网格划分的例子。



(a) 被细粒度正方形网格化的城市区域 (b) 被基站以蜂窝状网格化的城市区域

图 1 不同应用中的两种网格化方式

Fig. 1 Two ways of gridding in different applications

3.2 基于网格化的建模方法

对于一组时空数据来说,其原始的表达方式通常为在网格间的转移序列 P ,而人流就可以从这样的数组轨迹集合 P 中得到。对于城市网格中第 m 行和第 n 列的一个网格来说,在时间序列中第 k 时刻的人流可以由两组流量表示,本文将将其命名为流入量和流出量,流入量表示在 k 时刻进入网格 (m, n) 的人流量,流出量表示在 k 时刻离开网格 (m, n) 的人流量,这两种流量被表示为:

$$x_k^{in,m,n} = \sum_{T_{rk} \in P} |\{i > 1 | g_{i-1} \notin (m, n) \cap g_i \in (m, n)\}| \quad (1)$$

$$x_k^{out,m,n} = \sum_{T_{rk} \in P} |\{i > 1 | g_{i-1} \in (m, n) \cap g_i \notin (m, n)\}| \quad (2)$$

其中, T_{rk} 表示在 k 时刻的轨迹集合 $g_1 \rightarrow g_2 \rightarrow \dots \rightarrow g_{|T_{rk}|}$, g_i 表示网格坐标, $g_i \in (m, n)$ 表示对象位于网格 (m, n) 中。在第 k 时刻,一个 $M \times N$ 大小网格的流入量和流出量可以被记作 $X_k \in \mathbb{R}^{2 \times M \times N}$,其中 $(X_k)_{0,m,n} = x_k^{in,m,n}$, $(X_k)_{1,m,n} = x_k^{out,m,n}$ 。因此在总时间长度为 Q 的一段时间内,整个区域的人流可以被记作一个张量 $X \in \mathbb{R}^{Q \times M \times N}$ 。

3.3 人流模型的进一步表达方法

在得到城市人流模型后,考虑到在特定场合下例如群智感知的用户交互中有着特定的需求,可以将人流模型进一步表达为图像序列,这样可以使时空数据有更直观的体现。例如,一个 3×3 网格区域在 t 时刻内的流量变化如图 2 所示。

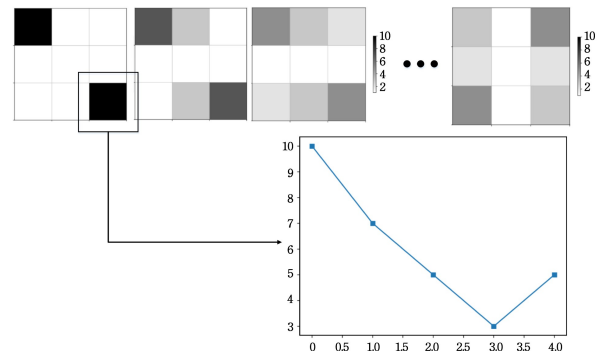


图 2 一组流量变化图像和单个网格中的流量变化

Fig. 2 Traffic changes in single grid from a set of traffic change images

在交通领域的许多应用中,城市人流模型还可以表达为起始地-目的地矩阵(Origin-Destination Matrix, OD 矩阵)。OD 矩阵是记录 N 个区域在某时间段内的流量情况的一个 $N \times N$ 的矩阵,一般以行为起始地,以列为目的地,第 a 行 b 列的值即为由区域 a 流向区域 b 的流量。对于一组城市中的连续轨迹来说,假设它在 t 时刻位于网格 n 内,通常它只对于网格 n 周围的 m 个网格是可达的,而由于传感器感知频率或数据收集周期等问题,低时间敏感度的数据通常会有跨区域的流动,这就导致有时一个网格对于所有其他网格都是可达的,此时 OD 矩阵可以在不丢失信息的情况下更好地表示城市人流的移动情况。部分子区域的 OD 矩阵实例如图 3 所示。

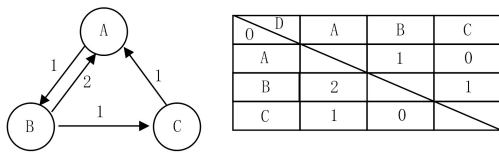


图 3 3 个子区域的 OD 矩阵实例

Fig. 3 Example of OD matrix of three subregions

4 基于聚类的人流移动模式挖掘方法

本节重点介绍在不同领域中,各类时空数据统一建模为人流后,基于“人流是由个体组成的群体结构”这一前提,提出了基于聚类进行人流移动模式挖掘的方法;考虑到传统聚类方法的不足,本文提出了改进的聚类方法来解决移动模式挖掘的问题。

4.1 基于聚类思想的移动模式挖掘

人流由多个个体的移动行为构成,对城市中的人流而言,最重要的特征就是人流的流量、流向与时间。近年来,许多学者,如 Lian 等,将移动模式定义为具有重叠的起源地和目的地区域的移动行为的集合^[16]。基于这样的思想,为了发现移动特征相似度较高的移动行为集合,本文选取了聚类的方法从人流中挖掘移动模式。

在众多的聚类方法中, Ester 提出了基于密度的聚类(Density-Based Spatial Clustering of Applications with Noise, DBSCAN),它将簇定义为样本中密度相连的点的最大集合,旨在发现空间平面上任意形状的高密度聚类。在本文对人流的移动模式进行挖掘的场景中,基于密度的聚类算法比一般的聚类算法更为合适,这是因为它不需要输入固定的簇的个数,且基于邻域的可达性也很适合本文发现相似的移动行为的需求。通常情况下, DBSCAN 算法的参数有两个,分别是邻域半径 c 和核心点邻域内最小个数 $MinPts$ 。事实上,任意两点间的距离的计算方法在算法的使用过程中需要根据不同的场景给出相应的定义。

4.2 基于密度的时空维度聚类算法

考虑到人流在时间和空间维度上的多种特征,传统的 DBSCAN 不足以对人流数据进行聚类,因此本文提出了基于密度的时空维度聚类(Spatial-Temporal Density-Based Spatial Clustering of Applications with Noise, ST-DBSCAN)算法,在定义邻域时,通过输入一组时空邻域半径 (c_s, c_t) 分别从时间和空间尺度上为样本划分领域。算法的相关定义如下。

定义 1(时空邻域) 有样本集 $D = (x_1, x_2, \dots, x_m)$, 对于任意 $x \in D$, 其时空邻域为包含样本集 D 中与 x 的时空距离不大于邻域半径阈值的子样本集, 即:

$$N_{c_s, c_t}(x_i) = \{x_j \mid spatioidist(x_i, x_j) < c_s, \text{temporal}dist(x_i, x_j) < c_t\} \quad (3)$$

定义 2(核心对象) 有样本集 $D = (x_1, x_2, \dots, x_m)$, 对于任意 $x \in D$, 如果其时空邻域 $N_{c_s, c_t}(x_i)$ 内至少包含 $MinPts$ 个样本, 则称 x_i 为核心对象, 即:

$$|N_{c_s, c_t}(x_i)| > MinPts \quad (4)$$

定义 3(时空距离) 有样本集 $D = (x_1, x_2, \dots, x_m)$, 其中 $x = (o_{i,x}, o_{i,y}, d_{i,x}, d_{i,y}, t_i, t_{i,off})$, 其时空距离分别为:

$$spatioidist(x_i, x_j) = \sqrt{\|o_i - o_j\|_2^2 + \|d_i - d_j\|_2^2} \quad (5)$$

$$\text{temporal}dist(x_i, x_j) = |t_{i,o} - t_{j,o}| |t_{i,d} t_{j,d}| \quad (6)$$

其中:

$$\|o_i - o_j\|_2 = \sqrt{(o_{i,x} - o_{j,x})^2 + (d_{i,x} - d_{j,x})^2} \quad (7)$$

根据以上 3 个定义, 本文提出的 ST-DBSCAN 的算法如算法 1 所示。

算法 1 ST-DBSCAN

输入: $(D, c_s, c_t, MinPts)$

输出: Labels

1. 初始化核心对象集合 $cores = \emptyset$, 种子集合 $seeds = \emptyset$, 初始化类别数 $k = 0$, 初始化大小等于样本集大小的标签集合 $Labels = (-1, -1, \dots, -1)$ 。
2. 遍历样本集 D 中的元素, 如果是核心对象, 则将其加入至种子集 $seeds$ 中。
3. 如果核心对象集 $cores$ 中的元素都已经被访问, 则算法结束, 输出结果 Labels; 否则转入步骤 4。
4. 在核心对象集合 $cores$ 中, 随机选择一个未被访问过的核心对象 x_i , 将 x_i 标记为已访问, 然后根据 x_i 在 $cores$ 中的索引对应地将 Labels 中的 l_i 标记为类别 k , 最后将 x_i 的时空邻域内的所有未被访问的对象存放至种子集合 $seeds$ 中。
5. 如果种子集合 $seeds = \emptyset$, 则当前第 k 类生成完毕, $k = k + 1$, 跳转至步骤 3。否则, 从种子集中任选一个种子点 $seed_j$, 将其标记为已访问, 根据索引对应地将 Labels 中的 l_j 标记为类别 k , 判断 $seed_j$ 是否为核心对象, 如果是, 则将 $seed_j$ 的时空邻域中的所有点加入至种子集合 $seeds$ 中。重复步骤 5。

5 实验方案设计

本文所提出的 ST-DBSCAN 通过聚类的方法来挖掘人流的移动模式, 而移动模式的挖掘结果事实上不能直观地从数值上进行比较。为了评价算法的好坏, 本文提出了一个人流移动模式的应用场景, 并在该场景中提出一个对算法的结果进行评价的方法。

5.1 应用场景设计

在现今的城市中, 交通资源是城市资源中的重要组成部分, 人们的出行行为都会占用城市中的交通资源。而对于同一移动模式的人流中的多个移动行为来说, 由于它们的时空特征都很相似, 所以移动所使用的交通资源实际上有很大部分是重复消耗的。在这一前提下, 本文提出一种为同一移动模式中的不同行为个体分配交通资源来进行服务的应用场景, 从而减小城市交通系统的压力。

同一移动模式中的不同移动行为通常有两个主要特点: 1) 不同行为的起点和终点距离较近, 随机地分布在起始区域和终点区域; 2) 起始区域和终点区域之间距离较长, 区域之间没有停止点。基于这样的特点, 本文通过依次为个体分配交通资源, 然后统一提供交通服务的方法来降低交通系统的成本。一个交通路线的规划模型如图 4 所示。事实上, 在智慧城市的发展中, 许多新型的交通应用如定制公交^[17]、拼车服务^[18]等都采取了类似的方法来缓解城市交通系统的压力。

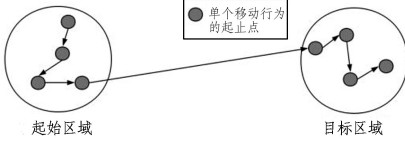


图 4 统一提供的交通服务模式

Fig. 4 Unified transportation service model

在这样的应用场景下, 相比于统一提供交通服务之前, 统一之后的交通服务可以大幅节省不同个体各自进行移动时所消耗的交通资源, 而统一服务产生的额外成本则是将每个个体聚集起来所产生的成本。本文将在下一节对交通资源成本进行建模, 来对移动模式的合理性进行评价。

5.2 基于移动成本的移动模式评价模型

一次交通服务所产生的成本可以分为固定成本和可变成本两种。其中固定成本包括提供服务的车辆本身、驾驶员人工成本等, 本文统称为车次成本; 可变成本包括随路程而增加的成本, 如油耗、对道路产生的交通压力等, 本文统称为路程成本。所以一次交通服务的成本可以建模为:

$$COST_t = c_t + L \times c_{t1} \quad (8)$$

其中, c_t 表示车次成本, c_{t1} 表示单位距离内的路程成本, L 为一次交通服务所经历的距离。而在本文提出的统一交通服务中, 本文将产生的额外成本例如个体等待的耗时和个体间的额外距离等统称为个体成本, 可以建模为:

$$COST_b = c_b + L \times c_{b1} + \sum_i^n c_{b2i} \quad (9)$$

其中, c_{b2i} 表示同一移动模式中所有个体产生的个体成本。在次应用场景中, 本文使用统一分配服务后的成本节省量来对移动模式的准确性进行评价。由式(8)和式(9)可得, 成本节省量可以表示为:

$$\Delta COST = (\sum_i^n t_0 - 1) \cdot c_0 + (t_1 \sum_i^n L_i - L) \cdot c_1 - \sum_i^n c_i \quad (10)$$

其中, c_0 为统一交通服务中的车次成本, t_0 为将 c_0 视为单位成本后, 个体移动产生的车次成本系数; 相似地, c_1 为统一交通服务中的路程成本, t_1 表示个体移动的路程成本系数; c_i 为额外的个体成本。本文将在实验部分应用此成本差公式对城市人流移动模式的挖掘结果进行评价。

6 实验方案设计

6.1 数据设置

本文实验所使用的出租车 GPS 数据收集自中国某城市 100 辆出租车在 2018 年 10 月 1 日到 2018 年 10 月 26 日期间产生的约 120 万条数据, 表 1 展示了此数据集中的一个数据样例。

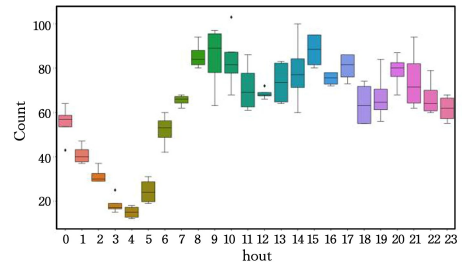
表 1 一个 GPS 数据的样例

Table 1 Sample of GPS data

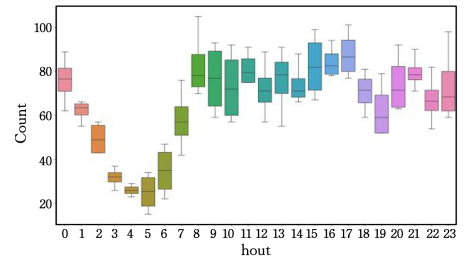
车辆 ID	38801486
时间戳	1539409634
经度	118.75
纬度	32.003
车速	16
状态	空车

6.2 时空特征分析

本文首先对 GPS 数据进行处理, 提取出城市人口的有效出行行为, 然后进行统计, 最后分析城市人流的时空特征。本文首先选取了 2018 年 10 月 8 日至 10 月 21 日两周的每日人流数据, 在时间维度上按工作日和休息日分别统计, 结果如图 5 所示; 在空间维度上绘制热力图, 结果如图 6 所示。



(a) 工作日的城市人流量



(b) 休息日的城市人流量

图 5 两天的整日人流移动情况

Fig. 5 Flow distribution of two days

从图 5 可以看出, 城市人流的移动行为在一天内的时间分布有一定特点。在工作日, 城市人流的移动集中在早晚通勤高峰期, 且在中午时段活动量最少; 而在休息日, 由于没有通勤上的时间压力, 因此整体的分布较为均匀, 且夜间人流量显然比工作日高。

从图 6 可以看出, 城市的人流移动有数个明显的热点区域, 本文结合了真实的地理信息, 对人流量最大的 4 个区域进行了标注, 可以发现热点区域符合城市中居民的出行规律。

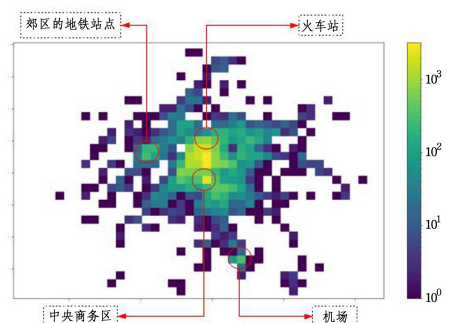


图 6 城市的人流量热力图

Fig. 6 Heat map of city traffic

6.3 基于聚类的城市人流移动模式挖掘

本文选取了 2018 年 10 月 8 日至 10 月 12 日的数据进行实验,聚类得到的初步结果以 xy 轴为空间平面, z 轴为时间切片,结果如图 7 所示。

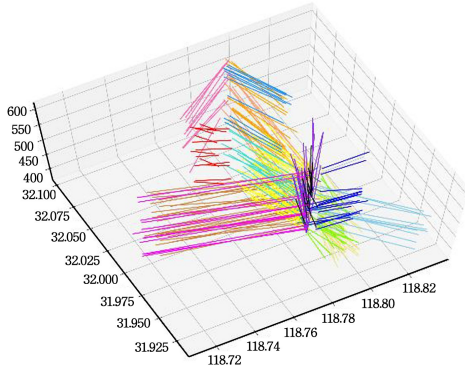
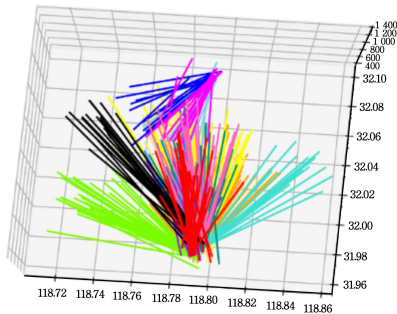


图 7 聚类结果概览

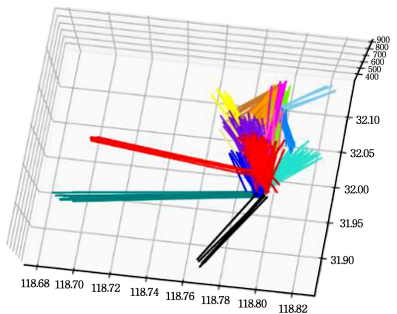
Fig. 7 Overview of clustering results

相较于传统的 DBSCAN,本文提出的 ST-DBSCAN 的主要改进在于通过重新选择距离公式,重新定义了邻域。接下来在时间和空间两个维度上通过对比实验来说明 ST-DBSCAN 对聚类结果的改进。

在空间维度上,由于基于密度的聚类算法是通过密度相连关系分簇的,会导致一些结构较为分离的簇会因几个点而达成相连关系,所以,基于密度的聚类通常会尽量选择较小的空间邻域来避免这种关系。普通的 DBSCAN 算法由于要计算起始地与目的地间的距离,因此会出现放射状的结果簇,如图 8(a)所示,这会导致一个移动模式的某一端覆盖范围过大而丢失实际意义。通过重新定义邻域,分别计算起始地与目的地间的空间距离,可以重新使得到的结果簇有正常的覆盖范围,如图 8(b)所示。



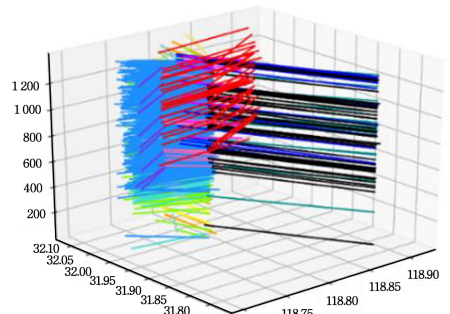
(a) 普通 DBSCAN



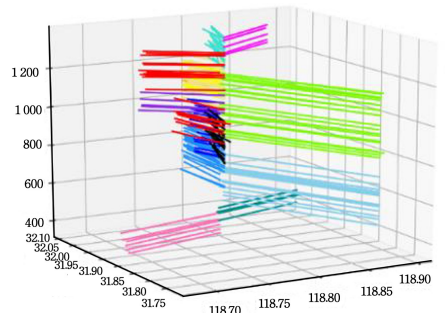
(b) ST-DBSCAN

图 8 空间维度上的聚类结果对比

Fig. 8 Comparison of clustering results in spatial dimensions



(a) 普通 DBSCAN



(b) ST-DBSCAN

图 9 时间维度上的聚类结果对比

Fig. 9 Comparison of clustering results in temporal dimensions

在时间维度上,通过添加时间邻域,可以在时间上划分移动模式,这样做的现实意义是找到一对热点出行区域间的出行集中高峰。为了体现时间邻域的差别,本节选取了移动距离超过 5km 的长距离出行数据集,并对这个子数据集进行聚类,聚类的对比结果如图 9 所示。可以发现,添加时间邻域会显著降低簇中所含的对象数,由于时间邻域的限制,移动地点相近但时间差距较大的点会被视为离群点。但这种方法的好处是可以明显发现相同起始地和目的地间的出行存在时间聚集点。

6.4 基于移动成本的移动模式评价

由式(10)可知,在第 5 节提出的场景下,对得到的每一组移动模式可以计算得出一个交通成本差,差值越大代表交通系统的资源消耗越少,同时可以反映出得到的人流移动模式越符合实际情况且可以为应用提供支撑。基于 6.3 节得到的结果,成本的差值,即交通资源节省量如图 10 所示,可以发现,在移动模式中的个体数越多,统一的交通模式节省的交通成本就越多,在个体数为 4 时成本差为 0,且在个体数为 7 及以上时可以节省 25% 以上的交通资源。该结果证明了本文提出的方法的合理性,以及为城市应用提供支撑的可能性。

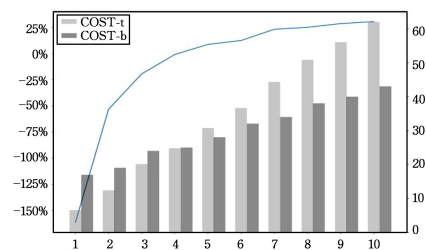


图 10 交通资源节省量

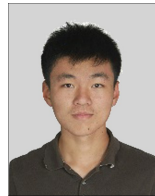
Fig. 10 Traffic resource savings

结束语 本文在智慧城市的背景下,研究了基于智慧城市所感知的时空数据对人流的移动模式进行挖掘的方法。首先对异构的时空数据进行统一的建模处理,将其转化为统一的人流;然后设计了一种改进的基于密度思想的聚类算法,针对传统 DBSCAN 无法解决时空数据的问题,对 DBSCAN 的输入参数进行扩展,并重新在时空域上定义了邻域,从而实现了对人流的聚类;其次针对移动模式提出了一个应用场景以及基于该场景的评价标准,利用交通成本差来对移动模式进行评价;最后在真实的出租车数据集上进行实验,对结果进行了分析和评价。

本文基于时空数据仅对移动模式进行了初步的挖掘,事实上,在各个智慧城市的应用场景中,对移动模式的应用不尽相同,本文方法可能并不普适于所有的实际应用。如何合理地移动模式建模,从而贴合实际应用,仍然是值得研究和需要解决的问题。

参 考 文 献

- [1] ITOH M, YOKOYAMA D, TOYODA M, et al. Visual Exploration of Changes in Passenger Flows and Tweets on Mega-City Metro Network [J]. *IEEE Transactions on Big Data*, 2016, 2(1):85-99.
- [2] ALESSANDRINI A, GIOIA C, SERMI F, et al. WiFi positioning and Big Data to monitor flows of people on a wide scale [C] // 2017 European Navigation Conference (ENC). IEEE, 2017.
- [3] JING Y, YU Z, XING X. Discovering regions of different functions in a city using human mobility and POIs [C] // 18th SIGKDD conference on Knowledge Discovery and Data Mining.
- [4] LYU Y, CHOW C Y, VICTOR C. S, et al. T2CBS: Mining Taxi Trajectories for Customized Bus Systems [C] // *Computer Communications Workshops*. IEEE, 2016.
- [5] ZHANG J, ZHENG Y, QI D. DeepSpatio-Temporal Residual Networks for Citywide Crowd Flows Prediction [C] // *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. AAAI Press, 2017:1655-1661.
- [6] HASHEM I A T, CHANG V, ANUAR N B, et al. The role of big data in smart city [J]. *International Journal of Information Management*, 2016, 36(5):748-758.
- [7] STEENBRUGGEN J, TRANOS E, NIJKAMP P. Data from mobile phone operators [M]. Oxford: Pergamon Press, 2015, 39(3):335-346.
- [8] DEMISSIE M G, PHITHAKKITNUKON S, KATTAN L. Trip Distribution Modeling Using Mobile Phone Data; Emphasis on Intra-Zonal Trips [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2018, 20(7):2605-2617.
- [9] DEMISSIE M G, PHITHAKKITNUKON S, SUKHVIBUL T, et al. Inferring Passenger Travel Demand to Improve Urban Mobility in Developing Countries Using Cell Phone Data: A Case Study of Senegal [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2016, 17(9):2466-2478.
- [10] FEI X, GKOUNTOUNA O. Spatio temporal clustering in urban transportation; a bus route case study in Washington D. C [J]. *Sigspatial Special*, 2018, 10(2):26-33.
- [11] ZHENG Y. Trajectory data mining: an overview [J]. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2015, 6(3):29.
- [12] LIU Z C, YU J B, XIONG W T, et al. Using mobile phone data to explore spatial-temporal evolution of home-based daily mobility patterns in Shanghai [C] // *International Conference on Behavioral, Economic and Socio-Cultural Computing (BESCI)*. IEEE, 2016.
- [13] YABE T, TSUBOUCHI K, SEKIMOTO Y. CityFlowFragility: Measuring the Fragility of People Flow in Cities to Disasters using GPS Data Collected from Smartphones [J]. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2017, 1(3):1-17.
- [14] DO C X, TSUKAI M. Exploring Potential Use of Mobile Phone Data Resource to Analyze Inter-regional Travel Patterns in Japan [M] // *Data Mining and Big Data*. 2017:314-325.
- [15] LI Y, BAO J, LI Y, et al. Mining the Most Influential k-Location Set from Massive Trajectories [J]. *IEEE Transactions on Big Data*, 2017, 4(4):556-570.
- [16] LIAN J, LI Y, GU W, et al. Joint Mobility Pattern Mining with Urban Region Partitions [C] // *Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems; Computing, Networking and Services*. ACM, 2018:362-371.
- [17] LIU T, CEDER A A. Analysis of a new public-transport-service concept: Customized bus in China [J]. *Transport Policy*, 2015, 39:63-76.
- [18] MA S, ZHENG Y, WOLFSON O. Real-Time City-Scale Taxi Ridesharing [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2015, 27(7):1782-1795.



SUN Tian-xu, born in 1995, M. S. candidate, is a member of China Computer Federation. His main research interests include urban computing and pervasive computing.



ZHAO Yun-long, born in 1975, Ph. D., professor, is a member of China Computer Federation. His main research interests include wireless network, collective computing, pervasive computing, data mining, wearable computing, etc.