

基于交通指数聚类的路网区域动态划分

侯军军, 龙佰超, 王洪钰, 肖建力
(上海理工大学 光电信息与计算机工程学院, 上海 200093)

摘要: 针对宏观路网区域交通状态预报需要首先产生路网区域的需求, 提出了一种新的基于交通指数聚类的路网区域动态划分方法。首先对整个城市路网进行网格化划分, 将路段划分为从属于某个网格的子路段; 然后, 计算每个网格的交通指数, 提取网格特征, 从而得到样本特征矩阵; 接着, 利用 k -means++ 聚类算法对样本特征矩阵进行聚类, 得到初始聚类标签, 并对其中奇异网格的聚类标签加以修正; 最后, 得到划分后的路网区域。为了验证该方法的性能, 利用上海市的 GPS 数据对上海市进行了路网区域的划分, 并与不同聚类方法的结果进行了对比。结果表明, 新方法对路网区域划分的精度及稳定性均有所提高。

关键词: 路网区域; 动态划分; 交通指数; k -means++ 聚类算法
中图分类号: U 491 **文献标志码:** A

Dynamic division of road network areas based on traffic index clustering

HOU Junjun, LONG Baichao, WANG Hongyu, XIAO Jianli
(School of Optical-Electrical and Computer Engineering, University of Shanghai
for Science and Technology, Shanghai 200093, China)

Abstract: As road network areas need to be generated firstly for traffic state forecasting of macro road network areas, a new dynamic division method of road network areas based on traffic index clustering was presented. The entire city's road networks were first divided into grids, in which each road section belonged to one certain grid. Following by this, the traffic index for each grid was computed. Then, features for each grid were extracted to get sample feature matrix. The k -means++ clustering algorithm was used to cluster the sample feature matrix. Consequently, the initial clustering labels were generated. For better clustering results, the grid labels with singularity were modified. Finally, the completed road network areas were obtained. In order to verify the performance of the proposed method, the GPS data of Shanghai was utilized to divide road network areas, and the results of the proposed method were compared with the results obtained by other clustering methods. Experimental results show that the proposed method has improved the division accuracy and stability.

收稿日期: 2020-11-06
基金项目: 国家自然科学基金资助项目 (61603257)
第一作者: 侯军军 (1993-), 男, 硕士研究生. 研究方向: 智能交通. E-mail: 182560476@st.usst.edu.cn
通信作者: 肖建力 (1982-), 男, 副教授. 研究方向: 数据挖掘与大数据技术、机器学习与计算机视觉、智能交通与智慧城市. E-mail: audyxiao@sjtu.edu.cn

Keywords: road network areas; dynamic division; traffic index; *k*-means++ clustering algorithm

现有的交通状态预报方法通常以较小空间尺度的路段作为研究对象。例如,高德地图和百度地图显示的交通状态就是以路段的不同颜色来表示。然而,在某些情形下,大空间尺度的路网交通状态预报往往更为重要,特别是对宏观路网区域交通状态的预报。例如,假设需要选择一个交通状态良好的区域举办一场演唱会,选择徐家汇还是人民广场区域,此时,需要对区域的交通状态进行评价。然而,对区域交通状态进行评价,首先必须产生路网区域。因此,本文主要关注路网区域划分方法的研究。

早期的城市路网区域划分主要根据行政区域及土地性质等因素来进行^[1]。该方法主要用于研究区域的演变及其空间格局的变化,其特点是将乡镇街道的变化等因素融入路网区域划分的研究中。近年来,路网交通状态的预报对路网区域划分的目标提出了更高的要求,需要进行路网区域划分后得到的区域尽量能够反映本区域的交通特性,如果仍然基于之前的思路进行划分,显然已无法满足现代路网区域划分的要求^[2]。国内外学者对此展开了深入的研究。

目前,国内学者对城市路网区域划分的研究,大体上可以分为两种思路:一种是以路网内交叉口的交通特性作为划分标准,将交通特性相近的交叉路口划分为同一个交通区域,这种路网划分方法称为聚类分析方法^[3];第二种是考虑交通路网中所有路段的交叉口,再通过已经规定好的搜索规则,对符合搜索规则的交叉口路段统一考虑后划入同一个区域,这种路网划分方法称为遍历搜索法^[4]。Moore等^[5]提出两步划分法,将路网划分分成两步:初次划分时依据交叉路口的流量大小以及信号控制参数作为划分的标准,通过聚类分析完成初次的区域划分;然后加入最大绿波优化模型,并以信号配时优化为目标,在初次划分的结果上对路网区域实现二次划分。Geroliminis等^[6]提出了宏观基本图的概念,通过应用示例数据区交通流进行数据分析,发现城市路网中均存在“回滞环”的现象,即宏观基本图客观存在且具有不随时间变化的共性,这为城市路网区域划分提供了新的研究思路。

总体来看,现有的路网区域划分方法的研究还是不够成熟,不能够适应现代路网区域划分的

要求。为此,本文提出了基于交通指数聚类的路网区域动态划分方法。为验证城市路网区域划分方法的性能,采用基于*k*-means++聚类算法^[7]的路网区域划分方法进行区域划分,并与其他聚类方法进行对比。结果表明,本文提出的路网区域划分方法的精度要优于其他对比方法,且具有良好的稳定性。

1 城市路网区域动态划分方法

图1呈现了基于聚类的路网区域划分方法的流程,其具体步骤为:第一步,对整个城市区域进行网格划分,路网中的路段被网格切分成不同的子路段,每个子路段都从属于某个特定的网格;第二步,计算每个网格的交通指数;第三步,对每个网格提取特征,建立样本特征矩阵;第四步,对样本特征矩阵进行聚类,得到初始聚类标签;第五步,对奇异网格的标签进行修正,得到网格的最终标签,同时对面积过小的区域进行合并;最后,将具有相同标签的网格组成一个路网区域,对城市路网区域划分的结果进行可视化,得到最终的划分结果。

1.1 地图网格划分

本文提出的路网区域动态划分方法,其主要思路是先将整个路网区域划分成相同大小的网格,然后对网格进行聚类,从而得到路网区域。要实现这一想法,首先需要对整个城市区域进行网格划分。先选定整个城市的路网区域,该区域为一个矩形区域。然后,将其划分为 $N \times N$ 的网格,其中, N 为矩形区域的长或宽的等分数且 N 为正整数。在本文的研究中,将整个上海区域划分成 50×50 的网格。自然地,一条长的路段会被网格切分成不同的子路段,每个子路段将从属于一个唯一的网格。得到网格后,在接下来的小节中将为每个网格计算交通指数。

1.2 网格交通指数计算

交通畅通指数^[8]被定义为在给定时间段内,城市路网或区域的总体畅通程度的相对数,它是一个无量纲的量。交通畅通指数的区间为 $[0, 100]$,数值大小与畅通程度成正比,即数值越大表示状态越畅通,数值越小则越拥堵^[9]。本文将交通畅通指数简称为交通指数。

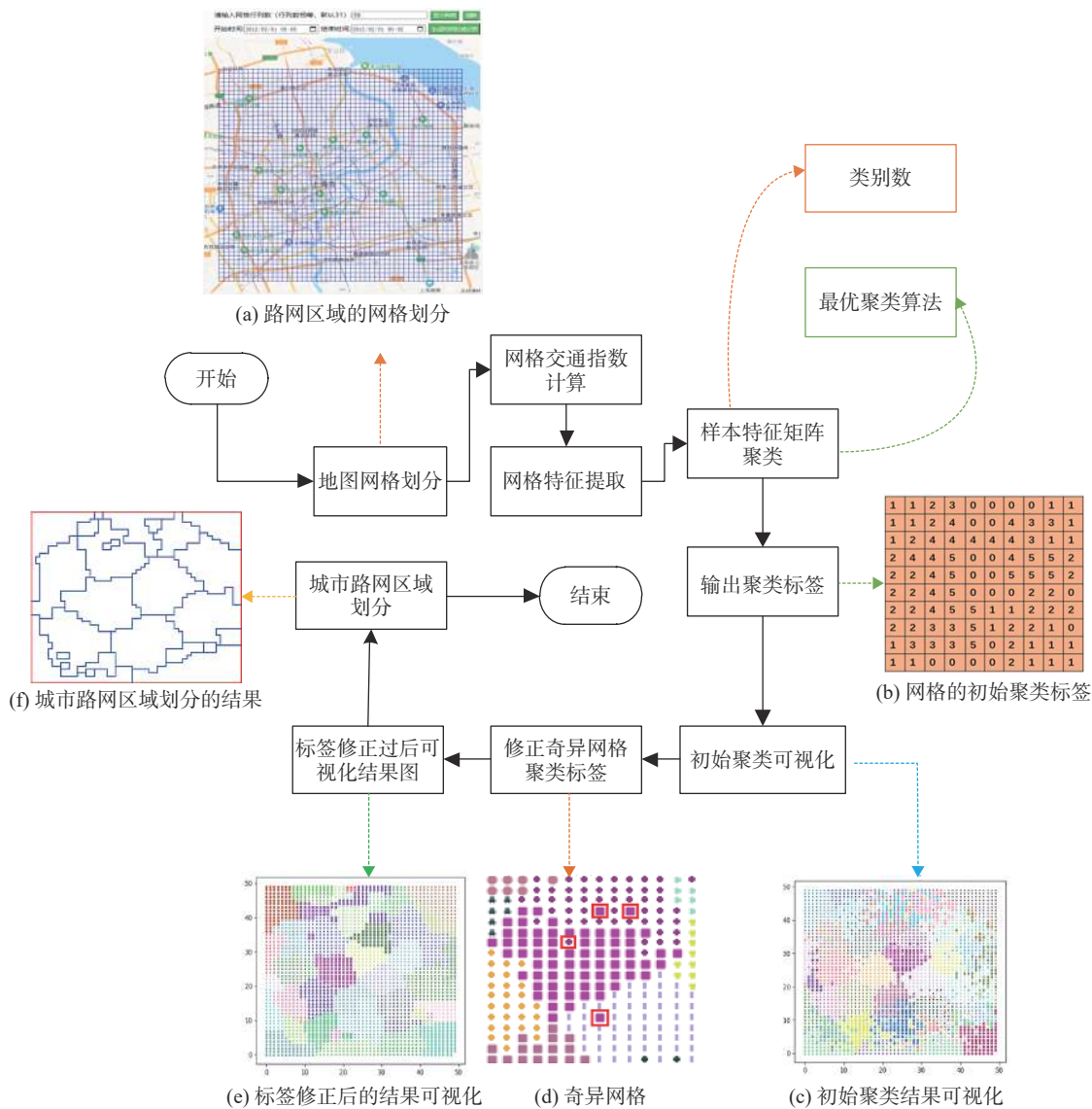


图1 路网区域动态划分方法流程

Fig.1 Flowchart of the dynamic division method of traffic road network area

一个路网区域的交通指数可由式(1)计算得到,即

$$TFI = \frac{\sum_{r=1}^R w_r \left(\sum_{t=1}^T w_t \left(\sum_{i=1}^N k_i l_i \frac{v_i}{v_{fr}} \right) \right)}{\sum_{r=1}^R w_r \left(\sum_{t=1}^T w_t \left(\sum_{i=1}^N k_i l_i \right) \right)} \times 100 \quad (1)$$

式中： TFI 表示一个网格的交通指数； i 代表当前路网区域中的第*i*条子路段； r 代表当前区域属于第*r*种路网； t 代表当前时间属于第*t*个时间段； v_i 为路段实际交通流速度； v_{fr} 为不同路网下的路段的自由流车速，根据路网为快速路、地面主干道、次干道、支路以及高速公路取不同的值； l_i 为路段*i*的里程长度； k_i 为路段的车道数； w_t 为时间权重系

数，按高峰时段和平峰时段取不同的权重； w_r 为路网权重系数，根据路网属于快速路、地面主干道、地面次干道、地面支路、高速公路取不同的权重。

1.3 网格特征提取

为了后续对网格进行聚类，必须先对每个网格进行特征提取。可以由网格连续一段时间内的交通指数以及网格的坐标来构造该网格的特征向量。例如，取一个小时内网格的交通指数，由于指数的计算周期是10 min/次，即一共取6组交通指数加入到网格的特征向量中。对于第*i*个网格，其特征向量可以表示为

$$f_i = (x,y,TFI_{i,1},TFI_{i,2},\cdots,TFI_{i,6}) \quad (2)$$

式中： f_i 表示第*i*个网格的特征向量； (x,y) 表示第*i*个网格的坐标； $TFI_{i,1}$ 表示第*i*个网格的第1个计算

周期的交通指数,其余的5个交通指数以此类推。

对所有网格均作类似处理,可以得到由所有网格的特征向量组成的样本特征矩阵。由于本文中路网划分是按照 50×50 来进行划分的,即一共有2500个网格。于是,样本特征矩阵可表示为

$$F = \begin{bmatrix} x & y & TFI_{i,1} & TFI_{i,2} & \cdots & TFI_{i,6} \\ 1 & 1 & TFI_{1,1} & TFI_{1,2} & \cdots & TFI_{1,6} \\ 1 & 2 & TFI_{2,1} & TFI_{2,2} & \cdots & TFI_{2,6} \\ 1 & 3 & TFI_{3,1} & TFI_{3,2} & \cdots & TFI_{3,6} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ m & n & TFI_{m \times n,1} & TFI_{m \times n,2} & \cdots & TFI_{m \times n,6} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 50 & 50 & TFI_{2500,1} & TFI_{2500,2} & \cdots & TFI_{2500,6} \end{bmatrix} \quad (3)$$

可以将样本特征矩阵 F 表示为列向量的形式,即有

$$F = \begin{bmatrix} f_1 \\ \vdots \\ f_i \\ \vdots \\ f_{2500} \end{bmatrix} \quad (4)$$

1.4 样本特征矩阵聚类

a. 初始聚类。

得到样本特征矩阵后,需要对其进行聚类,从而产生初始的聚类标签。一些常见的聚类算法都可以用来执行聚类任务,例如, k -means 聚类算法^[10]、 k -means++聚类算法、Birch 聚类算法^[11]和AGNES 聚类算法^[12]等。本文采用 k -means++聚类算法进行样本特征矩阵的聚类。

在 k -means++聚类算法中,数据点与聚类中心之间的最近距离用 $D(x)$ 表示,其中 $\mu_o(o=1,2,\cdots,n)$ 表示 n 个聚类中心点, $D(x)$ 的具体计算公式^[13]如下:

$$D(x) = \arg \min \|x_i - \mu_o\|_2^2 \quad (5)$$

k -means++聚类算法在聚类中心点的选取过程中引入了概率的思想:将每一个被选中的样本数据点概率与距其最近的已选中心点的距离相互联系,这样距离越大被选中的概率就越大。选取好 k -means++聚类中心点后,还需要对选定的聚类中心点进行不断更新,直到聚类中心点不再发生变化。假定类别数为 k ,则聚类中心 G 可以表示为 $G = \{g_1, g_2, \cdots, g_k\}$,则聚类中心的数学表达式可以表示为^[13]

$$g_k = \frac{1}{|G_k|} \sum_{x_i \in A_k} x_i \quad (6)$$

式中: g_k 表示第 k 类样本簇的中心点; $|G_k|$ 表示第 k 类样本簇的样本总数; A_k 表示第 k 类样本簇的样本集合; x_i 表示第 k 类样本簇的第 i 个样本。 k -means++聚类算法的具体步骤如下:

步骤1 从输入的样本特征数据 $F = \{x_1, x_2, \cdots, x_n\}$ 中随机选择一个数据点作为第一个聚类中心点 μ_1 ;

步骤2 对于数据点 x_i ,计算数据点与第一个聚类中心点的最近距离;

步骤3 选择新的数据点作为新的聚类中心点,其中选取的原则为 $D(x)$ 中距离较大的点,其被选为聚类中心点的概率较大;

步骤4 重复上述步骤2和步骤3,选出 k 个聚类中心点,然后对其更新,直到满足所有条件,输出聚类标签。

通过 k -means++聚类算法对样本特征矩阵进行初始的聚类后,可以得到网格的聚类标签。图1(b)为网格数为10、类别数为6的带有聚类标签的可视化效果图,数字代表每个小网格的聚类标签。注意,该图只是一个示意图,为给读者一个直观的展示。

b. 奇异网格标签的修正。

通过初始聚类得到聚类标签,然后对具有相同类别标签的网格着上同一种颜色,得到初始聚类的可视化结果图,如图1(c)所示。为了更清楚地观察到初始聚类可视化结果图中存在的奇异网格,对图1(c)进行局部放大,结果如图1(d)所示。对其中存在的部分奇异网格用红色方框标识出来。所谓奇异网格是指其标签(即颜色)与周围邻域内网格的绝大多数标签存在明显差异的网格。奇异网格的标签可能是错误的,因此,需要采用正确的方法对奇异网格的标签进行修正。

图1(d)展示了中心网格的四连通和八连通区域内的奇异网格。可以看到奇异网格与四连通和八连通区域内绝大多数网格的颜色明显不同。为便于后续论述,先给出四连通与八连通区域的基本概念。

根据数字图像处理^[14]中的相关理论,四连通区域是指处于中心网格上、下、左、右,紧邻的4个网格区域,如图2(a)所示。共有4个方向,所以称为四连通区域,又称四邻域。假设中心网格 p 的坐标为 (x, y) ,则四连通数学含义可定义为

$$N4(p) = (x-1, y), (x+1, y), (x, y-1), (x, y+1) \quad (7)$$

同理,八连通区域,是指在中心网格的上、下、左、右、左上、右上、左下、右下,是紧邻的网格和斜向相邻的网格区域组成的全体,共

8个方向，所以称之为八连通区域或八邻域，如图2(b)所示。假设中心网格 p 的坐标为 (x,y) ，则八连通数学含义可定义为

$$N8(p)=N4\bigcup_{(x-1,y+1),(x+1,y+1),(x-1,y-1),(x+1,y-1)}(x,y)$$

(8)

本文提出一种基于八连通多数投票的奇异网格标签修正方法，其具体步骤如下：

步骤1 输入初始聚类标签；

步骤2 判断在中心网格的八连通区域中是否存在奇异网格，如果存在奇异网格，则将奇异网格聚类标签修正为八邻域内数目最多的标签，同时中心网格右移一个网格，直到所有奇异网格的聚类标签都得到修正，否则重复步骤1和步骤2；

步骤3 输出修正后的聚类标签。

经过基于八连通多数投票的奇异网格标签修正后，即可得到初始的路网区域划分，观察各个区域，一旦发现某个区域的面积过小，且其颜色与周围面积较大区域的颜色相近，则将该区域进行合并。

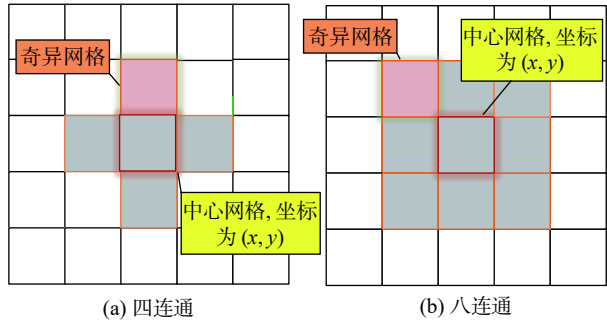


图2 奇异网格示意图
Fig.2 Examples of singular grids

1.5 产生路网区域

对于上述步骤得到的区域，还需要用轮廓线进行标识，具体方法为：通过比较各网格的聚类标签来画出城市路网区域的轮廓线。首先确定轮廓线起点，接下来如果两个聚类标签不相同就在两个聚类标签中间画一条直线，相同则不作处理，依次迭代下去，直到将所有聚类簇的轮廓线都画出来为止。最终得到城市路网区域划分结果，如图1(f)所示。

2 相关实验

2.1 数据来源

本文实验部分的GPS数据来源于上海市交通

信息中心，时间为2012年2月1日，一天24h的GPS数据。

2.2 交通指数颜色映射

为了后续验证路网区域划分方法的精度，将网格的交通指数映射为不同的颜色，颜色映射的规则为：将交通指数由0到100分别映射为深红到深绿色，颜色会随着交通指数数值由小到大呈现出不断渐变的过程^[15-16]。其中，某些缺少交通指数的网格则用白色表示，效果图如图3所示。

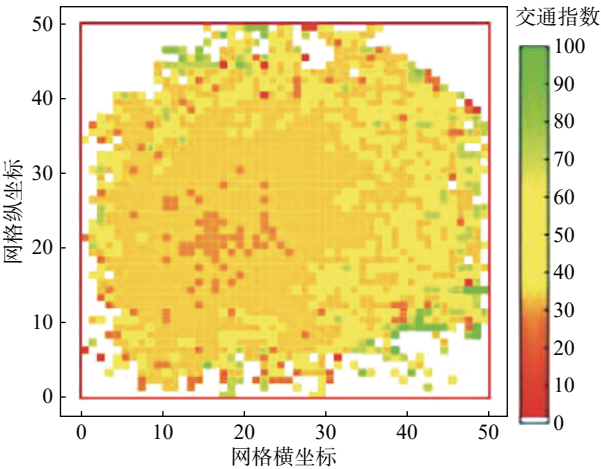


图3 交通指数颜色映射图
Fig.3 Map generated by mapping the traffic indexes into the color space

2.3 聚类算法精度评价指标

将城市路网区域划分结果图1(f)叠加到图3中得到图4。在图4中红色线段所框出来的区域中，组成区域的网格主要由两部分组成，主基色网格以及辅基色网格。所谓主基色网格是指一个区域内颜色接近且网格数目占绝大多数的网格，辅基色网格则是指区域内除了主基色网格之外的其他颜色网格。一般而言，辅基色网格的颜色与主基色网格颜色不同，且其网格数量占该区域的少数。区域面积则可由该区域内的网格总数来表示，也即主基色网格数量和辅基色网格数量的总和。

为了评价不同城市路网区域划分方法的精度，创新性地提出了一个衡量某个区域划分精度的指标 P ，其计算公式为

$$P=\left(\frac{1}{H}\sum_{i=1}^nh_i+\frac{1}{S}\right)$$

(9)

式中： P 表示当前区域的划分精度， P 值越小表示划分精度越好， P 值越大表示划分精度越不理想； H 表示路网区域中的主基色网格数量； h_i 为第 i

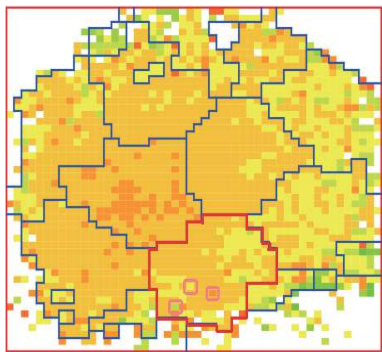


图4 路网区域及区域中的主基色网格和辅基色网格

Fig. 4 Traffic road network areas and the major color grids and minor color grids in these areas

种辅基色网格的数目, n 为辅基色颜色的总数, $\sum_{i=1}^n h_i$ 表示路网区域中辅基色网格总数; S 表示当前区域的面积大小, 即当前区域内网格的总数。

为了衡量整个路网划分的精度, 可以计算 P 的平均值 P_{ave} , 即有

$$P_{ave} = \frac{\sum_{k=1}^N P_k}{N} \quad (10)$$

式中: N 表示城市路网中的区域总数; P_k 表示第 k 个区域的划分精度; P_{ave} 指标用来衡量整个城市路网划分的平均精度。

2.4 实验验证

实验部分根据如下思路进行组织: 实验1分别对不同的路网区域划分方法的精度进行定性和定量分析, 实验2对不同的路网区域划分方法的稳定性进行验证。

将 k -means++ 聚类算法与 k -means 聚类算法、AGNES 聚类算法以及 Birch 聚类算法 4 种方法对比验证。 k -means 聚类算法流程如下: 首先设置聚类的类别数 k , 然后从样本数据中随机选取 k 个样本作为初始的聚类中心, 通过计算聚类中心与样本数据点之间的距离, 把每个数据点分配到距离它最近的聚类中心; 再将同一类别的所有样本数据特征取平均值, 并将平均后所得的数据点作为新的聚类中心; 最后, 重复上述步骤直到聚类中心满足终止条件, 输出聚类标签。 AGNES 聚类算法采取自下而上的聚类思想。首先将样本数据作为一个聚类簇, 然后根据一定的相似性度量将这些聚类簇逐渐合并成一个大的聚类簇, 直到达到初始设定的类别数为止。 Birch 聚类算法引入聚类特征和聚类特征树两个新的概念来概括对聚类的描述。其中, 聚类特征树概括了聚类中的有用信

息, 占用的空间比原来的数据集要小, 这节约了内存空间, 提升了 Birch 算法在大型数据集上的聚类速度^[17-20]。

a. 方法精度验证。

分别选取以下 4 个时间段: 04:00—04:20、17:40—18:00、08:00—08:20、00:40—01:00, 通过定性分析和定量计算验证基于 k -means++ 聚类算法的路网区域划分方法的精度。

定性分析, 以图 5(a)~(d) 的结果, 基于 k -means++ 聚类算法的路网区域划分方法的精度, 对其展开定性分析。通过观察图 5(a)~(d) 发现各个路网区域中网格的颜色一致性较强, 这表明路网区域划分的精度较高。

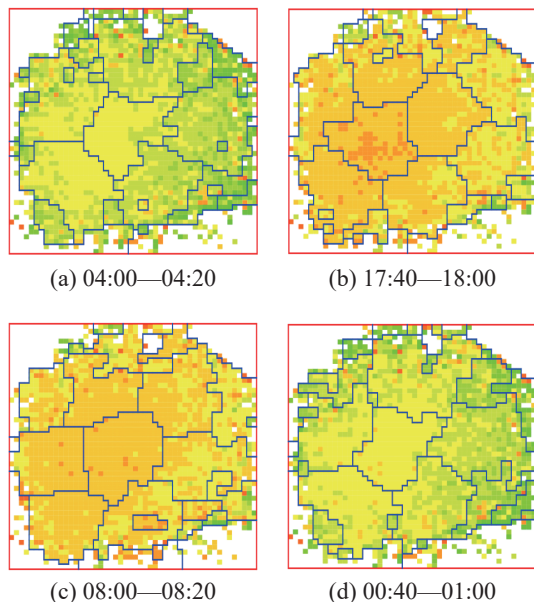


图5 k -means++ 聚类算法的精度验证

Fig.5 Accuracy verification of k -means++ clustering algorithm

定量分析, 通过对城市路网的各区域分别计算精度性能评价指标 P , 然后计算 P 的均值 P_{ave} 来评价路网区域划分方法的精度, P_{ave} 值越小精度越高。选取的交通指数时间段为 17:40—18:00。图 6(a)~(d) 分别为基于 k -means++, k -means, AGNES, Birch 聚类算法的路网区域划分方法的划分结果, 由图 6(a)~(d) 分别计算各方法划分结果的 P_{ave} 值, 并总结于表 1 中。由表 1 可知, 基于 k -means++ 聚类算法的路网区域划分方法精度最高。

b. 方法稳定性验证。

如图 7(a)~(i) 所示, 对基于 k -means++ 聚类算法的路网区域划分方法在不同时间段上的稳定性进行验证, 选取的 9 个时间段分别是: 12:00—

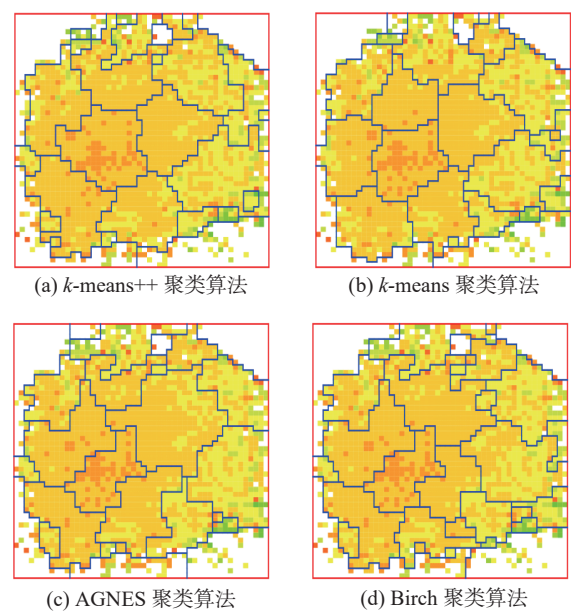


图 6 不同聚类算法对比图

Fig. 6 Comparison of different clustering algorithms

表 1 不同方法的 P_{ave} 值
Tab.1 P_{ave} values of different methods

算法	P_{ave}
k -means	0.55
k -means++	0.49
Birch	0.54
AGNES	0.51

12:20、13:00—13:20、14:00—14:20、15:00—15:20、16:00 —16:20、 17:00 —17:20、 18:00 —18:20、19:00—19:20 和 20:00—20:20。纵观整个路网区域，划分结果会随着不同时间段的交通指数发生一定的变化。但是，总体上同一区域的划分结果不会发生太大的空间区域变化。实验结果证实了基于 k -means++聚类算法的路网区域划分方法其稳定性较为良好。

分别在 9 个不同时间段的路网区域划分结果

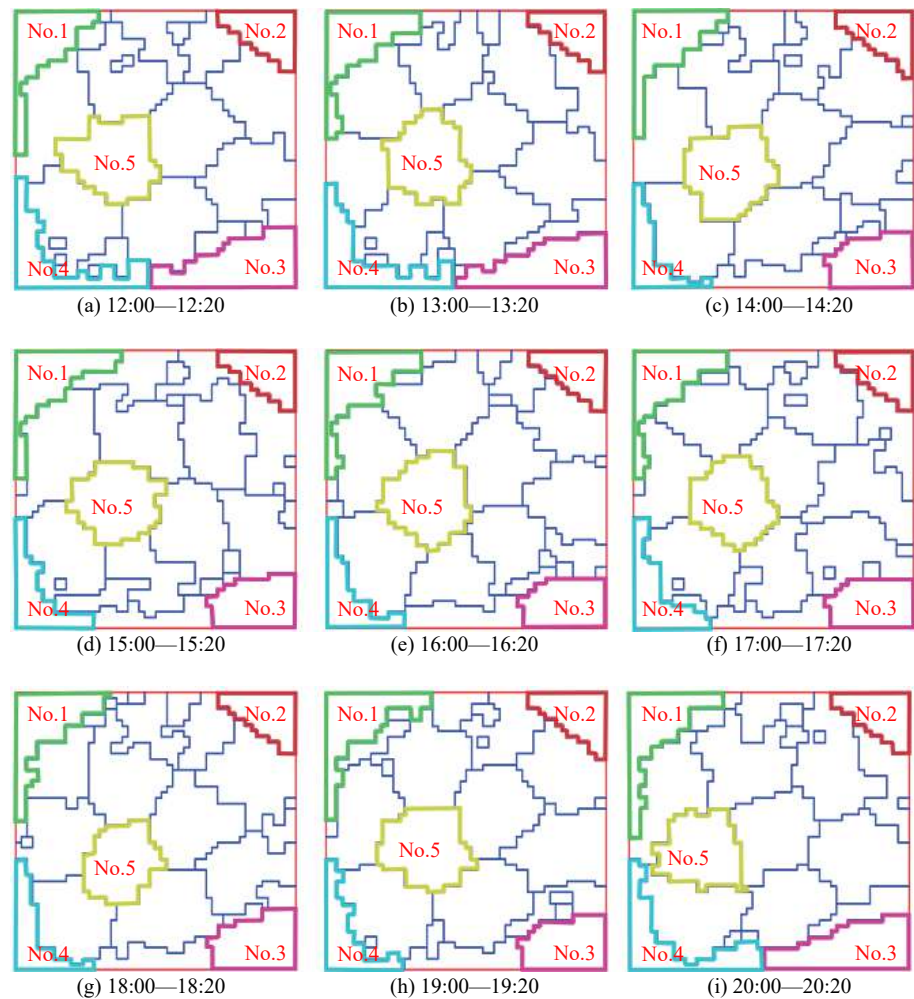


图 7 k -means++聚类算法稳定性验证

Fig.7 Stability verification of k -means++ clustering algorithm

图中选取 5 个空间地理位置接近的区域,通过观察这 5 个区域在不同时间段上形状的变化,来定性分析算法的稳定性。如图 7(a)~(i)所示,选取的区域为:区域 No.1、区域 No.2、区域 No.3、区域 No.4 和区域 No.5,采用不同颜色的轮廓线对 5 个区域进行标注。通过对比 9 个时间段上的城市路网区域划分的结果,发现在不同的时间段上,5 个区域面积大小、划分区域的路网拓扑结构没有发生较大变化。这从局部细节上反映了基于 k -means++ 聚类算法的路网区域划分方法的稳定性较为良好。

3 结 论

提出了一种基于交通指数聚类的路网区域动态划分方法,该方法能够对城市路网区域实现动态划分,并具有良好的划分精度和稳定性。本文创新性地提出了衡量不同城市路网区域划分方法的精度衡量指标,使对路网区域划分方法的精度进行定量分析成为可能。通过将基于 k -means++ 聚类算法的路网区域划分方法与其他 3 种方法进行实验对比,实验结果表明,基于 k -means++ 聚类算法的路网区域划分方法相比于其他 3 种方法具有更好的划分精度和稳定性。在今后的研究中,可以将性能更加优越的聚类算法引入到城市路网区域划分中,进一步提高路网区域划分的性能。

参考文献:

- [1] LI Y F, DANG A R, HUANG T H. Evolution of administrative division in Beijing and analysis on its spatial pattern based on GIS[C]//Proceedings of 17th International Conference on Geoinformatics. Fairfax, VA: IEEE, 2009: 1–6.
- [2] 邹文杰,翁剑成,荣建,等.基于空间相关性分析的路网评价区域划分方法[J].北京工业大学学报,2012,38(4): 564–569.
- [3] 谭飞刚,刘伟铭,黄玲,等.混合交通交叉路口跨摄像机自行车再识别研究[J].中国公路学报,2017,30(5): 132–138.
- [4] 张爽.城市路网交通控制子区自动划分方法研究[D].南京:东南大学,2018.
- [5] MOORE J E, JOVANIS P P. Statistical designation of traffic control subareas[J]. *Journal of Transportation Engineering*, 1985, 111(3): 208–223.
- [6] GEROLIMINIS N, DAGANZO C F. Existence of urban-scale macroscopic fundamental diagrams: some

- experimental findings[J]. *Transportation Research Part B: Methodological*, 2008, 42(9): 759–770.
- [7] LIN T, ZHAO C. Nearest neighbor optimization k -means++ clustering algorithm[J]. *Computer Science*, 2019, 46(S2): 216–219.
- [8] XIAO J L, WANG H Y. Traffic index cloud maps for traffic flow analysis with big traffic data[C]//Proceedings of the 5th International Conference on Computer and Communication Systems (ICCCS). Shanghai: IEEE, 2020: 20–23.
- [9] 肖建力.智能交通中的多核支持向量机与分类器集成方法研究[D].上海:上海交通大学,2013.
- [10] SINAGA K P, YANG M S. Unsupervised k -means clustering algorithm[J]. *IEEE Access*, 2020, 8: 80716–80727.
- [11] LANG A, SCHUBERT E. BETULA: numerically stable CF-Trees for BIRCH clustering[C]//Proceedings of the 13th International Conference on Similarity Search and Applications. Copenhagen, Denmark: Springer, 2020: 281–296.
- [12] SINGH A, YADAV A. Effect of similarity measures on AGNES based hierarchical clustering[J]. *International Journal of Engineering and Management Research*, 2014, 4(3): 257–263.
- [13] ARTHUR D, VASSILVITSKII S. K-means++: the advantages of careful seeding[C]//Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. New Orleans: Society for Industrial and Applied Mathematics, 2007: 1027–1035.
- [14] GONZALEZ R C, WOODS R E, EDDINS S L. 数字图像处理[M].阮秋琦,译.北京:电子工业出版社,2014.
- [15] XIAO J L, WEI C, LIU Y C. Speed estimation of traffic flow using multiple kernel support vector regression[J]. *Physica A: Statistical Mechanics and its Applications*, 2018, 509: 989–997.
- [16] 王忠浩,张静,肖建力.基于交通指数云图的宏观交通流分析方法综述[J].上海理工大学学报,2017,39(4): 353–357.
- [17] 张宪超.数据聚类[M].北京:科学出版社,2018.
- [18] SANO A V D, IMANUEL T D, CALISTA M I, et al. The application of AGNES algorithm to optimize knowledge base for tourism chatbot[C]//Proceedings of 2018 International Conference on Information Management and Technology. Jakarta, Indonesia: IEEE, 2018: 65–68.
- [19] 崔世斌,牛智勇,李若楠,等.基于 AGNES 聚类算法的城市交通运行状态分析研究[J].西昌学院学报(自然科学版),2020,34(4): 62–67.
- [20] 王梦遥,王晓晔,洪睿琪,等.基于改进 BIRCH 聚类算法的评价对象挖掘[J].软件,2019,40(11): 9–12, 61.