引文格式：韩崔燕,郝越,李德勇,等. 面向多元时空对象聚类分析的道路工程统筹[J]. 地理空间信息, 2024, 22(11): 49-52. doi:10. 3969/j.issn. 1672-4623.2024.11.009

面向多元时空对象聚类分析的道路工程统筹

韩崔燕[1]，郝　越[2]，李德勇[1*]，罗　乐[1]，文　星[1]，雷　豁[1]
（1. 武汉众智鸿图科技有限公司，湖北 武汉 430223；2. 中国地质大学（武汉） 地理与信息工程学院，湖北 武汉 430078）

摘　要：传统时空聚类分析主要面向轨迹数据和点状事件，在实际应用中具有局限性。提出了一种基于ST-DBSCAN的多元时空对象聚类分析方法，针对点、线、面等多元化时空对象的空间形态和存续时间，构建度量对象相似性的时空距离模型，并提供关键参数——时空邻域半径的确定策略。在时空距离模型方面，分别计算时空对象的时间距离和空间距离，从而建立综合性的时空距离；在参数确定方面，综合了K-dist图和曲线拐点求取方法，使得聚类结果更加合理和准确。针对城市道路工程的应用结果表明，该方法能获得具有很高内部时空相似性的聚类簇，可有效指导道路工程统筹协调。

关键词：多元时空对象；时空邻域半径；ST-DBSCAN；道路工程
中图分类号：P208　　**文献标志码：**B　　**文章编号：**1672-4623（2024）11-0049-04

Integrated Road Engineering Planning for Multiple Spatio-temporal Objects Clustering Analysis

HAN Cuiyan[1], HAO Yue[2], LI Deyong[1], LUO Le[1], WEN Xing[1], LEI Huo[1]
(1. Wuhan HopeTop Technology Co., Ltd., Wuhan 430223, China;
2. School of Geography and Information Engineering, China University of Geosciences (Wuhan), Wuhan 430078, China)

Abstract: The traditional spatio-temporal clustering analysis mainly focuses on trajectory data and point events, which is limited in practical application. In this paper, we proposed a clustering analysis method for multiple spatio-temporal objects based on ST-DBSCAN. We constructed a novel spatio-temporal distance model to measure the similarity of spatio-temporal objects according to the spatial morphology and existence time of multiple spatio-temporal objects such as point, line and surface, and provided the determination strategy of spatio-temporal neighborhood radius. In the aspect of spatio-temporal distance model, we calculated the temporal distance and spatial distance of spatio-temporal object respectively, so as to establish the comprehensive spatio-temporal distance. In terms of parameter determination, we integrated K-dist figure and curve inflection point calculating method to make the clustering results more reasonable and accurate. The application results of urban road engineering show that the spatio-temporal clustering method can obtain clustering clusters with high internal spatio-temporal similarity, and effectively integrate all kinds of road engineering.

Key words: multiple spatio-temporal object, spatio-temporal neighborhood radius, ST-DBSCAN, road engineering

城市道路工程包括地下管线铺设维修、地铁轻轨等公交设施建设维护、道路改建扩建等，通常需要开挖路面，对城市交通和市民生活具有较大影响。建设单位各自为政，独立规划实施道路工程，极易形成“马路拉链”现象，降低道路通行效率，给城市交通管理带来巨大不便。面对不同的道路工程，各级管理部门需要汇总各单位上报的道路工程施工计划，无法统筹时间空间相近的工程，实现集中破路、集中覆土。为了解决该问题，本文采用时空聚类方法进行道路工程统筹，有效提高了工程实施效率。

时空聚类在空间聚类的尺度上增加了时间维度，在时空域中寻找具有相似特征或行为的时空对象簇，是地理信息科学和计算机科学领域的研究热点。目前时空聚类的研究对象主要集中在轨迹数据和点状事件。聚类对象形态单一难以在实际中得到有效应用，迫切需要一种融合多种对象形态的聚类方法。时空聚类方法包括时空扫描统计[1-2]、时空混合距离[3-4]、时空密度聚类[5-6]等。刘辉[7]等提出了一种基于自适应迭代的空间聚类方法，利用自适应迭代聚类和相关性分析探索微博空间分布和城市公共服务设施的时间活跃

收稿日期：2023-08-29。
项目来源：贵州省科技计划资助项目（2023YB169）；中国地质大学（武汉）研究生联合培养基地资助项目（YJC2021520）。
第一作者简介：韩崔燕（1984—），本科，高级工程师，从事时空信息处理及大数据分析工作，E-mail：hancuiyan@ecitychina.com。
通信作者：李德勇（1989—），本科，工程师，从事时空信息处理及大数据分析工作，E-mail：lideyong@ecitychina.com。

度；秦昆[8]等利用行为轨迹大数据进行时空轨迹探测，得到行为轨迹的时空热点区域；Lamb Z W[9]等利用车辆轨迹数据模拟用户车载网络在时间上的潜在关联，以合理规划道路网内信号基站布设；董一强[10]将耦合速度和方向的轨迹数据抽象为点状事件，以简化算法复杂度；张忠贵[11]等研究了供水管网爆管点数据的时空聚类算法，分析了事故高发区域；郭乃琨[12]等在空间聚类算法的基础上引入时间距离度量方法，实现了船舶轨迹的时空聚类。上述研究将空间实体理想化为实体点且赋予简单的瞬时时间状态，基于对象的时空聚类往往以某个时刻或观测地点的属性作为实体点进行聚类，属性维护扩展性差[13]。实际中市政工程、污染事故、交通拥堵、管理网格等时空对象都具有特定的延展空间和存续时间，针对点状瞬时对象的时空聚类方法并不适用，需要进一步研究新方法。时空聚类算法的效果与聚类参数的选取密切相关，许多学者对时空聚类的参数优化策略进行了研究。基于密度的带噪声的时空聚类算法（ST-DBSCAN）应用广泛，邻域半径与密度阈值等关键参数的确定成为研究热点[14-15]。Strauss C[16]等采用基于时间滑动窗口和标准核密度估计的方法改进了时空聚类效果；石陆魁[17]等直接采用经验法确定密度阈值，并进行多次试验求取最佳结果；刘勇[18]等考虑数据重心点分布密度不均衡的问题，提出时空分布重心点转移方案，改进了聚类结果；徐子惠[19]采用K近邻法，通过频次统计得到距离半径；王慧东[20]等通过绘制时空对象距离频数柱状图来设定自适应阈值，以提高聚类的准确性和有效性。总体来看，聚类参数的确定主要基于经验和试验，缺乏普遍适用的方法。本文提出了一种基于ST-DBSCAN算法的多元时空对象聚类分析方法，用于处理点、线、面等空间形态多样化且有一定存续时间的聚类对象；并给出了相应的参数确定策略。

1 ST-DBSCAN时空聚类算法

ST-DBSCAN算法是DBSCAN算法在时间维度上的延展，以时空密度作为时空对象相似性的度量标准，无需提前规定分类数量，且可得到带有边缘噪声的任意大小类别。ST-DBSCAN算法的主要参数为时空邻域半径Eps和时空密度阈值Minpts。一个时空对象的时空邻域是与其时空距离不超过Eps的时空对象的集合，而时空簇是聚类过程中被划分为同一类的时空对象集合。ST-DBSCAN算法的目标是将每个时空对象按所处位置的密度归类到某个时空簇，或直接将其标记为不归属于任何时空簇的噪声。

针对每个不归属任何时空簇的对象 d_i，校验其时空邻域内对象个数是否达到Minpts，若成立则创建新的时空簇 C_k，将 d_i 标记为时空簇 C_k，将 d_i 时空邻域内所有时空对象加入 C_k，持续校验 C_k 中新加入对象的时空邻域，将满足密度要求的邻域并入 C_k，直至 C_k 不再增长为止。该过程针对全部时空对象循环执行，直至每个时空对象被归属为某个时空簇或被标记为孤立的时空噪声为止。

2 多元时空对象距离模型

本文中的时空对象可以是任意点、线、面形态，且具有一定持续时间，称之为多元时空对象。例如，对于道路工程而言，工程实施对象可以是单一的点状，如井盖；可以是线状，如道路；也可以是面状，如一块区域。当道路工程有了时间维度，在三维空间下点状对象表现为空间线段，线状对象表现为空间面，面状对象表现为空间体。多元时空对象的聚类首先需要建立有针对性的距离模型来衡量对象间的相似性，以判别是否将其归为一类。本文首先计算多元时空对象的时间距离和空间距离，再建立综合性时空距离模型。

将具有既定长度的一段持续时间记为时间窗口 tw，将起始时间点记为 $Ts(tw)$，结束时间点记为 $Te(tw)$；多元时空对象在时间维度上的持续时间即其生存周期。若针对 tw，多元时空对象 i 和 j 的生存周期满足式（1），则称 i 和 j 归并受限于 tw。

$$\begin{cases} \mathrm{st}_i \geqslant Ts(tw) \\ \mathrm{st}_j \geqslant Ts(tw) \\ \mathrm{et}_i \geqslant Te(tw) \\ \mathrm{et}_j \geqslant Te(tw) \end{cases} \tag{1}$$

式中，st_i、st_j 分别为 i 和 j 的开始时间点；et_i、et_j 分别为 i 和 j 的结束时间点。

若针对 tw，多元时空对象 i 和 j 的生存周期满足式（2），则称 i 和 j 的重叠超越 tw。

$$\begin{cases} \max(\mathrm{st}_i, \mathrm{st}_j) \leqslant Ts(tw) \\ \min(\mathrm{et}_i, \mathrm{et}_j) \geqslant Te(tw) \end{cases} \tag{2}$$

若针对某一特定时长阈值 ΔT 和多元时空对象 i、j，存在时间窗口 tw，满足 $Te(tw)-Ts(tw)=\Delta T$，且 i 和 j 归并受限于 tw 或重叠超越 tw，则认为它们相对于 ΔT 而言，时间属性差异很小，时间距离 TD_{ij} 为0。在其他情形下，多元时空对象 i 和 j 时间距离的计算公式为：

$$TD_{ij} = \left| \mathrm{st}_i - \mathrm{st}_j \right| \tag{3}$$

多元时空对象 i 和 j 的空间形态表现为点、线或

面，都可抽象为空间点集合，两点集 D_i 、 D_j 间最近点对的欧氏距离即为 i 和 j 的空间距离 SD_{ij}，计算公式为：

$$SD_{ij}=\min_{\substack{(x_i,y_i)\in D_i\\(x_j,y_j)\in D_j}}\sqrt{(x_i-x_j)^2+(y_i-y_j)^2} \tag{4}$$

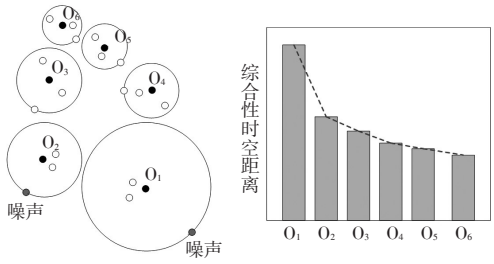
将时间距离和空间距离作为两个独立维度加以综合，定义 i 和 j 的综合性时空距离，即

$$\mathrm{STD}_{ij}=\sqrt{SD_{ij}^{\ 2}+TD_{ij}^{\ 2}} \tag{5}$$

3　多元时空对象聚类方法参数分析

除上述多元时空对象距离模型外，还需确定基于ST-DBSCAN算法的关键参数，即Minpts和Eps。本文参照既往研究结论，将Minpts设定为ln N [14]，N 为多元时空对象的数量。合适的Eps应有效分隔噪声与聚集区，首先计算每个多元时空对象与其第 K 邻近对象间的时空距离（这里 K 的取值即为Minpts），再将这些距离由大到小排序形成柱状图，连接柱状图顶部可以得到一条下降曲线，即K-dist图，下降曲线为K-dist曲线。若K-dist曲线从快速下降到趋于平缓，则变化拐点表示开始形成较明显的聚集状态，以拐点处的距离值作为Eps，则可较好地筛除噪声。

图1a是一个简单的例子，6个对象 O_1 ~ O_6 的3-邻域范围用周边圆圈表示。图1b为对应的K-dist图（$K=3$），K-dist曲线趋于平稳的拐点对应的是 O_2 与其第3邻近对象的时空距离。欲将 O_2 的第3邻近对象纳入聚类簇，则邻域半径需在上述时空距离的基础上继续扩张，将使聚类簇数量过少，影响整体聚类效果，因此以K-dist曲线的拐点来确定Eps是适宜的。



a 邻域范围　　b K-dist图

图1　邻域范围及其K-dist图

在K-dist图中仅以目视判别寻找拐点，易产生误差。为提高精确度，本文采用二阶求导的方法求取图中拐点。针对K-dist图中的离散点，分别计算两点间的斜率和斜率变化率，找到斜率变化率趋近于0的点，即为拐点。结合时空距离模型和Eps的确定策略，多元时空对象聚类算法的流程见图2。其中时长阈值可根据特定应用场景选取。若特定应用场景不易明确该阈值时，可通过比较不同时长阈值下的K-dist图来确定：K-dist曲线快速下降到趋于平缓的拐点处曲线形态变化越显著，求得的拐点值作为邻域半径的聚类效果越好，选择该曲线对应的时长阈值就越合适。

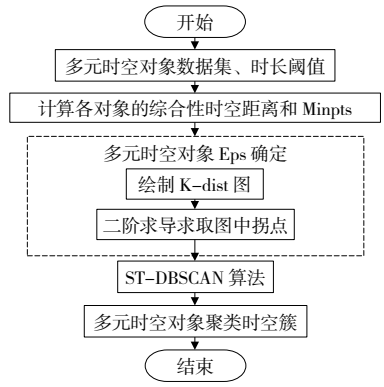


图2　算法流程图

4　实例应用

道路工程在空间分布上存在点、线、面多种形态，在时间上具有持续性，是典型的多元时空对象，对它们进行时空聚类，即可形成道路工程统筹方案，作为管理部门集中协调的基础。为验证前述聚类方法的有效性，将本文方法应用于丹阳市某市政公用集团2020年度的道路工程规划中，道路工程相关属性信息见表1。

表1　道路工程属性信息

道路工程属性	属性表字段	属性表字段类型	详细说明
唯一标识	ID	int	可唯一标识工程的字段
地理位置	Position	geometry collection	点、线、面3种类型
计划起始时间	Start_time	date	
计划竣工时间	End_time	date	
计划工期	Duration	int	单位（日）
权属单位	Ownership	text	工程实施或承办单位

将道路工程数据作为多元时空对象数据集，数据集大小为87，Minpts为ln（87）。设定 K =Minpts=4，以10日、20日、30日和40日为时长阈值分别绘制K-dist图（图3），可以看出，在4个K-dist曲线中30日时长阈值对应曲线的拐点最明显，因此在时空距离计算中采用的时长阈值为30日。

采用多元时空对象聚类分析方法的三维时空聚类结果见图4，图中不同颜色代表算法形成的不同时空簇，横纵坐标代表地理位置，Z 轴代表时间，在此三维时空空间内，点、线、面状道路工程分别表现为时空线段、时空曲面和时空体，可以看出，大量工程孤

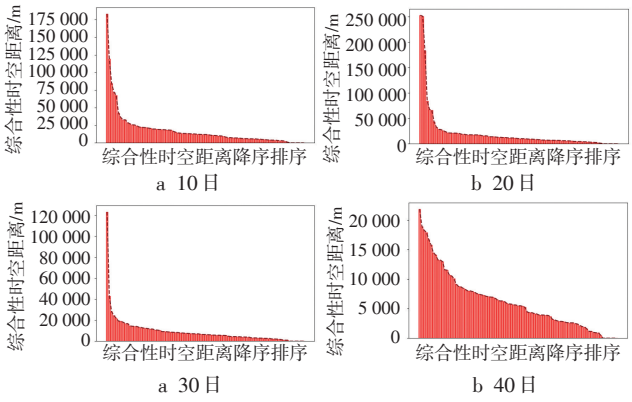


图3 不同时长阈值的K-dist图

立存在（对应于聚类噪声），表示无法与其他工程统筹，可独立安排施工；但有相当部分道路工程被归并到大小不一的6个时空簇，各簇内的道路工程具有较高的时空相似性，工程建设单位可结合实际工程情况统筹安排施工。

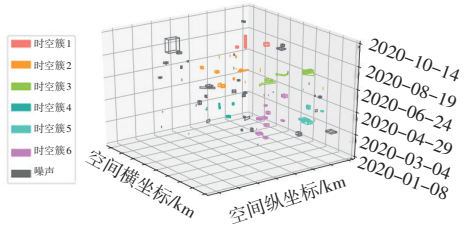


图4 时空聚类结果

5 结 语

本文将具有点、线、面等多种空间形态和一定存续时间的对象统一定义为多元时空对象；并针对多元时空对象构建了多元时空对象距离模型，提出了基于ST-DBSCAN算法的时空聚类分析方法。采用K-dist图与曲线拐点求取方法相结合的方式确定算法的主要参数——Eps，有效保证了聚类结果的合理性和准确性，形成了多元时空对象聚类的有效方法。利用该算法对道路工程进行聚类分析，结果表明地理位置邻近且工程持续时间相近的道路工程被归并到同一集合，验证了该算法对于多元时空对象聚类的有效性和工程适用性。该算法不仅可减轻管理部门之间的协调统筹工作，还可减少大量重复工程造成的人力财力浪费。

参考文献

[1] 张漪. 基于时空扫描统计量的异常聚类方法研究[D]. 昆明:昆明理工大学,2019

[2] Shino S,Narushige S. Network-based Space-time Scan Statistics for Detecting Micro-scale Hotspots[J]. Sustainability,2022,14(24):16 902-16 921

[3] 翟婷,宋文爱,富丽贞,等. 基于路网感知的时空轨迹聚类[J]. 计算机工程与设计,2016,37(3):635-642

[4] 周洋,杨超. 基于时空聚类算法的轨迹停驻点识别研究[J]. 交通运输系统工程与信息,2018,18(4):88-95

[5] 吴笛,杜云艳,易嘉伟,等. 基于密度的轨迹时空聚类分析[J]. 地球信息科学学报,2015,17(10):1 162-1 171

[6] Aryal A M, Wang S J. Discovery of Patterns in Spatio-temporal Data Using Clustering Techniques[C]. The 2nd International Conference on Image, Vision and Computing (ICIVC),Chengdu,2017

[7] 刘辉,黄新,王京晶. 基于位置数据和POI的聚类方法[J]. 地理空间信息,2017,15(11):46-49

[8] 秦昆,王玉龙,赵鹏祥,等. 行为轨迹时空聚类与分析[J]. 自然杂志,2018,40(3):177-182

[9] Lamb Z W, Agrawal D P. Spatio-temporal Clustering of GPS Traces for Energy-efficiency in Vehicular Ad-hoc Networks[C]. IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData),Halifax,NS,2018

[10] 董一强. 基于DBSCAN改进算法的时空轨迹聚类分析与研究[D]. 天津:天津大学,2017

[11] 张忠贵,王玉岚,邓洋,等. 供水管网爆管时空聚类分析研究[J]. 湖北第二师范学院学报,2019,36(2):36-38

[12] 郭乃琨,陈明剑,陈锐. 一种顾及时间特征的船舶轨迹DBSCAN聚类算法[J]. 测绘工程,2021,30(3):51-58

[13] 杨振凯,李响,陈达. 时空对象的聚类方法与应用初探[J]. 地理信息世界,2018,25(2):40-44

[14] Birant D, Kut A. ST-DBSCAN: An Algorithm for Clustering Spatial-temporal Data[J]. Data and Knowledge Engineering,2007,60(1):208-221

[15] Choi C, Hong S Y. MDST-DBSCAN: A Density-based Clustering Method for Multidimensional Spatiotemporal Data[J]. International Journal of Geo-information, 2021, 10(6):391-406

[16] Strauss C, Rosa M B, Stephany S. Spatio-temporal Clustering and Density Estimation of Lightning Data for the Tracking of Convective Events[J]. Atmospheric Research, 2013, 134(1):87-99

[17] 石陆魁,张延茹,张欣. 基于时空模式的轨迹数据聚类算法[J]. 计算机应用,2017,37(3):854-859

[18] 刘勇,何婧,姚绍文,等. 基于重心点转移的St-DBSCAN改进算法[J]. 计算机技术与发展,2018,28(11):6-11

[19] 徐子惠. 基于云GIS的城市交通路况时空聚类特征研究[D]. 大连:辽宁师范大学,2019

[20] 王慧东,宋耀莲,田榆杰. 一种多属性的时空数据聚类算法分析研究[J]. 重庆邮电大学学报(自然科学版),2021,33(4):661-668