



# 基于机器学习技术的耙吸挖泥船施工行为识别

徐 婷，戴文伯，张晴波，周雨森

(中交疏浚技术装备国家工程研究中心有限公司，上海 201208)

**摘要：**耙吸挖泥船施工区域轨迹密集，有明确的抛泥区和挖泥区，但轨迹密度不同，传统的轨迹识别技术对于其施工行为模式识别困难，难以有效应用。针对该问题，提出一种无监督的耙吸挖泥船施工行为识别框架。首先，基于卡尔曼滤波算法解决轨迹跳变问题，提升轨迹数据的质量；然后，基于 HDBSCAN 算法同时识别出密度不同的挖泥和抛泥轨迹，解决了传统 DBSCAN 算法在类间密度不均衡的情况下参数设置困难的问题；最后，基于航向因素建立高斯混合模型 GMM 可进一步识别出运泥轨迹和返回轨迹。结果表明，上述方法能够快速、有效地实现耙吸船施工轨迹的精准识别。

**关键词：**耙吸挖泥船；HDBSCAN 算法；轨迹聚类；行为识别

中图分类号：U 616+.2

文献标志码：A

文章编号：1002-4972(2022)12-0221-04

## Construction behavior recognition of trailing suction hopper dredger based on machine learning technology

XU Ting, DAI Wen-bo, ZHANG Qing-bo, ZHOU Yu-miao

(CCCC National Engineering Research Center of Dredging Technology and Equipment Co., Ltd., Shanghai 201208, China)

**Abstract:** The construction area of a trailing suction hopper dredger has dense trajectories and separate mud dumping and dredging areas. However, these trajectories have different densities, and traditional trajectory recognition technology fails to effectively recognize the construction behavior and mode of the dredger and thus cannot be successfully applied. In view of these problems, this paper proposes an unsupervised framework for recognizing the construction behavior of the dredger. Firstly, the paper solves the problem of trajectory jump based on the Kalman filter algorithm and improves the quality of trajectory data. Then, the paper uses the HDBSCAN algorithm to identify mud dredging and dumping trajectories with different densities simultaneously and solves the problem of difficult parameter setting by the traditional DBSCAN algorithm in the case of uneven density between classes. Finally, the paper establishes a GMM model based on directional factors, so as to further identify the mud transportation and return trajectories. The results show that the above method can quickly and accurately identify the construction trajectory of a trailing suction hopper dredger.

**Keywords:** trailing suction hopper dredger; HDBSCAN algorithm; trajectory clustering; behavior recognition

耙吸挖泥船是一种装备有耙头挖掘机具和水力吸泥装置的大型自航挖泥船，广泛应用于疏浚和吹填工程中<sup>[1]</sup>。耙吸挖泥船施工区域包含明确的挖泥区和抛泥区<sup>[2]</sup>，其施工行为主要有 4 类：挖泥、运泥、抛泥、返回。自动识别

系统 (automatic identification system, AIS) 数据中蕴含着丰富的施工作业信息<sup>[3]</sup>，从中识别出耙吸挖泥船各种施工行为，可及时预警异常行为、评估作业效率，对施工效率优化具有重要意义。然而，快速、有效识别耙吸船各种施工行为较为困难，

一方面由于 AIS 数据存在的轨迹跳变和缺失情况影响识别算法的效果；另一方面是由于耙吸船施工的特有属性决定了挖泥区和抛泥区密度不同，使得传统的基于密度聚类的 DBSCAN 算法在类间密度不均衡的情况下设置参数十分困难，同时，聚类算法中通常使用欧式距离，而耙吸船施工经度、纬度、速度、航向均为重要信息，用多维数据进行距离计算较为复杂，权重难以设置。

本文提出层次密度聚类 HDBSCAN (hierarchical density-based spatial clustering of applications with noise) 算法，改善了 DBSCAN 算法的不稳定性，具有可发现任意形状的簇和异常值点不敏感的优点，能够实现类密度不平衡的精准聚类。

## 1 耙吸船施工行为识别框架

为了快速有效地识别出耙吸挖泥船施工行为模式，设计如图 1 所示识别框架：首先，通过卡尔曼滤波算法，解决轨迹跳变问题；然后，通过 HDBSCAN 算法同时识别出轨迹密度不同的抛泥和挖泥轨迹；最后，基于航向因素建立高斯混合模型 (GMM) 可进一步识别出运泥和返回轨迹。框架可以不考虑航速特征，实现轨迹快速辨识，解决了不同工况下挖泥航速不同导致参数设置困难的问题。本文所述的挖泥轨迹指抛泥区所有轨迹，包括减速驶入、掉头、抛泥和提速驶离等过程。

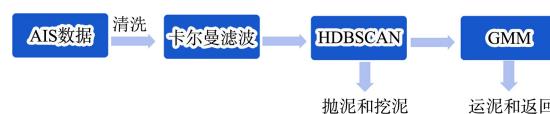


图 1 耙吸挖泥船施工行为识别框架

### 1.1 卡尔曼滤波

卡尔曼滤波本质上是一种数据融合算法<sup>[4-5]</sup>，通过系统状态方程和观测信息的融合算法得到更精确的测量值，分为预测过程和参数更新 2 个过程。

#### 1.1.1 预测过程

预测过程公式为：

$$\hat{x}_t = \mathbf{F}\hat{x}_{t-1} + \mathbf{B}u_{t-1} \quad (1)$$

$$P_t^- = \mathbf{F}P_{t-1}^-\mathbf{F}^\top + Q \quad (2)$$

式中： $\hat{x}_t$  为状态方程  $t$  时刻的估计值； $\mathbf{F}$  为状态转移矩阵； $\mathbf{B}$  为控制转移矩阵； $u_{t-1}$  为控制变量； $P_t^-$  为  $t-1$  时刻对  $t$  时刻系统误差方差的预测值； $P_t$  为  $t$  时刻系统误差协方差的最优估计值； $Q$  为系统估计的偏差。

#### 1.1.2 更新过程

更新过程公式为：

$$K_t = P_t^- \mathbf{H}^\top (\mathbf{H} P_t^- \mathbf{H}^\top + \mathbf{R})^{-1} \quad (3)$$

$$\hat{x}_t = \hat{x}_t^- + K_t (Z_t - \mathbf{H}\hat{x}_t^-) \quad (4)$$

$$P_t = (\mathbf{I} - K_t \mathbf{H}) P_t^- \quad (5)$$

式中： $\mathbf{I}$  为单位矩阵； $\mathbf{H}$  为观测状态转移矩阵； $\mathbf{R}$  为观测值噪声的协方差矩阵； $Z_t$  为  $t$  时刻观测值； $K_t$  为卡尔曼增益率； $\hat{x}_t$  为最优估计值。

## 1.2 HDBSCAN 算法

HDBSCAN 是由 Campello、Moulavi 和 Sander 开发的聚类算法<sup>[6-7]</sup>。它将分层聚类算法思想引入以改进 DBSCAN 算法，可以处理密度不同的聚类问题；通过空间变换，使得密集区域的样本点距离不受影响，而稀疏区域的样本点距离被放大，从而增加了聚类算法的适用性。具体过程如下：

1) 空间变换。用互达距离  $d_{\text{mreach-}k}$  表示 2 个样本点的距离，计算公式为：

$$d_{\text{mreach-}k} = \max \{ \text{core}_k(a), \text{core}_k(b), d(a, b) \} \quad (6)$$

其中： $\text{core}_k(x)$  样本与第  $k$  个最近样本点的距离称为核心距离，计算公式为：

$$\text{core}_k(x) = d[x, N^k(x)] \quad (7)$$

2) 构建互达距离加权图的最小生成树。

3) 根据最小生成树，构造一个完整的含有连通分支的聚类层次结构。

4) 根据最小簇大小压缩集群层次结构。

5) 从压缩树中提取稳定群集。

### 1.3 GMM 算法

GMM 是一种混合概率分布模型，概率密度由多个单高斯分布加权平均得到，数学形式如下：

$$P(x) = \sum_{k=1}^C \pi_k g_k(x; \mu_k, \sigma_k) \quad (8)$$

式中： $g_k(x; \mu_k, \sigma_k)$  为高斯分布； $\pi_k$  为权重； $C$  为该模型中高斯成分数量，计算样本属于  $C$  个高斯

分布的概率, 选取概率最大的类为判决结果。

## 2 实证分析

为了验证算法框架的可行性, 选取某耙吸挖泥船 2021-12-20—2021-12-03 在某地施工的 AIS 数据进行验证, 该船初始的轨迹路径见图 2。

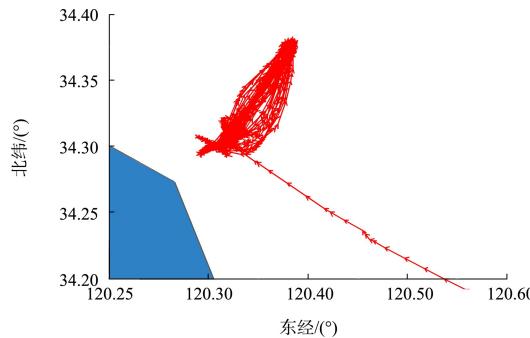


图 2 某耙吸船初始轨迹路径

### 2.1 数据清洗

船舶位置信息不仅可以从 AIS 信息中直接读取, 也可以根据航速、航向等信息建立状态方程预测, 由于 AIS 数据有时出现的位置跳变问题影响建模效果, 故利用卡尔曼滤波算法修正 AIS 数据的位置信息十分必要, 图 3 分别为经度和纬度的卡尔曼滤波清洗后效果。

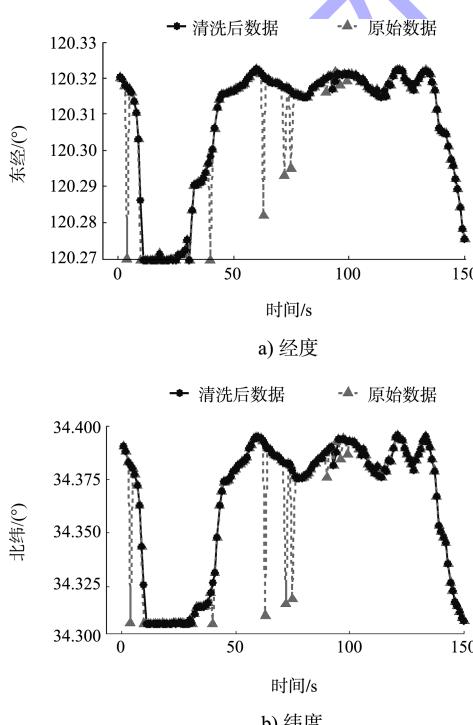


图 3 卡尔曼滤波清洗效果

在施工区域中, 不仅存在轨迹跳变情况, 还会存在锚泊行为, 需要识别并剔除, 否则影响建模效率。将航速  $< 0.5 \text{ kn}$  且持续时间超过 30 min 的轨迹识别为抛锚轨迹。图 4 为经过卡尔曼滤波且去除锚泊和入场轨迹的路径。

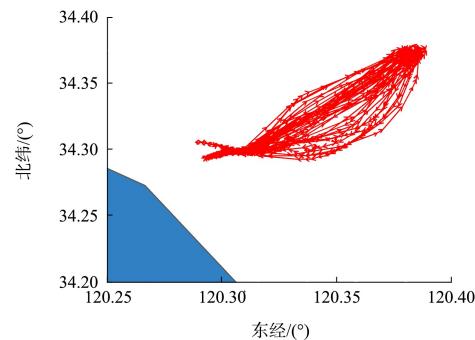


图 4 某耙吸船清洗后的轨迹路径

### 2.2 模型建立

基于清洗后数据的经、纬度信息建立 HDBSCAN 模型, 设置最小簇样本量为 10, 聚类结果见表 1。图 5 为标识稳定簇的聚类结果, 系统判定簇 1、4 为稳定簇, 并用矩形框标出。由于施工区轨迹更为密集, 将期望距离较小的簇 4 识别为施工区, 簇 1 为抛泥区。

表 1 基于经、纬度信息的 HDBSCAN 聚类结果

簇编号	样本量	稳定性分值
0	417	-
1	136	36 208.67
2	10	76.78
3	15	845.21
5	13	1 567.02
6	16	1 577.28

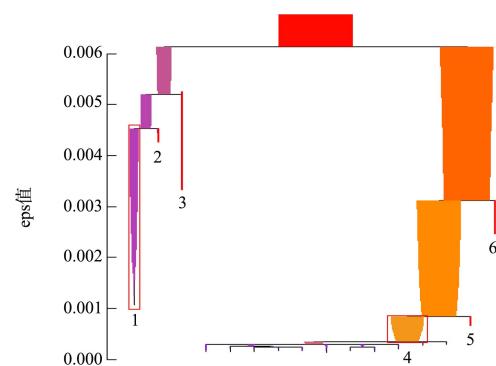


图 5 HDBSCAN 聚类结果

图 6 为根据 HDBSCAN 聚类结果绘制的挖泥区和抛泥区的轨迹路径, 剔除抛泥和挖泥轨迹, 将

剩余轨迹基于速度特征建立 GMM 模型, 见表 2。由于挖泥区到抛泥区的轨迹夹角为锐角, 可判断运泥方位角为锐角, 簇 1 均值为  $52.67^\circ$ , 识别为运泥轨迹; 而簇 2 识别为返回轨迹。最终的聚类轨迹见图 7。图 8 为随机截取一段聚类结果的速度时序, 虽然聚类过程并未考虑时间因素, 但其相关性明显, 可作为人工判断船舶行为的辅助工具。

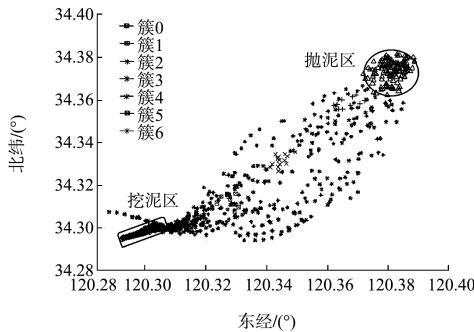


图 6 挖泥区和抛泥区的轨迹路径

表 2 基于速度的 GMM 聚类算法结果

项目	样本量	均值	标准差
簇 1	216	52.67	5.12
簇 2	255	233.07	20.46

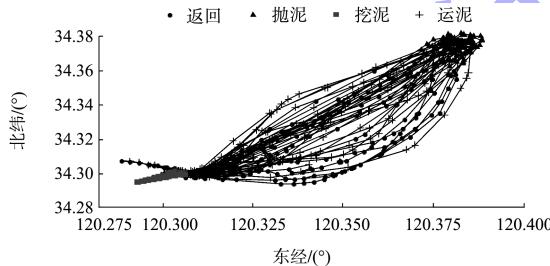


图 7 最终聚类结果

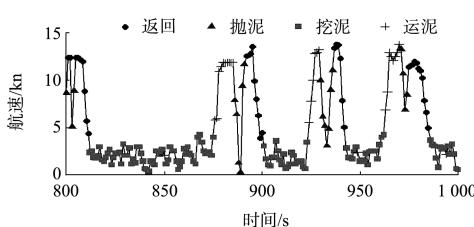


图 8 一段基于航速聚类的时序

### 2.3 模型验证

采用算法判断和人工判断对比分析的方法进行模型验证, 结果见表 3。挖泥、运泥、抛泥、返回预测的准确率分别为: 98.8%, 96.8%, 94.1%, 97.3%, 总体准确率为 97.5%, 表明算法精准度良好。

表 3 算法判断与人工判断对比矩阵

		人工判断			准确率/%
		挖泥	运泥	抛泥	
预 测 值	挖泥	491	1	0	5 98.8
	运泥	4	209	3	0 96.8
	抛泥	0	6	128	2 94.1
	返回	2	0	5	248 97.3
总计		497	216	136	255 97.5

### 3 结语

- 利用卡尔曼滤波算法能够很好地解决 AIS 数据存在着轨迹跳变的问题, 大大提升数据质量。
- 通过某耙吸挖泥船 1 a 的轨迹数据, 验证了 HDBSCAN 算法框架的有效性, 与人工判断结果比对的准确率高达 97.5%, 该算法能够很好地识别出不同密度数据集簇。
- HDBSCAN 算法聚类过程并未使用航速这一参数, 因此该算法适用于施工航速不同的工况, 扩大了模型的适用范围。

### 参考文献:

- [1] 秦亮, 刘长海, 李金峰, 等. 超大型耙吸船舱内泥泵汽蚀性能分析[J]. 水运工程, 2020(10): 178-181.
- [2] 庄素婕, 杨星, 刘克中, 等. 基于数据挖掘的耙吸式挖泥船行为辨识方法[J]. 中国航海, 2021, 44(4): 112-117.
- [3] 徐婷, 戴文伯, 鲁嘉俊. 基于自动识别系统大数据的船舶施工轨迹识别与预测[J]. 水运工程, 2019(12): 119-122, 126.
- [4] 杜志强, 谭玉琪, 仇林遥. 基于卡尔曼滤波的船舶轨迹异常行为快速检测方法[J]. 地理信息世界, 2021, 28(4): 112-118.
- [5] 张群, 洪志强. 抗差自适应无迹 Kalman 滤波在 GPS/INS 组合导航中的应用[J]. 北京测绘, 2021, 35(11): 1440-1446.
- [6] WANG L H, CHEN P F, CHEN L Y, et al. Ship AIS trajectory clustering: an HDBSCAN-based approach [J]. Journal of marine science and engineering, 2021, 9(6): 566.
- [7] MELVIN R L, XIAO J J, GODWIN R C, et al. Visualizing correlated motion with HDBSCAN clustering [J]. Protein science, 2018, 27(1): 62-75.