A Collaborative Anomaly Detection Approach of Marine Vessel Trajectory (Short Paper)

Zejun Huang[1], Jian Wan[1,2], Jie Huang[1(✉)], Gangyong Jia[1], and Wei Zhang[1]

[1] School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China
huangjie@hdu.edu.cn
[2] School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China

Abstract. Trajectory anomaly detection plays a very important role in navigation safety. Most trajectory anomaly detection methods mainly detect the spatial information of the vessel's trajectory. These methods neglect a vessel's dynamic behavior characteristics, such as course, speed, and acceleration. In this paper, a vessel trajectory multi-factor collaborative anomaly detection (VT-MCAD) approach is proposed to realize the anomaly detection of vessels at sea by studying the trajectory characteristics of vessels. Firstly, the trajectory behavior of historical vessels is identified, and the trajectory characteristics, such as course, speed, and acceleration, are extracted for different trajectory behaviors. Then, the current trajectory behavior is identified when the trajectory anomaly is detected. Based on the TRAjectory Outlier Detection (TRAOD) method, the corresponding trajectory feature model components, including instantaneous angle acceleration, average angle acceleration, instantaneous velocity, and the average velocity and acceleration, are used to detect the anomaly trajectory, and trajectory's suspicious degree of each component in VT-MCAD are obtained. Finally, the suspicious degree of each component is combined to calculate the final suspicious degree. VT-MCAD can change the weight of components according to the detection effectiveness of different components and avoid excessive dependence on one component, which results in better robustness and reliability. The experimental results based on real-world vessel data showed that VT-MCAD could effectively capture anomaly trajectories.

Keywords: Vessel trajectory · Collaborative anomaly detection · Multi-factor · Marine

1 Introduction

With the development of vessel monitoring systems in recent years, more and more maritime surveillance data is collected. Among them, vessel trajectory data is one of the most important data. The loss of personnel and property can be reduced through the detection and analysis of trajectory data by not only making early warnings for piracy, drug smuggling, and other situations but also aiding in timely rescues when vessels encounter bad weather, reef strikes, collisions, or other situations. Due to the large

X. Wang et al. (Eds.): CollaborateCom 2019, LNICST 292, pp. 279–294, 2019.
https://doi.org/10.1007/978-3-030-30146-0_19

number of vessels, the number of historical trajectories generated by them is also very large, and thus it is difficult to detect abnormalities manually. Therefore, it is necessary to use an intelligent algorithm to automatically analyze the abnormalities of the trajectories to help the monitors realize the abnormal monitoring of vessels.

A spatial trajectory is a trace generated by a moving object in geographical spaces, usually represented by a series of chronologically ordered points, where each point consists of a geospatial coordinate set and a time stamp, such as $p = (x, y, t)$. Trajectory anomalies can be items that are significantly different from the other items in terms of some similarity metric. They can also be events or observations that do not conform to an expected pattern [1]. Existing trajectory clustering or frequent pattern mining methods are usually used in trajectory anomaly detection. A trajectory may be abnormal if it cannot adapt to any cluster, or if it is infrequent. The factors affecting trajectory anomalies are not only reflected in unusual location points or sub-trajectories in the spatial domain but also hidden in the sequence of movements associated with a moving object [2].

In the maritime domain, most research of trajectory anomaly detection only considers the trajectory's spatial information and ignores the characteristics of vessel motion. These motion characteristics are the key attributes that describe the behavior of vessels and they become important factors for the anomaly detection of vessel trajectories. This paper proposes a collaborative anomaly detection approach of marine vessel trajectory, called vessel trajectory multi-factor collaborative anomaly detection (VT-MCAD). Based on TRAjectory Outlier Detection (TRAOD) [3], VT-MCAD takes the motion characteristics of the vessel's speed, direction, and acceleration as the components of vessel trajectory anomaly detection, and the anomaly detection result of each component is finally integrated into the trajectory's comprehensive anomaly trend score to achieve anomaly detection. The proposed method combines the spatial models and motion models of the trajectories in a multi-factor framework to generate more accurate detection.

2 Related Work

Most existing anomaly trajectory detection techniques are based on the distance, direction, and density of trajectories. However, classification and historical similarity-based techniques have also been proposed [4].

A distance-based approach was originally proposed by Knorr [5]. Lee [3] proposed the TRAOD algorithm which consists of two phases: partitioning and detection. In the first phase, each trajectory is partitioned first in coarse granularity and then in fine granularity. In the second phase, the outlying trajectory partitions are detected mainly using distance, and thus this phase is intuitive and efficient.

Ge [6] considered outliers in terms of direction and density. An evolving trajectory outlier detection method was provided, named TOP-EYE, which continuously computes the outlying score for each trajectory in an accumulating way. In TOP-EYE, a decay function is introduced to mitigate the influence of the past trajectories on the evolving outlying score, which is defined based on the evolving moving direction and density of trajectories. This decay function enables the evolving computation of the accumulated outlying scores along the trajectories.

Classification-based approaches establish a normal trajectory model and abnormal trajectory model through statistical information of the historical trajectory and generate a trajectory classifier. Li [7] proposed a Motion-Alert classification method for trajectory outlier detection that consists of the following three steps: (1) Object movement features, called motifs, are extracted from the object paths. Each path consists of a sequence of motif expressions associated with the values related to time and location. (2) Motif-based generalization is performed to discover anomalies in object movements, which clusters similar object movement fragments and generalizes the movements based on the associated motifs. (3) With motif-based generalization, objects are put into a multi-level feature space and are classified by a classifier that can handle high-dimensional feature spaces.

The historical similarity-based approach establishes a global feature model according to the frequent patterns of historical data mining, and then the data different from the global feature model are identified as abnormal trajectories. This method can usually be used for training data sets without labels. The historical similarity-based approach is widely used in navigation, road network traffic, and other fields. Lei [2] proposed a framework for maritime trajectory modeling and anomaly detection, called MT-MAD. The model considers the fact that anomalous behavior manifests in unusual location points and sub-trajectories in the spatial domain as well as in the sequence and manner in which these locations and sub-trajectories occur.

Currently, there is little research on trajectory anomaly detection for marine vessels and its application to the real world is still at an immature stage [8–12]. To promote the safety of marine navigation, there is an urgent need to design a more robust trajectory anomaly detection method.

3 Collaborative Trajectory Anomaly Detection Framework

3.1 The Framework

The framework of VT-MCAD is shown in Fig. 1. It may be quite difficult to directly identify whether a trajectory is an anomaly trajectory since anomalous trajectories are rare in the trajectory data set. Therefore, VT-MCAD assumes that each vessel trajectory is suspicious to varying degrees and it achieves anomaly detection by identifying the vessel trajectory's degree of suspiciousness [2]. When a trajectory anomaly detection is performed, VT-MCAD intercepts a trajectory segment of the vessel's adjacent time, and then recognizes the behavior of the trajectory segment, calculates the suspicious degree of its location, speed, direction, acceleration and so on, and integrates them to realize anomaly detection. As shown, the VT-MCAD method can be divided into two phases (shown in Fig. 1): trajectory modeling and anomaly detection.

3.2 TRAOD

TRAOD [3] is a basic component of VT-MCAD. The algorithm divides the trajectory into a set of trajectory segments and then calculates the distance between the trajectory segments to realize trajectory anomaly detection.

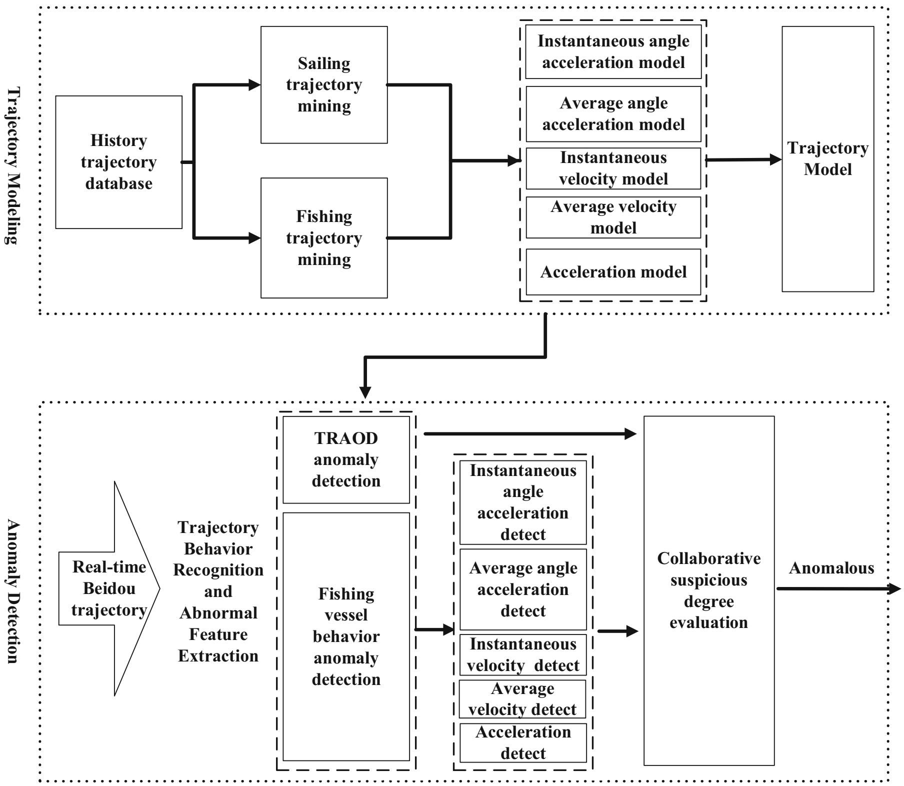


Fig. 1. The framework of VT-MCAD

TRAOD uses the Hausdorff distance, commonly used in pattern recognition, to define the distance between trajectory segments [13], which mainly consists of the perpendicular distance $(d_{\perp})$, parallel distance $\left(d_{\|}\right)$, and angle distance (d_{θ}).

Suppose there are two trajectory segments $S_1 = s_1e_1$, $S_2 = s_2e_2$, where s_i and $e_i(\mathrm{i} = 1, 2)$ are the starting and ending points of the trajectory segment, respectively. S_1 is the shorter one of the two trajectory segments. $d_{\perp}$, $d_{\|}$, and d_{θ} are shown in Formulas (1), (2), and (5), respectively. Among them, suppose that the projection points of the points s_1 and e_1 onto S_2 are p_s and p_e, respectively. $l_{\perp 1}$ is the Euclidean distance between s_1 and p_s and $l_{\perp 2}$ is the Euclidean distance between e_1 and p_e. $\|S_1\|$ is the length of S_1 and $\theta\left(0^{\circ} \leq \theta \leq 180^{\circ}\right)$ is the smaller intersection angle between S_1 and S_2. Finally, the distance between the two trajectory segments can be determined through Formula (6).

$$d_{\perp}(S_1, S_2) = \frac{l_{\perp 1}^2 + l_{\perp 2}^2}{l_{\perp 1} + l_{\perp 2}} \tag{1}$$

$$d_{\|}(S_1, S_2) = MIN\left(l_{\|1}, l_{\|2}\right) \tag{2}$$

$$l_{||1} = \mathrm{MIN}(\|p_s s_2\|, \|p_s e_2\|) \tag{3}$$

$$l_{||2} = \mathrm{MIN}(\|p_e s_2\|, \|p_e e_2\|) \tag{4}$$

$$d_\theta(S_1, S_2) = \begin{cases} \|S_1\| \times \sin(\theta), & \text{if } 0^\circ \le \theta < 90^\circ \\ \|S_1\|, & \text{if } 90^\circ \le \theta \le 180^\circ \end{cases} \tag{5}$$

$$\mathrm{dist}(S_1, S_2) = w_\perp \cdot d_\perp(S_1, S_2) + w_\| \cdot d_\|(S_1, S_2) + w_\theta \cdot d_\theta(S_1, S_2) \tag{6}$$

The adjacent trajectory, anomaly segment and anomaly trajectory are defined respectively below. Table 1 introduces the notation meanings used in the following definitions.

Table 1. Notation meanings

SYMBOL	Meaning
$\mathrm{len}(S_i)$	The length of the segment S_i
$\mathrm{P}(TR_i)$	The set of all segments of TR_i
$\mathrm{CP}(TR_i, S_j, \mathrm{D})$	The set of TR_i's segments within the distance D from $\mathrm{S_j} \in \mathrm{P}(\mathrm{TR_j})$ $(\mathrm{TR_i} \neq \mathrm{TR_j})$, i.e., $\{S_i \mid S_i \in \mathrm{P}(TR_i) \wedge \mathrm{dist}(S_i, \mathrm{S_j}) \le \mathrm{D}\}$
$\mathrm{CTR}(S_i, \mathrm{D})$	The set of trajectories close to S_i
$\mathrm{OP}(TR_i, \mathrm{D}, \mathrm{p})$	The set of outlying segments of TR_i

Definition 1: Adjacent Trajectory. $\mathrm{TR_i}$ is $\mathrm{S_j}$'s adjacent trajectory when inequality (7) is true, where $\mathrm{S_j} \in \mathrm{P}(\mathrm{TR_j})(\mathrm{TR_i} \neq \mathrm{TR_j})$.

$$\sum\nolimits_{\mathrm{S_i} \in \mathrm{CP}(\mathrm{TR_i}, \mathrm{S_j}, \mathrm{D})} \mathrm{len}(\mathrm{S_i}) \ge len(\mathrm{S_j}) \tag{7}$$

Definition 2: Anomaly Segment. $S_i \in P(TR_i)$ is an anomaly segment when inequality (8) is true. Here, $|\Gamma|$ denotes the total number of trajectories, and p is a parameter given by a user.

$$|CTR(S_i, D)| \le (1 - p)|\Gamma| \tag{8}$$

Definition 3: Anomaly Trajectory. TR_i is an anomaly trajectory when inequality (9) is true. In other words, when the proportion of the anomaly segments in the trajectory is not less than the threshold F (defined by the user), TR_i is an anomaly trajectory.

$$\frac{\sum_{S_i \in OP(TR_i, D, p)} len(S_i)}{\sum_{M_i \in P(TR_i)} len(M_i)} \ge \mathrm{F} \tag{9}$$

3.3 Trajectory Behavior Recognition

Trajectory behavior recognition of a fishing vessel is the basis of trajectory modeling [14–17]. In this study, the MSC-FBI algorithm proposed by Zhang et al. [18] is used to recognize fishing vessel behavior. The biggest difference between MSC-FBI and other fishing vessel behavior recognition algorithms is that the time distance between the trajectory points is added to the trajectory point distance measurement, and the distance between the trajectory points is considered comprehensively through the four dimensions of time, space, velocity, and direction. The time distance between two points is shown in Formula (10); the closer the time interval between the trajectory points $\mathrm{P_i}$ and $\mathrm{P_j}$ is, the smaller their time distance is. The spatial distance, velocity distance, and direction distance between $\mathrm{P_i}$ and $\mathrm{P_j}$ are shown in Formulas (11), (12), and (13), respectively, where $\mathrm{d(i)}$ is the direction change times within a given period of time $\{\mathrm{time_i} - \mathrm{t_d}, \mathrm{time_i} + \mathrm{t_d}\}$. Finally, the temporal-spatial distance between $\mathrm{P_i}$ and $\mathrm{P_j}$ is shown in Formula (14), where W is a weight vector.

$$\mathrm{T(i,j)} = \frac{\left|time_i - time_j\right|}{\max(T) - \min(T)} \tag{10}$$

$$\mathrm{S(i,j)} = \frac{\sqrt{\left(lon_i - lon_j\right)^2 + \left(lat_i - lat_j\right)^2}}{\max(S) - \min(S)} \tag{11}$$

$$\mathrm{V(i,j)} = \frac{\left|Speed_i - Speed_j\right|}{\max(V) - \min(V)} \tag{12}$$

$$\mathrm{DIR(i,j)} = |d(i) - d(j)| \tag{13}$$

$$\mathrm{D(i,j)} = W[T(i,j)\ S(i,j)\ V(i,j)\ DIR(i,j)]^T \tag{14}$$

The following is a brief introduction to the implementation steps of the MSC-FBI algorithm. Data set D contains a historical trajectory that needs behavioral recognition. The specific steps are as follows:

(1) Use DBSCAN and the space-time distance measurement method to cluster the trajectory in D and obtain the trajectory segments of the different behaviors.
(2) Use K-Means to cluster the trajectory segments that were obtained in Step (1), and the trajectory segments of the same behavior pattern are clustered into one cluster.
(3) After obtaining the trajectory segments of same behavior pattern, model the trajectory of different behavior patterns.
(4) Use the behavior model obtained in Step (3) to identify the trajectory behavior.

3.4 Trajectory Modeling

Trajectory modeling is used to establish the behavior model of the fishing vessels, including five trajectory characteristic models: instantaneous angle acceleration, average angle acceleration, instantaneous velocity, average velocity, and acceleration. It is necessary to distinguish between behaviors when modeling due to the different behavior patterns of fishing vessels' sailing state and fishing state. Figure 2 shows the trajectory modeling process of the fishing vessel, in which the behavior of the fishing vessel is divided into two categories: sailing behavior and fishing behavior.

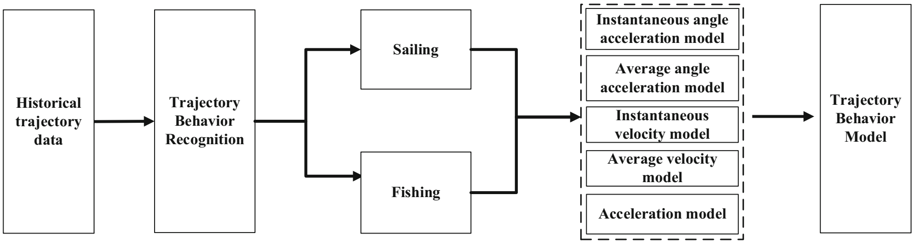


Fig. 2. Trajectory modeling process of the fishing vessel

The trajectory data of the fishing vessel discussed in this paper were positioned and transmitted by the BeiDou navigation satellite system (BDS). The trajectory information includes longitude, latitude, timestamp, direction, speed, and temperature of the fishing vessel. The behavior patterns of fishing vessels in different states can be extracted and the trajectory model can be established through the study and analysis of these attributes.

Traditional anomaly detection modeling methods can be divided into supervised learning and unsupervised learning. Supervised learning needs to label the data in the training set before training the model, and then the anomaly detection model is built by classifying training set labels. In contrast, unsupervised learning does not need to label the training set data before beginning training but implements anomaly detection through data features. Because anomalies are rare in the trajectory data of this study, it was not possible to construct a rich training set for supervised learning, and therefore an unsupervised one-class SVM was used to establish the trajectory model.

One-class SVM [19] is a common method in the field of anomaly detection. It is a variant of the SVM algorithm. SVM is a supervised algorithm, and its essence is to find a hyperplane with the largest classification interval to realize data classification. SVM's training set is divided into two parts: metadata and the classification label. One-class SVM is an unsupervised algorithm in which the metadata does not need any data labels in its training set. The goal of one-class SVM is to find a hyperplane in the feature space so that most of the training patterns are in front of the hyperplane and the distance between the hyperplane and the origin is maximized. In other words, one-class SVM actually finds a hyperplane in the feature space to separate the origin and training patterns, and the distance between the hyperplane and the origin is maximized.

The maximum edge problem is the core of one-class SVM. Therefore, the problem can be formulated as follows:

$$\begin{gathered} min_{w,\varepsilon,\rho}\frac{1}{2}\|w\|^2 + \frac{1}{mC}\sum_i \varepsilon_i \\ w\cdot\varnothing(x_i)\geq\rho-\varepsilon_i, \varepsilon_i\geq 0, \forall i=1,\ldots,m \end{gathered} \tag{15}$$

where w is a vector orthogonal to the hyperplane; C represents the fraction of training patterns that are allowed to be rejected; x_i is the i^{th} training pattern; m is the total number of training patterns; $\varepsilon = \varepsilon_1,\ldots,\varepsilon_m$ is a vector of slack variables used to "penalize" the rejected patterns; and ρ represents the margin, that is, the distance from the origin to the hyperplane.

The trajectory modeling algorithm is summarized as follows:

Algorithm 1 Trajectory modeling.

Input: Sailing trajectory training set $D_s = \{ST_1, ST_2, \ldots, ST_n\}$,
Fishing trajectory training set $D_f = \{FT_1, FT_2, \ldots, FT_n\}$
Output: Trajectory Model M
// Using instantaneous direction, average direction,
// instantaneous velocity, average velocity and acceleration
// established corresponding models respectively.
for each $ST_i(FT_i)$ **in** $D_s(D_f)$**:**
  for each P_j **in** $ST_i(FT_i)$**:**
    $S^s_{MD}(S^f_{MD}) \leftarrow$ **Calculate Moment Direction for** P_j
    $S^s_{AD}(S^f_{AD}) \leftarrow$ **Calculate Average Direction for** P_j
    $S^s_{MS}(S^f_{MS}) \leftarrow$ **Calculate Moment Velocity for** P_j
    $S^s_{AS}(S^f_{AS}) \leftarrow$ **Calculate Average Velocity for** P_j
    $S^s_{A}(S^f_{A}) \leftarrow$ **Calculate Acceleration for** P_j
    // Modeling instantaneous angle acceleration by
    // instantaneous direction
  $M^s_{MD}, M^f_{MD} \leftarrow$ `Model`(S^s_{MD}), `Model`(S^f_{MD})
  // Modeling average angle acceleration by average
  // direction
  $M^s_{AD}, M^f_{AD} \leftarrow$ `Model`(S^s_{AD}), `Model`(S^f_{AD})
  // Modeling of Instantaneous velocity
  $M^s_{MS}, M^f_{MS} \leftarrow$ `Model`(S^s_{MS}), `Model`(S^f_{MS})
  // Modeling of average velocity
  $M^s_{AS}, M^f_{AS} \leftarrow$ `Model`(S^s_{AS}), `Model`(S^f_{AS})
  // Modeling acceleration
  $M^s_{A}, M^f_{A} \leftarrow$ `Model`(S^s_{A}), `Model`(S^f_{A})
  M $\leftarrow \{M^s_{MD}(M^f_{MD}), M^s_{AD}(M^f_{AD}),, M^s_{MS}(M^f_{MS}),$
  $M^s_{AS}(M^f_{AS}), M^s_{A}(M^f_{A})\}$
  `return` **M**

The instantaneous angle acceleration model M^s_{MD}, average angle acceleration model M^s_{AD}, instantaneous velocity model M^s_{MS}, average velocity model M^s_{AS}, and acceleration model M^s_{A} are the five independent components of the trajectory model. When the anomaly trajectories of fishing vessels are detected, the suspicious degree of each component (including TRAOD) are calculated separately, and then the suspicious degrees are combined to realize anomaly detection.

3.5 Anomaly Trajectory Detection of Vessels

In the anomaly detection phase, this method integrates the suspicious degree of the trajectory model's different components to obtain more reliable anomaly detection results. Among them, the suspicious degree of the TRAOD component is calculated using Formula (16), and the suspicious degrees of the other components are calculated by Formula (17), where $OPS(TR_i)$ represents the anomaly point set in trajectory TR_i and $PS(TR_i)$ represents all the points in TR_i.

$$\text{Suspicious Degree}_{TRAOD} = \frac{\sum_{S_i \in OP(TR_i, D, p)} len(S_i)}{\sum_{M_i \in P(TR_i)} len(M_i)} \tag{16}$$

$$\text{Suspicious Degree}_{others} = \frac{OPS(TR_i)}{PS(TR_i)} \tag{17}$$

Different from the traditional anomaly detection algorithm, the motion characteristics of vessel trajectory data are modeled and treated as different components of VT-MCAD. After obtaining the suspicious degree of each component, the final anomaly detection results are obtained by using the predefined combination strategy. In addition, in the process of suspicious degree combination, different characteristic models of the algorithm are given different weights according to the model availability so as to deal with the anomaly detection sensitivity of different characteristics.

Specifically, the different characteristics of vessel trajectory have different meanings. It is necessary to effectively deal with the suspicious degrees while combining them. Traditional combination methods mainly include cumulative sum and sorting methods, but both methods have shortcomings. For example, the cumulative sum method may cause the final result to be excessively dependent on the component and weaken the influence of other components when the suspicious degree of a component is abnormally large. Additionally, the sorting method may cause conflicts in the individual component results. To solve these problems, each component of VT-MCAD is weighted based on the cumulative sum. The weight of each component is determined by the availability of the component in different application scenarios. The weighted cumulative sum algorithm is as follows:

Algorithm 2 Weighted cumulative sum.

Input: Component suspicious degree set S = $\{S_1, S_2, \ldots, S_n\}$
Output: Final suspicious degree S_f

1: **Foreach i=1 to n**
2: **Assign** w_i **to** S_i
3: $S_f = \sum_{i=1}^{n} w_i \cdot S_i$
4: **Return** S_f

The above algorithm assigns a weight w to the suspicious degree obtained by the different components in VT-MCAD, and $w_1 + w_2 + \ldots + w_n = 1$, where n is the number of VT-MCAD components. Figure 3 shows a flow chart of the vessel trajectory anomaly detection. After inputting the trajectory of the vessel, the six component models of VT-MCAD calculate the suspicious degree separately, and then the components of each component are integrated to acquire the final suspicious degree of the current trajectory.

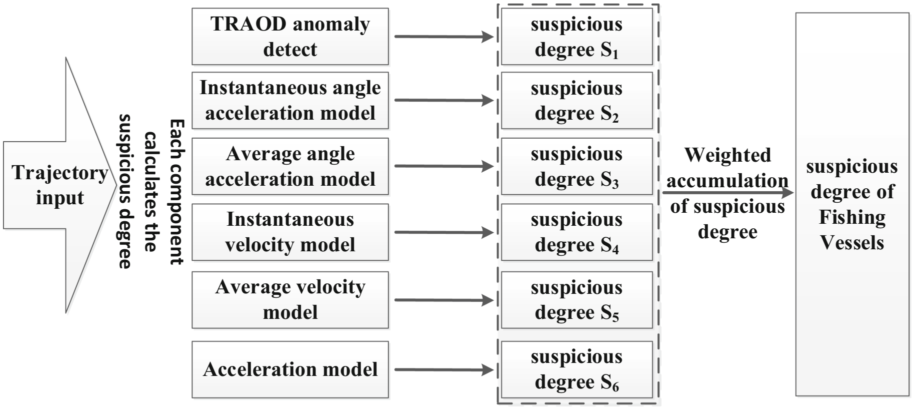


Fig. 3. Flow chart of vessel trajectory anomaly detection

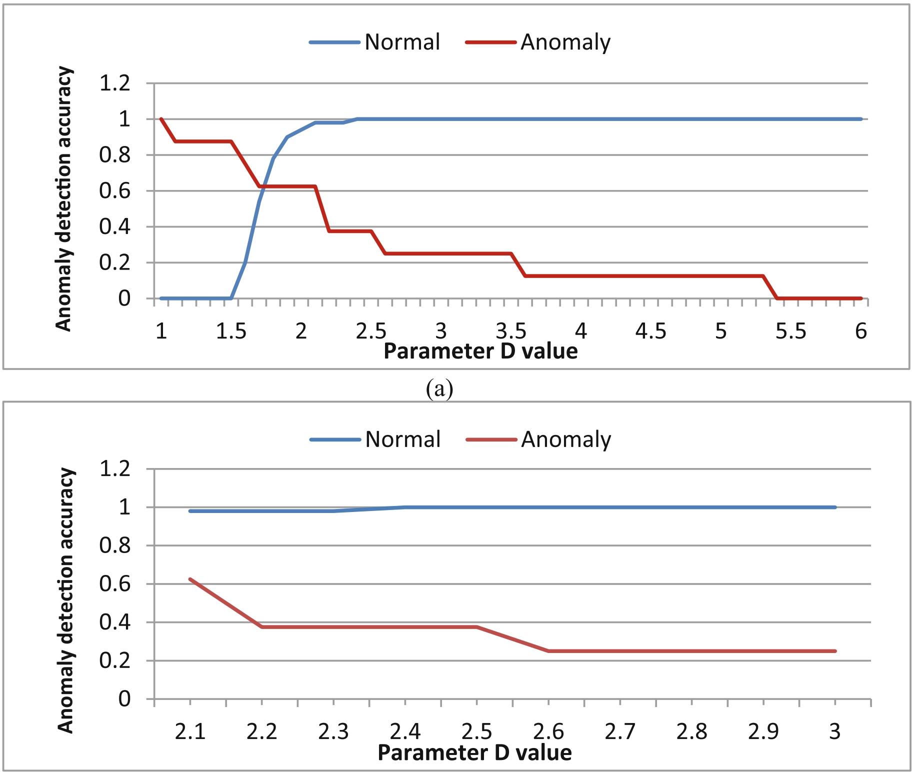
4 Experiments

The experiment used the trajectory data set in a dynamic fishing vessel management system operated in Zhoushan, China. The system utilizes BDS to locate vessels and transmit real-time data. Its main functions include vessel inquiry, track playback, alarm rescue, and voyage statistics. The experimental data were collected from December 22, 2016, to November 8, 2018, from a total of 220 vessels, and was about 0.6 GB in size.

The effectiveness of the VT-MCAD method proposed in this paper was verified by comparing the effect of TRAOD and VT-MCAD on anomaly detection of the above trajectory dataset. Since the detection methods of the two algorithms are basically the same under the two behaviors of fishing vessel sailing and fishing, this study only detected anomalies of the fishing vessel's sailing trajectory.

4.1 TRAOD Anomaly Detection

After cleaning and filtering the trajectory data, 6,110 trajectory data were obtained. The TRAOD trajectory anomaly detection effect was verified on a test set containing 50 normal trajectories and eight anomaly trajectories. During the experiment, the parameter p was set to 0.9 and the parameter F was set to 0.2. Figure 4(a) shows the accuracy of TRAOD for the detection of different parameters D values. As the value of D increases, the detection accuracy of the normal trajectory continuously increases, and the detection accuracy of the anomaly trajectory continuously decreases. Figure 4(b) shows the anomaly detection accuracy of the parameter D from 2 to 3 interval 0.1. As shown, when the parameter D is greater than 2.4, the normal trajectory detection accuracy is 100%, and the accuracy of anomaly trajectory detection is 25%–37.5%.



(b)

Fig. 4. Anomaly detection accuracy of TRAOD with different parameter *D* values

The experimental results showed that TRAOD is not ideal for anomaly detection of a fishing vessel's trajectory. In the case of not tolerating the misjudgment of the normal trajectory, the final detection ratio of the TRAOD to the anomaly trajectory was only 37.5%, and the risk of the algorithm erroneously judging normal trajectories was large. When the parameter D was increased to reduce misjudgment, the accuracy of the anomaly trajectory detection was further reduced to 25%. Therefore, TRAOD can detect a part of the anomaly trajectory, but the detection accuracy is low and the effect is poor.

4.2 VT-MCAD Anomaly Detection

Tables 2 and 3 show the suspicious degree of VT-MCAD's different components. There were ten normal trajectories and eight anomaly trajectories in the test set, of which the table respectively describes the suspicious degree for the normal and anomaly trajectories. The parameter D in the TRAOD component was 2.6. Since the components in the VT-MCAD had similar performances under the experimental scenario, this study used the simple averaging method to integrate the component results [20]. It can be seen that although the trajectory suspicious degrees of different trajectory features were different, the anomaly detection accuracy of each component was generally higher. Finally, the suspicious degree distribution of each trajectory is shown in Fig. 5. The suspicious degrees of the anomaly trajectories in the graph were significantly higher than those of normal trajectories.

Table 2. VT-MCAD components suspicious degree distribution of the normal trajectory

Components	Normal trajectory									
	1	2	3	4	5	6	7	8	9	10
Instantaneous angle acceleration	0.31	0.05	0.19	0.07	0.07	0.07	0.11	0.00	0.15	0.05
Average angle acceleration	0.31	0.05	0.00	0.21	0.27	0.00	0.17	0.00	0.05	0.32
Instantaneous velocity	0.23	0.21	0.43	0.07	0.33	0.07	0.00	0.19	0.10	0.05
Average velocity	0.31	0.05	0.25	0.07	0.27	0.07	0.00	0.38	0.15	0.16
Acceleration	0.23	0.26	0.38	0.00	0.13	0.07	0.00	0.00	0.15	0.11
TRAOD	0.28	0.32	0.22	0.22	0.21	0.54	0.19	0.41	0.47	0.43

Table 3. VT-MCAD components suspicious degree distribution of the anomaly trajectory

Components	Anomaly trajectory							
	1	2	3	4	5	6	7	8
Instantaneous angle acceleration	0.35	0.17	1.00	0.38	0.50	0.21	0.64	0.23
Average angle acceleration	0.24	0.28	0.33	0.44	0.38	0.14	0.27	0.15
Instantaneous velocity	0.41	0.11	0.67	0.88	0.88	0.79	0.72	0.62
Average velocity	0.29	0.06	1.00	0.81	1.0	0.36	0.91	0.54
Acceleration	0.29	0.06	0.07	0.31	0.25	0.50	0.27	0.38
TRAOD	0.87	0.09	0.21	0.40	0.63	0.24	0.59	0.95

The experimental results showed that compared with TRAOD, the VT-MCAD algorithm proposed in this paper had a higher accuracy of anomaly detection, and the algorithm considers various trajectory features, such as position, speed, and direction, which make VT-MCAD more robust and reliable.

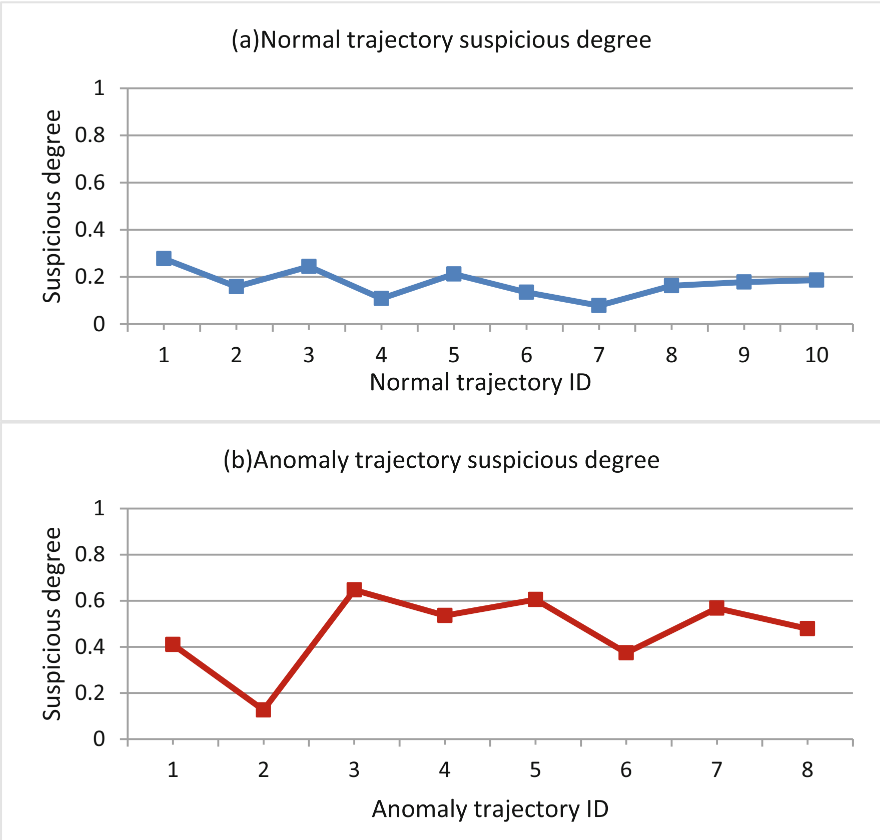


Fig. 5. VT-MCAD suspicious degree

4.3 Algorithm Evaluation

To compare the effectiveness of TRAOD and VT-MCAD, the Precision, Recall, and F-measure evaluation criteria are used, as shown in Formulas (18), (19), and (20), where R represents the known anomaly trajectory data set; D represents the anomaly result set obtained after the algorithm is executed; Precision and Recall are used to evaluate the accuracy and completeness of the algorithm anomaly detection result, respectively; and the F-measure index integrates the Precision and Recall evaluation results.

$$\text{Precision} = \frac{|R \cap D|}{|D|} \tag{18}$$

$$\text{Recall} = \frac{|R \cap D|}{|R|} \tag{19}$$

$$\text{Recall} = \frac{|R \cap D|}{|R|} \tag{20}$$

The algorithm anomaly detection performance evaluation is based on the maximum order of the anomaly detection results in Tables 2 and 3 and Fig. 5, and the top ten trajectories of the maximum ordering were defined as anomaly trajectories. Figures 6, 7 and 8 show the results of VT-MCAD and the Precision, Recall, and F-measure of its various components. It can be seen from the figure that under the current trajectory data set, the anomaly detection accuracy of the TRAOD was 0.5, which was 0.2–0.3 lower than the other components of VT-MCAD, and the final accuracy of VT-MCAD was 0.2 higher than that of TRAOD. In terms of the completeness of anomaly detection, the proportion of anomaly trajectories detected by other components of VT-MCAD to the total number of anomaly trajectories exceeded 0.8 while that of TRAOD was only 0.6. By combining the above two evaluation indicators, the F-measure evaluation results of VT-MCAD were significantly better than TRAOD.

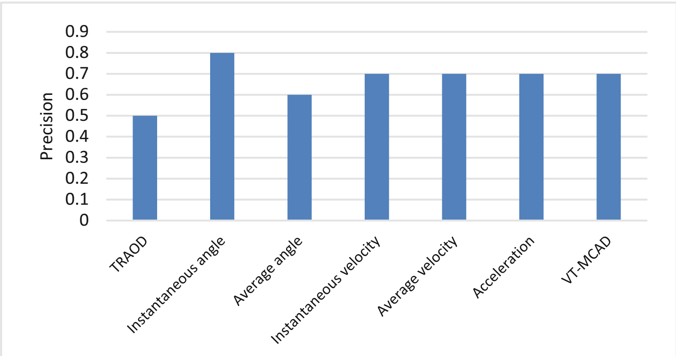


Fig. 6. Precision evaluation

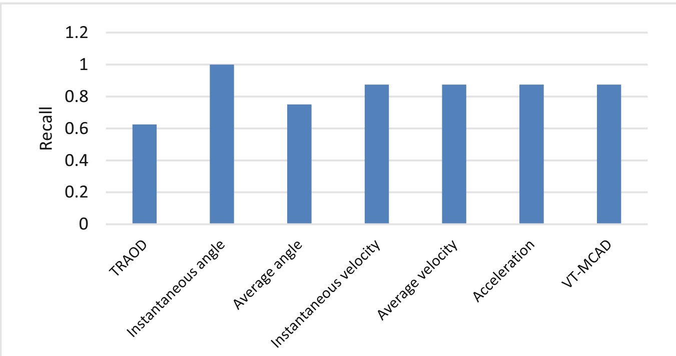


Fig. 7. Recall evaluation

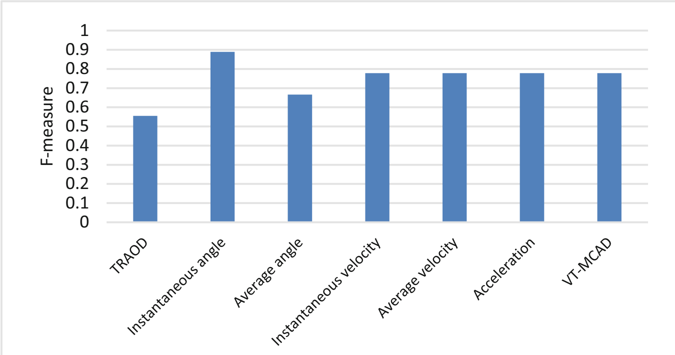


Fig. 8. F-measure evaluation

5 Conclusion

In this paper, a multi-factor collaborative approach to detecting anomaly trajectories of marine vessels is proposed. This approach simultaneously considers the vessel's speed, direction, and acceleration. Our VT-MCAD approach was evaluated using data from the Zhoushan fishing vessel management system. Results showed that the accuracy and completeness of the VT-MCAD in the experiments trajectory dataset were significantly better than the TRAOD, which verifies the effectiveness of the VT-MCAD. Future work shall consider more factors by exploiting the knowledge of marine meteorology and hydrology.

Acknowledgment. This work was supported in part by the Key Research and Development Project of Zhejiang Province (Grant No. 2017C03024) and the National Natural Science Foundation of China (Grant No. 61572163, No. 61872119).

References

1. Zheng, Y.: Trajectory data mining: an overview. ACM Trans. Intell. Syst. Technol. **6**(3), 29 (2015)
2. Lei, P.R.: A framework for anomaly detection in maritime trajectory behavior. Knowl. Inf. Syst. **47**(1), 189–214 (2016)
3. Lee, J.G., Han, J., Li, X.: Trajectory outlier detection: a partition-and-detect framework. In: IEEE International Conference on Data Engineering, Cancun, pp. 140–149. IEEE (2008)
4. Gupta, M., Gao, J., Aggarwal, C.C.: Outlier detection for temporal data: a survey. IEEE Trans. Knowl. Data Eng. **26**(9), 2250–2267 (2014)
5. Knorr, E.M., Raymond, T.N., Vladimir, T.: Distance-based outliers: algorithms and applications. VLDB J. **8**(3–4), 237–253 (2000)
6. Ge, Y., Xiong, H., Zhou, Z., Ozdemir, H., Yu, J., Lee, K.: Top-Eye: top-k evolving trajectory outlier detection. In: Proceedings of the 19th ACM Conference on Information and Knowledge Management, Toronto, pp. 1733–1736. ACM (2010)

7. Li, X., Han, J., Kim, S.: Motion-Alert: automatic anomaly detection in massive moving objects. In: IEEE International Conference on Intelligence and Security Informatics, San Diego, vol. 3975, pp. 166–177. IEEE (2006)
8. Chang, S., Yeh, K., Peng, G.: From safety to security - pattern and anomaly detections in maritime trajectories. In: International Carnahan Conference on Security Technology, Taipei, pp. 415–419. IEEE (2015)
9. Zhen, R., Jin, Y., Hu, Q.: Maritime anomaly detection within coastal waters based on vessel trajectory clustering and naive bayes classifier. J. Navig. **70**(3), 648–670 (2017)
10. Tu, E., Zhang, G., Rachmawati, L.: Exploiting AIS data for intelligent maritime navigation: a comprehensive survey from data to methodology. IEEE Trans. Intell. Transp. Syst. **19**(5), 1559–1582 (2018)
11. Zhao, L., Shi, G.: A trajectory clustering method based on Douglas-Peucker compression and density for marine traffic pattern recognition. Ocean Eng. **172**, 456–467 (2019)
12. Pallotta, G., Vespe, M., Bryan, K.: Vessel pattern knowledge discovery from AIS data: a framework for anomaly detection and route prediction. Entropy **15**(6), 2218–2245 (2013)
13. Chen, J., Leung, M.K., Gao, Y.: Noisy logo recognition using line segment Hausdorff distance. Pattern Recogn. **36**(4), 943–955 (2003)
14. Marzuki, M.I., Gaspar, P., Garello, R.: Fishing gear identification from vessel-monitoring-system-based fishing vessel trajectories. IEEE J. Oceanic Eng. **43**(3), 689–699 (2018)
15. Fernando, T., Mercedes, V., Antonio, F.S.: A complex event processing approach to detect abnormal behaviours in the marine environment. Inf. Syst. Front. **18**(4), 765–780 (2016)
16. Dabrowski, J.J., de Villiers, J.P., Beyers, C.: Context-based behaviour modelling and classification of marine vessels in an abalone poaching situation. Eng. Appl. Artif. Intell. **64**, 95–111 (2017)
17. de Souza, E.N., Boerder, K., Matwin, S.: Improving fishing pattern detection from satellite AIS using data mining and machine learning. PLoS ONE **11**(7), e0158248 (2016)
18. Zhang, J., Geng, J., Wan, J., et al.: An automatically learning and discovering human fishing behaviors scheme for CPSCN. IEEE Access **6**, 19844–19858 (2018)
19. Perdisci, R., Gu, G., Lee, W.: Using an ensemble of one-class SVM classifiers to harden payload-based anomaly detection systems. In: International Conference on Data Mining, Hong Kong, pp. 488–498. IEEE (2006)
20. Zhou, Z.H.: Ensemble Methods - Foundations and Algorithms. Taylor & Francis, Milton Park (2012)