TRANSNAV
http://www.transnav.eu
the International Journal
on Marine Navigation
and Safety of Sea Transportation
Volume 7
Number 1
March 2013

DOI: 10.12716/1001.07.01.13

A Clustering Analysis for Identifying Areas of Collision Risk in Restricted Waters

Z. Xiang, Q. Hu & C. Shi
Merchant Marine College, Shanghai Maritime University, China

ABSTRACT: The identification of areas of collision risk in restricted waters could play an important role in VTS services. Based on the concept of ship domain, this paper introduces a model for identifying collision risk between vessels in restricted waters, then puts forward an improved DBSCAN clustering algorithm for identifying areas of high collision risk, finally, the visualization algorithm is presented. The experimental results in this paper show the algorithm is capable of identifying and rendering areas of collision risk in restricted waters.

1 INTRODUCTION

Identification of areas of collision risk in restricted waters is not only important to measure marine traffic safety, but also helpful for marine traffic guiding and controlling.

Many methods have been put forward for measuring the collision risk between vessels, such as fuzzy theory [1], ANN[2], ship's DCPA and TCPA [3], traditional clustering algorithm [4], etc. Generally, these methods are suitable for the high seas or open waters because they do not consider ships' dimension, which is a quite important factor for evaluating the collision risk between vessels in restricted waters.

AIS technology makes it practical to collect ships' dimension information. Ship's domain, which is closely related to the ship's dimension, is applied to measure and identify the collision risk between vessels in restricted waters, and then an improved clustering algorithm based on DBSCAN [5], is put forward for gathering the ships of collision risk and areas of collision risk will be developed, finally, the border of the areas is smoothed.

2 AN IMPROVED DBSCAN ALGORITHM FOR INDENTIFING AREAS OF COLLISION RISK

For measuring collision risk between vessels in restricted waters, this paper images ships in the restricted waters as each independent polygon object. these objects are clustered to identify area of collision risk. The so-called clustering is the method grouping data object into several classes or cluster, the objects in same class has high similarity, otherwise has high difference. Clustering algorithm is an important part in Data Mining, commonly used clustering algorithm are DBSCAN algorithm, OPTICS algorithm, DENCLUE algorithms, CLIQUE algorithm, K - MEANS algorithm, etc [4].

In the above algorithms, DBSCAN is typical clustering algorithms which based on density clustering method, its high dimensional data handling and noise object eliminating effect is better than the other clustering algorithm, so after improving it is suitable to clustering geometry objects. In this paper the traffic flow in restricted waters has the same characteristic (geometry type), so we choose DBSCAN algorithm for prototype clustering algorithm.

In DBSCAN algorithm, the data object to be clustering is seen as a particle, obviously it does not fit the situation in the restricted waters. Taking this problem into account, this paper introduces the concept of the ship domain [6,7,8,9,10].

Ship domain is the area that every ship hopes to maintain an independent region from otherness around itself in order to avoid collisions. Factors affecting the ship domain include: Ship's size, speed, state of motion, manipulation and movement of the ship's performance, encounter posture, manipulator's psychological factors, traffic density in water, traffic environment, and then hydro-meteorological, etc.

This paper combines ship domain concept and DBSCAN clustering algorithm, proposed a improved DBSCAN algorithm.It selects the model of ship domain as follow:

1 Ship domain is oval.(it is simplified as octagon when calculate in computer).
2 Ship domain can modify according to ship's size,speed,etc.
3 Ship itself locates in rearward position of its ship domain.
4 Only when two ships enter into each other's ship domain, there is a collision risk.

Figure 2.1 show the ship domain model used in this paper.

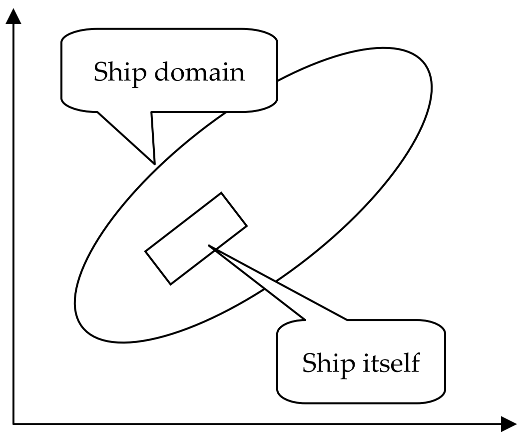

Figure 2.1. Ship domain domain

After selected model of ship domain, this paper described the improved DBSCAN algorithm as below.

Definition 1 Objects:in this paper, the objects means ships.

Definition 2 ε neighborhood: the region(polygon) of objects' ship domains can be called object's ε neighborhood.

Definition 3 density: the density of a object p is that the number of objects be contained in object p's ε neighborhood.

Definition 4 Core object: if an object's ε neighborhood at least contains the minimum number (MinPts) objects, it named core object.

Definition 5 Directly density-reachable: in a given set of objects marked D, for a given ε neighborhood, if the object p and the object q both in each other's ε neighborhood ,and q is a core object, then we call the object p is directly density-reachable from the object q.

Definition 6 density-reachable: for exist objects chain $p_1, p_2, \ldots p_n, p_1 = q, p_n = p$,and $p_i \in D, (1 \le i \le n)$,if every p_{i+1} is directly density-reachable from p_i in the condition of given ε neighborhood and MinPts .we call the object p is density-reachable from the object q.(transmission).

Definition 7 density-connected: in a given set of objects D , ε and MinPts, object p and object q are both density-reachable from the object O, then the object p and object q is density-connected.

Definition 8 class with noise: in given ε and MinPts, a class C is non-empty subset of given set of objects D satisfies the following three conditions:

1 For any p, q ∈ D, if q ∈ C, and P is density-reachable from q, then p ∈ C;
2 For any p, q ∈ C, p and q are density-connected;
3 Do not belong to any class of objects is considered noise.

Improved DBSCAN algorithm through ε and MinPts as input parameters to control the density of the class, only class the density more than MinPts can be retain . The clustering process is based on the following procedures:

(1) Given parameters ε and MinPts ,if exists core object P , every objects density-reachable from P satisfies ε and MinPts clusters a class and P belongs to this class;

(2) Assuming that class C is satisfies ε and MinPts, P is a core object of class C, then the class C is equivalent to the set contains objects density-reachable from P.

Algorithm's idea is that we trace and find points(ships) to cluster through checking database (AIS database).Firstly, we take different time data synchronous to the same time, this step involving dead reckoning. If ε neighborhood of point P (ship P) contains more than MinPts number points, then create a new class (collision risk area) which core is P. Next, repeatedly looking up to some objects which is density-reachable from the core P, in this process may involve some class combined. When there is no new point can be added to any class, it is the end of the clustering process. Figure 2.2 and figure 2.3 respectively shows the DBSCAN's schematic diagram and the improved DBSCAN algorithm's schematic diagram.

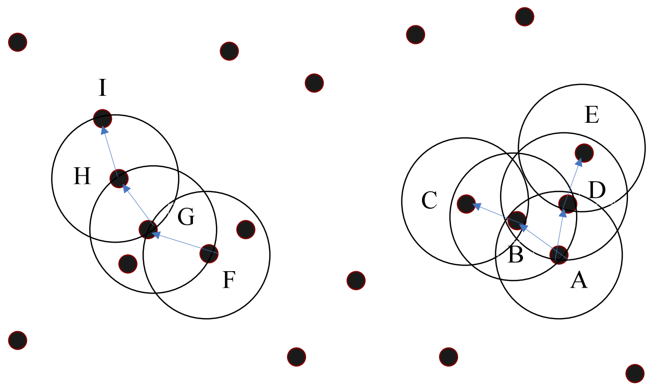

Figure 2.2. DBSCAN's schematic diagram

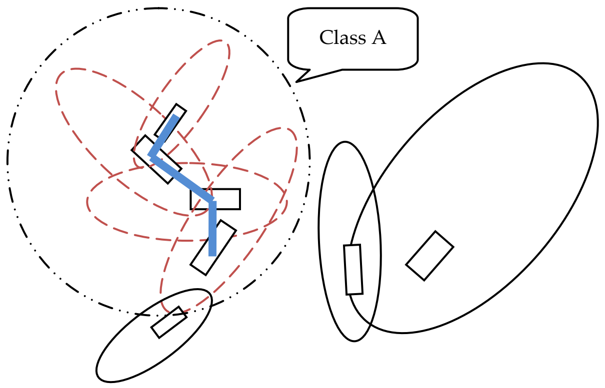

Figure 2.4. Improved DBSCAN algorithm's schematic diagram

It seems that improved DBSCAN algorithm has the best recognition effect among three algorithm above, specifically in the restricted waters.

3 ALGORITHM FOR COMBINING AREAS OF COLLISION RISK

Clustering results in the shape of the high-risk areas of the ship presented overlapping geometric and disorganized, in order to let the drawing area to contain the entire polygon outer envelope for identifying collision risk areas, we need to design a polygon merge algorithm.

This paper used the method of traversing coordinates of spatial point to combine collision risk areas, and allowed the points arrange in clockwise or counterclockwise. Its purpose is formed through set of points form vertices of multiple polygon (named G)

$$G=\{p_1(x_1,y_1),p_2(x_2,y_2)\ldots\ldots,p_i(x_i,y_i)\}(i\in N)$$

acquired the queue of points in connection diagram(named T)

$$T=\{p_1(x_1,y_1),p_2(x_2,y_2)\ldots\ldots,p_j(x_j,y_j)\}(j\in N)$$

and make the queue T can include all points of G.

The algorithms described are as follows:

The set of points in the Cartesian coordinate system is

$$G=\{p_1(x_1,y_1),p_2(x_2,y_2)\ldots\ldots,p_i(x_i,y_i)\}(i\in N)$$

(1) Calculated maximum and minimum two points in G sorting x,named $p_{\max}$ and $p_{\min}$.

$$p_{\max}=(x_{\max},y_{\max}),x_{\max}=\max(x_1,x_2\ldots\ldots,x_i),p_{\max}\in G \quad (3.1)$$

$$p_{\min}=(x_{\min},y_{\min}),x_{\min}=\min(x_1,x_2\ldots\ldots,x_i),p_{\min}\in G \quad (3.2)$$

Point $p_{\max}$ as the first point of T queue.

(2) Make a line linking $p_{\max}$ and $p_{\min}$,its equation is:

$$y=((y_{\max}-y_{\min})/(x_{\max}-x_{\min}))(x_k-x_{\min})+y_{\min} \quad (3.3)$$

It can divide G into upper part and lower part:

$$G_{up}=\{p_1(x_1,y_1),p_2(x_2,y_2)\ldots\ldots,p_k(x_k,y_k)\}$$
$$(k\in N,y_k>((y_{\max}-y_{\min})/(x_{\max}-x_{\min}))(x_k-x_{\min})+y_{\min}) \quad (3.4)$$

$$G_{down}=\{p_1(x_1,y_1),p_2(x_2,y_2)\ldots\ldots,p_t(x_t,y_t)\}$$
$$(t\in N,y_t>((y_{\max}-y_{\min})/(x_{\max}-x_{\min}))(x_t-x_{\min})+y_{\min}) \quad (3.5)$$

(3) Let set G_{up} sorted in x descending order, sequentially transferred the points in G_{up} to the queue T until G_{up} is empty, and then added point $p_{\min}$ into T for the end point as an upper part boundary;

(4) Let set G_{down} sorted in x ascending order, sequentially transferred the points in G_{down} to the queue T until G_{down} is empty, and then added point $p_{\max}$ into T for the end point as an lower part boundary and also end point of the whole queue T.

Algorithm schematic diagram as follows:

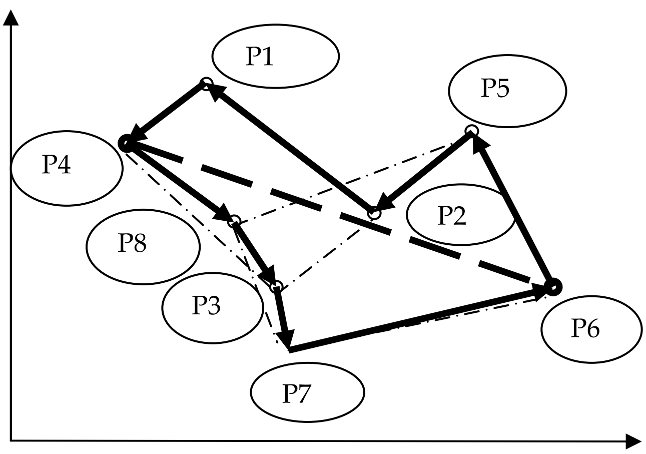

Figure 3.1. Algorithm of combining collision risk areas

In figure 3.1, the dotted line around the two quadrilateral vertex,they are $G_1 = \{p_1, p_2, p_3, p_4\}$ and $G_2 = \{p_5, p_6, p_7, p_8\}$.

The set of points form vertices of G_1 polygon and G_2 polygon named

$$G = \{p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_1\}$$

can accordance with the above algorithm to acquire a queue of points in connection diagram

$$T = \{p_6, p_5, p_2, p_1, p_4, p_8, p_3, p_7, p_6\}$$

In Figure 3.1, the solid line in the direction of the direction of the envelope. P_4 and P_6 are $p_{\min}$ and $p_{\max}$, queue T's sequence is counterclockwise.

4 ALGORITHM FOR SMOOTHING THE BORDER OF AREA OF COLLISION RISK

The combined polygon's boundary is always rough and exists pits, this paper introduced a algorithm to smooth polygons, this method through calculating polygon's area to remove surplus pits and make polygon mellow and looks good.

The algorithm's purpose is that through the connection diagram (named T)

$$T = \{p_1(x_1, y_1), p_2(x_2, y_2)......, p_j(x_j, y_j)\}(j \in N)$$

to find another connection diagram R, and cause R including all points of T and have maximum area.

The recursive algorithm is described as follows:

(1) input connection diagram T, and calculate its spatial area called T_{area}.

(2) Sequentially removed a point $p_k(x_k, y_k)$ in

$$T = \{p_1(x_1, y_1), p_2(x_2, y_2)......, p_j(x_j, y_j)\}(j \in N),$$

make it

$$T_{temp} = \{p_1(x_1, y_1), p_2(x_2, y_2)..., p_{k-1}(x_{k-1}, y_{k-1}), p_{k+1}(x_{k+1}, y_{k+1})..,p_j(x_j, y_j)\}(k, j \in N) \quad (4.1)$$

calculate the area of T_{temp},named $T_{temparea}$, if $k = j$ then the process is completed.

(3) compare T_{area} and $T_{temparea}$, if T_{area} less than $T_{temparea}$,then replaced T into T_{temp} as input of step (1) and return to perform the step (1) ,step (2) ; if not go to step(2).

Algorithm schematic diagram as follows:

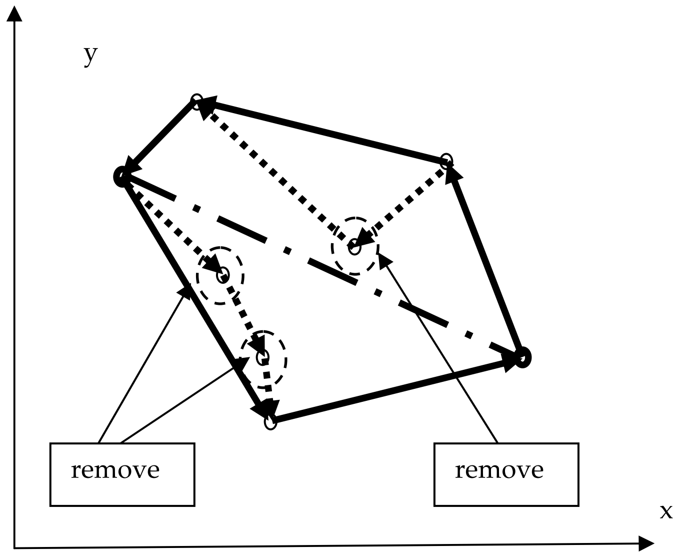

Figure 4.1. Border smoothing algorithm

Obviously, removing the removed points in figure 4.1 can surplus T's area.

5 EXPERIMENTAL RESULTS AND CONCLUSIONS

Experiment used AIS data collected from Wusongkou (Shanghai), Waigaoqiao (shanghai) which raw AIS data show as figure 5.1, successfully identified areas of collision risk as figure 5.2 shows.(Ships' information in figure 5.2 correspond to raw AIS data in figure 5.1)

mmsi	lon	lat	course	length	width	speed
413762146	7291.2842960328207	1883.7410641210	245.599999999999	64.000000000000	12.000000000000	2.700000000000000
413952225	7291.3033975615098	1883.765434384778	223.099999999999	45.000000000000	9.0000000000000	3.799999999999999
413800364	7291.2871121086509	1883.701121079184	256.300000000000	100.000000000000	16.000000000000	1.500000000000000
412375570	7291.3559063324919	1883.61947180651	80.500000000000	32.000000000000	6.0000000000000	22.800000000000
413773727	7291.4761939892596	1883.7792878031	190.300000000000	80.000000000000	14.000000000000	3.299999999999999
413767959	7291.6861835481440	1883.835718093026	68.0999999999999	73.000000000000	13.000000000000	6.799999999999999
413812575	7291.7243932365918	1883.874429811929	198.000000000000	20.000000000000	10.000000000000	3.000000000000000
413764858	7291.7021506332366	1883.754338702387	60.100000000000	45.000000000000	8.0000000000000	6.700000000000000
413766217	7291.7485974425171	1883.788457441704	242.900000000000	95.000000000000	15.000000000000	1.899999999999999
600000924	7291.8715958477487	1883.786218551814	55.500000000000	20.000000000000	10.000000000000	8.099999999999999

Figure 5.1. AIS data

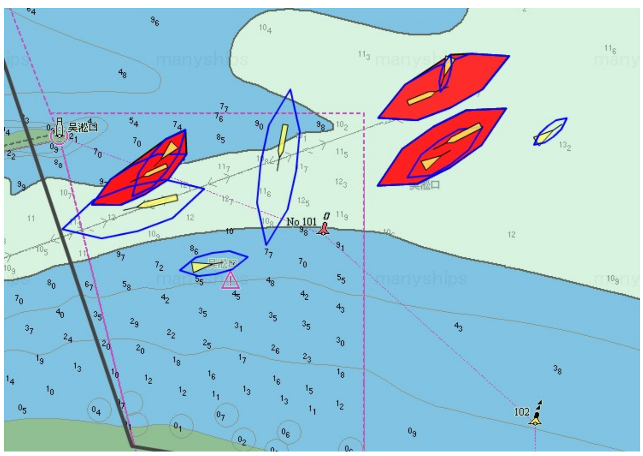

Figure 5.2. Experiment result

The results show that the algorithm put forward in this paper is able to identify and render areas of collision risk in restricted waters.

REFERENCES

[1] Xuwu Su, Guangyou Yang, Guozhu Zhou. Comparison of fuzzy mathematics in pattern recognition applications. Journal of Hubei University, 2005

[2] Xide Cheng, Zhuyuan Liu. Using improved BP neural network caculate open water ship collision risk [C]. Naval Architecture and Ocean Engineering Research Album.2001,143:91-93.

[3] Xiaolin Zhu, Hanzhen Xu. A ship collision risk model based on neural network [J]. Journal of Huazhong Science University, 2005, 28(10).

[4] Lin He, Linda Wu, Yizhao Cai. Summary of clustering algorithms in data mining. Application and Research of Computers, 2007(10).

[5] Jian Li, Wei Yu, Baoping Yan. Memory effect in DBSCAN algorithm. 2009 4th International Conference on Computer Science and Education(ICCSE 2009), 2009(31-36).

[6] E.M.Goodwin, A statistical study of ship domain, The Journal of Navigation, 1975, vol.28, no.3 : 328-344

[7] E.M.Goodwin, Determination of ship Domain Size． Proceedings of International Conference on Mathematical Aspect of Marine Traffic． London： Academic Press，1977：103-127

[8] E.M.Goodwin, J.f.Kemp, A survey of marine traffic in the southern north sea, The Journal of Navigation, 1977, vol.30, no.4: 378

[9] P.A.Davis, M.J.Dove, C.T.Stockel, A computer simulation of marine traffic using domains and areas, The Journal of Navigation, 1980, vol.33, no.2: 215-222

[10] Tao Liu, Qinyou Hu, Chun Yang. cluster analysis and recognition in heavy water traffic area [J]. China Navigation. 2010.33(04)