Pirates Attack Space Division Based on Spatial Clustering

Maojin Sun, Jing Lu, Tianhang Gao, Lequn Zhu

School of Traffic and Transportation, Dalian Maritime University, Dalian 116026, China

Abstract
Because the pirates gathering areas, merchant shipping routes and pirates style can influence the pirates attack, the Pirate Event Special Division Algorithm based on the DBSCAN is constructed to divide the space of the pirates attack positions. A case study in East Africa is calculated and the results are analyzed, and 9 clusters are got. The analysis proves that the Pirate Event Special Division Algorithm conforms to the reality. It indicates that the method is practical and scientific. pirates attack has always been an important factor threatening international maritime transportation. Nowadays, there are numerous areas of international waters affected by pirates attack, mainly including the Gulf of Aden, Indian Ocean, the west African coast, Strait of Malacca, South China Sea and some other sea areas. Starting from the report on pirates attack geographical location, the paper established pirates attack density-based spatial clustering space division algorithm and explored the features of pirates attack events of different areas. The algorithm not only considered the geographical coordinates of the historical pirates attack events, but also the number of person being killed, missing or injured in pirates attack which can reflect the factors of pirate behavior styles.

Key words: Pirates Attack, Space Division, Density-Based Spatial Clustering, Clustering Analysis

1. INTRODUCTION
Pirates Attack, Space Division, Density-Based Spatial Clustering, Clustering Analysis

With the acceleration of the globalization of world economy, marine transportation has become the most important international transport mode. And how to ensure the safety of marine transportation has become one of the issues mostly concerned by national governments, enterprises and ship staff. Among them, pirates attack has always been an important factor threatening international maritime transportation. Nowadays, there are numerous areas of international waters affected by pirates attack, mainly including the Gulf of Aden, Indian Ocean, the west African coast, Strait of Malacca, South China Sea and some other sea areas. The analysis proves that the Pirates Attack Event Special Division Algorithm conforms to the reality. It indicates that the method is practical and scientific. pirates attack has always been an important factor threatening international maritime transportation. Nowadays, there are numerous areas of international waters affected by pirates attack, mainly including the Gulf of Aden, Indian Ocean, the west African coast, Strait of Malacca, South China Sea and some other sea areas. Starting from the report on pirates attack geographical location, the paper established pirates attack density-based spatial clustering space division algorithm and explored the features of pirates attack events of different areas. The algorithm not only considered the geographical coordinates of the historical pirates attack events, but also the number of person being killed, missing or injured in pirates attack which can reflect the factors of pirate behavior styles.

The existing research orientations on pirates attack are relatively diverse. Hewei conducted discussion and analysis on the formation, operation and international countermeasures of Somali pirates specific to pirates attack(Chen, Wang and Li, 2006), Guha studied economic loss caused by marine transportation ships specific to pirates attack(Elio, Shane and Alan, 2014). Elio considered pirates attack as a dynamic behavior (Guha and Ashok, 2011), analyzed pirates attack number and activity area by adopting a dynamic model, and conducted a case study by setting the Gulf of Aden as an example. Gao Tianhang proposed an escort method for new warships to combat piracy to improve navy efficiency by adopting the thought of key area coverage(Gao and Lv, 2016). Zhu Lequan simulated the game process between the navy and pirates by utilizing Stackelberg game and evolutionary game respectively (Ge, John, Pu, Zhao and Li, 2015), selected the escort area of navy and detailed the decision making process between the two(Samiotis, Psarrou and Pazarzis, 2013). Although current researches have conducted extensive studies on pirate attacks, there is no research on pirate attacks space division to analyze the event rules. For space division, other relevant areas already have more practical experience. For example, Geying utilized Ripley’K function based on spatial point pattern, conducted quantitative measure combining with the marginal analysis method of microeconomics on urban agglomeration effect, and performed economic space division(He, Liu and Zhuang, 2006). Considering the global spatial autocorrelation characteristics, Lin Jinyao proposed an improved clustering method based on genetic algorithm, and divided main functional areas by setting Dongguan as an example. Besides, for the research of space division, DBSCAN is one of the most representative density-based spatial clustering algorithms at present which has derived a series of improved algorithms. For example, Liu Zongtian et. al. proposed a parallel DBSCAN algorithm based on data partition to be used for large spatial database thus to improve clustering efficiency. Org Sander, Martin Ester proposed a density-based spatial clustering algorithm (GDBSCAN) for
multidimensional space database (Sander and Martin, 1998). Chen Zhiping and Wanglei proposed a clustering algorithm based on density gradient (Lin and Li, 2014). The algorithm is to seek for original clustering center along the direction with high variation of density through analyzing data sample and peripheral point density changes, merging of sub-culture group according to distribution of border points between classes, and disposal irregular shape data points of clustering. The existing researches have provided research thought and method for molding. The paper put forward a new research thought and method on the basis. The paper set DBSCAN algorithm as a basis, proposed a new density-based clustering algorithm according to the characteristics of pirate attacks, and performed space division for pirates attack to detect the existing special and event rules.

2. SPACE DIVISION MODEL FOR PIRATES ATTACK

Space division of pirates attack is aimed to distinguish all pirates attacks converged on an oceanic map according to certain features and rules. Each group is corresponding to a specific spatial region thus to master the high-risk areas deployment of targeted countermeasures of pirates attack. Among them, category information is not contained in the historical data of pirates attack used for space division. Therefore, in order to realize the above mentioned space division, it is considered to adopt unsupervised clustering method. The essence of clustering is to divide the similarity of the samples not containing label according to certain definition. The similar sample points are classified into one class. The essence of space division is to cluster all similar features as a homogeneous region. The similarity is determined by pirates attack.

A common center clustering method for clustering analysis is K-means. The sample points are divided by setting several points as the center. The samples close to certain center point belong to the same category. Considering the huge marine space and uneven pirate attacks locations distribution, as well as the sensitiveness of the method on initial values of central coordinates. The method may lead to worthless results of clustering. Consequently, the paper adopted DBSCAN (Density-based spatial clustering of applications with noise) in density-based spatial clustering. The main idea of the method is to specify a scanning radius and density threshold. If the number within the effective range exceeds the threshold, they are to be incorporated to focused core points. Among the final clustering results, a clustering is to form a group with adjacent core sets and surrounding reachable points. The density of all clusters is close to or exceeds the density threshold. The other points are peripheral points, not participating in the clustering. The advantage of DBSCAN is that as being affected by noisy points. In case of any error point existing in the record, the distance is far which will be summarized to peripheral point not affecting the core clustering set; in addition, due to the setting of density threshold value, the final clustering results focus on high density point set. Namely relative to high incidence area, for effective revelation of spatial laws, the interference generated from low incidence areas will be ignored.

Specific parameters of the model are as following:

- \( p_i \): Pirates attack clustering
- \( D_i(x, y) \): The coordinate of the \( i \) pirates attack occurrence place
- \( ||D_i, D_j|| \): The earth's surface ground distance of the occurrence place of \( i \) pirates attack and \( j \) pirates attack
- \( \varepsilon \): Scanning radius
- \( \text{minPts} \): Density threshold value
- \( N_i \): Neighbor of \( p_i \)
- \( C \): Core points clustering
- \( R \): Density reachable points set

It can be seen that \( C \) and \( R \) can be divided to several clusters \( C\), namely the final clustering result. Different set members are non-adjacent. One group is as a cluster. Intuitively, a cluster constitutes a high density area in the space. The dot density within all highly density area is obviously higher than other regions thus to realize regional division on density. The specific process of DBSCAN and category expansion algorithm is as following:

| Data: Space coordination of pirates attack, effective radius and density threshold |
| Result: Clustering result |
| Initialization: |
| For data concentration points not visited \( p \) |
| Set the status of \( p \) as visited; |
| Find all neighbors \( \text{Neighbor}\, p \) of \( p \); |
If the number of members $\text{NeighborP}$ is greater than $\min P \text{ ts}$, then
The distributed new category $C$;
Category expansion algorithm is adopted to expand the category;
else
Mark $p$ as noisy point;
end
end

**Figure 1.** DBSCAN density-based spatial clustering

Data: Core point $P$, category $C$, neighbor $\text{NeighborP}$ and other coordinates
Result: Category expansion result
Add $P$ to $C$;
For the points $P'$ in $\text{NeighborP}$ do
If $P'$ is not visited
Set the status of $P'$ as visited;
Find all neighbors $\text{NeighborP'}$ of $P'$;
If the number of members $\text{NeighborP'}$ is greater than $\min P \text{ ts}$, then
Add $\text{NeighborP'}$ to $\text{NeighborP}$;
end
end
If $P'$ does not belong to any category, then
Add $P'$ to $C$;
end
end

**Figure 2.** Category expansion algorithm

However, traditional DBSCAN method has two related issues. The first one is parameters selection. And the second one is too less information can be utilized. The algorithm needs two parameters, namely scanning radius and density threshold, which are set artificially. The method doesn’t contain the selection of the two parameters. On the other aspect, it is far insufficient by utilizing the spatial position and density. Pirate attacks are associated with case casualties and loss etc.. We will analyze below to utilize the information thus to reasonably determine the parameters.

Clustering is aimed to gather similar samples as far as possible. Among ideal clustering results, the corresponding events of a cluster are relatively close to other feature dimensions. DBSCAN only considers the spatial features. Set a total of $l$ other dimensional characteristics. Use $y_{kl}$ to represent all values of the cluster $k$ of the feature $l$. The non-uniformity of the feature $l$ of the cluster $k$ can be expressed with.

$$s_{kl} = \sqrt{\frac{1}{n} \sum_{u=1}^{n} (y_{kl} - \bar{y}_{kl})^2}$$

(1)

$n$ refers to the number of the cluster $k$. $\bar{y}_{kl}$ is the average value of $y_{kl}$. $s_{kl}$ is used to describe the dispersion degree of certain feature. When the size of elements $y_{kl}$ is close to each other, $s_{kl}$ is close to 0, and is not sensitive to the absolute size. However, it is sensitive to the relative size of the elements. The non-uniformity of the general features of cluster can be measured with.

$$L_{U} = \frac{1}{|C|} \sum_{k} \sum_{l} w_{l} \cdot s_{kl}$$

(2)

Namely, perform weighted average for nonuniformity arising from all the features of each cluster thus to get the average value of the clusters. Intuitively, the more similar among cluster members, the smaller general nonuniformity and the better classification effects will be. However, if set the nonuniformity characteristics as optimization target, two extreme cases will appear. One possibility is that in case of no category or small amount of categories without clustering, $L_{U}$ is close to 0; the other possibility is excessive clustering, namely that one or more events constitute a category. $L_{U}$ still will be low. Obviously, the two results are excessively optimized. In order to prevent excessive optimization, unreasonable loss of categories will be introduced.
\[ L_c = \frac{A}{1 + \|C\|} + B\cdot\|C\| \]  

(3)

In case of excessive or undersized categories, unreasonable loss of categories will be increased sharply to prevent parameters changing to unreasonable directions. Therefore, we can merge the two losses. The total clustering loss is.

\[ L = L_u + L_c \]  

(4)

Select a group of optimal parameters by setting minimum as the target so that the corresponding total loss \( L \) will be the minimum under the group of parameters. PESDA (Pirate Event Spatial Division Algorithm) was proposed based on this idea, namely piracy attack space division algorithm:

| Data: Historic data and sample feature data of pirates attack |
| Result: Space division result |
| Initialization; |
| For parameters \( \varepsilon \) and \( \min \) \( \psi \) do of parameter space |
| Density-based spatial clustering is adopted for clustering; |
| Calculate the total clustering loss \( L \) and keep record; |
| end |
| Find the minimum \( L \) of the record; |
| Return to corresponding parameters and the result of clustering; |

**Figure 3.** Pirates event space division algorithm (PESDA)

3. LEAST SQUARES SUPPORT VECTOR MACHINES

Initially designed by Suykens et al., LSSVM can tackle the problems in the traditional SVM in the field of computational time and simplicity in implementation (Zhu and Li, 2014). The main idea of LSSVM algorithm lies in that it can map the input data which is non-linear to high dimensional feature space (Zhu, Lv and Li, 2015). LSSVM is designed based on the primal dual formulation. Supposing the dataset is represented as \( \{x_i, y_i\}_{i=1}^N \), and then LSSVM aims to estimate the following model.

\[ y(x) = w^T \cdot \zeta(x) + b + e_i \]  

(1)

In eq.1, \( e_i \) represents as the noise, and the conditions \( x \in R^n \) and \( y \in R \) are satisfied. Particularly, \( \zeta(x) \) refers to a mapping function. Afterwards, the optimization problem is described as follows.

\[ J(w, e) = \frac{w^T w}{2} + \frac{\lambda}{2} \sum_{i=1}^{N} e_i^2 \]  

(2)

where the following equality constraints should be satisfied.

\[ y(x) = w^T \cdot \zeta(x_i) + b + e_i, i \in \{1, 2, \cdots, N\} \]  

(3)

Afterwards, the Lagrangian function is defined as follows.

\[ \delta(w, b, e, \alpha) = \frac{w^T w}{2} + \frac{\lambda}{2} \sum_{i=1}^{N} e_i^2 + \sum_{i=1}^{N} \alpha_i \left(w^T \zeta(x_i) + b + e_i - y_i\right) \]  

(4)

In Eq.4, \( \alpha_i \) represents denotes the Lagrange multipliers that can either be positive or negative state under the equality constraints in the LSSVM classifier. Next, differentiating the above equation with different parameters, the conditions for optimality of the oil productivity forecasting can be got via setting all the derivatives to 0 as follows.

\[ \frac{\partial \delta}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{N} \alpha_i \cdot \zeta(x_i) \]  

(5)

\[ \frac{\partial \delta}{\partial b} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i = 0 \]  

(6)
\[
\frac{\partial \delta}{\partial e_i} = 0 \rightarrow \alpha_i = \lambda_i \cdot e_i, \quad i \in \{1, 2, \cdots, N\} \tag{7}
\]

\[
\frac{\partial \delta}{\partial \alpha_i} = 0 \rightarrow y_i = w^T \cdot \zeta(x_i) + b + e_i \tag{8}
\]

From the above formal description, it can be seen that LSSVM is a modified form of SVM and it is a more simple technology than the standard SVM. Furthermore, LSSVM can solve linear or nonlinear multivariable calibration problems quickly.

4 MODELING AND SIMULATION

In order to better indicate that pirates event space division model based on density clustering can be applied in practice better, the paper selected the pirates attack in east African regional waters according to GISIS database as the research object. The sea areas mainly include Arabian sea water and certain red sea. The influence factors associated with the model include: geographical coordinates of attack location, if anyone was killed, the number of people being killed, if anyone was missing, missing people amount, if anyone was injured, the number of injuries, if any material was stolen, if any material is damaged, if any gunslinging was associated, if any damage was caused, if any hijacking event occurred and if anyone was kidnapped. Specific calculation results are as shown in Fig. 4:

![Pirates attack density-based spatial clustering diagram](image)

**Figure 4.** Pirates attack density-based spatial clustering diagram

**Table 1.** Pirates attack density-based spatial clustering results table

<table>
<thead>
<tr>
<th>Classification No.</th>
<th>If anyone was killed</th>
<th>The number of people being killed</th>
<th>If anyone was missing</th>
<th>Missing people amount</th>
<th>If anyone was injured</th>
<th>The number of injuries</th>
<th>If any material was stolen</th>
<th>If any material is damaged</th>
<th>If any hijacking event occurred</th>
<th>If anyone was kidnapped</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.005</td>
<td>0.005</td>
<td>0.000</td>
<td>0.016</td>
<td>0.030</td>
<td>0.037</td>
<td>0.007</td>
<td>0.350</td>
<td>0.092</td>
<td>0.130</td>
</tr>
<tr>
<td>2</td>
<td>0.034</td>
<td>0.138</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.448</td>
<td>0.138</td>
<td>0.138</td>
<td>0.138</td>
</tr>
<tr>
<td>3</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.059</td>
<td>0.118</td>
<td>0.000</td>
<td>0.412</td>
<td>0.059</td>
<td>0.118</td>
<td>0.059</td>
</tr>
<tr>
<td>4</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.556</td>
<td>0.167</td>
<td>0.111</td>
<td>0.111</td>
</tr>
<tr>
<td>5</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.357</td>
<td>0.000</td>
<td>0.286</td>
<td>0.286</td>
</tr>
<tr>
<td>7</td>
<td>0.111</td>
<td>0.111</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.333</td>
<td>0.000</td>
<td>0.222</td>
<td>0.000</td>
</tr>
<tr>
<td>8</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.083</td>
<td>0.000</td>
<td>0.083</td>
<td>0.083</td>
</tr>
<tr>
<td>9</td>
<td>0.000</td>
<td>0.035</td>
<td>0.035</td>
<td>0.000</td>
<td>0.000</td>
<td>0.175</td>
<td>0.035</td>
<td>0.175</td>
<td>0.053</td>
<td>0.211</td>
</tr>
</tbody>
</table>
It can be observed through calculation that there are 9 types of the selected historic pirates attack events at targeting sea area according to the influencing factors. The remaining events are pirates attack not forming a category (Figure 5). The paper conducted specific analysis specific to the 9 types of products combining with actual features (Figure 6).

Figure 5. 3D Results of Preferences
Figure 6. Density of the All types of Products Combining With Actual Features

Where, the most majority clustering pirates attack events are of type 1. The pirates attack events are distributed around the Gulf of Aden and Bab-el-Mandeb. It is the most serious area with pirates attack in the past few years. First of all, for medium and small sized transport ships tripping back and forth between the Far East and Europe, approaching the Suez Canal will save more costs comparing to bypass the Cape of Good Hope. Consequently, most such ships will choose to pass through the Suez Canal so that the ships have to go pass the Gulf of Aden and Bab-el-Mandeb. So there is large number of ships move in the area every day which provides a number of targets for pirate attack. Secondly, The Somalia Democratic Republic is located at the south of the Gulf of Aden(Figure 7). The political situation of the area is unstable which will lead that numerous personnel will survive for life as a pirate. Beside the narrow water areas, the traffic route of transport ships is relatively closer to the coastline. The pirate can attack without the advanced equipment. To sum up, the transport ship base is high. The pirate rampant and lower hijack difficulty will jointly lead to pirates attack clustering gathering. Reviewing from clustering results, the pirates attack of the area will lead to low probability casualties, damage and loss. However, high probability armed attacks, hijacked ship and kidnap hostages will occur. It is visible that the pirates in the region are not aimed to threaten the crew life and robbing, but armed ships hijacking and kidnapped for ransom hostage, which is consistent with the actual situation of Somali pirates style (Figure 8).

Figure 7. Actual Situation on Density of the All types of Products Combining With Actual Features
Secondly, type 2 and 3 shall be analyzed simultaneously. The above two kinds of pirate attacks are located at the entrance to the Gulf of Aden on geographical position. In comparison, type 2 is closer to the Gulf of Aden. And type 3 is close to off-lying sea. The main reason causing the phenomenon is that in case of clustering of the first type of pirates attacks, countries established a combined fleet for escort which has stoked the pirates around the Gulf of Aden. In order to avoid the navy, the pirates in the area will turn the hijacking object to transport vessels without escort beyond the Gulf of Aden. While reviewing from the clustering results, the pirates of area 2 and 3 are from the same area of that of type 1. Therefore, their behaviors are similar, presenting hijacking, and kidnapped for ransom hostage as the main objects(Figure 9-10). Relatively speaking, the pirates of type 2 areas are crueler. The kill probability and number are obviously higher than those of other areas. The pirates of type 3 areas will lead to higher personnel injury probability. The damage, hijacking probability and kidnapping probability are slightly lower than the first two types(Figure 11).
And then, the section analyzed type 4, 6 and 8 simultaneously. Geographically, no special geographical features appeared on the three types of pirates attacks which are located at off-lying sea far away from the coastline. Certain similarities exist from clustering results. First of all, the first two types of the three pirates attack are free of personnel casualties and property loss. Although there is any property loss of the third type, the probability is not high. Secondly, the three will be subject to higher probability of kidnapped hostages with hijacking ships and armed attack. The difference is that although the armed attack probability of type 4 pirates attack is high, the successful rate of hijacking and kidnap hostages is slightly lower. The situation of type 6 pirates attack is on the contrary, with lower armed attack probability but higher successful rate of hijacking and kidnap hostages. And type 8 pirates attack shows lower armed attack probability as well as successful rate of hijacking and kidnap hostages.

The value of each pirates attack clustering result is 0 without relevant information provided. There are mainly two reasons leading to such situation. The first one is the lack or improper reports on pirates attack which will lead to no result available; the second one is the pirate style of the area will not cause damage to transport ships. Comparatively, the possibility of the former case is higher.

Type 7 is the most particular type of the above mentioned 9 types of pirates attack events. Geographically, the type of pirates attack events are distributed along the coast of Somalia, which is comparatively closer to the coastal line. While reviewing from the clustering results, the type of pirates attack shows higher death, property stealing and armed attack. And no hijacking and ship kidnap hostages appeared. Based on the analysis by combining with the geographical characteristics and clustering results, the type of pirates attack events are mainly aimed to fishing boats and vessels at the southern coastal area of Somalia for the property on board. In addition, the pirates will carry out weapons in most cases. In case of being detected, fighting against or other special circumstances, they will kill the staff and threaten the lives of crew safety. It is one of the most dangerous pirate attacks clustering.

Finally, type 9 pirates attack events are around the coastline of Kenya and Tanzania while analyze the pirates attack location geographically. However, the pirates attack events of the region didn’t show particularly distinctive features. Casualties, property losses, ship hijacking and hostages kidnapping occurred in this area. The probability of staff missing, stolen property, items damaged, armed attack and ship hijacking is comparatively higher, followed by that of hostages kidnapping and damage, and then deaths and injuries.

5 CONCLUSIONS

In order to better study the features of pirates attack events, countermeasure are put forward specifically. Starting from the report on pirates attack geographical location, the paper established pirates attack density-based spatial clustering space division algorithm and explored the features of pirates attack events of different areas. The algorithm not only considered the geographical coordinates of the historical pirates attack events, but also the number of person being killed, missing or injured in pirates attack which can reflect the factors of pirate behavior styles. On such basis, the paper calculated, discussed and analyzed by setting the historic pirates attack events in east African regional waters of the GISIS database. The analysis results show that after eliminating the peripheral points which cannot be clustered, the pirates attack events in east African regional waters can be classified to 9 types. Each type has its unique features. The analysis results also indicate that the features of each type are comparatively in line with the actual situation.
REFERENCES


