

On a Method to Detect the Dynamic Hotspot

Takahiro Saito ^{1,a,*} and Isamu Watanabe ^{1,b}

¹ Artificial Intelligence Laboratory, Fujitsu Laboratories LTD, Kawasaki, Japan
a. saitou.takah-02@jp.fujitsu.com, b. watanabe.isamu@jp.fujitsu.com

*corresponding author

Keywords: Dynamic Hotspot, Collision Risk, Moran Scatter Plot, Multiple Encounter

Abstract: Even if there is a risk of two ships colliding, this risk can usually be averted using avoidance maneuvers. However, if this avoidance maneuver is restricted, for example, because it affects the course of a third ship, then it is difficult to reduce this risk. We call the point where such a situation occurs the dynamic hotspot. In this paper, we propose a method to detect this dynamic hotspot in the near future, from current AIS data. We also discuss the nature of dynamic hotspots calculated from historical AIS data in two respects. First, in respect of spatiotemporal statistics using the Moran scatter plot. The second is more practical; we discuss the relation to the spatial distribution of a Multiple Encounter, which is implied to have a relation to the above situation.

1. INTRODUCTION

In maritime traffic control tasks, it is important to quickly identify vessels which may be on a collision course and intervene appropriately by providing information. However, each ship is also striving to navigate safely. A captain who recognizes that there is a risk of collision on his ship, for example, makes avoidance maneuvers. As a result, if there is no difficulty in making the avoidance maneuver, safe navigation may be maintained by calling for attention. However, an avoidance maneuver of a ship for which a collision risk has occurred may interfere with the safe navigation of another ship. The danger of such a situation has been previously identified, for example, [1] stressed that this situation is often difficult to solve and forcefully strains the captain/pilot. For this reason, control operators are expected to anticipate dangerous situations like these and take preventive measures. Nevertheless, a great number of skills are required to do such control operations. A system to support a predictive response is required.

A dynamic hotspot is a location where there is a high possibility that dangerous situations like the one described above may occur. This paper discusses the method of detecting the dynamic hotspots that may appear in the very near future. In this paper, we first explain this method. Our method includes parameters, however, so the next section describes how to determine the suitable parameters using actual data. In the last section, we report to what extent the dangerous situation presented above can be predicted by our method.

2. METHOD

2.1. Notations

- If there is a possibility of collision between a ship-A and a ship-B, which occurs at time t , we write this probability $p(A,B :t)$. This collision point is denoted by $l(A,B : t)$.
- The grid's divided target area is denoted by $g(i)$, where i is an index of grids. The estimation value of hotspots in a grid $g(i)$ is denoted by $e(i)$ or $e(i :T)$. The $e(i :T)$ is the estimation value calculated by probabilities existing at time T .

2.2. Method to calculate the estimation values

We assume that we can estimate the hotspots based on the probability of a ship which has the possibility of colliding with multiple ships. As a result, the method to calculate the estimation value of hotspot $e(i :T)$ is as follows:

1. Calculation of $p(A,B :t)$
We calculate the probabilities of the ships colliding in the target area at time T , using AIS data at T . However, this calculation is very difficult, so we set up the hypothesis and calculated approximate values based on this hypothesis. This approximation is discussed in the next section.
2. Selection of $p(A,B : t)$
First we set the time window T , and selected probabilities $p(A,B :t)$ such that t is in the interval $(T, T + \Delta T)$. Next we selected the probabilities of ships whose multiple probabilities of colliding with other ships are in the selected probabilities above. The result of this selection is expressed by Ω .
3. Calculation of $e(i : T)$
We calculated the probability of collision by the probabilities whose collision points are in grid $g(i)$, and this probability is defined as an estimation value of the hotspots. As a result, $e(i: T)$ is calculated using (1):

$$e(i; T) = 1 - \Pi(1 - p(A, B : t)) \quad (1)$$

where $p(A, B : T) \in \Omega$, and $l(A, B : t) \in g(i)$

2.3. Approximation of $p(A,B :t)$

Various methods to evaluate the collision risk have been proposed. Of course, the risk indices are strongly related to the probability of collision, so we assume the function of the risk index based on the probability of collision. If the risk index is 0 when the probability of collision is 0, and 1 when the probability of collision is 1, then this function is a monotonically increasing function from $[0,1]$ to $[0,1]$. In the set of functions which have this characteristic, we choose the n -th power function as a translation function. This function has a parameter n , so we will find a suitable n for the experiment using real data.

It is necessary to calculate the probability and $p(A,B :t)$, and estimate the collision point and time. We assume the midpoint of the closest points of approach as a collision point, and the time to the closest point of approach as the time of collision. In this calculation, we assume that both ships hold the current speed and the current course.

3. RISK INDEX

In this paper, we assume the probability of collision can be calculated using the risk index of collision. However, there are many methods to evaluate the collision risk. We adopt the “ensemble”

risk index of the risk indices calculated by the multiple methods. In our method, we should decide the ensemble weights of each risk indices. In order to get these weights, we apply factor analysis. Factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variable called factors. In applying factor analysis, we assumed that each risk index evaluates the mixture of various aspects of risk, and this mixing ratio varies depending on the method. This assumption seems to be appropriate, for example, it can be said that “CJ” method [2] tends to evaluate the degree of difficulty of risk avoidance because this evaluation formula contains explicitly the term related to difficulty of risk avoidance. Under this assumption, factors in factor analysis correspond to each aspect of risk. Using a principal factor analysis algorithm, the factor that has the largest influence overall can be extracted as a primal factor.

The first step in factor analysis is to determine the number of factors. There are several ways to determine this value, but we adopt the method using “scree plot”, which may be most common. In addition, factor analysis has degree of freedom about rotation. In our case, independence between factors cannot be assumed. So we adopted “promax rotation”, which may be most common in the methods of oblique rotation. Also there is a point to note when extracting the primal factor using factor analysis. If the risk index of a specific method is hardly affected by the primal factor, the loading of primal factor may be calculated as a negative value. In this case, we can remove this risk index from the candidates of the ensemble.

According to the above consideration, we determined that our ensemble risk index is calculated using the following steps:

1. Using the AIS data including collisions, we calculate the risk indices by the risk evaluation method of each, and modify the values to $[0,1]$, where 1 means the collision and 0 means that the probability of collision seems to be 0. Furthermore, there are methods which do not output the scalar value as a result of evaluation. For example, “HRL” method [5] outputs the categorical value like a “Moderate risk”. Even in such a method, we tried to output the value in $[0,1]$, however these quantification methods may be improved in the future. In light of this, the contents of these methods are omitted in this paper.
2. We apply factor analysis to the risk indices and calculate the loadings of the primal factor. If there is a method whose loading is negative, then this method is removed from the candidates of the ensemble, and we once again apply a factor analysis.
3. A loading of each risk index divided by the sum of loadings is regarded as the ensemble weight of each risk index.

Table I shows the labels of methods incorporated into the ensemble risk index and weights calculated by the above processing. Of course, this result depends on data and rotation method, does not indicate superiority or inferiority of each method. This result only means that the weighed sum of risk indices using weights in this table can evaluate the primal factor that most affect each risk index overall. As a matter of fact, we think that the method explained in this section should be improved. However, it must be beneficial to evaluate the hotspots detected using current method.

Table 1: Risk evaluation methods and weights of ensemble

Label and Citation	Weights
CJ [2]	0.200
HRL [3]	0.262
SYSROC [4]	0.258
OZT [5]	0.113
DCPA	0.100
SJS [6]	0.066

4. EXPERIMENT TO FIND A SUITABLE PARAMETER

4.1. Moran Scatter Plot and Evaluation

Assume that the value of the normalized variable X is assigned to each grid in the space divided grids. This value assigned to grid $g(i)$ is denoted by $x(i)$.

The Moran scatter plot shows the horizontal axis in the normalized variable X , and in the normalized ordinate spatial delay of that variable. The normalized ordinate spatial delay of grid $g(i)$ is denoted by $y(i)$, so $y(i)$ is defined by (2):

$$y(i) = Nx(i)\sum w_{ij}x(j)/(S_0\sum x(k)^2) \quad (2)$$

where N is a number of grids, w_{ij} is the weight between grids $g(i)$ and $g(j)$, and S_0 is the sum of w_{ij} .

This w_{ij} represents the spatial proximity between the grids $g(i)$ and $g(j)$. In this experiment, w_{ij} is defined by (3):

$$w_{ij} = \begin{cases} 0, & d_{ij} = 0 \\ 1, & 0 < d_{ij} \leq 500 \\ 1 - (d_{ij} - 500)/1500, & 500 < d_{ij} \leq 2000 \\ 0, & d_{ij} > 2000 \end{cases} \quad (3)$$

where d_{ij} is the distance([m]) between the centers of grids $g(i)$ and $g(j)$.

In this scatter plot, one grid is represented as one point, and each grid is classified as follows by $x(i)$ and $y(i)$.

- H-H, if $x(i) > 0$ and $y(i) > 0$

The grid in this class has the feature that the value is high like the neighboring grids. From the point of view of geospatial statistics, this grid may be called the “hotspot”.

- L-L, if $x(i) < 0$ and $y(i) < 0$

The grid in this class has the feature that the value is low like the neighboring grids. This grid may be called the “cool spot”, just like the “hotspot” above.

- H-L, if $x(i) > 0$ and $y(i) < 0$

The grid in this class has the feature that the value is specifically high, despite the low value of the neighboring grids.

- L-H, if $x(i) < 0$ and $y(i) > 0$

The grid in this class has the feature that the value is specifically low, despite the high value of the neighboring grids.

If $x(i)$ is the probability of collision in grid $g(i)$, we cannot assume that there are many grids in classes “H-L” or “LH”, so we can regard the ratio of grids in class “H-L” or “L-H” as one of the evaluation values.

On the other hand, it is also desirable for the ratio of the grids in class “HH” to be lower. This is easy to understand assuming the extreme case as follows: if the values of all of the grids are unique except one grid whose value is lower, the ratio of the grids in class “H-H” is very high and the ratio of the grids in classes “H-L” or “L-H” is very low.

Based on the above discussion, we adopted the product of the above two ratios as an evaluation value.

4.2. Data

In this experiment, we use AIS data including collisions for discussion. This data is as follows:

- Area
The range of latitude is in [1.15, 1.38], and the range of longitude is in [103.48, 104.20]. This area includes the Singapore Strait and port area
- Interval
3 hours from 2014-01-29 08:00:00 to 11:00:00(UTC).
The collision recorded in this data occurred about 10:36(UTC). We denote the grid including this collision point as $g(c)$. We created data at intervals of 10 seconds by linear interpolation from raw AIS data
- Type of Ships
“Tanker”, “Cargo” and “Passenger”. Passenger ships shorter than 60[m] were excluded from the data. Ships whose speed is slower than 3.0[kn] were also excluded from the target ships whose collision risk was estimated at this time
In this data, there were 4653 pairs of ships estimated to have had a collision risk between them.

4.3. Parameters

Our hotspot calculation method contains several parameters. In fact, there is an optimal combination of the parameters. On the other hand, these parameters should be determined considering practicality, so we determined the values of the parameters as follows:

- Grid size
The width (longitude) and height (latitude) of the grid are both 0.0025 degrees. In this data, both are approximately 280[m].
- Time window
We set the time window of calculation to 15 minutes, so the detected hotspots are considered as the probability of collision within 15 minutes
- Power of translation function
This parameter has nothing to do with practicality, as a result of which we will find the value which produces the best evaluation value.

4.4. Procedure

We calculate the evaluation values using AIS data at 5 minute intervals and the parameter n ($n = 1$ to 9), and find the n which makes the mean of these minimum evaluation values.

4.5. Results

Fig. 1 shows the mean of the evaluation values calculated with parameter n . We can find that this is best in the case of $n = 7$. In this case, the total number of “H-H” grids is 1536 for all 869093 (= 0.177%).

Table II shows the rank of the $e(c:T_c)$ which is the evaluation value of the grid of the collision which occurred and T_c is the time just before this collision occurred (“10:35:00(UTC)”).

Based on the results from this table, it may be said that $n \geq 4$ is preferable. From the above two results, we adopted $n=7$. Using this parameter, the average number of grids detected as hotspots at each time is about 42.

Fig. 2 shows an example of hotspots at the time when most hotspots are detected. In this figure, there are 65 gray points which mean hotspots. This figure shows the following:

- There are hotspots not only in the port area but also in the Singapore Strait.
- The hotspots are agglomerated. In practice, rather than paying attention to each point of the hotspot, it is possible to pay attention to the area where many hotspots are clustered.

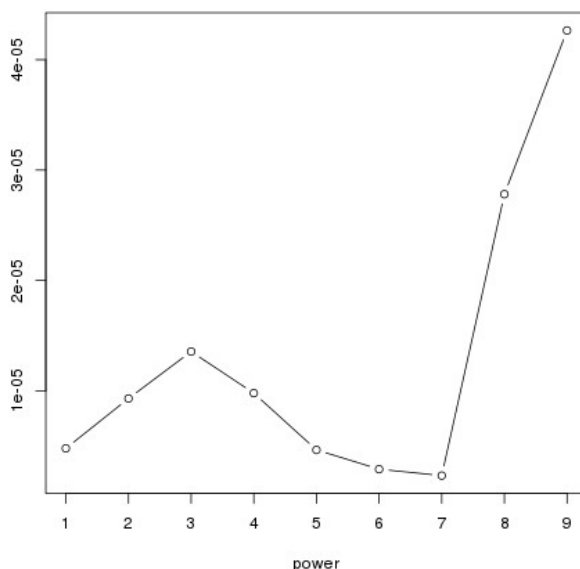


Figure 1: Result of Evaluation

Table 2: Rank of The Collision Grid

<i>n</i>	1	2	3	4	5	6	7	8	9
rank	25	10	3	2	2	2	2	2	2



Figure 2: Example of hotspots

5. Experiment of Evaluation

5.1. Multiple Encounter

Fujii indicated the presence of the effective domain around a ship into which other ships avoid entering. This domain is approximately elliptic with a long radius of $8L$ and short radius of $3.2L$ under ordinary navigation conditions. This domain is called the “bumper”, and a multiple encounter is defined as simultaneously overlapping three or more bumpers. He reported that a multiple encounter is a dangerous situation, and it is closely related with the probability that the avoidance maneuver interferes with the navigation of the third ship [7].

Of course, this probability is also closely related to the hotspot of interest in this paper, so we attempt to validate the relation between the multiple encounter and hotspots.

5.2. Procedure

1. Extraction of multiple encounters

Using AIS data, we extract the ships which are in a multiple encounter. In this step, we adopt the rectangle model which Fig. 3 shows as a bumper (Fujii discussed that using this rectangle model instead of the original model did not pose problems).

2. Extraction of the grids

We analyze the trajectories of ships in a multiple encounter, and extract the grids and transit times passed by those ships.

3. Evaluation of hotspots

If there is a grid which satisfies the following condition, this multiple encounter can be regarded as having been predicted by the hotspots (See Fig. 4).

- It is adjacent within 2 grids of the extracted grid by the trajectory of the ship.
- It belongs to class “H-H”, evaluated within 15 minutes immediately before the transit time.

We apply this evaluation for each ship which is in a multiple encounter, and calculate the number of ships whose multiple encounters are predicted

5.3. Results

1. Extraction of multiple encounters

From this data, 179 multiple encounters are extracted.

2. Extraction of grids

2996 grids where there was at least one ship in a multiple encounter are extracted.

3. Evaluation of hotspots

162 multiple encounters are regarded to be predicted, so 90% of multiple encounters occurred near the hotspots pointed out beforehand.

Regarding the “prediction time” defined in Fig. 4, Fig. 5 shows the maximum amount of prediction time and the cumulative frequency of predicted multiple encounters. For example, point (600, 59) means that 59 multiple encounters are predicted within 10 minutes of occurrence. The jump at the right side of the graph means that the predicted multiple encounters are mainly predicted 15 minutes before.

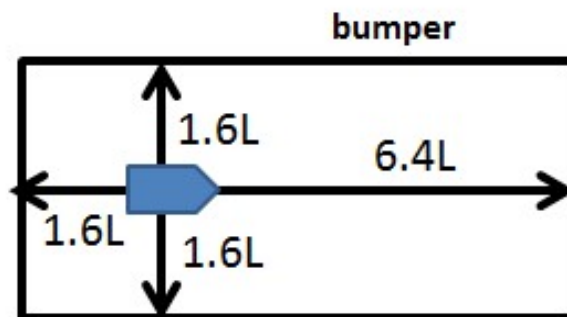


Figure 3: Rectangle bumper

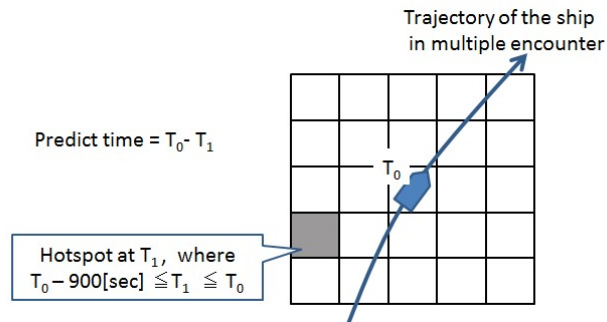


Figure 4: Predicted multiple encounter

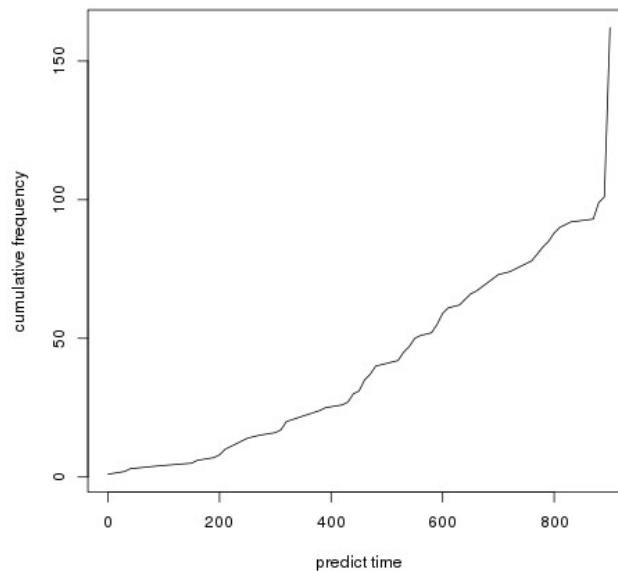


Figure 5: Result of predict time

6. CONCLUSION

This paper proposes a method to detect the dynamic hotspot. In experiments using real AIS data, results were obtained that 90% of multiple encounters occurring near the hotspot were pointed out beforehand by our method.

In future work, we would like to develop a method of detecting a dangerous situation earlier by combining trajectory prediction. In order to implement this, it is believed that it is necessary to process taking into consideration prediction errors. For example, it is conceivable to incorporate the trajectory prediction result as a probability distribution. However, it seems that there are still many problems about how to properly combine the spatiotemporal probability distribution of predicted trajectories and the collision probability.

References

- [1] A. Jensen, "Safety at sea Problem", *Accid. Anal and Prev.*, Vol.1, pp. 1-7, 1967
- [2] H. Kobayashi and M. Endoh, "Analysis of collision avoiding action of ship - from viewpoint of man-machine system Analysis -", *J. Navigation*, 56, pp. 101-109, 1976
- [3] D. Achim Kumer and Chin, Hoong Chor, "Hierarchical modeling of perceived collision risks in port fairway", *Transportation Research Record*, 2010, pp. 68-75, 2010

- [4] T. Kubota and N. Kouguchi, "Navigator's decision-making and action model ii: synthetically subjective risk of collision", *J. Navigation*, 122, pp. 21-26, 2010
- [5] H. Imazu, J. Fukuto and M. Numano, "Obstacle Zone by Target and its Expression", *J. Navigation*, 107, pp. 191-197, 2002
- [6] K. Inoue, "Evaluation Method of Ship-handling Difficulty for Navigation in Restricted and Congested Waterways", *J. Navigation*, 53, pp. 167-180, 2000
- [7] Y. Fujii, "A consideration on multiple encounter", *J. Navigation*, 65, pp. 179-184, 1981