

# Anomaly Detection in Vessel Tracking Using Support Vector Machines (SVMs)

Dini Oktarina Dwi Handayani<sup>1</sup>, Wahyu Sediono<sup>2</sup>, Asadullah Shah<sup>1</sup>

<sup>1</sup>Department of Computer Science, Information and Communication Technology

<sup>2</sup>Department of Mechatronics, Engineering

International Islamic University of Malaysia

Gombak, Selangor Darul Ehsan, Malaysia

e-mail: dini.handayani@gmail.com, {wsediono, asadullah}@iium.edu.my

**Abstract**—The paper is devoted to supervise method approach to identify the vessel anomaly behavior in waterways using the Automated Identification System (AIS) vessel reporting data. In this work, we describe the use of SVMs to detect the vessel anomaly behavior. The SVMs is a supervised method that needs some pre knowledge to extract the maritime movement patterns of AIS raw data into information. This is the basis to remodel information into a meaningful and valuable form. The result of this work shows that the SVMs technique is applicable to be used for the identification of vessel anomaly behavior. It is proved that the best accuracy result is obtained from dividing raw data into 70% for training and 30% for testing stages.

**Keywords:** Maritime Surveillance, AIS, SVMs, Interpolation, Anomaly Detection

## I. INTRODUCTION

The number of problems that threaten vessels at sea and the security of coastal countries increase daily in the form of collision, illegal fishing, smuggling, pollution, and piracy on cruise liners. For those problems, some fall prey to human error and some are victims of nature's wrath. One may assume that mishaps might be the downhill of technological advancement.

With recent advances in surveillance technology and the imminent need for better protection of homeland security, automated solutions for detecting anomaly behavior in moving objects, such as vessels, planes and vehicles, have become an important issue in intelligence and security informatics.

This is a very challenging since the pattern of movement that links with the environment can become very complex. With the technological advancement, the process of analyzing raw data becomes more effective in the decision support process.

In the Straits of Malacca and Singapore, more than 70,000 vessels each year and over 150 vessels a day transit at this strategically important international waterway. Thus, the International Chamber of Shipping (ICS) believes in placing safety and security as priorities for nations across the globe [1].

There are many typical sensors and device that are used for maritime surveillance and port, such as coastal radars,

very high resolution (VHR) optical monitoring, satellite synthetic aperture radar (SAR) surveillance and AIS reporting system. It can be installed on fixed ground locations or mounted on border patrol vessel and aircraft [2].

For the purpose of our research, we utilize the AIS raw data which consist of the vessel id, position, course, speed, etc. The AIS is a self reporting maritime system, which has been recently made compulsory by the international convention for the safety of life and sea (SOLAS) for vessel of above 300 gross tons and most commercial vessels such as cargo, passenger vessel, tankers, etc. [3].

In this paper, the supervised method approach is proposed to identify the vessel anomaly behavior. The proposed algorithms for identifying the anomaly behavior assume that relevant motion patterns have already been extracted from the historic (training) data during the data processing stage.

The anomaly behavior is defined as an unusual behaviour of vessel, such as vessels with random movement in the middle of water, vessel with unexpected stops, vessels with close track in the middle of water, vessels with very short tracks, vessel tracks with many interaction, vessel tracks with many loops, travel over land, deviations from standard routes, speeding, traffic direction violation etc. [4][5]. The main goal of this research is to interpret and analyze the AIS raw data for detection and identification of the vessel anomaly behavior by developing a model that represents the vessel behavior.

The paper is organized as follows. Sec. II describes the SVMs approaches for detection and identification of vessel anomaly behavior. Sec. III presents the simulation process. Sec. IV describes the data that we used for the simulation process. Sec. V displays the experimental results. Sec. VI describes the discussion and Sec. VII denotes the conclusion.

## II. SVMs APPROACHES TO ANOMALY DETECTION

We explore the use of SVMs for analyzing the AIS raw data for detection and identification of vessel anomaly behavior. While the SVMs have been widely applied for maritime surveillance and identification of vessel anomaly behavior [6][7], to date, there has not been any extensive study with regards to the applications of SVMs to maritime anomaly detection.

The SVMs is a supervised method that needs some pre knowledge before classification. The main idea of SVMs is mapping the training data which consist of nonlinear data into higher dimensional feature space through the kernel function [8]. The SVMs method is implemented as a pattern classification technique that measures the similarity between input tracking data and the tracking data stored in the database. The whole anomaly detection system has two important phases, namely, the training phase and testing phase. The training phase is the process of developing models of AIS raw data which will be used as the database for the anomaly detection system. The testing phase is a process of evaluating the performance of the proposed system. In this research we will compare the results of simulation with and without interpolating the AIS raw data.

### III. SIMULATION

Figure 1 shows the first step in the cleaning of the AIS raw data and the assigning of each record to separate tracks based on the Maritime Mobile Service Identity (MMSI). The next step is duplicating the data, in which the first group continues the process without interpolating, and the remaining group continues to the interpolating process. The interpolating process interpolates each row's value into the nearest three minutes interval and eliminates the duplicated data.

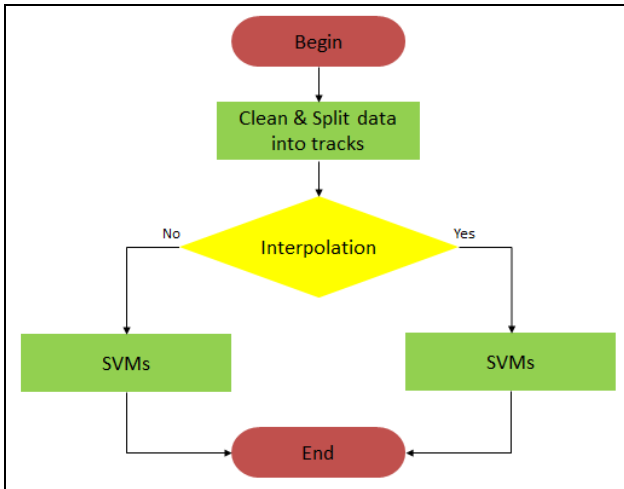


Figure 1. Flowchart of the simulation process

As mentioned before, AIS raw data are used for detection and identification of the vessel behavior. The AIS raw data consist of many cases for every single vessel route, with no explanation of how many journeys it took for some period. Track splitting was performed so that each single track corresponds better to a single route. We split a track record into multiple routes when the vessel is stopped or not transmitting.

In this paper, we use visual analysis method for detection of vessel anomaly behavior. While visual analysis approach generally produces very good results for small

data sets, it typically failed when the data set is too large to be captured by a human analyst [9].

The vessel track data are categorized into normal and anomaly tracks data based on the visual analysis process. Each groups of tracking data will be randomized and divided into two small groups for training and testing processes. Figure 2 shows the detail of this experimental stages.

#### Algorithm 1 anomaly detection process

1. *Categorize* data, separate and label into two groups, *normal* tracks data and *anomaly* tracks data
2. *Randomize* normal tracks data
3. *Randomize* anomaly tracks data
4. *Divide* normal tracks data into two groups
5. *Divide* anomaly tracks data into two groups
6. *Combine* first group of normal tracks data and first group anomaly tracks data, and randomize, for *training process*.
7. *SVMs classification*, get the accuracy of training data.
8. *Combine* second group normal tracks data and second group anomaly tracks data, and randomize, for *testing process*.
9. *SVMs classification*, get the accuracy of testing data.

Figure 2. Algorithm of anomaly detection process

### IV. THE DATA

We used the AIS raw data of Port Klang from July to September 2013. They consist of 9,845 rows of data, including the vessel's MMSI, status, speed, longitude, latitude, course, heading and timestamp (see table 1). All information is obtained from the public website marinetraffic.com.

TABLE I. AN EXAMPLE OF SEVEN DIFFERENT VESSEL DATA FROM THE ORIGINAL AIS DATA, WITH INFORMATION REMOVED TO PRESERVE ANONYMITY.

MMSI	Status	Speed	LON	LAT	Course	Heading	TIMESTAMP (UTC)
47724****	1	142	102.2369	1.900327	115	117	7/10/2013 9:20:00 PM
52501****	9	118	99.867	3.012243	352	511	7/31/2013 5:15:00 AM
53301****	0	70	103.6514	1.273675	214	511	8/23/2013 9:22:00 AM
53301****	0	0	101.3507	3.05836	226	0	8/5/2013 5:03:00 PM
53301****	1	1	103.7945	1.241248	47	267	7/12/2013 8:26:00 AM
53313****	5	1	101.3565	3.042735	338	357	7/18/2013 6:32:00 AM
53386****	1	85	101.7891	2.396002	142	303	7/31/2013 9:44:00 PM

The above data produced 367 tracks across 7 unique MMSI averaging 1,400 rows each. The vessel behavior models have been limited to 7 variables, such as status, speed, location, course, heading and timestamp (including hour of day and time), coupled with the type of vessel, as shown in figure 3. Figure 3 shows the Straits of Mallaca digital map with the longitude from 101.25<sup>0</sup> to 102.36<sup>0</sup>, and latitude from 2<sup>0</sup> to 3<sup>0</sup>.

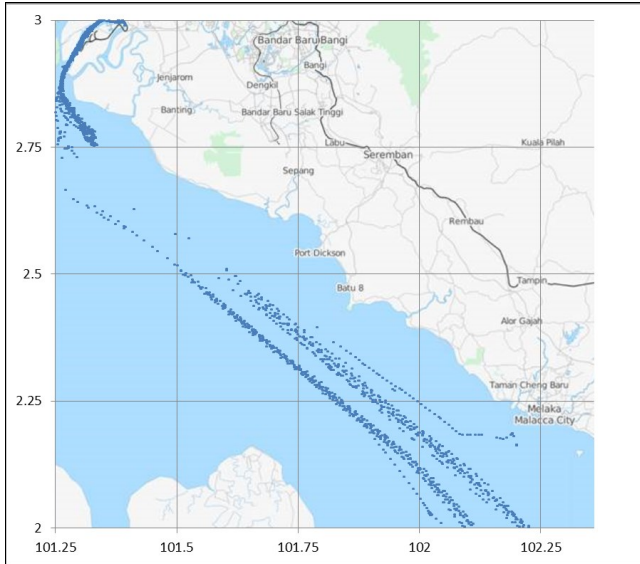


Figure 3. AIS trajectories in Port Klang

Figure 3 denotes the vessel history data without the interpolation process. As shown, some data are missing. As comparison, figure 4 shows the vessel history data with the interpolation process. It generates new data to the nearest three minutes interval.

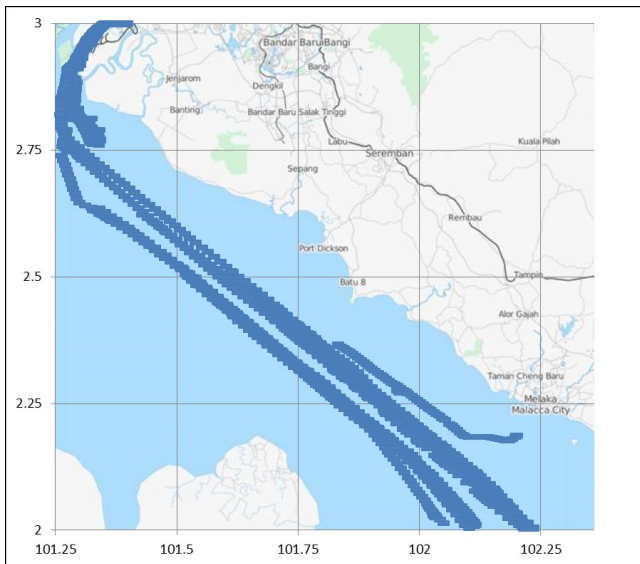


Figure 4. AIS trajectories in Port Klang with interpolation

## V. RESULTS

Previous study shows that many cases can be called as anomaly behavior, such as vessels with random movement in the middle of water, vessel with unexpected stops, vessels with a close track in the middle of water, vessels with very short tracks, vessel tracks with many interaction, vessel

tracks with many loops, travel over land, deviations from standard routes, speeding, traffic direction violation etc. [4][5].

Here, we present examples of the vessel anomaly behavior in two different scenarios. The first scenario is the vessel anomaly behavior with the U-turn route, while the second anomaly behavior is the vessel with the unusual stop points.

### A. Scenario of anomaly behavior with the U-turn route

Figure 5 illustrates the scenario of vessel anomaly behavior in the case of a U-turn route. The digital map shows the Straits of Mallaca with the longitude from  $101.25^0$  to  $101.55^0$ , and latitude from  $2.75^0$  to  $3^0$ .

As shown in the figure 5, the vessel starts and moved away from Port Klang. The vessel initially moves within the normal route. However, the vessel makes U-turn which makes the pattern become anomalous. The normal route is shown with blue stripes. On the other hand, the anomaly route shown with a black circle.

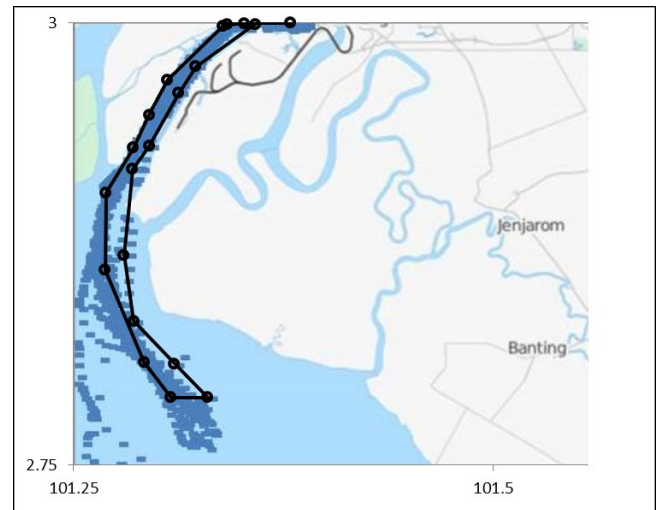


Figure 5. Scenario of U-turn route

### B. Scenario of anomaly behavior with the unusual stop route

The following figure (figure 6) displays another vessel anomaly behavior. Here, the vessel route denotes the anomaly in which vessel stops at variant unusual points. This may happen when the vessel run out of fuel.

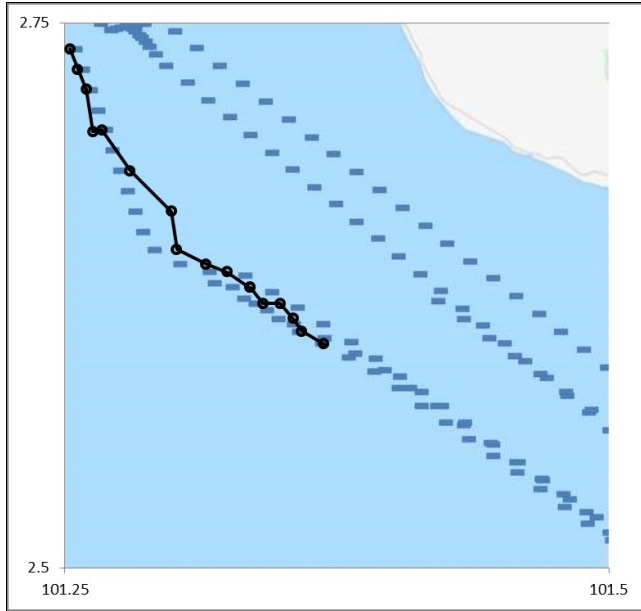


Figure 6. Scenario of anomaly behavior with the unusual stop route

## VI. DISCUSSION

The aim of our research is to propose a supervised method approach to identify the vessel anomaly behavior. In relation to that, we choose the SVMs as a technique to classify the data and identify the normal and anomaly behavior of the vessel. The data consist of Port Klang vessel tracking data for a three month period. In the experimental stage, we use holdout method. With the holdout method, the data is separated into two groups. The first group is for training purposes and the second groups is for testing. It will give significantly different results, depending on the division for training data and testing data [10].

We choose three types of experiments:

- (i). 60% of the data is used for training, and 40% of data for testing.
- (ii). 70% of the data is used for training, and 30% of data for testing.
- (iii). 80% of the data is used for training, and 20% of data for testing.

TABLE II. ACCURACY OF RAW DATA CLASSIFICATION.

	60-40	70-30	80-20
Training	92.03%	90.47%	92.16%
Testing	90.55%	91.63%	89.50%

The result is presented in table 2. As shown, the best accuracy result appears from dividing the raw data into 70% for training and 30% for testing.

As a comparison, we do the similar experiment using the interpolated data. The result is shown in table 3.

TABLE III. ACCURACY OF INTERPOLATION DATA CLASSIFICATION.

	60-40	70-30	80-20
Training	99.69%	99.63%	99.77%
Testing	99.53%	99.81%	99.06%

As shown in table 3, the result from interpolation data also shows that, the best accuracy result is obtained from dividing the raw data into 70% for training and 30% for testing.

## VII. CONCLUSION

In this paper, we have investigated the vessel anomaly behavior with visual analytic approach and vessel tracking data.

From the experimental stage, we found that the SVMs method is applicable to be used for vessel anomaly detection. It is shown that the classification using interpolation data give better accuracies than the classification without interpolation data.

Plan to extend the variable related to the ship itself (including type, dimension and weight) and the weather (such as temperature, wind speed etc.) are the suggested future works.

## ACKNOWLEDGMENT

The author would like to thank the marinetrffic.com, especially Miluše Tichavska, for providing access to their data and generously answering many questions about the AIS raw data.

## REFERENCES

- [1] MaritimeSecurityAsia, "ICS calls for improved navigation safety in Malacca, Singapore Straits," 2011. [Online]. Available: <http://maritimesecurity.asia/free-2/strait-of-malacca-free-2/ics-calls-for-improved-navigation-safety-in-malacca-singapore-straits/>. [Accessed: 12-Mar-2013].
- [2] SSR Engineering, "Maritime Security 2013," 2013. [Online]. Available: <http://www.maritimesecurity2013.com/west/index.php/exhibitors/42-ssr>. [Accessed: 20-Nov-2013].
- [3] B. Ristic and B. La Scala, "Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction," *11th International Conference on Information Fusion*, 2008.
- [4] S. Mascaro, A. E. Nicholso, and K. B. Korb, "Anomaly detection in vessel tracks using Bayesian networks," *International Journal of Approximate Reasoning*, Apr. 2013.
- [5] R. Laxhammar, "Anomaly detection for sea surveillance," pp. 55–62, 2008.
- [6] S. Mirghasemi, "A Parallel Approach Based on Combining SVM , Edge and Corner Detection Methods for Target Detection," pp. 1–6, 2012.
- [7] X. Li, J. Han, and S. Kim, "Motion-alert: automatic anomaly detection in massive moving objects," *Intelligence and Security Informatics*, 2006.

- [8] C. M. Bishop, *Pattern Recognition and Machine Learning*, 9th ed. Springer, 2009.
- [9] D. Keim, F. Mansmann, and J. Thomas, “Visual analytics: how much visualization and how much analytics?,” *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 2, pp. 5–8, 2010.
- [10] J. Schneider, “Cross Validation,” 1997. [Online]. Available: <http://www.cs.cmu.edu/~schneide/tut5/node42.html>. [Accessed: 07-Dec-2013].