

# Pattern Recognition of Ship Navigational Data Using Support Vector Machine

Joo-Sung Kim and Jung Sik Jeong

Department of Maritime Transportation System, Mokpo National Maritime University, Mokpo, Korea

---



---

## Abstract

A ship's sailing route or plan is determined by the master as the decision maker of the vessel, and depends on the characteristics of the navigational environment and the conditions of the ship. The trajectory, which appears as a result of the ship's navigation, is monitored and stored by a Vessel Traffic Service center, and is used for an analysis of the ship's navigational pattern and risk assessment within a particular area. However, such an analysis is performed in the same manner, despite the different navigational environments between coastal areas and the harbor limits. The navigational environment within the harbor limits changes rapidly owing to construction of the port facilities, dredging operations, and so on. In this study, a support vector machine was used for processing and modeling the trajectory data. A K-fold cross-validation and a grid search were used for selecting the optimal parameters. A complicated traffic route similar to the circumstances of the harbor limits was constructed for a validation of the model. A group of vessels was composed, each vessel of which was given various speed and course changes along a specified route. As a result of the machine learning, the optimal route and voyage data model were obtained. Finally, the model was presented to Vessel Traffic Service operators to detect any anomalous vessel behaviors. Using the proposed data modeling method, we intend to support the decision-making of Vessel Traffic Service operators in terms of navigational patterns and their characteristics.

**Keywords:** Vessel Traffic Service, Traffic pattern, Route extraction, Decision-making support system, Support vector machine

---

## 1. Introduction

Based on Article 36 of the Maritime Safety Act (20), Article 28 of the Open Ports Rules and Regulations, and the IMO RESOLUTION A.857(20) [1], a Vessel Traffic Service (VTS) can be established and operated by relevant or authorized governments as a way to prevent accidents and protect marine environments. The goals of a VTS are ship safety and efficiency at sea. However, the risk of marine accidents is increasing owing to larger and faster ships and increased volumes of marine traffic. VTS operators (VTSOs) monitor the traffic situations and provide proper information 24 hours a day to prevent potential accidents. Although such tasks are extremely important aspects of VTS operations, they are made entirely based on the specific capabilities of the VTSO [2, 3]. The decision-making of a VTSO is a matter of how they collect and utilize significant amounts of information. Kim et al. [4, 5] tried to apply a ship's dead reckoning position (DRP) on its traffic route; however, an advanced DRP requires an accurate traffic route model. Ship trajectories and navigational data are collected,

---

Received: Nov. 18, 2015  
Revised : Dec. 24, 2015  
Accepted: Dec. 25, 2015

Correspondence to: Jung Sik Jeong  
(jsjeong@mmu.ac.kr)  
©The Korean Institute of Intelligent Systems

---

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

stored, and used for traffic and accident analyses by VTS centers.

However, current traffic analyses have been conducted for simple reasons such as passage usage or frequency. In addition, traffic analyses within harbor limits have also been performed in similar ways in coastal areas [6-8]. In particular, the traffic patterns and navigational environments within harbor limits differ in different coastal areas owing to frequent changes in course and speed, a large density of vessels, the geographical conditions, construction of the port facilities, dredging operations, and so on. Depending on the ship destinations, different traffic patterns among the vessels that navigate the same passage exist, and there are few datasets classifying each specific leg [9-11]. Owing to these characteristics, it is necessary to develop methods applicable to particular harbor limits.

In this study, we present a data modeling method for vessels that are the last have called, the results of which can be calculated using a small dataset. A support vector machine-regression (SVM-R) was used to develop the optimal-route and navigational-pattern model. A K-fold cross-validation and grid search were used for selecting the appropriate parameters. A simulation was conducted using a virtual channel and ships to validate the proposed modeling method. The channel complexity used is similar to the circumstances found in actual harbor limits. A group of vessels was composed within the channel, where each vessel was given various changes in speed and course. As a result of machine learning, the optimal route and voyage data model were obtained. Finally, the proposed model was presented to VTSOs to detect anomalous vessel behaviors.

## 2. Support Vector Machine-Regression

An SVM is a tool used for classifying data consisting of a hyperplane, and generates the maximum margin among the data using a supervised learning method [12-14]. An SVM was originally developed for solving classification problems, but has recently been extended to address problems associated with regression and probability density estimations [15-17].

An SVM assumes that the training dataset,  $(x_1, y_1), \dots, (x_N, y_N) \in R^M \times R$ , are given. When learning datasets ( $N$ ) are given in input space  $R^M$ , a linear function  $f(x)$  that has spaces  $\varepsilon$  from output vector  $y$  against input vector  $x$  is

$$f(x) = \langle w, x \rangle + b \text{ with } w \in R^M, b \in R. \quad (1)$$

As shown in Figure 1, when training datasets are given out-

side of space  $\varepsilon$ , a solution to the following optimization problem using slack variables  $\xi_i, \xi_i^*$  is required [12, 15]:

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \\ \text{subject to} \quad & y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ & \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i \\ & \xi_i, \xi_i^* \geq 0. \end{aligned} \quad (2)$$

Meanwhile, it is necessary to solve the separation of a non-linear dataset. In general, a linearly separable condition may occur if the dataset transfers to a higher-level space using a non-linear transfer function. Figure 2 shows a case of linear data that cannot be classified [12]. Non-detachable data are transferred to a non-linear case using a non-linear transfer function (mapping)  $z_k = (x_k, x_k^2)$ .

For two-dimensional data, it is possible to obtain a linearly separable plane when mapping as a three-dimensional space. The kernel function is defined as follows:

$$K(x, x') = \phi(x)^T \phi(x'). \quad (3)$$

As shown in Figure 3, function  $f(x)$  is composed in a non-linear form but can be transformed into a linear form using a kernel function [12, 15]:

$$\begin{aligned} y &= \sum_{i=1}^N \alpha_i + \alpha_i^* \cdot \langle \phi(x_i), \phi(x) \rangle + b \\ &= \sum_{i=1}^N (\alpha_i + \alpha_i^*) \cdot K(x_i, x) + b. \end{aligned} \quad (4)$$

The types of kernel functions include splines, polynomials, hyperbolic tangents, sigmoid functions, and radial basis functions (RBFs) [18]. In this study, we used an RBF, which is generally known to have an excellent performance capability:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right). \quad (5)$$

## 3. Data Extraction

Data extraction is a three-step process:

- 1) Data collection
- 2) Data classification
- 3) Data modeling

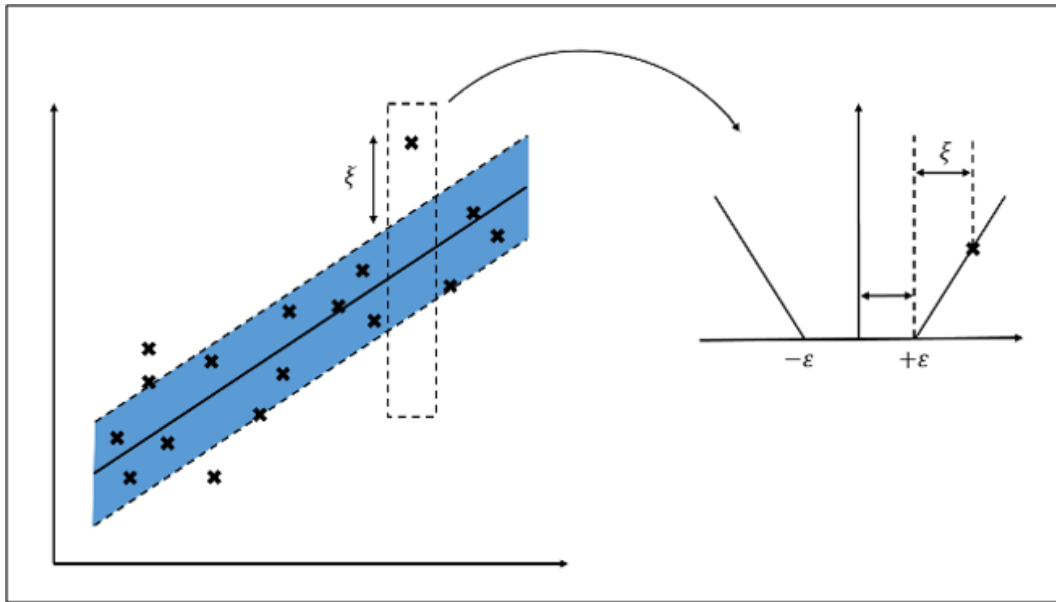


Figure 1.  $\epsilon$ -Insensitive loss function.

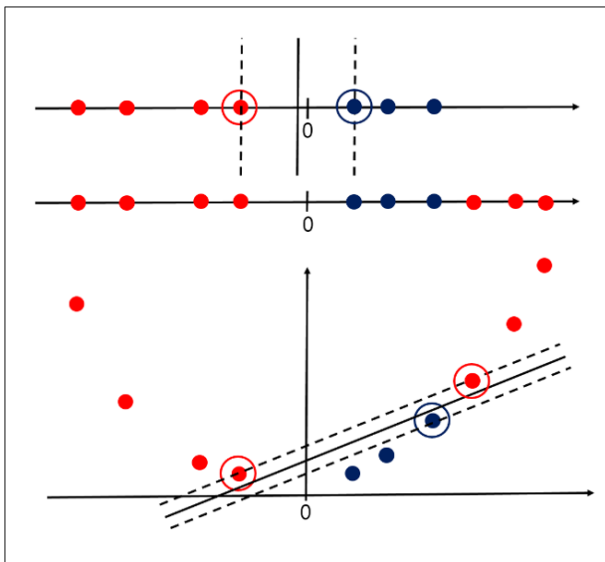


Figure 2. Non-detachable data mapping to two-dimensional space.

In this section, we present the proposed data modeling method. The collected trajectory data are categorized as regional separations and the separated data are classified as particular routes based on the gate line at each passage [19, 20]. The divided datasets need to be converted into a type of column vector to match the SVM-R algorithm before the modeling process is conducted. The modeling process can be described through the following steps [18]:

- 1) Dataset transformation
- 2) Data scaling
- 3) Kernel function selection
- 4) Parameter selection
  - 4-1) K-fold cross-validation
  - 4-2) Grid search
- 5) Selected model training
- 6) Model validation

For the above modeling process, it is necessary to separate the datasets into sub-datasets and transform them into sets of column vectors such as time  $T(n) = [t_1, t_2, \dots, t_n]^T$ , longitude  $Lon(t_n) = [x(t_1), x(t_2), \dots, x(t_n)]^T$ , latitude  $Lat(t_n) = [y(t_1), y(t_2), \dots, y(t_n)]^T$ , course  $Co(t_n) = [c(t_1), \dots, c(t_n)]^T$ , and speed  $Spd(t_n) = [s(t_1), s(t_2), \dots, s(t_n)]^T$ .

According to Hsu et al. [18], scaling prior to applying an SVM is very important. The main advantage of scaling is to avoid attributes within greater numeric ranges dominating those within smaller numeric ranges. Another advantage is to avoid numerical difficulties during the calculation.

The vector  $T(n) = [t_1, t_2, \dots, t_n]^T$  can be transformed into

$$T_{scaling} = [(t_1 - t_1)/(t_k - t_1), (t_2 - t_1)/(t_k - t_1), \dots, (t_k - t_1)/(t_k - t_1)]^T$$

after scaling from zero to 1. The other vectors (i.e.,  $Lon(t_n)$ ,

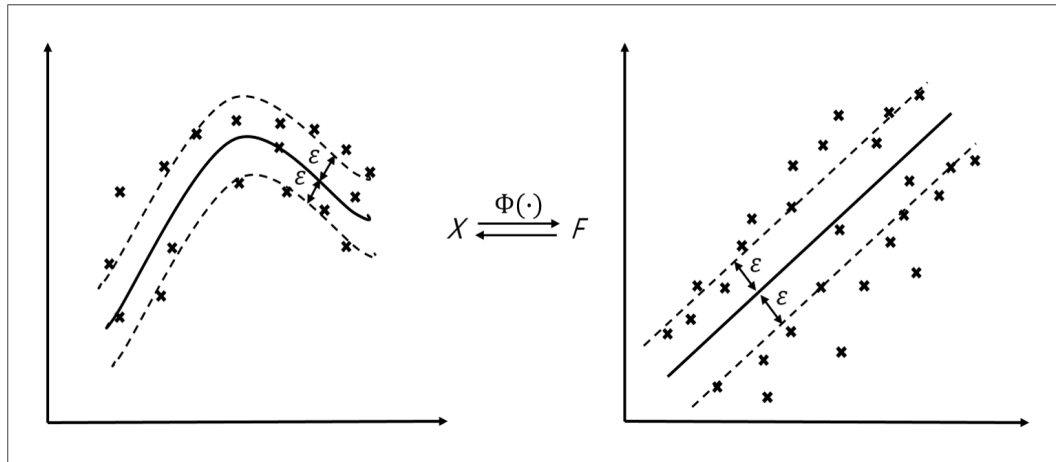


Figure 3. Kernel functions for transforming data into higher dimensional feature space.

$Lat(t_n)$ ,  $Co(t_n)$ , and  $Spd(t_n)$ ) can also be transformed into

$$T_{scaling} = \frac{T - Min[T]}{Max[T] - Min[T]} \tag{6}$$

For the optimal parameter search, the training sets need to be divided into ‘k’ sub-sets of equal size. One sub-set is tested using the algorithm and trained on the remaining ‘k-1’ sub-sets. The cross-validation procedure can prevent an over-fitting problem.

Figure 4 shows a k-fold cross-validation, which can be explained as below:

$$Min_{validation\ error} = Min \left( \frac{q_1 + q_2 + q_3 + \dots + q_k}{k} \right) \tag{7}$$

As a result of each instance, a validation errors  $(q_1, \dots, q_k)$ , and the minimum validation error can be found.

### 4. Simulation

We reproduced a curved virtual channel and ship trajectories with various characteristics. The circumstances are similar with the actual harbor limits. As shown in Figure 5, the channel is in the form of a letter ‘R’ and the ship tracks are within the channel. Before conducting the simulation, the whole datasets were separated into sub-datasets and transformed into sets of column vectors. The simulation conditions are as follows.

- 1) Sharp curves
- 2) Narrow areas
- 3) Various directions
- 4) High density

- 5) Different course changes
- 5) Different speeds
- 6) Navigating within the channel only
- 7) Same starting and end points
- 8) No detours or short-cuts through the channel

The separated datasets are divided into sub-datasets again, which consist of a validation dataset and four learning datasets. A 5-fold cross-validation and grid search were conducted to select the optimal parameters, and an RBF was used as a kernel function. The results of the calculations are listed in Table 1.

As shown in Figures 6(left) and 7(left), each sub-dataset has a certain pattern, but the datasets are necessary to be adjusted because the ships are sailing within the same location but at different speeds.

After scaling the data, machine learning was conducted for whole datasets. As a result of the learning, red curves were obtained, as shown in Figures 6(right) and 7(right).

As shown in Figure 8, the optimal route is obtained after modeling. The entire datasets were learned to build the model by applying the selected parameters. The data conversion and SVM-R were performed for all sub-datasets.

The final model can be adjusted based on the number of training datasets or samplings. It is also possible to reduce the calculation time of the modeling process by downsizing the number of support vectors used.

As shown in Figure 9, the course-change model was obtained through machine learning. The extracted navigational model can be compared to data values for changes in course and speed at a certain position through a combination with the coordinates. It can also be used for preventing accidents by identifying

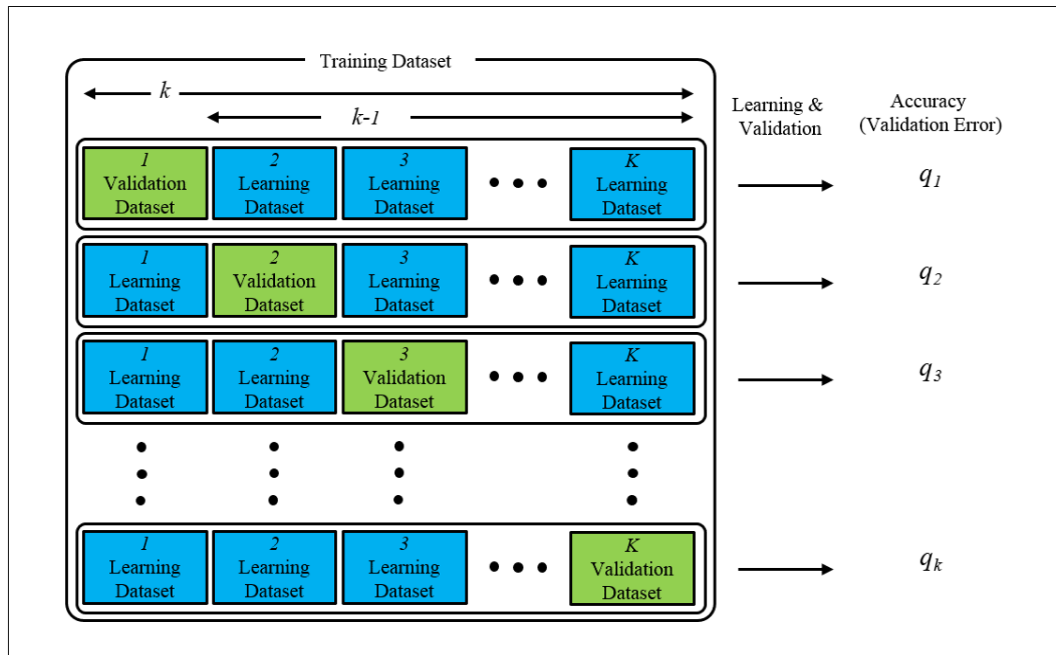


Figure 4. K-fold cross-validation.

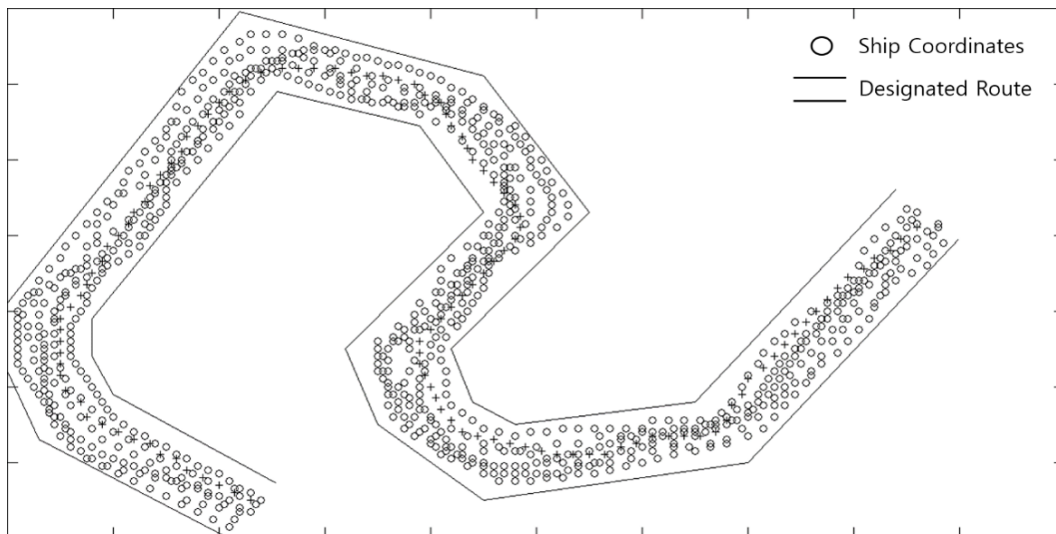


Figure 5. Virtual curved channel and ship tracks.

the anomalous movements of ships and predicting changes in navigation before they occur. If any deviation in a ship's position or navigational data, such as course and speed, occurs it means that the uncertainty about the ship's future position is increasing. Therefore, the proposed model can be used for the prediction of future navigation.

## 5. Conclusion

It is important to understand a ship's navigational intention and its location in order to predict traffic situations and organize traffic plans and determine dangerous traffic situations. In particular, the identification of a ship's anomalous navigational behaviors within VTS areas is available only through the constant monitoring of traffic states by VTSOs. In addition, the

Table 1. Parameter selection results

g\ε	0	1	2	3	4	5
6	4.4662	4.2730	4.2620	4.7694	5.2301	6.2994
7	4.5501	4.2347	4.2148	4.6951	5.2297	6.4380
<b>8</b>	4.6137	4.2476	<b>4.1977</b>	4.6256	5.6486	8.1237
9	4.6653	4.2805	4.2111	4.8137	6.2200	9.2866
10	4.8653	4.4474	4.4471	5.0990	6.6824	10.0518

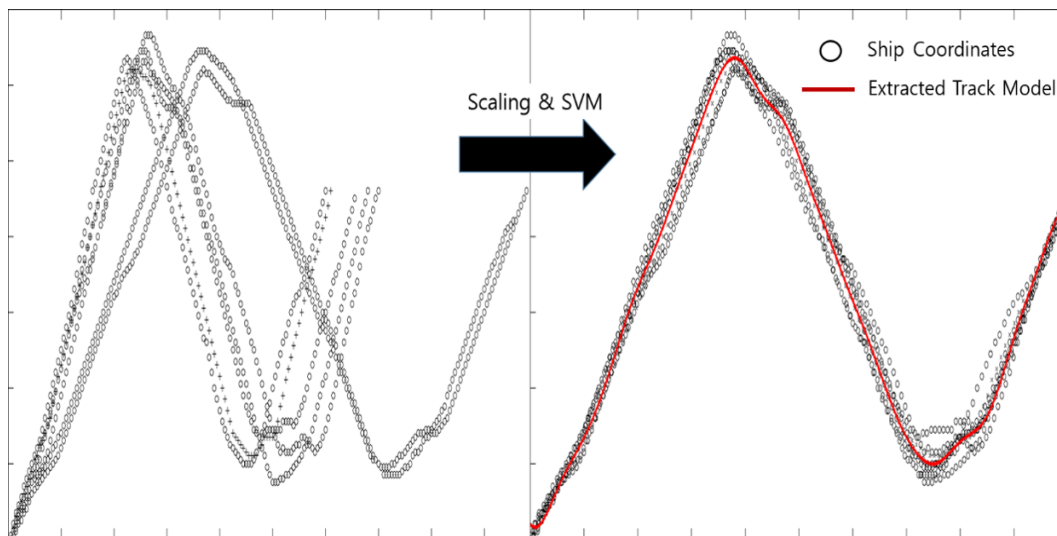


Figure 6. Plotted latitude components (left), and results of machine learning after scaling (right). SVM, support vector machine.

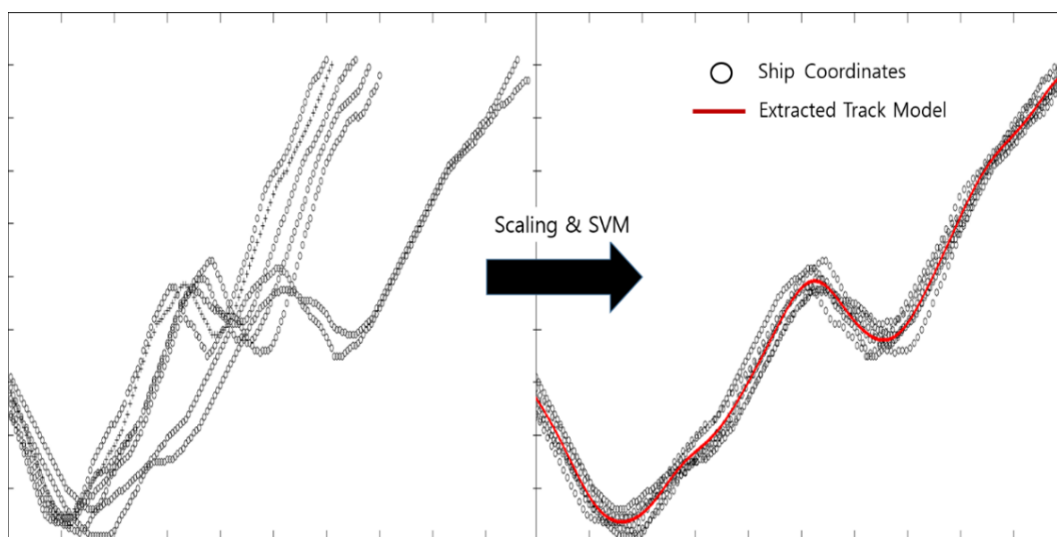


Figure 7. Plotted longitude components (left), and results of machine learning after scaling (right). SVM, support vector machine.

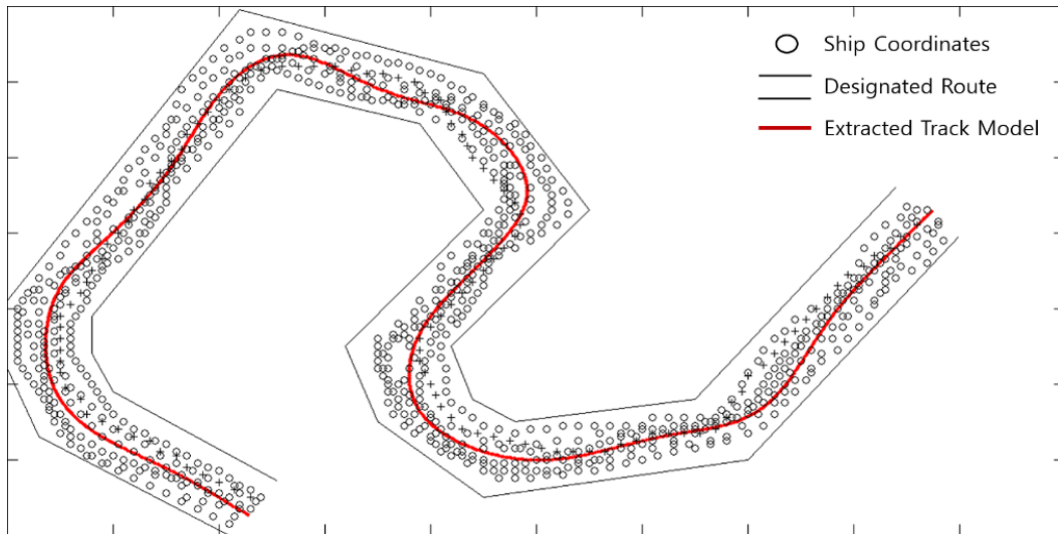


Figure 8. Extracted traffic route.

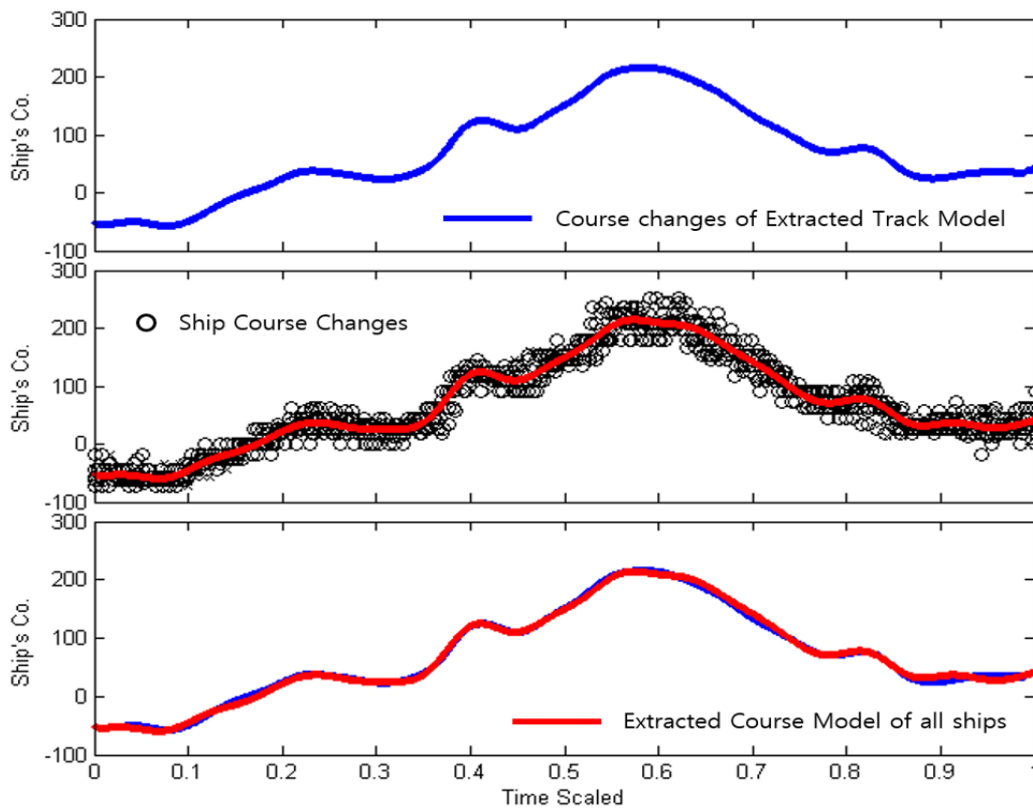


Figure 9. Course changes of extracted route (upper), course data of all ships (middle), and extracted course model (lower).

workload and stress levels of VTSOs are increasing in VTS areas owing to a high density of traffic and narrow maritime waters. This study was conducted to prevent accidents to marine traffic by reducing errors of VTSOs caused by individual

differences in ability. The proposed method can be used for identifying a ship's deviation and irregular changes in speed and course. It may also be used for supporting the decision-making of a VTSO under complex situations by improving a ship's

dead reckoning position. Moreover, it is expected to assist a proper Traffic Organization Service (TOS) or Information Service (INS) when multilateral relationships are encountered in a VTS area. As a future study, we plan to develop additional methods that can be applied to a large number of ships with different types and navigational environments. Additionally, an application needs to be developed for investigating marine accident in VTS areas.

## Acknowledgements

This work was supported by ETRI through Maritime Safety & Maritime Traffic Management R&D Program of the Ministry of Public Safety and Security/Korea Institute of Marine Science & Technology Promotion (No. 2009403, Development of next generation VTS for maritime safety).

## References

- [1] International Maritime Organization, *Guidelines for Vessel Traffic Services [IMO Resolution A.857(20)]*. London: International Maritime Organization, 1997.
- [2] J. S. Kim, J. S. Jeong, and G. K. Park, "Prediction table for marine traffic for vessel traffic service based on cognitive work analysis," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 13, no. 4, pp. 315-323, 2013. <http://dx.doi.org/10.5391/ijfis.2013.13.4.315>
- [3] J. S. Kim, J. S. Jeong, and G. K. Park, "Utilization of planned routes and dead reckoning positions to improve situation awareness at sea," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 14, no. 4, pp. 288-294, 2014. <http://dx.doi.org/10.5391/ijfis.2014.14.4.288>
- [4] J. S. Kim, J. S. Jeong, and G. K. Park, "Application of planned routes and dead reckoning positions for supporting decision-making of vessel traffic services," in *Proceeding of the International Symposium on Advanced Intelligent Maritime Safety and Technology*, Mokpo, Korea, 2014, pp. 121-124.
- [5] J. S. Kim, J. S. Jeong, G. K. Park, J. Y. Jung, J. H. Lee, T. Oh, and H. Han, "Extraction of ship's tracking model for improving situation awareness at vessel traffic service areas," in *Proceeding of the Korean Institute of Intelligent Systems (KIIS) Spring Conference*, Ansan, Korea, 2015, pp. 51-52.
- [6] T. Hong, "Development of a system for transmitting a navigator's intention for safe navigation," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 14, no. 2, pp. 130-135, 2014. <http://dx.doi.org/10.5391/ijfis.2014.14.2.130>
- [7] E. K. Kim, J. S. Jeong, G. K. Park, and N. K. Im, "Characteristics of ship movements in a fairway," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 12, no. 4, pp. 285-289, 2012. <http://dx.doi.org/10.5391/ijfis.2012.12.4.285>
- [8] J. S. Jeong, K. I. Kim, and G. K. Park, "A quantitative collision probability analysis in port waterway," *Journal of Korean Institute of Intelligent Systems*, vol. 22, no. 3, pp. 373-378, 2012. <http://dx.doi.org/10.5391/jkiis.2012.22.3.373>
- [9] B. T. Moon, Y. H. Ryu, and G. J. Joe, "Study on software architecture for intelligent vessel navigation analysis," *Journal of Korean Institute of Intelligent Systems*, vol. 22, no. 5, pp. 583-589, 2012. <http://dx.doi.org/10.5391/jkiis.2012.22.5.583>
- [10] K. I. Kim, J. S. Jeong, and G. K. Park, "Assessment of external force acting on ship using big data in maritime traffic," *Journal of Korean Institute of Intelligent Systems*, vol. 23, no. 5, pp. 379-384, 2013. <http://dx.doi.org/10.5391/jkiis.2013.23.5.379>
- [11] D. Y. Kim, G. K. Park, and H. Y. Kim, "A study on the ship information fusion with AIS and ARPA Radar using by blackboard system," *Journal of Korean Institute of Intelligent Systems*, vol. 24, no. 1, pp. 16-21, 2014. <http://dx.doi.org/10.5391/jkiis.2014.24.1.016>
- [12] S. R. Gunn, "Support vector machines for classification and regression," School of Electronics and Computer Science, University of Southampton, UK, Technical Report, 1998.
- [13] G. H. Kim, Y. W. Kim, S. J. Lee, and G. J. Jeon, "A hierarchical clustering method based on SVM for real-time gas mixture classification," *Journal of Korean Institute of Intelligent Systems*, vol. 20, no. 5, pp. 716-721, 2010. <http://dx.doi.org/10.5391/jkiis.2010.20.5.716>
- [14] H. H. Choi, G. H. Lee, J. G. Kim, Y. B. Joo, B. J. Choi, K. H. Park, and B. J. Yun, "A defect inspection method in TFT-LCD panel using LS-SVM," *Journal of Korean*



*Institute of Intelligent Systems*, vol. 19, no. 6, pp. 852-859, 2009. <http://dx.doi.org/10.5391/jkiis.2009.19.6.852>

- [15] V. N. Vapnik, *Statistical Learning Theory*. New York, NY: Wiley, 1998.
- [16] T. H. Jo, "Modified version of SVM for text categorization," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 8, no. 1, pp. 52-60, 2008. <http://dx.doi.org/10.5391/ijfis.2008.8.1.052>
- [17] M. S. Kim and S. Y. Lee, "Pattern classification for biomedical signal using BP algorithm and SVM," *Journal of Korean Institute of Intelligent Systems*, vol. 14, no. 1, pp. 82-87, 2004. <http://dx.doi.org/10.5391/jkiis.2004.14.1.082>
- [18] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," Department of Computer Science, National Taiwan University, Taipei, Taiwan, 2003.
- [19] J. S. Kim and J. S. Jeong, "Vessel trajectory and route detection in vessel traffic service areas using machine learning theories," in *Proceedings of International Symposium on Advanced Intelligent Maritime Safety and Technology*, Daejeon, Korea, 2015, pp. 142-145.
- [20] J. S. Kim, J. S. Jeong, and J. Y. Jung, "Recreating ship's reference routes to predict traffic situations in vessel traffic service areas," in *Proceeding of the Korean Society of Marine Environment and Safety (KOSOMES) Fall Conference*, Busan, Korea, 2014, pp. 295-297.



**Joo-Sung Kim** is a Ph.D. candidate in International Maritime Transportation Sciences at Mokpo National Maritime University in Korea, and a VTSO at Kyeong-in VTS Center in Korea. His research interests include maritime traffic engineering, ship collision avoidance, maritime information and communication network. He received his B.S. degree in Nautical Science from Mokpo National Maritime University in Korea in 2004 and his M.S. degree in International Maritime Transportation Sciences from Mokpo National Maritime University in Korea in 2014. His research areas include intelligent system, fuzzy system, human factors engineering, work analysis, vessel traffic services, maritime transportation system, etc.  
E-mail: [jskim81@korea.kr](mailto:jskim81@korea.kr)



**Jung Sik Jeong** is a professor in the Department of International Maritime Transportation Sciences at Mokpo National Maritime University in Korea. His research interests include intelligent system, fuzzy system, intelligent navigation control system and maritime information. He received his B.S. degree in Nautical Science from Korea Maritime University in 1987, his M.S. degree in Communication and Electronic Engineering from Korea Maritime University in 1993, and his Ph.D. degree in Electrical and Electronic Engineering from Tokyo Institute of Technology in 2001. He worked at Korea Telecom at 1996. His research areas include maritime traffic engineering, ship collision avoidance, maritime information and communication network, etc.  
E-mail: [jsjeong@mmu.ac.kr](mailto:jsjeong@mmu.ac.kr)