Extraction of Reference Seaway through Machine Learning of Ship Navigational Data and Trajectory

Joo-Sung Kim¹ and Jung Sik Jeong²

¹Kyeong-In VTS Center, Ministry of Public Safety and Security, Incheon, Korea
²Department of Maritime Transportation System, Mokpo National Maritime University, Mokpo, Korea

Abstract

Vessel Traffic Services operators have kept sharp monitoring and provided the appropriate information to ensure safe and effective navigation. While attending the tasks, analysis of traffic patterns and navigational data is required to conduct accurate situation assessment in decision-making process of VTS operators (VTSO). Unfortunately, there are problems in the process of data analysis such as appropriateness of time, VTSO’s personal error and improper judgment. Therefore, objective and proper data analysis is necessary to solve above matters. However, it is virtually impossible to monitor all vessels because there are many vessels in the VTS area and at the same time complex traffic situations are produced. In this study, we proposed a machine learning algorithms for objective and accurate pattern recognition and data modeling. Support Vector Regression algorithm was used for data learning and modeling. The optimal parameters were selected through v-fold cross validation and grid search. The machine learning was conducted with virtual route and ship tracks that are similar with real navigational environment. As a result, we presented reference route and navigational patterns. We expect that the proposed modeling methods could be utilized for relevant tasks as the useful information to VTSO and/or ship’s mater.

Keywords: Vessel Traffic Services, Machine learning, Sea traffic route, Decision making, Pattern recognition

1. Introduction

The navigational environment in harbour limits are complex and changeable due to concentrated traffic density, development of port facilities and other geographical and environmental factors. Therefore, required information is becoming increasingly diverse to support safe and effective marine traffic [1][2]. According to the e-navigation strategy implementation, in addition, the role of shore based information is expected to gradually increase and Vessel Traffic Services (VTS) will be the part of important roles [3][4]. VTS has established by authorized governments in necessary places in order to promote safe and efficient navigation and prevent marine accidents based on ‘IMO RESOLUTION A.857 (20) on Guidelines for Vessel Traffic Services’, Article 36 of ‘Maritime Safety Act’, Chapter 4 of ‘Act of Ship Arrival and Departure’ and ‘rules of implementation for Vessel Traffic Services’ [5][6]. In recent
years, the role of VTS is expanded to adjust traffic situations in VTS areas as providing effective information to vessels based on collected and selected navigational data [9][10].

Understanding traffic patterns is one of the most important parts for situational awareness of maritime traffic in order to predict traffic situations and ship’s positions. In addition, VTS operators (VTSO) who have to monitor wide areas should keep a capability of accurate assessment within limited time [11][12]. Therefore, it is necessary for development of decision-making support tools to relieve cognitive workload and stress of VTSOs.

In this study, we aimed to help VTSOs and ship’s masters by providing information on route usage through the data analysis and pattern recognition of navigational data. In addition, we intended to contribute to the traffic monitoring and prediction. The machine learning algorithm was used to construct a reliable reference route with the small number of data set consisting of data for the most recent sailing. Support Vector Regression (SVR) algorithm was used for pattern recognition and modeling. The optimal parameters were selected through \(v\)-fold cross validation and grid search. The machine learning was conducted with virtual route and ship tracks that are similar with real navigational environment. As a result, we presented reference route model and utilization method of this study.

2. Modeling Methods

2.1 Data Processing

For building a learning model in SVR, the most important thing is to choose the kernel function and optimal parameters. In the relevant study, Hsu et al. [13] proposed LIBSVM algorithm with cross validation and grid search. They presented following six steps in the data learning and SVM model design process.

1. Transform data
2. Conduct simple scaling on the data
3. Consider the RBF kernel as a kernel function
4. Use cross-validation to find the best parameter
5. Use the best parameter to train the whole training set
6. Test

In this paper, we configured modeling process with reference to parameter selection that proposed in the relevant articles. Modeling process for extracting reference route can be summarized in the following steps:

(1) Data collection
(2) Data classification
   - Classification of target area
   - Classification of target route
   - Construction of sub-data sets
(3) Data learning and model extraction
   - Conversion of input data sets
   - Data scaling
   - Parameter range selection
   - Optimal parameter selection
   - Optimal model selection
(4) Database construction

The simulation was performed with the process from (2)-③ to (3)-⑤ because classified virtual data sets were used for the simulation. Therefore, the classification steps were omitted in the simulation.

Meanwhile, same steps were applied for route extraction and pattern recognition of navigational data.

2.2 Construction of Input Dataset

Configuration of data set is the structure of \((M \times N)\) matrix for each individual ship. Because the structure is easy to transfer with each component to column vectors for learning [14][17].

First, the program selects any one of the vessels to learn entire trajectory and divides total distance to \(k\) equal parts while moving the entire trajectory to constitute the \(k\) sets of data. Each sub-data set is composed of a distance of \(2k\) with overlapping the distance \(k\). Therefore, the number of sub-data sets is \(k - 1\). Each divided data sets are learned twice and models of the overlapping legs are composed of one final model after final learning process.

Due of each trajectory data obtained from individual vessel, the number of constituent elements and data range are different according to the ship speed. The data scaling is required in order to effectively learn the trajectory without being affected by the differences. When the data set \(X^t = [x^t_1 x^t_2 x^t_3 \cdots x^t_i]\) scales in the input variables as a value between 0 and 1,

\[
X_{scaling}^t = \frac{X^t_i - Min[ X^t ]}{Max[ X^t ] - Min[ X^t ]}.
\]
2.3 Support Vector Regression

The Support Vector Machine (SVM) is a classifying technique to configure the hyper-plane to maximize the margin through supervised learning. Although the SVM originally developed for classification issue, it has been extended to problems with regression and probability density estimation [18–20].

The SVM can be applied to regression model by introducing the loss function. When the training data set \((x_1, y_1), \ldots, (x_N, y_N) \in \mathbb{R}^M \times \mathbb{R}\) are given, \(\varepsilon\)-SVR that was proposed by Vapnik is to find \(f(x)\) with minimum \(w\) and insensitive parameter \(\varepsilon\) against all training data sets \([21, 22]\). Where \(N\) is the number of training data, \(\mathbb{R}^M\) represents the input space. The linear function is as below,

\[
 f(x) = \langle w, x \rangle + b \text{ with } w \in \mathbb{R}^M, \ b \in \mathbb{R}.
\]

As Figure 1, the slack variable \(\xi_i, \xi_i^*\) are used to accept the training data locating the outside of space \(\varepsilon\), the \(f(x)\) is presented to

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

Therefore, the linear equation can be expressed the output SVR,

\[
 f(x) = \sum_{i=1}^{N} (\alpha_i + \alpha_i^*) \cdot \langle x_i, x \rangle + b.
\]

In SVR, the training data in the input space can be mapped in the high-dimensional space using a non-linear mapping function \(\varphi\) to approximate to a non-linear function \((\varphi : \mathbb{R}^M \rightarrow F)\). This non-linear mapping function is called the kernel function \([21, 22]\) and

\[
 K(x, x') = \langle \varphi(x), \varphi(x') \rangle.
\]

Therefore, the non-linear equation can be expressed the output SVR,

\[
 f(x) = \sum_{i=1}^{N} (\alpha_i + \alpha_i^*) \cdot k(x_i, x) + b.
\]

In this study, we used Gaussian Radial Basis Function (RBF) as the kernel function to solve the non-linear problem. Gaussian RBF presents successful performance of the various sector and it can be expressed that

\[
 K(x, y) = \exp \left(-\frac{(x-y)^2}{2\sigma^2}\right).
\]

2.4 Optimal Parameter Selection

Parameter values \(\varepsilon\) and \(\sigma\) are needed to be selected in the model selection. The parameter \(\varepsilon\) is a value to adjust the complexity of the model by adjusting the number of support vector. Because small value \(\varepsilon\) takes a lot of training data as support vectors, output model becomes complex. On the other hand, larger \(\varepsilon\) reduces the number of support vectors and model can be simply determined \([14–17]\). Parameter value \(\sigma\) is the standard deviation of the kernel function Gaussian RBF. Lager \(\sigma\) favors models that close to straight lines and small value of \(\sigma\) takes curved lines.

The \(v\)-fold cross validation and grid search were used for avoiding the over-fitting of models presented by SVR and selecting optimal parameters of route model. First, whole data sets are needed to be divided into the training data set and the evaluation data set. At this time, the number of these sub-data sets will be \(v\). One of the sub-data set \(v\) can be the validation set and the others \((v - 1)\) are training sets for a validation of a divided group. These procedure will be repeated as \(v\) times. The average value of these accuracies can be an objective indicator of the performance of the whole data sets. It can be explained as

\[
 Mean_{\text{validation error}} = k^{-1} \sum_{j=1}^{k} \text{Min}(q_j).
\]
Figure 2. Simulation dataset and data division.

Figure 3. Data learning on sub-dataset.

Figure 4. Extracted model of sub-dataset.
The $q_i$ represents the validation error, and the number of models for the verification is determined based on the number of $v$ and the range of parameter selection.

3. Simulation

3.1 Outline of Simulation

The virtual tracks are composed in order to conduct machine learning for ship trajectories and navigational data. The shape of passage is similar with character of alphabet ‘S’ that includes tight bends. Fifteen ships are composed as data sets and they have various characteristics while sailing on the route with different course and speed. The data sets are transformed to the learning formation and divided into eight sub-data sets. The virtual passage has following features;

1. It has tight bends,
2. narrow channels,
3. designated route track,
4. and diverse changes on course and speed.

Among the data sets, a vessel is showing anomaly behavior and the vessel break away from the designated route twice. The data sets and sub-data sets are presented on Figure 2.

4. Data Learning

Data learning on divided sub-data sets was carried out through each learning engine. Here, we presented the learning results of latitude and longitude components in the progress of the learning steps, and the results of these components is shown in Figure 3.

Meanwhile, the sub-models are extracted after data learning and the results are shown in Figure 4.

After extracting the each sub-model that has overlapped sections, the models are arranged and scaled as similar coordinates. After the process, a final model can be obtained with learning process on the whole sub-model data. The extraction steps of the sub-models and the final model are in same process and it will be repeated until finding the best model. The final route model for the virtual route is shown in Figure 5.

Navigation data are needed to be compared with the specific position of each ship. Therefore, the data are given to the approximate point of the model. The given data are differences among the data taken from a specific location of each vessel.

Figures 6-10 show comparison of data between the target vessel and others.

As the results of comparison of each vessel’s deviation, the anomaly behavior can be detected when a ship has sudden changes on course or speed. Users can easily recognize ship’s deviation and it would be a reference information to predict the above behavior in advance. When the position of the target vessel deviated from the specified path, the difference of ship’s speed and course began to increase. In other words, when an abnormality of ship’s speed or course occurs, the target vessel leaves the designated route and a dangerous situation occurs. Therefore, if the abnormalities of ship’s speed or course are monitored together with the position of the ship, it is possible to detect in advance whether the ship has deviated or not.

5. Conclusion

It is very important to understand the traffic route of a ship in order to prevent marine accidents. It is an indispensable task in VTS to predict the ship’s traffic and analyze the navigational data to determine the ship’s abnormal behavior. In this study, we presented reference route model using SVR. The SVR algorithm was used for pattern recognition and modeling. The optimal parameters were selected through $v$-fold cross validation and grid search. The machine learning was conducted with virtual route and ship tracks that are similar with real navigational environment. The proposed method is able to compose the dynamic route model based on recent and a few ship’s navigational data. And it was possible to extract reference data not only the route of the ship but also the speed and the course that can be used for judging the abnormality of the ship. The proposed model can be used to prevent accidents and provide reference routes through the prediction and decision-support information to VTSOs and mariners. For further study, we need to develop the appropriate method for data division in case that the legs become long and complex. In addition, we conduct simulations with users based on real navigational data and develop various applications related with data learning.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.
Acknowledgements

This work was conducted as the Research for Development Strategy to Future Maritime Traffic Environments and Applications to Maritime Safety Technology which was supported by KRISO from November 1, 2016 to February 28, 2017 (Project No. 2016-0096).
Figure 8. Comparison of speed changes.

Figure 9. Relationship between deviation and course differences.

Figure 10. Relationship between deviation and speed changes.

References

[1] 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]

[2] 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]

[3] International Maritime Organization, Strategy for the Development and Implementation of e-Navigation (MSC85-report, Annexes 20 and 21). Available [http://www.imo.org/en/OurWork/Safety/Navigation/Documents/enavigation/MSC%2085-%20annex%20-%20Strategy%20for%20the%20Development%20and%20Implementation%20of%20e-Navigation/pdf]
[4] 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]

[5] International Maritime Organization, Guidelines for Vessel Traffic Services (Resolution A.857(20)), International Maritime Organization, London, UK, 1997. Available [http://www.maritime-vts.co.uk/A857.pdf]

[6] Korea Ministry of Government Legislation, “Act of Ship Arrival and Departure,” 2015, Available [https://www.moleg.go.kr/english/]

[7] Korea Ministry of Government Legislation, “Maritime Safety Law,” 2015, Available [https://www.moleg.go.kr/english/]

[8] Ministry of Public Safety and Security, Vessel Traffic Service Operational Manuals, 2015. Available [http://www.mpss.go.kr/en/]

[9] 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]

[10] 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]

[11] 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]

[12] 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]

[13] C. W. Hsu, C. C. Chang, and C. J. Lin, A Practical Guide to Support Vector Classification. Taipei, Taiwan: Department of Computer Science, National Taiwan University, 2003. Available [http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf]

[14] J. S. Kim and J. S. Jeong, “Pattern recognition of ship navigational data using support vector machine,” International Journal of Fuzzy Logic and Intelligent Systems, vol. 15, no. 4, pp. 268-276, 2015. [http://dx.doi.org/10.5391/IJFIS.2015.15.4.268]

[15] J. S. Kim and J. S. Jeong, “Vessel trajectory and route detection in vessel traffic service areas using machine learning theories,” in Proceedings of 2015 International Conference on Advanced Intelligent Maritime Safety and Technology, Daegon, Korea, 2015, pp. 142-145.

[16] J. S. Kim and J. S. Jeong, “Vessel trajectory analysis for setting reference route model within harbor limits,” Proceedings of Korea Institute of Intelligent Systems Spring Conference 2016, vol. 26, no. 1, pp. 163-164, 2016.

[17] J. S. Kim, “A design of reference route model based on SVR through reconstruction of ship trajectories in VTS area,” Ph.D. dissertation, Department of Maritime Transportation System Graduate School of Mokpo Maritime University, Mokpo, Korea, 2016.

[18] V. N. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, NY, 1995.

[19] V. N. Vapnik, Statistical Learning Theory, Wiley, New York, NY, 1998.

[20] 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]

[21] S. R. Gunn, Support Vector Machines for Classification and Regression: Technical Report. Southampton, UK: School of Electronics and Computer Science, University of Southampton, 1998. Available [http://users.ecs.soton.ac.uk/srg/publications/pdf/SVM.pdf]

[22] H. Y. Han, Introduction to Pattern Recognition, Hanbit Academy, Inc., Seoul, Korea, 2014.
Joo-Sung Kim is a vessel traffic services operator 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, his M.S. degree in International Maritime Transportation Sciences from Mokpo National Maritime University in Korea in 2014 and his Ph.D. degree in International Maritime Transportation Sciences from Mokpo National Maritime University in Korea in 2016. 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

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