International Journal of Digital Information and Wireless Communications (IJDIWC) 9(4): 203-207
The Society of Digital Information and Wireless Communications, 2019 ISSN: 2225-658X (Online); ISSN 2412-6551 (Print)

A Study for Development of Route Prediction Method for Vessel using Machine Learning

Jumpei MATSUE [1] Kotetsu KAWAKAMI [1] Ryota ISOZAKI[2]
Shunsuke INADA[2] Masashi SUGIMOTO[2∗] Shinji TSUZUKI[2]
Kazuhiko NAGAO [3]
[1]Department of Electrical and Electronic Engineering,
Faculty of Engineering, Ehime University
3 Bunkyo, Matsuyama C., Ehime Pref. 790-8577 Japan
[2]Department of Electrical and Electronic Engineering,
Graduate School of Science and Engineering, Ehime University
3 Bunkyo, Matsuyama C., Ehime Pref. 790-8577 Japan
E-mail: sugimoto.masashi.du@ehime-u.ac.jp
(∗ : *Corresponding author*)
[3]Information Science and Technology Department
National Institute of Technology,Yuge College (Yuge National College of Maritime Technology)
1000 Shimoyuge, Kamijima T., Ehime Pref. 794-2593 Japan

ABSTRACT

Nowadays, vessels such as cargo vessels, ferries, and fishing boats are constantly coming and going on all over the sea. On the other hands, unfortunately, collisions among vessels account for about one-quarter of the total accidents. In this paper, small boats such as fishing boats and pleasure boats that is not required to mount Automatic Identification System (AIS) has been focused on. In detail, a module for LoRa (Long Range) wireless, which is one of the LPWA (Low Power, Wide Area) standards, and a GPS module are mounted on a small vessel. Moreover, machine learning is used to determine by predicting the navigation route, a system to avoid collisions between ships has been considered.

KEYWORDS

Route Prediction, Vessel Collision Avoidance, LoRa, Machine Learning, Random Forest, k-Means Clustering.

1 INTRODUCTION

Nowadays, vessels such as cargo vessels, ferries, and fishing boats are constantly coming and going on all over the sea. On the other hands, unfortunately, collisions among vessels account for about one-quarter of the total accidents [1]. Large vessels are obliged to have an Automatic Identification System (AIS) which is used to display the position of other vessels on the radar. From the reason, there are few cases to collision with each other large vessels. On the other hand, small boats such as fishing boats and pleasure boats are not required to mount AIS. Therefore, a small vessel can only confirm the position of another vessel by visual observation by passengers or pilot. In addition, the AIS equipment itself is very expensive, and a license for a radio station is required for activation. However, most of accidents had been caused by small vessels, therefore, to deal this issue is important [2]. Therefore, in this study, a module for LoRa (Long Range) wireless, which is one of the LPWA (Low Power, Wide Area) standards, and a GPS module are mounted on a small vessel, and machine learning is used to determine by predicting the navigation route, a system to avoid collisions between vessels has been considered. In this paper, as the first step of the collision avoidance system between vessels, the route prediction of large vessels by Random Forest and the identification of small vessels by k-means clustering were performed. From the verification experiment, the proposed method has been confirmed that identify each vessels and predicted each vessel's route in case of except the route get crossed with another vessel's route.

This paper is organized as follows: In section 2 gives an outline of the proposed method. Section 3 describes the verification experiment contents, experimental results, and considerations when identifying vessels using GPS data of actual small vessels using k-means clustering. Next, section 4 describes the experimental contents, experimental results, and considerations when using the Random Forest to predict the route of the ship using actual AIS data. Finally, section 5 describes the conclusions of this study and the future works.

2 SUBJECTION AREA AND ITS PROBLEM

The route prediction system for verification field work in this study takes the form of mounting a board computer such as Raspberry Pi on a vessel. It not only transmits and receives GPS information to and from other vessels and land-based base stations, but also uses its own vessel identification and linear prediction for small vessels. However, it is difficult to predict the route using machine learning in addition to the above process on a scale such as a board computer. It will be performed on the Japan Gigabit Network system [3]. Figure 1 shows a schematic diagram of communication between ships and between vessels and base stations.

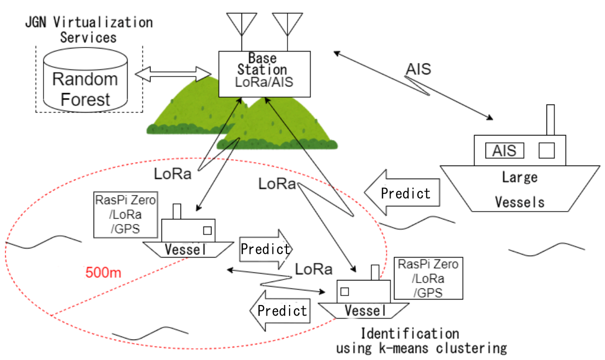

Figure 1. A Covered Area the Proposed Systems Followed.

When establishing a collision prediction system for vessels at sea, if small vessels can be assigned identification numbers such as MMSI numbers (marine mobile service identification codes) for large vessels, identification number will be assigned each one vessel and route data will be tied it. However, when an identification number is assigned to a small boat such as a fishing boat, the navigation route and fishing ground information of that boat are recorded. In this regard, the GPS module is used for fishermen who want to keep the fishing ground confidential. In this case, because of they want to keep secret these point, there was a possibility that consent could not be obtained for the installation of wireless devices, including those on board. Therefore, in order to predict the route of a small vessel, it was necessary to temporarily navigate during navigation based on GPS data to which no identification number was assigned. There is a problem that the ship must be identified and the route must be predicted from the results.

3 VERIFICATION EXPERIMENT – VESSEL IDENTIFICATION AND ROUTE PREDICTION USING MACHINE LEARNING

As mentioned above, when constructing a collision prediction system for vessels on the sea, in order to predict the route of a small vessel, the identification of the vessel that is temporarily navigating based on GPS data to which no identification number is assigned, is performed. Therefore, it was necessary to predict the route based on the results. As a first step, the following method was applied;

(1) Identification of small vessels by combining k-means clustering and GPS information.

(2) Navigation of large vessels using Random Forest.

3.1 Experiments on Identification of Small Vessels by k-Means Clustering

In the verification experiment, we used GPS data obtained near the Shimanami Kaido in

Japan while actually boarding a small vessel 2. Here, only a part of the GPS data of two small vessels are extracted and tested whether it is possible to identify each of them. The original GPS transmission is every 10 seconds, however, in this experiment does not consider time data, but treats it as only latitude and longitude data.

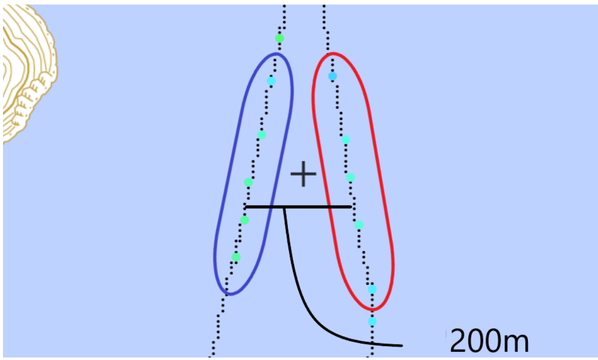

Figure 2. Target Object for k-Means Clustering.

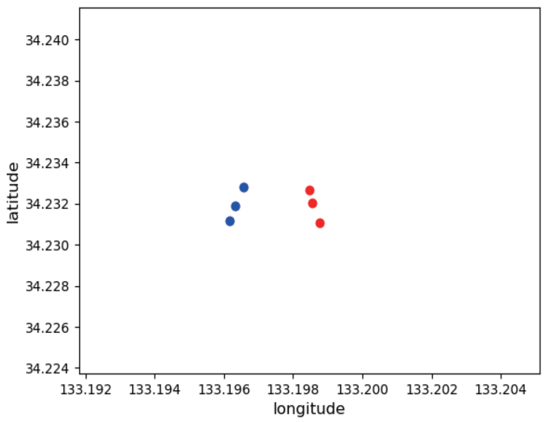

Figure 3. Result of k-Means Clustering and Identification.

Figure 3 shows the result of executing k-means clustering on the GPS information shown in fig. 2 with 2-clusters. In figure 3, it can be seen that the GPS data was correctly clustered as red and blue data for each route. Here, the GPS data of the small ship could be divided into two types of clusters using k-means clustering. However, it may that the proposed method alone will not always result in correct clustering such as the time interval for acquiring GPS data is different from that of other vessels, or if the vessels are navigating at short distances, etc. When identifying a vessel, it is necessary to consider a system that can reliably identify the vessel, taking into account various situations such as intersection and separation.

3.2 Experiments for Predicting the Route of Large Vessels using Random Forest

Compared to small vessels such as fishing boats and pleasure boats, the time and route of large vessels such as liner are predetermined. The route prediction method of large vessels by Random Forest was implemented.

(1) Daily data

(2) 8 days of data

(3) 14 days data

Random Forest was applied to the three types of data. The input data used for Random Forest is the time difference [s] between 15 positions from a certain point in time and 15 pieces of position information from a certain point in time. The output data is the position at a certain point in time. If there is a difference of 200 seconds or more in the data, it is assumed that this ship has called and is not included as training data for navigation prediction. $(\text{Total number of data} - 15)$ sets were created and used as training data.

Here, the liner between Mitsuhama Port, Matsuyama City and Nakajima Port, Matsuyama City:

(1) January 26, 2020

(2) January 19, 2020 $\sim$ 8 days of January 26, 2020

(3) January 13, 2020 $\sim$ 14 days January 26, 2020

The above three data sets were divided into learning data for the data up to 19:00 on March 26 and 256 data (19:00~23:59) on 26th as test data. For the data of 19:10~19:15 in the test data, the route was predicted based on each learning model. Of these, 00:00∼January 2020 on January 13, 2020 The number of data at 19:00 on March 26 is 28,064, which is used as training data. In this case, too, 10[%] of the training data is used to evaluate how much the learning model matches the original data.

The actual predicted results for each longitude are superimposed on the test data, as shown in figs. 6 through . The more data, the more accurate the prediction was. In the case of the data set (1), the prediction result was significantly different from the actual data, but this was due to the time (19:10∼19:15), the data of the place where the liner passed was not included much in the training data (00:00∼19:00 on March 26), and a highly accurate learning model could be obtained using Random Forest.
The data set in (2) and (3) was able to acquire data of 19:10∼19:15 over several days, and more accurate learning was performed using Random Forest. Since the model was created, it is considered that the prediction was more accurate than the data set in (1).

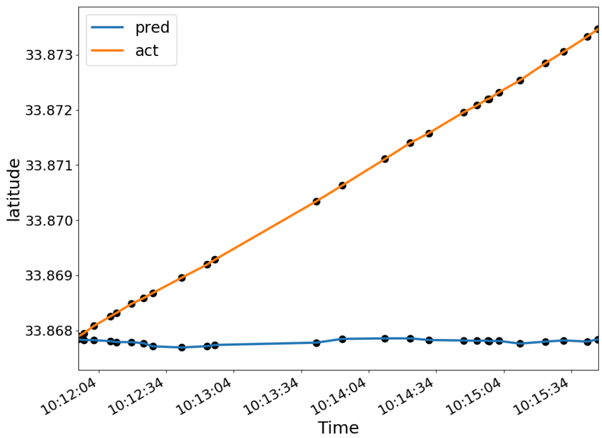

Figure 4. Prediction Result using Random Forest (Daily Data, Time vs. Latitude).

4 CONCLUSION

In this paper, as a first step toward realizing a vessel collision avoidance system between ves-

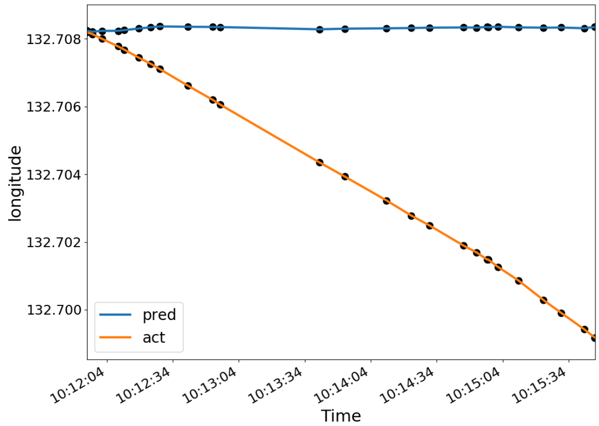

Figure 5. Prediction Result using Random Forest (Daily Data, Time vs. Longitude).

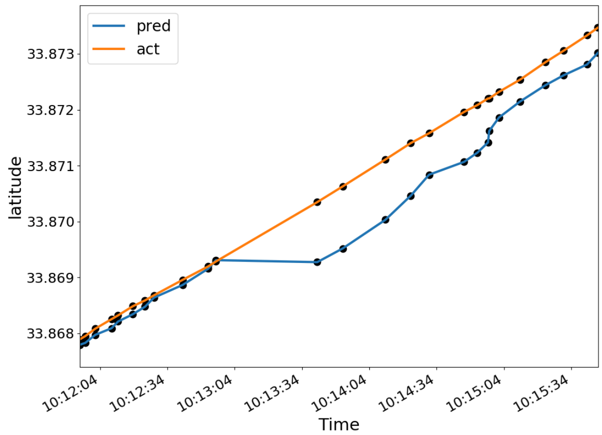

Figure 6. Prediction Result using Random Forest (8 Days of Data, Time vs. Latitude).

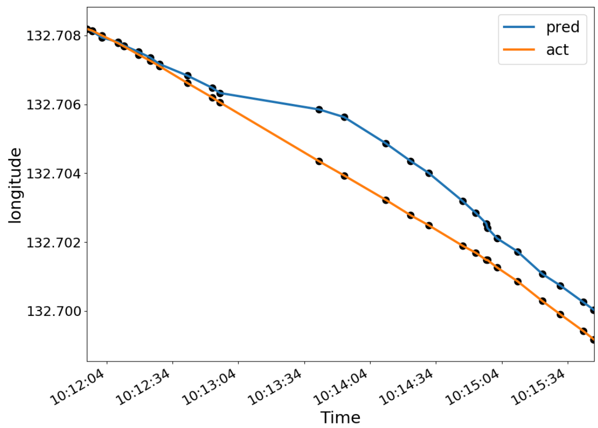

Figure 7. Prediction Result using Random Forest (8 Days of Data, Time vs. Longitude).

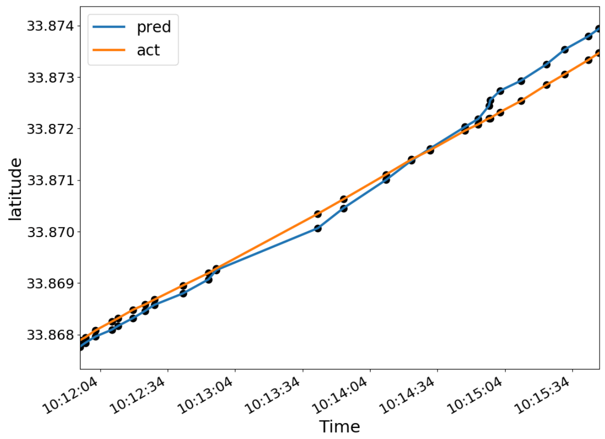

Figure 8. Prediction Result using Random Forest (14 Days of Data, Time vs. Latitude).

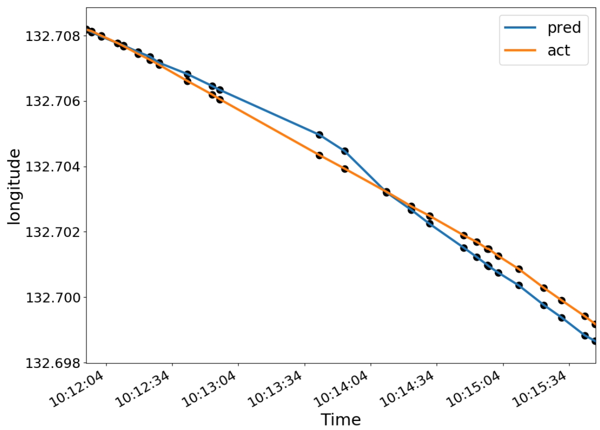

Figure 9. Prediction Result using Random Forest (14 Days of Data, Time vs. Longitude).

sels on the sea, we verified by applying two indices and two methods, identification of small vessels and prediction of the route of large vessels. ,

(1) By applying k-means clustering to GPS information, it was confirmed that small vessels can be identified under conditions where vessels do not cross each other.

(2) By applying Random Forest to large vessels, it was confirmed that the more data to be learned, the more accurate the prediction of the route.

In addition, in order to actually operate them as a vessel collision avoidance system, it is necessary to combine each function. There are many issues such as how to bring prediction results to a collision avoidance system.

ACKNOWLEDGEMENTS

This work was supported by the Strategic Information and Communications R & D Promotion Programme (SCOPE) projects from the Ministry of Internal Affairs and Communications (MIC) of 191609005.

REFERENCES

[1] Japan Transport Safety Board. Statistics of Marine Accident. `https://www.mlit.go.jp/jtsb/annualreport2019e/E11chapter5.pdf`, 2019.

[2] Chai, T., Weng, J., and Li, G. Estimation of vessel collision frequency in the Yangtze River estuary considering dynamic ship domains. J Mar Sci Technol 2019. `https://doi.org/10.1007/s00773-019-00693-6`

[3] National Institute of Information and Communications Technology (NICT). Virtualization services A nationwide fundamental virtual environment with virtual routers and virtual storages. `https://testbed.nict.go.jp/jgn/english/services/virtualization.html`, 2016.