

Automatic Identification System based Fishing Trajectory Data Preprocessing Method using Map Reduce

Jin-wan Park, Keon Myung Lee, Kwang-il Kim

Abstract: Many countries use vessel monitoring system (VMS) data to monitor their fishery activities. However, VMS data is limited in terms of distinguishing operations involving illegal fishing gear. Recently introduced automatic identification system (AIS) data is advantageous for tracking fishing ship behaviors. AIS data include various types of information about a ship, such as its state of navigation and its broadcast rate on the radio channel. We interpolate AIS trajectory data with a regular time interval and extract the ship velocity and course change data for fishing ship gear classification. To simplify and condense the data, the course change index (CCI) and ship speed index (SSI) are applied to the ship velocity and course data. The proposed mapper combines CCIs and SSIs into key words, while the proposed reducer collects fishing ship gear type values that are of the same key. By using the proposed key-value dataset from the MapReduce procedure, we can classify fishing gear type. We evaluated the performance of the proposed model by using a test dataset. The results showed that the proposed model achieved 76.2% accuracy in the classification of fishing ship trajectories against the test dataset.

Keywords : Automatic Identification System, Course Change Index, Fishing Activity, Fishing Gear Classification, MapReduce, Vessel Monitoring System

I. INTRODUCTION

Today, approximately 2.6 billion of the world's 7 billion people depend on seafood, such as fish and seaweed, as a source of protein. However, this dependence is not sustainable because the ever-increasing human population and the threat of global warming are expected to drastically reduce food supplies from the oceans and rivers in the near future [1]. The Food and Agriculture Organization (FAO) has been regulating fishery activities to protect fish from illegal, unreported, and unregulated (IUU) fishing [2]. To preserve fish stocks, most countries have declared Exclusive Economic Zones (EEZ) in accordance with the United Nations Convention on the Law of the Sea and enforced the Common Fisheries Policy to license fishing gear types such as seiners, longliners, and trawlers [3]. They also monitor fishing

activities with a vessel monitoring system (VMS) using an onboard fishing ship position transmitter to prevent fishing activities in restricted fishery zones as well as the use of illegal fishing gear. VMS data includes the movements of fishing ships spatiotemporally collected at long intervals of 30 min or 1 h and transmitted through MF/HF radio devices or satellites such as Iridium, ARGOS, and Inmarsat [4, 5]. However, it is extremely difficult to identify the fishing activity and gear type by using VMS trajectory data with such long intervals.

Identifying specific fishing ship trajectory information such as fishing activity or gear type is useful for monitoring illegal fishing activities. The competent authorities for fishing ships set a closed season in which fishing activities are temporally banned for the preservation of fish resources. Thus far, illegal fishing trajectory patterns have been detected on the basis of fishing ship positions regardless of whether they are in prohibited areas. Most countries grant fishing gear licenses for different gear types such as stow nets, drift nets, and trawlers to fishing ships. Fishing ships require a certain fishing license to operate with licensed fishing gear. Under the existing method, it is difficult to monitor whether a fishing operation is conducted with licensed fishing gear. Moreover, because each country only monitors its own fishing fleet, the VMS data of other countries' fishing ships are not received. Presently, to detect operations involving illegal fishing gear, marine patrol ships enter fishing areas and check whether the fishing gear is licensed. To monitor fishing operations effectively, a method is required to identify fishing activity with regard to fishing gear patterns.

Prior to the dissemination of automatic identification system (AIS) devices, most researchers proposed fishing and nonfishing activity classification using VMS data. A VMS fishing trajectory has a temporal resolution of 30 min or 1 h [6]. In general, fishing ship information, such as ship type, licensed gear type, and Maritime Mobile Service Identity (MMSI) number, is managed in the Vessel Register (VR) database. To distinguish the fishing activities or gear type, the ship trajectory interpolation method for fishing ships is applied to the VMS data over a long period, e.g., 1 or 2 months. Fishing gear type is extracted from the VR database by using the fishing ship's identity, e.g., the MMSI and IMO number.

Some studies have attempted to perform fishing pattern classification using ship speed data.

Revised Manuscript Received on July 22, 2019.

* Correspondence Author

Jin-wan Park, Graduate of Maritime Transportation System, Mokpo National Maritime University, Mokpo, Republic of Korea. Email: pjinwan2@gmail.com

Keon Myung Lee*, Faculty of Computer Science, Chungbuk National University, 1 Chungdae-ro Cheongju-si, Republic of Korea. Email: kmlee@cbnu.ac.kr

Kwang-il Kim, Division of Marine Industry and Maritime Police, Jeju National University, Jeju, Republic of Korea. Email: kki@jejunu.ac.kr

Marzuki analyzed the fishing activity patterns of longliners by using fishing ship speed and identified fishing and nonfishing patterns [7]. Fishing activities are divided into three types or speed ranges, namely setting (4–6 knots), hauling (2–4 knots), and others. This approach is only applicable to longliners. There are no guarantees that other fishing ship activities and speeds are consistent with the proposed speed profiles. Piet et al. classified fishing activities by using fishing operation statistics and speed profiles from VMS data. He defined three levels [8]. The first level is the number of fishing ships in a fleet. The second level is the number of hours spent fishing at sea. The third level includes fishing parameters such as fishing speed and gear characteristics. As mentioned above, VMS data is applicable to fishing activity identification owing to its long transmission interval period. However, it is difficult for this method to distinguish fishing activities involving illegal gear.

Recently, most fishing ships have been equipped with AIS devices, and they broadcast their position, course, and speed at intervals of 2 to 12 s. These facilities can analysis the ship movements and activities. VMS data has a longer interval than AIS data, and the trajectory resolution of AIS is better. AIS data is more suitable for identifying fishing activity and gear type and is thus preferable for finding illegal fishing patterns and preserve fishing resources. At present, a large amount of AIS data is stored in shore-based stations such as vessel traffic service (VTS) stations. Such data can be used in big data analysis or machine learning applications [9]. In this study, we propose a fishing gear classification method using AIS trajectory data based on the MapReduce technique.

The remainder of this paper is organized as follows. Section 2 describes the characteristics of the AIS data and the interpolation method. Section 3 introduces the proposed AIS data preprocessing method with MapReduce. Section 4 presents some experimental results obtained with the MapReduce model. Finally, Section 5 concludes the paper.

II. AUTOMATIC IDENTIFICATION SYSTEM DATA PREPARATION

A. Characteristic of AIS Data

The AIS is an automatic tracking device for broadcasting the state of a ship and displaying the information of nearby ships. A ship carrying more than 300 tonnage of cargo is equipped with AIS transponders to broadcast dynamic, static, and navigation information on the ship [10]. Table 1 lists the data items in AIS messages and their broadcast rates. The AIS message contains dynamic information on the ship's motion, including its location, speed, heading, and navigation. This dynamic information message is distributed according to the state of the ship's movement. The International Convention on the Safety of Life at Sea (SOLAS) has established minimum safety standards, according to which merchant ships with over 300 t of cargo and all passenger ships must be equipped with an AIS transponder. With regard to navigation safety, some countries have recently stipulated that fishing vessels over 10 m in length should be equipped with an AIS transponder. The AIS message consists of dynamic, quasi-static, and static information. Static information includes fixed information such as ship name, call sign, ship type, and ship

specifications. The information on the voyage is quasi-static because a voyage from the starting point to the destination cannot be switched. It includes cargo details, vessel draught, and estimated time of arrival (ETA). Dynamic information in the AIS message includes speed, direction, and GPS coordinates. A ship's static and semi-static information is transmitted in a low-flow AIS message, e.g., every 6 min. Dynamic information is broadcast at different speeds depending on the sailing state of the ship [11]. Table 1 lists the typical transmission rates and dynamic information for AIS messages.

Table-I: Transmission Interval of AIS Dynamic messages with Navigation Status

Navigation Status	Ship Speed	Transmission Interval
At anchor or moored (Non-navigation)	< 3 knots	3 minutes
	> 3 knots	10 seconds
Cruising (Navigation)	0–14 knots	10 seconds
	0–14 knots and changing course	3.3 seconds
	14–23 knots	6 seconds
	14–23 knots and changing course	2 seconds
	>23 knots	2 seconds

The stored AIS data contain all type ship such as merchant ship, fishing ship and other ship according ship type code. Originally, the AIS device has been developed for the commercial ship like cargo ship, tanker ship. Fishing ship account for only one (code 30) of 100 AIS ship type categories. Fishing ship also has code 0 of AIS ship type category because most fishing ships have class-B AIS device whose performance is worse than class-A AIS device. We examined actual ship type and AIS ship type code of 543 ships as shown in table 2. These results showed that 414 ships of 543 ships were fishing ship. This means that most fishing ships do not enter their ship type, so a default value of zero was used.

Table-II: Actual Ship Type Distribution with Zero AIS code ship

Ship Type	AIS code =0
Cargo ship	23(4.2%)
Fishing ship	414(76.2%)
Others	106(19.5%)
Total	543

B. Interpolation of AIS Trajectory

The AIS message contains various types of information about ships and their navigation status, which is broadcast at different rates on the radio channel. AIS messages can be contaminated by noise signals generated by radio wave interference due to adjacent ships' radio signals, islands, and other geographical obstructions.

An AIS base station receives broken messages with some missing, duplicated, incorrect, or delayed AIS messages. It needs to recover the broken or omitted trajectory section from an AIS trajectory sequence. Therefore, the AIS message data is first filtered to exclude unreliable, erroneous, or meaningless messages. Then, we delete the AIS messages that are close to ports or water areas where fishing is impossible and exclude AIS messages for which the speed is 0 knots or greater than a predetermined threshold, e.g., 10 knots.

After the cleaning process, the remaining AIS messages appear to be separated by random time intervals. The proposed method assumes that the AIS messages are generated at the same time intervals. Therefore, it needs to interpolate the trajectory of the AIS message sequence at a specific time interval. The reference point of time shows a sampling time in which the sampling period is Δt . To improve the efficiency of calculation, linear interpolation is generally used to interpolate adjacent positions [12]. The interpolation equations of a new position at a reference time are shown in (1). Here, R is the radius of the earth and m_k is the angle between the x-axis and the course direction θ .

$$\begin{aligned} lat_{i+\Delta t} &= \arcsin \left[\sin(lat_i) \cdot \cos \frac{v \cdot \Delta t}{R} + \frac{\cos(m_k)}{\Delta t^2} \cdot \cos(lat_i) \cdot \sin \frac{V \cdot \Delta t}{R} \right] \\ lon_{i+\Delta t} &= lon_i + \arctan \left[\frac{\frac{\sin(m_k)}{\Delta t^2} \cdot \sin \frac{V \cdot \Delta t}{R} \cdot \cos(lat_i)}{\cos \frac{v \cdot \Delta t}{R} \cdot \sin(lat_i) \cdot \sin(lat_{i+\Delta t})} \right] \end{aligned} \quad (1)$$

III. PROPOSED AIS DATA PREPROCESSING METHOD WITH MAPREDUCE

Fishing trajectory patterns are mainly affected by ship vector data. To compare the trajectory patterns, interpolation should be performed with the same interval. In this study, we mainly use the ship velocity and course change data for fishing ship gear classification. The ship course change and velocity data are continuous values. It is difficult to apply key data to such continuous values, which are mainly used for categorical-type data such as “1”, “A”, and “A1”. Therefore, these data should be converted into course change index (CCI) and ship speed index (SSI). CCI consists of 16 direction steps with a range of 22.5° , and SSI consists of 5 speed steps with a range of 2 knots. Fig. 1 shows the course and speed indices and their data ranges.

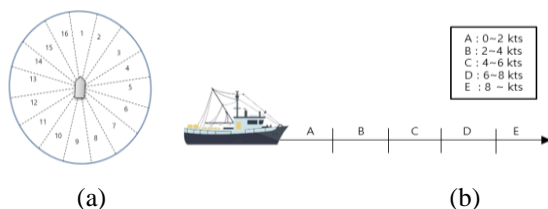


Fig 1. Ship Trajectory Index : (a) Course Change Index, (b) Ship Speed Index

With this AIS information, the fishing ship status identification can extract a fishing ship with a shipping line code of 0 or 30 because these codes are the candidate data for the fishing ship fleet. Because the AIS data does not contain information about fishing gear type, it is necessary to extract

information on how each fishing boat operates. To identify each fishing ship gear type, we extract the entire ship trajectory data of each ship. Then, a group of experts surveys the fishing area with a speed of less than 6 knots for each entire fishing trajectory and classifies the fishing gear type with labels. Figure 2 shows the labeling process for fishing gear type from the AIS database.

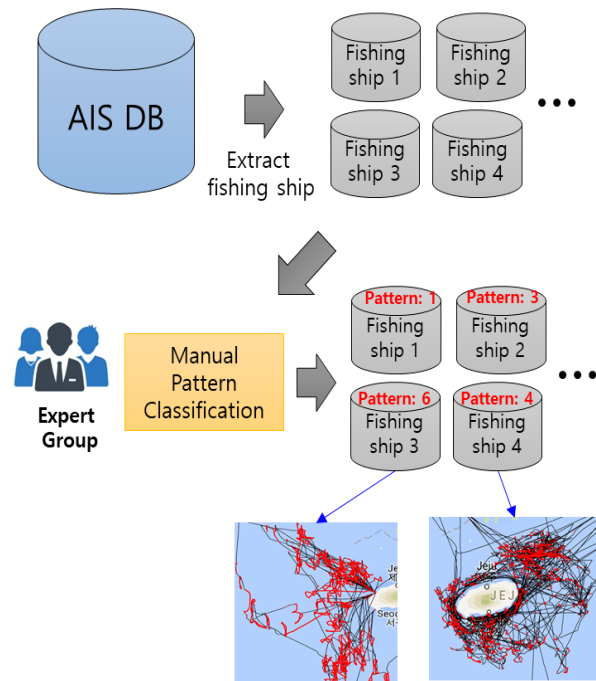


Fig 2. The Labeling Process of Fishing Gear Type from the AIS DB

In this study, MapReduce techniques, which are distributed programming techniques, are applied to efficiently process and store separated trajectory data. MapReduce is a distributed computing programming framework in a fast time for big data process, which is a sub-project framework of Hadoop developed by Google in 2004. This MapReduce framework has distributed to bigdata analyzer with variable application programming languages such as C, Java, Scala. MapReduce has two processing steps of Map and Reduce. The Map step separates input data into several pieces of data with key-value pairs. And the Reduce function computes the final value by collecting in a list for values with the same key. The produced key-value data is stored in the Hadoop Distributed File System or the NoSQL database. This technique enables faster processing and analysis of data than traditional databases in multiple general server or cluster environments without high-performance servers [13].

Fig. 3. The Proposed MapReduce Procedure and Results

To prepare data for MapReduce processing, we extract the speed and course data from each interpolated fishing ship trajectory at a regular timestamp interval. In the mapping phase, each course and speed data of a ship trajectory data can be converted into CCI and SSI.

The key value is set by combining all CCIs and SSIs into one character for a regular timestamp interval. For instance, a fishing ship trajectory data is converted into CCI string “112113” and SSI string “EEDBAA”, and these strings are combined into key string value “1E1E2D1B1B3A”. This value corresponds to a fishing gear type label. In the reducing phase, the reducer searches for a key value that is the same as the reference key from the intermediate database and stores it in a one-dimensional vector. Fig. 3 shows the proposed MapReduce procedure and the results obtained.

IV. EXPERIMENTS

To verify the performance of the proposed method, we collected AIS trajectory data over two years (2015-2016) in the southern part of the Korean peninsula. This area generated more than 10 GB of data every day. We extracted the fishing ship trajectory data which their AIS code is 0 or 30 and organized a separate data table based on the fishing ship identity. A group of experts classified fishing ship trajectories into six patterns. The Fig. 4 shows the fishing ship trajectory pattern according to their fishing gear type.

In order to validate proposed method, the fishing ship trajectory data were divided into training and testing datasets. All the training data were fed into the MapReduce procedure, which yielded a fishing gear type classification model consisting of the trajectory pattern (key) – fishing ship gear type (value). We evaluated the performance of the proposed model by using a test dataset. The results showed that the proposed model achieved 76.2% accuracy in the classification of fishing ship trajectories against the test dataset. Figure 5 shows the results of predicting the fishing gear type by comparing the actual fishing ship gear type with the results of the proposed model. The blue dotted line is where the actual gear type matches the predicted gear type, and the red dotted line does not match.

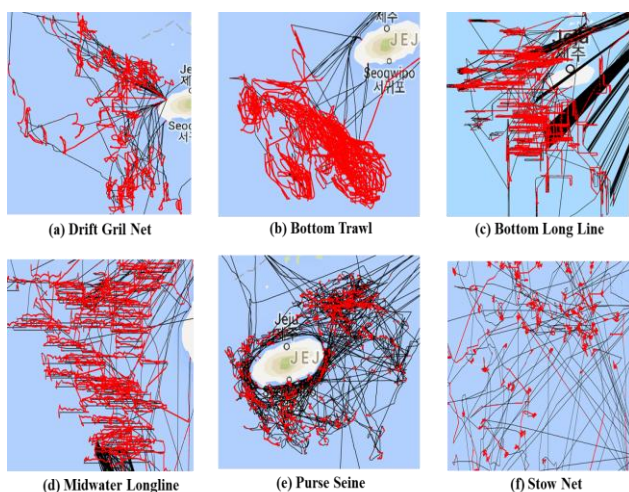


Fig.4. The Classified Fishing Trajectory Patterns

V. CONCLUSION

In terms of fishing ship trajectory classification, AIS data trajectories are more detailed than VMS data. In this study, we contributed to the several improvements for AIS data based fishing ship trajectory classification. First, we proposed AIS data preparation method by synchronizing interpolation

interval. Second, AIS data doesn't have fishing ship gear type, so we invited a group of fishery experts and they found out each fishing's gear type. Finally, because the size of AIS data is very huge we proposed the MapReduce processing method to process ship trajectory data effectively.

Using proposed key-value dataset from MapReduce procedure, we obtained 76.2% accuracy in the classification of fishing type. However, selecting the length of the AIS trajectory section and improving the accuracy of the fishing trajectory classifier remain as future research tasks.

ACKNOWLEDGMENT

This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation (NRF) of Korea (Grant no.: NRF-2017M3C4A7069432) and by a grant (NRF-2016R1A6A3 A11935806) of the National Research Foundation funded by the Korean Government.

REFERENCES

1. B. Fagan, Fish on Friday: feasting, fasting, and the discovery of the New World, Basic Books, 2008.
2. K. Cochrane, C. Young, D. Soto, T. Bahri, "Climate change implications for fisheries and aquaculture," FAO Fisheries and aquaculture technical paper 530, 2009.
3. S. R. Yoo, J. Y. Jeong, J. C. Jeong, "A Study on Spatiotemporal Distribution of Offshore Trap for the Maritime Safety" Ocean Policy Research, vol.32, no.1, 2017, pp.143-161.
4. R. Deng, C. Dichmont, D. Milton, M. Haywood, D. Vance, N. Hall, D. Die, "Can vessel monitoring system data also be used to study trawling intensity and population depletion? The example of Australia's northern prawn fishery." Canadian Journal of Fisheries and Aquatic Sciences., vol.62, no.3, 2005, pp.611-622. <https://doi.org/10.1139/f04-219>
5. C. M. Mills, S. E. Townsend, S. Jennings, P. D. Eastwood, C. A. Houghton, "Estimating high resolution trawl fishing effort from satellite-based vessel monitoring system data." ICES Journal of Marine Science., 64(2), 248-255, 2007.
6. M. I. Marzuki, "VMS data analyses and modeling for the monitoring and surveillance of Indonesian fisheries", Doctoral dissertation, Ecole nationale supérieure Mines-Télécom Atlantique, 2007.
7. M. I. Marzuki, P. Gaspar, R. Garelo, V. Kerbaol, R. Fablet, "Fishing Gear Identification From Vessel-Monitoring-System-Based Fishing Vessel Trajectories." IEEE Journal of Oceanic Engineering, vol.43, no.3, 2018, pp. 689-699.
8. G. J. Piet, F. J. Quirijns, L. Robinson, S. P. R. Greenstreet, "Potential pressure indicators for fishing, and their data requirements." ICES Journal of Marine Science., vol. 64, no.1, 2007, pp.110-121.
9. J. S. Kim, "Vessel Target Prediction Method and Dead Reckoning Position Based on SVR Seaway Model." International Journal of Fuzzy Logic and Intelligent Systems., vol.17, no.4, 2017, pp.279-288.
10. K. I. Kim, K. M. Lee, "Dynamic Programming-Based Vessel Speed Adjustment for Energy Saving and Emission Reduction." Energies., vol.11, no.5, 1273, 2018.
11. International Maritime Organization (IMO), Guidelines for the Onboard Operational Use of Shipborne Automatic Identification Systems (AIS), 2002.
12. K. I. Kim, K. M. Lee, "Deep learning-based caution area traffic prediction with automatic identification system sensor data." Sensors., vol.18, no.9: 3172, 2018.
13. J. Dean, S. Ghemawat. "MapReduce: simplified data processing on large clusters.", Communications of the ACM., vol.51, no.1, 2008, pp.107-113.

AUTHORS PROFILE



Jin-Wan Parkisa Ph.D.student inMarine Information System Engineering, Mokpo National Maritime University,Korea. And heworks at Mokpo VTS center, Western Regional Coast Guard Headquarters, as a supervisor.



Keon Myung Lee is a professor in the Department of Computer Science, ChungbukNational University, Korea. He received hisB.S., M.S., and Ph.D. degrees in computerscience from KAIST, Korea and was a postdoctoratefellow at INSA de Lyon, France. He was a visitingprofessor at the University of Colorado at Denver and a visitingscholar at Indiana University, USA. His principal research interestsare data mining, machine learning, soft computing, bigdata processing, and intelligent service systems.



Kwang-II Kimis a professor in the College of Ocean Science, Jeju National University, Korea. He received his B.S., M.S., andPh.D. in Maritime Information System Engineeringfrom Mokpo National Maritime University,Mokpo Korea, in 2005, 2010, and2012, respectively. He was a research professor at Chungbuk National University.