Journal of Engineering and Applied Sciences 13 (Special Issue 4): 3803-3808, 2018
ISSN: 1816-949X

Maritime Big Data Analysis of Ship Route Traffic Characteristics with MapReduce Processing

[1]Kwang Il Kim, [2]Keon Myung Lee and [3]Jung Sik Jeong
[1]Department of Computer Science, Chungbuk National University, 28644 Cheongju, Republic of Korea
[2]Department of International Maritime Transportation Science, Mokpo National Maritime University, 58628 Mokpo, Republic of Korea

Abstract: Previously, the assessment of ship route traffic was carried out using Automatic Identification System (AIS) data. However, the analysis of the AIS data of ship routes became problematic because of the volume of data and the difficulties associated with data access. We propose the use of data acquired via. the Port Management Information System (PORT-MIS) to overcome the aforementioned problems with data properties. Maritime big data is processed by, firstly, setting several gate lines in the ship route. These gate lines are then saved as key-value pairs. Secondly, these ship movement data based on the port facility are processed by the PORT-MIS DB mapper and reducer. Using the key-value results, hereafter, the authors conduct a variety of statistical analyses on the shipping route traffic. PORT-MIS data is more appropriate to use as maritime big data for ship route traffic than AIS data because PORT-MIS data makes it possible to prepare gate lines. A conversion algorithm for shipping route traffic is also presented. The results of this study can be used to analyze ship route traffic and carry out analyses of other big data from the ship route key-value database.

Key words: Maritime big data, port management information system, MapReduce, automatic identification system, ship route traffic analysis

INTRODUCTION

In recent year, studies related to maritime big data began to emerge. In this regard, e-Navigation is the most prominent study area in the field of maritime big data and is supported by the International Maritime Organization (IMO) and many countries. It promotes the construction of ship navigation platforms including integrated bridge systems, a maritime cloud and an enhanced wireless communication network (IMO., 2013) by many researchers. However, this attempt is mainly focused on infrastructure preparation for the future and big data processing studies about ship route traffic have rarely been reported.

In the water area within a port, ship traffic data is used for the operational control of ship traffic and to assess the risk of traffic congestion. The development of maritime information technology has resulted in ship traffic data being gathered by port authorities and Vessel Traffic Service (VTS) centers for more than 10 years. These data are sufficient to utilize in ship traffic analysis and enable a statistical approach to be used to process ship traffic big data generated by ships in port (Kim *et al*., 2013a, b).

To date, the assessment of ship traffic characteristics on route was performed by analyzing Automatic Identification System (AIS) data which consist of both static and dynamic data of ships. This data is appropriate for ship traffic analysis in a port water area. However, the amount of AIS data generated per day is excessively large and access to the AIS DB is limited by the security policy (Kim *et al*., 2013).

Another type of data that is available to evaluate the characteristics of ship route traffic is Port Information Management System (PORT-MIS) data which comprise recorded ship movements and port management and logistics statistics. In addition, these data are open and are available to everyone who needs access to raw data and statistical information, for example the ship entry history and sum of the specific traffic volume. However, PORT-MIS data are based on the use of port facilities and anchorage, thus its applicability to ship route traffic analysis was limited (IMO., 2013).

Maritime big data for ship route traffic characteristics

Automatic identification system trajectory data: The Vessel Traffic Service (VTS) center is a shore-based

Corresponding Author: Keon Myung Lee, Department of International Maritime Transportation Science, Mokpo National Maritime University, 58628 Mokpo, Republic of Korea

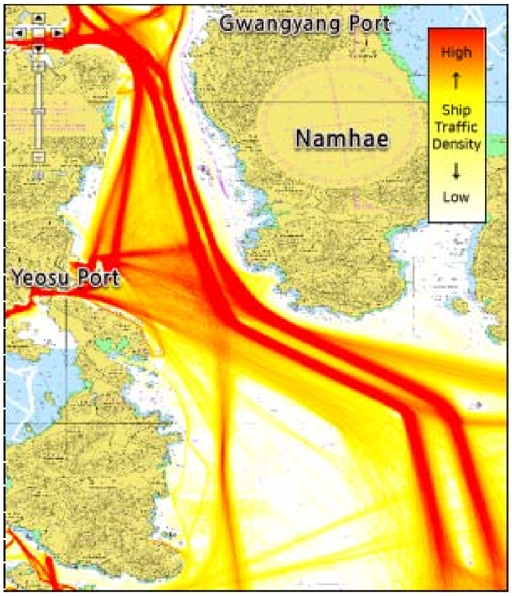

Fig. 1: The example of AIS ship trajectory distribution

station that automatically collects information about ship movements by using an AIS device. This AIS device transmits messages containing dynamic and static information in accordance with the update rate. Static data including the name, type, length, width and draft information of a ship is transmitted every 6 min or upon request from a shore station. The dynamic data consist of the position, course, speed, heading direction, rate of turn, etc. of a ship and its transition period is 3~180 seconds (IMO., 2002).

According to a related study on the analysis of AIS data, these data are mainly used in ship traffic analysis, and for the assessment and estimation of ship collision risk (Jeong *et al*., 2012). Figure 1 represents the AIS ship trajectory distribution.

Port management information data: All ships entering or departing from a port facility are required to report the event time and target facility information to the port authority or Vessel Traffic Service (VTS) Center. In the case of South Korea, this information is entered into PORT-MIS and stored in the database. PORT-MIS is an integrated information system in which the movement of ships and cargo in and out of ports is recorded for the purpose of collecting port and facility fees. The PORT-MIS database can help users who are interested in ship and cargo operations to assess real-time and statistical information about ship entries/departures, port facilities and cargo and so on. The functions of PORT-MIS are explained in Table 1.

Table 1: Operational functions of the PORT-MIS

Types	Operational functions
Ship control	Ship arrival/departure and berth control, pilot/tugboat control
Cargo control	General cargo, container cargo, transshipment cargo, dangerous goods
Billing	Levy data generation, levy calculation, bill publication, billing error/arrear management
Statistics	Ocean-going ship, coastwise ship, export/import cargo, domestic cargo

However, this system does not include ship route traffic information such as the path of the ship route and the passing time. Information that could possibly be used for estimating ship route traffic is the time of entry, time of departure and movements between port facilities only. Therefore, it is necessary to develop an algorithm for extracting ship route traffic data from the PORT-MIS database.

Application of port big data to ship traffic analysis: The AIS data is convenient for identifying the actual movement pattern of the ship and it is applicable to utilization for risk assessment of ship collision, ship traffic density analysis and real time monitoring (Kim *et al*., 2017). However, since, this data contains vessels and personal ship identification information with security policy, it is difficult to obtain data for a long time. In addition, data out of the ship route are also collected and the data size is stored in excess of 10 GB and also it takes a long time to analyze and process the data.

In contrast as noted in this study, PORT-MIS data are of a simple type, namely the recorded time of arrival or departure of a ship based on the port facility. The volume of PORT-MIS data is less 10 MB per day and most importantly, this data has been made public which enables anyone to access and utilize the data.

Clearly, PORT-MIS data is more suitable than AIS data in terms of analyzing ship route characteristics using maritime big data. The former of these two types of data allows faster processing and unlimited data access. However, as this study mainly focuses on ship route traffic, it is necessary to convert port facility arrival/departure records to ship route traffic. This study proposes a method of inputting ship movement data on each route gate line using PORT-MIS data.

MATERIALS AND METHODS

Preparation of the gate line database

Preparation of gate line information: In general, a ship should navigate along the designated ship route to the destination by following the shortest route. Calculation of the ship route traffic data using PORT-MIS data requires route gate lines to be designated according to the ship

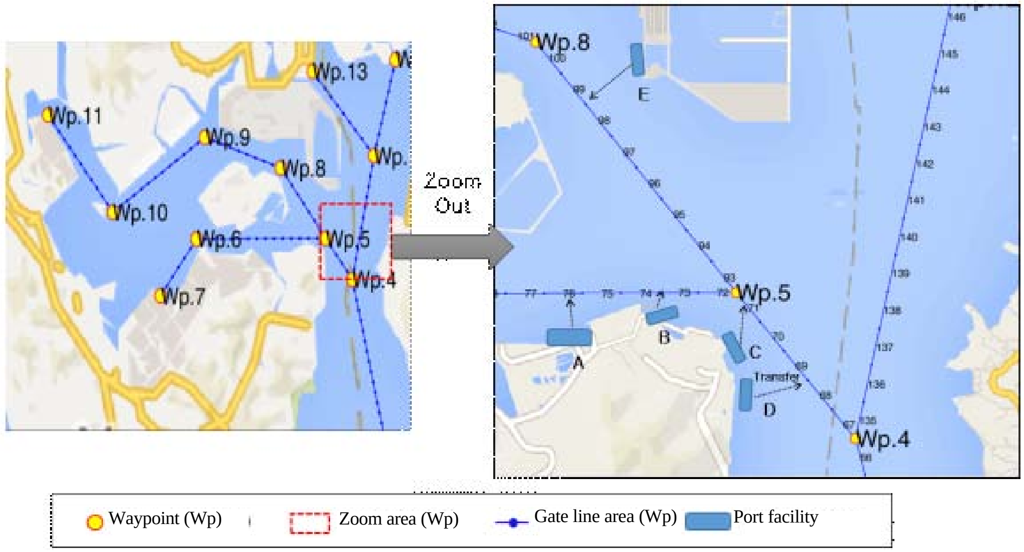

Fig. 2: Waypoint nodes, gate lines and port facility in yeosu port area

traffic pattern. For this reserach, information about Waypoints (Wp), according to which the course followed by a ship is changed to allow it to enter another connected ship route is prepared based on the AIS ship trajectory pattern as seen in Fig. 1. Gate lines are created by dividing a ship route connected by adjacent waypoints into unit lengths. Subsequently, each gate line is assigned a gate number in descending order from the inside toward the outside direction in a port. Figure 2 represents the results of the gate lines that were created in this way.

When processing the port facility data, the position of the port facility (A~E) is shifted to the adjacent gate line. Thus, ship movement data in a certain port facility is transferred to the adjacent route gate line.

Calculation of the transit time of a ship requires information of the average passing speed on each gate line. Hence, this study used a table containing the average passing speed results consisting of the ship type in the navigational section for 3 months. All ships navigating to the target area are limited to a speed of 12 knots or less, especially in the limited section of the harbor (gatelines 66~166) the speed of ships is restricted to 10 knots. Table 2 represents the average ship speed in the target area.

Preprocessing of PORT-MIS data: The use of PORT-MIS data to determine the path of a ship on the route requires both the arrival and departure facility. In the case of arrival, the No.1 gate which means the port entrance gate, is set as the start gate and the arrival gate is set as the end gate. In the case of departure, the departure gate is set as the start gate and the No.1 gate is set as the end gate. In

Table 2: Average ship speed in Yeosu Port

Variables	Inner harbor area (Gate line 66~166) (Knots)	Outer harbor area (Gate line 1~65) (Knots)
Cargo ship	9	11.0
Container ship	10	13.0
Tanker/LNG	8.5	10.5
Towing ship	4	6.0
Other ship	6	8.5

the case of movement, the start gate is obtained from the previous unberthing event records and the arrival gate is set as the end gate. These start and end gates, including the berth/unberth time are used as input data for mapreduce.

Ship route path analysis using mapreduce: MapReduce is a software framework developed by Google in 2004. It uses distributed computing to perform calculations on large data sets using clusters of computers (Dean and Ghemawat, 2010). Mapreduce is a programming model for processing large data sets with various programming languages. The map function is used to process raw data and to generate a set of key-value pairs. The reduce function merges all intermediate values in accordance with the same key (Kang *et al.*, 2015).

The input data, the preparation of which is described in this section, consists of the start and end gate numbers, passing time and the ship identification data. Firstly, the proposed PORT-MIS mapper imports the input data and verifies the consistency of the data. Secondly, the path connecting the gates along the ship route is calculated by using Dijkstra's algorithm (Cho, 2009) which is a shortest-path algorithm. Its path weight w is set by using the distance between gates. Then, the time at which the ship passes the gates is added to the initial time to determine the traveling time to the gate by dividing the gate length by the gate speed. Thirdly, two gate numbers are compared to obtain the sailing direction because the gate numbers are listed in descending order from the inside in the direction of the outside of the port. Finally, data consisting of information about the ship such as the ship type and size are connected to information related to the ship in the database. After processing by the algorithm, key contains [gate number, ship type, ship size, movement of PORT-MIS, sailing direction] and Value contains the vector of the passing time. The pseudocode of the PORT-MIS Mapper is shown in Algrothim 1.

Algrothim 1; The pseudocode of the port-mis db mapper Procedure mapper (s, e, mmsi, mp):

```
Procedure Mapper (s, e, mmsi, mp)
// Input data: start gate s, end gate e, ship's id mmsi, movement of PORT-MIS mp
// Output data: key, value
// Finding Shortest Gate Path set S
Initialize-Source G = (V, E)
S←∅, Q←V [G]
While Q≠ do
    Q'←V-Q
    u←min{d (w)| w ∈ Q'
    S←SU {u}
    remove u from Q
    for i = 1 to end (u) do
        if length (u (i))+distance (u (i), v) <
          length (u (i))//distance (u,v) get distance  from current node u and v
          length (u (i))←length (u (i))+distance (u (i),v)
// Decision of Sailing Direction sd
if S (i) < S(i+1)
    sd←0//inbound case
else
    sd←1//outbound case
// Connection DB of ship type and ship size
st←shiptype (mmsi)  // get ship type number (1~9) using ship's id mmsi
ss←shipsize (mmsi)  // get ship size number (1~6) using ship's id mmsi
// Calculation of Key and Value (Passing Time t (S (i)) )
for  i = 1  to  end (S) do
    t←t+length (S (i))/spd (S (i))//spd (S (i)) get speed at gate line S (i)
    value←t
    key←[S (i), st, ss, mp, sd ]
```

The PORT-MIS DB reducer collects the answers to all the sub-mapper results and combines them according to the same key domain. Algrothim 2 displays the pseudocode of the Port-MIS DB reducer.

Algrothim 2; The pseudocode of the port-mis db reducer:

```
Procedure mapper (s, e, mmsi, mp)
// Input data: key, value, Iterator set Q, OutKeyValueStore
// Output data: key, value
int result = 0
While Q≠∅ do
    value←load (Q (1))
    result [n+1]←value
    Q←Q-Q (1)
AddKeyValue (OutKeyValueStore, key, result)//save result to the Key-Value store for MapReduce
```

RESULTS AND DISCUSSION

Experiment

Overview of experiment: This experiment was conducted by utilizing the data of ships entering and departing from the Yeosu port from 1998 to 2015. There are 230,000 port facility usage records per year for the target area. The input data for this experiment were prepared by preprocessing the PORT-MIS data. The preprocessed input data were separated into key-value pairs by the PORT-MIS DB mapper and the PORT-MIS DB reducer was then used to save the time data in vector form for the same key. The procedure of this experiment is presented in Fig. 3.

Application of the route statistical analysis: A statistical analysis of the ship route was carried out on the large PORT-MIS database. The value data needs to convert other time units such as the month and day of the week in accordance with the purpose. As a result of the analysis in Fig. 4, three groups of ship route traffic characteristics were obtained (Fig. 5).

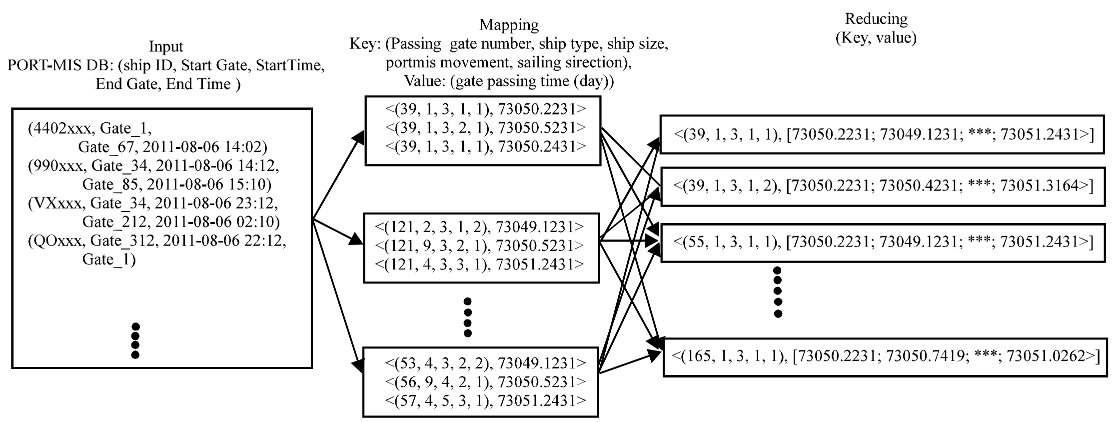

Fig. 3: Procedure of the MapReduce experiment

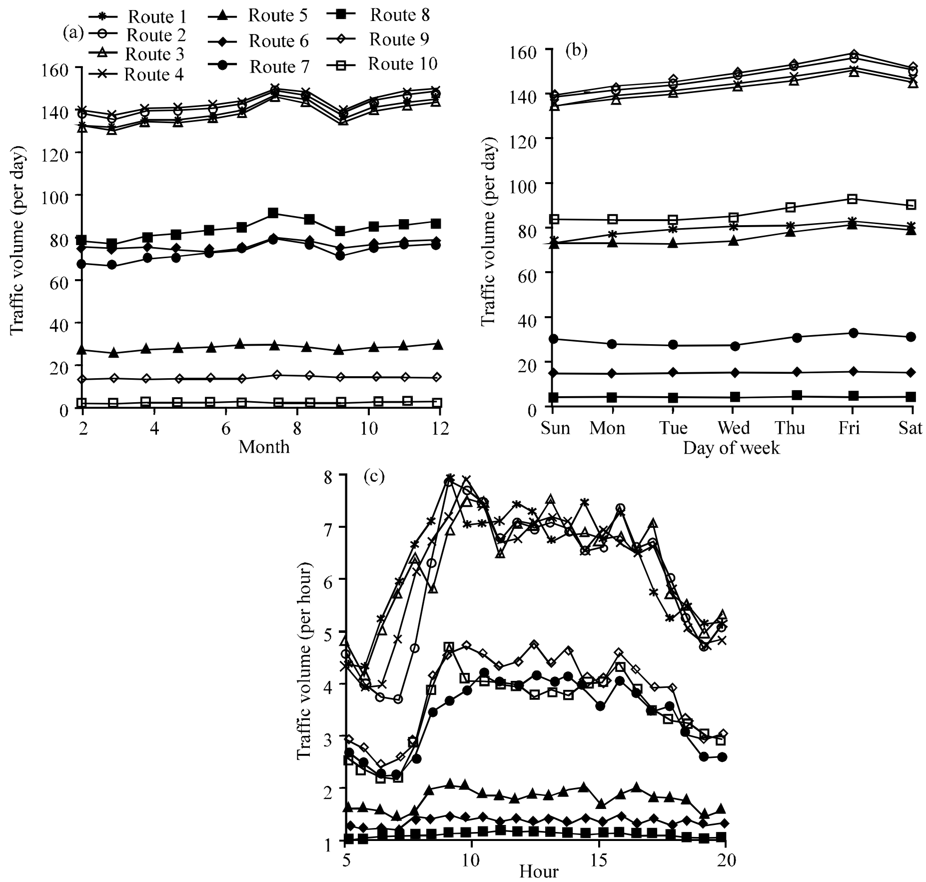

Fig. 4: Results of the route statistical analysis in monthly (left), day of weekly (middle), hourly (right)

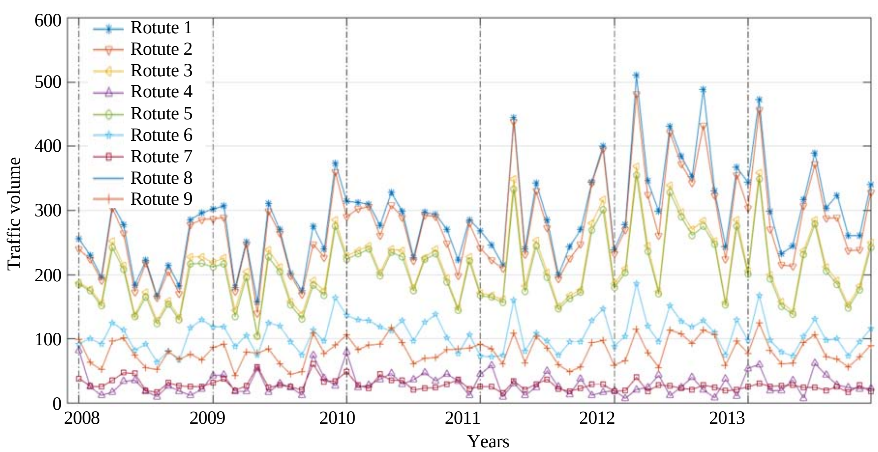

Fig. 5: Time series analysis of ship traffic volume grouped by route

The first group (route 1-5) refers to route entrance traffic. The traffic volume is larger than that of the other groups with 130~150 ships entering per day and difference between the daytime and nighttime traffic is 4-5 traffic volumes per hour. The second group (route 6-8) indicates route separation traffic. The traffic volume is 60 to 90 traffic volumes per day and the difference between the day and nighttime traffic is 2-3 traffic volumes per hour. The third group (route 6-8) includes approach route traffic near the port facility. The traffic volume is 5-30 traffic volumes per day and the difference between the day and nighttime traffic is <1 traffic volume per hour.

Figure 5 shows the trend of the ship traffic volume for each route from 2008-2014. Routes 1-5 indicate the traffic pattern outside the harbor limit and routes 6-9 represent the traffic pattern within the harbor. All vessels enter and leave the port via route 1 thus, this route has the highest ship traffic volume. The deviation of ship traffic trends is high in route 1-5 because more than half of the vessels on this route are proceeding to waiting anchorage for bunkering or to receive supplements. On the other hand, the traffic volume on route 6-9 is constant at an average of 100 ships per day.

CONCLUSION

Even though PORT-MIS data is an important source of data for big data analysis in the maritime field, it was not attempted to use these data for the analysis of ship route traffic. That is because the ship movement data in PORT-MIS is collected based on the port facility.

This study led to the proposal of methods to convert ship movement data among port facilities to ship route traffic data using PORT-MIS DB MapReduce. This was achieved by firstly setting the gate lines along the ship route to save key-value pairs. Then, the ship arrival and departure data based on the port facility were extracted from the PORT-MIS DB and converted to ship route traffic data using the PORT-MIS DB mapper and reducer.

RECOMMENDATIONS

Future plans involve using the key-value results to conduct a variety of statistical analysis on ship route traffic.

ACKNOWLEDGEMENTS

This research was supported by basic science research program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01061062) and by Next-Generation Information Computing Development Program through the National Research Foundation (NRF) of Korea (Grant no.: NRF-2017M3C4A7069432).

REFERENCES

Cho, M.G., 2009. A dynamic shortest path finding with forecasting result of traffic flow. J. Korea Inst. Inf. Commun. Eng., 13: 988-995.

Dean, J. and S. Ghemawat, 2010. MapReduce: A flexible data processing tool. Commun. ACM., 53: 72-77.

IMO., 2002. Guidelines for the onboard operational use of shipborne Automatic Identification Systems (AIS). International Maritime Organization, London, England.

IMO., 2013. Development of draft Software Quality Assurance (SOA) guidelines for E-navigation. International Maritime Organization, London, UK.

Jeong, J.S., G.K. Park and K.I. Kim, 2012. Risk assessment model of maritime traffic in time-variant CPA environments in waterway. J. Adv. Comput. Intell. Inf., 16: 866-873.

Kang, S.J., S.Y. Lee and K.M. Lee, 2015. Performance comparison of OpenMP, MPI and mapreduce in practical problems. Adv. Multimedia, 2015: 1-9.

Kim, J.S., J.S. Jeong and G.K. Park, 2013a. Prediction table for marine traffic for vessel traffic service based on cognitive work analysis. Intl. J. Fuzzy Logic Intell. Syst., 13: 315-323.

Kim, K.I., J.S. Jeong and B.G. Lee, 2017. Study on the analysis of near-miss ship collisions using logistic regression. J. adv. Comput. Intell. Inf., 21: 467-473.

Kim, K.I., J.S. Jeong and G.K. Park, 2013b. Assessment of external force acting on ship using big data in maritime traffic. J. Korean Inst. Intell. Syst., 23: 379-384.