International Journal of Innovative Computing, Information and Control
Volume 17, Number 2, April 2021
ICIC International ©2021 ISSN 1349-4198
pp. 725–734

PREDICTION OF VESSEL ARRIVAL TIME USING AUTO IDENTIFICATION SYSTEM DATA

Hyeonho Kwun[1] and Hyerim Bae[2,*]

[1]Department of Industrial Engineering
[2]Major in Industrial Data Science and Engineering, Department of Industrial Engineering
Pusan National University
2, Busandaehak-ro 63beon-gil, Geumjeong-gu, Busan 46241, Korea
khh3230@pusan.ac.kr; *Corresponding author: hrbae@pusan.ac.kr

Received October 2020; revised February 2021

Abstract. *The purpose of this study is to accurately predict the ship's arrival time using a data-driven methodology. In previous studies, studies have calculated the arrival time based on distance after generating a route using ship data to predict the ship's arrival time and a study that applied trajectory data and weather forecast data to machine learning methodology. These studies have a limitation in not considering the maritime situation information in predicting the arrival time. This study proposes a data-driven methodology that uses vessel trajectory data to consider different maritime situations depending on the location. In this approach, AIS data are preprocessed using the trajectory mining technique and based on this, a pathfinding algorithm is applied and arrival time is estimated. As a result of comparing this method to the actual ship data of a terminal in Busan, Korea, the average error was improved by about 20% compared to the benchmarking methodology. This information helps improve port monitoring and berth planning and is expected to increase port operational efficiency.*
Keywords: AIS data, Trajectory mining, Pathfinding, ETA

1. **Introduction.** About 90% of world trade is carried out by the international shipping industry, so the importance of ocean transportation is growing, and expectations for growth in the shipping industry are rising [1]. In such an environment, the need increases for efficient transportation and economical logistics. Accordingly, there is a growing need for research on safer and more efficient port operations, such as terminal efficiency, port safety and security [2]. In particular, the efficiency of port operations is recognized as an essential factor for securing the port's cargo volume and for smooth performance of the national economy.

Today, port-related data are actively collected through vessel reporting systems on port efficiency. In particular, the development of vessel reporting technology and remote sensing systems provides high-quality vessel data, including time and geographic information. Among vessel reporting systems, the Auto Identification System (AIS) designed for collision prevention is essential in marine monitoring and safety [3]. AIS provides not only static information (e.g., name of the vessel, length, type) but also dynamic information (position, speed, heading, etc.). It has been suggested that such AIS data will be useful not only for collision prevention but also for predicting various indicators related to vessel operations [4,5].

The purpose of this paper is to accurately estimate a vessel's arrival time at the port for efficient port operational logistics. A vessel that is not punctual causes congestion in the terminal reducing terminal operations' efficiency and the aftermath lasts up to three days

DOI: 10.24507/ijicic.17.02.725

[6]. In addition, inaccurate ETA causes excessive variability in the berth plan, causing operational inefficiency of the entire port. Also, the vessel's inaccurate arrival time can waste a lot of time for those engaged in the tugboat service. Therefore, predicting the vessel's arrival time more accurately is a problem that must be improved to increase the port's overall operational efficiency.

Vessels are heavily affected by geographical requirements on the sea, regulations on the operation of vessels, and the conditions of the sea such as waves, tide, and weather. In areas with shallow water or obstacles such as islands, the speed of vessels is restricted to prevent accidents. Furthermore, environmental factors on the sea, which are wave height and strong winds, affect vessels' resistance and movement. Due to these various factors, the characteristics of maritime traffic vary depending on each location [7,8]. The AIS data, which is the trajectory data, is a record of sailing taking account of the above factors. Thus, to consider these factors, this paper proposes a data-driven methodology using AIS data.

This paper proposes a new method for predicting the time a vessel arrives at its destination by using vessel track information. The proposed method derives an optimal path based on past tracks and uses this predicted path to more accurately predict the arrival time. To solve this problem, a new methodology using track data, which is a record of voyages, is needed. In this paper, to verify the method, performance is evaluated by comparing the actual operating time of the vessel with the predicted arrival time in connection with the port's operational data.

This paper is organized as follows. Section 2 briefly describes the background, and Section 3 discusses the overall method of the ETA framework. Section 4 applies the algorithm to actual data and evaluates it through experiments. Finally, Section 5 draws conclusions and looks ahead to future work.

2. Background.

In this paper, an algorithm that shows more efficient computational performance is developed to find the optimal path from past vessel tracking data, and an accurate ETA is calculated based on the path.

2.1. Trajectory mining.

Trajectory, which generally means the history of spatial movement of an object, is a trace created by an object moving in geographic space. At this time, one trace is expressed as the connection of multiple points that the vessel has passed in a time sequence. Each point is composed of spatial coordinates and time, and also includes metadata such as speed, direction, and object properties. To use trajectory data, trajectory preprocessing must proceed through stages such as noise filtering, segmentation, stay point detection, and map matching. The goal of noise filtering is to remove error points that may occur due to poor signals from positioning systems such as a GPS. Trajectory segmentation is grouping by time intervals, trajectory shapes, or semantics for further processes, such as clustering and classification. Stay point detection identifies a location where a moving object stays for a particular time or longer within a specific distance threshold [9].

2.2. Pathfinding.

Pathfinding is a process of finding an optimal path when moving from one point to another without colliding with an obstacle in the path [10]. Dijkstra [11] and A* [12] algorithms are examples of graph-based pathfinding algorithms. The A* algorithm, an extension of Dijkstra's algorithm, uses heuristics to improve temporal performance and to increase efficiency. This algorithm aims to find a path that minimizes the estimated total cost by using an evaluation cost function, $f(x) = g(x)+h(x)$ of node x. Cost function $f(x)$ is divided into two parts: path cost function $g(x)$ and future path cost function $h(x)$. The cost function expressed as $g(x)$ is the actual path cost from the start node to node x, and $h(x)$ is a heuristic function that estimates the cost of the cheapest path from node x to the destination node. The calculation of $h(x)$ depends on different situations. In geographic problems, straight distance from the node x to the destination node is often used.

2.3. **Estimate time of arrival.** Predicting the arrival time is an important task in a vessel's voyage. From a traveler's point of view, there are advantages: selecting a route to travel and securing safety. In logistics, there are advantages such as increasing the reliability of delivery and reducing delivery costs. In the port field, there are various studies on predicting the arrival time of a ship using the seq2seq approach [13], on predicting the arrival time of a container ship by applying location data and weather forecasts to machine learning [14], on predicting the arrival time of a ship's liner and tramp [15] and so on.

These previous studies are data-based arrival time prediction methodologies using AIS data. AIS implicitly contains the assumptions of vessels' movement in the past. Therefore, previous studies were in considerations of characteristics of the sea while predicting the arrival time. However, since these studies only used AIS as input data, it is difficult to say that different maritime traffic characteristics were taken care of for each location on the sea. Therefore, this study pre-processes AIS data by dividing it into a spatial grid and considers different maritime characteristics depending on each point's location.

3. **Methodology of ETA Prediction.** Before we predict the arrival time, it is common to find an appropriate path between the departure and the arrival points, and calculate the travel time by considering the distance along the path and the speed of the vessel. To use this method, it is necessary to predict the path. The detailed path prediction procedure and the ETA calculation are schematically illustrated in Figure 1.

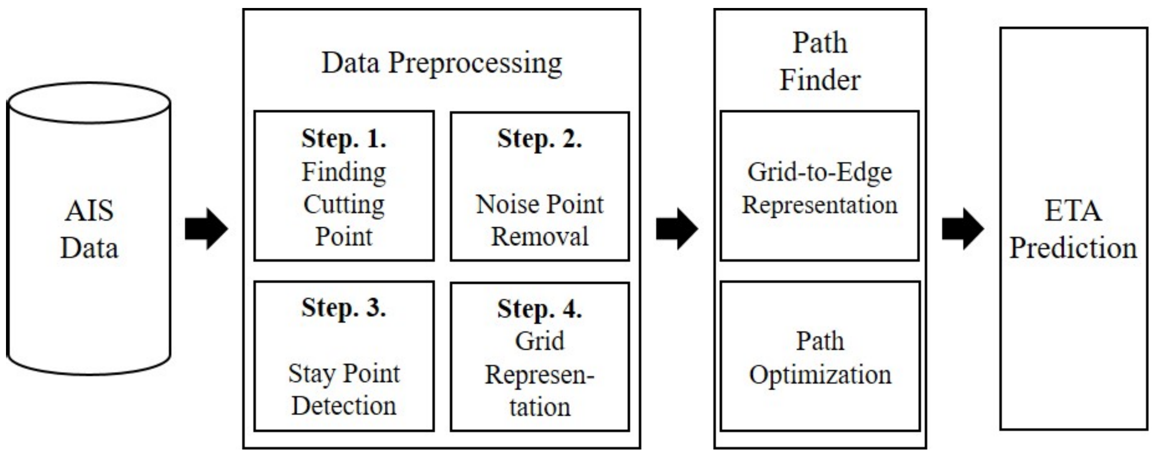


FIGURE 1. Flow chart for the prediction of vessel times of arrival in port areas

3.1. **Data preprocessing.** Before predicting the path, the AIS data, including the vessel's position data, are preprocessed. The preprocessing method used in this paper is based on the method used in the trajectory mining technique. The path is the course that the object traveled between two points. Therefore, the vessel's AIS data are classified as moving data using trajectory preprocessing. Trajectory preprocessing proceeds through four steps, as follows.

Step 1. Finding Cutting Points: Classify the trajectory using the time differences between data.

Step 2. Removing Noise Points: Find and remove noise points using a distance threshold within an appropriate time difference between trajectory data.

Step 3. Detecting Stay Points: Classify the trajectory by finding the points where the vessel was staying in one place.

Step 4. Grid Representation: The trajectory is expressed in grid format.

In Figure 2, p_i is a cutting point since the time difference between p_i and p_{i+1} exceeds the threshold, ΔT_c. In this case, p_i and p_{i+1} are considered to be on different trajectories.

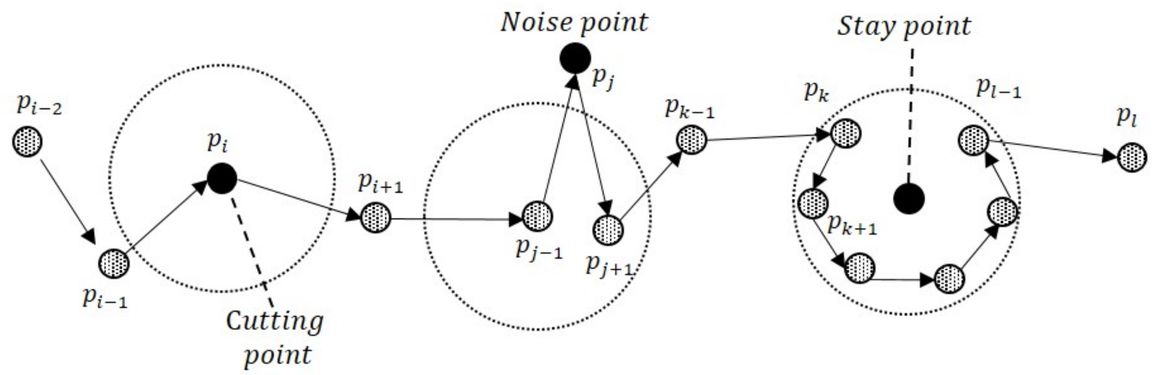


Figure 2. Cutting point, noise point, and stay point in a trajectory

Point p_j in Figure 2 is a noise point (since it has a low possibility of being on the path), which was caused by a sensor error or incorrect position signal reception from the positioning system. Therefore, noise filtering is required, especially in preprocessing trajectory data. In this paper, noise points are simply removed using the heuristics-based outlier detection noise filtering method. The moving speed of each point is calculated based on the time interval and distance between two consecutive points. The point where the velocity was higher than the velocity threshold value, such as p_j in Figure 2, is removed.

A stay point in Figure 2 is the midpoint of the rightmost circle, where the moving object stayed at a specific place for a certain period of time. When the distance between consecutive points from p_k to p_{l-1} in trajectory is smaller than distance threshold ΔD, and the time difference between p_k and p_{l-1} is greater than threshold ΔT, a stay point is generated. The stay point is the centroid of p_k to p_{l-1} in Figure 2. After we found the stay point, we had two separate trajectories: one before the stay point and the other after the stay point.

The sub-trajectories classified through the above three steps are expressed as a grid in order to apply a graph-based pathfinding algorithm. All consecutive trajectory data in geographic space are represented by a grid. The grid is arranged in rows and columns and contains metadata indicating position, time, speed, and direction angle. The trajectory on the grid in Figure 3 shows that one of the trajectories classified through the previous three processes, Tr^k, is placed in the same $[N \times M]$ grid cell. The reduced trajectory is a representation of points placed in one grid cell as one point, grp_j^k, where grp_j^k is the data of grid j on the k-th trajectory. Time value t of grp_j^k is expressed as $grp_j^k.t$.

3.2. **Pathfinding algorithm.** The main purpose of pathfinding is to find the path with the least cost among the paths between two points on the map. Before actual pathfinding, the cost between consecutive grids is calculated based on graph theory. First, convert $grp_j^k.t$ and $grp_l^k.t$ of the k-th trajectory into $ed_{j,l}^k.td$, like the edge matrix in Figure 3. The notation $ed_{j,l}^k.td$ represents the time difference between j, l grids. This becomes the edge of the graph connecting each grid. Edge data are created for all k trajectories preprocessed

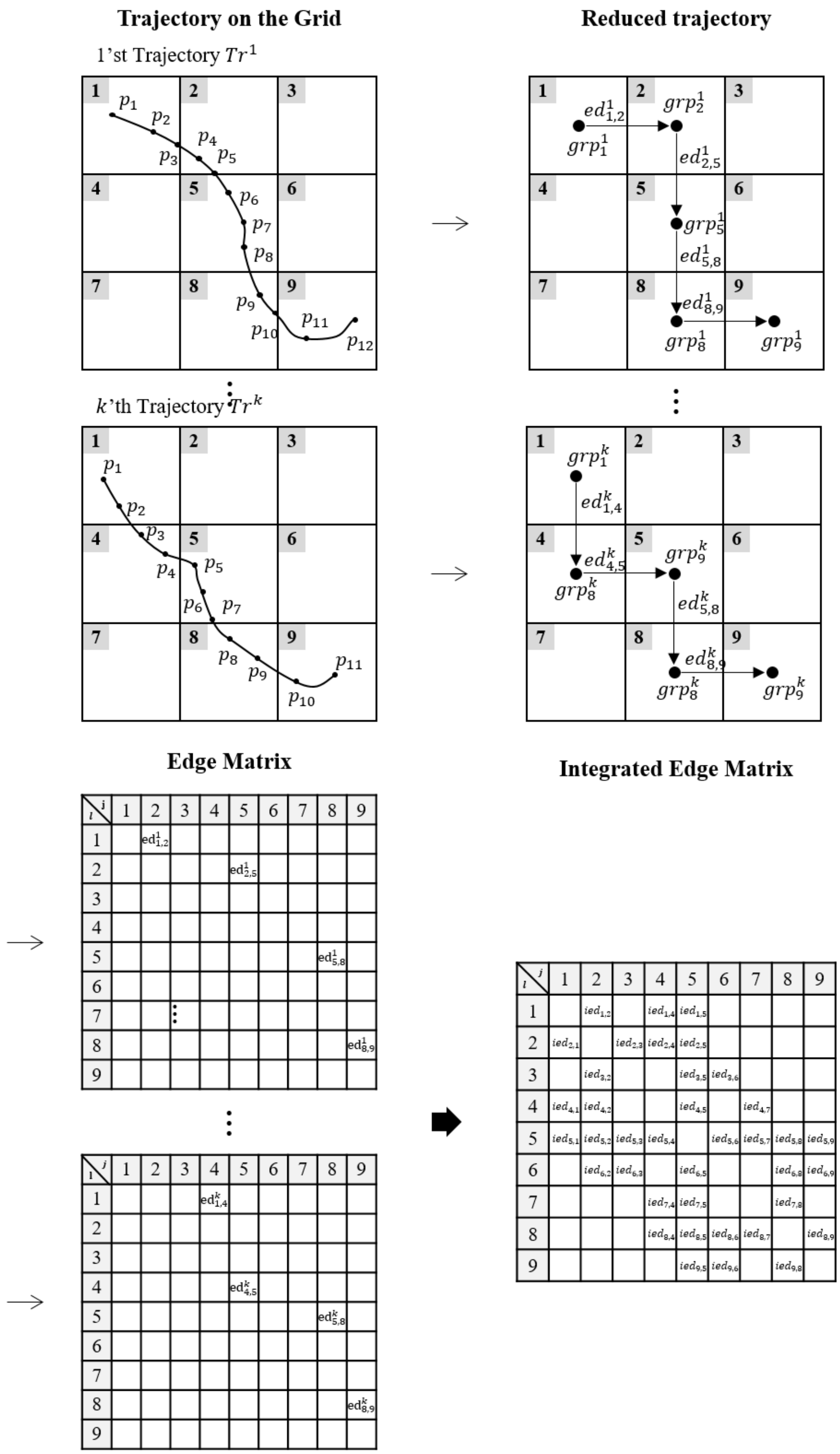


FIGURE 3. Trajectory to graph representation process

per Section 3.1. After that, the integrated edge $ied_{j,l}.tm$ is the average value of $ed^k_{j,l}.td$, like the integrated edge matrix in Figure 3. And, the $ied_{j,l}.den$ edge is created using the non-null number of $ed^k_{j,l}.td$ corresponding to the same j, l edge. The integrated edge created through this process represents connections from one grid to all grids except for the grid corresponding to an obstacle among the adjacent eight grids.

In this paper, we propose a weight optimization A* algorithm that is similar to the actual path by changing the existing A* algorithm, which is a pathfinding algorithm. The weight optimization A* algorithm has basically the same structure of the A* algorithm, but gives weight w_g, w_d to $g(x)$, $h(x)$ in evaluation function $f(x) = g(x)+h(x)$. And then, $f_w(x) = w_g \times g(x) + w_h \times h(x)$ is used as an evaluation function, and a prediction path is generated using it. In this paper, since the purpose of creating a path is to find a path that a normal vessel travels, weights are given to heuristic function $h(x)$ and actual path function $g(x)$ to make an evaluation function that fits the purpose. At this time, the grid point created in this paper is expressed as node x. Taking the first reduced trajectory, Tr^1, in Figure 3 as an example, grp^1_j becomes the nodes. The actual path function, $g(x)$, means the cost between node x and starting node x_0, and in this paper, the combination of the integrated edge, $ied_{x,x_0}.tm, ied_{x,x_0}.den$ is used. Heuristic function $h(x)$ uses the great circle distance, the shortest distance connecting two points on the surface of the sphere, as the cost between node x and the destination node. This can be expressed as follows:

$$\begin{aligned} f_w(x) &= w_g \times g(x) + w_h \times h(x) \\ &= < w_t, w_d > \times < ied_{x,x_0}.tm, ied_{x,x_0}.den > + w_h \times h(x) \\ &= (w_t \times ied_{x_0,x}.tm + w_d \times ied_{x_0,x}.den) + w_h \times h(x) \end{aligned} \tag{1}$$

Using the above $f_w(x)$ as the evaluation cost function of the A* algorithm, a predicted path, $\widehat{tr^k}$, from the departure to the destination is generated based on the actual path, tr^k. Predicted path $\widehat{tr^k}$ generated at this time, and the actual path tr^k, are compared to optimize the weights w_t, w_d, and w_h. At this time, the given optimization equation is as follows:

$$\begin{aligned} &\min_{w_t,w_d,w_h} \left[\sum_{k=1}^{m} d_H\left(tr^k, \widehat{tr^k}\right) \right], \\ &\text{Subject to } \; w_t + w_d + w_h = 1, \\ &\qquad\qquad\quad w_t > 0,\ w_d > 0,\ w_h > 0 \end{aligned} \tag{2}$$

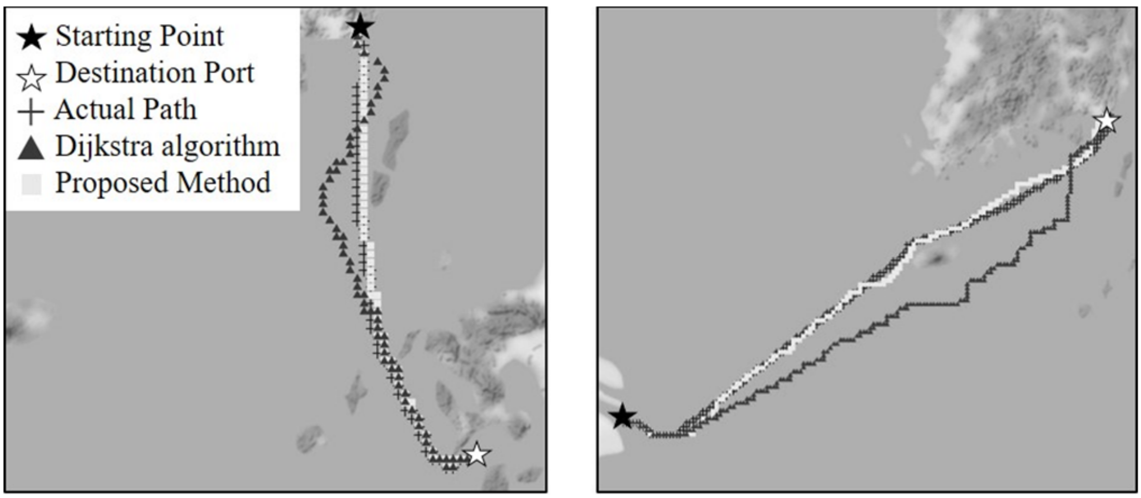
Solving the above, first, extract a random actual m path $\left(\left\{ tr^k \middle| k = 1, \ldots, m\right\}\right)$ and apply it to the A* algorithm to generate predicted path $\left\{\widehat{tr^k}\right\}_{k=1}^{m}$; d_H is a mathematical metric that measures the degree of similarity between one set and another set, called the Hausdorff distance [16], and it measures the degree of similarity between the predicted path $\widehat{tr^k}$ and actual path tr^k. Using d_H, the degree of similarity of the m paths is measured, and w_t, w_d and w_h are optimized using their sum. Since the above optimization problem goes through a process of finding each prediction path through the path and comparing it again, it may take a considerable amount of time to directly access it. Therefore, the optimization problem is solved by finding the optimal combination of w_t, w_d and w_h in increments of 0.05.

3.3. **Prediction of vessel arrival time.** A method of predicting the time of arrival from a point on the map to a port uses the time it took for the vessel to pass the same point in the past. At sea, the degree of congestion in maritime traffic is different for each point of each path [8]. To consider the characteristics of maritime traffic and use the average speed [17] representing the characteristics of normal/congested traffic, the average time difference between each grid and the grid generated per Section 3.2 is used. As an example, from Figure 3, the proposed method in the paper is explained. The Tr^k, which is a reduced trajectory from Figure 3, is the predicted path from grid 1 to grid 9 using the method described in Section 3.2. The predicted path can be expressed as $\widehat{tr^k} = \left\{ grp_i^k \middle| i = 1, 4, 5, 8, 9 \right\}$. At this time, the successive points of $\widehat{tr^k}$ become the edges, and the cost of each edge uses $ied_{j,k}.tm$ in Section 3.2. This can be used to calculate the remaining time for the vessel to reach its destination:

$$\sum_{i=1}^{4} \left(ied_{p_i, p_{i+1}}.tm \right) \tag{3}$$

The above calculation means the sum of $ied_{j,k}.tm$, which is the average time difference for each grid. Through this, it is possible to predict the vessel's arrival time considering the characteristics of the maritime traffic, which are different for each point of the path.

4. **Results.** Figure 4 shows a comparison of the paths from two points to the destination, B-Terminal located in Busan, Korea. They are the real path, the path predicted by the Dijkstra method, and the path predicted by the method proposed in this paper. Figure 4(a) shows the path from the Nagasaki Prefecture in Japan to the Port of Busan. Through the algorithm proposed in this paper, the path from the point marked with a black star to the point marked with a white star was generated and is represented with a white square. To compare the method proposed in this paper, the path created by the Dijkstra method is shown as a triangle. Figure 4(b) shows the path from the Port of Shanghai in China to the Port of Busan. Both figures show that the proposed method in this paper is more similar to the actual path than the Dijkstra method.

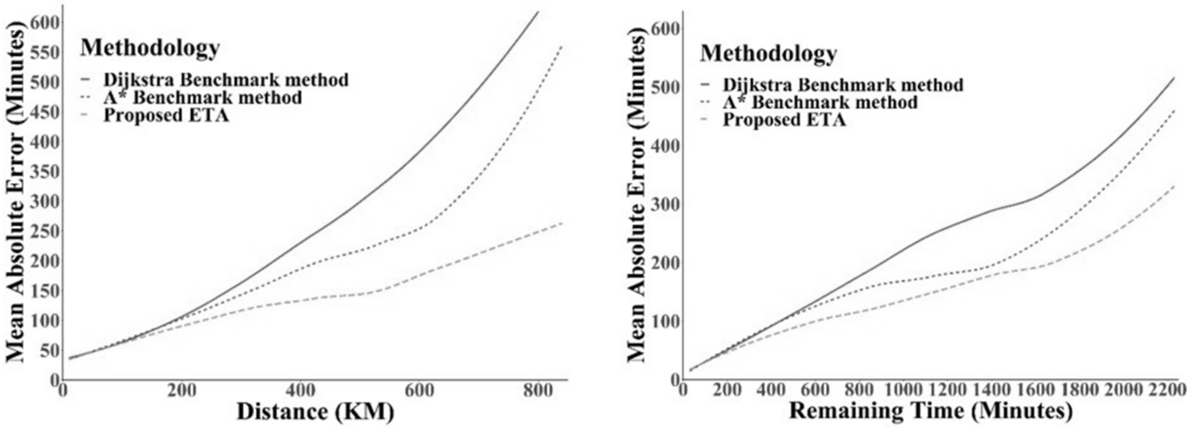


(a) (b)

Figure 4. Pathfinding result

To evaluate the performance of the vessel arrival time algorithm, the algorithm was applied to the vessel track extracted from Section 3.1. A destination and a source are required as input variables to the pathfinding algorithm. The destination is fixed at B-Terminal in the Port of Busan; the starting point was changed from the first point of the track to the destination port, and the pathfinding algorithm was applied. Then, for each point on the track, the arrival time was estimated based on the predicted path. Estimated arrival times can be calculated for all points on the vessel tracking, and the error compared to the actual arrival time can be calculated using this.

Figure 5(a) shows the absolute error between the predicted arrival time and the actual arrival time for the remaining distance to the destination for the vessel. The dotted lines in Figure 5(a) are benchmarks of the method proposed in [18] and represent a method for calculating the arrival time by dividing the predicted path's distance by the average value of SOG (speed in AIS data). The detailed dotted line is the distance of the path found using the Dijkstra method divided by the average speed of the vessel, and the wide dotted line is the distance of the path obtained by weight optimization A* proposed in this paper divided by the average speed of the vessel. Finally, the solid line is the arrival time as calculated by the method proposed in this paper. In all cases, the error increases as the remaining distance to the destination increases. When comparing two source points, even if the distance between the destination and the source is the same, if the distance is not large, the probability that the two source points are in the same space is high. Conversely, if the distance is great, the probability is high that even points at the same distance are in a different space. This means that if the distance to the destination is great, the data density is low because there are many points, even at the same distance. Accordingly, as the distance increases, the data density is relatively low, so the error in the predicted arrival time increases. The remaining distance varies from 0 to 800 km, and the results of each algorithm in Figure 5(a) show that the method proposed in this paper has better performance than the benchmark method.



(a) (b)

Figure 5. Absolute error between ATA and predicted ETA

In the following experiment, the absolute error between the predicted arrival time and the actual arrival time was compared with several methods. Figure 5(b) shows the results of the experiment. Like Figure 5(a), the solid line represents the method proposed in this paper, and the dotted lines represent the comparison methodologies. In all remaining time intervals, the proposed method showed better results.

A similar path was created based on the past path, and the arrival time was calculated from the result of pre-processing the AIS data. These AIS data imply the marine environment such as wind direction, wind speed, and waves. By considering these marine environment factors form the AIS data, the result outperformed by the proposed method.

5. **Conclusions.** This paper presents a new method for predicting a vessel's time of arrival at its destination. This method uses the vessel's track data and approaches it based on trajectory mining and pathfinding algorithms. Also, a methodology using AIS data, which is a record of past voyages, was presented. In the pathfinding problem, an optimization equation was introduced to find a path similar to a past track. In this paper, the new arrival time prediction method and methods proposed in other papers were compared through experiments. As a result, the proposed method showed better performance in the error between the actual arrival time and the predicted arrival time.

As much as a vessel sails in the sea, it is affected by marine conditions such as wind direction, wind speed, wave height, and weather. The method proposed in this paper has one limitation in that it does not consider the ocean conditions. Therefore, if an element that reflects the marine situation is added, it is expected to show even better performance. Also, the problem of optimizing the weight of pathfinding, or the problem of introducing other variables as an evaluation function of A*, can be a future research project.

If we reconsider slightly better performance in the results shown in this paper, it can be used for RTA occurrence and accurate ETA prediction. If this is applied to the port system to improve monitoring activities and further establish an accurate berth plan, it will help to improve port productivity.

Acknowledgment. This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2020-2016-0-00318) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation) and in part of the research projects of 'Development of IoT Infrastructure Technology for Smart Port' funded by the Ministry of Oceans and Fisheries, Korea.

REFERENCES

[1] S. F. Ahamed, Application of alternative fuels in maritime industry, *International Research Journal of Engineering and Technology (IRJET)*, vol.5, 2018.

[2] Y.-T. Park, H.-S. Kim and G. Cho, Study on the system development for port logistics system diagnostics, *ICIC Express Letters, Part B: Applications*, vol.10, no.7, pp.605-612, 2019.

[3] ITU-R, *Technical Characteristics for an Automatic Identification System Using Time-Division Multiple Access in the VHF Maritime Mobile Band*, M Series, 2010.

[4] D. Yang et al., How big data enriches maritime research – A critical review of automatic identification system (AIS) data applications, *Transport Reviews*, vol.39, no.6, pp.755-773, 2019.

[5] B. J. Tetreault, Use of the automatic identification system (AIS) for maritime domain awareness (MDA), *Proc. of Oceans 2005 MTS/IEEE*, Washington, D.C., pp.1590-1594, 2005.

[6] H. Hasheminia and C. Jiang, Strategic trade-off between vessel delay and schedule recovery: An empirical analysis of container liner shipping, *Maritime Policy & Management*, vol.44, no.4, pp.458-473, 2017.

[7] W. H. Lee, Determination of optimal ship route in coastal sea considering sea state and under keel clearance, *Journal of the Society of Naval Architects of Korea*, vol.56, no.6, pp.480-487, 2019.

[8] Y. S. Park and J. Y. Jeong, A study on the marine traffic congestion by analysis of ship's domain, *Journal of the Korean Society of Marine Environment & Safety*, vol.20, no.5, pp.535-542, 2014.

[9] Y. Zheng, Trajectory data mining: An overview, *ACM Trans. Intelligent Systems and Technology (TIST)*, vol.6, no.3, 2015.

[10] X. Cui and H. Shi, An overview of pathfinding in navigation mesh, *International Journal of Computer Science and Network Security*, vol.12, no.12, pp.48-51, 2012.

[11] E. W. Dijkstra, A note on two problems in connexion with graphs, *Numerische Mathematic*, vol.1, no.1, pp.269-271, 1959.

[12] S. Russel and P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, 1995.

[13] D.-D. Nguyen, C. L. Van and M. I. Ali, Vessel destination and arrival time prediction with sequence-to-sequence models over spatial grid, *Proc. of the 12th ACM International Conference on Distributed and Event-Based Systems*, pp.217-220, 2018.

[14] I. Parolas, ETA prediction for containerships at the port of Rotterdam using machine learning techniques, *TU Delft*, 2016.

[15] T. Mestl and K. Dausendschön, Port ETA prediction based on AIS data, *The 15th International Conference on Computer and IT Applications in the Maritime Industries*, Lecce, 2016.

[16] H. Alt et al., Approximate matching of polygonal shapes, *Annals of Mathematics and Artificial Intelligence*, vol.13, nos.3-4, pp.251-265, 1995.

[17] S. M. Kim and S. H. Cheon, A study on the estimation methodology of traffic congestion cost based on individual vehicle speed data, *KOTI*, vol.26, pp.45-57, 2019.

[18] A. Alessandrini et al., Estimated time of arrival using historical vessel tracking data, *IEEE Trans. Intelligent Transportation Systems*, vol.20, no.1, pp.7-15, 2018.