International Journal of Architectural, Civil and Construction Sciences
ISSN: 2415-1734
Vol:15, No:8, 2021

Destination Port Detection for Vessels: An Analytic Tool for Optimizing Port Authorities Resources

Lubna Eljabu, Mohammad Etemad, Stan Matwin

Abstract—Port authorities have many challenges in congested ports to allocate their resources to provide a safe and secure loading/unloading procedure for cargo vessels. Selecting a destination port is the decision of a vessel master based on many factors such as weather, wavelength and changes of priorities. Having access to a tool which leverages Automatic Identification System (AIS) messages to monitor vessel's movements and accurately predict their next destination port promotes an effective resource allocation process for port authorities. In this research, we propose a method, namely, Reference Route of Trajectory (RRoT) to assist port authorities in predicting inflow and outflow traffic in their local environment by monitoring AIS messages. Our RRo method creates a reference route based on historical AIS messages. It utilizes some of the best trajectory similarity measures to identify the destination of a vessel using their recent movement. We evaluated five different similarity measures such as Discrete Fréchet Distance (DFD), Dynamic Time Warping (DTW), Partial Curve Mapping (PCM), Area between two curves (Area) and Curve length (CL). Our experiments show that our method identifies the destination port with an accuracy of 98.97% and an f-measure of 99.08% using Dynamic Time Warping (DTW) similarity measure.

Keywords—Spatial temporal data mining, trajectory mining, trajectory similarity, resource optimization.

I. Introduction

DUE to the advancement in information and communications technologies present in maritime transportation, a massive amount of vessel trajectory data can be captured and processed [1]. Analysis of these trajectories reveals typical mobility patterns of vessels and provides an overview of the maritime traffic [2], [3]. Such analysis facilitates more advanced tasks on trajectory mining, such as path planning for autonomous vessels [4]. Detecting the destination port is a piece of focal information for path planning and simulation of vessel movements. By early detection of the destination port, port authorities could allocate their resources more efficiently and effectively. This early detection could lower the waiting time for each vessel to load/unload in ports, and assist the port authorities to decrease their cost by optimizing their resource usage.

AIS is an automated tracking system installed onboard to record vessels' trajectories and provide unique identification for a specific vessel, such as timestamp, position, course, speed, AIS message 5, etc. The AIS message 5 includes the vessel's destination port and Estimate Time of Arrival (ETA) [5]. The AIS text field for destination allows for "free text" up to 20 characters, resulting in numerous variations in the spelling of the same port [6]. However, the manually filled fields of destination in AIS message 5 are not always available or filled with mistakes [6]. The erroneously entered information affects the accuracy rate in ascertaining the correctness of destination reports observed in AIS data. The lack of accurate information of vessels' destinations would subject port authorities to challenges like arranging port activities for safe and efficient vessel operations and guiding traffic routes to ensure the safety and efficiency of the maritime traffic environment. Therefore, the research to predict vessels' destinations would be of great value for port authorities to automatize and to make timely and efficient decisions to allocate their resources and ensure a safe and secure maritime traffic environment. In summary, our approach can be utilized in scenarios where a system is going to generate some events based on the changes in the destination for each vessel and they do not require any human intervention.

Detecting the destination port can be seen from the trajectory path that normally traversed using AIS historical data, then compared to traveling trajectories to predict the destination. Thus, the similarities between traveling and historical trajectories can be measured and utilized to classify and predict the vessel's destination. Similarity analysis plays a significant role in solving many movement patterns recognition problems such as classification, clustering and anomaly detection. Vries et al. [7] proposed a similarity measure based on edit distance and applied this measure in a classification task to predict the type of vessel. Alizadeh et al. [8] suggested a point-based model for vessel location and traffic predictions. The location prediction procedure was setup based on similarity analysis of historical AIS data. Zhen et al. [9] offered an anomaly detection method for vessel's behaviour based on similarity analysis. A similarity measure between vessel trajectories is designed based on the spatial and directional features in their work. Then they applied this measure in clustering and classification tasks to detect anomalous vessel sailing behaviour [9].

Meaningful sailing patterns can be extracted from semantic trajectories [10]. A semantic trajectory of a vessel is generated by integrating some background datasets such as maps and geographical layers [11]. However, analyzing these trajectories of moving vessels is challenging because the volume of data to be processed is very large and the computation is complex [12]. The movement of a vessel can be identified based on the characteristics of the surrounding environment and underlying landscape such as anchoring areas, sea lanes

L. Eljabu is with the Institute for Big Data Analytics, Dalhousie University, Halifax, NS, Canada (e-mail: l.eljabu@gmail.com).
M. Etemad is with the Institute for Big Data Analytics, Dalhousie University, Halifax, NS, Canada.
S. Matwin is with the Institute for Computer Science, Polish Academy of Sciences, Warsaw and Postcode, Poland.

and harbours. Adding this geographical domain information to the trajectories can be used as a discriminator and incorporate it with the trajectory data in similarity measure allows more complex analysis and a better understanding of vessel routes and changes in movement behaviour [7], [8], [12], [13].

Trajectory segmentation methods provide basics for detecting changes in vessel movement behaviour [14]. There are many approaches such as SWS, WSII, CBSMoT, GRASP-UTS, and SPD available for trajectory segmentation [15]–[18]. The segmentation process partitions trajectories into segments that enable us to discover different underlying patterns.

In this work, we propose to combine vessel trajectory data with geographical knowledge to identify the ports where the vessel stops. Then, we perform trajectory segmentation based on the defined ports. The trajectory segmentation facilitates an understanding of the purpose of the trip. Also, enables tracking in large water area to identify vessels activities. Once the segmentation of trajectories is completed, some semantic labels are added to each segment such as the path number, to distinguish the routes with the same start and end ports, and segment identifier, a unique number for each segment. Once the segmentation is completed and annotations are applied, the resulted data will be incorporated in the similarity measure and reference route construction. We studied five different similarity measures and investigated which measure is the best similarity measure for comparing routes in vessel navigation. Then, we utilized the trajectory segments to generate a summarized reference trajectory. After that, we implemented the selected similarity measures as classification techniques to predict the destination port of vessel routes and find the best similarity measure performance of estimating the similarity between short segments and reference routes. The contributions of this work are summarized below:

- We proposed a geographical knowledge annotation for the generated segments to distinguish trajectories by their spatial feature and their context.
- A method to generate a reference route of trajectory (RRoT) is proposed that summarizes a set of segments with the same source and destination into a summarized reference trajectory.
- We compared five different similarity measures to understand the best similarity measure for comparing vessel trajectories.
- We suggested a method to predict the destination of a vessel by processing a short and recent segment.

The remainder of this work is structured as follows. In Section II, we provide definitions revolving around trajectory data mining. Section III presents previous work in related domains. Section IV explains the four steps of our proposed methodology, i.e. data preparations, reference route construction, comparing the proposed similarity measures and destination port prediction. The experimental results are reviewed in Section V. Finally, we conclude our discussion in section VI and share some future work.

II. Definitions

Definition 1. A trajectory point, l_i, is a geolocation of object o at time i, and is defined as, $l_i^o = \langle x_i^o, y_i^o \rangle$, where x_i^o represents the longitude of the location which varies from $0°$ to $\pm 180°$, and y_i^o represents the latitude of the location that varies from $0°$ to $\pm 90°$.

Definition 2. A *raw trajectory*, or simply *trajectory*, τ, is a time-ordered sequence of trajectory points of a moving object o,

$$\tau^o = \langle l_0^o, l_1^o, .., l_n^o \rangle \tag{1}$$

Because the trajectory data is from AIS, the trajectory points could be provided with additional information, including vessel identity, course/speed over ground, ship type which are considered as segment features.

Definition 3. A Port [19] is a polygon defining a circular area of radius r centered on the geographical coordinates of a sea port.

Definition 4. Partitioning Position is the last trajectory point of a trajectory segment where the segment movement behaviour changes.

Definition 5. Trajectory Segment s_i is a set of consecutive trajectory points belonging to a raw trajectory $\tau^o = \langle l_0^o, l_1^o, .., l_n^o \rangle$,

$$s^o = \langle l_j^o, \cdots, l_k^o \rangle, \;\; j \geq 0, \; k \leq n \text{ and } \mathrm{s^o} \text{ is a subsequence of } \tau^o \tag{2}$$

The process of generating segments from a trajectory is called trajectory segmentation.

Definition 6. Origin and Destination. The nearest *Port* to the first trajectory point in a segment is the Origin Port and the nearest *Port* to the last trajectory point of a segment is the Destination Port.

Definition 7. A reference route of trajectory A reference route of trajectory between point A and B is a trajectory segment that shows the average behaviour of all trajectory segments starting from point A and ending at point B. There can be various way of calculating this reference route of trajectory. We explains the calculation of route of trajectory for this paper in Algorithm .1.

Definition 8. Segment Label or Segment Feature A Segment Label or Segment Feature is an annotation given to a segment to define the shipping lanes with different origin and destination ports and we use segment label or Segment feature for ground truth in our work [20].

Definition 9. Trajectory Similarity is a path distance function δ, that measures a distance between pair of trajectories P, Q. δ is zero when P equals Q and grows positively as the curves become more dissimilar.

III. Related Work

In this section, we review related state-of-the-art trajectory data preparations and trajectory similarity measures. In trajectory data preparations, stop-points are required to be identified first. Then, trajectories are segmented to measure the similarity.

A. Trajectory Segmentation

There are many trajectory segmentation algorithms available for segmenting trajectory in different situations. We selected SPD, CBSMoT, and SWS to review here for the following reasons. SPD is the simplest basic algorithm for trajectory segmentation and has been provided reasonable results for segmentation [14]. CBSMoT is an approach that reported reasonable results in the vessel navigation domain. We also selected SWS which is the most recent and simple trajectory segmentation algorithm based on detecting changes in the behaviour of movement. Li et al. [21] propose stay Point Detection (SPD) algorithm that detects stay points by observing a moving object stay within a certain spatial region for a period exceeding a certain threshold. Palma et al. [22] propose a spatiotemporal clustering method based on the DB-SCAN algorithm, is called the CB-SMoT (Clustering-Based Stops and Moves of Trajectories). CBSMoT discovers the stops and moves based on the speed variation of the moving object. In this method, stops are considered the most important part of the trajectory. The stops are clustered where the trajectory speed is lower than a given threshold for a minimal period, then match the clusters with relevant geographic places [22]. Etemad et al. [16] proposed Sliding Window Segmentation (SWS). It is the most recent algorithm for trajectory segmentation. This algorithm produces segments based on the position of trajectory point where the moving object changes its behaviour. This algorithm has some extensions such as WSII benefiting from using machine learning approaches to detect segments [15].

In summary, we applied trajectory segmentation for the pre-processing of the trajectories of vessels. The differences between the trajectory segmentation algorithms were insignificant since we calculated the nearest Port as an origin and destination for the segments. Therefore, we decided to apply SPD in our research. Future work can be carried on to investigate the difference between using SPD and other approaches in the port detection scenario.

B. Trajectory Similarity

Generally, the similarity of trajectories is evaluated using the distance between their points. A review of the literature reveals that with the plethora of trajectory similarity measures available for movement data, there are several measures that are used frequently. Popular measures of trajectory similarity include: Euclidean distance, Dynamic Time Warping [23], Fréchet Distance [24], Discrete Fréchet Distance [25]. Due to the widespread use of these popular measures, functions for calculating these measures have already been implemented in commonly used statistical software, such as R and Python [26]. In this work we utilize the Python similarity measures library developed by Jekel et al. [27]. This library includes five methods: (i) Discrete Fréchet Distance, (ii) Dynamic Time Warping, (iii) Partial Curve Mapping, (iv) Area between two curves and (v) Curve length.

The Python similarity measures library is selected for the following reasons: it is the most recent python library for measuring the similarity of trajectories and the included similarity measures can be used on trajectories of different length, as vessels trajectories are typically different in temporal length, distance traveled and the number of data points. Also, these similarity measures can handle the non-linearity of vessel movement. Furthermore, the Discrete Fréchet Distance and the Dynamic Time Warping are of the most common methods used for measuring the similarity of trajectories. The Partial Curve Mapping, the Area between two curves and the Curve length methods are used for the first time to measure the similarity of vessels trajectories. These five methods will be compared and evaluated to find the best similarity measure for port detection problem.

IV. Methodology

In this section, we introduce our proposed approach for destination port prediction. We present the sequence of steps used in this work to predict the destination port of the vessel route (Fig. 1). This framework has four main steps: 1- Data preparation, 2- Reference route construction, 3- Similarity measurements, 4- Destination port prediction. We describe these steps in detail in the following sections.

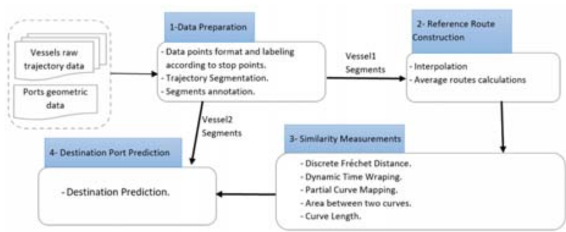


Fig. 1 Framework of vessel destination port prediction

A. Data Preparation

The data preparation first step is to form the vessel *trajectory* by sorting a set of trajectory points collected from AIS messages based on date and time of collection. Then we utilize semantic layer information that contains the related *ports* data (Fig. 2) to annotate the trajectory points. Annotating the trajectory points provides a more detailed description of the sailing behaviour of vessels. The trajectory points are annotated according to the defined *ports*; each trajectory point is checked whether or not it is located within the region of one of the defined *ports*. If the trajectory point locates in a port area, it is annotated as a *stop* trajectory point. If it locates outside of a port area it is annotated as a *move* trajectory point.

port_name	geometry
port1	POLYGON ((-63.569431 44.649993, -63.56943966749199 44.64981656914741, -63.56946558649528 44.64964183742038, -63.569508507395
port2	POLYGON ((-63.568048 44.663875, -63.56805666749199 44.66369856914741, -63.56808258649527 44.66352383742037, -63.568125507395
port3	POLYGON ((-63.555828 44.662453, -63.55583666749199 44.66227656914741, -63.55586258649527 44.66210183742037, -63.555905507395
port4	POLYGON ((-63.547843 44.648763, -63.54785166749199 44.64858656914741, -63.54787758649528 44.64841183742038, -63.547920507395

Fig. 2 A depiction of the ferry four terminals *(ports)* data

Then, the annotated data are used to partition the trajectory of a vessel into segments. This process is called trajectory segmentation. The main advantages of trajectory segmentation are to facilitate an understanding of the purpose of the trip and to track vessels across a large area to identify their activities.

The segmentation process detects the partitioning positions in the trajectory and uses them to divide the trajectory into distinct segments. These segments accurately capture the underlying movement patterns which aid in mining richer knowledge. Then, the centers of origin and destination port are added to the segment, which represent the start and endpoints of the segment. Each of these segments represents movements from the origin port to the destination port. Fig. 3 shows segments resulting from the segmentation.

Fig. 3 Segments between $stop_1$ and $stop_2$. The red and green segments represent the movement of vessel from $stop_1$ to $stop_2$. The Blue and yellow segments represent the movement of vessels from $stop_2$ to $stop_1$

The segmentation step provides information on partitioning positions $(ports)$ along the trajectory, without any further ecological context [28]. Thus, it is not possible to directly associate an individual segment to a specific sea lane or specific sailing behaviour. To facilitate the ecological interpretation of the segments, we propose to assign an identifier annotation, VID, to each segment to distinguish the segments. Moreover, we assign another feature to segments, $route$, and annotate segments with the same origin and destination ports with $Path$ number to facilitate classifying segments based on these labels. i.e. Fig. 3 shows four segments represent trips through the area of the same shipping lane (e.g., $stop_1$-$stop_2$), but the segments are separated by $routes$ with different start and endpoints (e.g., The red and green segments represent the movement of a vessel from $stop_1$ to $stop_2$ and they are annotated as $Path_1$. The Blue and yellow segments showing the movement of a vessel from $stop_2$ to $stop_1$ and they are annotated as $Path_2$. To sum up, these segments may appear visually similar whereas the ecological context is not similar.

B. Reference Route Construction

After the trajectories are partitioned into segments and these segments are labelled we construct *Reference Route of Trajectory (RRoT)*. The reference route is a mean segment represents the trajectory segments between every two ports (step 2). The reference route construction phase consists of two steps:

Step one is to interpolate trajectory points in the segments based on time. Linear interpolation is used with the time and longitude/latitude data points because it is the simplest and consumes the lowest computational power. Algorithm .1 shows the steps of generating reference route. We interpolate the longitude values concerning time. The input of this algorithm is all segments that belong to the same route (segments that have the same origin port and same destination port). For each segment (line 1), the time is transformed to an increasing number (line 2), then apply the spacing to create a list of an evenly spaced sequence in the specified period of time of required data points (line 3). Then, we pass the list of an evenly spaced sequence created from the spacing method to the linear interpolation method and interpolate independently for the longitude values (line 4) and the latitude values (line 5).

In step two, each mean segment is calculated from many segments between two ports that belong to the same $route$ (lines 10, 11, 12). Finally, return the mean segment as a reference route. The interpolated reference routes result in no missing data points, consistent and usable formatting of the trajectory data.

Fig. 4 shows the reference routes of the vessel trajectory that is illustrated in Fig. 6. Since ferryboats trips rarely follow a random trajectory, these reference routes represent the travel history and movement paths that are normally traversed by the ferry.

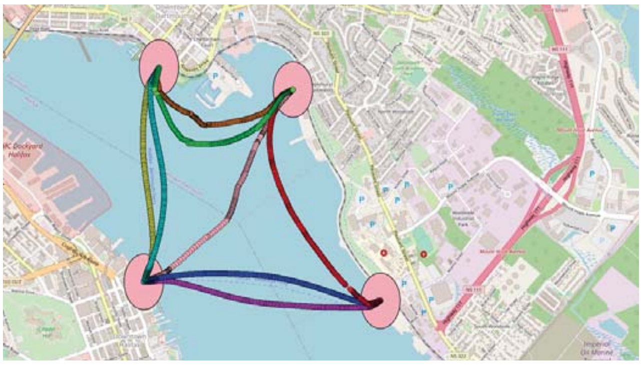


Fig. 4 Depiction of the constructed reference routes of trajectory

C. Similarity Measurements

In phase three, five similarity measurements are explored to find the best similarity measure for our trajectory data. As mentioned in Section III, the five similarity measurements are 1) Discrete Fréchet Distance, 2) Dynamic Time Warping, 3) Partial Curve Mapping, 4) Area between two curves and 5) Curve length. We assess the performance of these five measurements by observing the distributions of similarity and dissimilarity scores. It is a simple way to quantify the difference between the similarity and dissimilarity distributions of each method. If a method has both distributions overlap, the method will not perform well in quantifying the differences between the compared routes and it will be eliminated.

D. Destination Port Prediction

Finally, the best similarity measurements are implemented as pattern classification techniques that measure the similarity

Algorithm .1 Reference Route Algorithm

Require: $Segments\_of\_the\_same\_route$ {All segments that has the same origin and destination ports}
Ensure: $Reference\_Route$ {An average segment represents the mean of $Segments\_of\_the\_same\_route$}

for $segment$ **in** $Segments\_of\_the\_same\_route$ **do**
  tm $\leftarrow$ *segment['Time']* converted to increasing numbers;
  /*create an evenly spaced sequence in the specified period of tm*/
  interp_time $\leftarrow$ *linspace(tm[0], tm[-1], num_points);*
  /*functions return one-dimensional piecewise linear interpolated lon/lat with given discrete data points (tm, lon/lat_points), evaluated at interp_time */
  interp_longitude $\leftarrow$ *interpolate (interp_time, tm, Longitude_points);*
  interp_latitude $\leftarrow$ *interpolate (interp_time, tm, Latitude_points);*
  /*sum the time and coordinates values of the segments*/
  sum_interp_time;
  $sum\_interp\_longitud$;
  $sum\_interp\_latitude$;
end for
/*compute the average time and average coordinates*/)
avg_time $\leftarrow$ *mean(sum_interp_time)*
avg_longitude $\leftarrow$ *mean(sum_interp_longitude)*
avg_latitude $\leftarrow$ *mean(sum_interp_latitude)*
/*Concatenate the averages and store the results in Reference_Route*/
Reference_Route $\leftarrow$ *(avg_time, avg_longitude, avg_latitude, label, route)*
return $Reference\_Route$

between the input tracking data and the stored reference routes data (Algorithm.1) in phase four. The tested segments will be assigned to a reference route according to the minimum similarity score. Then the destination port of the reference route will be the result of the prediction.

V. Experiments

This section provides the details of our experiments and the experimental results of our proposed method. The experiments are performed in the sequence of three steps: 1) data and data preparation, 2) selecting the best similarity measure, and 3) measuring the accuracy of destination prediction.

First, we do an experiment on the distribution of destination ports in AIS message 5 in order to see how much the data embedded in this field is useful. So, it is practical for port authorities to be able to use this data efficiently. Fig. 5 shows a distribution of the destination field of AIS message 5 in the AIS data. The majority of distribution shows that the destination port is not entered, or the destination port field is unknown where the name of a small town, bay, anchorage, or shipyard is entered. All these variations cause ambiguity in destination reports and leading to confusion and data interchange inefficiency.

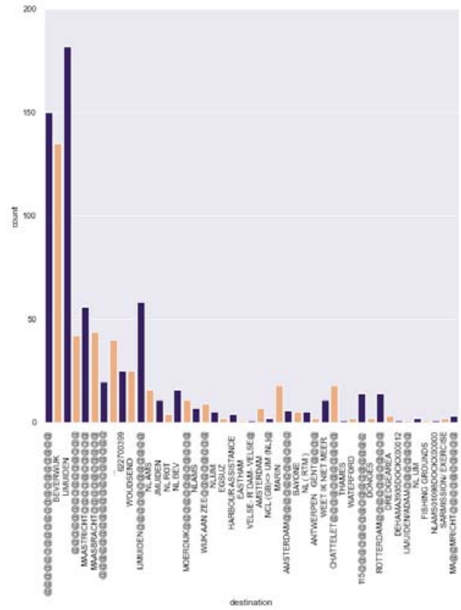


Fig. 5 The distribution of destination ports in AIS message_5 shows most of the data submitted in this field is not accurate and not a representation of the real world

A. Data and Data Preparation

The data used for this work cover the AIS messages captured in the Halifax harbour area in Nova Scotia province in Canada. For our experiments, we explore trajectories of two transit ferries navigating in the Halifax area. The first vessel, V_1, trajectory points were collected from 18-March-2019 to 10-July-2019. The dataset has 27,028 trajectory points. The second vessel, V_2, trajectory points were collected from 05-March-2019 to 12-July-2019 including 103,161 trajectory points.

As shown in Fig. 6 a ferry can sail to different destination ports. Therefore, if we have the last trajectory points of a segment (sub-segment) that are near to the destination port, we can calculate the similarity with the constructed reference routes (algorithm.1) and predict the destination port.

Understanding the context of the dataset is the key part of any research in ecology and Geographical Information Science (GIScience). GIS is one of the available tools that have some application to analyze spatial locations and organize information into visualizations using graphical maps, which reveal deeper insights into data. For this work we use the QGIS application to compose a semantic layer from OpenStreetMap and polygons represent the related ports for visualization. Fig. 6 depicts the integration of a vessel raw trajectory with a semantic layer using QGIS application. The geographical context makes it easier for users to identify the sailing behaviour of a vessel.

The trajectory points are ordered based on date and time and defined its semantic information by:

- Define Label feature that represents which trajectory point is a stopping point and which one is not. Check each

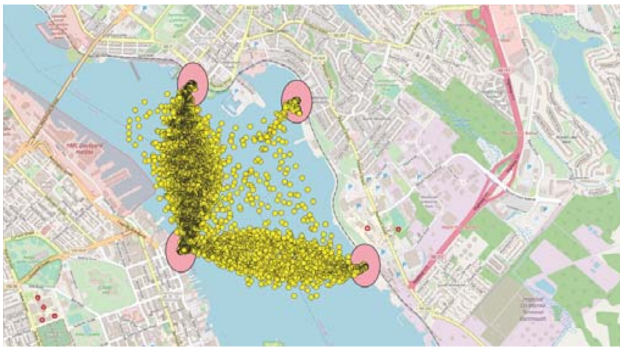


Fig. 6 An overview of a vessel trajectory of 27028 trajectory points generated by transit ferry voyaging in Halifax water and a semantic layer representing the Halifax-Dartmouth ports that the transit ferry voyaging in between

trajectory point, if it locates in one of the four port areas assign: $stop_1$ or $stop_2$ or $stop_3$ or $stop_4$ to the Label. If the trajectory point locates outside the port areas, we assign $move$ to the Label feature.

- Pass the labelled trajectory data to the segmentation function where the trajectory of a vessel is partitioned into segments based on partitioning positions. Then, the centers of origin and destination ports are added as the start and endpoints of the segment.
- Define $route$ feature so similar segments with the same origin and destination ports will be given a distinct value. For example, a segment from $stop_1$ to $stop_2$ is annotated with $Path_1$, a segment from $stop_2$ to $stop_1$ is annotated with $Path_2$ etc.

The output data that are used to investigate the best similarity measure, 1186 segments are produced by V_1 and 4263 segments are produced by V_2. The 1186 segments of V_1 are used to construct the reference routes by using our proposed *Reference Route Algorithm* (algorithm .1). While the 4263 segments of V_2 are used as test data for route classification and destination port prediction.

B. Best Similarity Measure

The five similarity measurements we examine are: 1) Discrete Fréchet Distance (DFD), 2) Dynamic Time Warping (DTW), 3) Partial Curve Mapping (PCM), 4) Area between two curves (Area) and 5) Curve length (CL). We investigate the similarity measurements' performance using the distribution of the quantified differences between all segments of V_1. For similar segments that have the same origin and destination ports, we calculate their similarity differences and make a distribution of the acquired scores. For the dissimilar segments, we choose a segment that represents one $route$ and compare it with segments representing other $routes$ and make a distribution of the acquired dissimilarity scores.

We visualize both the similarity and dissimilarity distributions side-by-side to infer each method performance as shown in Fig. 7. It is obvious from the plots that in the plot (c) PCM and the plot (d) Area methods, their similarity and dissimilarity distributions are overlapped. This means the proposed models using these methods will have a poor performance in the classification and prediction of destination ports. Therefore, these two methods are eliminated.

Discrete Fréchet Distance (DFD), Dynamic Time Warping (DTW) and Curve Length (CL) performed the best across our simulation in terms of being able to distinguish the different distributions of the similarity and dissimilarity scores. These results guiding which similarity measure is the 'best' similarity measure to use with this AIS data among similarity measures that are studied.

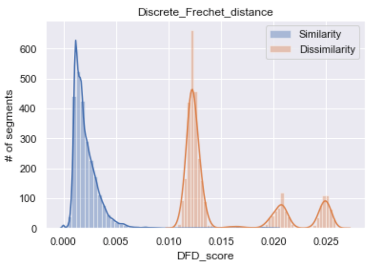
C. The Accuracy of Destination Prediction

As vessel trips rarely follow a random trajectory, the prediction of destination port could be considered as a classification problem. We use the 4263 segments of V_2 as test data. First, We create a list of actual labels $(i.e. path_1, path_2, path_3, path_4, path_5, path_6)$ which represents *the actual routes* of the 4263 segments. Each segment's route feature is compared to the constructed reference routes and the corresponding route_id is appended to the list. Then, We choose the last six trajectory points of each segment (sub-segment) before the destination port. Based on the capturing rate, six trajectory points provide the minimum information we need to calculate the similarity. Then, we calculate the similarity between these six trajectory points and the reference routes.

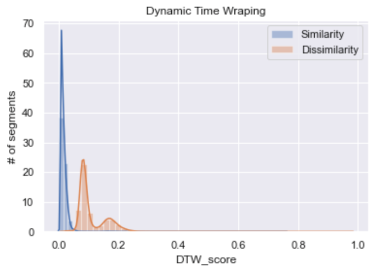
Each reference route consists of 300 trajectory points. Using the three selected similarity measurements: Discrete Fréchet Distance (DFD), Dynamic Time Warping (DTW) and Curve Length (CL). The sub-segment will be classified according to the minimum similarity score with one of the constructed reference routes. The destination port will be predicted according to the selected reference route destination port. Therefore, we first visualize the performance of each method using the multi-class confusion matrix. Then, to evaluate the performance of the three selected similarity measures we used the Accuracy and the F1 measure.

Fig. 8 demonstrates a visual interpretation of the confusion matrices where the prediction output for the similarity measurements models have six routes: $path_1$, $path_2$, $path_3$, $path_4$, $path_5$, $path_6$. The diagonal elements represent the correct predictions per route, the lighter the color the greater the number. While off-diagonal elements are those that are mislabelled. These matrices show the percentage prediction of each destination port made by each model for the actual routes. As shown in Fig. 8, confusion matrix in (a) represents the percentage prediction of each destination port made by the model using Discrete Fréchet Distance (DFD) similarity measure for the actual routes. Diagonal elements for $path_1$, $path_2$, $path_4$, $path_5$, $path_6$ are perfectly predicted, but it performs comparatively not good for the $path_3$, all true destination ports for $path_3$, it only predicts 55% of them correctly.

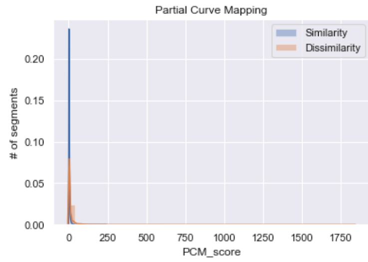
Confusion matrix in (b) represents the percentage prediction of each destination port made by the model using Dynamic Time Warping (DTW) similarity measure for the actual routes. Diagonal elements for $path_1$, $path_2$,$path_5$, $path_6$ are perfectly



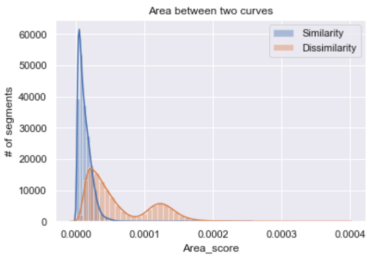
(a) Discrete Fréchet Distance



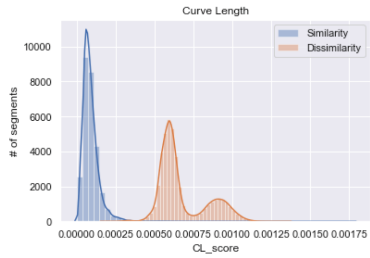
(b) Dynamic Time Warping



(c) Partial Curve Mapping

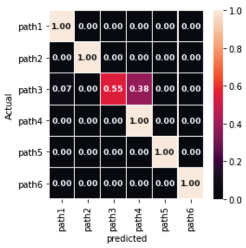


(d) Area between two curves

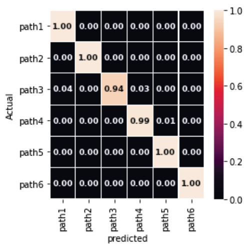


(e) Curve Length

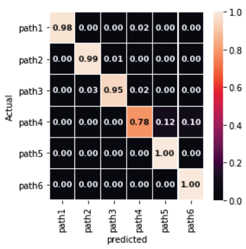
Fig. 7 A depiction of similarity and dissimilarity distributions of the five selected methods



(a) Destination port prediction performance using Discrete Fréchet Distance



(b) Destination port prediction performance using Dynamic Time Warping



(c) Destination port prediction performance using Curve Length

Fig. 8 A depiction of confusion matrices of destination port prediction performance using similarity methods: (a)Discrete Fréchet Distance, (b) Dynamic Time Warping, (C) Curve Length

predicted. Next comes $path_4$ with 99% correct predictions. Then comes $path_3$ with 94% correct predictions which is for this particular route this model of DTW outperforms the model of (DFD).

The confusion matrix in (c) represents the percentage prediction of each destination port made by the model using the Curve Length (CL) similarity measure for the actual routes. Diagonal elements for $path_5$, $path_6$ are perfectly predicted with 100% correct predictions. Next comes $path_2$ with 99% correct predictions. Then $path_1$ with 98% correct predictions. Then comes $path_3$ with 95% correct predictions, which is for this particular route this model of CL outperforms the model of (DFD) and slightly better than the model of DTW. Then, $path_4$ could predict the only 78% of them correctly.

From the results of the confusion matrices, we can infer that because the discrete Fréchet Distance (DFD) is a max-measure; defined as the maximum distance measured at each position. The dependence on the maximum value of distance leads to non-robust behaviour, where some variation in the sub-segments related to $path_3$ distorts the distance function by a large amount. Thus, the percentage prediction for $path_3$ is significantly low. Dynamic Time Warping (DTW) is a sum-measure, defined as the sum of the distance measured at each position. Hence, this measure smooths the distortion that occurred in the DFD model. Thus, the percentage prediction of the DTW model for $path_3$ is significantly improved. Curve Length (CL) is a measure of the i^{th} point of the sub-segment to the corresponding equivalent length of the curve of the compared reference route. The i^{th} point of the sub-segment does not correspond to the same abscissa, as is the case in the DFD and DTW models, where some variation in the sub-segments can distort the distance function by some amount as the case in $path4$ but notably less than the DFD model.

Table I shows the destination port accuracy and f1 measure for Discrete Fréchet Distance, Dynamic Time Warping, and Curve Length. The model of Dynamic Time Warping surpasses the other two methods in both accuracy and f1. As a result, Dynamic Time Warping is a very robust technique to compare the peaks and troughs by taking into account the varying lags and phases in the trajectories.

TABLE I
ACCURACY AND F1 MEASURE OF THE THREE SELECTED MODELS: DISCRETE FRÉCHET DISTANCE, DYNAMIC TIME WARPING AND CURVE LENGTH

Discrete Fréchet Distance		Dynamic Time Warping		Curve Length	
Acc.	f1	Acc.	f1	Acc.	f1
95.82%	95.31%	98.97%	99.08%	89.75%	93.58%

VI. CONCLUSION

We proposed a method to predict the destination port for a voyage using a recent trajectory segment of a vessel. This prediction can bring economic value to a higher level of maintenance for port authorities. The insight of future traffic in a port can help port authorities to dedicate their resources more effectively and efficiently. Our approach contains four major phases: 1) data preparations, 2) reference route construction 3) similarity measurement and 4) destination port prediction. We prepared data by extracting stop points of the trajectories and annotating the extracted segments. Then we proposed a *RRoT* method to find reference route which is the key for our solution. After that we explored five trajectory similarity measures (Discrete Fréchet Distance (DFD), Dynamic Time warping (DTW), Partial curve mapping (PCM) Area between two curves (Area) and Curve Length (CL)) to find the best trajectory similarity method for this domain. We applied the best similarity measures to detect the destination port.

Simulations suggested that DFD, DTW and CL performed best on the AIS data. In contrast, PCM and Area were generally the worst performing measures and lacked some of the stricter time-series constraints that characterize the other measures. By applying Dynamic Time warping (DTW) we identified the destination ports of 4263 trajectories with an accuracy of 98.97%. These studies suggest that the success of effective prediction of destination port depends upon the constructed reference route of trajectory and the distance function of the similarity measure that is used for the classification. The most important factor in this success is using a sufficient number of trajectory points with a robust similarity measure, that allows the model to better recognize the destination of a recent trajectory segment of a vessel.

In the future, we plan to expand this study by developing a graph for predicted traffic of vessels and identify routes with high demand in resources. Also, we plan to advance reference route detection to identify multiple paths with the same source and destination and propose a visualization tool for representing these paths on the map.

REFERENCES

[1] L. Qi and Z. ZHENG, "A measure of similarity between trajectories of vessels," *Journal of Engineering Science and Technology Review*, vol. 9, pp. 17–22, 03 2016.

[2] E. Carlini, V. M. de Lira, A. Soares, M. Etemad, B. B. Machado, and S. Matwin, "Uncovering vessel movement patterns from ais data with graph evolution analysis." in *EDBT/ICDT Workshops*, 2020.

[3] I. Varlamis, I. Kontopoulos, K. Tserpes, M. Etemad, A. Soares, and S. Matwin, "Building navigation networks from multi-vessel trajectory data," *GeoInformatica*, pp. 1–29, 2020.

[4] M. Etemad, N. Zare, M. Sarvmaili, A. Soares, B. B. Machado, and S. Matwin, "Using deep reinforcement learning methods for autonomous vessels in 2d environments," in *Canadian Conference on Artificial Intelligence*. Springer, 2020, pp. 220–231.

[5] "Ais class a ship static and voyage related data (message 5)," https://www.navcen.uscg.gov/?pageName=AISMessagesAStatic, accessed: 2021-04-02.

[6] S. Kos, M. Vukić, and D. Brcic, "Use of universal protocol for entering the port of destination in ais device," 2013.

[7] G. K. D. de Vries, W. R. van Hage, and M. van Someren, "Comparing vessel trajectories using geographical domain knowledge and alignments," in *2010 IEEE International Conference on Data Mining Workshops*, 2010, pp. 209–216.

[8] D. Alizadeh, A. A. Alesheikh, and M. Sharif, "Prediction of vessels locations and maritime traffic using similarity measurement of trajectory," *Annals of GIS*, vol. 0, no. 0, pp. 1–12, 2020. [Online]. Available: https://doi.org/10.1080/19475683.2020.1840434

[9] R. Zhen, J. Yongxing, Q. Hu, Z. Shao, and N. Nikitakos, "Maritime anomaly detection within coastal waters based on vessel trajectory clustering and naïve bayes classifier," *Journal of Navigation*, vol. 70, pp. 1–23, 01 2017.

[10] L. Alvares, V. Bogorny, and B. Kuijpers, "Towards semantic trajectory knowledge discovery," 10 2007.

[11] R. d. S. Mello, V. Bogorny, L. O. Alvares, L. H. Z. Santana, C. A. Ferrero, A. A. Frozza, G. A. Schreiner, and C. Renso, "Master: A multiple aspect view on trajectories," *Transactions in GIS*, vol. 23, no. 4, pp. 805–822, 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/tgis.12526

[12] P. Sheng and J. Yin, "Extracting shipping route patterns by trajectory clustering model based on automatic identification system data," *Sustainability*, vol. 10, p. 2327, 07 2018.

[13] M. Buchin, S. Dodge, and B. Speckmann, "Similarity of trajectories taking into account geographic context," *Journal of Spatial Information Science*, vol. 9, 11 2014.

[14] M. Etemad, "Novel algorithms for trajectory segmentation based on interpolation-based change detection strategies," 2020.

[15] M. Etemad, Z. Etemad, A. Soares, V. Bogorny, S. Matwin, and L. Torgo, "Wise sliding window segmentation: A classification-aided approach for trajectory segmentation," in *Canadian Conference on Artificial Intelligence*. Springer, 2020, pp. 208–219.

[16] M. Etemad, A. Soares, E. Etemad, J. Rose, L. Torgo, and S. Matwin, "Sws: an unsupervised trajectory segmentation algorithm based on change detection with interpolation kernels," *GeoInformatica*, pp. 1–21, 2020.

[17] Z. Xiu-Li and X. Wei-Xiang, "A clustering-based approach for discovering interesting places in a single trajectory," in *2009 second international conference on intelligent computation technology and automation*, vol. 3. IEEE, 2009, pp. 429–432.

[18] Y. Zheng, X. Xie, W.-Y. Ma *et al.*, "Geolife: A collaborative social networking service among user, location and trajectory." *IEEE Data Eng. Bull.*, vol. 33, no. 2, pp. 32–39, 2010.

[19] E. Carlini, V. Monteiro, A. Soares, M. Etemad, B. Machado, and S. Matwin, "Uncovering vessel movement patterns from ais data with graph evolution analysis," in *EDBT/ICDT Workshops*, 01 2020.

[20] M. Etemad, "Novel algorithms for trajectory segmentation based on interpolation-based change detection strategies," *Dalhousie Faculty of Graduate Studies Online Theses 2020*. [Online]. Available: http://hdl.handle.net/10222/79941

[21] Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, and W.-Y. Ma, "Mining user similarity based on location history," in *Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. GIS '08. New York, NY, USA: Association for Computing Machinery, 2008, p. 10. [Online]. Available: https://doi.org/10.1145/1463434.1463477

[22] A. Palma, V. Bogorny, B. Kuijpers, and L. Alvares, "A clustering-based approach for discovering interesting places in trajectories," 03 2008, pp. 863–868.

[23] P. Senin, "Dynamic time warping algorithm review," 01 2009.

[24] M. Fréchet, "Sur quelques points du calcul fonctionnel," *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, vol. 22, pp. 1–72.

[25] T. Eiter and H. Mannila, "Computing discrete frechet distance," 05 1994.

[26] I. Cleasby, E. Wakefield, B. Morrissey, T. Bodey, S. Votier, S. Bearhop, and K. Hamer, "Using time-series similarity measures to compare animal movement trajectories in ecology," *Behavioral Ecology and Sociobiology*, vol. 73, 11 2019.
[27] C. Jekel, G. Venter, M. Venter, N. Stander, and R. Haftka, "Similarity measures for identifying material parameters from hysteresis loops using inverse analysis," *International Journal of Material Forming*, vol. 12, 05 2019.
[28] H. Edelhoff, J. Signer, and N. Balkenhol, "Path segmentation for beginners: An overview of current methods for detecting changes in animal movement patterns," *Movement Ecology*, vol. 4, 09 2016.