Published in Network Science : 7th International Winter Conference, NetSci-X 2022, Porto, Portugal, February 8–11, 2022, Proceedings, pp. 80-93, which should be cited to refer to this work.
DOI: 10.1007/978-3-030-97240-0_7

Vessel Destination Prediction Using a Graph-Based Machine Learning Model

Racha Gouareb[1], Francois Can[2], Sohrab Ferdowsi[1,3], and Douglas Teodoro[1,3](✉)

[1] University of Geneva, Geneva, Switzerland
{racha.gouareb,douglas.teodoro}@unige.ch
[2] Riverlake Shipping SA, Geneva, Switzerland
can@riverlake.ch
[3] HES-SO University of Applied Sciences and Arts of Western Switzerland, Geneva, Switzerland
sohrab.ferdowsi@hesge.ch

Abstract. As the world's population continues to expand, maritime transport is critical to ensure economic growth. To improve security and safety of maritime transportation, the Automatic Identification System (AIS) collects real-time data about vessels and their positions. While a large portion of the AIS data is provided via an automatic tracking system, some key fields, such as destination and draught, are entered manually by the ship navigator and are thus prone to errors. To support decision making in maritime industries, in this paper we propose a data-driven vessel destination prediction algorithm based on heterogeneous graph and machine learning models. We design the task as a multi-class classification problem, where the destination port is the category to be predicted given the vessel and origin information. Then, we use a link prediction model in a weighted heterogeneous graph to predict the vessel destination. Experimental comparison against baseline methods, such as logistic regression and k-nearest neighbors, showed that our model provides a robust performance, outperforming the baseline algorithms by 9% and 33% in terms of accuracy and F1-score, respectively. Thus, heterogeneous graph models provide a powerful alternative to predict port destination, and could support enhancing AIS data quality and better decision making in maritime transportation industries. AQ1

Keywords: Destination prediction · Maritime transportation · Machine learning · Graph model · Link prediction · AIS · Heterogeneous graph

1 Introduction

Maritime shipping is one of the main pillars of freight transportation around the world. Due to its economic and environmental advantages, 90% of commodity shipment travels by the sea. With the expected world's population increase of 3.3 billion people by the end of the century [35], maritime traffic will continue

P. Ribeiro et al. (Eds.): NetSci-X 2022, LNCS 13197, pp. 1–14, 2022.
https://doi.org/10.1007/978-3-030-97240-0_7

to expand due to strong commodity needs. In turn, this expansion will lead to high transportation demand and increased traffic congestion, collisions, and accidents [18]. Thus, it is important to enhance available maritime data quality and explore different solutions for decision-making in the maritime industry.

The Automatic Identification System (AIS) is an automatic tracking system used by vessel traffic services and boats [19]. AIS uses a transceiver placed on ships to transmit their data. As of December 2004, installing AIS aboard vessels of a specific size and tonnage has become mandatory [20]. This regulation facilitated in the past two decades the collection of vessel information, including static data, such as vessel size, in addition to voyage information, such as position and destination. Since 2008, satellites equipped with receivers are able to receive AIS signals sent by the transceivers and easily collect AIS data [45]. By automatically sharing this information between ships and coastal authorities, the safety of ship management can be improved [4]. AIS information can be divided into three subcategories [1]. First, static data contain vessel-related information that defines the vessel's identity, such as MMSI and IMO, and are specified when the AIS system is installed on the ship. Second, navigational data, such as position coordinates, are transmitted automatically to track vessel movements every two to ten seconds, depending on the type and speed of the vessel. Finally, voyage data give general information about the voyage, including destination port, estimated time of arrival, and draught, and are entered manually before each journey.

The quality of AIS data can vary depending on the class of AIS equipment, that is, class A or class B. The choice of equipment is based on the type and size of vessels and the type of voyages a ship makes [43]. Despite its tabular format, AIS data is complex, requiring significant processing before it can be useful. For example, voyages' start and end flags are not readily available from the data. Moreover, due to technical failures, such as instability of the signal transmission rate, data transmission congestion [7], or environmental and human factors, it is estimated that as much as 80% of AIS messages contain errors [2,49], resulting in incorrect vessel name, Maritime Mobile Service Identity (MMSI) number, International Maritime Organization (IMO) number, position, and speed over ground [16], among others. Yang *et al.* [44] estimated that 40% of the data are wrongly entered on purpose or involuntarily, while Wu *et al.* [42] estimated that 62% of AIS destinations are mistaken and not always updated. For some ports, some studies showed that the accuracy of the reported destination information can be as low as 4% [27].

To improve the quality of AIS data and support maritime shipping decision-making, in this paper, we propose a link prediction algorithm in a heterogeneous graph model to address the problem of predicting voyage destinations using historical AIS data. Historical AIS data, such as latitude, longitude, and speed over ground (SOG), are used to construct the voyages. The resulting segments are used to create the navigation network, which is modeled as a heterogeneous graph and used to train and validate the prediction models. We defined the problem as a multi-label classification task. The algorithm's goal is to predict one of the destination ports from a pre-defined list extracted from the navigation network. Then, using the graph model, a link prediction algorithm is used to predict the next port for a vessel.

The remainder of the paper is divided into the following sections. Section 2 summarizes previous and most recent research work related to AIS data, including destination and trajectory predictions. Section 3 illustrates the data pre-processing and voyages creation process and describes the data and the proposed prediction models. Section 4 shows the results of prediction algorithms, followed by Sect. 5 that illustrates limitations and potential extensions of this work and concludes the study.

2 Related Work

Due to the high demand for shipping services [22], the development of maritime industries, and the increase in maritime traffic, accidents, and collisions [23], AIS data-driven solutions have received considerable attention from researchers. Thus, several studies were conducted to investigate research questions in the field of maritime traffic using historical AIS data. Examples of these studies include the detection of abnormal ship behavior [48], prediction of vessel trajectory [32, 47], data analysis, such as outlier detection [5] and collision risk analysis [29], and the application of machine learning algorithms [46] to improve the quality of AIS data and enhance the performance of handling maritime processes.

AIS data were used to develop various destination prediction models using both classic and deep learning-based machine learning approaches. Zhang *et al.* [46] used a random forest-based model supported by historical AIS data to create a destination prediction model based on similarities between trajectories. The data were labeled using a data clustering algorithm, the density-based spatial clustering of applications with noise method [14]. Their results showed better model accuracy when predicting cities rather than ports. Wang *et al.* [39] also used a random forest-based model combined with a port frequency-based decision strategy for destination prediction problems for ships. The authors highlighted different approaches used to pre-process and construct ship trajectories from raw AIS data and noise filtering methods, such as the average and Kalman filters and heuristic-based outlier detection methods. Lin *et al.* [24] used deep learning models for destination and arrival time prediction for different ship types. They proposed an incremental majority filter, which captures the most frequently predicted port instead of the last predicted one.

Artificial neural networks was applied to trajectory prediction using AIS data [25,33,37,38]. Chen *et al.* [7] highlighted the noise issue affecting the quality of AIS data and proposed a method to predict trajectories using neural networks. They followed a three-step pre-processing approach: *i)* organised the data, *ii)* removed outliers and *iii)* normalised the data into data samples using cubic spline interpolation and a moving average model. Their study is limited in the number of trajectories and vessels used to validate the results (only two). Zhang *et al.* [34] proposed an ensemble learning model for AIS trajectory prediction using a 200-segments sample from AIS data. They trained models on clusters of patterns to improve the prediction accuracy, where each cluster represents a boat trajectory. Similarly, Suo *et al.* [33] presented a real-time ship track prediction model using different recurrent neural network (RNN) architectures [13], focusing on the port of Zhangzhou in China. The authors showed that the vanilla RNN had similar accuracy to that of long short-term memory (LSTM) architecture [17], while the gated recurrent unit (GRU) [9] model outperformed the LSTM in terms of computational time. Wang *et al.* proposed a trajectory prediction model for multiple vessels simultaneously sharing the same area. The authors used a generative adversarial network with attention and interaction module [15]. They improved the accuracy compared to sequence to sequence, plain GAN, and the Kalman models by a minimum of 20%.

More recently, graph-based models have been proposed to improve predictive outcomes by representing data as a graph, such as the work in [26]. Carlini *et al.* [6] presented a network analysis using an AIS dataset to build a set of voyage graphs and capture the evolution of networks based on several topological features. Another example of a graph-based method is the work proposed by Magnusen *et al.* [26]. The authors represented the sea traffic in a graph, where vertices represent sea areas that can be a turning or staying point, and links are created by splitting a trajectory into several sub-trajectories. A port-to-port trajectory is described in this work by a sequence of vertices used to train a recurrent neural network model to predict destinations for oil tankers on both port and regional levels. The proposed model achieved 41% accuracy when predicting destination ports versus 87% predicting regions.

To the best of our knowledge, little attention has been given to heterogeneous graph methods for voyage destination prediction, despite being a powerful framework for modelling maritime navigation networks and capturing relations between heterogeneous entities, such as ports and vessels. Furthermore, the graph modelling approach allows the destination prediction task to be designed as a link prediction algorithm, which is also new and little explored.

3 Methods

This section describes the AIS data pre-processing, voyage creation algorithms, the voyage destination prediction model, and the evaluation approach. We modeled the destination prediction problem as a multi-classification problem to predict the destination port. Given a vessel, a departure port, and the list of destination ports available in the network, the algorithm predicts the most likely destination for the ship. We will first describe the cleaning, filtering, and organizing methods applied to AIS data. This process is critical, particularly as we cannot use the destination information found in AIS messages as a validation gold-standard [27,42]. Therefore, we propose a heuristic algorithm to create different voyages per vessel, and positional-based validated moored ports.

3.1 AIS Dataset

In our experiments, voyage segments were created using the publicly available historical AIS data from the Danish maritime authority website (www.dma.dk). This dataset covers the region around Danish waters. Nevertheless, destination ports can cover ports outside of the specified area. We have processed a snapshot from January 2014 until March 2021 containing around 10TB. However, for computational reasons, we are using a randomly generated sample containing 2757 tanker vessels, 58690 voyage samples, and 620 ports (see Table 1). We have focused on ships of type tankers due to their high rate of data completion and availability for most attributes.

Table 1. Statistics for the training and test sets.

	Training	Test
Number of vessels	2399	1713
Number of unique source ports	539	413
Number of unique destination ports	499	391
Number of segments	35214	11738
Median segments per vessel	3	2
Minimum segments per vessel	1	1
Maximum segments per vessel	5614	1897

3.2 AIS Pre-processing Approach

In this work, we define a vessel voyage segment as a voyage from a source port A to a destination port B and describe every voyage by a unique id, departure date and port, and arrival date and port. We used the attributes of AIS messages, such as coordinates, speed, and navigational status (under way using engine, at anchor, moored, etc.) to generate the voyages. Speed is used since ships will slow down when approaching a port and then stop at the voyage destination.

AIS historical data offer numerous dynamic attributes related to voyages, such as draught, estimated arrival time, and destination. However, as such data are entered manually, human errors often occur. We defined vessel stops using different AIS attributes such as speed and position to determine the actual moored ports to overcome this issue. Additionally, we used the World Port Index (WPI) 2019 database [40] to link vessel positions to the closest ports.

Draught is the only AIS data that provides information about the activity of a ship in a port. If it increases, the boat is heavier and therefore loaded commodities in the port. If it decreases, the boat unloaded in the last port. While we cannot trust the value of the draught at every signal as it is entered manually, every ship must have the correct value of draught when entering a port. Thus, at every stop, we get the valid value of the draught related to the previous voyage and correct the draught value if it is different.

3.3 Voyage Creation

To generate port stops and construct voyages for every vessel, we calculated the distance between every vessel stop position and ports listed in the WPI using the Haversine Formula [8]. The port with the minimum distance to the vessel position is defined as the closest (moored) port. Then, using the nearest defined ports, we create the voyage following Algorithm 1. For each ship, we traverse its positions. Based on the speed of the vessel at every position, we predefined $VesselMoving$. If the boat has stopped ($VesselMoving = 0$), we define the current timestamp as the date and time of arrival and the nearest port as the arrival port of the current voyage. Once the vessel starts moving away from the current port ($VesselMoving = 1$), we define it as the departure port of the next voyage and set the departure date as the current timestamp.

To avoid ill-defined segments, e.g., as a result of ships travelling outside the coverage area, an empirical 12 min no-signal threshold is defined (that is, twice the maximum time span of shared static data). A voyage is then suppressed if the time interval between two consecutive signals exceeds the defined threshold.

Algorithm 1. Voyage creation

```
For Each Vessel
  VesselDeparted = 0
  InPort = 0
  For Each VesselPosition
    If VesselMoving == 0
      If (VesselDeparted == 0) and (InPort == 0)
        Assign DepartPort←ClosestPort
        InPort = 1
      Else If VesselDeparted == 1
        Assign ArrivalPort←ClosestPort
        Assign ArrivalDate←TimeStamp
        VesselDeparted = 0
    Else If InPort == 1
      Assign DepartDate←TimeStamp
      VesselDeparted = 1
      InPort = 0
  EndFor
EndFor
```

3.4 Proposed Graph-Based Machine Learning Model

The heterogeneous graph abstraction proposed to model the maritime transportation network is described in Fig. 1. A heterogeneous graph is denoted by $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A}, \mathcal{R})$ where $\mathcal{V}$ and $\mathcal{E}$ denote the node and link sets, respectively. Each node $v, p \in \mathcal{V}$ and each link $e \in \mathcal{E}$ is associated with a mapping function, where $\phi(v) : \mathcal{V} \leftarrow \mathcal{A}$ and $\rho(e) : \mathcal{E} \leftarrow \mathcal{R}$ represent the node mapping function and edge mapping function, respectively. A graph is defined as heterogeneous if it contains more than one node type and/or more than one edge type. Therefore, $\mathcal{A}$ and $\mathcal{R}$ denote the sets of node and edge types satisfying $|\mathcal{A}| + |\mathcal{R}| > 2$.

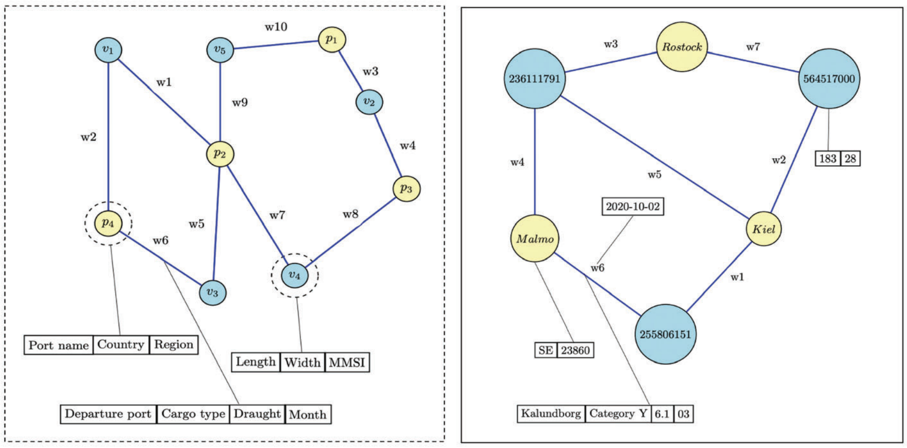
Figure 1a shows that nodes represent vessels (blue) and ports (yellow), while edges w_i define links between the vessel node v_i and a destination port node p_j for a specific voyage. Vessel nodes are described by three features - length, width, and MMSI - while port nodes' features include port name, country, and region id information. On the other hand, link features describe specific vessel voyage information, including the departure port, month, draught, and cargo type. The departure time is added as weights to the link to represent voyages of the same vessel with the same source-destination occurring at different dates.

Figure 1b shows a real example of seven voyages related to three vessels and three ports. The vessel with MMSI *255806151* is traveling in *March* (03) from *Kalundborg* to *Malmo* with a cargo type of *Category Y* and a draught of value *6.1*. Each node type is defined by different features. Port node *Malmo* is described by the country *SE*, which represents Sweden and a region id *23860*. Vessel node *564517000* is described by its length (*183* m) and width (*28* m). The weight of each link is defined by the departure date of a voyage. For example, w1 is the weight of the link representing a voyage of the boat with MMSI *255806151* traveling on the *2020-10-02* to the *Kiel* port.

Following the methodology described in [12,28], we use word2vec [30] to perform link prediction task. To create a low dimensional representation of a node, that is, a node embedding, random walks are computed using the heterogeneous graph model. The node embedding shall ensure that the distance between nodes is preserved in the embedding space. If two nodes are close to each other in the graph, their closeness shall be maintained in the embedding space. The resulting list of paths created by the random walk for a node is then provided to a word2vec model to generate the node embedding for the respective nodes. Then, using the vessel and destination node embeddings, link embedding is computed for the voyage segments. Negative voyage segments are randomly generated using possible vessel-port connections available in the network to provide negative examples to the learning algorithms. Link embeddings are similarly created for the negative samples. Finally, link embedding is used to train the predictive model. The entire destination prediction pipeline is shown in Fig. 2.

3.5 Experiments

We divided the data into training, dev, and validation sets (60% training, 20% dev, and 20% validation), as shown in Table 1, where the test set statistics include both dev and validation samples. Scikit-learn and Stellargraph were used to build the machine learning models. We use the Stellar graph library [11] to create the heterogeneous graph, and node and link embeddings. A k-nearest neighbors (kNN) algorithm was used as the machine learning model for our graph-based methodology (after an empirical comparison with other classic machine learning models). The graph-based model was compared to different traditional machine learning approaches (logistic regression, kNN, random forest, and Catboost [3, 10,31,41]) using only the vessel-, port- and voyage-related features, without the link embeddings. The experiments were conducted on a server with 40 Intel® Xeon® CPU E5-2690 v2 @ 3.00 GHz cores and 756 GB RAM.



(a) (b)

Fig. 1. A heterogeneous graph with blue and yellow nodes referring to vessels and ports respectively, and edges between vessels and destination ports. (a) Feature types related to vessels, ports and voyages are shown for nodes p4 and v4, and for edge w6, respectively. (b) A real example of seven voyages for three vessels and three ports. (Color figure online)

4 Results

Macro-averaged results for the destination prediction models are presented in Table 2. As we can notice, the graph-based machine learning model outperforms all the classic models that do use graph-based features. It outperforms the best baseline algorithm - random forest - with an increase in accuracy of approximately 3%, the precision almost doubled, going from 34% to 69%, and recall and F1-score improved by 7% and 23% respectively. While our best model is able to predict the correct destination port nearly 70% correct, it is only able to do so for around 1/3 of the ports in the network. If we only compare among the baseline models, the random forest algorithm performs the best, with an accuracy of 61%, precision of 34%, recall of 29%, and F1-score of 31%, followed closely by Catboost. Surprisingly, the logistic regression, despite being a strong classification method, performs the worst.

Lastly, we can verify the power that the graph-based features bring to the model by comparing the performance of the kNN model (without graph-based features) and our model, that is, a kNN enhanced with graph-based features. As we can see from Table 2, there is a significant increase in precision, more than doubling with a subsequent impact on the F1-score (which is also almost double). We believe that the addition of the topological features derived from the heterogeneous graph are thus able to better characterise a voyage segment.

Table 2. Destination prediction models results

Model	Accuracy	Precision	Recall	F1-score
Logistic regression	0.5574	0.2225	0.2211	0.2217
kNN	0.5822	0.2995	0.2774	0.2880
Catboost	0.6036	0.3136	0.2620	0.2854
Random forest	0.6133	0.3369	0.2873	0.3101
Graph-based (ours)	0.6472	0.6877	0.3604	0.5426

4.1 Comparison with AIS Manually Entered Destinations

To have a better reference, we compare the destination information available in the AIS message with the destination derived by our voyage reconstruction algorithm, which uses automatic AIS position and speed data, and an external port database (WPI). Before comparing the datasets, we cleaned AIS destinations by removing samples with meaningless or random destination values, such as *HERE WE GO AGAIN*, *HOME*, etc. We also created various rules to link AIS destination codes with WPI port names. We can cite examples of ports in AIS data with values *SEGOT* and *SE GBG* destinations, both equivalent to *GOTEBORG* in the generated voyage dataset.

The resulting comparison shows that we covered around 48% of AIS destination ports, which means that we generated the same AIS destination for almost half of our data. As a comparison, the graph-based model is accurate in 65% of its predictions. If we relax the matching process between AIS and the port names of WPI, using a fuzzy search with a minimum similarity percentage of 90%, still only 55% of AIS destinations matches, which is a similar accuracy to the worst baseline model. These can be explained by the high risk of errors within manually entered AIS data.

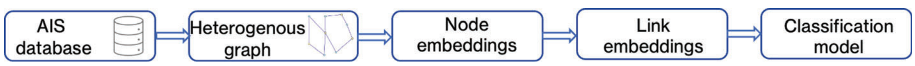


Fig. 2. Overview of the proposed graph-based voyage destination prediction model.

4.2 Error Analysis

We present a summary of prediction results for our graph-based model in the confusion matrix of Fig. 3. Due to the high number of ports available in the test set, we show the results only for the 10 top destination ports. We added "Other ports" to represent any port that is not in the list of ports displayed in the confusion matrix. As we can notice, most of the confusions of the top ports are with ports lower-destination ports ("Other ports"), e.g., *Nysted*, *Rostock*, etc. Among the top ports, *Karsto* is confused often with *Nykobing (MOR)* (20%), and *Skudeneshavn* with *Karsto* (22%) and *Nykobing (MOR)* (22%). We believe this might be due to the fact that *Skudeneshavn* and *Karsto* are very close geographically and also visited by the same boat *257144700* in Fig. 4.

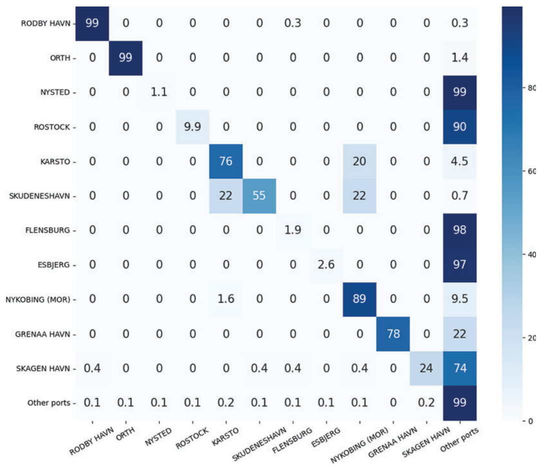
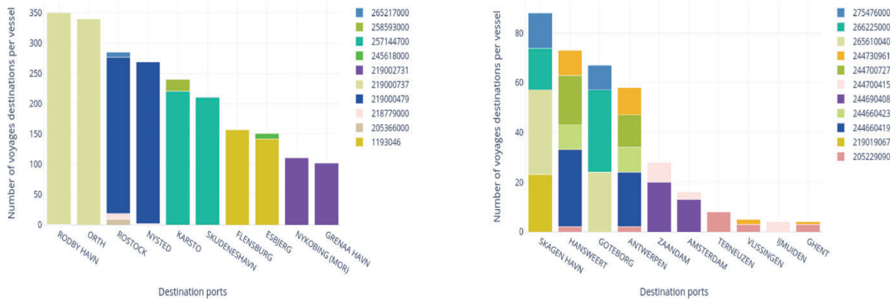


Fig. 3. Link prediction confusion matrix for the top destination ports. "Other ports" represent the remainder predictions.

Figure 4a shows an overview of the distribution of data related to the top 10 destination ports with the highest number of voyages in test set. We notice that vessel with MMSI 219000737, during 2000 voyages, has been visiting *Rodby Havn* port 50% of the time and *Orth* port the rest of the voyages. This means that the probability of predicting the right port is 0.5. The dominant ports have little diversity in terms of visiting boats. Only three of the five ports have been visited by one to three boats. This explains why ports, such as *Rodby Havn* and *Orth*, have such good performance as shown in Fig. 3. Figure 4b shows the distribution of data in the test set related to the top destinations that a maximum number of vessels has visited. We can see that for ports such as *Skagen Haven*, a much higher confusion is found (see Fig. 3), indicating that the algorithm is biased towards the majority classes, having significantly better performance for ships that always go to the same destination than for ships that make voyages to different destinations.



(a) (b)

Fig. 4. Test set. (a) Number of port destinations per vessel for the top 10 destinations with the maximum number of voyages. (b) Number of port destinations per MMSI for top destinations shared by a maximum number of vessels.

5 Discussion

In this paper, we designed a novel method based on graph-based machine learning models to predict voyage destination. Such methods can be used to improve AIS data quality and promote better decision making in the maritime transportation. In addition to the usual tasks of cleaning and removing conflicting data, and filling in missing information wherever is possible, we created voyage segments by combining historical AIS data with a world port database (WPI). In our experiments, we cover the region around Danish waters, nevertheless, the methodology is readily applied to the world maritime transportation network.

In the proposed model, we organized AIS data by vessels and created voyages based on a set of rules using different AIS attributes combined with the WPI dataset. Then, voyages are abstracted using a heterogeneous undirected graph, which is used to train a machine learning model that solves the voyage destination port prediction problem as a multi-class classification based on a link prediction algorithm. The graph-enriched model was compared to baseline models that do not exploit the network properties, achieving significant performance improvement upon them. While more complex graph-based models exist, such as those based on graph neural networks, e.g., graph convolution neural network [21] and graph attention network [36], the objective here is to demonstrate that the topological features can contribute positively to the performance of the predictive models. Their investigation is left for a future work.

To represent the time factor in the proposed model, we conducted two experiments. First, the one presented in this paper, where time is added as a link property, that is, new travels have a higher weight. Second, we evaluated a recurrent model, where previous voyage features (time - n), e.g., previous voyage departure and destination ports, are added to the current set of features. However, adding recurrent information did not make any significant change to the performance of the models, only improved their complexity. Therefore, these results were not presented in this paper, but they could still be relevant in a wider coverage database.

Our work has some limitations. The AIS used only covers a region around the Danish waters. The destination mentioned can be outside the covered region. Therefore, we do not capture the position information of the ship at the final destination. Using data covering the whole world and considering other types of ships will increase the diversity of the data and improve the analysis of the ships' behaviours. Moreover, enhanced learning models, such as graph neural networks, as aforementioned, could be also analysed. Finally, some features, such as ships' id, while might help to enhance the learning of ships' behaviours, they open the risk of overfitting. Therefore, more robust evaluation methods, such as a cross-validation, could be employed.

To conclude, by approaching the voyage destination prediction problem as a multi-class link prediction task, we have explored the possibility of using the network features, such as link embeddings, in a graph-based model to improve the predictive power of learning algorithms. Despite the significant performance enhancement, voyage destination prediction remains as a challenge. Nevertheless, our results show that the performance of the graph-based predictive model outperforms the manually entered AIS destination data. Therefore, they could be used to augment AIS data quality and support data-driven maritime support systems.

References

1. AISM: IALA guidelines on the universal automatic identification system (2002). https://ucakikazlambasi.com.tr/wp-content/uploads/2017/04/iala-guidelines.pdf. Accessed 8 Jan 2022
2. Bailey, N.J.: Training, technology and AIS: looking beyond the box. In: SIRC Symposium 2005, Seafarers International Research Centre (SIRC), pp. 108–128 (2005)
3. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
4. Bye, R., Almklov, P.G.: Normalization of maritime accident data using AIS. Marine Policy (2019)
5. Cai, M., Zhang, J., Zhang, D., Yuan, X., Soares, C.G.: Collision risk analysis on ferry ships in Jiangsu section of the Yangtze river based on AIS data. Reliab. Eng. Syst. Safety **215**, 107901 (2021)
6. Carlini, E., de Lira, V.M., Júnior, A.S., Etemad, M., Machado, B.B., Matwin, S.: Uncovering vessel movement patterns from AIS data with graph evolution analysis. In: EDBT/ICDT Workshops (2020)
7. Chen, X., et al.: Ship trajectory reconstruction from AIS sensory data via data quality control and prediction. Math. Probl. Eng. 2020 (2020)
8. Chopde, N.R., Nichat, M.: Landmark based shortest path detection by using a* and haversine formula. Int. J. Innov. Res. Comput. Commun. Eng. **1**(2), 298–302 (2013)

AQ2

9. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555 (2014)
10. Cunningham, P., Delany, S.J.: k-nearest neighbour classifiers: (with python examples). arXiv preprint arXiv:2004.04523 (2020)
11. Data61, C.: Stellargraph machine learning library. Publication Title: GitHub Repository. GitHub (2018)
12. Dong, Y., Chawla, N.V., Swami, A.: metapath2vec: scalable representation learning for heterogeneous networks. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 135–144 (2017)
13. Elman, J.L.: Finding structure in time. Cogn. Sci. **14**(2), 179–211 (1990)
14. Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: KDD (1996)
15. Goodfellow, I., et al.: Generative adversarial nets. Adv. Neural Inf. Process. Syst. **27** (2014)
16. Harati-Mokhtari, A., Wall, A., Brooks, P., Wang, J.: Automatic identification system (AIS): data reliability and human error implications. J. Navig. **60**(3), 373–389 (2007)
17. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
18. Hu, Y., Park, G.K.: Collision risk assessment based on the vulnerability of marine accidents using fuzzy logic. Int. J. Naval Archit. Ocean Eng. **12**, 541–551 (2020)
19. Ifrim, C., Wallace, M., Poulopoulos, V., Mourti, A.: Methods and techniques for automatic identification system data reduction. In: Pop, F., Neagu, G. (eds.) Big Data Platforms and Applications. CCN, pp. 253–269. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-38836-2_12
20. IMO: AIS transponder (2020). https://www.imo.org/en/OurWork/Safety/Pages/AIS.aspx. Accessed 11 Dec 2021
21. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
22. Kisialiou, Y., Gribkovskaia, I., Laporte, G.: Robust supply vessel routing and scheduling. Transp. Res. Part C: Emerg. Technol. (2018)
23. Li, B., Lu, J., Lu, H., Li, J.: Predicting maritime accident consequence scenarios for emergency response decisions using optimization-based decision tree approach. Marit. Policy Manage. 1–23 (2021) AQ3
24. Lin, C.X., Huang, T.W., Guo, G., Wong, M.D.: MtDetector: a high-performance marine traffic detector at stream scale. In: Proceedings of the 12th ACM International Conference on Distributed and Event-Based Systems (2018)
25. Ma, S., Liu, S., Meng, X.: Optimized BP neural network algorithm for predicting ship trajectory. In: 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), vol. 1, pp. 525–532. IEEE (2020)
26. Magnussen, B.B., Bläser, N., Jensen, R.M., Ylänen, K.: Destination prediction of oil tankers using graph abstractions and recurrent neural networks. In: Mes, M., Lalla-Ruiz, E., Voß, S. (eds.) ICCL 2021. LNCS, vol. 13004, pp. 51–65. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87672-2_4
27. Mestl, T., Dausendschön, K.: Port eta prediction based on AIS data. In: 15th International Conference on Computer and IT Applications in the Maritime Industries, Lecce, pp. 9–11 (2016)
28. Metapath2Vec: Link prediction with metapath2vec (2019). https://stellargraph.readthedocs.io/en/stable/demos/link-prediction/metapath2vec-link-prediction.html. Accessed 10 Jan 2022

29. Mieczyńska, M., Czarnowski, I.: K-means clustering for SAT-AIS data analysis. WMU J. Marit. Aff. **20**(3), 377–400 (2021)
30. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
31. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A.: CatBoost: unbiased boosting with categorical features. arXiv:1706.09516 (2017)
32. Stogiannos, M., Papadimitrakis, M., Sarimveis, H., Alexandridis, A.: Vessel trajectory prediction using radial basis function neural networks. In: IEEE EUROCON 2021–19th International Conference on Smart Technologies. IEEE (2021)
33. Suo, Y., Chen, W., Claramunt, C., Yang, S.: A ship trajectory prediction framework based on a recurrent neural network. Sensors **20**(18), 5133 (2020)
34. Tu, E., Zhang, G., Mao, S., Rachmawati, L., Huang, G.: Modeling historical AIS data for vessel path prediction: a comprehensive treatment. CoRR abs/2001.01592 (2020). http://arxiv.org/abs/2001.01592
35. UN: Global issues (2019). https://www.un.org/en/global-issues/population
36. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
37. Wang, C., Fu, Y.: Ship trajectory prediction based on attention in bidirectional recurrent neural networks. In: 2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT). IEEE (2020)
38. Wang, S., He, Z.: A prediction model of vessel trajectory based on generative adversarial network. J. Navig. **74**, 1–11 (2021)
39. Wang, W., Zhang, C., Guillaume, F., Halldearn, R., Kristensen, T.S., Liu, Z.: From AIS data to vessel destination through prediction with machine learning techniques. Artif. Intell. Models, Algor. Appl. 1 (2021)
40. WPI: (2019). https://msi.nga.mil/Publications/WPI. Accessed 13 Nov 2021
41. Wright, R.E.: Logistic regression. In: Reading and Understanding Multivariate Statistics, pp. 217–244. American Psychological Association (1995)
42. Wu, L., Xu, Y., Wang, F.: Identifying port calls of ships by uncertain reasoning with trajectory data. ISPRS Int. J. Geo-Inf. **9**, 756 (2020)
43. Xiao, F., Ligteringen, H., Van Gulijk, C., Ale, B.: Comparison study on AIS data of ship traffic behavior. Ocean Eng. **95**, 84–93 (2015)
44. Yang, D., Wu, L., Wang, S.: Can we trust the AIS destination port information for bulk ships?-Implications for shipping policy and practice. Transp. Res. Part E: Logistics Transp. Rev. **149**, 102308 (2021)
45. Yang, D., Wu, L., Wang, S., Jia, H., Li, K.X.: How big data enriches maritime research-a critical review of automatic identification system (AIS) data applications. Transp. Rev. **39**(6), 755–773 (2019)
46. Zhang, C., et al.: AIS data driven general vessel destination prediction: a random forest based approach. Transp. Res. Part C: Emerg. Technol. **118**, 102729 (2020)
47. Zhang, T., Liu, C., Wen, B.: Abnormal ship behavior detection after the closure of AIS based on radar data (2021). https://doi.org/10.21203/rs.3.rs-551597/v1
48. Zhang, Z., Suo, Y., Yang, S., Zhao, Z.: Detection of complex abnormal ship behavior based on event stream. In: 2020 Chinese Automation Congress (CAC), pp. 5730–5735 (2020). https://doi.org/10.1109/CAC51589.2020.9327793
49. Zhao, L., Shi, G., Yang, J.: Ship trajectories pre-processing based on AIS data. J. Navig. **71**(5), 1210–1230 (2018)

Author Queries

Chapter 7

Query Refs.	Details Required	Author's response
AQ1	This is to inform you that corresponding author has been identified as per the information available in the Copyright form.	
AQ2	Kindly provide the volume number and page range for Refs. [7,22], and page range for Ref. [15] if applicable.	
AQ3	Kindly provide the volume number for Refs. [23,39], if applicable.	