



LJMU Research Online

Li, G, Liu, M, Zhang, X, Wang, C, Lai, KH and Qian, W

Semantic Recognition of Ship Motion Patterns Entering and Leaving Port Based on Topic Model

<http://researchonline.ljmu.ac.uk/id/eprint/23941/>

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

Li, G, Liu, M, Zhang, X, Wang, C, Lai, KH and Qian, W (2022) Semantic Recognition of Ship Motion Patterns Entering and Leaving Port Based on Topic Model. Journal of Marine Science and Engineering, 10 (12).

LJMU has developed **LJMU Research Online** for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

<http://researchonline.ljmu.ac.uk/>

Article

Semantic Recognition of Ship Motion Patterns Entering and Leaving Port Based on Topic Model

Gaocai Li ¹, Mingzheng Liu ² , Xinyu Zhang ^{1,*} , Chengbo Wang ¹ , Kee-hung Lai ^{2,*}  and Weihuachao Qian ¹

¹ Maritime Intelligent Transportation Research Team, Navigation College, Dalian Maritime University, Dalian 116026, China

² Shipping Research Centre, PolyU Business School, The Hong Kong Polytechnic University, Hung Hum, Kowloon, Hong Kong 999077, China

* Correspondence: zhangxy@dmlu.edu.cn (X.Z.); mike.lai@polyu.edu.hk (K.-h.L.)

Abstract: Recognition and understanding of ship motion patterns have excellent application value for ship navigation and maritime supervision, i.e., route planning and maritime risk assessment. This paper proposes a semantic recognition method for ship motion patterns entering and leaving port based on a probabilistic topic model. The method enables the discovery of ship motion patterns from a large amount of trajectory data in an unsupervised manner and makes the results more interpretable. The method includes three modules: trajectory preprocessing, semantic process, and knowledge discovery. Firstly, based on the activity types and characteristics of ships in the harbor waters, we propose a multi-criteria ship motion state recognition and voyage division algorithm (McSMSRVD), and ship trajectory is divided into three sub-trajectories: hoteling, maneuvering, and normal-speed sailing. Secondly, considering the influence of port traffic rules on ship motion, the semantic transformation and enrichment of port traffic rules and ship location, course, and speed are combined to construct the trajectory text document. Ship motion patterns hidden in the trajectory document set are recognized using the Latent Dirichlet allocation (LDA) topic model. Meanwhile, topic coherence and topic correlation metrics are introduced to optimize the number of topics. Thirdly, a visualization platform based on ArcGIS and Electronic Navigational Charts (ENCs) is designed to analyze the knowledge of ship motion patterns. Finally, the Tianjin port in northern China is used as the experimental object, and the results show that the method is able to identify 17 representative inbound and outbound motion patterns from AIS data and discover the ship motion details in each pattern.

Keywords: AIS; trajectory segmentation; semantic transformation; topic model; motion pattern recognition



Citation: Li, G.; Liu, M.; Zhang, X.; Wang, C.; Lai, K.-h.; Qian, W. Semantic Recognition of Ship Motion Patterns Entering and Leaving Port Based on Topic Model. *J. Mar. Sci. Eng.* **2022**, *10*, 2012. <https://doi.org/10.3390/jmse10122012>

Academic Editor: Mihalis Golias

Received: 29 November 2022

Accepted: 14 December 2022

Published: 16 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Shipping is the predominant mode of international trade, accounting for 90% of world trade [1]. With the continuous growth of international trade, the number and size of ships and marine infrastructure are increasing, making the maritime traffic environment increasingly severe. Once a maritime accident occurs, it will cause not only serious economic losses and casualties, but will also bring devastating damage to the marine ecological environment [2–4]. Recognition and understanding the movement patterns of ships play an important role in building intelligent, safe, and green shipping, such as ship trajectory prediction [5], maritime risk assessment [6,7], route customization and optimization [8], etc.

When a ship is at sea, it is subject to a combination of its own inherent properties (ship type, ship size, tonnage, etc.), the external environment (traffic rules, channel scale, weather, etc.), and the pilot's ship-handling skills, making it difficult to provide a complete and accurate description of the ship's movements [9]. Traditional methods of ship behavior analysis are mainly achieved through ship traffic surveys. This method cannot record ship traffic information in a large scale and sustainable manner and is characterized by low

efficiency and high cost in information acquisition. With the widespread deployment of AIS (Automatic Identification System) and VTS (Vessel Traffic Service), real-time supervision and service of ship navigation is realized. Massive, multidimensional ship trajectory data are collected, providing new ideas and means to uncover useful ship motion patterns. A number of quantitative methods have been developed to identify ship motion patterns from ship trajectories. These include statistical methods [10], clustering algorithms [11,12], deep learning [13], visual analysis [14], and semantic analysis [15,16].

Statistical methods are used to analyze and study trajectory data using statistical methods to obtain ship movements obeying some probability distribution, such as the Gaussian model, kernel density estimation, etc. Kowalska et al. [17] proposed a data-driven Bayesian model based on the Gaussian process and used the model to learn the motion behavior of model ships from AIS data. Laxhammar et al. [18] proposed and implemented a multivariate mixture Gaussian model to extract normal ship traffic patterns. The clustering method was used to classify ship trajectories into different movement patterns based on the distance between trajectories. Li et al. [19] proposed an improved spectral clustering algorithm, which can effectively and robustly extract the movement patterns of ships in the waters of Chengshantou. Li et al. [20] proposed an improved DBSCAN clustering algorithm for clustering ship trajectories in inland waters to extract typical ship motion patterns. Deep learning has a strong ability to capture data features, making deep learning models applied to trajectory pattern extraction research. Chen et al. [13] proposed a convolutional neural network ship movement classification algorithm (CNN-SMMC) by transforming different motion trajectories into images and labeling them, using a convolutional neural network model to learn and can recognize the trajectory as the corresponding movement patterns. Zhang et al. [21] established a convolutional neural network-based shape recognition model to identify a variety of typical ship drifting behaviors. The methods above were able to classify different trajectory cluster classes from the trajectory dataset. However, the models were usually constructed by considering only the motion information in the trajectory data and lacking consideration of the external environment, resulting in difficulty in understanding the motion patterns.

Visualization analysis is the application of various visualization elements and representations to explore potential pattern knowledge from trajectories [22]. Jin et al. [23] proposed a visualization analysis framework to extract features of ship motion from trajectory data in an interactive manner. Li et al. [24] proposed an interactive visual analysis method with multiple views to explore and understand the pattern knowledge of ship behavior in the Yangtze estuary. The visualization analysis provides a more intuitive understanding of pattern knowledge, but the method also struggles to consider the influence of external factors, and the method relies on human–interaction operation.

Semantic analysis can add ship motion-related information, such as traffic rules, and navigation environment, to the model to identify ship motion patterns [15,16]. In this context, ship motion is represented as a series of concepts with attributes and logical relationships by constructing a knowledge database of ship motions. The datAcron project, as part of the European Union's Horizon 2020 Program, has proposed the datAcron semantic model and applied it to fishing monitoring [25]. Yan et al. [26] proposed a method for extracting ship-stopping patterns by combining trajectory features and geographic scene semantics. Based on random forest method, a classification model of the stopping behaviour patterns of ships at berth and anchorage was constructed to achieve stopping pattern recognition. Although the semantic analysis method is able to add the external environment to the ship motion model, the method requires expert knowledge in building the knowledge base and consumes a lot of human and material resources.

Semantic recognition is one of the key components of natural language processing technology. Semantic recognition techniques can analyze large amounts of text data from web pages, emails, and social media to achieve an understanding of the meaning of words and the interpretation of the ideas represented by the word in paragraphs and chapters. Topic modeling is an important approach in semantic recognition technology and has

been widely used in industries such as healthcare, education, and finance. The model maps the high-dimensional word space to the low-dimensional topic space through a probabilistic generative model, effectively discovering the underlying structure and hidden semantic information of the document [27]. Recently, the topic modeling approach has been successfully applied to the study of traffic pattern exploration [28,29]. Liu et al. [30] proposed a spatial interaction pattern recognition method for vehicle movements in urban road networks. The “strokes” (i.e., natural streets) were chosen as the geographical units to represent vehicle moving paths, and analogies between geographical elements (i.e., stroke, moving path) and textual elements (i.e., word, document) were established to apply the topic model to identify the spatial interaction patterns of road networks. To balance modeling efforts with the interpretability of movement patterns, Huang et al. [31] used a topic model to identify ship movement patterns in inland waters. A semantic transformation method was used to convert the ship’s position, heading, and speed characteristics into the corresponding motion semantic information, construct a text document of the ship’s trajectory, and mine the potential ship motion patterns with the help of the topic model. However, the study ignored the influence of external factors on ship motion, especially the influence of traffic rules on ship motion, in the process of semantic transformation of trajectories. In port waters, maritime authorities usually divide port waters into different functional waters in order to maintain traffic order and improve traffic efficiency, and they require ships to obey traffic rules to navigate in specific waters. Therefore, when studying the behavior of ships in port waters, port traffic rules need to be taken into account as an important factor.

In summary, in order to fill the available research gaps revealed in the literature, this study proposes a semantic recognition method for ship motion patterns entering and leaving port based on a probabilistic topic model. Unlike previous studies, the influence of port traffic rules on ship motions is considered in the semantic transformation, and the multidimensional ship motion information is semantically transformed and enriched. In topic modeling, two indicators, topic coherence and topic correlation, are introduced to optimize the number of topics to achieve the extraction of typical ship motion patterns from a large amount of AIS data. In addition, a ship traffic visualization and analysis platform based on ArcGIS and Electronic Navigation Charts (ENCs) is constructed to realize the spatial visualization and analysis of ship motion patterns. The research contributions of this paper can be summarized as follows.

- A semantic recognition method for motion patterns of ships entering and leaving port based on a probabilistic topic model is proposed, and the effectiveness of the method is verified using real port data.
- Based on the types and characteristics of ship activities in harbor waters, a multi-criteria ship motion state recognition and voyage division (McSMSRVD) algorithm is proposed to identify and divide the raw ship trajectory into sub-trajectories of three motion types: hoteling, maneuvering, and normal-speed sailing.
- In semantic modeling, the influence of port traffic rules on ship movement is considered to construct a ship trajectory semantic document. At the same time, two metrics, topic coherence and topic correlation, are introduced to optimize the number of topics.
- A ship traffic visualization and analysis platform based on ArcGIS and ENCs was constructed to realize the visualization and analysis of ship movement characteristics.

The remainder of this paper is organized as follows. The framework and principles of the proposed method are presented in Section 2. Case validation and result analysis are presented in Section 3. Conclusions and research perspectives are given in Section 4.

2. Semantic Recognition Method for Ship Motion Patterns Entering and Leaving Port

2.1. Methodological Framework

In this paper, a semantic recognition method of ship motion patterns entering and leaving the port based on a probabilistic topic model is proposed, as shown in Figure 1. The method consists of three modules: trajectory preprocessing, semantic process, and knowl-

edge discovery. In the preprocessing module, a multi-criteria ship motion state recognition and voyage division (McSMSRVD) algorithm is designed by analyzing the activity types and characteristics of ships in port waters, realizing the division of ship trajectories into three sub-trajectories of hoteling, maneuvering and normal-speed sailing, and dividing normal-speed sailing trajectory sets into different voyages. In the semantic process module, considering the influence of port traffic rules on ship motion, the normal-speed sailing trajectory is semantically transformed and enriched, and the ship trajectory document set is constructed. The LDA topic model method is used to discover the semantic structure of the trajectory document set. Two indicators, topic coherence and topic correlation, are introduced to select the optimal number of topics. In the knowledge discovery module, the modeling results are semantically analyzed, and the details of ship motion entering and leaving the harbor are better understood by visualization with the help of ArcGIS and ENCs.

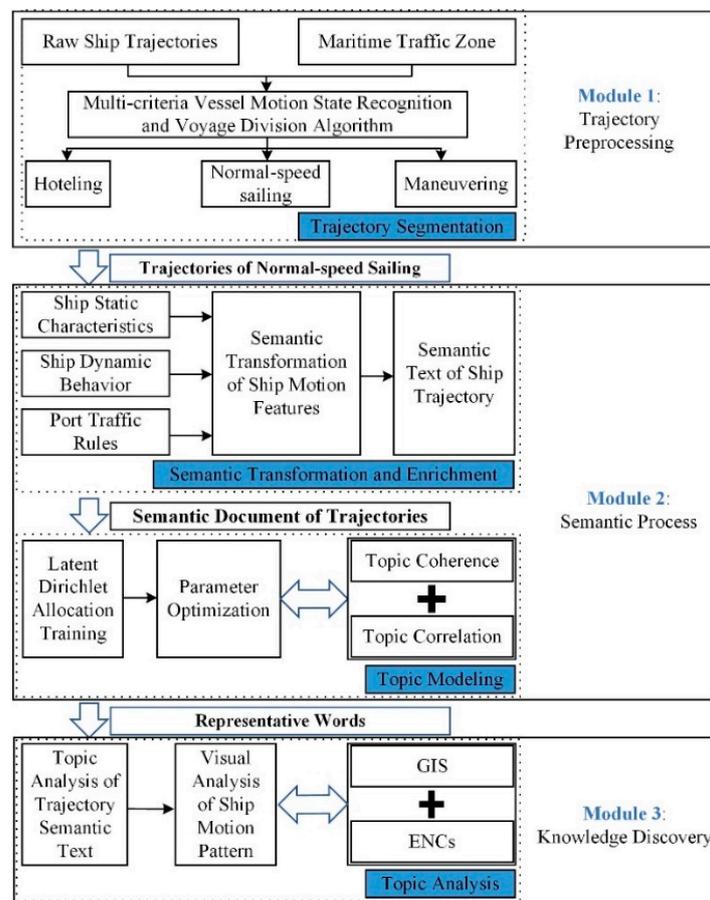


Figure 1. Framework of the semantic recognition method for ship motion patterns entering and leaving the port.

2.2. Ship Entering and Leaving Port Trajectory Segment Extraction

To recognize the ship motion patterns entering and leaving the harbor, the ship voyage data at this stated needs to be extracted. By analyzing the types of ship motion in the harbor waters and their characteristics, the differences in the spatiotemporal characteristics of ships in different motion states are clarified. Then, the identification method of ship motion states is proposed, and the classification of ship voyages is realized.

2.2.1. Types and Characteristics of Ship Activities

The activity chain of a ship generally includes three types of activities: hoteling, maneuvering, and normal-speed sailing. The voyage of a ship is one trip from one maritime

traffic zone (MTZ) to another traffic zone [32]. The MTZ includes two types of waters for seaports: basin and anchorage. The trajectory within the navigable waters is the route that connects the two traffic zones. Figure 2 depicts the general process of a ship entering and leaving a harbor. In the hoteling state, the ship is usually fixed with cables to receive services such as loading and unloading, waiting at anchor, or other. The ship will move slightly due to the wind and wave effect, and its speed will generally be less than 0.5 knots. The maneuvering state means that the ship moves slowly with the help of tug ships, which is different from entirely relying on the ship’s main engine to achieve normal-speed sailing. Therefore, the ship’s speed in the maneuvering state will be less than the speed of normal-speed sailing. Generally, the ship’s speed in the maneuvering state will not be more than 3 knots.

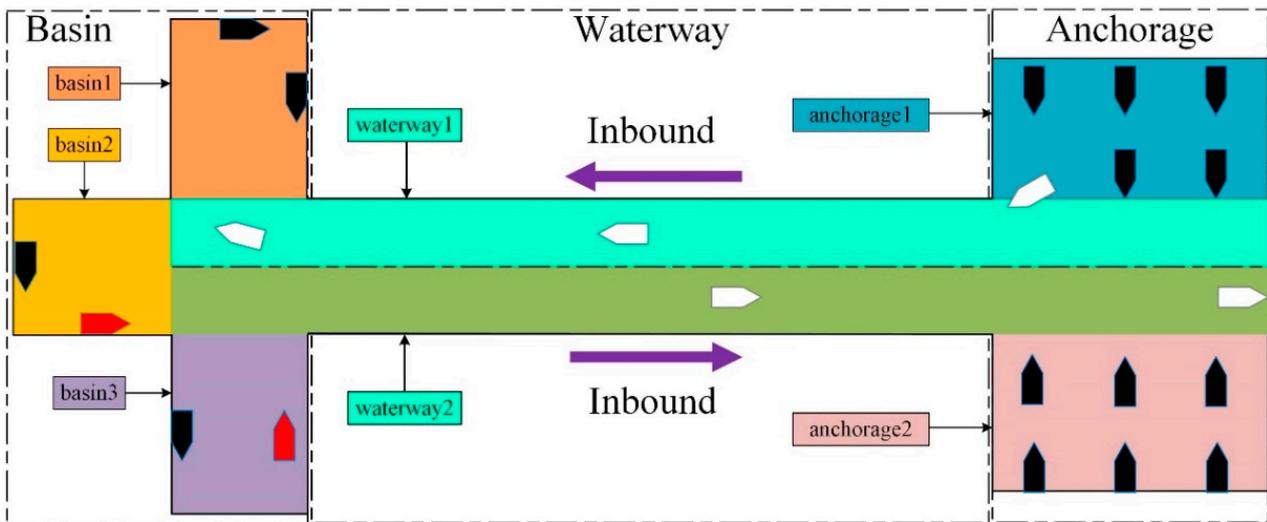


Figure 2. Schematic diagram of the flow of ships entering and leaving the port. Ships with red marks are in a maneuvering state; ships with white marks are in a normal-speed sailing state; ships with black marks are in a hoteling state.

2.2.2. Ship Motion States Recognition and Voyage Division

AIS automatically sends and receives information related to ship navigation, including static and dynamic information about the ship [33]. Table 1 shows the primary information of AIS. The AIS equipment automatically adjusts according to the motion state of the ship. The frequency of sending and receiving dynamic information of Class-A AIS is shown in Table 2. As mentioned above, when the ship is in the hoteling and maneuvering state, the ship speed is generally less than 3 knots, and the update frequency of dynamic information is 3 min. While the ship is in the normal-speed sailing state, the ship’s speed will be more incredible than 3 knots, and the update frequency of dynamic information will be significantly less than 3 min. Therefore, 3 min can be used as an essential discriminating basis between the ship’s normal-speed sailing and maneuvering state.

Table 1. Example of AIS sample data.

Date	MMSI	Type	Length	Longitude	Latitude	Direction	Speed
1,472,691,694	xxxxxx230	Cargo	114	118.0683	38.9092	317.1	3.1
1,633,518,105	xxxxxx950	Tanker	138	117.9754	38.8992	347	10.3
1,633,772,063	xxxxxx000	Cargo	197	118.0103	38.9150	311.2	7.8

Table 2. Update frequency of Class-A ship dynamic information [33].

Motion States	Update Frequency
Ship is anchored or moored and less than 3 knots	3 min
Ship is anchored or moored and greater than 3 knots	10 s
Ship with a speed between 0 and 14 knots	10 s
Ship with a speed between 0 and 14 knots and is turning	3 1/3 s
Ship with a speed between 14 and 23 knots	6 s
Ship with a speed between 14 and 23 knots and is turning	2 s
Ship with a speed greater than 23 knots	2 s
Ship with a speed greater than 23 knots and is turning	2 s

However, traffic flow, weather conditions, and operators easily affect AIS information transmission, resulting in missing phenomena and increasing the interval between adjacent AIS data [34]. Therefore, relying only on the update frequency as the basis for ship state recognition can easily lead to misclassification. After analyzing the ship activity type and its characteristics, we know that the ship activity state is closely related to the type of water in which the ship is located, the ship speed, and the dynamic information update frequency. Integrating the above multiple factors, we clarify three ship motion states and subdivide five kinds of ship motion in different environments, as shown in Table 3. Based on the above analysis, we propose a multi-criteria ship motion state recognition and voyage division algorithm (McSMSRVD). The algorithm combines the geographical information of the port traffic zone with AIS data, that is, ship location, the speed, and the update frequency as the conditions for recognition motion states. At the same time, it considers that the same ship may enter and leave the port several times, which means it has multiple voyages. In order to achieve the accurate division of different voyages, we divide different voyages by comparing the time interval of adjacent trajectory points with the voyage interval threshold Δt . The pseudocode for ship motion status identification and voyage extraction is given in Algorithm 1.

Table 3. Ship motion characteristics in port waters.

Time Interval	Maritime Traffic Zone	Speed/ <i>kn</i>	Motion State
≥ 3 min	Yes	[0, 0.5]	Hoteling
]0.5, 3]	Maneuvering
]3, +∞]	Normal-speed navigation
< 3 min	No]0, +∞]	Normal-speed navigation
	/]0, +∞]	Normal-speed navigation

2.3. Semantic Transformation and Textural Representation of Ship Trajectories

Topic modeling uses textual documents composed of meaningful words as data sources. However, in the raw trajectory data, the ship location, speed, and course are in numerical form. It is necessary to take semantic transformation processing for the raw trajectory. In harbor waters, ship navigation is not only influenced by its inherent characteristics but also constrained by port traffic rules. In order to comprehensively represent the semantic information of ship motion in the process of entering and leaving the harbor, this paper will combine three kinds of motion information and traffic rules, namely location, direction, and speed, to represent the semantic information of ship motion.

Algorithm 1 Multi-criteria Ship Motion State Recognition and Voyage Division Algorithm**Input:** T (raw trajectories), MTZ (maritime traffic zone), Δt (time interval threshold)**Output:** $Norsail_set$ (normal-speed sailing activity), $Mauver_set$ (maneuvering activity), $hoteling_set$ (hoteling activity)**Initialize:** $Norsail_set \leftarrow \emptyset$, $Mauver_set \leftarrow \emptyset$, $hoteling_set \leftarrow \emptyset$, $k \leftarrow 1$

```

[1] Sort  $T$  by MMSI and Date in ascending order.
[2] Extract and deduplicate MMSI to form ship  $MMSI\_set$ .
/* Motion state recognition */
1: for each MMSI in  $MMSI\_set$  do
2:   [3] Extract all the data of MMSI in  $MMSI\_set$  to  $temp$  and calculate  $temp$  size  $m$ .
3:   for  $i = 1$  to  $m$  do
4:     if  $temp[i + 1].Date - temp[i].Date \geq 180$  then
5:       if  $temp[i].location$  in  $MTZ$  then
6:         if  $temp[i].Speed \geq 0$  and  $temp[i].Speed < 0.5$  then
7:           [4] Add  $temp[i]$  to  $hoteling\_set$ 
8:         elif  $temp[i].Speed \geq 0.5$  and  $temp[i].Speed < 3$  then
9:           [5] Add  $temp[i]$  to  $Mauver\_set$ 
10:        else
11:          [6] Add  $temp[i]$  to  $temp\_Norsail\_set$ 
12:        end if
13:      else
14:        [7] Add  $temp[i]$  to  $temp\_Norsail\_set$ .
15:      end if
16:    else
17:      [8] Add  $temp[i]$  to  $temp\_Norsail\_set$ 
18:    end if
19:  end for
20: /* Voyage division */
21: [9] Sort  $temp\_Norsail\_set$  in ascending order of MMSI and Date, add a column of voyage, and
    calculate the size of  $temp\_Norsail\_set$  as  $n$ .
22:   for  $j = 1$  to  $n$  do
23:     [10]  $temp\_Norsail\_set[j].voyage = k$ 
24:     if  $|temp\_Norsail\_set[j + 1].Date - temp\_Norsail\_set[j].Date| \geq \Delta t$  then
25:       [11]  $temp\_Norsail\_set[j + 1].voyage = k + 1$ 
26:       [12]  $k = k + 1$ 
27:     else
28:       [13]  $temp\_Norsail\_set[j + 1].voyage = k$ 
29:     end if
30:   end for
31: [14] Add  $temp\_Norsail\_set$  to  $Norsail\_set$ .
32: [15]  $k = k + 1$ 
33: end for

```

In the process of semantic transformation, the spatial grid codes are created as so-called place names to describe the ship location semantic information. Fuzzy sets of velocity and direction based on the geographic information theory are designed to describe semantic information of speed and course. The ship type and functional water name are combined to represent the traffic rule semantic information, which enriches the ship's motion under traffic rules. In the textual representation of the trajectory, the transformed semantic information is constructed into a trajectory dictionary to generate a textual document of the ship's trajectory.

2.3.1. Semantic Transformation and Enrichment of Ship Motion Features

1. Ship Location Semantic

The location feature represents the position information of ships in navigable waters. The latitude and longitude coordinates in the trajectory represent the geospatial significance

of the ship, but it is not easy to intuitively understand the meaning of the ship’s motion. Therefore, the study waters are named by geocoding to describe the meaning of ship location features. Figure 3b shows that the waters are divided into $M \times N$ square grids. Each grid is recorded as G_{mn} , where m and n represent the grid’s row and column coding, respectively. The semantic information of the location feature can be represented by row encoding and column encoding, that is, $ST(lon_k, lat_k) = Row\#Col$. Assuming that the research waters is a rectangular area of $22\text{ km} \times 10\text{ km}$, the partition grid $1\text{ km} \times 1\text{ km}$, then the waters can describe 22×10 different semantic information of ship location features.

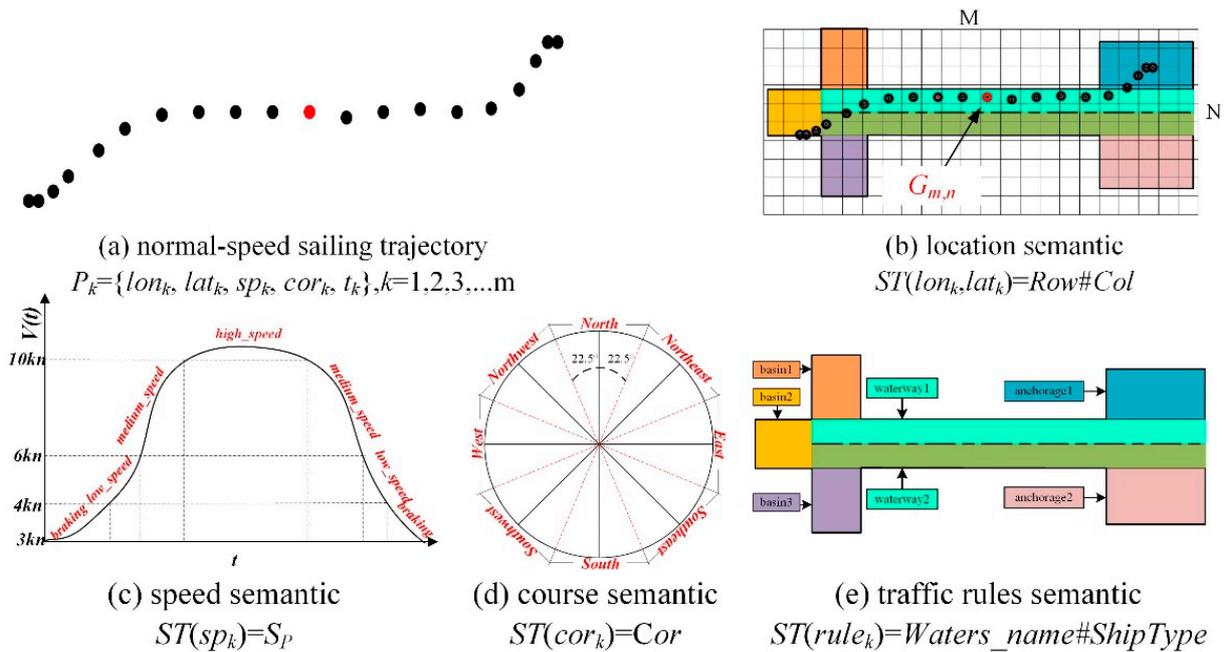


Figure 3. Semantic transformation and enrichment of ship motion features.

2. Ship Course Semantic

The course feature describes the navigation direction of the ship at the current moment. Similarly, the numerical direction angle cannot provide an intuitive perception of the course. To better describe the semantic information of the ship course, this paper utilizes a cone-based direction model to extract semantic information about course features. Figure 3d shows that the course is uniformly divided into eight cone directions, $Cor = \{North, Northeast, East, Southeast, South, Southwest, West, Northwest\}$. For a sample point, its course value will be compared to the angle range of each cone direction and labeled with the matched direction symbol. Then, the semantic description of the course feature can be represented as $ST(Cor_k) = Cor$.

3. Ship Speed Semantic

The speed feature describes how fast or slow a ship can travel in navigable waters. When entering the port, the main engine rotation speed is usually reduced step and step to lower the ship’s speed. According to the steering ability to control the ship’s direction, the speed-decreasing process is usually divided into four stages: braking stage, low-speed stage, medium-speed stage, and high-speed stage. The corresponding ship speed ranges are 3~4 knots, 4~6 knots, 6~10 knots, and more than 10 knots, respectively [35]. The ship outbound process is the opposite. Based on this characteristic, this study will use the discretization method to discretize the ship speed and represent it as $Sp = \{high_speed, medium_speed, low_speed, braking\}$, as shown in Figure 3c. For an arbitrary sample point, its speed value will be compared to the interval range of each velocity discretization and labeled with the matched speed symbol. Then the semantic description of the speed feature can be described as $ST(Sp_k) = Sp$.

4. Traffic Rules Semantic

Traffic rules are regulations and laws made by the traffic management departments to maintain traffic order, prevent and reduce traffic accidents, and improve traffic efficiency. In harbor waters, maritime authorities usually designate port waters as different functional waters and require ships to follow the traffic rules to navigate the specific functional waters. It is very difficult to directly describe that ships are affected by port traffic rules. By analogy with road traffic rules, pedestrians are required to walk on sidewalks and motor vehicles to drive on motor vehicle lanes. In order to describe the ship motion affected by the port traffic rules, this study will combine the ship type and the name of navigable waters passing through to describe the semantic information of the port traffic rules, that is, $ST(rule_k) = Waters_name\#ShipType$. As shown in Figure 3e, assuming that the port waters are designed with seven different waters, and that there are four types of ships, there will be 7×4 different semantic descriptions of traffic rules.

2.3.2. Textual Representation of Ship Trajectories

The raw trajectories represented by geometric representation are not suitable for the direct use of document-based topic modeling. Based on the semantic transformations method introduced in Section 2.3.1, we can obtain textual representations of ship trajectories by constructing a trajectory dictionary. The trajectory dictionary means that a series of motion words are used in writing to describe the motion state of the ship. In this study, each motion word is a comprehensive representation of ship location semantics, traffic rule semantics, ship speed semantics, and course semantics, namely $mw = Row\#Col\#Waters_Name\#ShipType\#Cor\#Sp$. Therefore, a total of $N_p \times N_r \times N_c \times N_s$ motion words are contained in the trajectory dictionary. N_p is the number of words describing the location feature of the ship, which is determined by the size of the study waters and the size of the space grid division; N_r is the number of words describing the port traffic rules, determined by the division result of the research waters and the classification of ships; N_c and N_s are the constant numbers of words depicting ship course features and speed features.

A trajectory document is a set of motion words describing a ship's continuous motion. This study takes each inbound and outbound voyage trajectory as the goal and creates the continuous motion within each voyage as a trajectory document. The document set is a collection of individual ship behaviors at different times and voyages. Assuming that the research waters include M trajectories of inbound and outbound voyages, that is, M trajectory documents, it can be expressed as $TD = \{td_1, td_2, \dots, td_M\}$. For the trajectory document td_i , it can be represented by limited and continuous motion words, $td_i = \{mw_1, mw_2, \dots, mw_N\}$.

2.4. Topic Modeling of Ship Trajectory Documents

Topic modeling is able to identify hidden semantic structures from large document sets. It considers a semantic feature space, the topic, between the word-space at the bottom and the document-space at the top. Therefore, topics have the same meaning as cluster centers in traditional clustering methods. A topic represents the main idea of a sentence, paragraph, or article, and is represented by a series of words. Each document D_i consists of a finite combination of multiple topics in the topic model, called the document-topic distribution $Di = \{T_1, T_2, \dots, T_m\}$. At the same time, each topic is composed of a distribution of finite words, called topic-word distribution $Ti = \{W_1, W_2, \dots, W_n\}$. During the modeling process, both document-topic and topic-word follow a specific probability distribution, and different words with high frequency in the document will be clustered into the same topic. The meaning of the topic depends on the meaning of the words. A trajectory can be seen as a document consisting of multiple motion topics for ship. A set of motion words can represent each topic. Here, the motion word represents a snapshot of the ship's motion state, and a set of snapshots can reflect the ship's motion pattern. Figure 4 reflects the analogy of discovering motion patterns from ship trajectories and learning topics from documents.

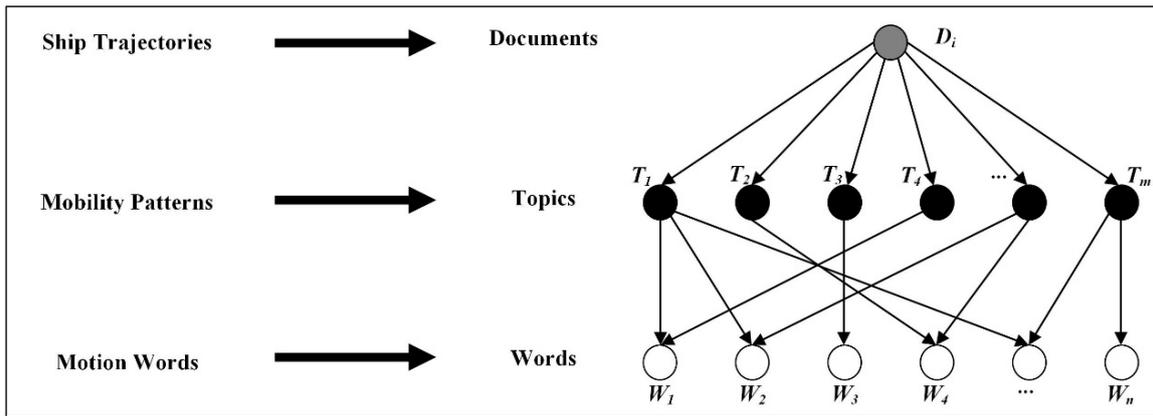


Figure 4. Analogy from trajectory-patterns discovering to document-topics learning [31].

2.4.1. Latent Dirichlet Allocation Model

The Latent Dirichlet Allocation (LDA) model is a classic, widely used, and effective probabilistic topic modeling method [27]. The model has the ability to capture multiple topics and efficiently derive topic features. The LDA model introduces a sparse Dirichlet distribution on the document-topic and topic-word distributions; that is, each document contains a small number of topics, and semantically related words can describe each topic. The graphical model of the latent Dirichlet distribution is shown in Figure 5. Latent variables are represented by single-circle nodes, while double-circle nodes represent observed variables. Boxes represent sets of variables. The specific and detailed generation process of the LDA model can be described as follows:

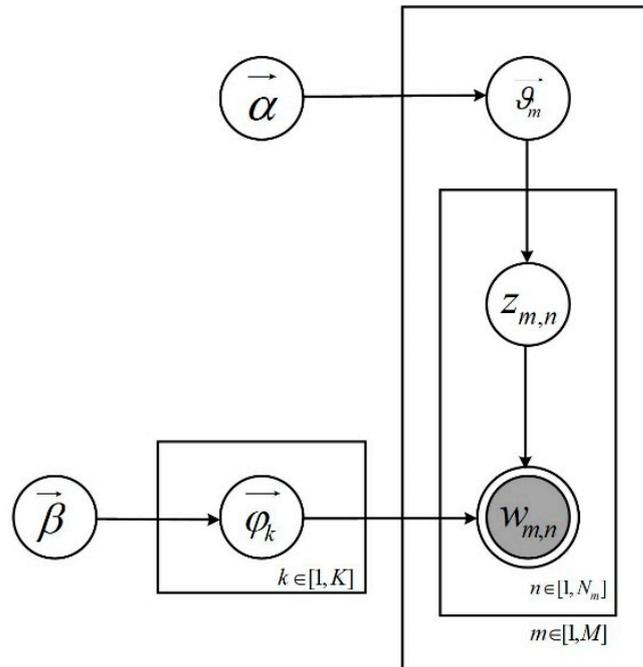


Figure 5. Geographical representation of the Latent Dirichlet Allocation model. K represents the number of topics, N represents the number of all words in the document corpus, and M represents the number of documents in the corpus. $\vec{\alpha}$ and $\vec{\beta}$ represent the Dirichlet prior parameters of the document-topic distribution and the topic-word distribution prior parameters, respectively, $\vec{\theta}_m$ represents the topic distribution for document m , which is a K -dimensional vector. $\vec{\varphi}_k$ represents the word distribution for topic k , which is an N -dimensional vector. $z_{m,n}$ is the topic of the n -th word in document m , and $w_{m,n}$ represents a specific word.

```

//topic plate:
For all topic  $k \in [1, K]$  do
Sample mixture components  $\vec{\varphi}_k \sim \text{Dirichlet}(\vec{\beta})$ 
//document plate:
For all document  $m \in [1, M]$  do
Sample mixture proportion  $\vec{\vartheta}_m \sim \text{Dirichlet}(\vec{\alpha})$ 
//word plate:
For all words  $n \in [1, N_m]$  in document  $m$  do
Sample topic index  $z_{m,n} \sim \text{Multinomial}(\vec{\vartheta}_m)$ 
Sample term for word  $w_{m,n} \sim \text{Multinomial}(\varphi_{z_{m,n}})$ 

```

In this study, we use the LDA model and select appropriate prior parameters to perform topic modeling analysis on the trajectory document. We can obtain two probability distributions, specifically, the trajectory-topic probability distribution and the topic-motion word probability distribution.

Topic-motion word probability distribution: Each motion pattern topic is composed of the probability distribution of a series of motion words, which can be expressed as $T_i = \langle P_{mw_1}^{T_i}, \dots, P_{mw_i}^{T_i}, \dots, P_{mw_n}^{T_i} \rangle$. $P_{mw_i}^{T_i}$ represents the probability that the motion word mw_i generates the topic T_i . If the probability threshold α is defined, when the probability is greater than the threshold β , the motion word is considered to be the representative motion of the topic T_i .

Trajectory-topic probability distribution: A ship trajectory may have one or more motion topics. Each trajectory document has a probability distribution over multiple topics, and the contribution rate of this motion topic T_i to a given motion document td_i is denoted as $P(T_i | td_i)$. The sum of a ship's trajectory probabilities to generate all topics is 1; that is, $\sum(T_i | td_i) = 1$.

2.4.2. Optimization of Topics Number

In topic modeling, setting the appropriate number of topics is a very important and challenging task [36]. Usually, the number of topics is mainly set by relying on human experience, which is inefficient for large-scale datasets. Currently, some indicators are proposed from different perspectives to find the optimal number of topics, including perplexity [27], topic coherence [37], and topic correlation [38]. This paper adopts joint topic correlation and topic coherence indicators to optimize the number of topics to improve the quality and interpretability of topics.

Cao et al. [38] studied the relationship between the performance of the LDA model and the topic correlation. The results showed that when the average cosine similarity between topics was the smallest, the topic model results were optimal and representative. Therefore, the optimal number of topics can be determined by iteratively calculating the average cosine similarity. Suppose that the distribution of topic-word of topic i in N -dimensional space is marked as a vector z_i , where N is the number of words in the corpus; the value of its t^{th} dimension is z_i^t . Then the calculation steps of topic relevance are as follows:

- (1) Calculate the correlation between topic i and topic j using cosine similarity. When the value is smaller, which indicates that topic i and topic j are more independent.

$$correlation(i, j) = \frac{\sum_{t=1}^V z_i^t z_j^t}{\sqrt{\sum_{t=1}^V (z_i^t)^2} \sqrt{\sum_{t=1}^V (z_j^t)^2}} \tag{1}$$

- (2) Calculate the topic correlation among K topics. When the value is smaller, it indicates that the topic model has better generalization performance and representation.

$$correlation = \frac{\sum_{i=1}^{K-1} \sum_{j=1}^K correlation(z_i, z_j)}{K \times (K - 1) / 2} \tag{2}$$

Mimno et al. [37] proposed a topic coherence score, which can automatically evaluate the coherence among each topic. Topic coherence is a measure of word co-occurrence. The main idea of topic coherence is that words with the same semantics as the topic words often appear simultaneously in the document. Specifically, given a topic z , the T words most relevant to the topic are denoted as $V^z = \{v_1^z, v_2^z, \dots, v_T^z\}$, denotes the number of occurrences of the word in the document, and $D(v_i^z, v_j^z)$ denotes the number of co-occurrences of the word v_i^z and v_j^z in the document. The calculation steps of topic consistency are as follows:

- (1) Calculate the consistency of topic z .

$$coherence(z) = \sum_{t=2}^T \sum_{l=1}^t \log_{10} \frac{D(v_l^z, v_t^z) + 1}{D(v_l^z)} \tag{3}$$

- (2) Calculate the topic consistency of K topics where a greater value indicates a higher quality of topics.

$$coherence = \frac{\sum_{z=1}^K coherence(z)}{K} \tag{4}$$

3. Experiment and Result Analysis

3.1. Experimental Area and Data

To verify the effectiveness of the proposed method, a case study was conducted at the Tianjin port in northern China. The layout of the Tianjin port is shown in Figure 6. It includes 3 anchorages, 11 terminal operation areas, 2 basins, and 6 waterways marked with different colors. Tianjin port is a compound waterway, including one main waterway, two boat waterways, and three warning waterways. The boat waterway includes the South Boat Waterway and North Boat Waterway. The warning waterway includes three waterways: South, North, and Main Warning Waterway. The waterway of the warning area extends westward to the West Basin and northward to the North Basin, forming a Y-shaped intersection. In total, 11 terminal operation areas are distributed around the West and North basin. Three anchorages are distributed on both sides of the Main Waterway to provide anchorage for ships waiting to enter the port. From September–November 2016, more than 4 million AIS trajectory data were collected.

3.2. Experimental Design

3.2.1. Ship Motion States Recognition and Voyage Division

According to the layout of Tianjin port, the scope and coordinates of the anchorage and basin waters of Tianjin port are extracted by ENCs. According to the type of terminal operation area in Tianjin port, this paper investigates three types of ships, namely cargo ships, tankers, and container ships, to identify and analyze the motion patterns of the inbound and outbound processes. Based on the algorithm proposed in Section 2.2, the raw trajectory data in the port waters are divided into hoteling, maneuvering, and normal-speed sailing. For the time analysis of the ship receiving service in the anchorage and berth of Tianjin port, we find that the ship stays in the above waters for more than 1 h. Therefore, in this study, the voyage time interval threshold Δt is set to 1 h. As a result, we obtained trajectory data for a total of 1956 voyages.

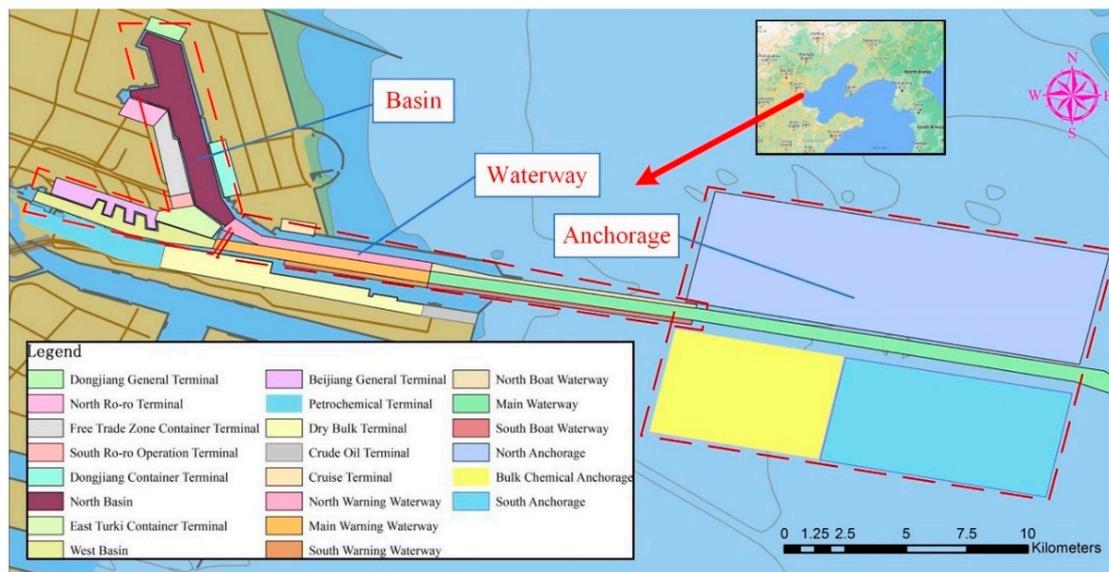


Figure 6. Layout of water and land facilities in Tianjin Port.

3.2.2. Semantic Transformation and Enrichment of Ship Trajectories

Tianjin port has a complex layout of waters with several waterways, anchorage waters, and basins. The port authority has formulated rules for ships to enter and leave the port in order to ensure the navigation safety of ships. To facilitate the semantic transformation of traffic rules, we rename the above navigable waters according to the port layout. As a result, we can obtain the names of three anchorages, six waterways, and two basins. Moreover, for other navigable waters, we use uniform names for them, as shown in Table 4. Based on the method proposed in Section 2.3, the Tianjin port is divided into square grid cells with side lengths of 150 m. The voyage trajectories data are semantically transformed and enriched by combining the water names, ship types, speed, and course information. Through the semantic transformation operation, a total of 1956 ship motion semantic documents are generated, including 45,438 motion words. Table 5 shows the motion words for two different types of ships.

Table 4. Renaming navigable waters.

Waters	Waters_Name	Waters	Waters_Name
North Anchorage	North_anchorage	North Warning Waterway	North_warning_waterway
South Anchorage	South_anchorage	Main Warning Waterway	Main_warning_waterway
Bulk Chemical Anchorage	Bulk_chemical_anchorage	South Warning Waterway	South_warning_waterway
North Boat Waterway	North_boat_waterway	North Basin	North_basin
Main Waterway	Main_waterway	West Basin	West_basin
South Boat Waterway	South_boat_waterway	Other Navigation Water	Other

Table 5. Examples of textual representations of ship trajectories.

ID	MMSI	Motion Word
1	xxxxxx360	87325#28860#Main_waterway#Cargo#West#high_speed
2	xxxxxx360	87323#28860#Main_waterway#Cargo#West#high_speed
3	xxxxxx280	87319#28861#Main_waterway#Container#West#high_speed
4	xxxxxx280	87319#28861#Main_waterway#Container#West#high_speed
5	xxxxxx280	87319#28861#Main_waterway#Container#West#high_speed

By analyzing the frequencies of the words in the document set, we found that the rank-frequency distribution displays a long tail, as shown in Figure 7. Moreover, the cumulative distribution function indicates that ships usually adopt relatively fixed actions

when sailing in specific water. This statistical result is consistent with the word distribution in natural language texts, where a few words are used frequently while most other words are used only once in a while [39]. The result shows that such motion words can maintain the basic statistical assumptions in NLP, justifying the use of motion words to represent ship motion and applying the topic model to the trajectory documents.

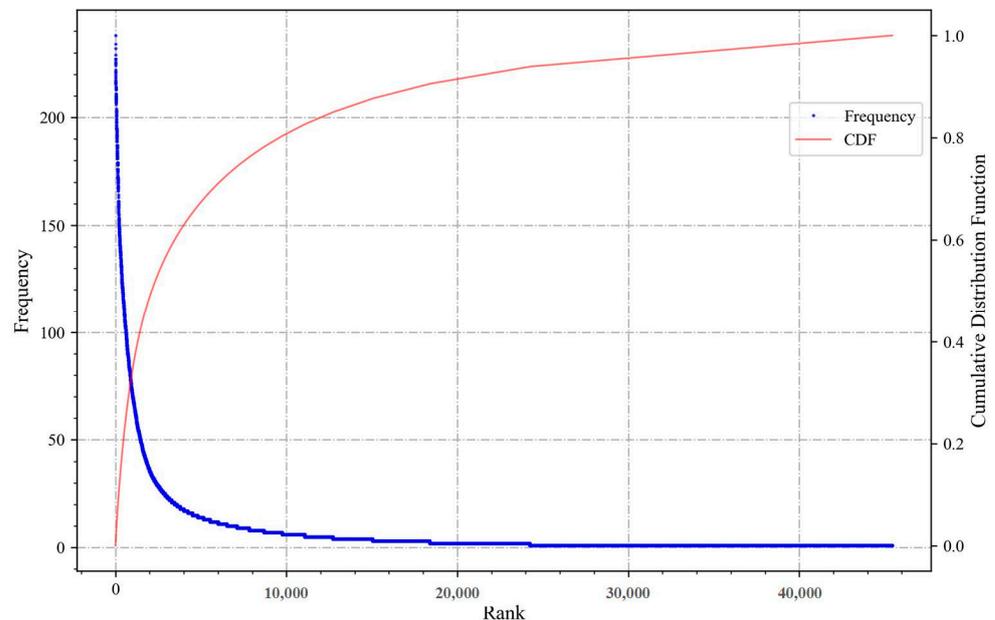


Figure 7. Rank-frequency and CDF of motion words in the trajectory documents.

3.2.3. Semantic Recognition of Ship Trajectory Documents

We execute LDA model training in this study using a variational Bayesian algorithm to estimate the prior parameters. The number of topics was set to 2–50 with an interval of 1, and 100 iterations of training were performed each time. The topic-motion word threshold value (δ) for each motion pattern topic was set to 0.01 to identify representative motion words and calculate the corresponding topic coherence and topic correlation score. As shown in Figure 8, with the increasing topic number, the topic coherence increases first and reaches a peak when the topic number is 17, and then slowly decreases. Figure 9 illustrates that the topic correlation value decreases as the number of topics increases and then remains at a low level. To ensure the topic quality and the interpretability of the topic words, we chose 17 as the optimal number of topics for the document set, considering these two indicators. As a result, we can obtain two valuable probability distributions to discover and interpret meaningful ship motion patterns.

3.3. Result Analysis

After topic modeling, we can obtain the topic probability distribution for each trajectory. Table 6 illustrates probability distributions for three different ship trajectories in a total of 17 topics. As can be seen from the table, the ship's trajectory may have one or more motion pattern topics. That is because the ship may have adopted a variety of motion patterns to complete the entire process of entering and leaving the port. The contribution of each motion pattern topic to the ship trajectory is also different, but the sum of all topic probabilities for each ship trajectory is 1. In the table, topic 2 contributed 17.2% to the ship (Ship Trajectory: xxxxxx000), topic 8 contributed 82.1%, and topics 9 and 12 contributed the remaining 0.7%. The ship (Ship Trajectory: xxxxxx030) participated in as many as six themes, indicating that the ship may have adopted flexible operation with multiple modes of motion to comply with the port traffic rules. In contrast, the ship (Ship Trajectory: xxxxxx680) was mainly involved in topic 10 and slightly involved in the motion patterns of topics 5, 6, 7, 16, and 17.

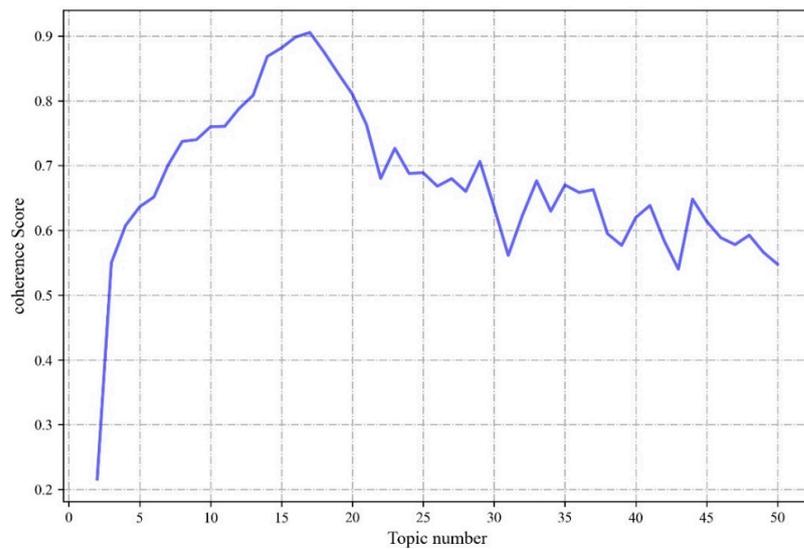


Figure 8. The change of topic coherence with topic number.

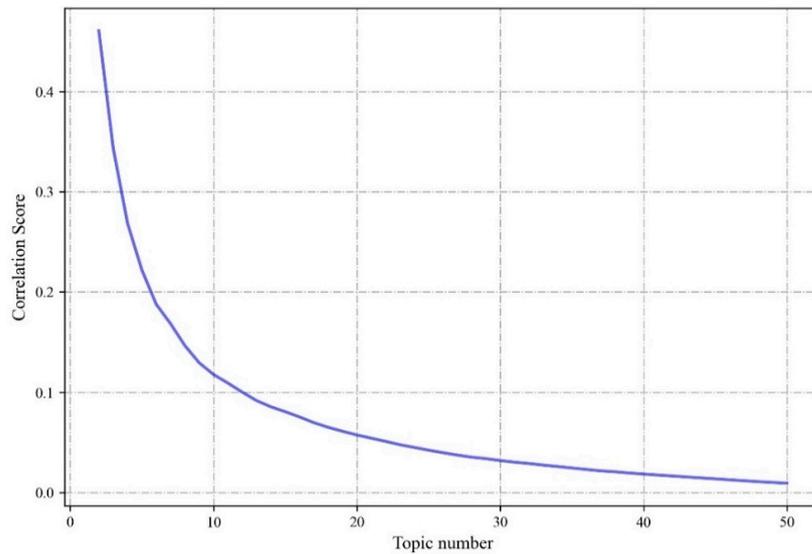


Figure 9. The change of topic correlation with topic number.

Table 6. Trajectory probability distribution in topics.

Ship Trajectory	Ship Type	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9
xxxxxx680	Container	0	0	0	0	0.037	0.046	0.081	0	0
xxxxxx030	Cargo	0	0	0.006	0.375	0	0.189	0	0	
xxxxxx000	Cargo	0	0.172	0	0	0	0	0	0.821	0.003
Ship Trajectory	Ship Type	Topic 10	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15	Topic 16	Topic 17	
xxxxxx680	Container	0.737	0	0	0	0	0	0.045	0.047	
xxxxxx030	Cargo	0	0	0	0	0.240	0.169	0	0.021	
xxxxxx000	Cargo	0	0	0.004	0	0	0	0	0	

Meanwhile, we can also obtain the probability distribution of the topic words in each topic. Each topic is composed of a group of motion words with different frequencies of occurrence. The more times the motion word appears in the topic, the higher the contribution of the motion word to the generation of the motion pattern, as shown in

Table 5. Four marked topics are listed in the table, five topic-motion words are selected for each topic, and the motion words are ranked according to the location semantic information. Combining Figures 6 and 10, we can observe that Topic 4 and Topic 11 are inbound patterns, and Topic 8 and Topic 9 are outbound patterns by analyzing each motion word’s location semantics and course semantics information. For Topic 4, the cargo ship first enters the North Boat Waterway and then travels along the North Warning Waterway to the North Basin, always sailing at medium speed during the process. For Topic 11, the container first enters the Main Waterway, then passes through the Main Warning Waterway, the North Warning Waterway to the North Basin, and maintains a medium speed during this process. For Topic 7, the container ship enters the North Warning Waterway from the North Basin, then enters the Main Warning Waterway, and finally leaves the port from the Main Waterway, and the speed changes from medium speed to high speed in the process. For Topic 9, the cargo ship enters the Main Warning Waterway from the West Basin, then enters the South Warning Waterway, and finally departs from the South Boat Waterway.

Topic4	Topic7
Representative movement words	Representative movement words
87404#28853#Other#Cargo#West#medium_speed	87258#28880#Northern_basin#Container#Southeast#medium_speed
87367#28856#Norther_boat_channel#Cargo#West#medium_speed	87264#28875#Northern_warning_channel#Container#Southeast#medium_speed
87344#28859#Norther_boat_channel#Cargo#West#medium_speed	87299#28863#Warning_main_channel#Container#East#high_speed
87322#28863#Norther_boat_channel#Cargo#West#medium_speed	87340#28857#Main_channel#Container#East#high_speed
87309#28865#Northern_warning_channel#Cargo#West#medium_speed	87365#28854#Main_channel#Container#East#high_speed
87287#28868#Northern_warning_channel#Cargo#West#medium_speed	87390#28850#Main_channel#Container#East#high_speed
87250#28898#Northern_basin#Cargo#North#medium_speed	87428#28844#Main_channel#Container#East#high_speed
Topic11	Topic9
87408#28848#Main_channel#Container#West#medium_speed	87241#28873#Western_basin#Cargo#East#medium_speed
87370#28854#Main_channel#Container#West#medium_speed	87269#28868#Warning_main_channel#Cargo#East#medium_speed
87352#28856#Main_channel#Container#West#medium_speed	87280#28865#Warning_main_channel#Cargo#East#medium_speed
87319#28861#Main_channel#Container#West#medium_speed	87292#28862#Southern_warning_channel#Cargo#East#medium_speed
87310#28862#Warning_main_channel#Container#West#medium_speed	87316#28858#Southern_boat_channel#Cargo#East#medium_speed
87295#28865#Warning_main_channel#Container#West#medium_speed	87351#28854#Southern_boat_channel#Cargo#East#medium_speed
87264#28875#Northern_warning_channel#Container#Northwest#low_speed	87430#28836#Other#Cargo#East#medium_speed

Figure 10. Representative motion words in the topic probability distribution.

By analyzing the representative motion words, we can clearly understand the changes in the motion state of the ship when entering or leaving the port in each topic, such as how the ship sails through specific water under the port traffic rules and the course and speed taken when passing through this water.

3.4. Visual Analysis

In order to conveniently display the overview of ship motion for each topic, the top 50 motion words with the highest frequency in each topic are selected for visual representation, as shown in Figure 11. The x-axes and y-axes indicate the location semantic information of the motion words, and the z-axis is used to indicate the course and speed semantic information. The red and blue marked arrows and the names of the navigable waters passed through together indicate the motion of the ship under port rules. Through this visualization, the multidimensional semantic feature information can be clearly displayed.

The experimental results show 17 ship motion patterns in Tianjin port, as shown in Figure 11. The number of inbound and outbound patterns is 8 and 9, respectively.

- Topic 1 represents a container ship inbound pattern, passing through the Main Waterway, Main Warning Waterway, North Warning Waterway, and West Basin, during which the speed changes from high to medium and the heading changes from west to northwest.
- Topic 2 represents a cargo ship outbound pattern, passing through the North Warning Waterway, Main Warning Waterway, and Main Waterway, during which the speed changes from medium to high and the heading remains east at all times.

- Topic 3 represents a cargo ship inbound pattern, passing through the Main Waterway and Main Warning Waterway, maintaining high speed during this time and always heading west.
- Topic 4 represents a cargo ship inbound pattern, passing through the North Boat Waterway and North Warning Waterway, maintaining medium speed during this time and always heading west.
- Topic 5 represents a cargo ship outbound pattern, passing through the West Basin, Main Warning Waterway, and Main Waterway, during which time the speed changes from low to medium and the heading remains east at all times.
- Topic 6 represents a cargo ship inbound pattern, passing through the North Boat Waterway and North Warning Waterway, maintaining high speed during this time and always heading west.
- Topic 7 represents a container ship outbound pattern, passing through the North Basin, North Warning Waterway, and Main Waterway, during which time the speed from medium to high and always heading east.
- Topic 8 represents a cargo ship outbound pattern, passing through the West Basin, Main Warning Waterway, and Main Waterway, maintaining medium speed during this time and always heading east.
- Topic 9 represents a cargo ship outbound pattern, passing through the West Basin, Main Warning Waterway, South Warning Waterway, and South Boat Waterway, maintaining medium speed and heading is always east.
- Topic 10 represents a container ship outbound pattern; the pattern has two streams of vessel traffic. One stream passes through the North Basin, Main Warning Waterway, and the main channel, during which the speed changes from medium to high and the heading has a southeast to eastward change. The other stream passes through the West Basin, Main Warning waterway, and Main Waterway, during which the speed changes from medium to high and the heading changes from southeast to east.
- Topic 11 represents a container-inbound pattern, passing through the Main Waterway and Main Warning Waterway, maintaining medium speed, and during this time, the heading is always west.
- Topic 12 represents a cargo ship outbound pattern, passing through the Main Warning Waterway and Main Waterway, during which speed changes from medium to high speed and the heading always remains east.
- Topic 13 represents a cargo ship outbound pattern, passing through the North Basin, North Warning Waterway, Main Warning Waterway, Main Waterway, and South Boat Waterway, during which the speed changes from medium to high speed and the heading changes from southeast to east.
- Topic 14 represents a cargo ship inbound pattern, passing through the Main Waterway and Main Warning Waterway, during which speed is always medium and heading is always west.
- Topic 16 represents a container ship outbound pattern, passing through the North Basin, North Warning Waterway, Main Warning Waterway, and Main Waterway, maintaining medium speed during this time, with a southeast-to-east heading change.
- Topic 17 represents a cargo ship inbound pattern; the pattern has two streams. One stream passes through the Main Waterway and Main Warning Waterway, during which speed changes from high to medium, low and braking, and heading changes from west to northwest. The other stream passes through the Main Waterway, Main Warning Waterway, and West Basin, during which the speed changes from high to medium, low and braking, and the heading changes from west to northwest.

While one or more semantic features in motion words are very similar, the rest are significantly different. For example, topic 3 and topic 6 have the same heading semantics and speed semantics features. However, under the influence of traffic rules, the ship in topic 3 sails in the Main Waterway and the Main Warning Waterway, and the ship in topic 6 sails in the North Boat Waterway and North Warning Waterway. In the same way, due

to the difference in the semantic features of speed or course, they are finally identified as different motion patterns, such as topic 10 and topic 11. These pattern diagrams illustrate that the semantic recognition of motion patterns can be achieved by topic modeling for multidimensional semantic information of ship motion, and interpretable motion words can be extracted.

To understand the details of each topic more comprehensively, this paper combines ArcGIS and ENCs to visualize the representative motion words for each topic. Figure 12 shows spatial distributions of all 17 motion pattern topics. Visual elements, such as shape, size, and color, are used to display topics and motion words. In the figure, each topic is assigned a specific color. Motion word in the topic is represented by a directional triangle ship symbol with a tail at a specific location. Ship symbol size indicates the ship type, which depends on the ship type code specified in the AIS message standard. The length of the symbol tail represents the ship's speed. Long length, medium, medium-short length, and short length represent high speed, medium speed, low speed, and braking, respectively. Distributions of motion words in all topics can clearly show motion patterns and dynamics of ship traffic in the Tianjin port.

Firstly, seven typical topics are selected, which depict the representative motion patterns of ships entering and leaving Tianjin port, as shown in Figure 13. Topics 6, 11, and 15 are typical inbound motion patterns. Topic 15 (Cargo ship) represents the motion pattern along the Main Waterway and the north side of the Main Waterway into the West Basin. The ship entered the Main Waterway at high speed and gradually decreased its speed as it moved closer to the berth in the West Basin and finally sailed through the Main Warning Waterway at low speed. Topic 11 (Container) represents the motion pattern along the north side of the Main Waterway, through the North Warning Waterway, and into the North Basin. The ship always maintains medium speed. An interesting finding is that the motion pattern represented by topic 6 (Cargo ship) has two streams. When the ship enters the warning waters along the North Boat Waterway at high speed, one stream crosses the North Warning Waterway into the West Basin at high speed, and the other enters the West Basin at medium speed at the intersection of the warning waters. This situation may have occurred due to the presence of an adjacent bulk terminal operating area at the Y-shaped crossing. Due to the high frequency of the two streams sharing motion words, the two streams will be considered as one topic in topic modeling.

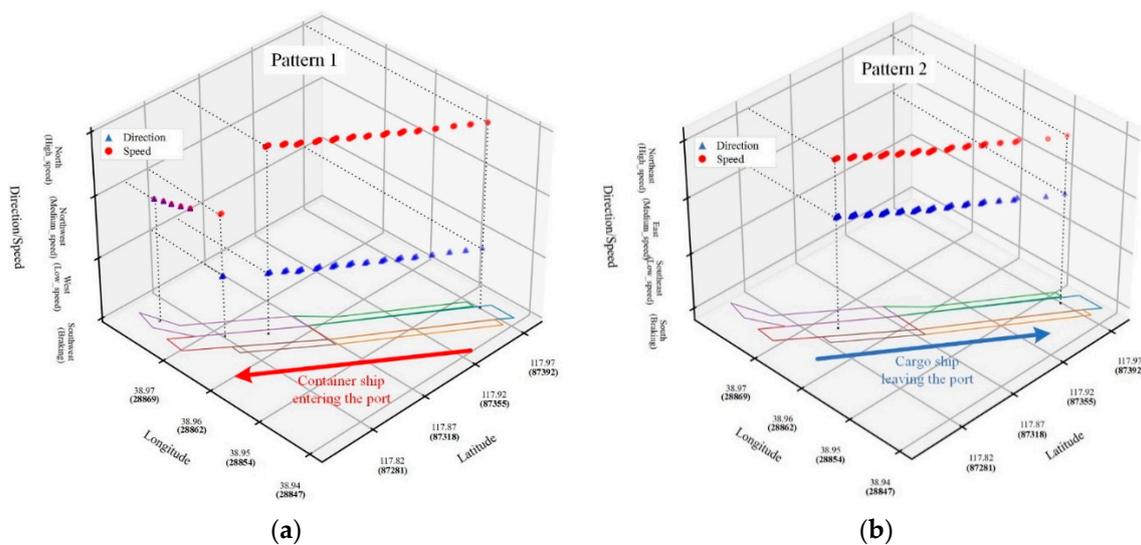


Figure 11. Cont.

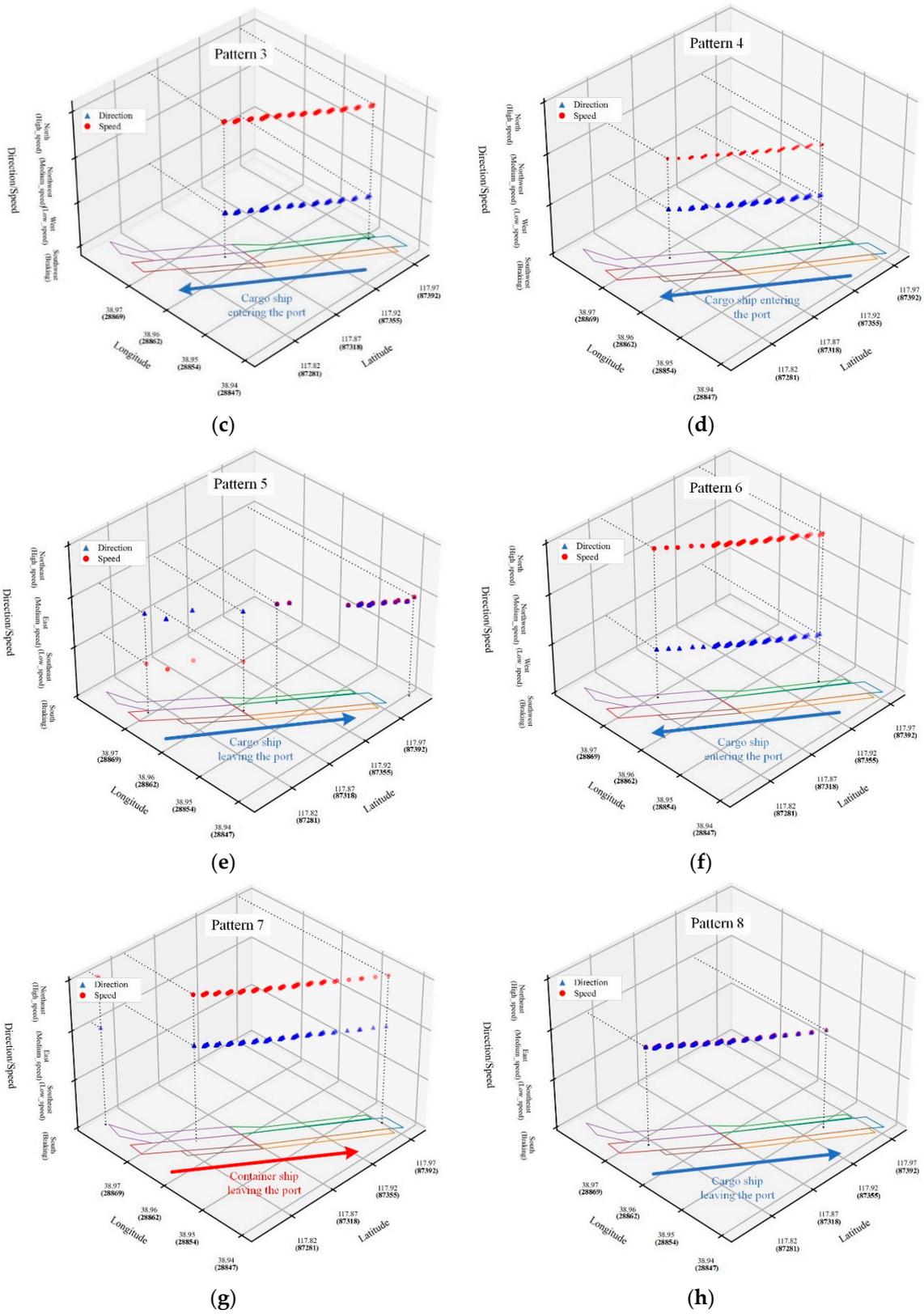


Figure 11. Cont.

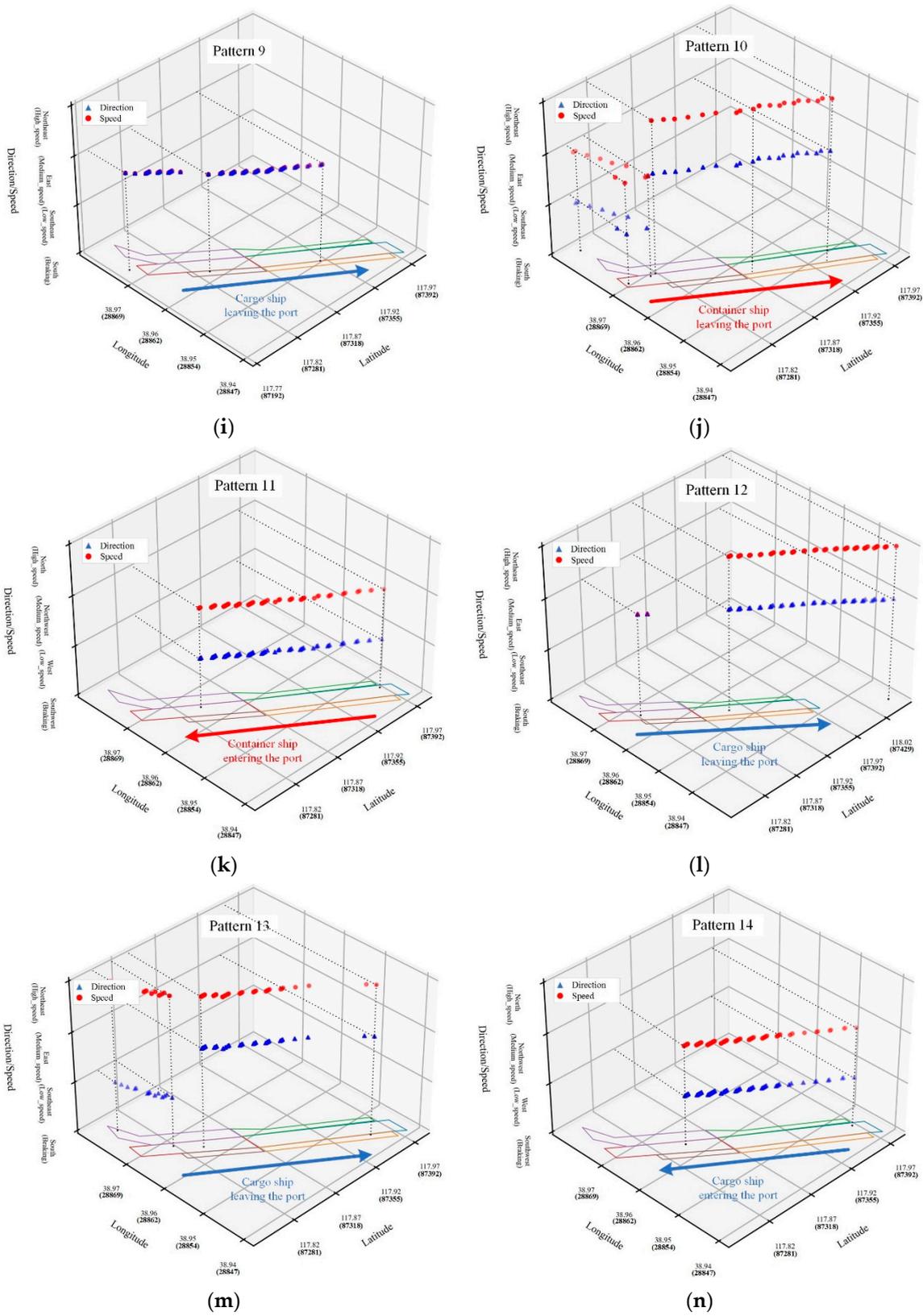


Figure 11. Cont.

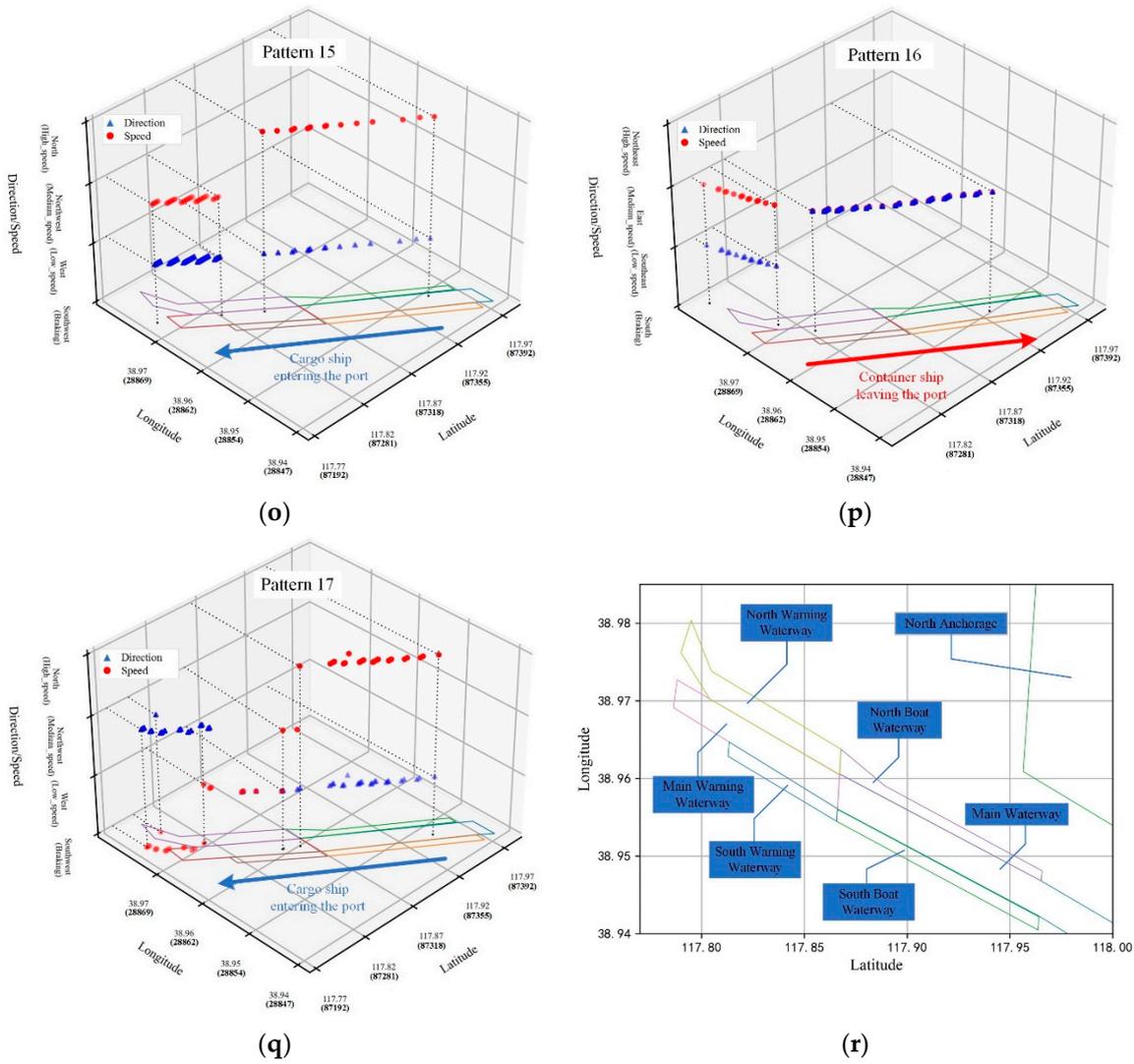


Figure 11. Visualization of the spatial distribution of high-frequency motion words for 17 motion pattern topics in Tianjin Port. (a) Topic 1; (b) Topic 2; (c) Topic 3; (d) Topic 4; (e) Topic 5; (f) Topic 6; (g) Topic 7; (h) Topic 8; (i) Topic 9; (j) Topic 10; (k) Topic 11; (l) Topic 12; (m) Topic 13; (n) Topic 14; (o) Topic 15; (p) Topic 16; (q) Topic 17; (r) Overhead view of the Tianjin Port waterway.

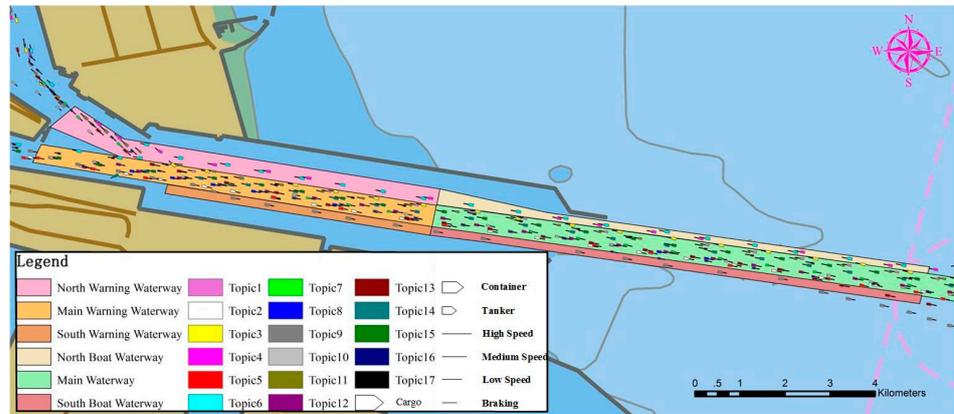


Figure 12. Visualization of the spatial distribution of 17 motion pattern topics entering and leaving the Tianjin port waterway.



Figure 13. Visualization of the spatial distribution of seven representation motion pattern topics entering and leaving the Tianjin port waterway.

Topics 7, 8, 9, and 13 are typical outbound motion patterns. Topic 7 (Container) represents sailing from the North Basin into the North Warning Waterway at medium speed and leaving along the south side of the Main Waterway at high speed. Topic 13 (Cargo ship) represents sailing from the North Basin into the North Warning Waterway at medium speed and crossing the Main Warning Waterway and the South Warning Waterway at the Y-shaped intersection at high speed, entering the South Boat Waterway and sailing out of the port. Topic 8 (Cargo ship) represents sailing out of the Main Warning Waterway at medium speed and maintaining medium speed out of the Main Waterway. Topic 9 (Cargo ship) represents sailing into the Main Warning Waterway at medium speed and crossing the South Warning Waterway and the South Boat Waterway at medium speed.

Then, we visualized the spatiotemporal distribution of the motion words in the inbound mode and outbound mode, respectively. We can find that when the ship enters the port, the ship will choose to enter the basin from the north side of the Main Waterway and the North Boat Waterway. The speed is gradually reduced to approach the berth safely based on ensuring rudder efficiency, as shown in Figure 14. When the ship leaves the port, the ship will choose to leave the port from the south side of the Main Waterway and the South Boat Waterway. The ship’s speed increases as it moves further away from the berth and maintains a medium speed and above as it leaves the harbor, as shown in Figure 15. The pattern knowledge uncovered is consistent with the traffic rules of the Tianjin port. In addition, we found that ships sailing in the North and South Boat Waterways have higher speeds than the Main Waterway. This is due to the fact that ships navigating these waterways are generally less than 100 m in ship length, and the maneuvering performance of such ships is relatively better. This is also in line with the traffic rules of Tianjin port, which adopts a compound waterway.



Figure 14. Visualization of the spatial distribution of 8 motion pattern topics entering the Tianjin port waterway.

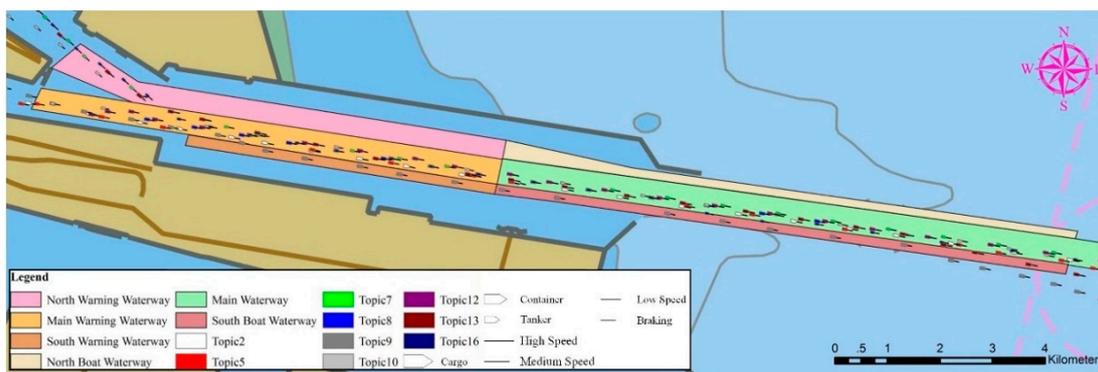


Figure 15. Visualization of the spatial distribution of nine motion pattern topics leaving the Tianjin port waterway.

Next, we interpret and analyze the motion words of each topic in combination with the water and land area facilities in the port. Topics 1, 7, 10, 11, 16, and 17 are container ship motion patterns, and the operating areas of container ships are mainly located on both sides of the North Basin. Topics 2, 3, 4, 5, 6, 8, 9, 12, 13, 14, and 15 are cargo ship motion patterns, and the operational areas for cargo ships are distributed in both the North Basin and West Basin. Thus, based on the spatiotemporal distribution of motion words for each topic, we can find that the motion trend of each pattern is consistent with the distribution of port operating areas. Meanwhile, we found no representative motion words for oil tankers in all patterns, which we infer is due to the fact that the types of vessels served by Tianjin port are mainly container ships and cargo ships, while the number of tankers is small.

Finally, we found that ships always navigate close to the east side of the channel when passing through the Y-shaped intersection due to the berthing operation areas on the west side of the Y-shaped intersection and the need for incoming and outgoing ships to avoid these areas. Meanwhile, we found various conflict patterns easily appear in the Y-shaped intersection when the ships enter and leave Tianjin port. For example, the crossing situation formed by topic 6 and topic 13, and the head-on situation formed by topic 15 and topic 5. The above knowledge can assist the driver in predicting the possible navigational hazards, driving carefully, and increasing ship-handling skills to avoid these conflict situations. Port managers will be able to monitor key areas and ensure that ships enter and leave the waterways safely.

4. Discussion

4.1. Advantages of the Method

It is well known that a ship is affected by a variety of factors when sailing on the ocean. Therefore, how these factors are considered when studying ship motions is crucial to accurately describe ship motions. However, when too many factors are considered too much, it can also make it difficult or require a lot of effort to construct complex ship motion models. In addition, the model results are not easy to understand.

With the wide application of new technologies, such as big data mining and artificial intelligence in the maritime field, continues to promote the development of the shipping industry. We propose a topic model-based method for recognizing ship motion patterns entering and leaving ports. Compared with other methods, this method has the following advantages:

- (1) More factors are considered and less modeling work is required

Compared with statistical methods, clustering methods, and deep learning methods, the proposed method is able to consider the influence of multiple external factors on ship motion. Different from semantic analysis methods, the proposed method does not need to build a knowledge base of ship motion but converts the semantics of trajectory data into corresponding semantic features, which reduces the tedious modeling workload.

- (2) Friendly to the inexperienced

Topic modeling is a semi-supervised learning approach to cluster the latent semantic structure of document sets, which usually requires manual setting of the number of topics and is not friendly to people without relevant practitioner backgrounds. The proposed method introduces two metrics, topic coherence, and topic correlation, to achieve the optimal number of topics, which is operator-friendly and excludes the influence of human factors.

(3) Easy-to-understand modeling results

By converting the numerical information that is not easy to understand into a semantically described trajectory document and using it as a data source for topic modeling, the semantic structure described in the form of vocabulary is output, making the model results easier to understand. In addition, applying GIS to the maritime field, combined with ENCs as a visual display platform, transforms textual information into intuitive and visual graphic information. Moreover, the geospatial information of port water and land facilities such as waterways, basins, and terminal operation areas are combined to achieve a further understanding of the motion patterns.

4.2. Limitations

The topic model uses text documents as the data source. When the trajectory text is too short, it leads to a too sparse document-term matrix, which makes it difficult to mine useful information. Therefore, before constructing ship trajectory documents, the trajectories should be preprocessed to eliminate trajectories containing fewer points or use interpolation to supplement the missing trajectory points.

5. Conclusions and Future Work

We propose a semantic recognition method based on a topic model for ship motion patterns entering and leaving the ports. The method can consider the influence of port traffic rules on ship motion and discover representative and interpretable ship motion patterns from a large amount of AIS data in an unsupervised manner. Validated by an example in the port of Tianjin in northern China, the method is able to mine 17 motion patterns and their pattern knowledge, which can help port managers to mine maritime traffic flow characteristics and identify traffic conflicts.

In future work, the obtained motion pattern results can be used in port vessel traffic modeling and simulation to assess the port's navigational efficiency and risk index. A ship abnormal behavior detector is constructed based on the pattern results to detect abnormal ship behavior. For example, abnormal acceleration and deceleration of ships, large angle turning, etc. In addition, hydrometeorology can be imported into the semantics to enrich the trajectories' semantic information and help find more meaningful conclusions.

Meanwhile, natural language processing techniques have good application prospects in the maritime traffic field. For example, (1) based on a corpus of maritime collision records, text mining methods are used to extract the ranking and weight relationships of relevant factors that cause maritime accidents, and the results of their models can be used to prevent maritime accidents [40]; (2) combining the rich information in unstructured formats (e.g., intelligence reports and news articles), defining entities and concepts and constructing their semantic relationships to achieve textual data from the maritime domain constructing probabilistic knowledge graphs [41]; (3) using multiple sources of ship traffic data, constructing sequential record texts of ships visiting ports and using natural language processing methods to construct port recommendation algorithms to provide suggestions for changing ship destinations under uncertainty [42].

Author Contributions: Conceptualization, G.L., X.Z. and K.-h.L.; methodology, G.L., M.L. and W.Q.; software, G.L.; validation, G.L., X.Z. and K.-h.L.; formal analysis, G.L., C.W. and M.L.; investigation, G.L., C.W. and M.L.; resources, X.Z. and K.-h.L.; data curation, G.L. and W.Q.; writing—original draft preparation, G.L.; writing—review and editing, G.L., X.Z., K.-h.L. and C.W.; visualization, G.L. and W.Q.; supervision, X.Z. and K.-h.L.; funding acquisition, X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Dalian Science and Technology Innovation Fund, grant number 2022JJ12GX015; the National Natural Science Foundation of China, grant number 51779028.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The article processing charge of this work is supported by Shipping Research Centre of The Hong Kong Polytechnic University.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Review of Maritime Transport. 2019. Available online: <https://unctad.org/webflyer/review-maritime-transport-2019> (accessed on 30 October 2022).
2. Chen, C.; Liu, X.; Chen, H.H.; Li, M.; Zhao, L. A rear-end collision risk evaluation and control scheme using a Bayesian network model. *IEEE Trans. Intell. Transp. Syst.* **2022**, *20*, 264–284. [[CrossRef](#)]
3. Chen, J.; Zhang, W.; Li, S.; Zhang, F.; Zhu, Y.; Huang, X. Identifying critical factors of oil spill in the tanker shipping industry worldwide. *J. Clean. Prod.* **2018**, *180*, 1–10. [[CrossRef](#)]
4. Xue, Y.; Lai, K.H. Responsible shipping for sustainable development: Adoption and performance value. *Transp. Policy.* **2023**, *13*, 89–99. [[CrossRef](#)]
5. Rong, H.; Teixeira, A.P.; Soares, C.G. Maritime traffic probabilistic prediction based on ship motion pattern extraction. *Reliab. Eng. Syst. Saf.* **2022**, *217*, 108061. [[CrossRef](#)]
6. Zhang, M.; Kujala, P.; Hirdaris, S. A machine learning method for the evaluation of ship grounding risk in real operational conditions. *Reliab. Eng. Syst. Saf.* **2022**, *226*, 108697. [[CrossRef](#)]
7. Zhen, R.; Shi, Z.; Liu, J.; Shao, Z. A novel arena-based regional collision risk assessment method of multi-ship encounter situation in complex waters. *Ocean Eng.* **2022**, *246*, 110531. [[CrossRef](#)]
8. Vejvar, M.; Lai, K.H.; Lo, C.K. A citation network analysis of sustainability development in liner shipping management: A review of the literature and policy implications. *Marit. Policy Manag.* **2020**, *47*, 1–26. [[CrossRef](#)]
9. Zhou, Y.; Daamen, W.; Vellinga, T.; Hoogendoorn, S. Review of maritime traffic models from vessel behavior modeling perspective. *Transp. Res. Part C-Emerg. Technol.* **2019**, *105*, 323–345. [[CrossRef](#)]
10. Smith, M.; Reece, S.; Roberts, S.; Rezek, I. Online Maritime Abnormality Detection using Gaussian Processes and Extreme Value Theory. In Proceedings of the 2012 IEEE 12th International Conference on Data Mining (ICDM), Brussels, Belgium, 10–13 December 2012; pp. 645–654. [[CrossRef](#)]
11. Zhao, L.; Shi, G. A trajectory clustering method based on Douglas-Peucker compression and density for marine traffic pattern recognition. *Ocean Eng.* **2018**, *172*, 456–467. [[CrossRef](#)]
12. Gao, M.; Shi, G.Y. Ship-handling behavior pattern recognition using AIS sub-trajectory clustering analysis based on the T-SNE and spectral clustering algorithms. *Ocean Eng.* **2020**, *205*, 106919. [[CrossRef](#)]
13. Chen, X.; Liu, Y.; Achuthan, K.; Zhang, X. A ship movement classification based on Automatic Identification System (AIS) data using Convolutional Neural Network. *Ocean Eng.* **2020**, *218*, 108182. [[CrossRef](#)]
14. Abreu, F.H.; Soares, A.; Paulovich, F.V.; Matwin, S. A trajectory scoring tool for local anomaly detection in maritime traffic using visual analytics. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 412. [[CrossRef](#)]
15. Svanberg, M.; Santén, V.; Hörteborn, A.; Holm, H.; Finnsgård, C. AIS in maritime research. *Mar. Pol.* **2019**, *106*, 103520. [[CrossRef](#)]
16. Riveiro, M.; Pallotta, G.; Vespe, M. Maritime anomaly detection: A review. *Wiley Interdiscip. Rev.-Data Mining Knowl. Discov.* **2018**, *8*, e1266. [[CrossRef](#)]
17. Kowalska, K.; Peel, L. Maritime Anomaly Detection using Gaussian Process Active Learning. In Proceedings of the 2012 15th International Conference on Information Fusion (FUSION), Singapore, 9–12 July 2012; pp. 1164–1171.
18. Laxhammar, R. Anomaly Detection for Sea Surveillance. In Proceedings of the 2008 11th International Conference on Information Fusion (FUSION), Cologne, Germany, 30 June–3 July 2008; pp. 1–8.
19. Li, H.; Lam, J.S.L.; Yang, Z.; Liu, J.; Liu, R.W.; Liang, M.; Li, Y. Unsupervised hierarchical methodology of maritime traffic pattern extraction for knowledge discovery. *Transp. Res. Part C-Emerg. Technol.* **2022**, *143*, 103856. [[CrossRef](#)]
20. Li, H. Typical Trajectory Extraction Method for Ships Based on AIS Data and Trajectory Clustering. In Proceedings of the 2021 2nd International Conference on Artificial Intelligence and Information Systems, Chongqing, China, 28–30 May 2021; pp. 1–8. [[CrossRef](#)]
21. Zhang, Z.; Huang, L.; Peng, X.; Wen, Y.; Song, L. Loitering behavior detection and classification of vessel movements based on trajectory shape and Convolutional Neural Networks. *Ocean Eng.* **2022**, *258*, 111852. [[CrossRef](#)]
22. Liu, H.; Chen, X.; Wang, Y.; Zhang, B.; Chen, Y.; Zhao, Y.; Zhou, F. Visualization and visual analysis of vessel trajectory data: A survey. *Vis. Inform.* **2021**, *5*, 1–10. [[CrossRef](#)]

23. Jin, L.; Luo, Z.; Gao, S. Visual analytics approach to vessel behaviour analysis. *J. Navig.* **2018**, *71*, 1195–1209. [[CrossRef](#)]
24. Li, Y.; Ren, H. Visual Analysis of Vessel Behaviour Based on Trajectory Data: A Case Study of the Yangtze River Estuary. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 244. [[CrossRef](#)]
25. Ray, R.C.; Elena, C.; Richard, D.; Anne-Laure, J.; Clément, I.; Maximilian, Z.; Melita, H. *Use Case Design and Big Data Analytics Evaluation for Fishing Monitoring*; OCEANS: Marseille, France, 2019; pp. 1–8. [[CrossRef](#)]
26. Yan, Z.; Cheng, L.; He, R.; Yang, H. Extracting ship stopping information from AIS data. *Ocean Eng.* **2022**, *250*, 111004. [[CrossRef](#)]
27. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn.* **2003**, *3*, 993–1022.
28. Zhao, Z.; Koutsopoulos, H.N.; Zhao, J. Discovering latent activity patterns from transit smart card data: A spatiotemporal topic model. *Transp. Res. Part C-Emerg. Technol.* **2020**, *116*, 102627. [[CrossRef](#)]
29. Tian, Z.; Yang, W.; Zhang, T.; Ai, T.; Wang, Y. Characterizing the activity patterns of outdoor jogging using massive multi-aspect trajectory data. *Comput. Environ. Urban Syst.* **2020**, *95*, 101804. [[CrossRef](#)]
30. Liu, K.; Gao, S.; Lu, F. Identifying spatial interaction patterns of vehicle movements on urban road networks by topic modelling. *Comput. Environ. Urban Syst.* **2019**, *74*, 50–61. [[CrossRef](#)]
31. Huang, L.; Wen, Y.; Guo, W.; Zhu, X.; Zhou, C.; Zhang, F.; Zhu, M. Mobility pattern analysis of ship trajectories based on semantic transformation and topic model. *Ocean Eng.* **2020**, *201*, 107092. [[CrossRef](#)]
32. Zhang, L.; Meng, Q.; Xiao, Z.; Fu, X. A novel ship trajectory reconstruction approach using AIS data. *Ocean Eng.* **2018**, *159*, 165–174. [[CrossRef](#)]
33. Technical Characteristics for an Automatic Identification System Using Time-Division Multiple Access in the VHF Maritime Mobile Band. Available online: <https://www.itu.int/rec/R-REC-M.1371/en> (accessed on 30 October 2022).
34. Harati-Mokhtari, A.; Wall, A.; Brooks, P.; Wang, J. Automatic Identification System (AIS): Data reliability and human error implications. *J. Navig.* **2007**, *60*, 373–389. [[CrossRef](#)]
35. Xiao, X.; Shao, Z.; Ji, X.; Chen, L. Speed control model of ships entering and leaving ports based on AIS data. *J. Shanghai Marit. Univ.* **2014**, *4*, 11–14.
36. Newman, D.; Noh, Y.; Talley, E.; Karimi, S.; Baldwin, T. Evaluating Topic Models for Digital Libraries. In Proceedings of the 10th Annual Joint Conference on Digital Libraries (JCDL), Gold Coast Queensland, Australia, 21–25 June 2010; pp. 215–224. [[CrossRef](#)]
37. Mimno, D.; Wallach, H.; Talley, E.; Leenders, M.; McCallum, A. Optimizing Semantic Coherence in Topic Models. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, UK, 27–31 July 2011; pp. 262–272.
38. Cao, J.; Xia, T.; Li, J.; Zhang, Y.; Tang, S. A density-based method for adaptive LDA model selection. *Neurocomputing* **2009**, *72*, 1775–1781. [[CrossRef](#)]
39. Manning, C.; Schütze, H. *Foundations of Statistical Natural Language Processing*; MIT Press: Cambridge, MA, USA, 1999; pp. 52–120.
40. Shi, S.; Zhang, D.; Su, X.; Zhang, M.; Sun, M.; Yao, H. Risk Factors Analysis Modeling for Ship Collision Accident in Inland River Based on Text Mining. In Proceedings of the International Conference on Transportation Information and Safety (ICTIS), Liverpool, UK, 14–17 July 2019; pp. 602–607. [[CrossRef](#)]
41. Shiri, F.; Wang, T.; Pan, S.; Chang, X.; Li, Y.; Haffari, R.; Nguyen, V.; Yu, S. Toward the Automated Construction of Probabilistic Knowledge Graphs for the Maritime Domain. In Proceedings of the International Conference on Information Fusion (FUSION), Sun City, South Africa, 1–4 November 2021; pp. 1–8. [[CrossRef](#)]
42. Mei, Q.; Hu, Q.; Yang, C.; Zheng, H.; Hu, Z. Port Recommendation System for Alternative Container Port Destinations Using a Novel Neural Language-Based Algorithm. *IEEE Access* **2020**, *8*, 199970–199979. [[CrossRef](#)]