Yang, J, Bian, X, Qi, Y, Wang, X, Yang, Z and Liu, J

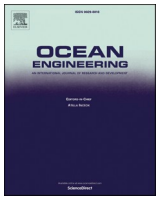
A spatial-temporal data mining method for the extraction of vessel traffic patterns using AIS data

Ocean Engineering 293 (2024) 116454

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng



Research paper

A spatial-temporal data mining method for the extraction of vessel traffic patterns using AIS data

Jiaxuan Yang [a,b], Xingpei Bian [a,b], Yuhao Qi [a,b], Xinjian Wang [a], Zaili Yang [c,*], Jiaguo Liu [d]

[a] Navigation College, Dalian Maritime University, Dalian, 116026, China
[b] The Key Laboratory of Navigation Safety Guarantee, Liaoning Province, Dalian, 116026, China
[c] Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool, L3 3AF, UK
[d] Shipping Economics and Management College, Dalian Maritime University, Dalian, 116026, China

ARTICLE INFO

Handling Editor: Prof. A.I. Incecik

Keywords:
Spatial-temporal analysis
Data mining
Vessel traffic pattern
Tensor factorization
Maritime safety
AIS data

ABSTRACT

Current traffic pattern mining methods fail to incorporate the temporal co-occurrence of traffic characteristics. To address this problem, a new spatial-temporal data mining method is developed involving three steps. Firstly, a three-dimensional traffic tensor is constructed utilizing AIS data. The AIS data is discretized and numbered so that each AIS data entry is represented by a one-dimensional array that includes region, time, ship type, and speed numbers. Then the AIS array is mapped to the three-dimensional ship traffic tensor. Second, non-negative tensor factorization (NTF) is used to break down the tensor into multiple sub-tensors (i.e., traffic patterns). The effect of the tensor rank (i.e., the number of traffic patterns) is discussed, and the appropriate value of the tensor rank is determined. Thirdly, the traffic patterns are derived from the three-dimensional traffic tensor. The ship traffic pattern is subsequently analyzed in accordance with the actual circumstances. To demonstrate the feasibility of the method, 9 traffic patterns are obtained from the AIS data of Tianjin port-Caofeidian waters. These patterns reveal the presentation of the spatio-temporal distribution of traffic activities of different ship types, and the distribution of navigation speed of different ship types in space, that are of strategic values for port planning, and maritime safety and sustainability.

1. Introduction

Since 2002, the International Maritime Organization (IMO) that all passenger ships and cargo ships over 300 gross tonnages must be equipped with the AIS (Tang et al., 2022). This measure provides the basis for using AIS data for traffic analysis. AIS data is composed of static and dynamic information, the static information includes Maritime Mobile Service Identity (MMSI), ship type, and sizes, which is used for ship identification, while the dynamic information includes ship speed, ship course, and geographical location relating to the interaction process between ships and the navigational environment (Cao et al., 2023b; Fang et al., 2023). In view of its strong availability and extensive information coverage, AIS data has gradually become a significant data source for regional water traffic situation assessment (Christensen et al., 2022; Huang et al., 2023; Valcalda et al., 2023). The AIS data within a specific period in a region can be used for vessel motion prediction (Zhang et al., 2023) and extraction of maritime traffic network (Liu et al., 2023), ship route planning (Filipiak et al., 2020; Han and Yang, 2020), waterway navigable capacity estimate (Liu et al., 2020), ship behavior research (Kabir et al., 2022; Wu et al., 2018) and marine traffic safety assessment (Chen et al., 2021; Xia, 2021; Montewka et al., 2022), etc. The availability of AIS data also provides an opportunity to understand vessel traffic patterns. In this study, the traffic pattern is defined as a brief representation of the vessel navigation rule (Xiao et al., 2020). It characterizes the vessel trajectories during their daily activity, which can represent voluminous real-world navigation instances. These patterns reveal the spatiotemporal relationships between ships, including the relative position of ships at specific time points, the regular patterns and variations in their movements over time, as well as the interactive behaviors exhibited between ships. Information such as traffic volume and the layout of shipping channels and other functional areas are also implied.

The extracted traffic pattern is not only supportive for the traffic prediction (Xiao et al., 2017) but also contributes towards maritime

* Corresponding author.
E-mail addresses: yangjiaxuan@dlmu.edu.cn (J. Yang), bianxp@dlmu.edu.cn (X. Bian), qiyuhao@dlmu.edu.cn (Y. Qi), wangxinjian@dlmu.edu.cn (X. Wang), Z.Yang@ljmu.ac.uk (Z. Yang), liujiaguo@gmail.com (J. Liu).

https://doi.org/10.1016/j.oceaneng.2023.116454
Received 6 April 2023; Received in revised form 5 October 2023; Accepted 26 November 2023
Available online 8 January 2024

traffic visualization (Zhen et al., 2017), anomaly detection (Ristic, 2014a), collision alert (Wu et al., 2017a), traffic planning (Filipiak et al., 2020) and other high level decision making support to VTS etc. Although showing some attractiveness, these methods still reveal some limitations in their applications (Cao et al., 2023a). First, the vector-based method focuses on the geometric topological characteristics in space and maritime traffic visualization. However, it often insufficiently takes into account other vessel traffic information such as ship speed changes. Secondly, the dynamic evolution of ship traffic is yet fully incorporated into the traffic pattern analysis. The correspondence relationship between each dimension (e.g. time and space dimension) and ship motion pattern therefore often remains unclear. Finally, the implicit relationship between the attributes of AIS data is at large ignored. In statistical-based methods, for multi-dimensional data, the existing studies often reduced the dimension of AIS data and imposed constraints such as a Gaussian distribution. As a result, they unavoidably change the original structure of AIS data. To promote the analysis of ship traffic patterns, it is necessary to find a new method that can support the analysis and expression of multidimensional dynamic data and better retain the AIS data structure. Based on these motivations, this study is conducted to make valuable explorations in maritime traffic patterns mining, providing support for subsequent studies on maritime traffic prediction and maritime safety planning. It will provide a theoretical basis for maritime authorities to enhance monitoring efficiency and reduce the occurrence and severity of maritime accidents.

In this study, the traffic tensors constructed from AIS data are decomposed into multiple sub-tensors (i.e., vessel traffic patterns). The tensor can address the association between multiple attributes simultaneously. This advanced NTF can therefore well fit the characteristics of the vessel traffic pattern containing multiple attributes.

In this study, the NTF method is used to explore vessel traffic patterns. The main contributions are as follows:

(1) A new method is proposed to construct three-dimensional vessel traffic tensors using AIS data. The research area is divided into irregular grids and assigned each of them a unique number, which corresponds to different grids where each ship is located. In addition, attributes such as time, ship types, and speed are separated and marked accordingly. Each AIS data entry is represented by a concise one-dimensional array that encapsulates the relevant information, including area, time, ship type, and speed numbers. By applying a mapping rule, the three-dimensional ship traffic tensors are constructed from one-dimensional AIS data.
(2) The NTF method is employed to directly analyze the distribution characteristics of AIS data, unveiling the concealed traffic pattern information. The influence of the tensor rank on the vessel traffic patterns is also discussed. A real case of Tianjin-Caofeidian waters is analyzed to verify the effectiveness of the vessel traffic patterns extraction method and the co-occurrence of traffic patterns in multiple attributes.

The rest of this work is organized as follows. Section 2 introduces the relevant work on the discovery of vessel traffic patterns. Section 3 describes the pre-processing of the AIS data, the establishment of the three-dimensional vessel traffic tensor and the detail of NTF algorithm. In Section 4, the experimental result of the Tianjin-Caofeidian water area and the choice of tensor rank are conducted, analyzed, and discussed. Section 5 discusses the application of the proposed method in practice. Section 6 gives the conclusion and future works.

2. Literature review

There are not lacking of empirical studies of using AIS data to analyze the vessel traffic patterns in the existing literature. Xiao et al. (2020) gave a comprehensive introduction to the analysis of maritime traffic patterns. It was found that the maritime traffic patterns mining methodologies can be categorized as three main methods: grid-based, vector-based and statistics-based methods. The grid-based approaches aggregate the raw AIS data into a set of indexed grids to facilitate retrieval and reduce storage. The vector method treats AIS data as a set of points and lines. The common statistics-based methods focus on the statistical analysis of the traffic features. Three traffic knowledge mining approaches are summarized respectively, including grid-based, vector-based and statistical-based methods (Animah and Shafiee, 2021).

The basic idea of the grid-based methods is to divide a target maritime area into a certain number of indexed grids. AIS data attributes are filled into the indexed grids, such as geographic locations (Lei et al., 2016; Wu et al., 2017b), ship courses (Osekowska et al., 2014), and ship speeds (Ristic, 2014b; Zhang et al., 2017). By comparing and combining the statistics of all grids, the traffic situation in the area such as the geographic distribution of ship density and speed can be obtained. However, In the grid-based methods, the study area is often divided into regular grids with the same size. When studying ship traffic characteristics, the study object is changed from countless AIS trajectory points to dozens of indexed grids. Therefore, the grid-based methods reduce the storage space of traffic patterns and facilitate retrieval by aggregating the traffic data into individual grids. It is also often used as a data processing method together with other pattern mining methods, to reduce the impact of noise.

In vector-based methods, a shipping route is abstracted as an ordered set of straight lines and points connecting the start and end points (Pallotta et al., 2013a), which is often necessary to cluster historical trajectories and waypoints. Therefore, clustering algorithms, such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Millefiori et al., 2017; Pallotta et al., 2013b; Xiao et al., 2020), K-means (Liu et al., 2020; Vespe et al., 2012) and their improvements (Liu et al., 2020; 2022) are used in vector-based methods. Rong et al. (2022) used an Ordering Points to Identify the Clustering Structure (OPTICS) algorithm to extract the ship motion mode and used the result for ship position prediction. Zhang et al. (2021a, 2021b) used the combination of K-means and DBSCAN to analyze the ships' trajectories, and provided the risk assessment of ship collision. The vector-based methods usually focus on the spatial position variation of vessels and traffic network visualization and overlook the effect of temporal data.

Most statistical-based methods do not explicitly differentiate the concepts like waypoints or turning points. These approaches represent traffic patterns through the distribution profile of traffic attributes such as speed and density. In this way, some statistic models, such as kernel density estimation (KDE) (Li et al., 2016; Millefiori et al., 2016; Ristic et al., 2008) and nonparametric regression (Wen et al., 2016) were used to estimate the distribution of important vessel traffic attributes. Statistical-based methods often appear as evaluation models. Wang et al. (2021) used an ordered logistic regression model to reflect the relationship between influencing factors and the severity of maritime accidents. Zhang et al. (2022) proposed a method to quantify the complexity of inland river traffic flows and evaluate the complexity of traffic flows in different areas. Gil et al. (2022) used ten years of AIS data to determine the empirical value of the Bow Crossing Range (BCR) and analyzed the influential factors affecting the BCR. In the relevant studies, a threshold is usually determined as a judgment standard, such as the risk assessment of ship collision (Silveira et al., 2013) and identification of traffic hotspots (Zhang et al., 2017). Wang et al. (2023) proposed a risk assessment model to quantify and rank the risk factors during the evacuation process of Human Evacuation from Passenger Ships. Statistical methods often provide the basis for decision-making, while other techniques are needed to implement decisions in specific applications.

In practical research, different methods are usually used together to formulate a comprehensive solution together. Wei et al. (2020) determined the hyper-parameters in a DBSCAN algorithm with the help of a KDE model and then used the DBSCAN algorithm to extract traffic patterns for anomaly detection. Yang et al. (2021) proposed a grid-based

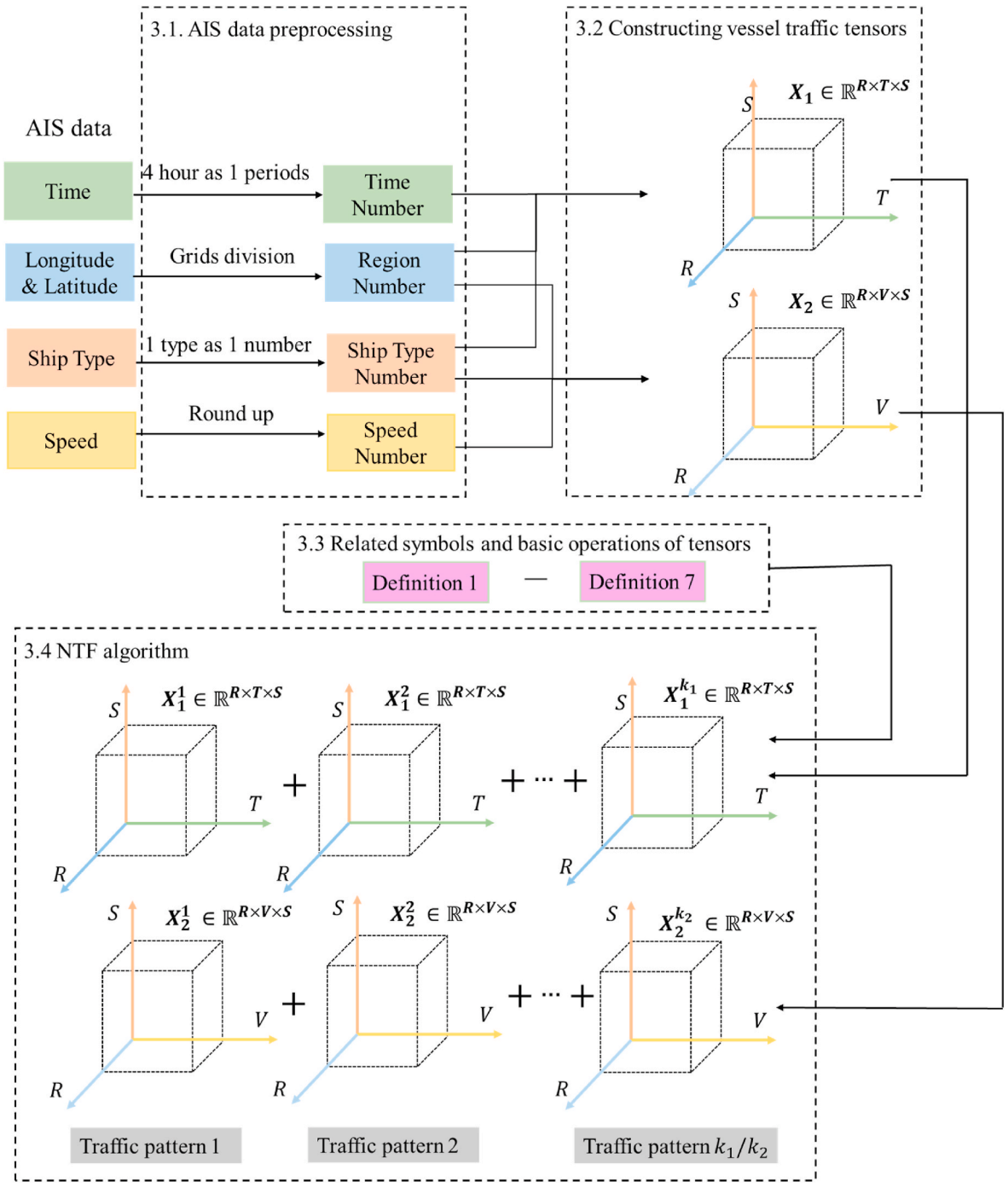


Fig. 1. The proposed framework for vessel traffic pattern extraction.

storage method for AIS data. The non-parametric KDE method was used to obtain the distribution pattern of ship density. Furthermore, such studies (Liu et al., 2021; Qi et al., 2021) used cellular automata for ship traffic behavior modeling.

It is noteworthy that tensor is a useful tool to excavate hidden structures of high dimensionality and stereoscopic information (Jiang et al., 2018). The advantage of the NTF method is that it can be used to directly analyze the distribution characteristics and internal structure of the data itself from a data-driven perspective, investigate the feature extraction of high-dimensional tensors (Cichocki et al., 2009). The NTF method can reproduce the spatio-temporal pattern and dynamic process characteristics of the original data, and reveal the characteristics and inner coupling relationship of each dimension. The tensor-based method has been applied to many maritime traffic studies. For instance, Liu and Chen (2014) drew a global sea surveillance image of traffic patterns in each region within a given timeframe with satellites AIS data using a time-link prediction model based on tensor CANDECOMP/PARAFAC (CP) decomposition. Using SAR image information, Biondi (2016, 2018, 2019) employed the CP tensor decomposition to detect and segment marine targets more accurately. Wang et al. (2017) extracted ship behavior patterns related to location, time, and ship type using spatio-temporal matrix decomposition and tensor decomposition. Although making some achievements in the field of maritime traffic analysis, they focused more on spatio-temporal knowledge and failed to incorporate other important information for a better result on traffic pattern recognition. Given the importance of three-dimensional traffic

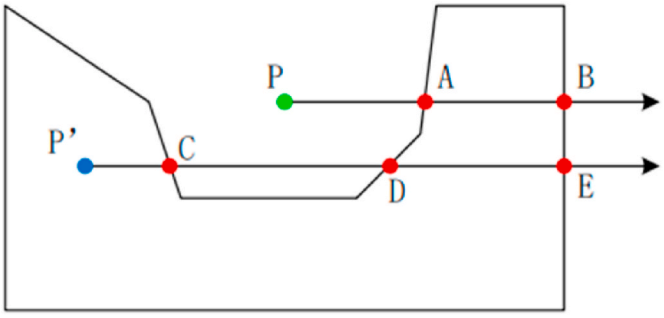


Fig. 2. The ray method of judging the relationship between the point and the polygon.

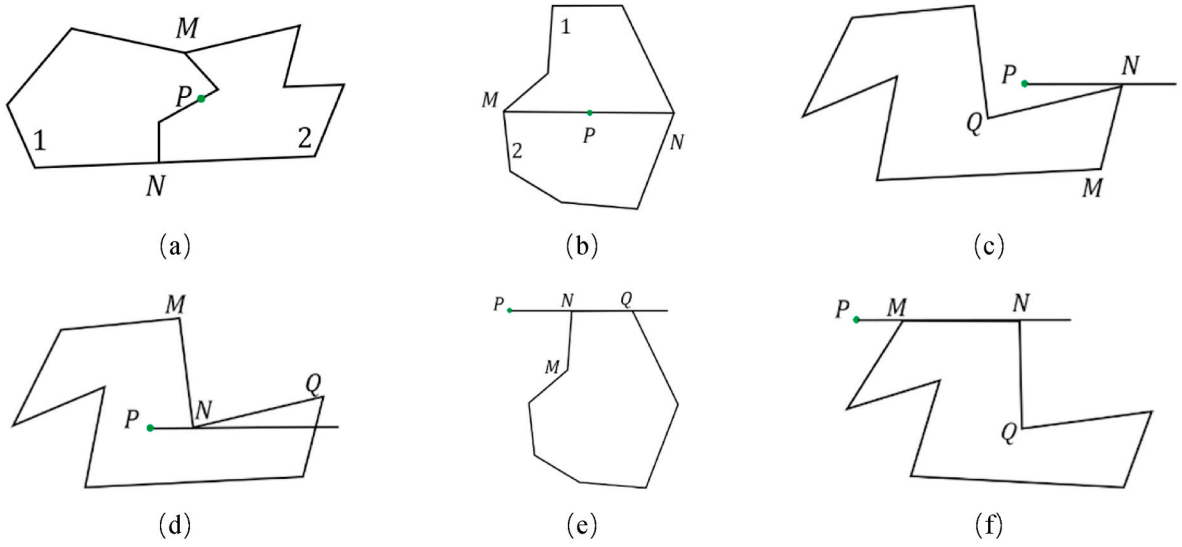


Fig. 3. Special circumstances in the ray method.

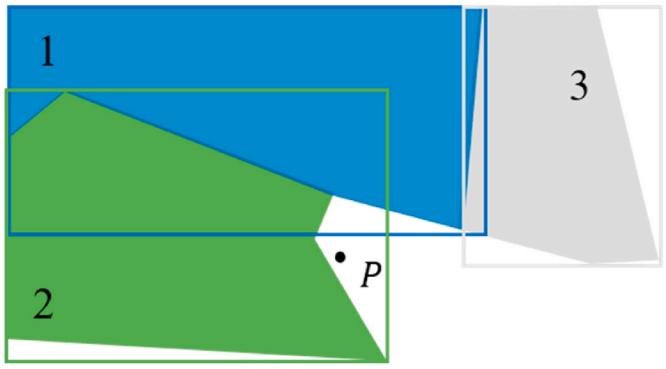


Fig. 4. Three polygons and their respective minimum circumscribed rectangles.

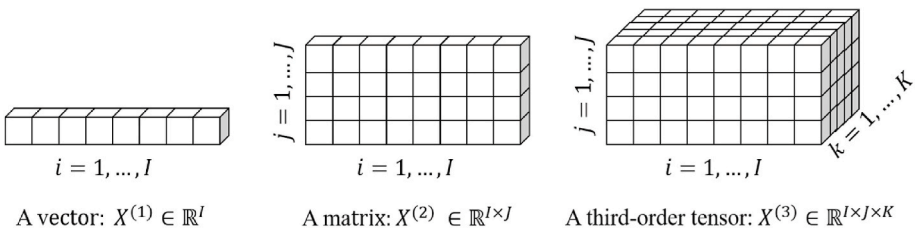


Fig. 5. Tensor diagram.

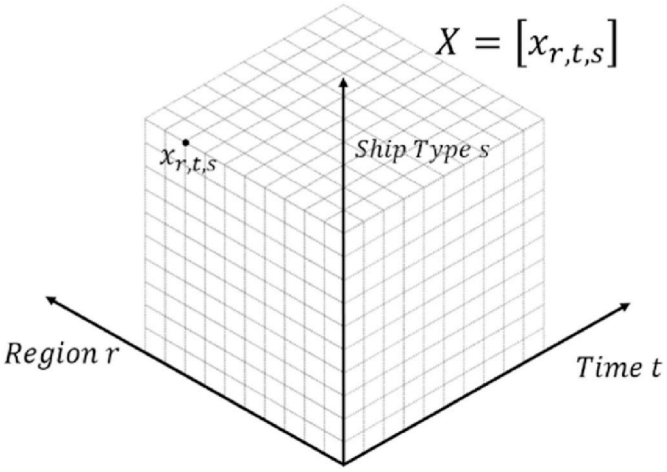


Fig. 6. Structure of 3D ship data tensor of < region, time, ship type >.

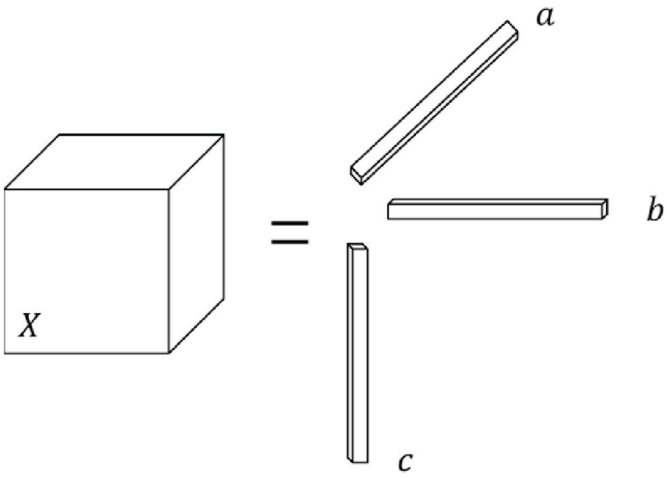


Fig. 7. Rank-1 tensor $X = a \circ b \circ c$.

patterns and K values, the explanation of three-dimensional traffic patterns becomes necessary and beneficial and the influence analysis of K value should also be further addressed in this study.

3. Methodology

This study uses AIS data and a NTF method to study ship traffic patterns. Section 3 explains the whole process of generating tensor using AIS data and obtaining traffic patterns in detail. Section 3.1 provides the process for discretizing and numbering various attributes of AIS data. Using the discretized AIS data, Section 3.2 specifies rules for mapping one-dimensional AIS array to three-dimensional vessel traffic tensors. In order to make the NTF algorithm easy to understand, Section 3.3 introduces the related symbols and basic operations of tensors. Section 3.4 describes the process of decomposing the three-dimensional traffic tensor into non-negative factors, which is the mathematical representation of the traffic pattern. Fig. 1 shows the framework for the proposed vessel traffic pattern extraction method.

3.1. The preprocessing of the AIS data

Before constructing the ship tensor, the AIS data need to be preprocessed. Section 3.1 introduces the separation and marking of the AIS data, describing the process of discretizing and marking the AIS data in detail. The attributes of position, time, ship type, and ship speed studied in this work are explained. In section 3.1.1, the discretization of region is described. Section 3.1.2 focuses on other attributes, such as time, speed and ship type.

3.1.1. Discretization of region

3.1.1.1. Irregular grids division. To better reflect the reality, the research water area is subdivided into irregular grids in this study. The grid-based methods aggregate the traffic data into individual grids, reducing the impact of noise and the burden of subsequent calculation. Some obvious functional regions can be pointed out with the help of experts' knowledge. The grids are divided with the following three pieces of information known: (1) Distribution of functional areas in the study waters, (2) Traffic flow directions in the study waters, and (3) Opinions of pilots and officers from the pilot station and the vessel traffic service (VTS) center. The division process is described in detail in section 4.1 within the

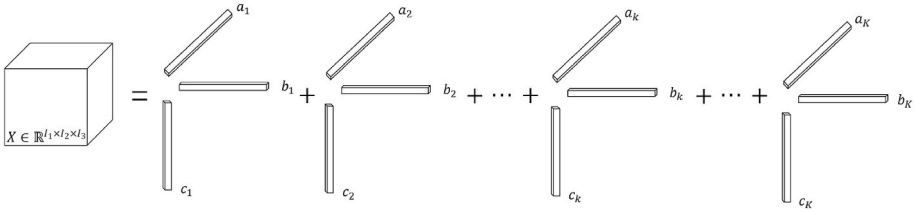


Fig. 8. Decomposition diagram of tensor CP decomposition.

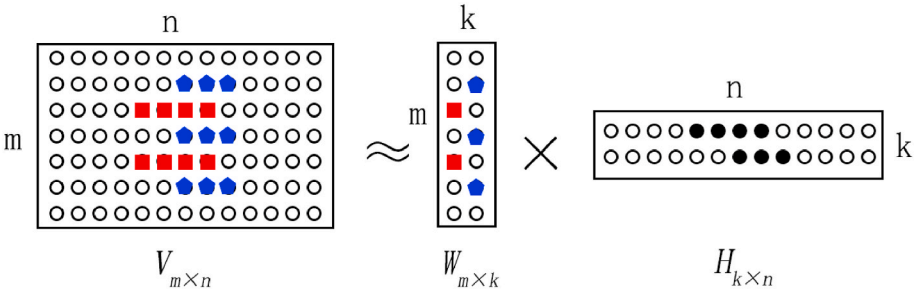


Fig. 9. Non negative matrix factorization diagram.

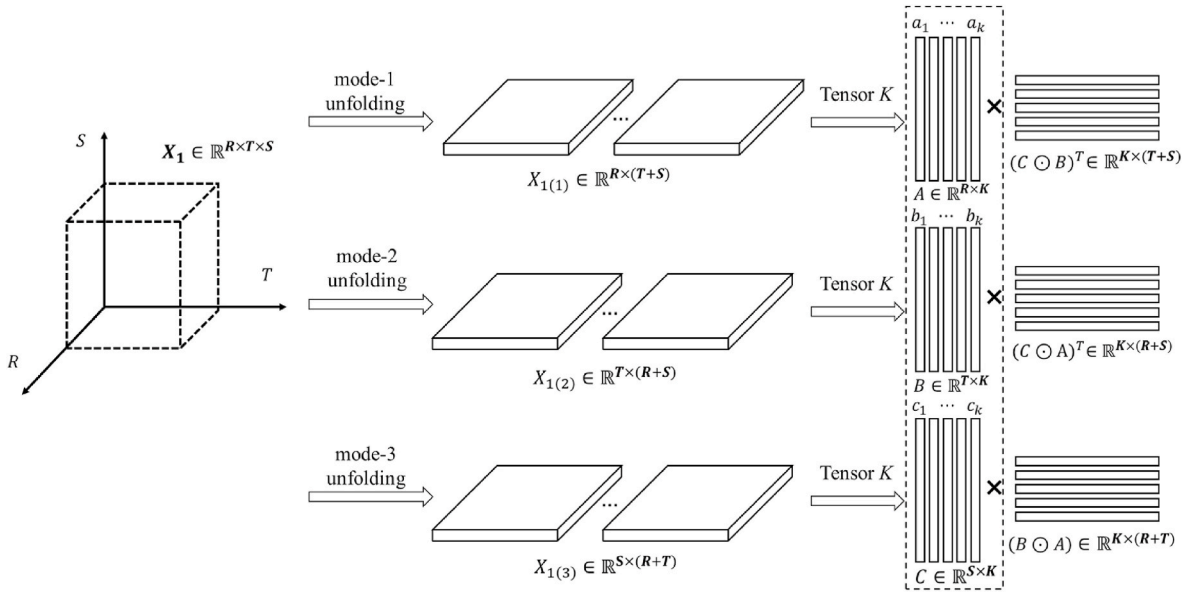


Fig. 10. NTF calculation diagram for $\boldsymbol{X} \in \mathbb{R}^{R\times T\times S}$ 4. Case Study.

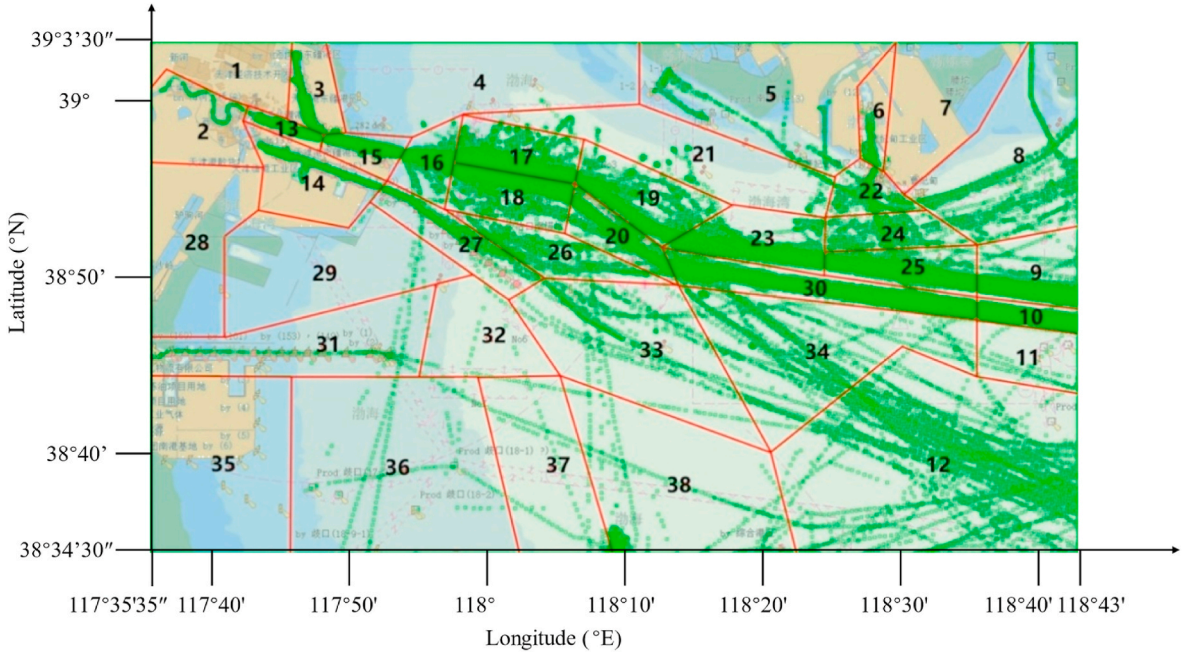


Fig. 11. Study water area irregular division.

Table 1
Ship type number.

Type of ship	Number
Fishing	1
Port tender	2
Cargo ship	3
Passenger ship	4
Recreational boat	5
Tugboat	6
Law Enforcement Craft	7
Oil Tanker	8
Pilot ship	9
High-speed ship	10

Table 2
Time discretization result and time labels.

Hour Date	0-4(h)	4-8(h)	8-12(h)	12-16(h)	16-20(h)	21-24(h)
01–01	1	2	3	4	5	6
01–02	7	8	9	10	11	12
01–03	13	14	15	16	17	18
01–04	19	20	21	22	23	24
01–05	25	26	27	28	29	30
01–06	31	32	33	34	35	36

Table 3
AIS data list before discretization.

Time	Longitude (°)	Latitude (°)	Ship type	Speed (knot)
01-01 08:37	117.8079	38.9663	Cargo ship	3.7
01-01 08:41	117.8134	38.9656	Cargo ship	4.3
01-03 04:59	117.8457	38.7044	Fishing	0.2
01-03 07:33	117.8501	38.6950	Cargo ship	5.6
…	…	…	…	…

Table 4
List of ships attribute label data after discretization.

Time	Region	Ship type	Speed
2	15	3	4
2	15	3	5
14	36	1	1
14	36	3	6
…	…	…	…

Table 5
Type of ships in research waters.

Types of ship	Number of ships	Number of AIS Data Information
Fishing	236	11023
Port tender	14	1847
Cargo ship	898	121361
Passenger ship	11	1010
Recreational boat	2	66
Tugboat	133	34090
Law Enforcement Craft	12	496
Oil Tanker	208	23588
Pilot ship	5	2287
High-speed ship	2	148
Total	1521	195917

Tianjin port-Caofeidian waters. The latitude and longitude coordinates of each grid are recorded with Electronic Chart Display and Information System (ECDIS).

3.1.1.2. The ray method. The latitude and longitude of ships selected from the AIS database are two-dimensional data. However, the spatial attribute is only one dimension in the three-dimensional vessel traffic tensor. Therefore, the two-dimensional latitude and longitude information can be represented by the region number where the ship is located. In this study, the ray method is used to determine the region.

The ray method determines the relationship between the point and the polygon according to the number of intersections, which is the number between a horizontal ray emitted from the point and the polygon. The point is inside the polygon when the number of intersections is odd, and the point is outside the polygon when the number of intersections is even. When the point is on one of the sides of a polygon, it is directly concluded to be inside the polygon.

In Fig. 2, the ray emitted from point P to the horizontal right direction forms two intersections with the polygon at point A and point B. The number of intersections is two, and the point P is outside the polygon. In the same way, the ray emitted from point P' to the horizontal right direction forms three intersections at points C, D, and E with the polygon. The number of intersections is odd, and point P' is inside the polygon.

There are special cases that should be noted when using the ray method. As shown in Fig. 3, polygon 1 and polygon 2 represent two adjacent irregular grids, point P is the point to be judged. It is assumed that the line segment MN is a common edge of two polygons, and M and N are the endpoints of the side respectively. When determining whether there is an intersection of the ray emitted from the point P with side MN:

(1) As shown in Fig. 3(a), when the side MN is the common edge of two polygons, it is considered that point P is in the right polygon 2. In Fig. 3(b), when the side MN is horizontal, then point P is in the upper polygon 1.
(2) If the ray emitted from point P passes one of the endpoints M or N, it must determine the relationship between the two edges of the polygon passing the endpoint and the ray. If the two edges are on the same side of the ray, then the intersection is ignored, while if on opposite sides, it is considered that the ray has an intersection with the edge MN. Just shown in Fig. 3(c), the ray passes through point N, and both edges MN and NQ are below the ray, then point N is not recorded as an intersection. In Fig. 3(d), the edge MN is above the ray and the edge NQ is below, then point N is recorded as an intersection. The edge is considered on the upper side of the ray if one edge coincides with the ray. In Fig. 3(e), edge NQ is considered above the ray emitted from point P. The edges NQ and MN are on both sides of the ray, then point N is recorded as an intersection of a ray with edge MN.
(3) If the ray emitted from point P coincides with the edge MN, as shown in Fig. 3, points M and N are not recorded as an intersection.

It is inefficient to determine the relationship between all the ship position points and each edge of all irregular regions due to the huge amount of data. To solve this problem, the minimum circumscribed rectangles of all regions are drawn, and then the points outside the minimum circumscribed rectangle are excluded. The size of the minimum enclosing rectangle is determined based on the maximum and minimum values of the latitude and longitude within the polygon region. The above processes significantly improve efficiency, as shown in Fig. 4.

In Fig. 4, the blue polygon 1, green polygon 2, and gray polygon 3 represent three irregular grids, and their outer borders form respective minimum enclosing rectangles. In determining the region where point P belongs, the minimum outer rectangle where point P is located has to be determined first. Point P is located in the green minimum enclosed rectangle of polygon 2 instead of polygons 1 and 3. Then it only needs to determine the positional relationship between point P and polygon 2 whose minimum enclosed rectangle contains point P. The relationship between point P and polygon 2 can be judged through the ray method.

3.1.2. Discretization of other attributes

3.1.2.1. Time discretization. According to the seafarer duty handover system, the ship implements an around-the-clock duty system in both sailing and berthing. During a voyage, each shift is fixed as 4 h. The chief officer is on duty from 04:00–08:00 and 16:00–20:00, the second officer from 00:00–04:00 and 12:00–16:00 and the third officer from 08:00–12:00 and 20:00–24:00. According to the duty schedule, every 4 h is defined an interval and a whole day is divided into six intervals.

3.1.2.2. Speed discretization. The speed changes slowly and the fluctuation is small. To reflect the difference in speed dimensions, each knot is treated as a speed interval. The value of ship speed will be rounded up to an integer and marked to facilitate the following analysis. For example, and the speed of 0–0.4 knots is marked as 0, and 0.5–1.4 knots marked as 1.

3.1.2.3. Number of ship types. A ship type is also an important attribute in the study of vessel traffic. Different types of ships sail at different speeds and through different routes. We therefore select the studied ship types and assign a number to each type.

After the pre-processing, AIS data is replaced by a set of the numbers of ship type, time, speed and region. The three-dimensional vessel traffic tensors are established from the set. An element in the set represents an AIS data point.

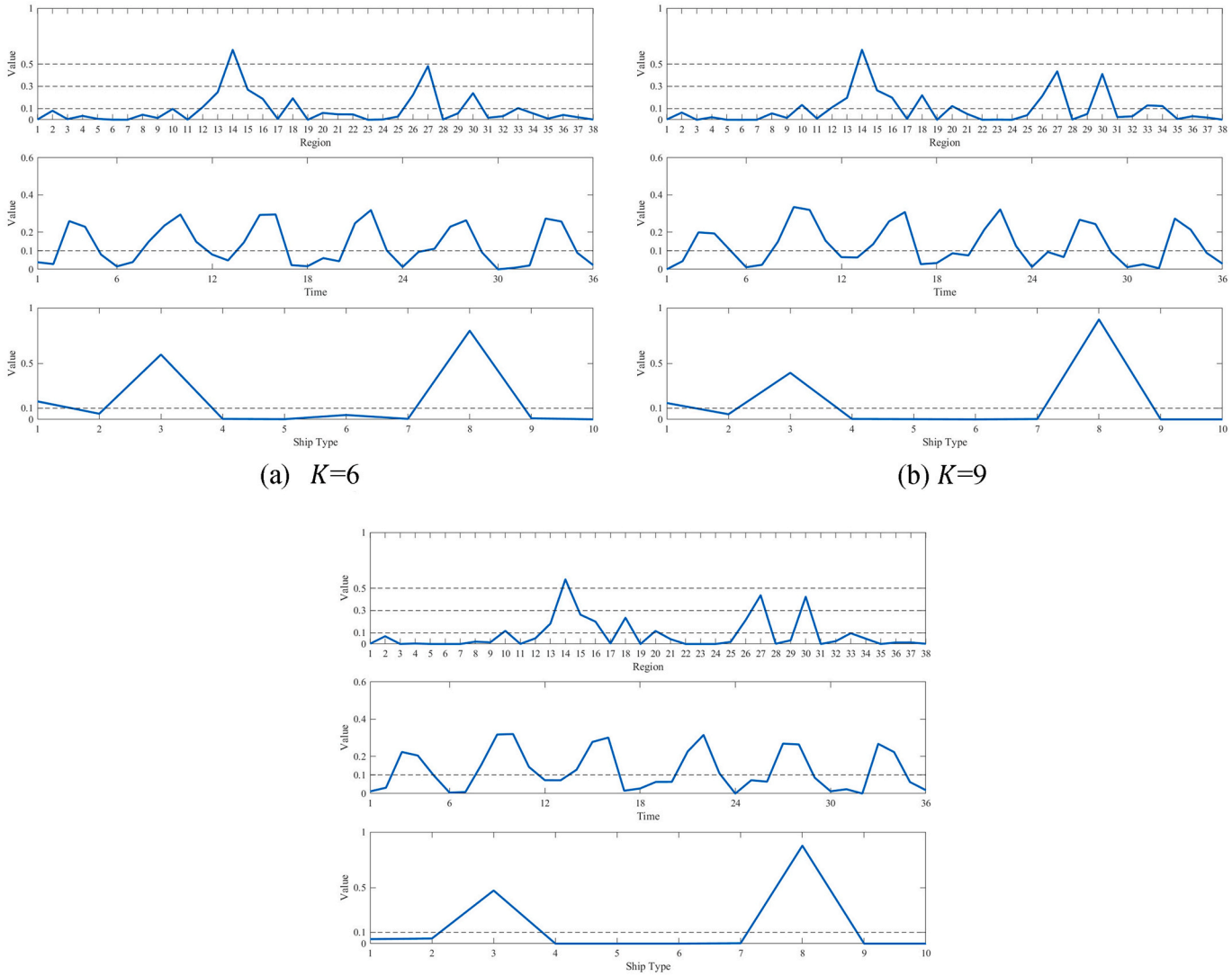


Fig. 12. Comparison of extraction results of the same feature under different K values.

3.2. Constructing three-dimensional vessel traffic tensors

A tensor is a multidimensional array of a data structure. A tensor of N order is an inherent structure of the tensor product of N vector spaces, each of which has its coordinate system. A first-order tensor is a vector, a second-order tensor is a matrix, and tensors of order three or higher are called higher-order tensors (Kolda and Bader, 2009), as shown in Fig. 5.

F is the mapping rule from the set of attributes numbers to vessel traffic tensors.

$$F{:}A=[a_1,\ldots,a_n,\ldots,a_N]\rightarrow\begin{cases} X_1\in\mathbb{R}^{R\times T\times S}=[x_{r,t,s}] \\ X_2\in\mathbb{R}^{R\times V\times S}=[x_{r,v,s}] \end{cases}$$

where $a_n=[r_n,t_n,v_n,s_n]$, A is the set of numbers of ship type, time, speed and region, which is obtained after the separation and marking of the AIS data. The element is represented by a_n and $n=1,\ldots,N$. $X\in\mathbb{R}^{R\times T\times S}$ means a third-order tensor with the structure of $R\times T\times S$, and the element (r,t,s) of $X\in\mathbb{R}^{R\times T\times S}$ is denoted by $x_{r,t,s}$. $X\in\mathbb{R}^{R\times V\times S}$ means a third-order tensor with the structure of $R\times V\times S$, and the element (r,v,s) of $X\in\mathbb{R}^{R\times V\times S}$ is denoted by $x_{r,v,s}$. r, t, v, and s denote the values of the region, time, speed, and ship type, respectively. The range of values of r is $r=1,2,\ldots,R$ and so do t, v and s. The rule F_1 from A to $X\in\mathbb{R}^{R\times T\times S}$ and the rule F_2 from A to $X\in\mathbb{R}^{R\times V\times S}$ are as follows.

Rule 1: $F_1{:}A=[a_1,\ldots,a_n,\ldots,a_N]\rightarrow X_1\in\mathbb{R}^{R\times T\times S}=[x_{r,t,s}]$.

$$f_1(r,t,s,a_n)=\begin{cases} 1, r_n=r \text{ and } t_n=t \text{ and } s_n=s \\ 0, \text{others} \end{cases} \tag{1}$$

$$x_{r,t,s}=\sum_{i=1}^{N} f_1(r,t,s,a_n) \tag{2}$$

Rule 2: $F_2{:}A=[a_1,\ldots,a_n,\ldots,a_N]\rightarrow X_2\in\mathbb{R}^{R\times V\times S}=[x_{r,v,s}]$.

$$f_2(r,v,s,a_n)=\begin{cases} 1, r_n=r \text{ and } v_n=v \text{ and } s_n=s \\ 0, \text{others} \end{cases} \tag{3}$$

$$x_{r,v,s}=\sum_{i=1}^{N} f_2(r,v,s,a_n) \tag{4}$$

Taking the construction of the tensor $X_1\in\mathbb{R}^{R\times T\times S}$ of $<$ region, time, ship type $>$ as an example. Fig. 6 shows the structure of a three-dimensional ship traffic tensor $X_1\in\mathbb{R}^{R\times T\times S}$ for $<$ region, time, ship type $>$. It is noteworthy that the value of the corresponding element pluses 1 when each point in the set of attributes numbers is mapped into the tensor. In other words, all trajectory points denoted by a ship with

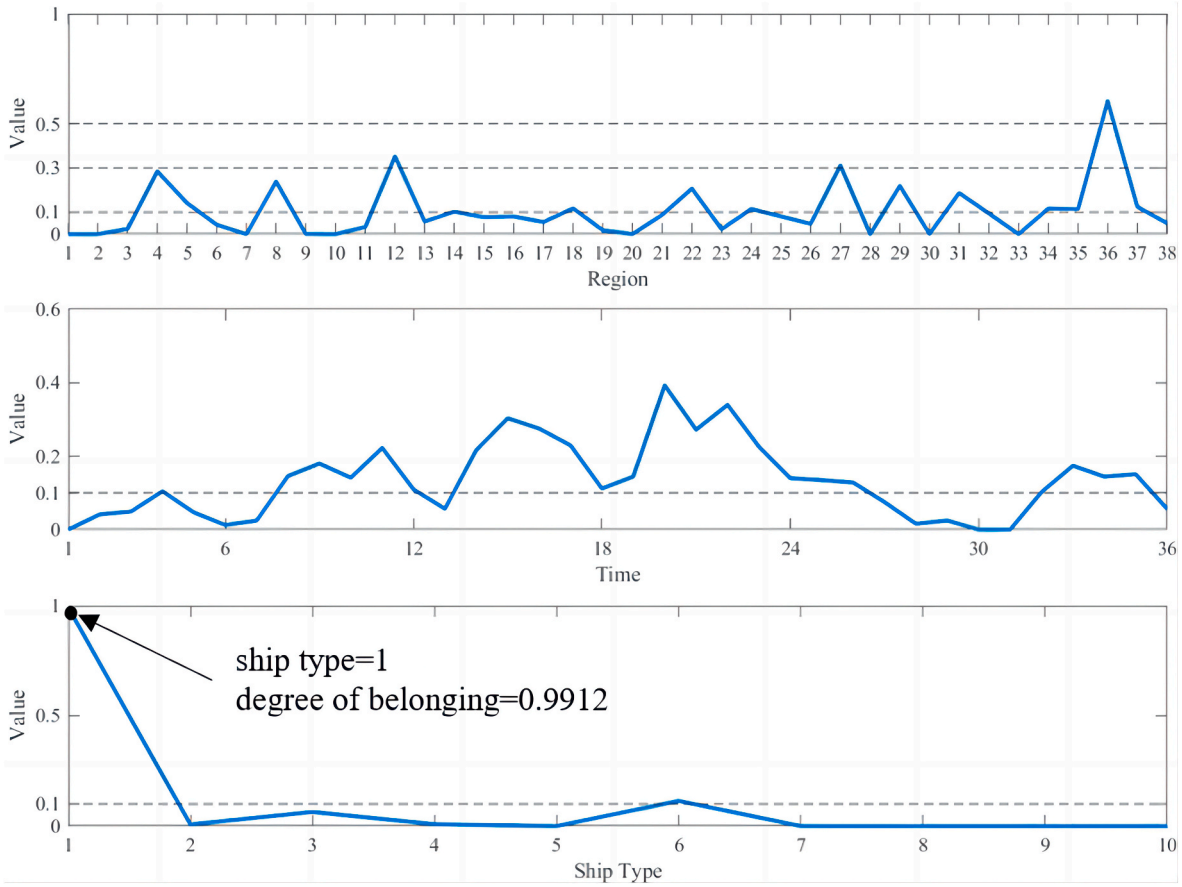


Fig. 13. A new pattern appears in the co-clustering results of $K = 10$.

ship type s located in region r in time t are selected as experimental data. $x_{r,t,s}$ records the number of all points with ship type s, region r, and time t.

The reason for specifying such a mapping rule is as follows :

(1) This study is expected to discover more information about vessel traffic patterns. All data points are mapped into the tensors to prevent the loss of any value information. The distributions of traffic attributes are also not changed.
(2) The simplification of data should refer to research requirements according to specific scenarios. In this study, the duplicated data will not significantly increase the burden of calculation, because they only result in the increased value of $x_{r,t,s}$, instead of the increased dimension of data. The increased computational burden of this case is therefore limited.

3.3. Related symbols and basic operations of tensors

The basic algebraic knowledge and operation symbols involved in tensor factorization are introduced below. Supposing there are N-order tensors $X, Y \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ (Kolda and Bader, 2009).

Definition 1. The F-norm of a tensor X is:

$$\|X\| = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \ldots i_N}^2} \tag{5}$$

Definition 2. The matricization, also known as the expansion of the matrix form of a tensor, is the process of reordering the N-order tensor into a matrix. Tensor's mode-n matricization $X \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is represented by $X_{(n)}$. The position of tensor element $x_{i_1 i_2 \ldots i_N}$ in a tensor is $(i_1, i_2, \ldots, i_n, \ldots, i_N)$ $(i_n \in [1, I_N])$. When being mapped to the mode-n matrix element (i,j) it is represented as:

$$\begin{cases} i = i_n \\ j = 1 + \sum_{\substack{k=1 \\ k \neq n}}^{N} (i_k - 1) J_k \text{ with } J_k = \prod_{\substack{m=1 \\ m \neq n}}^{k-1} I_m \end{cases} \tag{6}$$

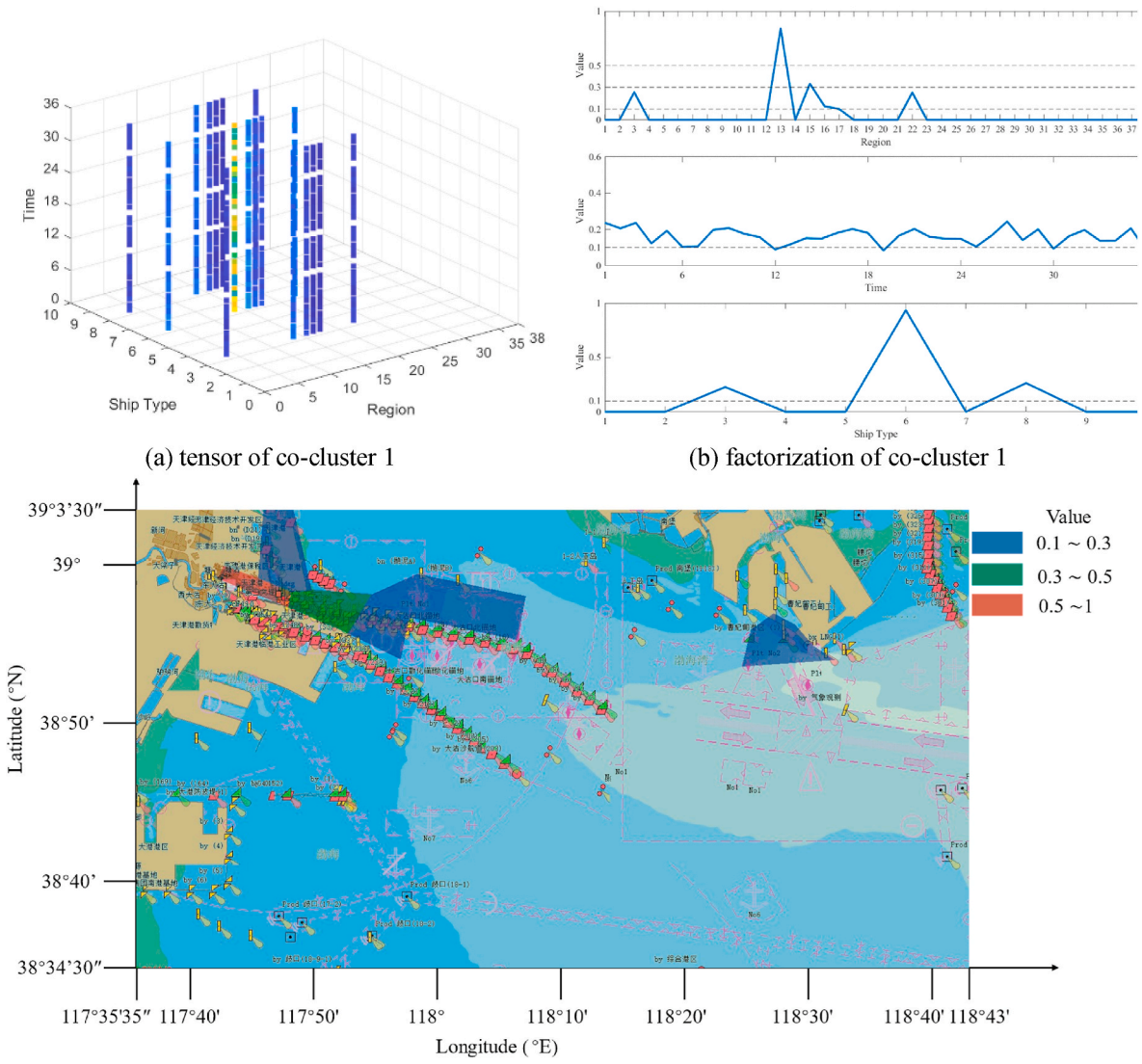
Slices are two-dimensional sections of a tensor, and the k th frontal slice of a third-order tensor is denoted as $X_{:k}$ or X_k. For example, let the frontal slices $X \in \mathbb{R}^{3 \times 4 \times 2}$ be

$$X_1 = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix}, X_2 = \begin{bmatrix} 13 & 14 & 15 & 16 \\ 17 & 18 & 19 & 20 \\ 21 & 22 & 23 & 24 \end{bmatrix}$$

Then the three mode-n unfolding are $X_{(1)} \in \mathbb{R}^{3 \times 8}$、 $X_{(2)} \in \mathbb{R}^{4 \times 6}$ and $X_{(3)} \in \mathbb{R}^{2 \times 12}$ are:

$$X_{(1)} = \begin{bmatrix} 1 & 2 & 3 & 4 & 13 & 14 & 15 & 16 \\ 5 & 6 & 7 & 8 & 17 & 18 & 19 & 20 \\ 9 & 10 & 11 & 12 & 21 & 22 & 23 & 24 \end{bmatrix}$$

$$X_{(2)} = \begin{bmatrix} 1 & 5 & 9 & 13 & 17 & 21 \\ 2 & 6 & 10 & 14 & 18 & 22 \\ 3 & 7 & 11 & 15 & 19 & 23 \\ 4 & 8 & 12 & 16 & 20 & 24 \end{bmatrix}$$



(a) tensor of co-cluster 1

(b) factorization of co-cluster 1

(c) Spatial distribution of co-cluster 1

Fig. 14. Co-cluster 1 of < region, time, ship type>.

$$X_{(3)} = \begin{bmatrix} 1 & 5 & 9 & 2 & 6 & \cdots & 11 & 4 & 8 & 12 \\ 13 & 17 & 21 & 14 & 18 & \cdots & 23 & 16 & 20 & 24 \end{bmatrix}$$

Definition 3. Considering a tensor $X \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, and a matrix $A \in \mathbb{R}^{J \times I_n}$, then the mode-n product of the tensor is $X \times_n A$, whose result is a tensor of size $I_1 \times \ldots \times I_{n-1} \times J \times I_{n+1} \times \ldots \times I_N$. The transformation of mode-n product satisfies Eq. (7).

$$Y = X \times_n A \Leftrightarrow Y_{(n)} = AX_{(n)} \tag{7}$$

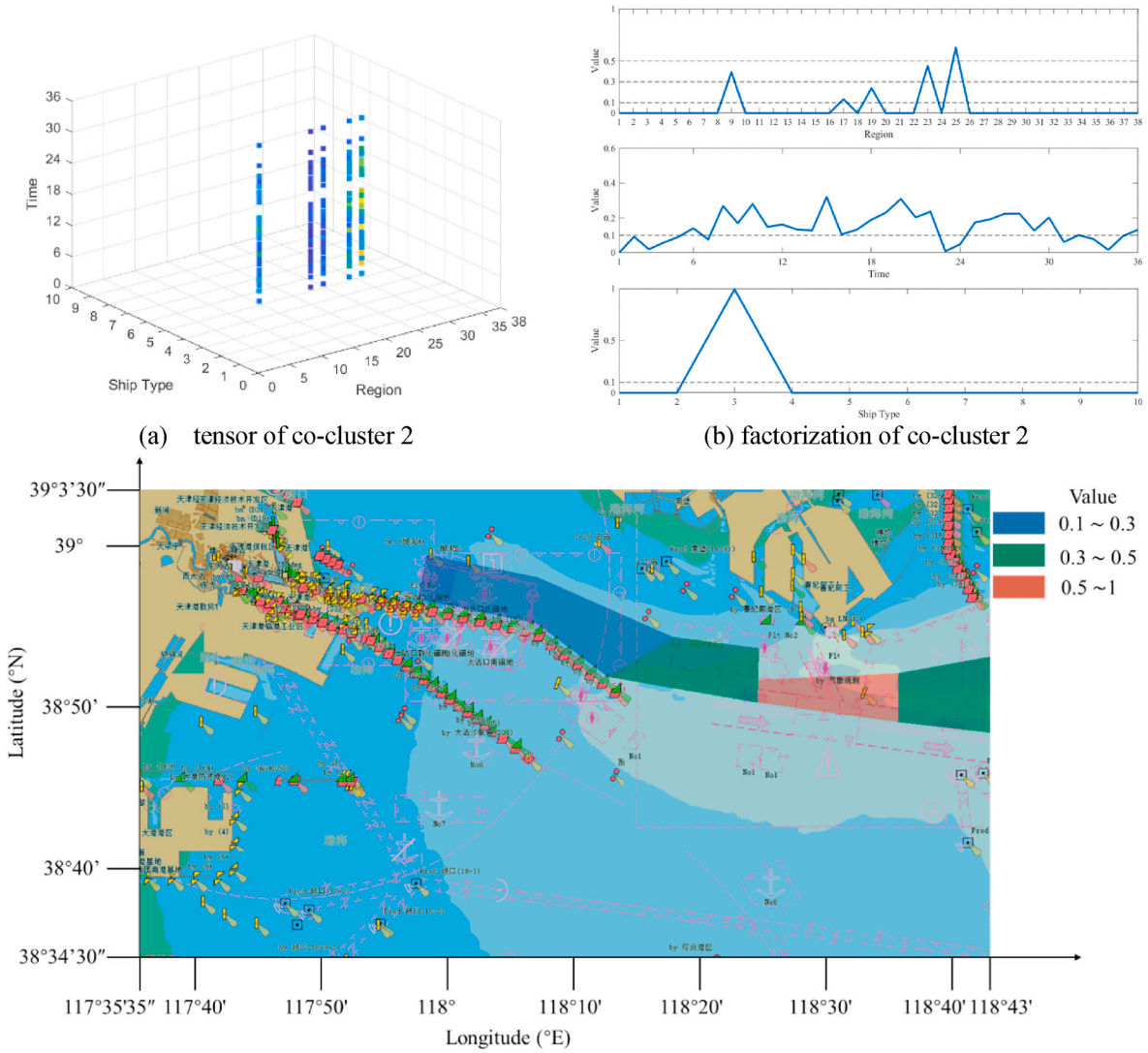
Definition 4. In general, the tensor product refers to the Kronecker product. The symbolic representation of the Kronecker product of a matrix is $\otimes$. Considering the matrices $A \in \mathbb{R}^{M \times N}$ and $B \in \mathbb{R}^{P \times Q}$, their Kronecker product is expressed as $A \otimes B$, whose calculation equation is shown as:

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1N}B \\ a_{21}B & a_{22}B & \cdots & a_{2N}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1}B & a_{M2}B & \cdots & a_{MN}B \end{bmatrix} = [a_1 \otimes b_1 \; a_1 \otimes b_2 \cdots a_N \otimes b_{Q-1} \; a_N \otimes b_Q] \tag{8}$$

Definition 5. The symbolic representation of the Khatri-Rao product (KR product) of a matrix is $\odot$. Considering matrices $A \in \mathbb{R}^{M \times N}$ and $B \in \mathbb{R}^{P \times N}$, KR product is denoted by $A \odot B$. The result is a matrix of size $(MP) \times N$ can be defined as:

$$A \odot B = [a_1 \otimes b_1 \; a_2 \otimes b_2 \cdots a_N \otimes b_Q] \tag{9}$$

Definition 6. Considering matrices A and B, both of size $I \times J$, their Hadamard product is denoted by $A * B$, which size is also $I \times J$ and defined as:



(a) tensor of co-cluster 2 (b) factorization of co-cluster 2

(c) Spatial distribution of co-cluster 2

Fig. 15. Co-cluster 2 of < region, time, ship type>.

$$A * B = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1N}b_{1N} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2N}b_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1}b_{M1} & a_{M2}b_{M2} & \cdots & a_{MN}b_{MN} \end{bmatrix} \tag{10}$$

Definition 7. Here † denotes the Moore–Penrose pseudoinverse of Matrix. If there is a Matrix B of the same type as the transpose matrix a of A, and these two matrices satisfy $ABA = A, BAB = B$, then the Matrix B is called the pseudo-inverse of the Matrix A, also known as the generalized inverse Matrix. The Usual Transformations for tensor operations are as:

$$\begin{cases} (A \otimes B)(C \otimes D) = AC \otimes BD \\ (A \otimes B)^{\dagger} = A^{\dagger} \otimes B^{\dagger} \\ A \odot B \odot C = (A \odot B) \odot C = A \odot (B \odot C) \\ (A \odot B)^{T}(A \odot B) = A^{T}A * B^{T}B \\ (A \odot B)^{\dagger} = \left((A^{T}A) * (B^{T}B)\right)^{\dagger}(A \odot B)^{T} \end{cases} \tag{11}$$

3.4. Non-negative tensor factorization algorithm

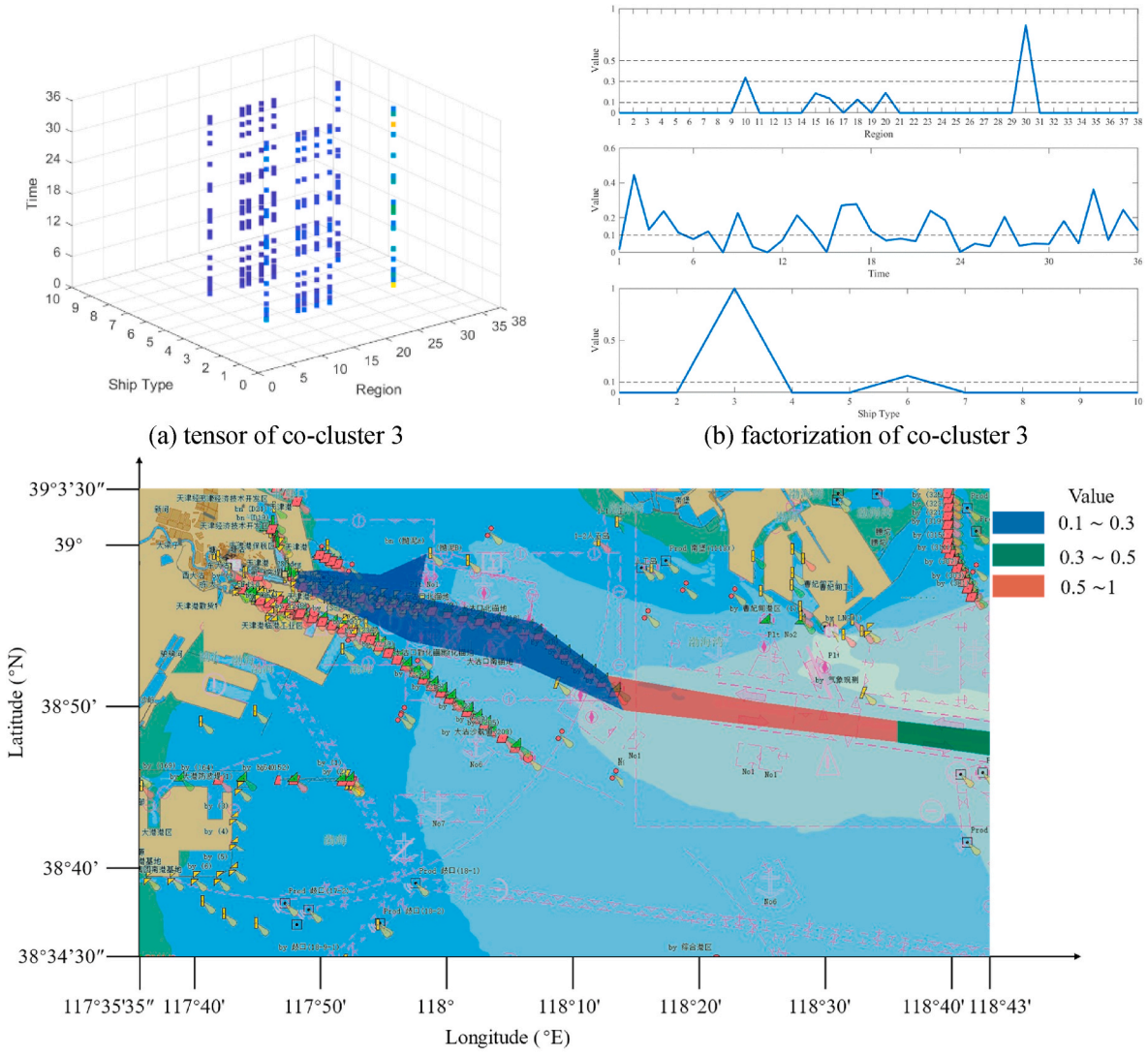
CP decomposition is a basic form of tensor decomposition. The NTF used in this study is a deformation of the PARAFAC model, which is both based on the decomposition of Rank-1. The difference is that the PARAFAC model rotates the entire space of the data, and the result may be negative. By imposing non-negative constraints on the coefficients, the NTF method in this study can better characterize the meaning and structure embedded in the data, which is expected to achieve an approximate approximation to the original data (Yu et al., 2011).

Considering an N-order tensor $X \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, which is called a rank-1 tensor when it can be expressed by the outer product of N vectors.

$$X = i_1 \circ i_2 \circ \ldots \circ i_N \tag{12}$$

The symbol ∘ is the outer product, which means that each element of the tensor is the product of the corresponding vector element. A third-order rank-1 tensor X is shown in Fig. 7, and $X = a \circ b \circ c$.

The main idea of tensor CP decomposition is to express a tensor as the sum of a finite number of rank-one tensors. For a third-order tensor $X \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, the CP decomposition is illustrated in Fig. 8. The rank of the tensor K can be understood as the number of rank-1 tensors obtained by decomposition. a_k is a vector with I_1 elements, b_k is a vector with I_2 elements, and c_k is a vector with I_3 elements. The tensor $X \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ is written as:



(a) tensor of co-cluster 3 (b) factorization of co-cluster 3

(c) Spatial distribution of co-cluster 3

Fig. 16. Co-cluster 3 of < region, time, ship type >.

$$X \approx \sum_{k=1}^{K} a_k \circ b_k \circ c_k \tag{13}$$

Euclidean distance is used to define the loss function. The optimization problem for the tensor CP decomposition is:

$$\min_{\widehat{X}} \|X - \widehat{X}\|_F^{2^2} \text{ with } \widehat{X} = \sum_{k=1}^{K} a_k \circ b_k \circ c_k \tag{14}$$

The results of tensor CP decomposition have negative values. To make the results of tensor decomposition more interpretable, the non-negativity needs to be guaranteed. Lee and Seung (2000) proposed the multiplication update rule for solving the Nonnegative Matrix Factorization (NMF). This method converts the additive form of gradient descent into the multiplicative form so that the iterative process and the result are guaranteed to be nonnegative as long as the initial matrixes are nonnegative. Detailed derivation and convergence proof of the method is given by Lee and Seung. In NMF, for a non-negative data matrix $V_{m\times n}$. , there is a non-negative matrix $W_{m\times k}$ and a non-negative matrix $H_{k\times n}$, their relationship is shown in Equation (15) :

$$V_{m\times n} \approx W_{m\times k} H_{k\times n} \tag{15}$$

where, W is the basis matrix, H is the coefficient matrix, k is the rank of the matrix obtained by decomposition. k is required to satisfy $k < \min(m,n)$. $W_{m\times k} H_{k\times n}$ represents the approximate reconstruction matrix, and the schematic diagram of non-negative matrix decomposition is shown in Fig. 9. The columns in the original matrix $V_{m\times n}$ can be represented by the weight sum of all columns of basis matrix $W_{m\times k}$. The weight coefficient is corresponding column in matrix $H_{k\times n}$. Each column in $V_{m\times n}$ is a local feature of the original data $V_{m\times n}$.The larger the value of an element in a column vector, the more prominent the element is in the feature and the more dominant it is in the feature. This value is defined as the relative degree of belonging in Section 5.

In this study, this method is extended to the three-dimensional tensor, exploring the iteration rules of non-negative tensor factorization and obtaining computational results.

Let the matrix $A \in \mathbb{R}^{I_1\times K} = [a_1, a_2, \cdots, a_K]$, Matrix $B \in \mathbb{R}^{I_2\times K} = [b_1, b_2, \cdots, b_K]$ and matrix $C \in \mathbb{R}^{I_3\times K} = [c_1, c_2, \cdots, c_K]$. When the tensor $X \in \mathbb{R}^{I_1\times I_2\times I_3}$ is unfolded in mode-1 form, the optimization problem can be written as Equation (16) :

$$\min_{A} \left\| X_{(1)} - A(C \odot B)^T \right\|_F^{2^2} \tag{16}$$

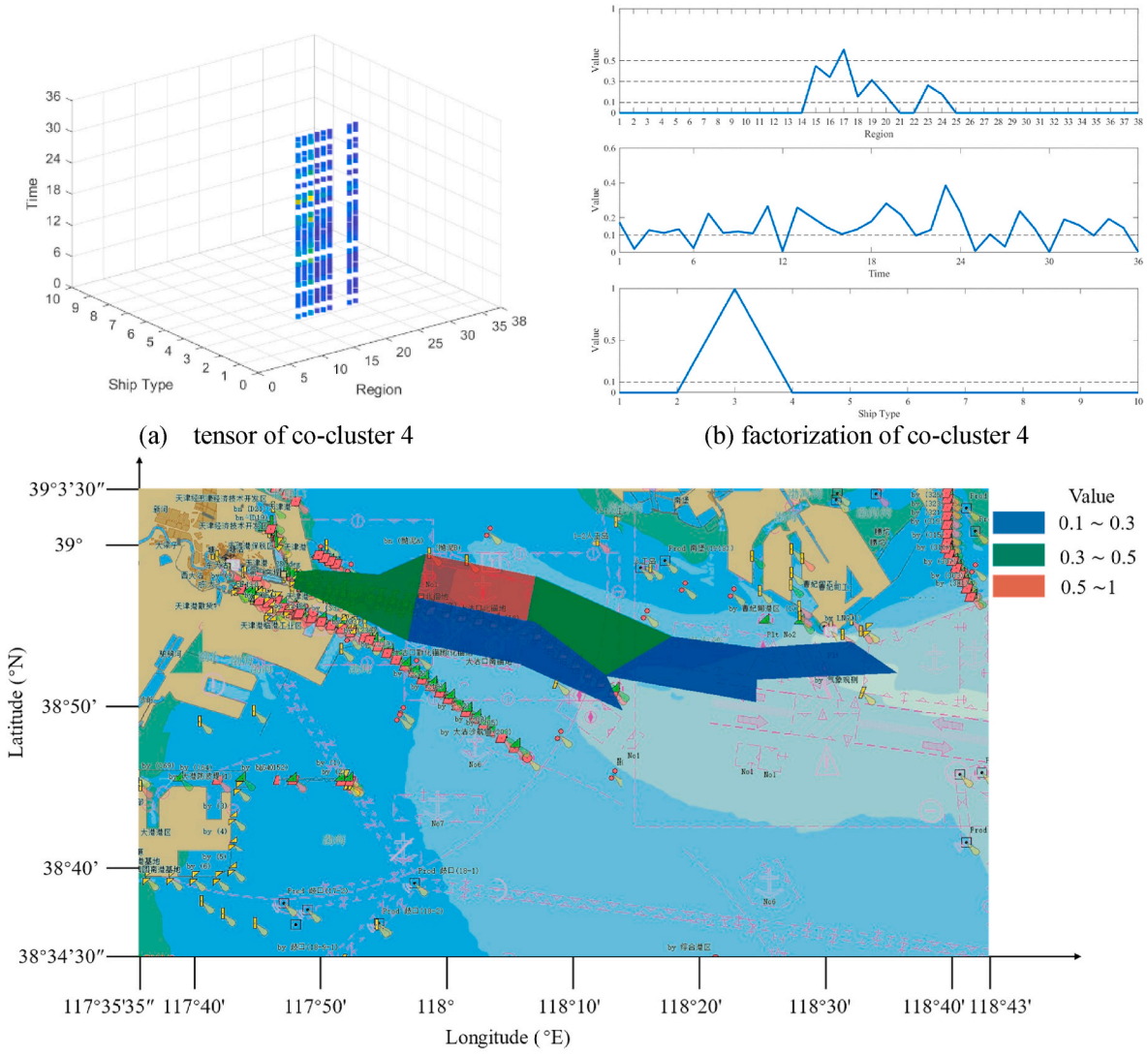


Fig. 17. Co-cluster 4 of < region, time, ship type>.

Through the expansion of the matrix form of a tensor, the problem of NTF is transformed into NMF. Matrix $A \in \mathbb{R}^{I_1 \times K}$ corresponding the basis matrix in NMF. Each column of matrix A is a local feature.

The minimum value is obtained when and only when $X_{(1)}$ and $A(C \odot B)^T$ are equal. The gradient H_A is obtained by taking the partial derivative of A. The step size in the gradient direction is η_A. The iterative equation for the matrix A is:

$$A \leftarrow A - H_A \eta_A \tag{16a}$$

H_A and η_A can be expressed as :

$$H_A = \left[X_{(1)} - A(C \odot B)^T\right][-(C \odot B)] \tag{17}$$

$$\eta_A = \frac{A}{A(C \odot B)^T (C \odot B)} \tag{18}$$

The iterative equation for the matrix A based on the multiplicative update rule can be obtained as:

$$A \leftarrow A * \frac{X_{(1)}(C \odot B)}{A\left(C^T C * B^T B\right)} \tag{19}$$

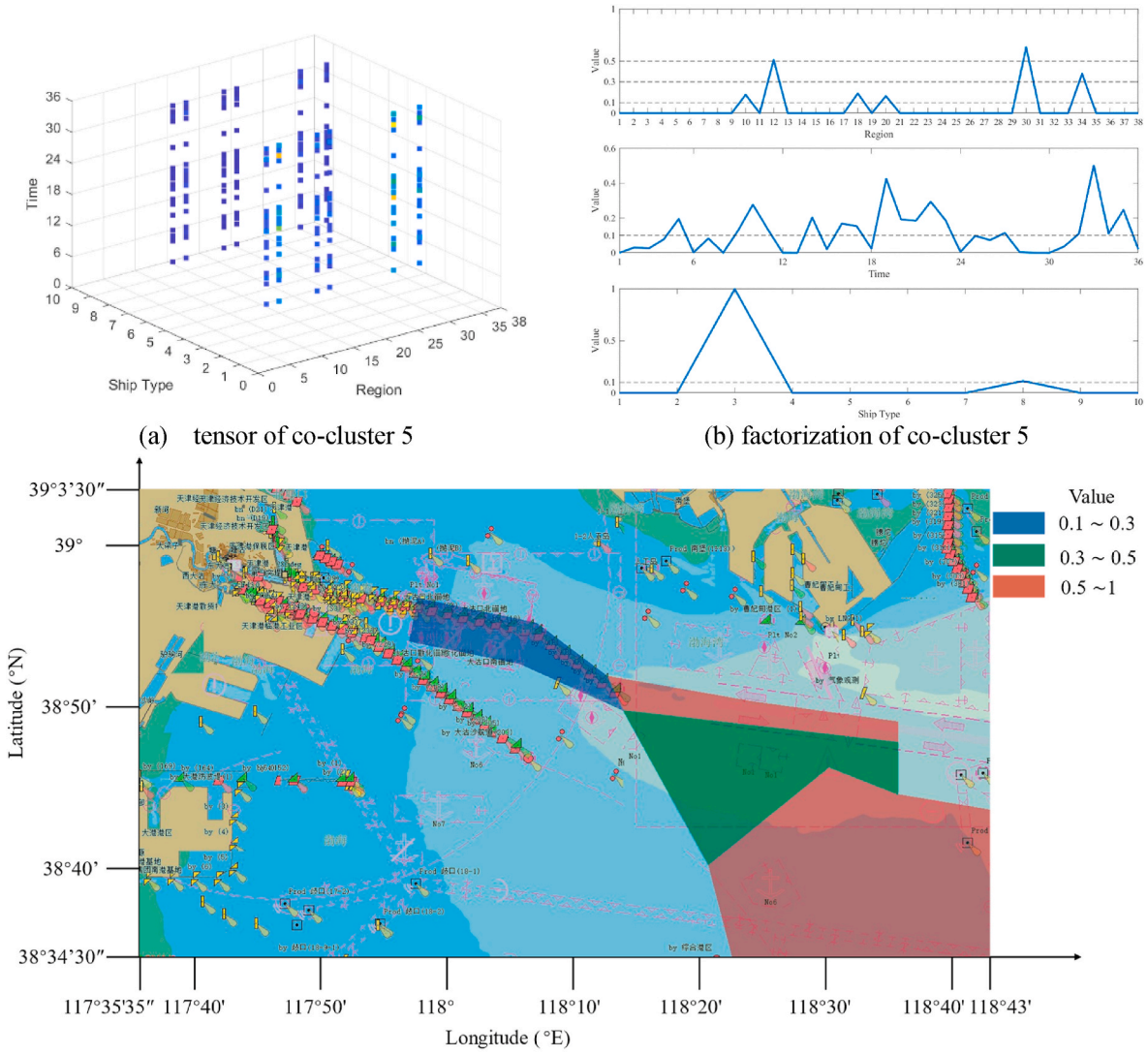
Similarly, the iterative equation for matrix B and matrix C are:

$$B \leftarrow \boldsymbol{B} * \frac{\boldsymbol{X}_{(2)}(\boldsymbol{C} \odot \boldsymbol{A})}{\boldsymbol{B}\left(\boldsymbol{C}^T \boldsymbol{C} * \boldsymbol{A}^T \boldsymbol{A}\right)} \tag{20}$$

$$\boldsymbol{C} \leftarrow \boldsymbol{C} * \frac{\boldsymbol{X}_{(3)}(\boldsymbol{B} \odot \boldsymbol{A})}{\boldsymbol{C}\left(\boldsymbol{B}^T \boldsymbol{B} * \boldsymbol{A}^T \boldsymbol{A}\right)} \tag{21}$$

In this way, only multiplication exists in the iterative process, and as long as the initial value of matrix $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are non-negative, the result of the iteration has to be non-negative.

For the tensor $\boldsymbol{X} \in \mathbb{R}^{\boldsymbol{R} \times \boldsymbol{T} \times \boldsymbol{S}}$, matrix $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{R} \times \boldsymbol{K}}$ represents the regional characteristics of the obtained traffic pattern. Matrix $\boldsymbol{B} \in \mathbb{R}^{\boldsymbol{T} \times \boldsymbol{K}}$ represents the temporal characteristics and matrix $\boldsymbol{C} \in \mathbb{R}^{\boldsymbol{S} \times \boldsymbol{K}}$ represents the composition of the ship type. Fig. 10 shows the NTF calculation diagram for $\boldsymbol{X} \in \mathbb{R}^{\boldsymbol{R} \times \boldsymbol{T} \times \boldsymbol{S}}$.



(a) tensor of co-cluster 5 (b) factorization of co-cluster 5

(c) Spatial distribution of co-cluster 5

Fig. 18. Co-cluster 5 of <region, time, ship type >.

The calculation process for the non-negative tensor factorization based is shown in Algorithm 1. The initial non-negative matrix $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are generated randomly. The maximum number of iterations is 500, and the calculation is terminated early when the change in relative error change is less than 0.0001 compared to the result of the previous iteration.

Result: Matrices $\boldsymbol{A} \in \mathbb{R}^{I_1 \times K}$、 $\boldsymbol{B} \in \mathbb{R}^{I_2 \times K}$、 $\boldsymbol{C} \in \mathbb{R}^{I_3 \times K}$

Input: The tensor $\boldsymbol{X} \in \mathbb{R}^{R \times T \times S}$， rank $\boldsymbol{K}$

1: Initialize $\boldsymbol{\varepsilon} = \mathbf{1}e^{-4}$

2: $\boldsymbol{X}_{(1)} \leftarrow$ unfold $\boldsymbol{X}$ in mode-1

3: $\boldsymbol{X}_{(2)} \leftarrow$ unfold $\boldsymbol{X}$ in mode-2

4: $\boldsymbol{X}_{(3)} \leftarrow$ unfold $\boldsymbol{X}$ in mode-3

5: Generated non-negative matrices $\boldsymbol{A_0}$、 $\boldsymbol{B_0}$、 $\boldsymbol{C_0}$

6: for $\boldsymbol{i} \boldsymbol{\epsilon} \{1, \ldots, \mathbf{499}\}$ do

7. $\boldsymbol{A}_i \leftarrow \boldsymbol{A}_{i-1} * \dfrac{\boldsymbol{X}_{(1)}(\boldsymbol{C}_{i-1} \odot \boldsymbol{B}_{i-1})}{\boldsymbol{A}_{i-1}(\boldsymbol{C}_{i-1}{}^T \boldsymbol{C}_{i-1} * \boldsymbol{B}_{i-1}{}^T \boldsymbol{B}_{i-1})}$

8. $\boldsymbol{B}_i \leftarrow \boldsymbol{B}_{i-1} * \dfrac{\boldsymbol{X}_{(2)}(\boldsymbol{C}_{i-1} \odot \boldsymbol{A}_{i-1})}{\boldsymbol{B}_{i-1}(\boldsymbol{C}_{i-1}{}^T \boldsymbol{C}_{i-1} * \boldsymbol{A}_{i-1}{}^T \boldsymbol{A}_{i-1})}$

9. $\boldsymbol{C}_i \leftarrow \boldsymbol{C}_{i-1} * \dfrac{\boldsymbol{X}_{(3)}(\boldsymbol{B}_{i-1} \odot \boldsymbol{A}_{i-1})}{\boldsymbol{C}_{i-1}(\boldsymbol{B}_{i-1}{}^T \boldsymbol{B}_{i-1} * \boldsymbol{A}_{i-1}{}^T \boldsymbol{A}_{i-1})}$

10. $\boldsymbol{\varepsilon}_i \leftarrow \left\| \boldsymbol{X} - \sum_{k=1}^{K} \boldsymbol{a}_k \circ \boldsymbol{b}_k \circ \boldsymbol{c}_k \right\|_F$

(*continued on next column*)

(*continued*)

11: if $|\varepsilon_i - \varepsilon_{i-1}| < 1e^{-4}$ then

12. break

13. end for 14. return Matrix $A_i \in \mathbb{R}^{I \times K}$、 $B_i \in \mathbb{R}^{I_2 \times K}$、 $C_i \in \mathbb{R}^{I_3 \times K}$

Algorithm 1. Non-negative Tensor Factorization

In this study, Tianjin Port is selected as the research water area to verify the effectiveness of the proposed method. Two perspectives < region, time, ship type > and < region, speed, ship type > are chose to analyze from the definition of the traffic pattern as they both present the increasing interests from the perspective of traffic management practice.

3.5. The separation and marking of the AIS data

The study area includes part of Tianjin Port and part of Caofeidian Port. It ranges from 117°35′35″ to 118°43′0″ in longitude and from 38°34′30″ to 39°3′30″ in latitude. The area is divided into 38 irregular regions as shown in Fig. 11. The green hollow squares indicate the ship trajectories.

First, according to the main anchorages, waterways, and the areas with traffic separation schemes, the following important functional

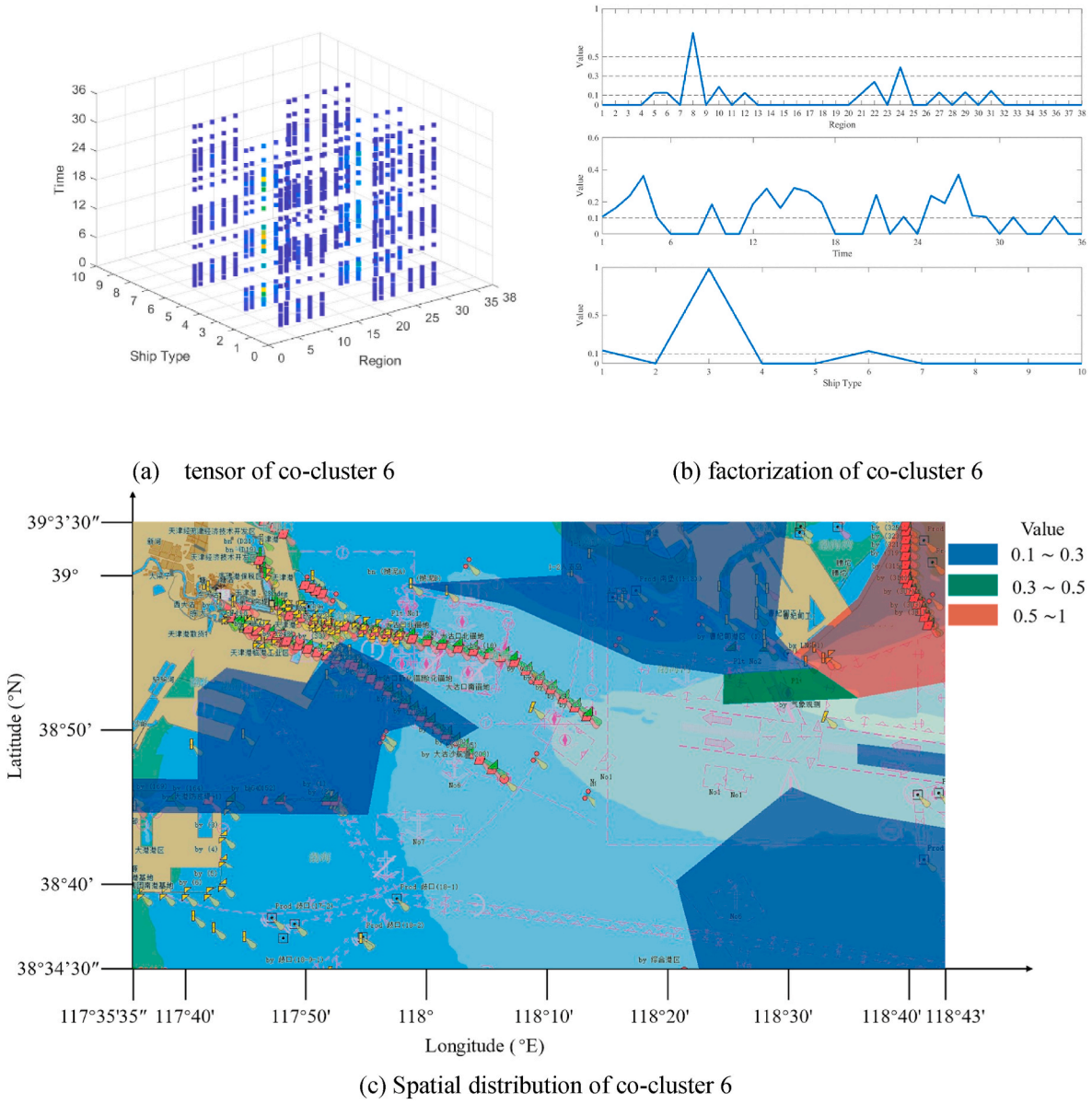


Fig. 19. Co-cluster 6 of <region, time, ship type>.

areas are divided and regulated. The main functional areas contained in the Tianjin Port include Dagukou north anchorage (region 17), Dagukou south anchorage and Dagukou bulking anchorage (region 18), the Beijiang port area (region 3), Nanjiang port area, Dongjiang port area and a channel between them (regions 13, 15 and16), Dagukou port area (region 14) and a waterway into it (region 27). In Caofeidian Port, there are an east anchorage (region 8) and a west anchorage (the adjacent parts of regions 21 and 23). In addition to the above, some regions are set up with a traffic separation scheme. Regions 9, 23, and 25 are located in the waterway into Tianjin Port, while regions 10 and 30 are located in the leaving one. Regions 22 and 24 are an exclusive waterway involving the traffic into and out of Caofeidian Port.

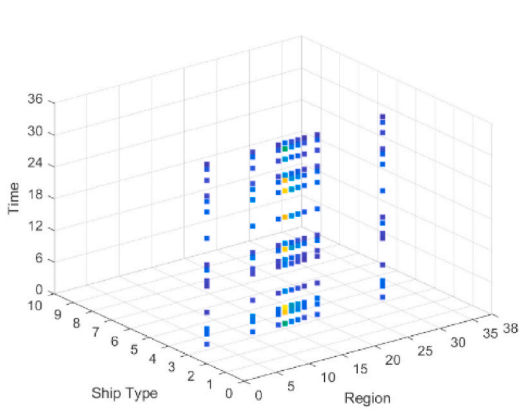
Then according to the traffic flow in the study waters, the Haihe river with a small amount of traffic flow in the upper left corner of Fig. 10 is divided into region 2. The traffic flow in the lower right corner of Fig. 9 is large, with a large number of ships coming in and out of Tianjin port. According to the trend of traffic flow, regions 11, 12, 22, and 34 are divided.

Finally, for the areas that have not yet been divided, the division is based on the opinion of the ship navigational officers. In the research waters of this study, this part is in the lower left part of Fig. 10, which has less traffic and fewer functional areas. It can also be seen that the grids in this part are more regular.

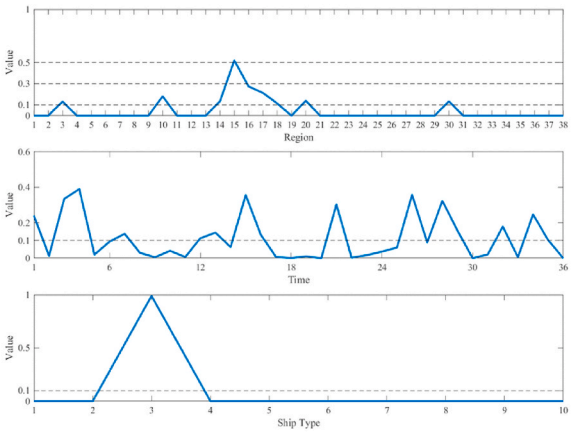
AIS data was used from 00:00:00 on January 1, 2019, to 23:59:59 on January 6, 2019. The AIS data are processed according to the method mentioned in Section 3.1. All ship types were selected for the study and the 10 ship types are numbered as shown in Table 1. In time, all AIS data are divided into 36 categories, as shown in Table 2. The AIS data before discretization and the ship attribute label data applied to tensor construction are shown in Table 3 and Table 4, respectively. Table 5 shows the number of ships and AIS data for each ship type after cleaning.

3.6. The selection of tensor rank

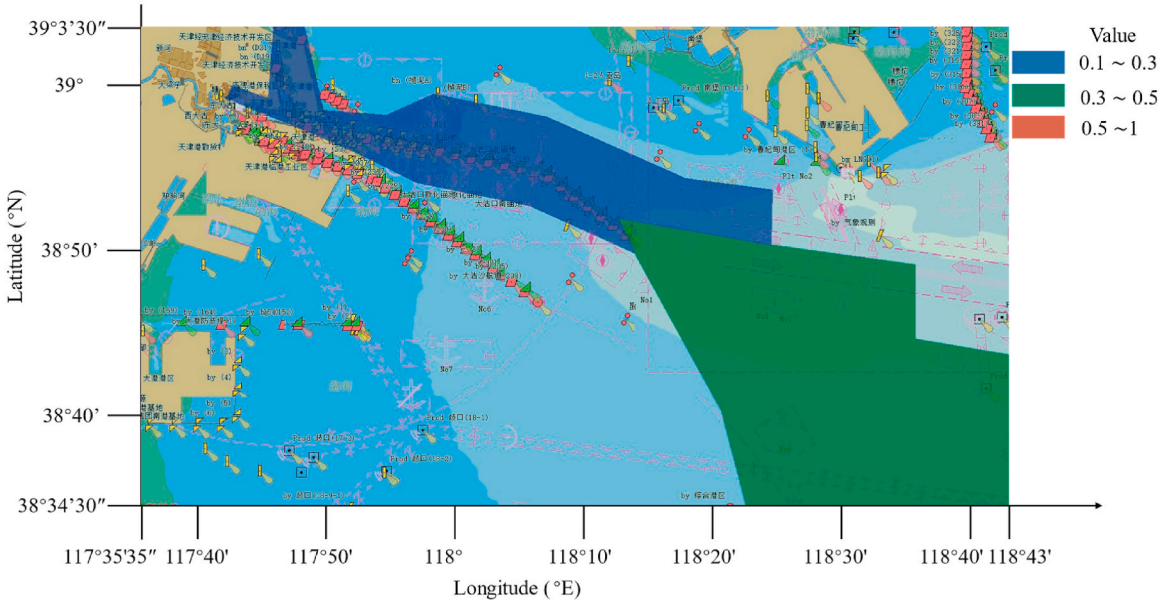
There is no exact way and only a limited boundary to select the value of the tensor factorization rank K. For tensor $X \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, tensor rank $K = \text{rank}(X) \leq \min\{I_1 I_2,\ I_1 I_3,\ I_2 I_3\}$. The factorization results vary depending on the value of K. The factorization result represents co-occurrence traffic patterns, which is called co-cluster in this study. To compare the differences in co-clusters under different values of K, this



(a) tensor of co-cluster 7



(b) factorization of co-cluster 7



(c) Spatial distribution of co-cluster 7

Fig. 20. Co-cluster 7 of <region, time, ship type>.

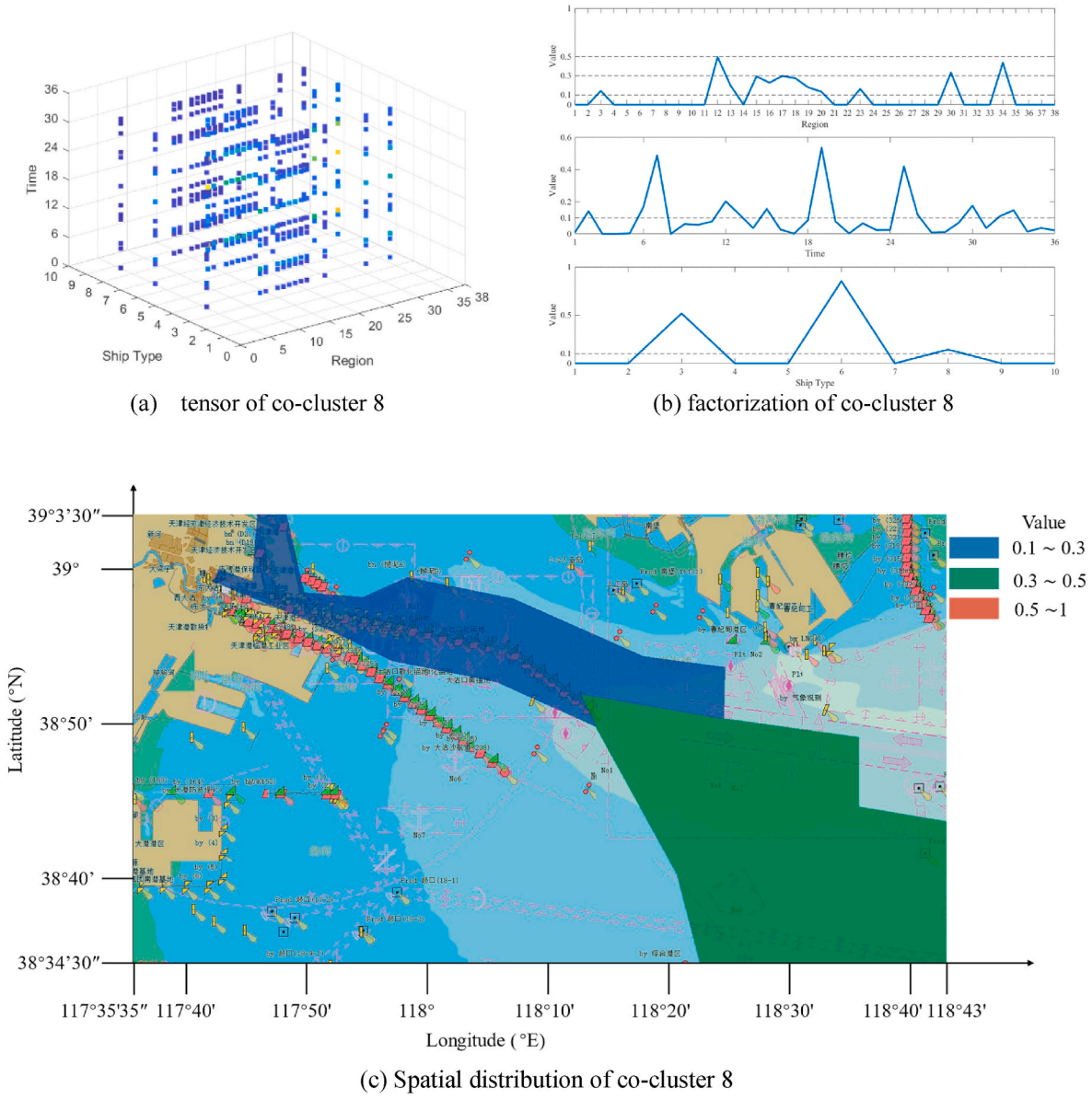
section makes a comparative experiment to illustrate the influence of K on the three-dimensional vessel traffic patterns. Tensor $X \in \mathbb{R}^{R\times T\times S}$ of mode < region, time, ship type > is selected for the comparative experiment. Tensor $X \in \mathbb{R}^{R\times T\times S}$ is decomposed non-negatively respectively, when the values of K are 6, 9, and 10. The co-clusters representing the same traffic pattern are selected for comparative analysis. Fig. 10 (a), (b), and (c) are the comparison of the co-clustering results showing a similar feature when K are 6, 9, and 10, respectively. The X-axis is the label of the attribute. The value of the Y-axis is the result of the l_2-norm normalized tensor factorization. It can also be interpreted as the degree of attribution to the cluster. It should be noted that the value of the Y-axis of each point represents the relative "degree of belonging" in the co-cluster. There is no quantity contrast relation and no addition relation among the values of the Y-axis.

From Fig. 12, it can be seen that under different K values, the co-clusters show similar trends for the same < region, time, ship type > pattern. The co-clusters have almost the same degree of belonging against each attribute. Therefore, the change in K does not affect the extraction of the main traffic patterns. The new patterns will appear as the increase of K. Data points with a small relative degree of belonging may form new co-clusters with some data points on the attribute in other patterns. Fig. 13 shows the phenomenon found in the experiment.

In the co-clustering results obtained with $K = 10$, a new pattern emerges. In Fig. 12 (b) and Fig. 12 (c), two patterns show the difference when ship type = 1. The difference is that the degree of belonging of the fishing boat in the traffic pattern shown in Fig. 12 (c) is 0.0141, which is close to 0, indicating that the distribution of fishing boats is no longer included in this pattern. At the same time, a new pattern that does not exist in the result of $K = 9$ is found and shown in Fig. 13. The fishing boat in this pattern has a degree of belonging of up to 0.9912, which is a vessel traffic pattern dominated by fishing boats. It is evident that most data points with ship type = 1 in the pattern as Fig. 12 (c) participate in the formation of a new pattern as Fig. 13.

The results can therefore be drawn as follows:

(1) The value of K does not affect the extraction of the main ship traffic patterns. There is barely difference in the three traffic patterns in Fig. 12 under different K values. As the value of K increases, more new ship traffic patterns may emerge. The emergence of new patterns may result in the decrease of the relative degree of belonging of some data points on the old



(a) tensor of co-cluster 8

(b) factorization of co-cluster 8

(c) Spatial distribution of co-cluster 8

Fig. 21. Co-cluster 8 of <region, time, ship type>.

pattern. It means that these points are more consistent with the traffic characteristics of the new pattern.

(2) Different K values determine different pattern extraction scales. With a higher tensor rank K, the knowledge contained in each traffic pattern is more individualized and more focused on the detailed features of the vessel traffic. On the contrary, the lower the tensor rank, the fewer number of the obtained traffic patterns is, and the information is more macroscopical. Increasing the tensor rank K for the same vessel traffic tensor may result in situations where only a few data points can form traffic patterns, leading to traffic patterns containing more microscopic knowledge. To address this issue, an appropriate tensor rank can be set, or high-dimensional data can be reduced to low-dimensional space using methods such as PCA or ICA for better feature and pattern extraction.

3.7. < region, time, ship type > traffic pattern tensor co-clustering experiment

The co-clustering result with $K = 9$ is obtained through repeated experiments. It can reflect the general ship traffic situation in Tianjin port waters. Refer to the information of navigational charts, the traffic information contained in the 9 patterns is analyzed. The result of co-clustering contains three attributes, including region, time, and ship types. The points with a degree of belonging being greater than 0.1 have the characteristic of the co-clusters. To make the results more intuitive, the values of the degree of belonging are set from less than 0.1 to 0 (see Fig. 21).

The co-clusters extracted from the < region, time, ship type > tensor show the spatio-temporal distribution of traffic activities of different ship types. However, the study of time has a certain complexity because of its continuity in sequence and discretization. Therefore, only part of the trend is observed here and there is no specific limitation on the degree of belonging of time. In summary, by analyzing the distribution of < region, time, ship type > of each co-cluster, the meaning hidden in each co-cluster can be concluded. The co-clusters extracted from < region, time, ship type > traffic tensor is shown in Figs. 14, 15, 16, 17, 18, 19, 20, 21 and 22. In each figure, (c) represents the characteristic regions distribution. The values on the region attribute are colored red (0.5 ~ 1), green (0.3 ~ 0.5), and blue (0.1 ~ 0.3).

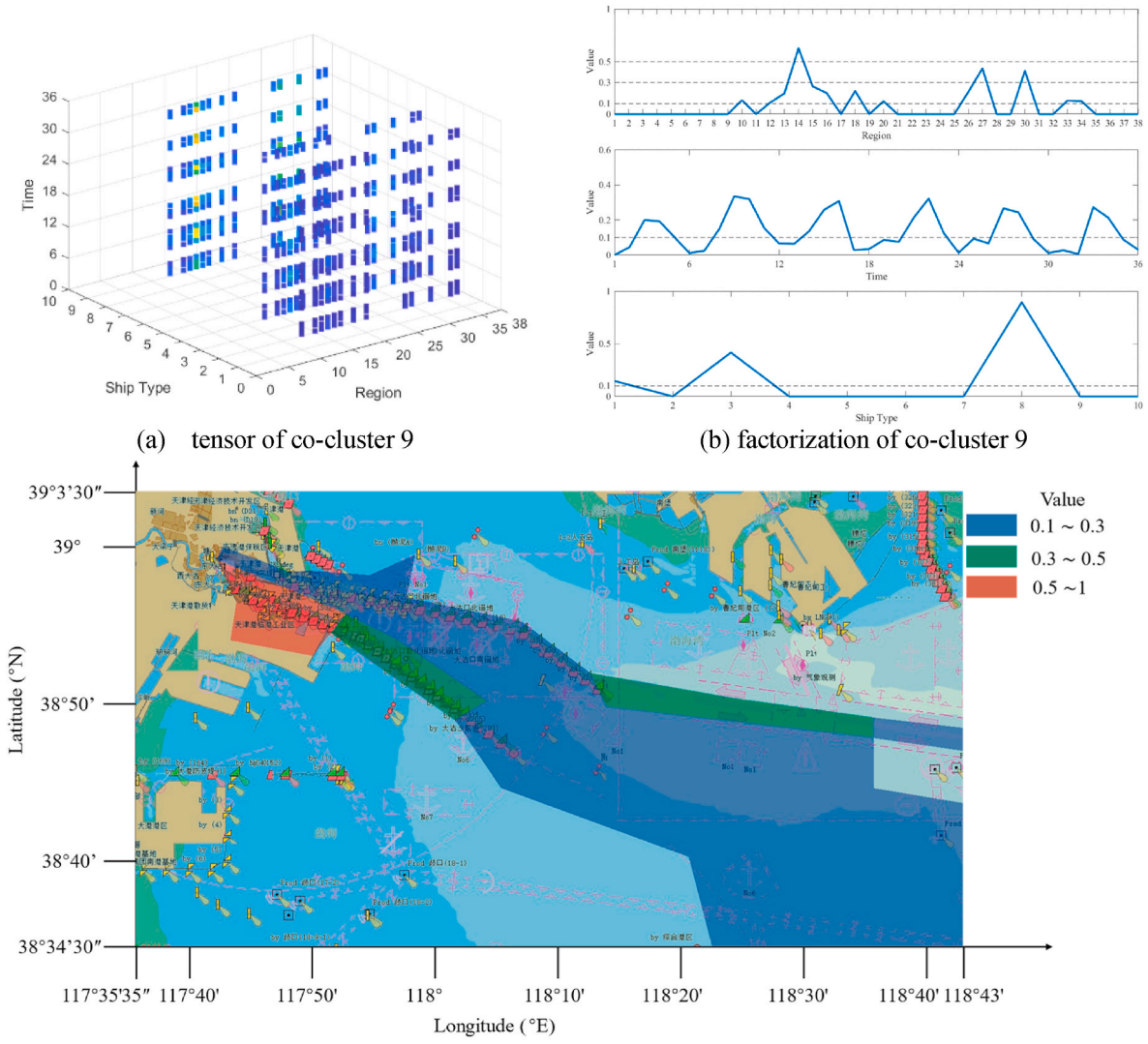


Fig. 22. Co-cluster 9 of <region, time, ship type>.

The characteristic regions of co-cluster 1 are 3, 13, 15, 16 and 22. These characteristic regions include the Nanjiang port area, Dongjiang port area, Dagukou port area, and the Haihe river. The largest number of ships are in Nanjiang port and Dongjiang port. This is a pattern that tugboats berth or work in the port area. Some cargo ships and oil tankers are also contained in the co-cluster.

In co-cluster 2, the characteristic regions are 9, 17, 19, 23, and 25. This is a pattern of cargo ships entering the Tianjin Port. Since the second day, almost every period, a large number of cargo ships entered the Tianjin Port through the exclusive waterway, passing through the Dagu anchorage.

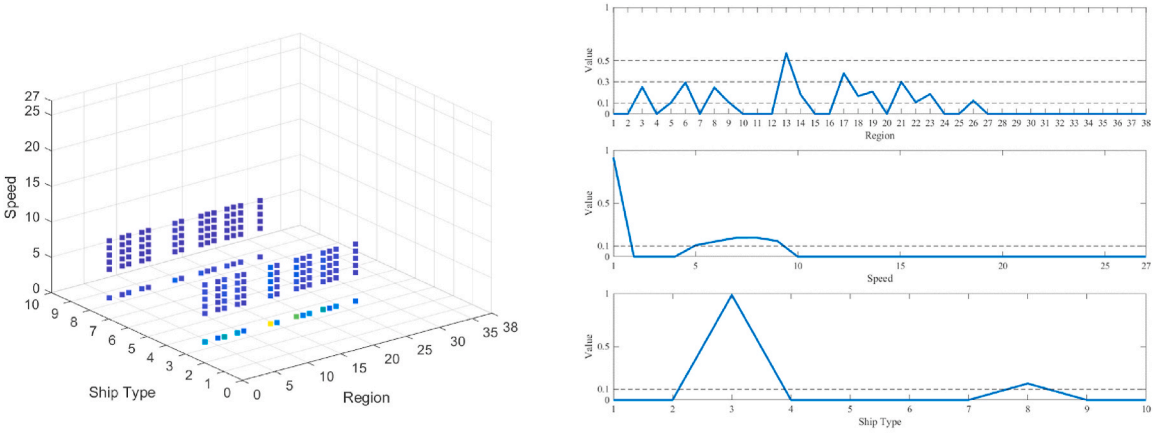
The characteristic regions in co-cluster 3 are 10, 15, 16, 18, 20, and 30. These regions are an exclusive route that leaves the Tianjin port, passing through the Dagukou south anchorage and Dagukou bulking anchorage. The Dagukou bulking anchorage is limited to cargo ships and oil tankers. Co-cluster 3 shows a pattern of cargo ships leaving the Tianjin Port. The cargo ships leave with the help of Tugboats.

In co-cluster 4, the characteristic regions are 15, 16, 17, 18, 19, 20, 23, 24, 25, and 30. The contained ship type is mainly cargo ships. The area with the highest density of ships is near the Dagukou north anchorage. The Dagukou north anchorage is limited to cargo ships. Co-cluster 4 represents a traffic pattern of cargo ships berthing at the Dagukou north anchorage. The Dagukou bulking anchorage and Nanjiang port area, also allow cargo ships to berth, so the two regions consisted in the pattern, too. The time characteristics are not obvious, but ships are berthing at all times except at night.

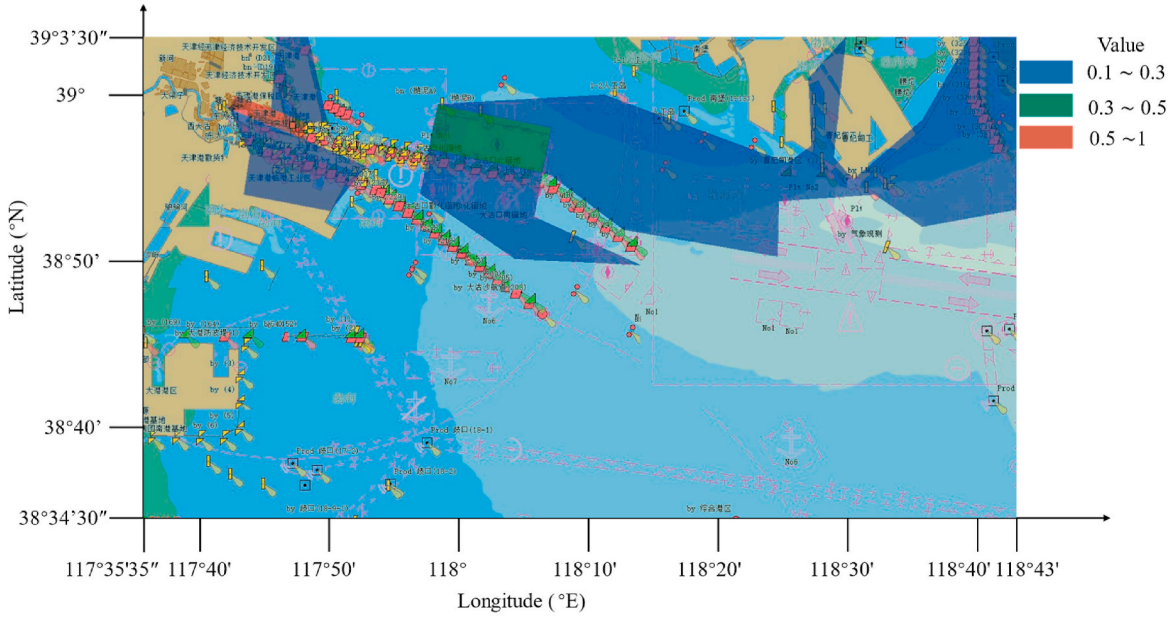
The characteristic regions in co-cluster 5 are 10, 12, 18, 20, 30, and 34. The information reflected in this co-cluster includes two routes of the cargo and oil tanker leaving the Tianjin port. One of the routes is the channel through the Caofeidian area, and the other is the channel at the southeast area of the Tianjin port. Ship activities are always frequent on January 4th and January 6th.

The co-cluster 6 includes fishing ships, cargo ships, and tugboats, and it can be observed that the labels of characteristic regions are 5, 6, 8, 10, 12, 21, 22, 24, 27, 29 and 31, respectively. There is little correlation among the characteristic regions except for region 8. To explain the reasons, the following analyses can be conducted based on the actual situations.

The answer can be obtained by correlating the regional and time attributes. On one hand, by analyzing the region attribute, it can be seen that region 8, has the highest degree of belonging. Region 8 includes the east anchorage of Caofeidian and channels in which ships enter and leave the Caofeidian third harbor (Caofeidian third harbor is not within the study area of this study). Meanwhile, the vessels waiting to enter the Caofeidian third harbor usually anchor in the northeast of the east anchorage outside the Caofeidian port. This co-cluster contains a large



(a) tensor of co-cluster 1 (b) factorization of co-cluster 1



(c) Spatial distribution of co-cluster 1

Fig. 23. Co-cluster 1 of <region, speed, ship type>.

number of fishing ships. The reason is that many fishing areas are distributed in Caofeidian third harbor. Region 27 is also an important fishing area in Tianjin Port. These fishing ships left a lot of trajectory information in region 8 when they enter and leave the harbor. These data are grouped into the same co-cluster according to the association of region 8. The co-cluster, which contains a variety of navigation conditions, results in the scattered regional distribution. On the other hand, the time distribution also enhances the relevance of these data. Although the time pattern is not prominent, it can still be seen that traffic volume is generally large from 0400 to 2000 every day.

The co-cluster 7 contains the characteristic regions 3, 10, 14, 15, 16, 17, 18, 20 and 30, respectively. It represents a pattern of cargo ships berthing in the Nanjiang port area and leaving the Tianjin port to the northeast in the morning.

The co-cluster 8 has the characteristic regions are 3, 12, 13, 15, 16, 17, 18, 19, 20, 23, 30 and 34, respectively. The cargo ships, tugboats, and a few oil tankers are included in this co-cluster. These characteristic regions indicate a route leaving the Tianjin port from a southeastern direction. Unlike co-cluster 5, the proportion of oil tankers is greater and the active time is mainly in the early morning. The regional distribution of co-cluster 8 also includes the pilotage area of Tianjin Port.

The co-cluster 9 includes fishing ships, cargo ships, and oil tankers. It can be observed that the relevant regions are 10, 12, 13, 14, 15, 16, 18, 20, 26, 27, 30, 33, and 34, respectively. The regional information reflected by this co-cluster is still the ship traffic on the route leaving the port. It is an oil tank's leaving route. In addition, some fishing boats are also included. Region 27 is the main fishing area in the Tianjin Port Area. A small amount of fishing ship data results in a less degree of belonging. The data have a strong characteristic association with temporal attributes. These regions are at the peak of traffic from 0800 to 1600 every day.

3.8. *< region, speed, ship type > traffic pattern tensor co-clustering experiment*

AIS contains data of multiple attributes, and in this study, the data tensors of < region, speed, ship type > are selected to carry on the co-cluster analysis with $K = 9$. The co-clustering results are illustrated by taking co-cluster 1–2 as an example. The co-cluster results are shown in Fig. 23 and Fig. 24. Each (c) represents the distribution of the

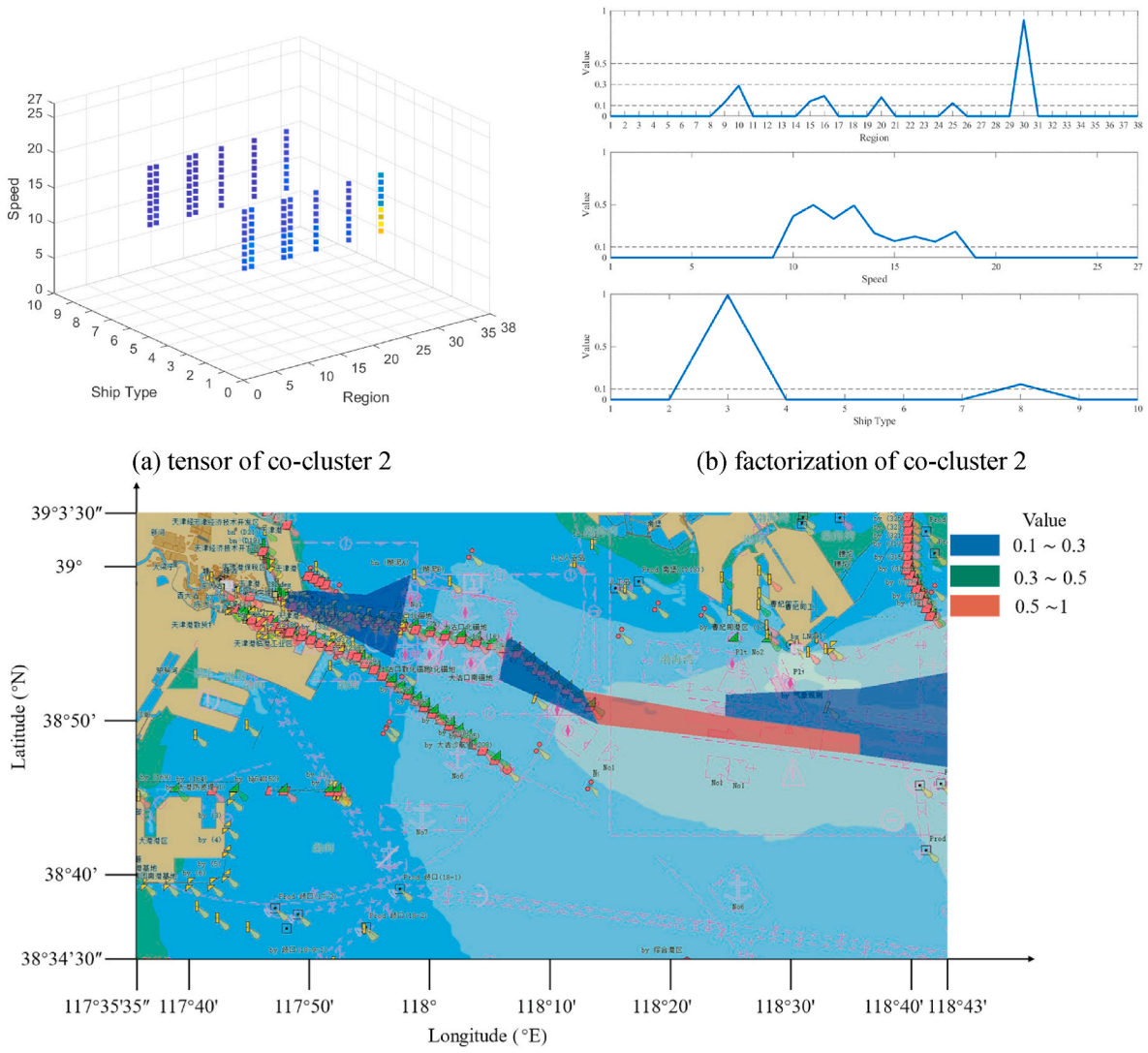


Fig. 24. Co-cluster 2 of < region, speed, ship type >

characteristic regions. The values on the region attribute are colored red (0.5 ~ 1), green (0.3 ~ 0.5), and blue (0.1 ~ 0.3).

The result of co-cluster 1 reveals that the characteristic regions are 3, 6, 8, 13, 14, 17, 18, 19, 21, 22, 23, 26, 33, and 34, respectively. The regions with the higher density of ships are the Nanjiang Port Area, the Beijiang Port Area, and the Dagukou North Anchorage. The speed is mainly distributed in 0–2 knots. Co-cluster shows a pattern that a big amount of cargo ships berths in port and anchorage. Except for regions 13 and 17, other regions are the anchorage and port areas.

The reasons why ships are moored in these locations and at speeds between 4 and 9 knots are twofold. Firstly, these regions include not only the anchorage but also the forbidden anchorage areas between anchorages. Secondly, the ships enter and leave these regions at a certain speed.

The characteristic regions of co-cluster 2 are 10, 16, 20, 25, and 30, respectively. Cargo ships and tankers in these regions have the speed distribution from 10 to 18 knots. These regions are the main channels entering and leaving the Tianjin port, where the speed is not allowed less than 5 knots.

4. Discussion

4.1. Validity discussion

The experimental results are consistent with the actual rules of Tianjin Port, indicating that the pattern mining method in this study is effective. The comparison with the existing research methods and results are as follows:

During the AIS data preprocessing stage, several steps are taken. Firstly, the AIS data is discretized into irregular grids based on the different functional and geographical attributes of actual sea areas. This method is superior over the traditional approach of regular grid processing (Lei et al., 2016; Wu et al., 2017b), as it enables clearer differentiation of the impact of different sea areas on vessel traffic. Secondly, this method considers the influence of time data and crew operation level on vessel traffic pattern extraction, in contrast to the vector-based clustering statistical method (Pallotta et al., 2013a). In the time discretization process, time intervals are divided according to the crew duty schedule to obtain more AIS data with similar characteristics. Lastly, ship type is an important attribute in vessel traffic research, as different types of vessels sail at different speeds and through different routes. Therefore, this method conducts clustering analysis on the data

based on different ship types, which is a novel approach in the study of vessel traffic patterns (Millefiori et al., 2017; Pallotta et al., 2013b; Xiao et al., 2020).

In the method of mining vessel traffic patterns, several steps are taken. Firstly, in contrast to the single grouping mode of the clustering feature extraction method (Liu et al., 2020; Vespe et al., 2012), the proposed method considers that AIS data points may contribute to multiple traffic patterns. Therefore, this method classifies the data points into multiple clusters with different degrees of belonging. The results in Section 4 show that ships in the same area are classified into different co-clusters. In actual sea areas, there may be ships with different motion patterns in the same area, which is consistent with reality. Secondly, unlike the statistical method (Gil et al., 2022), the original structure of AIS data is not altered in this method. The NTF method can reproduce the spatiotemporal patterns and dynamic process characteristics of the original data, which better describes the actual situation of vessel traffic. Moreover, the method has fewer parameters and is easy to adjust. Thirdly, to obtain better traffic pattern recognition results, this study establishes a three-dimensional traffic pattern based on the existing NTF method (Liu and Chen, 2014; Biondi, 2016, 2018, 2019). Experiments are conducted with different K values, and further analysis is performed. The range of K values needs to match the user's needs. Different needs of maritime management requirements, such as macro, meso, and micro, correspond to different K values. Experimental results indicate that the value of K should not be too large.

4.2. Limitations and uncertainties

Although showing some attractiveness, this study still reveals some limitations and uncertainties to be addressed in future studies.

(1) The balance between ship types should be considered. The number of cargo ships is large in study waters, resulting in a small degree of belonging of other ship types.
(2) In irregular grids division, there is no strict rule on the shape and size of the grids and the degree of utilization of prior knowledge. When the focus of the study is different, the division results of the same water area will be different. This leads to the uncertainty of the experimental results. There are also some limitations in using of prior knowledge.
(3) The parameter K also influences the experimental results. Different K values determine different pattern extraction scales, and a large value of K is not necessarily better. Increasing the value of K can result in fewer data points in each co-cluster, which cannot be accurately indicated in the NMF and NTF, a good choice of K value can also be determined by statistical measures, such as the cophenetic correlation coefficient or residual sum of squares (RSS) (Gao et al., 2019). In the RSS-based method, by plotting the variation of RSS with the variation of K value, the K value at the inflection point of the RSS curve is determined as the optimal one. We will also consider using other methods, to select appropriate K value in future studies.

5. Conclusion

In this study, a new three-dimensional traffic pattern extraction method is developed to incorporate the effect of the co-occurrence of multiple traffic attributes simultaneously. A real case study is carried out to analyze the AIS data from Tianjin Port-Caofeidian waters. More specifically, new findings of managerial insights from the analysis and case study are summarized as follows:

The tensor rank is set as 9 in this work when applying NTF for a traffic pattern study, the result shows that the vessels entering the Tianjin port are mainly through the exclusive channel in front of the Caofeidian and Dagukou north anchorage, and passing the Nanjiang port and Dongjiang port areas. Vessels leave the Tianjin port through two main routes: (a) from the Dagukou bulking anchorage towards the southeast; (b) from the Dagukou north anchorage and the exclusive waterway on the south side of Caofeidian, towards the east. Almost all of the ships on the routes have a speed of 10–18 knots. Furthermore, tugboats help cargo ships and oil tankers leaving the Tianjin port in the morning on route (b), and there are also a large number of tugboats working at the Nanjiang port and Dongjiang port throughout the day. Most oil tankers leave the Tianjin port through route (b). Cargo ships are the most active type of ships in these waters. Finally, in addition to the routes of entering and leaving the Tianjin port, there are a large number of cargo ships berthing at the Dagukou north anchorage all day, plan to enter the Tianjin port and Nanjiang port. Incorporation of the characteristics of the speed attribute reveals the new finding that the Nanjiang port has a significant amount of cargo ship berthing, which is overlooked in the existing literature, but has significant impact on port development and management.

CRediT authorship contribution statement

Jiaxuan Yang: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – review & editing. **Xingpei Bian:** Software, Investigation, Methodology, Data curation, Validation, Visualization. **Yuhao Qi:** Writing – editing, Visualization. **Xinjian Wang:** Writing – review & editing, Supervision, Writing – original draft. **Zaili Yang:** Writing-review. **Jiaguo Liu:** Formal analysis, Review.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Zaili Yang reports financial support was provided by European Research Council.

Data availability

Data will be made available on request.

Acknowledgement

This work is partially supported by a European Research Council project (TRUST CoG 2019 864724).

References

Animah, I., Shafiee, M., 2021. Maintenance strategy selection for critical shipboard machinery systems using a hybrid AHP-PROMETHEE and cost benefit analysis: a case study. Journal of Marine Engineering & Technology 20 (5), 312–323.
Biondi, F., 2016. Low Rank Plus Sparse Decomposition of Synthetic Aperture Radar Data for Maritime Surveillance. 2016 4th International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing. COSERA), Aachen, Germany.
Biondi, F., 2018. Low-rank plus sparse decomposition and localized radon transform for ship-wake detection in synthetic aperture radar images. Geosci. Rem. Sens. Lett. IEEE 15 (1), 117–121.
Biondi, F., 2019. A polarimetric extension of low-rank plus sparse decomposition and radon transform for ship wake detection in synthetic aperture radar images. Geosci. Rem. Sens. Lett. IEEE 16 (1), 75–79.
Cao, Y., Wang, X., Wang, Y., Fan, S., Wang, H., Yang, Z., Liu, Z., Wang, J., Shi, R., 2023a. Analysis of factors affecting the severity of marine accidents using a data-driven Bayesian network. Ocean. Eng. 269, 113563.
Cao, Y., Wang, X., Yang, Z., Wang, J., Wang, H., Liu, Z., 2023b. Research in marine accidents: a bibliometric analysis, systematic review and future directions. Ocean. Eng. 284, 115048.
Chen, P., Li, M., Mou, J., 2021. A velocity obstacle-based real-time regional ship collision risk analysis method. J. Mar. Sci. Eng. 9 (428), 428.
Christensen, M., Georgati, M., Arsanjani, J.J., 2022. A risk-based approach for determining the future potential of commercial shipping in the Arctic. Journal of Marine Engineering & Technology 21 (2), 82–99.
Cichocki, A., Zdunek, R., Phan, A.H., Amari, S., 2009. Nonnegative Matrix and Tensor Factorizations -Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation. John Wiley & Sons, London.

Fang, S., Liu, Z., Yang, X., Wang, X., Wang, J., Yang, Z., 2023. A quantitative study of the factors influencing human evacuation from ships. Ocean. Eng. 285, 115156.
Filipiak, D., Węcel, K., Stróżyna, M., Michalak, M., Abramowicz, W., 2020. Extracting maritime traffic networks from ais data using evolutionary algorithm. Business & Information Systems Engineering 62 (5), 435.
Gil, M., Kozioł, P., Wróbel, K., Montewka, J., 2022. Know your safety indicator - a determination of merchant vessels Bow crossing range based on big data analytics. Reliab. Eng. Syst. Saf. 220, 108311.
Han, P., Yang, X., 2020. Big data-driven automatic generation of ship route planning incomplex maritime environments. Acta Oceanologica Sinica -English Edition 8 (39), 113–120.
Huang, D., Liang, T., Hu, S., Loughney, S., Wang, J., 2023. Characteristics analysis of intercontinental sea accidents using weighted association rule mining: evidence from the Mediterranean Sea and Black Sea. Ocean. Eng. 287, 115839.
Jiang, Y., Ruan, Q., An, G., Jin, Y., 2018. Fast Non-negative Tensor Factorization Based on Cp Decomposition for 3D Facial Expression Recognition.
Kabir, M., Kang, M.J., Wu, X., Hamidi, M., 2022. Study on U-turn behavior of vessels in narrow waterways based on AIS data. Ocean. Eng. 246, 110608.
Kolda, T.G., Bader, B.W., 2009. Tensor decompositions and applications. SIAM Rev. 3 (51), 455–500.
Lee, D.D., Seung, H.S., 2000. Algorithms for non-negative matrix factorization. In: International Conference on Neural Information Processing Systems.
Lei, P., Tsai, T., Peng, W., 2016. Discovering Maritime Traffic Route from AIS Network. 2016 18Th Asia-Pacific Network Operations and Management Symposium (APNOMS), pp. 1–6.
Li, Y., Liu, R.W., Liu, J., Huang, Y., Hu, B., Wang, K., 2016. Trajectory compression-guided visualization of spatio-temporal AIS vessel density. In: 2016 8Th International Conference on Wireless Communications & Signal Processing (WCSP).
Liu, C., Chen, X., 2014. Vessel track recovery with incomplete AIS data using tensor candecom/parafac decomposition. J. Navig. 67 (1), 83–99.
Liu, C., Liu, J., Zhou, X., Zhao, Z., Wan, C., Liu, Z., 2020. AIS data-driven approach to estimate navigable capacity of busy waterways focusing on ships entering and leaving port. Ocean. Eng. 218, 108215.
Liu, J., Liu, Y., Qi, L., 2021. Modelling liquefied natural gas ship traffic in port based on cellular automaton and multi-agent system. J. Navig. 74 (3), 533–548.
Liu, Z., Gao, H., Zhang, M., Yan, R., Liu, J., 2023. A data mining method to extract traffic network for maritime transport management. Ocean Coast Manag. 239, 106622.
Millefiori, L.M., Braca, P., Arcieri, G., 2017. Scalable distributed change detection and its application to maritime traffic. In: 2017 IEEE International Conference on Big Data. BIGDATA), Boston, MA, USA.
Millefiori, L.M., Zissis, D., Cazzanti, L., Arcieri, G., 2016. A distributed approach to estimating sea port operational regions from lots of AIS data. In: 2016 IEEE International Conference on Big Data (Big Data).
Montewka, J., Manderbacka, T., Ruponen, P., Tompuri, M., Gil, M., Hirdaris, S., 2022. Accident susceptibility index for a passenger ship-a framework and case study. Reliab. Eng. Syst. Saf. 218, 108145.
Osekowska, E., Johnson, H., Carlsson, B., 2014. Grid size optimization for potential field based maritime anomaly detection. Transport. Res. Procedia 3, 720–729.
Pallotta, G., Vespe, M., Bryan, K., 2013a. Traffic knowledge discovery from AIS data. In: Proceedings of the 16th International Conference on Information Fusion, Istanbul, Turkey.
Pallotta, G., Vespe, M., Bryan, K., 2013b. Vessel pattern knowledge discovery from AIS data: a framework for anomaly detection and route prediction. Entropy 15 (12), 2218–2245.
Qi, L., Ji, Y., Balling, R., Xu, W., 2021. A cellular automaton-based model of ship traffic flow in busy waterways. J. Navig. 74 (3), 605–618.
Ristic, B., 2014a. Detecting anomalies from a multitarget tracking output. IEEE Trans. Aero. Electron. Syst. 50 (1), 798–803.
Ristic, B., 2014b. Detecting anomalies from a multitarget tracking output. IEEE Trans. Aero. Electron. Syst. 50 (1), 798–803.
Ristic, B., La Scala, B., Morelande, M., Gordon, N., 2008. Statistical analysis of motion patterns in ais data: anomaly detection and motion prediction. In: 2008 11th International Conference on Information Fusion, Cologne, Germany.
Rong, H., Teixeira, A.P., Guedes Soares, C., 2022. Maritime traffic probabilistic prediction based on ship motion pattern extraction. Reliab. Eng. Syst. Saf. 217, 108061.
Silveira, P.A.M., Teixeira, A.P., Soares, C.G., 2013. Use of ais data to characterise marine traffic patterns and ship collision risk off the coast of Portugal. J. Navig. 66 (6), 879–898.
Tang, H., Yin, Y., Shen, H., 2022. A model for vessel trajectory prediction based on long short-term memory neural network. Journal of Marine Engineering & Technology 21 (3), 136–145.
Valcalda, A., de Koningh, D., Kana, A., 2023. A method to assess the impact of safe return to port regulatory framework on passenger ships concept design. Journal of Marine Engineering & Technology 22, 1–12.
Vespe, M., Visentini, I., Bryan, K., Braca, P., 2012. Unsupervised learning of maritime traffic patterns for anomaly detection. In: Data Fusion & Target Tracking Conference, Stevenage, UK.
Wang, H., Liu, Z., Wang, X., Graham, T., Wang, J., 2021. An analysis of factors affecting the severity of marine accidents. Reliab. Eng. Syst. Saf. 210, 107513.
Wang, J., Zhu, C., Zhou, Y., Zhang, W., 2017. Vessel spatio-temporal knowledge discovery with ais trajectories using Co-clustering. J. Navig. 70 (6), 1383–1400.
Wang, X., Xia, G., Zhao, J., Wang, J., Yang, Z., Loughney, S., Fang, S., Zhang, S., Xing, Y., Liu, Z., 2023. A novel method for the risk assessment of human evacuation from cruise ships in maritime transportation. Reliab. Eng. Syst. Saf. 230, 108887.
Wei, Z., Xie, X., Lv, W., 2020. Self-adaption vessel traffic behaviour recognition algorithm based on multi-attribute trajectory characteristics. Ocean. Eng. 198, 106995.
Wen, R., Yan, W., Zhang, A.N., Chinh, N.Q., Akcan, O., 2016. Spatio-temporal route mining and visualization for busy waterways. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC).
Wu, L., Xu, Y., Wang, Q., Wang, F., Xu, Z., 2017b. Mapping global shipping density from ais data. J. Navig. 70 (1), 67–81.
Wu, X., Rahman, A., Zaloom, V.A., Craig, B.N., 2017a. Study of Vessel Travel Behavior at Hot Spots in Sabine-Neches Waterway. Transportation Research Board 96th Annual Meeting.
Wu, X., Rahman, A., Zaloom, V.A., 2018. Study of travel behavior of vessels in narrow waterways using AIS data – a case study in Sabine-Neches Waterways. Ocean. Eng. 147, 399–413.
Xia, H., 2021. Navigational risk analysis based on gis spatiotemporal trajectory mining: a case study in nanjing yangtze river bridge waters. Arabian J. Geosci. 14 (3), 229.
Xiao, Z., Fu, X., Zhang, L., Goh, R.S.M., 2020. Traffic pattern mining and forecasting technologies in maritime traffic service networks: a comprehensive survey. IEEE Trans. Intell. Transport. Syst. 21 (5), 1796–1825.
Xiao, Z., Ponnambalam, L., Fu, X., Zhang, W., 2017. Maritime traffic probabilistic forecasting based on vessels' waterway patterns and motion behaviors. IEEE Trans. Intell. Transport. Syst. 18 (11), 3122–3134.
Yang, J., Ma, L., Liu, J., 2021. Modeling and application of ship density based on ship scale conversion and grid. Ocean. Eng. 237, 109557.
Yu, Z., Yuan, L., Lv, G., Luo, W., Xie, Z., 2011. Coupling characteristics of zonal and meridional sea level change revealed by satellite altimetry data and their response to enso events. Chin. J. Geophys. 54 (8), 1972–1982.
Zhang, L., Meng, Q., Fang Fwa, T., 2017. Big ais data based spatial-temporal analyses of ship traffic in Singapore port waters. Transport. Res. E Logist. Transport. Rev. 129, 287–304.
Zhang, M., Conti, F., Le Sourne, H., Vassalos, D., Kujala, P., Lindroth, D., Hirdaris, S., 2021a. A method for the direct assessment of ship collision damage and flooding risk in real conditions. Ocean. Eng. 237, 109605.
Zhang, M., Montewka, J., Manderbacka, T., Kujala, P., Hirdaris, S., 2021b. A big data analytics method for the evaluation of ship - ship collision risk reflecting hydrometeorological conditions. Reliab. Eng. Syst. Saf. 213, 107674.
Zhang, M., Taimuri, G., Zhang, J., Hirdaris, S., 2023. A deep learning method for the prediction of 6-dof ship motions in real conditions. Proceedings of the Institution of Mechanical Engineers. Part M, Journal of Engineering for the Maritime Environment, 1629469542.
Zhang, M., Zhang, D., Fu, S., Kujala, P., Hirdaris, S., 2022. A predictive analytics method for maritime traffic flow complexity estimation in inland waterways. Reliab. Eng. Syst. Saf. 220, 108317.
Zhen, R., Jin, Y., Hu, Q., Shao, Z., Nikitakos, N., 2017. Maritime anomaly detection within coastal waters based on vessel trajectory clustering and naïve bayes classifier. J. Navig. 70 (3), 648–670.