IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 23, 2017, accepted April 14, 2017, date of publication April 27, 2017, date of current version June 7, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2698208

Finding Abnormal Vessel Trajectories Using Feature Learning

PEIGUO FU[1], HAOZHOU WANG[2], KUIEN LIU[1], XIAOHUI HU[1], AND HUI ZHANG[1]
[1]State Key Laboratory of Integrated Information System Technology, Institute of Software, Chinese Academy of Sciences, Beijing, China, 100190
[2]Pivotal Software Inc.

Corresponding author: Peiguo Fu (peiguo12@iscas.ac.cn)

This work was supported by the National Natural Science Foundation of China under Grant U1435220 and Grant 61503365.

ABSTRACT Global Positioning System technology has been widely used in vehicle tracking and road planning applications. An enormous amount of data concerning the trajectories of vehicles has been collected and stored for tracking purposes. A trajectory contains not only the footprints of a moving object but also additional information, such as speed and stopping points. Therefore, the large-scale trajectory data sets provide rich information and are currently attracting considerable attention; there have been many successful studies of event detection based on trajectory data. However, most of these studies have focused only on vehicles traveling in a road network and have note considered maritime trajectories. A maritime trajectory also contains auxiliary data (e.g., speed and rotation) in addition to the movements of a ship. However, ships are not bound to road networks, and consequently, it is difficult to apply traditional mining algorithms based on road networks. In addition, even if the amount of maritime trajectory data is very large, these data are also spatially sparse, which will significantly reduce the effectiveness of most existing mining algorithms. In this paper, we propose a new method of abnormal trajectory detection to address this problem. This method can detect abnormal vessel trajectories from Automatic Identification System (AIS), records for vessels via our feature learning algorithm. To reduce the search space, we invoke reference points as well as the Piecewise Linear Segmentation (PLS), algorithm to compress the trajectories without losing important information. A time-aware and spatially correlated collaborative algorithm is proposed to increase the density of the trajectories to improve the accuracy of the detection algorithm, which is based on Dynamic Time Warping (DTW). Finally, we report experiments conducted on a real-world data set, which demonstrate that the proposed detection method can detect anomalous trajectories effectively.

INDEX TERMS Feature extraction, trajectory compression, trajectory partitioning, similarity calculation.

I. INTRODUCTION

With the development of position sensing techniques and location-aware devices, the Global Positioning System (GPS) has now become widely used in many areas of industry, such as vehicle tracking and freight transport. These new technologies include the development of chips, wireless communication standards and so on that allow GPS devices to report their locations in real time. The trajectories that are collected by such devices contain important spatio-temporal information and play a significant role in these areas since trajectory data can be used to manage vehicles and predict their behavior. Moreover, these data can also be used for event detection to detect unusual events or accidents. Currently, not only industry agents are greatly interested in analyzing trajectory data; academic researchers are also focusing on designing effective trajectory indexing structures [1], [2] and methods of trajectory query processing [3], [4], trajectory uncertainty management [5], [6] and mining knowledge/patterns from trajectory data [7], [8].

Maritime trajectory data, another traditional type of trajectory data for moving objects, are also attracting increasing attention in both academia and industry. A maritime trajectory includes motion data as well as auxiliary records for a ship, and such trajectories have the potential to be used for anomalous event detection. This would require a system with the ability to mine patterns from such trajectory data. However, unlike vehicle-based trajectories, which are strictly bound to road networks, maritime trajectories are more likely to represent movement in free space, since vessels may be bound only to approximate routes and each ship must remain a long distance away from others for safety reasons. Accordingly, even if two maritime trajectories are

following the same pattern, the similarity between the two trajectories may be very low. This situation leads to the difficulty that traditional trajectory-based pattern mining algorithms are very difficult to apply directly to maritime trajectories because most trajectory-based pattern mining algorithms are based on similarity measures. To calculate the similarity between two trajectories, a distance measure such as the Euclidean distance, dynamic time warping (DTW) [9], longest common subsequence (LCSS) [10] or edit distance on real sequence (EDR) [3] is used. However, these distance measures are not applicable to maritime trajectory datasets because they are designed only to detect the ''shape'' similarity between given trajectories.

The challenge described above is not the only one encountered in this context. Maritime trajectories collected by GPS devices are also subject to uncertainty. Hence, another challenge is that for maritime trajectories, no additional knowledge, such as road network structures or points of interest (POIs), is available to help calibrate the trajectories, which reduces the accuracy of anomalous event detection. We can see from the sampled points of maritime trajectories shown in Fig. 1 that the distribution of maritime trajectories is highly sparse in an infinite free space, which makes pattern mining more difficult. Moreover, because of the slow motion of vessels and the high sampling rate, maritime trajectory datasets contain much redundant information, which not only reduces the processing speed but also wastes storage space and interferes with pattern mining algorithms.

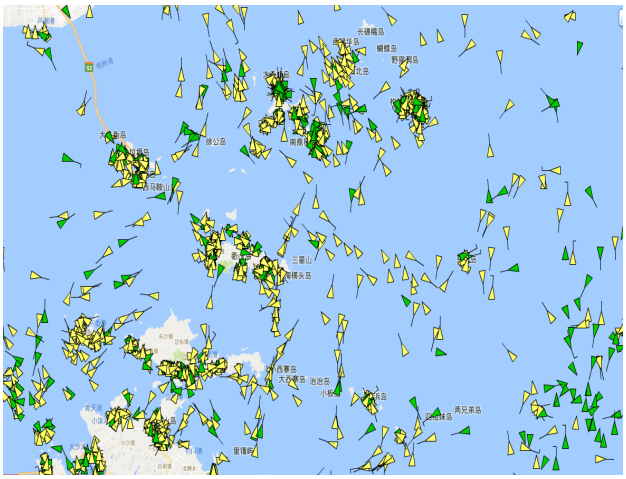


FIGURE 1. Raw trajectory data of vessels in a marine area.

To address these problems, we propose an abnormal trajectory detection method that leverages the features of raw trajectory data and extracts useful feature information to speed up the trajectory pattern mining algorithm. A spatio-temporal collaborative algorithm is used to fill in missing sample data in an entire dataset. Meanwhile, a compression method based on an entire trajectory is used to reduce data redundancy. To perform trajectory pattern mining, we propose a semantic trajectory similarity search method that is combined with the DBSCAN algorithm to extract the required features from the raw trajectory dataset.

In summary, the contributions of our work are three-fold.

1) We analyze the differences between vehicle trajectories and vessel trajectories and the challenges of mining trajectory patterns from a maritime trajectory dataset.

2) We propose a semantic trajectory similarity search method. We report a series of experiments conducted on a large-scale real-world trajectory dataset, the results of which demonstrate the effectiveness of our method of anomaly detection in maritime trajectories.

3) We design an abnormal trajectory detection system to report abnormal events that can be used in vessel monitoring.

The remainder of this paper is organized as follows. In Section 2, we present the notation, preliminary concepts and definitions used in this paper. Then, we present an overview of the proposed framework and introduce our algorithms in Section 3. Section 4 discusses experimental results obtained based on both real-world and synthetic maritime trajectory datasets. A literature review is presented in Section 5. Finally, we draw conclusions and propose future work in Section 6.

II. PROBLEM STATEMENT

In this section, we first define the problem in terms of formal concepts and definitions. The terminology and notation used in this paper are defined in Table 1. For a moving (marine) object tracked by a monitoring system such as AIS, its geographical location is sensed and recorded periodically. Such a sequence of spatial-temporal data points is called a trajectory.

TABLE 1. Summary of notation.

Notation	Definition
$\overline{T}$	A raw trajectory
T	A processed trajectory
p	A sampled point of a trajectory
$p.o$	The ID of p
$p.l$	The location of p, in the form $< latitude, longitude >$
$p.t$	The timestamp of p
$p.s$	The speed of p
$p.c$	The course of p
rp	A reference point
$\overline{RT}$	A reference trajectory
S	A set of reference trajectories
PC	A special feature of a moving object
E	A set of entry areas
$\overline{ST}$	A series of state transitions
λ_i, ω_i	Weight factors
α, β, γ	Function control factors
θ	An anomaly score threshold
$d(T_i, T_j)$	The distance between trajectories T_i and T_j

A. PRELIMINARY CONCEPTS

Definition 1 (Original Route): An original route of a moving object is a continuous mapping from the time domain into spatial coordinates (i.e., longitude and latitude), indicating the continuous path along which the object moves in practice.

No original route exists in a real database since few industrial positioning techniques can record continuous movements in a cyber space. Instead, only a set of samples

from the original route can be obtained and stored in the database.

Definition 2 (Raw Trajectory): A raw trajectory $\overline{T}$ is a finite sequence of locations sampled from the original route of a moving object in geographical space and denoted by a set of time-ordered points, i.e., $\overline{T} = \{p_1, p_2, \ldots, p_n\}$, where each point consists of a geospatial coordinate and a timestamp and other features, i.e., $p_i = (o, l, t, s, c), i \in \{1, 2, \ldots, n\}$.

Here, o is a moving object denoted by an unique ID; l is the location of p_i, denoted by $< latitude, longitude >$; t is the timestamp of p_i; s is the current speed at p_i; and c is the course at p_i. A raw trajectory is represented as a continuous sequence that contains spatial-temporal and other auxiliary information about the corresponding moving object. It is difficult to find a common path from a group of raw trajectories based on discrete samples; therefore, in this paper, the dataset is pre-processed to map each raw trajectory onto a road network to obtain a continuous trajectory. Throughout the remainder of the paper, we use ''trajectory'' and ''raw trajectory'' interchangeably.

Definition 3 (Reference Point): A reference point rp is a fixed location in the space that is independent of any trajectory data source.

A reference point may refer to either a geographical object (e.g., a PoI) or a landmark. Any entity in space can serve as a reference point as long as it satisfies the definition. However, in marine scenarios, such reference points (e.g., POIs in an urban area) are usually not available. In this paper, ports, turning points, centroids of clusters and other special points are all considered as reference points.

Definition 4 (Phase Change): A phase change, denoted by PC, is a special feature of object movement. A PC refers to the status of a moving object that is changing behavior states within a certain period of time and in a certain area. A PC is characterized by three values, id, t and a, where id is the ID of the moving object, t is the time span of the state change, and a is the corresponding area.

A change in speed or direction is a phase change of a moving object. For example, the typical states for a ship includes sailing, entering port, docking, and departure.

Definition 5 (Entry Area): An entry area is a fixed spatial area in marine space, in which a moving object undergoes a series of state transitions denoted by entering or leaving. In particular, it includes the area in which the moving object is in the berthing state (speed = 0).

The set of entry areas is denoted by E. The state of a moving object between two different entry areas e_i and e_{i+1} may change more than once, which means that there can a series of state transitions $\overline{ST} = \{st_1, st_2, \ldots, st_m\}$ can occur as a moving object travels from e_i to e_{i+1}. We can also treat each entry area as a special reference point.

Definition 6 (Reference Trajectory): A reference trajectory $\overline{RT}$ is a sequence of reference points and their corresponding timestamps, i.e., $\overline{RT} = \{rp_1, rp_2, \ldots, rp_{|T|}\}$ and $rp_i = (rp_i.o, rp_i.l, rp_i.t, rp_i.s, rp_i.c)$.

A reference trajectory can comprise reference points of different types. For example, a reference trajectory for a shipping line may consist of a set of ports; thus, we can rewrite a raw trajectory T into a trajectory $\overline{RT}$ based on phase changes. We call such a trajectory a reference trajectory. The size $|\overline{RT}|$ denotes the number of phase-change locations along $\overline{RT}$. For simplicity, in the following, we focus on reference trajectories whose reference points are of the same type. However, our technique can easily generalize to reference trajectories with different reference points.

Given the definitions of a trajectory and a reference trajectory, a trajectory distance function is proposed to measure the difference between a trajectory and a reference trajectory.

Definition 7 (Trajectory Distance Function): A trajectory distance function d computes a difference score between a trajectory T and a reference trajectory $\overline{RT}$.

Based on the above definitions, we present a formal definition of abnormal trajectory discovery.

Definition 8 (Abnormal Trajectory Discovery): Given a trajectory T, a reference trajectory set $S = \{\overline{RT_1}, \overline{RT_2}, ..., \overline{RT_k}\}$, a trajectory distance function d and an anomaly score threshold θ, we can calculate the anomaly score of T as follows:

$$Sim_T = \sum_{i=0}^{k} \omega_{\overline{RT_i}} * d(T, \overline{RT_i}) \quad (1)$$

where $\omega_{\overline{RT_i}}$ is the weight of $\overline{RT_i}$ in the reference trajectory set S. If $Sim_T > \theta$, then we call Sim_T a θ-outlier on S and d.

B. FEATURE EXTRACTION

In this subsection, we present some main features that will be used to describe trajectories. These features can be classified into two main types: routing features (which describe where a moving object travels) and movement features (which describe how it travels).

1) ROUTING FEATURES

Routing features describe where a moving object travels. Since we are focused on the trajectories of vessels, the natural routing features are those that provide information about the routes on which they travel. Route information directly affects the movement patterns of trajectories. In this paper, we define two kinds of route information ('type' and 'direction') as routing features, as shown in Table 2. These features can be extracted from the AIS data to distinguish different kinds of routes.

Type: The type of a vessel is an important feature that affects how its route changes. Different kinds of vessels follow different routes as they travel on the ocean.

Direction: The direction feature indicates the traffic direction of a route. We define two direction values, i.e., 1 (from start to destination) and -1 (from destination to start).

2) MOVEMENT FEATURES

Moving objects are characterized by different movement features that describe how the objects travel. Many articles

TABLE 2. Summary of notation.

Feature Type	Example	Numeric
type	cargo	No
direction	1(-1)	No

TABLE 3. Summary of notation.

Feature Type	Example	Numeric
speed	100	Yes
course	100	Yes
number of stop points	5	Yes
number of phase changes	10	Yes

have investigated the extraction of various types of movement information from trajectories. In this paper, we define four types of movement features, i.e., speed, course, number of stop points and number of phase changes in Table 3, to describe the motion behaviors of a moving object.

Speed: The speed of a moving object is one of its most important movement features. For instance, if the speed of a vessel is higher or lower than the average speed on the same trajectory, then we may distinguish the detected trajectory from others.

Course: The course of a moving object is another important movement feature of a trajectory. Similar to speed, if the course of a vessel is different from the historical scope of a trajectory, it may raise concerns about the personal liberty of its helmsman.

Number of stop points: Stop points are places where a moving object remains for a certain period of time at a speed close to 0. In this paper, stop points are different from start and destination points; they refer only to points at which the vessel halts along the trajectory between the start and destination. A stop point is an instantaneous phase change of the moving object, which can be treated as a potential anomaly. There are many causes of stop points.

Number of phase changes: A phase change is an important indicator of the movement status. If a trajectory changes phases too frequently, it may be abnormal compared with other trajectories.

III. LEARNING FEATURES FOR FINDING ABNORMAL VESSEL TRAJECTORIES

In this section, we present a detailed description of the proposed algorithm, which consists of three parts, i.e., data processing, similarity calculation and anomaly trajectory detection.

A. FRAMEWORK OVERVIEW

We present the structure of the proposed anomaly detection system in this section. Given a dataset of trajectories, we use this framework to pre-process the raw trajectory records, including compression of the raw trajectories, prediction of the missing data, and simultaneous construction of

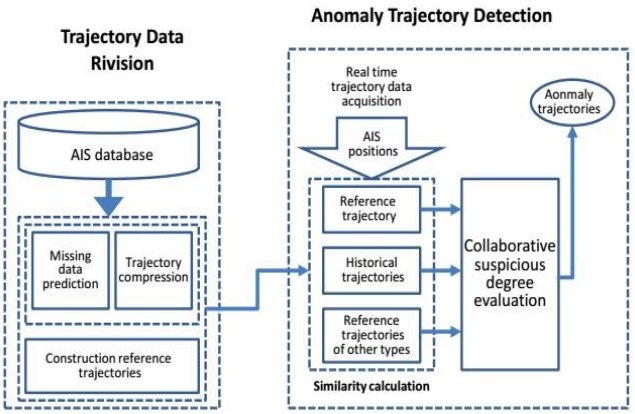


FIGURE 2. Overview of the framework.

a reference system. After the data are processed, we utilize the reference trajectories and the *DTW* algorithm to calculate the similarity between trajectories. To handle the novelty of newly arrived data, a time-aware approach that balances the updating of new trajectory data and the reference system is proposed. Fig 2 shows the details of the proposed framework. As can be seen from Fig 2, the framework consists of two modules: the trajectory data revision module and the anomalous trajectory detection module.

B. CONSTRUCTION OF THE REFERENCE SYSTEM

In this section, we discuss the use of reference points to construct a reference system. We cluster the raw trajectory data using the *DBSCAN* algorithm and use the geometric centers of the clusters as reference points. We use this method to build the reference system for two main reasons. First, although deviation is inevitable, most data records will lie in the vicinity of the route. Because of susceptibility to weather, water conditions, and other factors, a ship will inevitably deviate from its route; however, for historical records from a large number of ships, the data will be accurate. Second, the proportion of abnormal trajectories will be much smaller than that of normal trajectories. Among a large number of data records from AIS systems, as stated above, although data errors will inevitably exist, most data will represent normal trajectory records because in practice, even if one percent anomalous behavior is unacceptable, among the large-scale and large-volume daily data in the AIS dataset, the proportion of abnormal trajectories will be much smaller than that of normal trajectories. This is the basis of our reference system.

C. TRAJECTORY DATA REVISION

AIS data may be lost as a result of factors related to weather, hydrology, and transmission. Such data loss affects the accurate recording of a trajectory. In such a case, we will need to predict and analyze these missing but important data. In addition, since the navigation states of each vessel have been disseminated by the AIS system in real time, a large amount of data has been recorded and stored. The massive scale

of AIS data poses challenges of storage and computational capacity. Considering that a ship during actual navigation will show a certain regularity, we can compress the original AIS data to reduce the data size.

1) MISSING DATA PREDICTION

We divide the trajectories of ships of the same type that have the same origin and destination into a set, and we then analyze the trajectories in that set. However, because of susceptibility to weather, hydrology and other factors, signal transmission may be affected. The quality of ship trajectory records is not as high as that of the comprehensive data collected on land, and problems of missing traces may exist. To improve the accuracy of the similarity calculation, we need to fill in these data vacancies.

Suppose that there exists a long time interval or a long spatial distance between two adjacent points in a trajectory. To improve the prediction accuracy, we make three intuitive and reasonable assumptions:

(1) Ships of the same type will exhibit similar trajectory characteristics and navigational behaviors in neighboring areas.

In practice, as ships of the same type are navigating on adjacent routes, they may have to abide by certain navigation laws. Although they will inevitably be affected by weather as well as hydrological and other meteorological factors, the actual and nominal routes should be similar. Moreover, under normal circumstances, ships of the same type traveling the same route at the same place and time should exhibit similar navigation behavior. Intuitively, the mutual influence between trajectories is also related to the spatial distance and time interval between them.

As shown in Fig 3, suppose that the times at which the vessel passes through certain fixed regions are t_1, t_2 and $t_{current}$, respectively. The time interval between t_1 and $t_{current}$ is denoted by Δt_1. Similarly, the time interval between t_2 and $t_{current}$ is denoted by Δt_2.

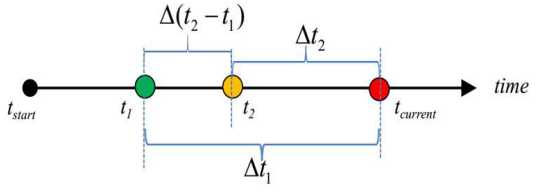


FIGURE 3. The effect of time factors.

If Δt_1 is long, then even if the recorded values for t_1 and $t_{current}$ are very similar, this does not imply high relevancy between t_1 and $t_{current}$ because other factors, such as speed or course, may change considerably over Δt_1. Therefore, a shorter Δt_1 (greater temporal closeness) generally indicates a greater contribution of the recorded values to the relevancy between the two records. Thus, the contribution of the value at t_1 can be approximately weighted by an exponential decay function of Δt_1, namely, a time control function, which is defined as follows:

$$f_1 = e^{-\alpha |t_{p_{predict}} - t_{p_h}|} \tag{2}$$

where $\alpha \geq 0$ is a non-negative decay constant, such that a larger α causes the value of f_1 (ranging from 1 to 0) to vanish more rapidly with an increasing time interval Δt_1; $t_{p_{predict}}$ is the time of the prediction; and t_{p_h} is the time of the historical record.

Analogously, we define a distance control function as follows:

$$f_2 = e^{-\beta \sqrt[2]{(la_{p_{predict}} - la_{p_h})^2 + (lon_{p_{predict}} - lon_{p_h})^2}} \tag{3}$$

where $\beta \geq 0$ is a non-negative decay constant, such that a larger β causes the value of f_2 (ranging from 1 to 0) to vanish more rapidly with increasing distance; $la_{p_{predict}}$ and $lon_{p_{predict}}$ are the latitude and longitude, respectively, of the predicted point; and la_{p_h} and lon_{p_h} are the latitude and longitude, respectively, of the historical record.

(2) The trajectory characteristics and navigation behaviors of ships of different types have a certain relevance to each other when the ships are traveling in neighboring areas during adjacent time periods.

Ships of different types exhibit different navigation behaviors even if they are traveling in the same channel. However, there still exist certain correlations between the trajectory characteristics and navigation behaviors of ships of different types in neighboring areas and adjacent time periods. Therefore, an adjustment function to calibrate the difference between vessels of different types is defined as follows:

$$f_3 = e^{-\gamma} \tag{4}$$

where $\gamma \geq 0$ is a non-negative decay constant, such that a larger γ causes the value of f_3 (ranging from 1 to 0) to vanish more rapidly.

(3) Historical routing features and movement features can assist in judging an undetermined trajectory.

Consider a situation such as that illustrated in Fig 4(a). Suppose that there are two historical trajectories, T_1 (in green) and T_2 (in yellow). T_3 (in red) is another trajectory, from which some data have been lost, and it is necessary to determine which path this third trajectory is following, T_1 or T_2. We can use the historical trajectories and the movement features of the vessel traveling on T_3. In Fig 4(b), the black records represent points obtained through calculation. From Fig 4(b), we can see that a more suitable trajectory can be found based on historical records.

Based on the discussion above, we propose a time-aware and spatially correlated collaborative method of solving the missing trace problem. We also use historical records and other types of ship records to calibrate trajectory data.

Suppose that a raw trajectory $T = \{p_1, p_2, \ldots, , p_n\}$ starts from p_{start} and ends at p_{end}. The time interval or spatial distance between two adjacent points p_i and p_{i+1} is much larger than a given value. If historical data exist for some locations between p_i and p_{i+1}, then we use these historical points to

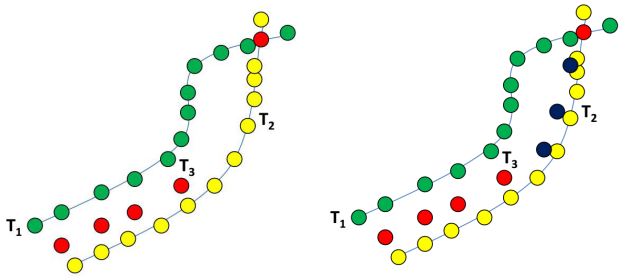


FIGURE 4. Using features to reduce data uncertainty. (a) The trajectory before processing. (b) The trajectory after processing.

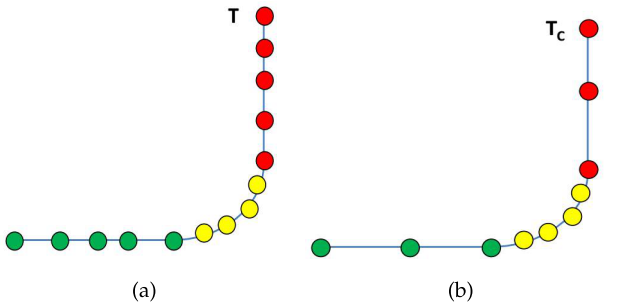


FIGURE 5. The result of trajectory compression. (a) The raw trajectory. (b) The compressed trajectory.

predict the missing points in this trajectory. First, we use the reference points at the locations of the missing points as a baseline and use an adaptive exponential function to calibrate the results. However, there may be no records or no historical points for routes of this type at the locations of the missing records; in this case, we will need to use records for routes of other types to predict the missing records. We again use an exponential function to adjust the results. Finally, the prediction formula is defined as follows:

$$p_{predict} = \lambda_1 * rp + \lambda_2 * p_h * f_1 * f_2 + (1 - \lambda_1 - \lambda_2) * p_o * f_3 \quad (5)$$

where $p_{predict}$ is a missing point that needs to be predicted; rp denotes any relevant reference points; p_h denotes relevant historical records; p_o denotes records for other types of vessels; λ_1 and λ_2 are the adjustment factors used to modify the proportions between the values of rp, p_h and p_o; f_1 and f_2 are the adjustment functions used to calibrate the differences in time and distance between $p_{predict}$ and p_h; f_3 is the adjustment function used to calibrate the difference between vessels of different types.

Once we have filled in and compressed the raw trajectory, we obtain a new trajectory with fewer points but without loss of the routing features and movement features of the raw trajectory. Next, we calculate the similarity between different trajectories, based on reference trajectories, to distinguish abnormal trajectories. When calculating the similarity between different trajectories, the newly processed trajectories are used.

2) TRAJECTORY COMPRESSION

Because of the high collection rate, the amount of AIS data is large. By virtue of the large amount of data and the regularity of vessel movements, we can compress the data into a much smaller volume without loss of important information. In fact, the fundamental criterion for high-quality compression is that it does not affect the subsequent use of the compressed trajectory data. For example, in Fig 3, the red, yellow and green points represent different PCs, as defined in the previous section. The raw trajectory records are shown in Fig 5(a). The red and green sub-trajectories can be compressed as shown in Fig 5(b) by applying a compression method based on the routing features and movement features. Through the trajectory compression method, the data volume of a trajectory can be greatly reduced. However, since the navigation behavior features are preserved, the new trajectory does not lose any important information. Therefore, in this paper, we use features extracted from raw trajectories to obtain compressed trajectories.

Using AIS data, the trajectory of a moving vessel can be constructed. We use the latitude, longitude and time to represent a trajectory. As described above, the trajectory of a moving object is denoted by a sequence of points: $T = \{p_1, p_2, \ldots, p_n\}$, where each point consists of a set of geospatial coordinates and a timestamp, i.e., $p_i = (o, l, t, s, c)$. $i \in \{1, 2, \ldots, n\}$.

In the field of databases, various techniques have been studied for compressing trajectory data. The most common method is Piecewise Linear Segmentation (PLS). The PLS algorithm compresses a trajectory T into linear segments by recursively retaining points that have a maximum error higher than a fixed threshold. Thus, the goal of the algorithm is to reduce the number of points in a trajectory while keeping the maximum deviation, or error, from the original trajectory within the specified threshold. We use the PLS algorithm for trajectory compression.

Compression using the PLS algorithm with any of these error measures can lead to problems with retaining stop points in trajectories. It is likely that a trajectory will be reduced in such a way that the vessel appears to be moving slowly in the compressed trajectory, whereas in the original, uncompressed trajectory, the vessel in fact stops moving for a period of time. Knowing whether a vessel stops or not is important for the behavior analysis of vessels; thus, we prefer to retain this kind of information during compression. The pseudo code for this algorithm is provided below (**Algorithm I**).

With the trajectory compression method, the data volume of the source trajectories is greatly reduced. However, since the navigation behavior features are preserved, the new set of trajectories does not lose any important information, and we can obtain a much smaller data set. In the next section, we use the newly compressed trajectories to describe trajectory characteristics and navigation behavior characteristics and to define data mining algorithms.

Algorithm 1 pls (T, ϵ)

Require: T, ϵ
Ensure: T_C
  $d_{max} \leftarrow 0, i_{max} \leftarrow 0$
  for $i \leftarrow 2to(end-1)$ **do**
    $d = E(T(i), T(1), T(end))$
    if $d > d_{max}$ **then**
      $i_{max} \leftarrow i$
      $d_{max} \leftarrow d$
    end if
  end for
  if $d_{max} \geq \epsilon$ **then**
    A = **pls**(T(1,i_{max}),ϵ)
    B = **pls**(T(i_{max},end),ϵ)
    T_C = A,B(2,end)
  else
    T_C = T(1),T(end)
    $N \leftarrow n$
  end if
  return T_C

D. ANOMALOUS TRAJECTORY DETECTION

The data in a static trajectory set usually satisfy an a priori distribution. In the absence of a labeled training data set, a global feature model can be constructed by excavating all trajectory information from the historical trajectory data. After we use the spatio-temporal relationship-based collaborative method to predict the missing records and use the trajectory compression algorithm to reduce the large size of the trajectory data, we exploit the newly processed trajectory data to detect anomalous trajectories. However, the continual updating of trajectory data may result in a new baseline and changes to the references for anomaly detection. Therefore, we propose a time-aware solution to balance the updating of the new trajectory data and the reference system to solve the problem of new information in anomaly detection.

1) SIMILARITY CALCULATION

There are many approaches to performing outlier detection tasks on vessel trajectory data. We use a similarity-based intelligent algorithm. We use similarities that are defined based on the alignments between two trajectories. Alignment measures are sufficiently flexible for dealing with trajectories of different lengths in terms of the number of points, time, and distance traveled. The compression method described in the previous section substantially reduces the number of points, which makes the alignment computation faster. However, compression may potentially have a negative influence on the quality of the alignments. We discuss this influence and the performance of alignment measures.

First, we cluster the raw trajectory data using the *DBSCAN* algorithm, and we use the geometric centers of the resulting clusters as reference points. Finally, we use these reference points to construct new reference trajectories. When calculating the similarity between two different trajectories, we use the reference trajectories as the reference system for ships of the same kind.

To calculate the similarity as defined above, given a trajectory T, we use the reference trajectories (denoted by $\overline{RT}$) of the same type as the reference system in addition to a reference trajectory set $S = \{\overline{RT_1}, \overline{RT_2}, ..., \overline{RT_k}\}$ that includes reference trajectories for other types of vessels and a trajectory distance function d. Then, we calculate the similarity between the reference trajectories and the new trajectory T. The specific formula is defined as follows:

$$Sim_T = \omega * d(T, \overline{RT}) + (1-\omega) * \sum_{i=0}^{k} \omega_{\overline{RT_i}} * d(T, \overline{RT_i}) \quad (6)$$

where ω is an adjustment factor to modify the proportions of the values and the $\omega_{\overline{RT_i}}$ are the weights of the reference trajectories, which satisfy $\sum_{i=0}^{k} \omega_{\overline{RT_i}} = 1$. In this paper, we use the *DTW* (dynamic time warping) algorithm to calculate the values of $d(T, \overline{RT})$ and $d(T, \overline{RT_i})$.

2) TIME-AWARE COLLABORATIVE INDICATOR OF THE DEGREE OF SUSPICION

Based on various outlying features, a unified evaluator is required to compute outlier scores. Thus, we propose a collaborative indicator of the degree of suspicion, which provides a collaborative approach to integrating the characterized outlying scores into a collaborative indicator for each detected trajectory. The continual updating of the trajectory data may result in new information for anomaly detection. Based on the above analysis, we propose a time-aware solution to balance the updating of new trajectory data and the reference system to address this issue. At the same time, to compute the similarity between trajectories of different sizes, the result must be normalized.

The continual updating of the trajectory data may may result in a new baseline and changes to the references for anomaly detection. In a practical situation, the new trajectories are also affected by other trajectories that are near to them in time, especially at relatively short time intervals. Therefore, the impact of these near-time trajectories on a detected trajectory must be considered. At the same time, we also use the historical records and other types of ship records to calibrate the similarity. Finally, we integrate these outlying scores into a collaborative indicator of the degree of suspicion as follows:

$$\begin{aligned} Sim_T = {} & \omega_1 * d(T, \overline{RT}) + \omega_2 * \sum_{i=0}^{k} \omega_{\overline{RT_i}} * d(T, \overline{RT_i}) \\ & + (1-\omega_1-\omega_2) * \sum_{i=0}^{k} f_1 * d(T, T_{current}) \end{aligned} \quad (7)$$

where ω_1 and ω_2 are the factors used to adjust the proportions of the values $d(T, \overline{RT})$, $d(T, \overline{RT_i})$ and $d(T, T_{current})$; the $\omega_{\overline{RT_i}}$ are the weights of the reference trajectories, as defined above;

TABLE 4. Analysis of the first port.

Cargo	Dredgers	Dredging or UW Ops	Fishing Vessels	High-Speed Craft	Law Enforcement
7.2%	0.3%	0.5%	2.2%	0.1%	17.4%
Other	Passenger Ships	Pilot Vessels	Pleasure Craft	Port Tender	Reserved
1.0%	0.01%	0.01%	0.01%	0.01%	0.01%
Tankers	Towing Vessels	Tugs	Unknown	Unspecified	WIG
9.6%	13.7%	13.6%	16.3%	0.7%	17.1%

TABLE 5. Analysis of the second port.

Anti-Pollution	Cargo	Dredging or UW Ops	Fishing Vessels	High-Speed Craft	Law Enforcement
0.02%	79.8%	0.08%	0.4%	0.2%	2.7%
Military Ops	Other	Passenger Ships	Pilot Vessels	Pleasure Craft	Port Tender
0.01%	0.7%	0.07%	0.02%	0.5%	0.01%
Tankers	Towing Vessels	Tugs	Unknown	Unspecified	WIG
3.0%	0.2%	0.4%	10.2%	1.1%	0.8%

and the function f_1 is the adjustment function used to calibrate the time impact between trajectories T and $T_{current}$, as defined in the previous section.

Additionally, to compute the similarity between trajectories of different sizes, we normalize the calculated similarity values as follows:

$$d(T, \overline{RT}) = \frac{DTW(T, \overline{RT})}{|T| + |\overline{RT}|} \tag{8}$$

$$d(T, \overline{RT_i}) = \frac{DTW(T, \overline{RT_i}))}{|T| + |\overline{RT_i}|} \tag{9}$$

$$d(T, T_{current}) = \frac{DTW(T, T_{current}))}{|T| + |T_{current}|} \tag{10}$$

where $|T|$, $|\overline{RT}|$, $|\overline{RT_i}|$ and $|T_{current}|$ are the numbers of points in trajectories T, $\overline{RT}$, $\overline{RT_i}$ and $T_{current}$, respectively. When we calculate the values of the collaborative indicator of the degree of suspicion, the normalized similarity values are used.

After we obtain the indicator values of the degree of suspicion for every detected trajectory, a lower score indicates a higher outlier degree. As stated in Definition 8, if $Sim_T > \theta$, we call Sim_T a θ-outlier.

IV. EXPERIMENT

In this section, we report a series of experiments conducted to validate the effectiveness of our algorithm. All experiments were performed using a real-world dataset on a workstation with 12 Intel Xeon 3.50 GHz CPUs, 32 GB of main memory and 64-bit Windows 10 as the operating system. The algorithm was implemented in JDK8.

A. EXPERIMENTAL SETUP

1) DATA PREPARATION

We used a real-world trajectory data set of vessels in the East China region over one month. This data set contains more than 12,000,000 records and more than 40,000 trajectories. Because of the enormous amount of data, only a portion of the data was extracted from this data set to be used as the experimental data. We clustered the raw data using the *DBSCAN* algorithm and used the geometric centers of the resulting clusters as the reference points. In our algorithm, we set $\lambda_1 = 0.8$, $\lambda_2 = 0.1$, $\omega = 0.9$, $\omega_1 = 0.8$, $\omega_2 = 0.1$, $\alpha = \beta = \gamma = 0.085$, $\theta = 0.8$ and $\epsilon = 200$.

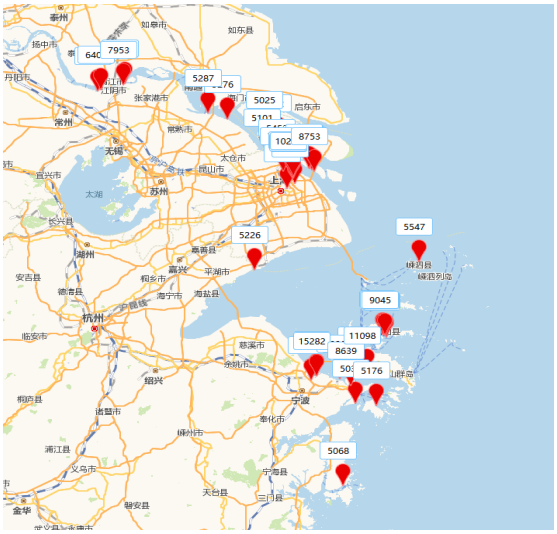
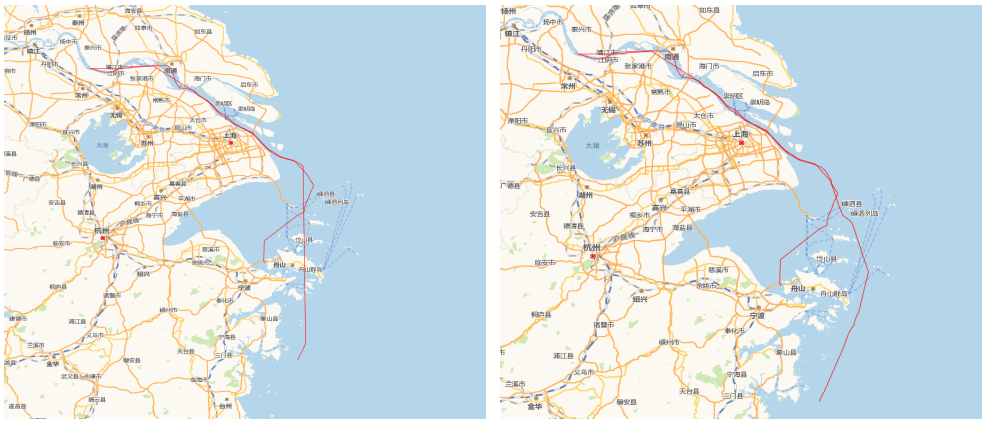


FIGURE 6. Important ports (POIs).

2) CLUSTERING SEMANTICS

First, we used the *DBSCAN* algorithm to cluster the raw data. In addition, when we applied certain rules, such as a vessel speed below a given threshold (in this paper, we use 0.5) and a number of neighboring points higher than a given threshold, we could identify several important ports, as shown in Fig 6. The value shown for each point in Fig 6 represents the number of vessels passing through this port during the given period. We also obtained some helpful semantic information concerning these ports. Examples are shown in Table 4 and Table 5. The two ports represented in these tables are typical cases. Based on this semantic information, we can gain a clear understanding of the ports. The business scope of the



(a) (b)

FIGURE 7. The results of missing data prediction. (a) The raw trajectory. (b) The processed trajectory.

first port includes Law Enforcement (17%), WIG (17%), Unknown Types (16%), Tugs (14%), Tankers (10%), Towing Vessels (9%), Cargo (7%) and others; therefore, it is a comprehensive port that hosts many different types of vessels. The business scope of the second port includes Cargo (80%), Unknown Types (10%) and others. This is a port with cargo transport as its main business. In this way, we gathered clear semantic information regarding the ports.

B. PERFORMANCE EVALUATION

1) EFFECTIVENESS OF MISSING DATA PREDICTION

When there exists a long time interval or a long spatial distance between two adjacent points of a trajectory, the missing record problem may arise. To improve the accuracy of the similarity calculation, we must fill in the missing records. As discussed above, we use the proposed time-aware and spatially correlated collaborative algorithm to solve the missing trace problem. We also use historical records and other types of ship records to calibrate the trajectory data. In this experiment, we applied the proposed method of predicting missing data. For the real trajectory dataset, Fig 7(a) shows a raw trajectory as a red line. Because of partial data loss, the trajectory of this ship crosses the island. Fig 7(b) shows the trajectory after processing using our prediction method, also as a red line. The processed trajectory is the real trajectory of the ship.

2) ANOMALOUS TRAJECTORY DETECTION

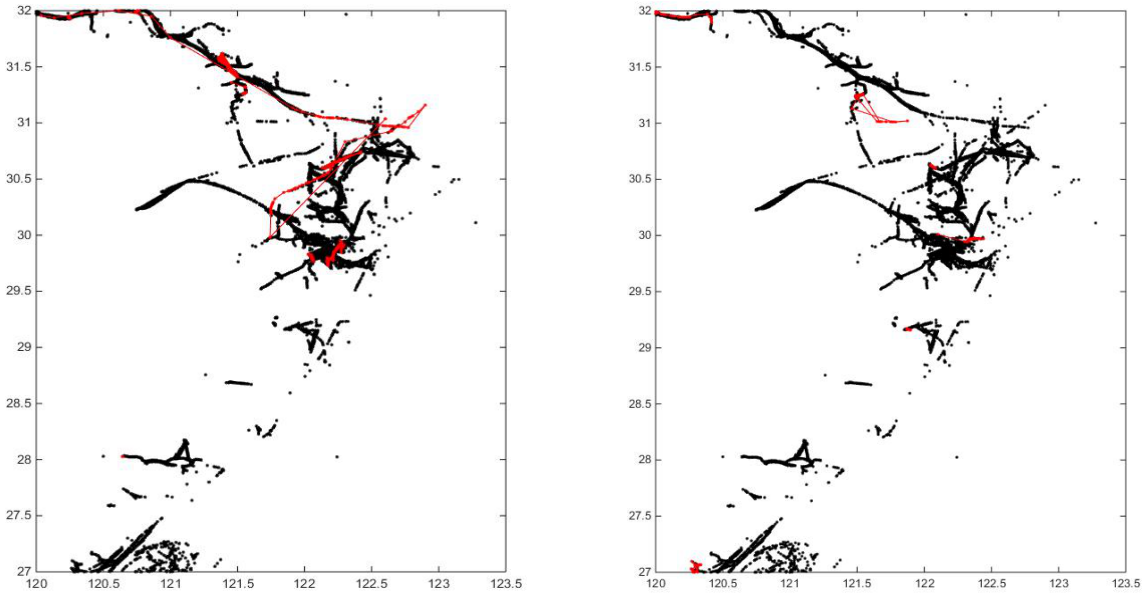
The primary goal of our work is to design an intuitive, accurate and easy-to-understand method of anomaly detection. Ideally, accurate abnormal trajectories should be displayed on the map in real time. Therefore, in this experiment, we tested the time complexity of our anomalous trajectory detection algorithm, which is especially important for online detection systems. Moreover, we tested and demonstrated the accuracy of our method. Fig 8(a) illustrates the effectiveness of our feature learning algorithm. Fig 8(b) shows the result obtained without our algorithm. The black dots represent normal trajectories, and the red lines represent abnormal trajectories. From these two images, we can see that our method significantly outperforms the basic approach. This is because we use the time-aware and spatially correlated collaborative algorithm to solve the missing trace problem, and we also use historical records and records of other types of ships to calibrate the trajectory data. After we have calculated more precise values to replace lost data, the accuracy of the similarity calculation is obviously increased. Simultaneously, because we use the time-aware and spatially correlated collaborative algorithm to calculate the similarity and use historical records and other types of ship records to calibrate the similarity values, we can obtain more precise trajectory similarities. This is why our algorithm can achieve more accurate results.

In addition, because we compress the large amount of trajectory data into a much smaller volume without affecting its further use, our algorithm can obviously reduce computation times. In fact, as the number of raw trajectories increases, the computation time of our approach increases very slowly compared with that of the basic approach. This is expected behavior because of the much smaller number of points considered in the computation.

V. RELATED WORKS

Our work involves two research problems concerning trajectory data: using feature learning to process trajectories and anomalous trajectory detection. First, we review some existing works on the mining of movement behavior from trajectory data. Then, we survey the techniques for detecting moving objects with anomalous movements in their trajectories.

Over the past few years, many studies have focused on trajectory analysis. Representative works include the design of effective trajectory indexing structures [1], [2] and methods for trajectory query processing [3], [4], uncertainty management [5], [6], and mining knowledge/patterns from trajectories [7], [8]. Effective index structures [1]–[4], [12]



(a) (b)

FIGURE 8. Comparison of the results obtained with and without our algorithm. (a) Result with our algorithm. (b) Result without our algorithm.

have been built to manage trajectories and support efficient trajectory queries. Data mining methods have been applied to trajectories to detect important PoIs and find popular routes [7], [13]. Attention has also been paid to the semantic representation or interpretation of trajectory data by associating or annotating GPS locations with semantic entities [14]–[17]. In addition, semantic representation is also a hot research area in the field of anomalous trajectory detection [18], [19].

Many articles have offered solutions for the management of uncertain data. Gao *et al.* [20] proposed an anchor-based calibration system and a more advanced calibration method based on a spatial-only model, which exploits the power of machine learning techniques to train inference models from historical trajectory data to improve calibration effectiveness. Given a set of trajectories, Evans *et al.* [21] proposed a solution for clustering the trajectories into several groups and representing each group by its most central trajectory. The assignment of trajectory sample points to certain semantic anchor points, i.e., PoIs, shares similar motivations with some studies on the construction of semantic trajectories; for example, Andrae [22] summarized a set of trajectories by providing a symbolic route to represent the cardinal trajectory directions.

Additionally, several works have focused on trajectory segmentation. In [23], the authors proposed a method of segmenting heterogeneous trajectories into several parts according to different means of transportation, e.g., by bike and by car. However, this trajectory segmentation method clearly cannot be applied to a trajectory that has been entirely generated by the same mode of transportation. In [11], the authors proposed a partition-and-summarization approach to using semantic information to segment and summarize individual trajectories. Dedicated algorithms have been independently designed for annotating trajectories with geographic regions or lines. Regarding annotation with geographic regions, some studies [15], [24] have focused on computing topological correlations (called spatial predicates) between trajectories and regions. Regarding annotation with geographic lines, many works [25]–[28] have focused on identifying the correct road segment on which a vehicle is traveling. Reference [25] used only the geometric information of the underlying road network and applied distance measurements to generate line annotations. Reference [26] accounted for the connectivity and contiguity of the road network in addition to the geometric distances. References [27] and [28] studied the generation of annotations for low-sampling-rate trajectories.

Many works have addressed missing data prediction. In [20], the author surveyed the key techniques for the data processing of trajectory data. In [29], the author took advantage of mobile phone networks, which offer enormous amounts of spatial and temporal communication data, and used a correlation-based clustering method for anomalous trajectory detection. In [30], the author proposed a time-aware approach for missing data prediction. [31] proposed

a tensor-based method for the completion of missing traffic data. This approach not only inherits the advantages of imputation methods based on matrix patterns for estimating missing points but also effectively mines the multi-dimensional inherent correlations in traffic data.

The existing algorithms for abnormal trajectory detection focus on how to find the most representative trajectory out of a set of trajectories. In [32], the author detailed the research progress in coping with enormous amounts of low-quality trajectory data for anomalous trajectory detection. Given a set of trajectory data, [33] applied a piecewise linear segmentation method to compress each trajectory and then used a similarity-based approach to perform clustering, classification and outlier detection using kernel methods. In [34], the author proposed an online method that is able to detect anomalous trajectories ''on the fly'' and to identify which parts of a detected trajectory are responsible for its anomalous nature. In [35], the authors proposed a framework called MT-MAD for maritime trajectory modeling and anomaly detection. They used the outlying features required for anomaly detection, including spatial, sequential, and behavioral features, and then explored the movement behavior indicated by historical trajectories and built a maritime trajectory model for anomaly detection. The proposed model accurately describes movement behavior and captures outlying features in trajectory data.

VI. CONCLUSION

In this paper, to solve the problems encountered when mining marine trajectory data, such as the uncertainty, sparseness, skewness, large scale and fast updating of such trajectory data, we propose a novel abnormal trajectory detection system. This system can detect abnormal vessel trajectories from the AIS records of the vessels by means of our feature learning algorithm. First, we use an effective method to pre-process the trajectory data by applying missing data prediction and compression to cope with the uncertainty and large volume of these data. Then, we use the *DTW* algorithm to calculate the similarity between trajectories and reference trajectories. Finally, a time-aware approach for balancing the influence of the updating of new trajectory data and the reference system is proposed to address the issue of new information. We used a real-world dataset to test our algorithm. The experimental results demonstrate that the proposed framework is capable of effectively detecting anomalous AIS trajectories. In the future, we will analyze the complex relationships between vessels of multiple types and attributes in a region, and we will study methods of analyzing ship group situations and predicting the behavior of ship groups.

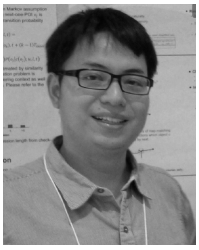
REFERENCES

[1] Y. Cai and R. Ng ''Indexing spatio-temporal trajectories with Chebyshev polynomials,'' in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2004, pp. 599–610.
[2] D. Wu *et al.*, ''Data engineering (ICDE),'' in *Proc. IEEE 27th Int. Conf.*, Jun. 2011, pp. 541–552.
[3] L. Chen, M. T. Özsu, and V. Oria, ''Robust and fast similarity search for moving object trajectories,'' in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2005, pp. 491–502.
[4] M. Vlachos, G. Kollios, and D. Gunopulos, ''Discovering similar multidimensional trajectories,'' in *Proc. 18th Int. Conf.*, 2002, pp. 673–684.
[5] K. Zheng *et al.*, ''Probabilistic range queries for uncertain trajectories on road networks,'' in *Proc. 14th Int. Conf. Extending Database Technol.*, 2011, pp. 283–294.
[6] K. Zheng *et al.*, ''Reducing uncertainty of low-sampling-rate trajectories,'' in *Proc. 14th Int. Conf. Extending Database Technol.*, Apr. 2012, pp. 1144–1155.
[7] H. Jeung *et al.*, ''Discovery of convoys in trajectory databases,'' *Proc. VLDB Endowment*, vol. 1, no. 1, pp. 1068–1080, 2008.
[8] K. Zheng *et al.*, ''On discovery of gathering patterns from trajectories,'' in *Proc. IEEE 29th Int. Conf. Data Eng. (ICDE)*, Apr. 2013, pp. 242–253.
[9] J. B. Kruskal, ''An overview of sequence comparison: Time warps, string edits, and macromolecules,'' *SIAM Rev.*, vol. 25, no. 2, pp. 201–237, 1983.
[10] J. K. Kearney and S. Hansen, ''Stream editing for animation,'' Dept. Comput. Sci., Iowa Univ., Iowa, IA, USA, Tech. Rep. TR-90-08, 1990.
[11] H. Su *et al.*, ''Making sense of trajectory data: A partition-and-summarization approach,'' in *Proc. IEEE 31st Int. Conf. Data Eng. (ICDE)*, Apr. 2015, pp. 963–974.
[12] H. Wang *et al.*, ''SharkDB: An in-memory columnoriented trajectory storage,'' in *Proc. 23rd ACM Int. Conf. Inf. Knowl. Manage.*, 2014, pp. 1409–1418.
[13] B. Zheng *et al.*, ''Efficient retrieval of top-K most similar users from travel smart card data,'' in *Proc. IEEE 15th Int. Conf. Mobile Data Manage. (MDM)*, Jul. 2014, pp. 259–268.
[14] Z. Yan and S. A. Spaccapietra, ''Towards semantic trajectory data analysis: A conceptual and computational approach,'' in *Proc. VLDB PhD Workshop*, 2009, p. 24.
[15] S. Spaccapietra *et al.*, ''A conceptual view on trajectories,'' *Data Knowl. Eng.*, vol. 65, no. 1, pp. 126–146, 2008.
[16] Z. Xu *et al.*, ''Building knowledge base of urban emergency events based on crowdsourcing of social media,'' *Concurrency Comput., Pract. Exper.* vol. 28, no. 15, pp. 4038–4052, 2016.
[17] Z. Xu *et al.*, ''Building the search pattern of Web users using conceptual semantic space model,'' *J. Int. J. Web Grid Services*, vol. 12, no. 3, pp. 328–347, 2016.
[18] Z. Xu ''Crowdsourcing based description of urban emergency events using social media big data,'' *IEEE Trans. Cloud Comput.*, to be published, doi: 10.1109/TCC.2016.2517638.
[19] Z. Xu *et al.*, ''From latency, through outbreak, to decline: Detecting different states of emergency events using Web resources,'' *IEEE Trans. Big Data*, to be published, doi: 10.1109/TBDATA.2016.2599935.
[20] Q. Gao, Z. Fl, W. Rj, and F. A. Zhou, ''Trajectory big data: Review of key technologies in data processing,'' *J. Softw.* [Online]. Available: http://www.jos.org.cn/1000-9825/5143.htm
[21] M. R. Evans *et al.*, ''Summarizing trajectories into, k-primary corridors: A summary of results,'' in *Proc. 20th In. Conf. Adv. Geograph. Inform. Syst.*, 2012, pp. 454–457.
[22] S. Andrae and T. Kaernten, ''Summarizing GPS trajectories by salient patterns,'' *Strobl*, 2005.
[23] Z. Yan *et al.*, ''SeMiTri: A framework for semantic annotation of heterogeneous trajectories,'' in *Proc. 14th Int. Conf. Extending Database Technol.*, 2011, pp. 259–270.
[24] M. E. Nergiz, M. Atzori, Y. A. Saygin, and B. Güç ''Towards trajectory anonymization: A generalization-based approach,'' in *Proc. SIGSPATIAL ACM GIS Int. Workshop Secur. Privacy GIS LBS*, 2008, pp. 52–61.
[25] D. Bernstein and A. Kornhauser, ''An introduction to map matching for personal navigation assistants,'' *Geometric Distrib.*, vol. 122, no. 7, pp. 1082–1083, 1998.
[26] C. E. White, D. Bernstein, and A. L. Kornhauser, ''Some map matching algorithms for personal navigation assistants,'' *Transp. Res. C, Emerg. Technol.*, vol. 8, nos. 1–6, pp. 91–108, 2000.
[27] P. Newson and J. Krumm, ''Hidden Markov map matching through noise and sparseness,'' in *Proc. 17th ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, 2009, pp. 336–343.
[28] Y. Lou, C. Zhang, Y. Zheng, X. Xie, W. Wang, and Y. Huang, ''Map-matching for low-sampling-rate GPS trajectories,'' in *Proc. 17th ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, 2009, pp. 352–361.

[29] S. Liu, L. Chen, and L. M. Ni, "Anomaly detection from incomplete data," *ACM Trans. Knowl. Discovery Data (TKDD)*, vol. 9, no. 2, p. 11, 2014.
[30] Y. Hu, Q. Peng, and X. Hu, "A time-aware and data sparsity tolerant approach for Web service recommendation," in *Proc. IEEE Int. Conf. Web Ser. (ICWS)*, Jun. 2014, pp. 33–40.
[31] H. Tan, G. Feng, J. Feng, W. Wang, Y.-J. Zhang, and F. Li, "A tensor-based method for missing traffic data completion," *Transp. Res. C, Emerg. Technol.*, vol. 28, pp. 15–27, Mar. 2013.
[32] J. L. Mao, C. Q. Jin, Z. G. Zhang, and A. Y. Zhou, "Anomaly detection for trajectory big data: Advancements and framework," *Ruan Jian Xue Bao/J. Softw.*, (in Chinese). vol. 28, no. 1, pp. 17–34, 2017. [Online]. Available: http://www.jos.org.cn/1000-9825/5151.htm
[33] G. K. D. de Vries and M. van Someren, "Machine learning for vessel trajectories using compression, alignments and domain knowledge," *Expert Syst. Appl.*, vol. 39, no. 18, pp. 13426–13439, 2012.
[34] C. Chen *et al.*, "iBOAT: Isolation-based online anomalous trajectory detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 806–818, Jun. 2013.
[35] P.-R. Lei, "A framework for anomaly detection in maritime trajectory behavior," *Knowl. Inf. Syst.*, vol. 47, no. 1, pp. 189–214, 2016.

PEIGUO FU received the master's degree in operational research and cybernetics from the Faculty of Science, Harbin Institute of Technology, China, in 2011. He is currently pursuing the Ph.D. degree with the Science and Technology on Integrated Information System Laboratory, Institute of Software, Chinese Academy of Sciences, Beijing, China. His main research interests include big data analytics, spatio-temporal data mining, and artificial intelligence.

HAOZHOU WANG received the Ph.D. degree in computer science from The University of Queensland in 2015. His main research interests include social media data analysis, spatiotemporal database, uncertain database, data mining, and bioinformatics.



KUIEN LIU received the Ph.D. degree from the Institute of Software, Chinese Academy of Sciences, Beijing, China. His main research interests include social media data analysis, spatiotemporal database, and data mining.

XIAOHUI HU is currently a Professor and a Ph.D. Supervisor with the Science and Technology on Integrated Information System Laboratory, Institute of Software, Chinese Academy of Sciences. His main research interests include information system integration, big data analytics, and artificial intelligence.

HUI ZHANG received the B.E. degree from the University of Science and Technology of China in 2007 and the Ph.D. degree from the Shanghai Institute of Technical Physics, Chinese Academy of Sciences, in 2012. He is currently an Associate Professor with the Institute of Software, Chinese Academy of Sciences. Her research interests include pattern recognition and remote sensing image processing.