IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 16, 2020, accepted November 30, 2020, date of publication December 4, 2020, date of current version December 16, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3042614

A Trajectory Classification Model Using Grammar Parsing

LEI BAO
Information Science and Technology Department, Xiamen University Tan Kah Kee College, Zhangzhou 363105, China
e-mail: blnj2000@163.com

ABSTRACT Trajectory classification is a hot topic in the field of spatiotemporal data mining. Existing models exert spatial or temporal computation on trajectory data, which require huge efforts and are often time consuming and lack of efficiency. This article proposes a model to classify unknown ship trajectories through a syntax recognition approach. By using the background semantic information in the rasterized sea chart, the model transforms the ship trajectories into symbolic sentences containing both spatiotemporal and semantic information, and reduces their scale. The class feature is expressed as a context-free grammar and the data classification is implemented through syntax parsing. The parsing requires less computation and is more efficient. Experiments are carried out to verify the model's practicability, and the results show that it is valid and effective.

INDEX TERMS Automatic identification system, spatiotemporal data mining, syntactic analysis, trajectory classification.

I. INTRODUCTION

With the continuous development on the acquisition technology of spatiotemporal data, the data mining on moving objects trajectories has become a hotspot in related research. There are two types of trajectory data according to the data formats and purposes. The first type of trajectory data is generated from tracking systems such as Automatic Identification System(AIS) or Global Positioning System (GPS), which is a sequence of location coordinates and timestamps. The second type of trajectory data is generated from image data or video data, which is a sequence of pixels in consecutive frames. Some studies focused on the preprocessing of these trajectory data including data cleaning [1], interpolation [2], [3], compression [4], [5], and segmentation [6]. Other works are devoted to explore the application of trajectory data mining in route recommendation [7]–[10], motion prediction [11], behavior understanding [12], [13], abnormal detection [14], [15] and traffic monitoring [16]–[18].

Among related works, trajectory classification is an efficient way to obtain information from trajectory data. Most of the classification approaches requires a lot of computation on the spatiotemporal data. However, the knowledge implied in trajectory data (such as purpose, intention, habit, relations) is hard to be utilized, and the computation on massive amounts of spatiotemporal data generates huge overhead. There are also semantic trajectory classification models which add semantic information into spatiotemporal trajectory to generate semantic trajectory that usually has less volume and higher description ability. Although the semantic trajectory model has advantages in classification, it still has some disadvantages. Some semantic models almost discard all the spatiotemporal information and only use the semantic data in classification computation. Whereas other semantic models still use the distance measurement between trajectory segments when they want to classify objected from their spatial or temporal features, which means heavy and inefficient computation. In short, the integration between spatiotemporal information and semantic information is still insufficient.

In this article, we propose a trajectory classification model based on grammar parsing. Its contribution is the design of a process using AIS data and geographic information to transform ship trajectories into symbolic sentences containing both semantic information and spatiotemporal information, while reducing the data size. The class feature is expressed as a context-free grammar, and the ship trajectory classification is implemented through syntax parsing, which is simpler and easier to implement. This is a preliminary study introducing text data mining technology to solve the trajectory mining problem.

II. RELATED WORKS

At present, the related trajectory classification models can be divided into two categories: the distance-based trajectory classification and the semantic trajectory classification.

The associate editor coordinating the review of this manuscript and approving it for publication was Keli Xiao.

The distance-based trajectory classification model usually extracts features from the locational or temporal data from trajectories to allocate the trajectories into cohesive classes according to their mutual similarities. There are many different distance-based similarity measurement approaches in these models, such as Euclidean Distance [19], Hausdorff Distance [20], Bhattacharyya Distance [21], Frechet Distance [22], Dynamic Time Warping (DTW) [23]–[25], Longest Common Subsequence (LCSS)[26] and Spatial Network Distance[27]. Based on the distance measurement results, these models use corresponding clustering, statistical, stochastic and deep learning algorithms in different applications. The clustering algorithms classify the trajectories via unsupervised approaches including models [28]–[30] derived from the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm and the hierarchical clustering models [31], [32]. The statistical or stochastic approaches in trajectory classification are the Nearest Neighbor(NN) models [33], Gaussian Mixture Model (GMM) [34], Bayesian models [35], [36] and Markov process models [37]. There are also models that use deep learning algorithms such as the Deep Neural Network(DNN) [38] to construct features of trajectory data which can improve classification performance.

The distance-based trajectory classification models work on the temporal or spatial attributes from the trajectory data. The similarity measurement between trajectories requires heavy computation on locational or temporal data, which becomes a huge burden as the size of data set increasing. The semantic trajectory usually has less volume and is more descriptive. And it has been widely used in trajectory data mining in recent years [39], [40].

The semantic trajectory is generated by adding semantic information such as the states, contextual information, and relationships of objects into spatiotemporal data sequences. There are four types of semantic trajectory approaches adopted in recent classification applications, the Stop/Move model, the trajectory segment model, the event-based semantic trajectory and the ontology-based semantic trajectory. The Stop/Move semantic trajectory model is currently the most widely used approach [41]–[43]. This model defines two basic movement states of stay and movement [44], and use the "stay" sub-trajectory and the "move" sub-trajectory to express the spatiotemporal trajectory. The trajectory segment model [45], [46] organizes the movement segments and their semantic relationships to form a trajectory semantic sequence. The event-based semantic trajectory annotates the changes of the state, position and attributes of the moving object in the form of events, then the spatiotemporal trajectory is transformed into a sequence of events for further process [47], [48]. These annotation methods can inject semantic information into spatiotemporal trajectories relatively freely, but the definition of relationships between attributes is not clear enough, which is not convenient for the inference, analysis and mining. There are also ontology-based semantic trajectory models that use ontology to formally describe concepts such as stay, movement, time, place, and mode in mobile behavior [49], so as to add semantic information to the original trajectory. The ontology-based semantic trajectory model can provide abundant and accurate semantic information, but it cannot be actually applied in real systems without an ontology library, which is still unavailable in most cases.

Comparing with the distance-based models, these semantic trajectory classification models have advantages in terms of semantics, interpretation and feasibility. The models usually reduce the amounts of computation on spatial and temporal raw data and introduce semantic information to assist the classification. However, the existing semantic trajectory models still have problems. On the one hand, there are still heavy distance computation overhead in Stop/Move derivative models when measuring the similarity between move segments. On the other hand, some semantic models almost discard all the spatiotemporal information and only use the semantic data in subsequent calculation.

As shown in Fig.1, the Stop/Move model is composed of stop points and move segments, the semantic information can be labeled on the stop points. However, when we are doing the classification algorithms, distance measurement on move segments which means huge computation on spatiotemporal data is still inevitable. Other semantic models such as the trajectory segments model and event-based model focus on the semantic sequence generated from the original trajectory, and some works even discard the locational and temporal attribute data completely. As a result, these algorithms can not accurately classify trajectories that have the same semantic stop area but with different behavior during their traveling stages. To cope with this problem, some trajectory segment model retains the original spatiotemporal trajectory as well as the semantic sequence, but the consequent classification still need to do the same distance measurement task as we mentioned above.

In this article, we propose a trajectory classification model using grammar parsing. It can be categorized as semantic trajectory model, but unlike the Stop/Move model or other semantic trajectory approaches, trajectory is described as a symbolic sentence and the pattern is a grammar. The classification is realized by syntactic parsing. Our motivation is to implement spatiotemporal data classification in a structured pattern recognition way.

III. MODEL FRAMEWORK

Since the construction of the AIS network, large amount of vessel trajectory data has been accumulated, which contains mainly the location data explicitly recorded in time series. AIS ship trajectory and Vessel Traffic Service(VTS) data records current or historical voyage data. These data sources are the main data sources for analysis, which have the characteristics of large amount, fragmented information, low accuracy, and semi-structured. To effectively classify the trajectories, it is not enough to use these data alone. It is necessary not only to analyze the location data of the ship's movement, but also to use heterogeneous data from other

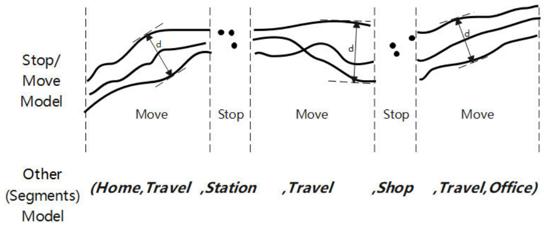


FIGURE 1. The Stop/Move trajectory and semantic sequence.

TABLE 1. Main Data Sources.

	Features	Uses
AIS data	Crowd source data, high coverage, easy access, low data quality	Trajectory data sources
VTS data	High coverage, low availability	AIS trajectory data supplement
Ship archives	Static information, rich semantic information	Reference source for ship behavior semantics
Port archives	Static information, rich semantic information	Sources of background area rasterization coding
Hydro meteorological data	Main environmental information	Background information for area rasterization and semantic annotation

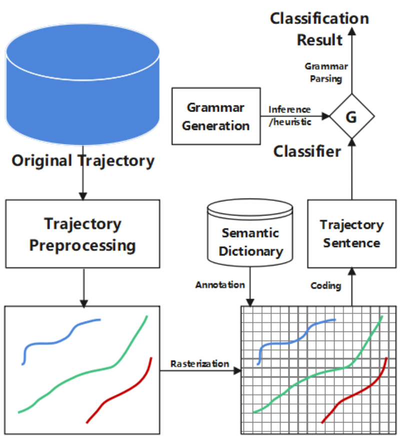


FIGURE 2. The framework of proposed model.

sources in computation, including the maritime geographic data and ship activity data, shown in Table 1. The maritime geographic data involves static data such as waterway data, meteorological data, and hydrological data, the data format is usually regular and the data changes slowly. Ship activity data includes ship archives and ship registration data. Generally, the data is heterogeneous and the data value density is low.

The framework of proposed model is shown in Fig. 2. Using maritime geographic data and domain knowledge, the background area is rasterized and semantically annotated. Then the trajectory can be expressed as a string composed of serial codes and semantic codes of the cell where moving object located at each sampling instant. The semantic codes of grid cells come from the semantic dictionary, which generated from port archive data and nautical sea chart. The serial code of grid cell contains spatial information, and the

TABLE 2. Preliminary Semantic Dictionary ***D***.

Name	Description	Category	Code
Passenger port	Port area for passenger	Port	P
Cargo dock	Port area for Cargo	Port	L
Bulk cargo dock	Port area for bulk domestic trading dock	Port	G
Berthing area	Open berthing waters	Port	T
Prescribed waterway	The waterway clearly defined by the authorities	Waterway	M
Agreed waterway	Seasonal waterway determined through negotiation	Waterway	W
Inward and outward waterway	The waterway connecting the main waterway and port	Waterway	E
Diversion channel	Secondary or diversion channel	Waterway	D
Inland waterway	Inland waterway navigation	Waterway	I
Fishing area	4 types of fishing operation areas defined by the FAO	Operation	F
Engineering operation area	Designated engineering operation area	Operation	O
Impassable	Obstacle， impassable area	Other	U
Not defined	Area with no background semantics set	Other	#

semantic code represents its semantic information. Although there are differences in format and accuracy in various trajectory data sources, their semantics and syntax are usually consistent, which are described as symbolic sentences. The grammar that characterizes the ship's behavior pattern can be obtained through syntax inference approaches or heuristic knowledge of experts. Later, grammar parsing is used to classify trajectories, or to measure the similarity of subsequence patterns in ship trajectories. The advantage of adopting this method is that the semantic grammar representing ship behavior is stable as the errors in the trajectory data vary. Although the trajectories are not exactly the same in time and space due to various measurement errors, their behavioral semantics are consistent.

IV. RASTERIZATION AND CODING

The rasterization has two steps: firstly, the whole area is divided into m^*m grid cells, and each cell gets a number according to its serial order. Secondly, using regional geographic background data, each cell is annotated with semantic information which is needed for subsequent sentences generation. As shown in formula (1):

$$H = \{(h_{ij}, g_{ij})|i = 1..m, j = 1..m, h_{ij} = 1 : m^2, g_{ij} \in D\} \quad (1)$$

Each grid cell is represented by a two-tuple, where $\boldsymbol{h_{ij}}$ is the serial code, $\boldsymbol{g_{ij}}$ is the semantic code representing its geographical background, and the code value comes from the dictionary $\boldsymbol{D}$, shown in Table 2.

In Fig. 3, a sea area is rasterized into 20*20 grid cells. Each cell has a serial number which refers to its spatial position.

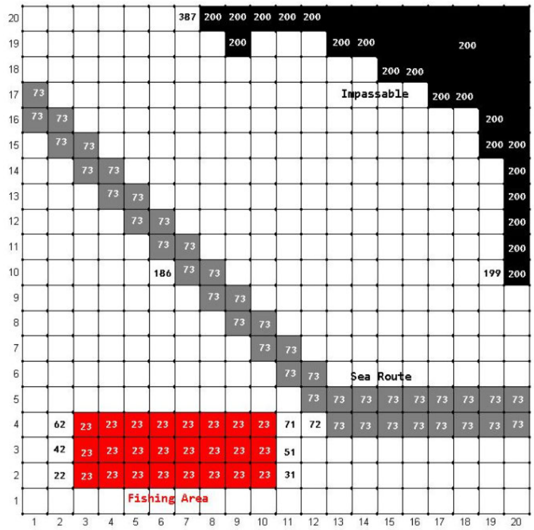


FIGURE 3. Rasterization of background area.

According to the geographic background information of the area itself, the background semantics are added to each grid cell. In order to reduce the number of effective grid cells and to improve the efficiency of classification, cells with the same semantic information can be merged according to the requirement of classification. The cells "23-30, 33-40, and 43-50" in the figure can be merged into one area, which is a fishing area, denoted by the serial number of the first cell "23". The gray area is a channel, which is represented by the code "73". In the nautical charts, the channels and various sea areas are relatively fixed, and the navigation of ships on these areas usually abide by certain fixed routines. When a ship moves in a certain channel area, its behavior has little to do with its specific location, but mainly related to the type of the area. As shown in the Fig. 3, our model assumes that the behavior semantics of ships located in all the cells in grid "73" are same, but the ship at grid "72" or group "23" has a different behavior rather than just sailing on the given channel. According to the information contained in the nautical chart, combined with common ship navigation routines, a preliminary background semantic dictionary ***D*** is established, as shown in Table 2. And the rasterization, coding and merging between grid cells on the area chart will be done before the trajectory sentence generation.

V. GENERATION OF TRAJECTORY SENTENCE STRING

The trajectory sentence string ***S*** is generated by sampling the position information of the ship trajectory on different time instant. ***S*** is composed of the grid serial codes and the its semantics codes. Given the original trajectory sequence ***T*** and the grid ***H***, the process of generating ***S*** is shown in formula (2).

$$\begin{aligned} &\mathrm{T} = \{(lat_i, long_i, t_i)\,|\mathrm{i} = 1..\mathrm{N}\} \\ &\mathrm{S} = \mathrm{s}_1\mathrm{s}_2..\mathrm{s}_i..\mathrm{s}_n, \\ &\mathrm{s}_i = h_i g_i, Inside\,(lat_i, long_i, h_i, t_i) = true \end{aligned} \tag{2}$$

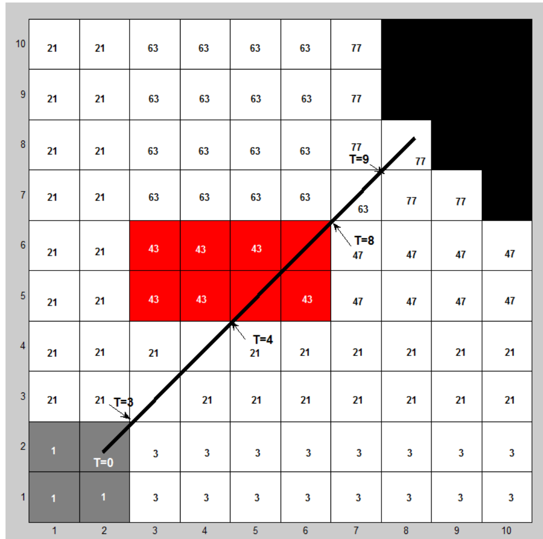


FIGURE 4. Trajectory sentence string generation on the raster.

Sequence ***T*** is ***N*** trajectory points sampled from the original trajectory, and each $\boldsymbol{h_i g_i}$ is the grid cell serial code and the semantic code where the trajectory point is located at the sampling time. ***Inside*** is the topological operation, which calculates the inclusion relationship between the point($\boldsymbol{lat_i, long_i}$) and the grid cell $\boldsymbol{h_i}$ at the time instant $\boldsymbol{t_i}$.

Fig. 4 shows a trajectory passed through multiple grid cells from the start point to the end point. The sampling number $N = 10$. The grid serial code where each sampling point located and its semantic code makes up the result sentence string:"***1A1A1A21M43W43W43W43W63E77P***". The symbols ***A***, ***M***, ***W***, ***E*** and ***P*** come from the dictionary ***D*** in Table 2.

It is beneficial for the algorithm to use the same samples number ***N*** for all trajectories. However, different ships have different tasks, attributes and navigation capabilities. It is not appropriate to equate the trajectories of ocean-going ships with the trajectories of small ships operating offshore. We only discuss the situation when the trajectories are in the same scale here. The preprocessing of trajectory data of different ship types and different trajectory scales can be carried out using ship archive data or other algorithms, which will not be discussed in depth here.

VI. CLASSIFICATION USING GRAMMAR PARSING

For each trajectory class, a context-free grammar ***G*** is used to establish its behavior pattern features, that is:

$$\begin{aligned} &\mathrm{G} = \{\mathrm{V_t}, \mathrm{V_n}, \mathrm{P}, \ \mathrm{S}\}: \\ &\text{S is the start symbol grammar,} \\ &V_t = \{h_i, g_i \,|\, \forall h_i \in H, \forall g_i \in D\}, \\ &V_n \text{ is the nonterminal symbol set,} \\ &\mathrm{P} = \{\mathrm{p}_i\} \text{ is the Production Rule Set.} \end{aligned} \tag{3}$$

In grammar ***G***, ***S*** is the start symbol, and the terminal symbol set $\boldsymbol{V_t}$ refers to the basic acceptable symbols that make up a sentence. The symbols in $\boldsymbol{V_t}$ include the valid serial code of grid cell $\boldsymbol{h_i}$ and its corresponding grid background semantic

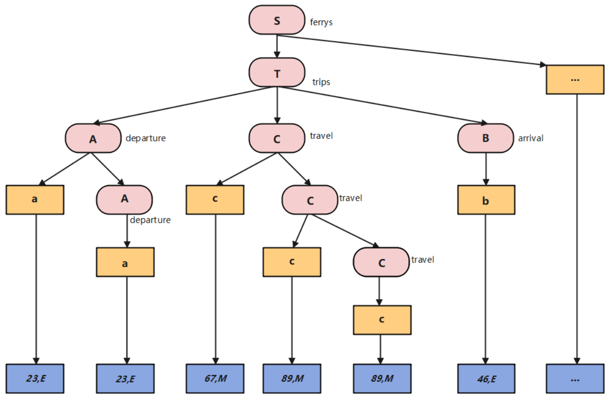


FIGURE 5. The grammar spanning tree of the class "'Ferries".

FIGURE 6. Selection area of the experiments.

TABLE 3. Example Records in Trajectory Data.

MMSI	BaseDateTime	LAT	LON
412457680	2015-07-01T12:00:05	24.2715	118.0392
477925800	2015-07-01T12:00:04	24.2485	118.0275
413046020	2015-07-01T12:00:04	24.2544	118.0518
412509645	2015-07-01T12:00:04	24.2487	118.0201

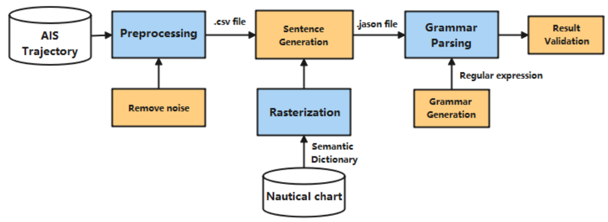


FIGURE 7. The experiment flowchart.

code g_i. $\boldsymbol{P}$ is the production rule set. The nonterminal symbol set V_n is the grammatical variables and intermediate symbols generated in the production process. When a trajectory $\boldsymbol{x}$ is to be identified, if $\boldsymbol{x}$ can be accepted by grammar $\boldsymbol{G}$, that is: $\boldsymbol{x \in L(G)}$, then $\boldsymbol{x}$ will be accepted by the classifier.

For example, the production rule set of the trajectory class "ferries" can be written as:

$$
\begin{aligned}
&\mathrm{p1: S \to T;\ p2: T \to tT;\ p3: T \to t;\ p4: T \to ACB;}\\
&\quad \mathrm{p5: A \to aA;\ p6: A \to a;}\\
&\quad \mathrm{p7: B \to bB;\ p8: B \to b;\ p9: C \to cC;\ p10: C \to c;}\\
&\quad \mathrm{a, b \in \{E\},\ c \in \{M, F, W\},\ t \in \{P\}}
\end{aligned}
$$

The grammar spanning tree is shown in Fig. 5:

According to this grammar pattern, if there is a string $\boldsymbol{x} =$ ***"23E23E67M89M89M46E"***, it can be accepted as a ferry.

The grammar production rule set $\boldsymbol{P}$ is a context-free grammar. The inference of linear grammar and simple deterministic grammar can be completed in polynomial time, whereas the inference of natural language grammar is proved to be a NP problem. It is a heavy task to infer the grammar from the historical trajectory data directly, which require a new article to discuss. In this article, we use expert knowledge to manually define the grammar of each pattern.

VII. EXPERIMENTS

The experiments use the real AIS data near XiaMen Port in July 2015, between the longitude ***117°57E*** to ***118°10E***, latitude ***24°30N*** to ***24°21N,*** on the east side of Shima Port in Xiamen. Fig. 6 shows a schematic representation of this area.

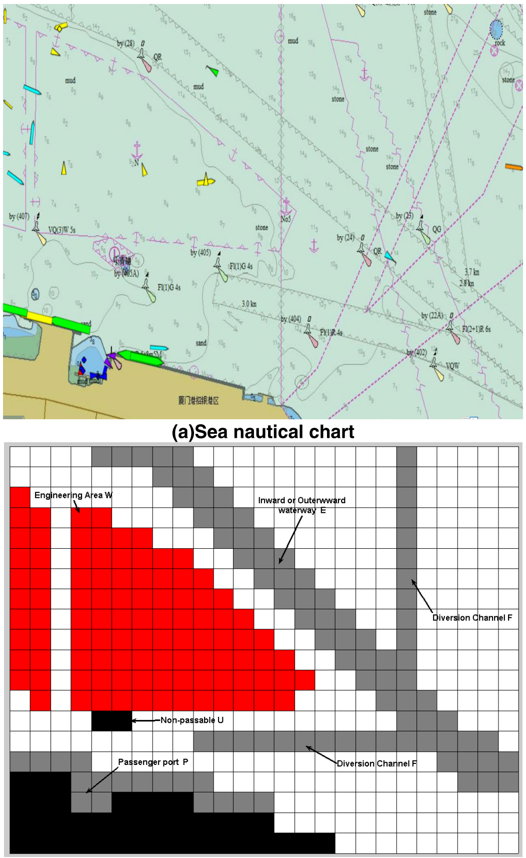
The trajectory data set comes from the website http://www.enclive.cn/Product/AISShipData.html, the data are not original AIS messages, but decoded trajectory points data. Only the MMSI number, timestamp, latitude and longitude fields are chosen, Table 3 shows a few records as examples.

The experiment flowchart is shown in Fig. 7. The trajectory data is preprocessed in a simple way, we first sort the data records by the MMSI of each object, and discard the duplicate records in it. If there exist blank fields in a record, we simply discard it. A 5-days ship AIS trajectory data set in the area was preprocessed for the experiments. They are stored in five CSV files, each file contains around 350 trajectories of various types of ships in one day, denoted as $\{\boldsymbol{D_1}, \boldsymbol{D_2}, \boldsymbol{D_3}, \boldsymbol{D_4}, \boldsymbol{D_5}\}$.

In the rasterization stage, the selected area is rasterized into 50*50 grid cells. According to the background data in the nautical chart, semantics codes denoting entry and exit channels, diversion channels, shoals, impassable areas, etc. are added to each grid cell, as shown in Fig. 8.

Not all grid cells have background semantics, there are also cells with undefined semantics between regions, denoted as the semantic code "***#***". The selected area contains 42 areas of various types. We sampled the AIS trajectory with a 24-hour time span and take $\boldsymbol{T}$ as the sampling time interval to generate semantic trajectory sentence strings. The sentence strings are stored in JASON format files.

We established three trajectory grammars, corresponding to three classes such as ferries, cargo ships and passing ships. The parsing approach we used here is the regular expressions match method provided in ***Javascripts***, and Table 4 shows the grammar pattern for three classes. Experiments are done to verify the practicability and to simulate the impact of different sampling interval $\boldsymbol{T}$ on the model correctness.



(a)Sea nautical chart

(b) Rasterization result

FIGURE 8. Rasterization of part of selected sea area.

TABLE 4. The Grammar for 3 Classes.

Ferry boat c_1	p1: S → T; p2: T → tT; p3: T → t; p4: T → ACB; p5: A → aA; p6: A → a; p7: B → bB; p8: B → b; p9: C → cC; p10: C → c; a, b ∈ {E}, c ∈ {M, F, W}, t ∈ {P}
Cargo ship c_2	p1: S → AC; p2: S → CA; p3: A → aA; p4: A → a; p5: C → cC; p6: C → c; a ∈ {L, G}, c ∈ {M, F, W}
Passing ship c_3	p1: S → dTd; p2: T → ACB; p3: T → ACBT; p4: A → aA; p5: A → a; p6: B → bB; p7: B → b; p8: C → cC; p9: C → c; a, b ∈ {P, L, G, T}, c ∈ {M, F, W, #}, d ∈ {P}

Table 5 shows the experimental results of the classification correctness of three trajectory classes.

In our experiments, due to the relatively fixed and simple behavior pattern, ferries C_1 can be recognized at highest accuracy. However, for cargo ships C_2, because the selected background area is small and the semantic information is poor, the distinction between passing ships and cargo ships leaving or docking is limited. The semantic of its berthed region affect the results. For the same reason, as well as the relatively free style behavior pattern, a certain amount of misjudgment has occurred on passing ships C_3, but the result is generally acceptable.

TABLE 5. The Result of Trajectory Classification.

	C_1 (Ferry)	C_2 (Cargo)	C_3(Passing)
D_1(2015-07-01)	94.3%	74.4%	56.3%
D_2(2015-07-02)	91.4%	75.7%	54.2%
D_3(2015-07-03)	97.1%	69.2%	57.2%
D_4(2015-07-04)	88.6%	71.8%	63.6%
D_5(2015-07-05)	97.1%	76.9%	58.6%

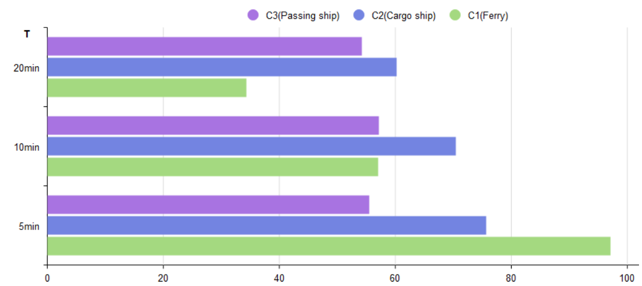


FIGURE 9. The classification accuracy under sampling interval T (5 minutes, 10 minutes, 20 minutes).

Through our experiments, it is found that the sampling interval T has a large impact on the classification accuracy. When the sampling interval is too large, the semantics of the trajectory will be lost, resulting in classification errors. When generating trajectory sentence strings, selecting a suitable time granularity for sampling is an important factor. Within a suitable granularity range, the classification algorithm is not sensitive to the position error and time error of the original trajectory data, but once the granularity chosen is bad, the classification accuracy will deteriorate rapidly. Furthermore, the grammar describing trajectory pattern should be adjusted according to the change of granularity too. The low classification accuracy is due to the lost of class features as the time granularity changing.

In order to improve the classification accuracy, more sophisticated data preprocessing technology must be adopted. Our experiments present here are to verify if the algorithm can classify trajectories properly, not how efficient it is. The data set we used is relatively small and comparison on the efficiency with other baseline algorithms need to be done in the future.

VIII. CONCLUSION

This article presents a trajectory classification model based on grammar parsing. It converts trajectories into sentences composed of symbolic codes from rasterization and semantic annotations. The class features are expressed as context-free grammars and the data classification is implemented through syntax parsing. Experiments were done and results verified the model effectiveness.

However, the following aspects need further studying:

A. ADAPTABILITY TO TIME AND SPACE GRANULARITY

The algorithm shows large performance differences under different sampling time intervals, reflecting its sensitivity to the time granularity. Similarly, it can be found that the algorithm will also produce performance differences under different spatial granularities. The model proposed does not have the ability to adapt different granularities. It should use heuristic information in preprocessing, or adopt certain multi-scale adaptive processing technology.

B. GRAMMAR TRAINING AND AUTOMATIC INFERENCE

The trajectory grammar given in the article is pre-defined, and it is a difficult task to infer the grammar through training data, we can't implement automatic reasoning algorithm yet. However, different from the grammar of natural language, the grammar of ship behavior pattern is relatively simple. There are already some feasible learning and training algorithms for simple deterministic grammar and linear grammar. Generation of the classifier from a large amount of training data should be a direction for future works.

C. IMPLEMENTATION ON BIG DATA ARCHITECTURE

AIS data is a source of big data. After selecting the appropriate analysis granularity, the corresponding parallel computing architecture and stream processing architecture should be adopted for the design of future classification algorithms.

REFERENCES

[1] W. Yang, T. Ai, and W. Lu, "A method for extracting road boundary information from crowdsourcing vehicle GPS trajectories," *Sensors*, vol. 18, no. 4, pp. 1261–1270, 2018.

[2] J. A. Long, "Kinematic interpolation of movement data," *Int. J. Geograph. Inf. Sci.*, vol. 30, no. 5, pp. 854–868, 2016.

[3] R. W. Liu, J. Nie, S. Garg, Z. Xiong, Y. Zhang, and M. S. Hossain, "Data-driven trajectory quality improvement for promoting intelligent vessel traffic services in 6G-enabled maritime IoT systems," *IEEE Internet Things J.*, early access, Oct. 5, 2020, doi: 10.1109/JIOT.2020.3028743.

[4] L. Bermingham and I. Lee, "A framework of spatio-temporal trajectory simplification methods," *Int. J. Geograph. Inf. Sci.*, vol. 31, no. 6, pp. 1128–1153, 2017.

[5] Y. Huang, Y. Li, Z. Zhang, and R. W. Liu, "GPU-accelerated compression and visualization of large-scale vessel trajectories in maritime IoT industries," *IEEE Internet Things J.*, vol. 7, no. 11, pp. 10794–10812, Nov. 2020, doi: 10.1109/JIOT.2020.2989398.

[6] F. Biljecki, H. Ledoux, and P. Van Oosterom, "Transportation mode-based segmentation and classification of movement trajectories," *Int. J. Geograph. Inf. Sci.*, vol. 27, no. 2, pp. 385–407, 2013.

[7] S. Ji, Z. Wang, T. Li, and Y. Zheng, "Spatio-temporal feature fusion for dynamic taxi route recommendation via deep reinforcement learning," *Knowl.-Based Syst.*, vol. 205, Oct. 2020, Art. no. 106302.

[8] Z. Ye, K. Xiao, Y. Ge, and Y. Deng, "Applying simulated annealing and parallel computing to the mobile sequential recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 2, pp. 243–256, Feb. 2019.

[9] Z. Ye, K. Xiao, and Y. Deng, "A unified theory of the mobile sequential recommendation problem," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2018, pp. 1380–1385.

[10] K. Xiao, Z. Ye, L. Zhang, W. Zhou, Y. Ge, and Y. Deng, "Multi-user mobile sequential recommendation for route optimization," *ACM Trans. Knowl. Discovery from Data*, vol. 14, no. 5, pp. 1–28, Aug. 2020.

[11] Z. Chen, H. T. Shen, X. Zhou, Y. Zheng, and X. Xie, "Searching trajectories by locations: An efficiency study," in *Proc. Int. Conf. Manage. Data (SIGMOD)*, 2010, pp. 255–266.

[12] H. Wang and C. O'Sullivan, "Globally continuous and non-Markovian crowd activity analysis from videos," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 527–544.

[13] T. Yao, Z. Wang, Z. Xie, J. Gao, and D. D. Feng, "Learning universal multiview dictionary for human action recognition," *Pattern Recognit.*, vol. 64, pp. 236–244, Apr. 2017.

[14] L. Wang and M. Dong, "Detection of abnormal human behavior using a matrix approximation-based approach," in *Proc. 13th Int. Conf. Mach. Learn. Appl.*, Dec. 2014, pp. 324–329.

[15] W. Zhao, Z. Zhang, and K. Huang, "Gestalt laws-based tracklets analysis for human crowd understanding," *Pattern Recognit.*, vol. 75, pp. 112–127, Mar. 2018.

[16] S. Atev, O. Masoud, and N. Papanikolopoulos, "Learning traffic patterns at intersections by spectral clustering of motion trajectories," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2006, pp. 4851–4856.

[17] S. Gurung, D. Lin, W. Jiang, A. Hurson, and R. Zhang, "Traffic information publication with privacy preservation," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, pp. 44–59, 2014.

[18] X. Li, J. Han, J.-G. Lee, and H. Gonzalez, "Traffic density-based discovery of hot routes in road networks," in *Proc. Int. Symp. Spatial Temporal Databases*. Berlin, Germany: Springer, 2007, pp. 441–459.

[19] M. Nanni and D. Pedreschi, "Time-focused clustering of trajectories of moving objects," *J. Intell. Inf. Syst.*, vol. 27, no. 3, pp. 267–289, 2006.

[20] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy-rate clustering: Cluster analysis via maximizing a submodular function subject to a mastoid constraint," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 99–112, Jan. 2014.

[21] X. Li, W. Hu, and W. Hu, "A coarse-to-fine strategy for vehicle motion trajectory clustering," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, 2006, pp. 591–594.

[22] J. A. Long and T. A. Nelson, "A review of quantitative methods for movement data," *Int. J. Geograph. Inf. Sci.*, vol. 27, no. 2, pp. 292–318, Feb. 2013.

[23] Z. Shao and Y. Li, "On integral invariants for effective 3D motion trajectory matching and recognition," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 511–523, Feb. 2016.

[24] R. Saini *et al.*, "Modeling local and global behavior for trajectory classification using graph based algorithm," *Pattern Recognit. Lett.*, to be published, doi: 10.1016/j.patrec.2019.05.014.

[25] L. Huanhuan, J. Liu, Z. Yang, R. W. Liu, K. Wu, and Y. Wan "Adaptively constrained dynamic time warping for time series classification and clustering," *Inf. Sci.*, vol. 534, pp. 97–116, Sep. 2020.

[26] M. Vlachos, M. Hadjieleftheriou, D. Gunopulos, and E. Keogh, "Indexing multidimensional time-series," *VLDB J.*, vol. 15, no. 1, pp. 1–20, Jan. 2006.

[27] E. Tiakas, A. Papadopoulos, A. Nanopoulos, Y. Manolopoulos, D. Stojanovic, and S. Djordjevic-Kajan, "Trajectory similarity search in spatial networks," in *Proc. 10th Int. Database Eng. Appl. Symp. (IDEAS)*, Dec. 2006, pp. 185–192.

[28] Y. Zheng, "Trajectory data mining: An overview," *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 3, pp. 29–41, 2015.

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

[30] H. Xu, Y. Zhou, W. Lin, and H. Zha, "Unsupervised trajectory clustering via adaptive multi-kernel-based shrinkage," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4328–4336.

[31] K. Zheng, Y. Zheng, N. J. Yuan, and S. Shang, "On discovery of gathering patterns from trajectories," in *Proc. IEEE 29th Int. Conf. Data Eng. (ICDE)*, Apr. 2013, pp. 242–253.

[32] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, "Mining interesting locations and travel sequences from GPS trajectories," in *Proc. 18th Int. Conf. World Wide Web (WWW)*, 2009, pp. 791–800.

[33] F. Solera, S. Calderara, and R. Cucchiara, "Socially constrained structural learning for groups detection in crowd," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 5, pp. 995–1008, Aug. 2016.

[34] M. A. Bautista, A. Hernández-Vela, S. Escalera, L. Igual, O. Pujol, J. Moya, V. Violant, and T. María Anguera, "A gesture recognition system for detecting behavioral patterns of ADHD," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 136–147, Jan. 2016.

[35] M. Devanne, S. Berretti, P. Pala, H. Wannous, M. Daoudi, and A. D. Bimbo, "Motion segment decomposition of RGB-D sequences for human behavior understanding," *Pattern Recognit.*, vol. 61, pp. 222–233, Jan. 2017.

[36] J. Yin, X. Y. Chai, and Q. Yang, "High-level goal recognition in a wireless LAN," in *Proc. 19th Nat. Conf. Artif. Intell.*, Palo Alto, CA, USA, 2004, pp. 578–584.
[37] C. Liu, Y. Ge, H. Xiong, K. Xiao, W. Geng, and M. Perkins, "Proactive workflow modeling by stochastic processes with application to healthcare operation and management," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2014, pp. 1593–1602.
[38] M. Hasan and A. K. Roy-Chowdhury, "A continuous learning framework for activity recognition using deep hybrid feature models," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 1909–1922, Nov. 2015.
[39] M. Boukhechba, A. Bouzouane, B. Bouchard, C. Gouin-Vallerand, and S. Giroux, "Online recognition of people's activities from raw GPS data: Semantic trajectory data analysis," in *Proc. 8th ACM Int. Conf. Pervasive Technol. Rel. Assistive Environ.*, Greece, 2015, pp. 11–18.
[40] C. Liu and C. Guo, "STCCD: Semantic trajectory clustering based on community detection in networks," *Expert Syst. Appl.*, vol. 162, Dec. 2020, Art. no. 113689.
[41] X. Wu, Z.-K. Tang, Y.-Y. Zhu, Y.-W. Peng, and Z.-Y. Peng, "Frequent pattern mining with approximate arrival-time in semantic trajectories," *J. Softw.*, vol. 29, no. 10, pp. 3184–3204, 2018, doi: 10.13328/j.cnki.jos.005418.
[42] Y. Cao, F. Xue, Y. Chi, Z. Ding, L. Guo, Z. Cai, and H. Tang, "Effective spatio-temporal semantic trajectory generation for similar pattern group identification," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 2, pp. 287–300, Feb. 2020.
[43] L. Xinyuan, "Semantic trajectory extraction framework for indoor space," *J. Tsinghua Univ., Sci. Technol.*, vol. 59, no. 3, pp. 186–193, 2019.
[44] S. Spaccapietra, C. Parent, and M. L. Damiani, "A conceptual view on trajectories," *Data Knowl. Eng.*, vol. 65, no. 1, pp. 126–146, 2008.
[45] D. Mountain and J. Raper, "Modeling human spatio-temporal behavior: A challenge for location- based services," in *Proc. 6th Int. Conf. GeoComputation*. Australia: Springer, 2001, pp. 231–237.
[46] J. Zhang, H. Shi, and L. Cui, "Location prediction model based on transportation mode and semantic trajectory," *J. Comput. Res. Develop.*, vol. 56, no. 7, pp. 1357–1369, 2019.
[47] A. Boulmakoul, L. Karim, A. Elbouziri, and A. Lbath, "A system architecture for heterogeneous moving-object trajectory metamodel using generic sensors: Tracking airport security case study," *IEEE Syst. J.*, vol. 9, no. 1, pp. 283–291, Mar. 2015.
[48] M. Manaa and J. Akaichi, "Ontology-based modeling and querying of trajectory data," *Data Knowl. Eng.*, vol. 111, pp. 58–72, Sep. 2017.
[49] M. Baglioni, J. A. F. de Macêdo, C. Renso, R. Trasarti, and M. Wachowicz, "Towards semantic interpretation of movement behavior," in *Proc. 12th AGILE Conf. Adv. GIScience*. Hannover, Germany: Springer, 2009, pp. 271–288.

LEI BAO received the B.S. degree in computer science from Southeast University, Nanjing, China, in 1995, and the Ph.D. degree in computer science from the Nanjing University of Aeronautics and Astronautics, in 2006. His current research interests include spatiotemporal data mining and crowded-source data analysis on GIS systems.