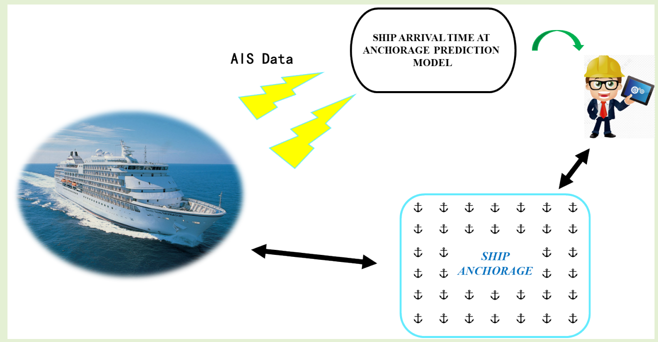
Research of AIS Data-Driven Ship Arrival Time at Anchorage Prediction

Meng Guan, Yu Cao, and Xu Cheng

***Abstract*—In today's time, maritime transport is becoming the mainstream. Ports have the problems of inefficient berthing and unreasonable allocation of shore and bridge resources. In this article, we propose a model to predict the ship arrival time based on automatic identification system (AIS) data and trajectory inflexion points to predict the anchor arrival time, aiming to solve the problems of low berthing efficiency and unreasonable allocation of shore and bridge resources. Experiments show that the minimum prediction error of the model is 8 min, the maximum error is 1 h, and the average error is 30 min; compared with the ship schedule data, the maximum error is seven days, the minimum error is one day, and the average error is 2.75 days, so the time got from this model is better than the ship schedule, which can effectively improve the berthing efficiency of the port and the reasonable allocation of shore and bridge resources. The model has good accuracy and effectiveness.**



***Index Terms*—Automatic identification system (AIS), big data analysis, designated berth efficiency, MeanShift clustering, shore bridge assignment, time to anchorage.**

I. Introduction

ALONG with the development of maritime transportation, the number of ships at sea is growing rapidly. The efficiency of the designated berth and reasonable allocation of shore and bridge resources are facing challenges. The general berthing rule is that a ship needs to apply to the port for berthing when it arrives near the port, and the port will give berthing instructions to the ship according to the availability of berths in the port; therefore, the arrival time of the ship at the anchorage is an important basis for the port berthing system and shore bridge resource allocation. At present, the port mainly relies on the time provided by the schedule to make the decision of berthing plan and shore and bridge allocation plan; however, the existence of uncertainties in practice leads to early or late arrival of ships, which makes the port unable to execute the plan as planned most of the time, reduces the designated berth efficiency and causes waste of shore and bridge resources. In order to reduce the uncertainty of ship arrival time at anchorage, it is urgent to use automatic identification system (AIS) information to predict the arrival time of ships at anchorage effectively to provide auxiliary decisions for ports.

At present, MeanShift clustering is mainly used in the field of image segmentation and big data analysis, Han and Wang [1] selected three categories of images based on MeanShift clustering for segmentation based on simple target and background, more obvious target region but the complex background, and multiple target instances with a complex background, and compared with common image segmentation algorithms, the results show that the segmentation effect of MeanShift clustering algorithm is better for all three categories of images; Li et al. [2] conducted a sensitivity study based on MeanShift algorithm for peak and valley tariff policies, dividing industries into multiple categories, and the difference in the acceptance of tariff adjustment by each industry can be well reflected by tariff sensitivity; Shen and Li [3] used MeanShift algorithm for WiFi fingerprint database clustering process, which effectively reduces the number of operations for WiFi fingerprint positioning. Sheng [4] provided a detailed introduction to the common principles of kernel functions and proposed more applicable kernel functions. Weights were assigned to each sample within a bandwidth,

Manuscript received 28 January 2024; accepted 23 February 2024. Date of publication 18 March 2024; date of current version 16 April 2024. This work was supported in part by the Applied Basic Research Program of Liaoning Province under Grant 2022JH2/101300272, in part by the Applied Basic Research Program of Liaoning Province under Grant 2023JH26/10300021, in part by the Applied Basic Research Program of Liaoning Province under Grant 2023JH26/10300013, in part by the Basic Scientific Research Project of Liaoning Provincial Department of Education under Grant LJKMZ20220754, and in part by the Basic Scientific Research Project of Liaoning Provincial Department of Education under Grant JYTMS20231450. The associate editor coordinating the review of this article and approving it for publication was Dr. Hassen Fourati. *(Corresponding author: Yu Cao.)*

Meng Guan and Yu Cao are with the School of Information Science and Control Engineering, Liaoning Petrochemical University, Fushun 113005, China (e-mail: gm_2072016@163.com; caoyu@lnpu.edu.cn).

Xu Cheng is with the School of Economics and Management, Shenyang Agricultural University, Shenyang 110866, China (e-mail: xucheng@syau.edu.cn).

Digital Object Identifier 10.1109/JSEN.2024.3370605

such that as the distance between the sample and the shifted point increased, the contribution of its shift to the mean shift vector also changed accordingly. The study conducted extensive experiments on the MeanShift algorithm in image segmentation, thoroughly validating its outstanding performance in segmentation tasks. These experiments demonstrated that the algorithm could adapt more flexibly when dealing with samples at different distances, thus providing strong support for applications in the field of image segmentation. In this article, based on previous research, we propose to apply MeanShift clustering to the clustering analysis of the ship's historical trajectory points so as to find the coordinates of the inflexion point where the ship arrives at the corresponding anchorage.

Previous studies on the prediction of arrival time at anchorage are still few, and most of them mainly focus on the study of arrival time prediction. For example, Pani et al. [5] used logistic regression, classification regression tree and random forest algorithm to predict the phenomenon of the early or late arrival of ships, respectively; Fancello et al. [6] used dynamic learning algorithm based on a neural network to predict the arrival time of ships; Xiao et al. [7] and Xiao [8] used a trajectory clustering algorithm to analyze AIS information in order to obtain the typical motion trajectory of ships; Tang et al. [9] used the core trajectory subsegment extraction algorithm based on DBSCAN algorithm to extract the core trajectory subsegments to form the typical motion trajectory of the ship; based on the position and speed information of the typical motion trajectory subsegments, the prearrival time of the ship was predicted. The authors find that most of the training model data used by previous authors are obtained based on the estimated time of arrival (ETA) of AIS data, and the data error rate is high, which will have some influence on the accuracy of the model. In this article, based on the previous research results, we propose the ship navigation inflexion point to determine the anchorage reached by the ship, and then combine the average speed of the ship to predict the arrival time at the anchorage by the method of mechanism analysis.

In this article, the ship arrival time prediction model is introduced in detail, and simulation experiments are conducted in the latter part of the article. First, the relevant theories involved in the model are introduced in Section II; second, the ship arrival time at anchorage prediction model is introduced in detail in Section III; third, simulation experiments are conducted to verify the model in Section IV; finally, a conclusion and outlook are given in Section V.

II. Theory Introduction

A. Related Definitions

Trajectory Point: The ship's trajectory point represented by p_i is a 4-D trajectory data, and the trajectory point is defined as

$$p_i = (x_i, y_i, t_i, v_i), \quad i = 1, 2, \ldots, n \tag{1}$$

where, x_i represents the longitude of the ship's trajectory point, y_i represents the latitude of the ship's trajectory point, t_i represents the time of the ship's trajectory point, and v_i represents the ship's speed to ground (SOG).

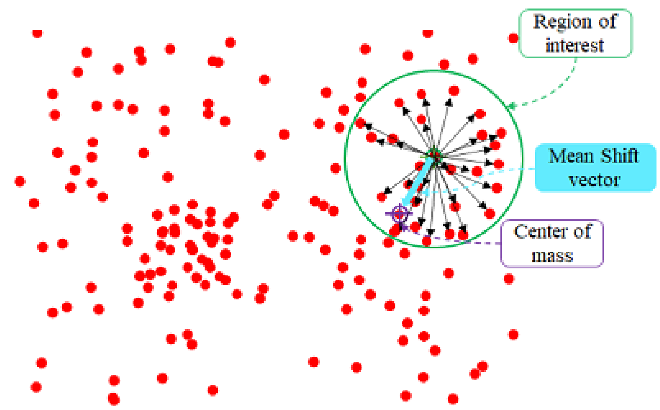


Fig. 1. Meanshift clustering schematic.

Trajectory: Trajectory is a point set consisting of several trajectory points that change continuously in time sequence, defined as

$$p = \{p_1, p_2, \ldots, p_n\}, \quad n \in N \tag{2}$$

where p represents a ship trajectory and n represents the number of trajectory points in this trajectory.

B. MeanShift Clustering Algorithm

MeanShift is composed of Mean and Shift; that is, there is a point, surrounded by many points and we calculate the sum of the offsets needed to move to each point and average them to get the Mean Shift. This offset contains the size and direction, and the direction is the direction of the dense distribution around. Then the points are moved in the direction of the average offset, and then, the endpoint is used as the new starting point, iterating until the offset Shift converges. As shown in Fig. 1. x represents the centroid, x_i represents the data points of clustering (dots), and the thick arrow is the average offset vector Shift, which is calculated by the formula (3), with the circle as the qualification

$$M_h = \sum_{x_i S_k} (x_i - x)/K, \quad i = 1, 2, \ldots, n \tag{3}$$

where S_k is a high-dimensional region of radius h, and k denotes the region in which k of these n sample points x_i fall into S_k, defined as follows:

$$S_h(x) = \{y : (y - x_i)_T (y - x_x i) < h^2\}, \quad i = 1, \ldots, n. \tag{4}$$

In this article, the kernel function (Gaussian kernel) is added to the Shift vector, and the formula is (5), and the Shift vector formula becomes (6)

$$K\{(X_1 - X_2)/(h)\} = e^{-(X_1 - X_2)^2/2h^2} / (h * \sqrt{2\pi}) \tag{5}$$

$$f_{h,K(x)} = c_{k,d} \cdot \sum_{i=1}^{n} k(\|(x - x_i)/h\|^2 / nh^d. \tag{6}$$

Next, make another circle with the endpoint of the Shift vector as the center, as shown in the figure, and repeat the above steps; finally, the Shift offset can converge to the place with the largest density in the distribution of points, as shown in Figs. 2 and 3.

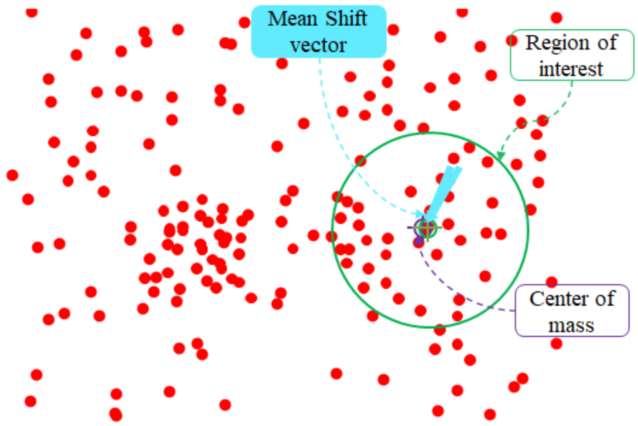


Fig. 2. Shift vector iterative shift process diagram.

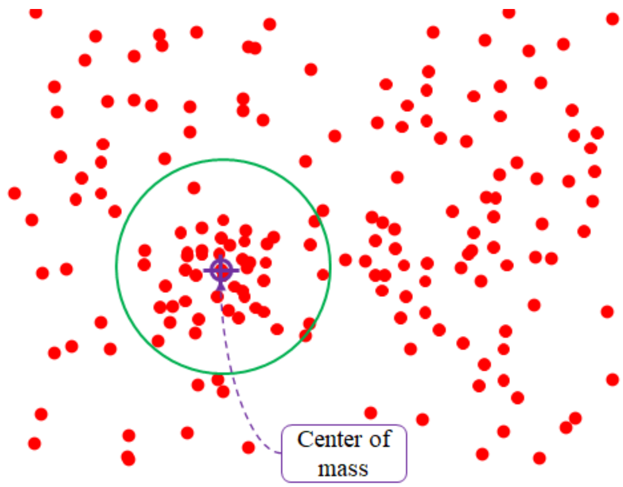


Fig. 3. Meanshift clustering results graph.

III. Ship Arrival Time at Anchorage Prediction Model

AIS system contains two parts: a shipboard system and a shore-based system. The shipboard system packages navigation information and transmits it outward by broadcasting; it also receives shore-based information and navigation information of all ships within the coverage area of the communication network; the shore-based system is the receiving and transmitting device of the AIS communication network, which receives real-time dynamic and static information of all ships within the communication area and sends commands from the base station to all ships within the control area.

AIS information has a fixed format, according to the international ITU-R standard, mainly including the information broadcasted by the ship to the shore base and other ships and the information sent by the shore base to the ship. Its fields include static data such as maritime mobile service identification (MMSI), International Maritime Organization (IMO) number, call sign, ship name, ship's length and breadth, call sign, ship's operation type, GPS antenna position, etc.; dynamic data fields include update time, ship's latitude and longitude position, SOG, etc.

Fig. 4 illustrates the overall framework of the ship arrival time prediction model at the anchorage. This model consists

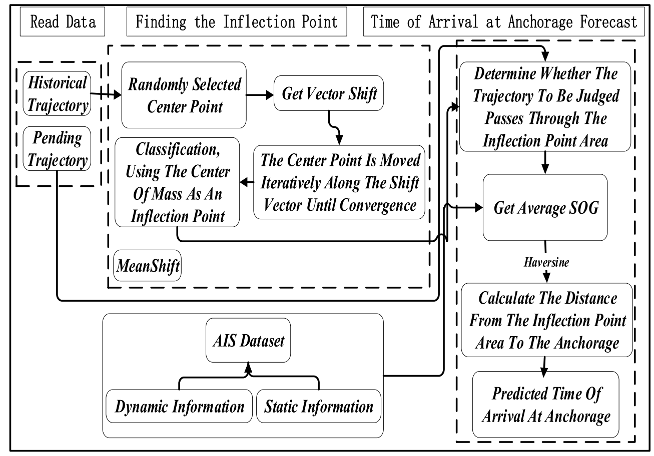


Fig. 4. Vessel arrival time at anchorage prediction model framework diagram.

of two main components: the model layer and the data layer, providing a comprehensive and efficient solution for accurate arrival time prediction at the anchorage. The following will delve into each component of the model to better understand its working principles and superior performance. First, the model layer comprises three main steps: data retrieval, turning point detection, and prediction of the arrival time at the anchorage.

In the data retrieval phase, real-time ship trajectory data is obtained from the data layer, providing necessary input for subsequent analysis. Next, by applying the MeanShift clustering algorithm, turning points in the trajectory can be accurately and adaptively identified during the turning point detection process. The MeanShift clustering algorithm, with its unique advantage in density estimation, can pinpoint potential turning points in the trajectory data, providing crucial time nodes for subsequent predictions. After obtaining the turning point region, the prediction of ship arrival time at the anchorage is carried out through mechanistic analysis. This step involves a thorough analysis of turning points and their nearby trajectories. By considering the features of historical trajectory data and turning point moments, the model can more accurately infer the time at which a ship is expected to arrive at the anchorage.

A. Dataset Description

The dataset used in this study was collected and established by the authors through the AIS Big Data platform, and noisy data, such as MMSI code anomalies, have been avoided in the collection process, so too many data preprocessing operations are no longer needed, and only the fields used in this study (TIME, MMSI, LON, LAT, SOG, and destination DES) need to be extracted. Table I shows some datasets used in this study.

B. Find the Inflexion Point Region Based on Meanshift Clustering Algorithm

The flowchart of the MeanShift clustering algorithm is shown in Fig. 5. Combining this chart and the steps of the algorithm, the historical trajectory points of ships heading to each port are clustered in order to find the inflexion point area.

TABLE I
DATA TABLES COLLECTED THROUGH THE AIS DATA PLATFORM (PARTIAL)

TIME	MMSI	LON	LAT	SOG	DES
2023-02-28 15:30:10	412208880	122°2′32″	38°31′7″	8.3	Tanker Anchorage
2023-02-28 16:00:01	412208880	121°59′41″	38°34′43″	8.6	Tanker Anchorage
2023-02-28 16:30:00	412208880	121°56′48″	38°38′27″	8.8	Tanker Anchorage
2023-02-28 17:00:00	412208880	121°53′53″	38°42′15″	8.7	Tanker Anchorage
2023-02-28 17:30:10	412208880	121°50′47″	38°45′53″	8.7	Tanker Anchorage
2023-02-28 18:00:09	412208880	121°48′4″	38°49′32″	8.6	Tanker Anchorage
2023-02-28 18:30:00	412208880	121°47′10″	38°53′33″	7.7	Tanker Anchorage
2023-02-28 18:38:27	412208880	121°46′35″	38°54′33″	8.4	Tanker Anchorage
2023-02-28 19:00:00	412208880	121°45′9″	38°57′2″	8.8	Tanker Anchorage

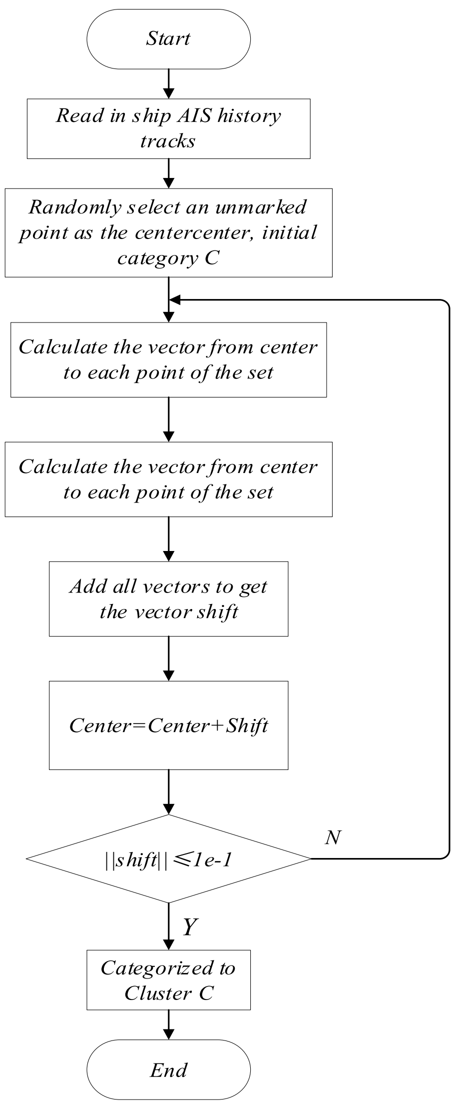


Fig. 5. Flowchart of MeanShift clustering algorithm.

1) Randomly select one of the unmarked data points as the starting centroid Center, categorized as initial category number C.
2) Using Center as the center point, calculate the vectors from Center to each element in the set, and sum these vectors to get the vector Shift.
3) Center = Center + Shift, Center moves in the direction of Shift by a distance ‖Shift‖, taking the endpoint of the vector Shift movement as the next center point Center.
4) Repeat steps 2, 3, and 4 until Shift is less than the clustering threshold, remembering that at this point Center. Note that all points encountered during this iteration should be grouped into cluster C.
5) If the distance between the Center of the current cluster C and the centers of other already existing clusters C2 is less than the distance threshold when Shift iterations converge, then merge C2 with C, and the number of data point occurrences are merged correspondingly; otherwise treat C as a new cluster.
6) Repeat 1–5 until all points have been marked as visited.
7) Classification: According to each class, the frequency of access to each point, the one with the greatest access frequency, is taken as the class to which the current point set belongs.

In this article, the clustering threshold is set to CLUSTER_THRESHOLD = 1e−1, the distance threshold is DISTANCE_THRESHOLD = 1e−4, and the Gaussian kernel is set to Gaussian_sigma = 0.15.

C. Ship Arrival Time at Anchorage Prediction Model Design

The flowchart of the model for predicting the arrival time of a ship at anchorage is shown in Fig. 6. Based on this chart and the inflexion area obtained above, the time difference for the ship to arrive at the anchorage destination from the inflexion area can be calculated to obtain the expected arrival time at the anchorage.

1) Enter the set of trajectory points of the vessel to be judged.
2) Judging whether the trajectory to be judged passes through this inflexion point area according to the range of the inflexion point area through which the trajectory of the ship arrives at each anchorage obtained in Section III-B by calculating the relationship between the spherical distance [formula (7)] between the point of the trajectory to be judged and the center of the circle of the inflexion point area, and then comparing it with the

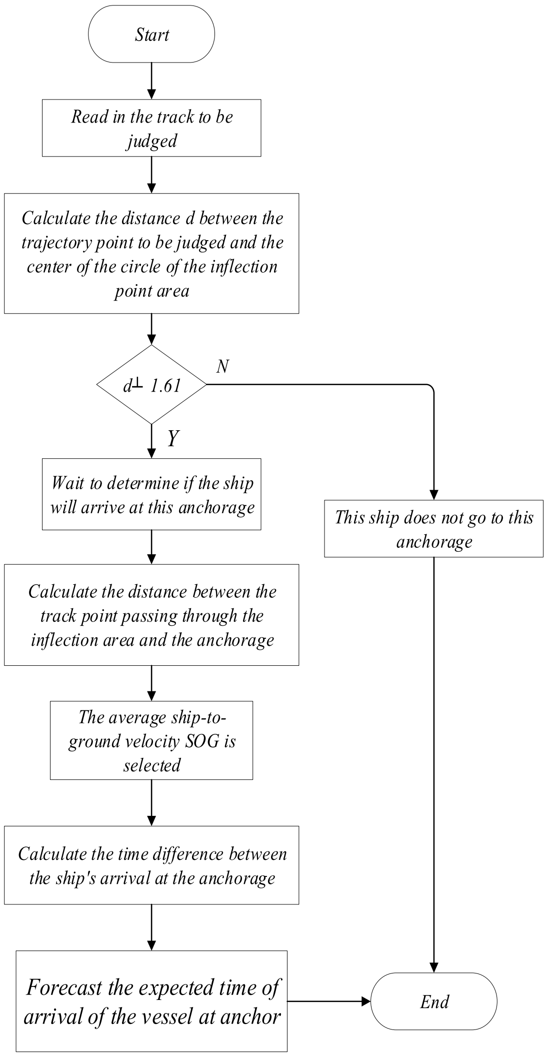


Fig. 6. Flowchart for predicting ship arrival time model.

radius, if it is within the circular area, the anchorage to which the ship arrives is judged.

3) Calculation of the spherical distance of the current trajectory point of the ship to be judged from the anchorage.
4) Since the ship's speed is basically constant or varies very little under normal conditions, this article selects the average speed of the ship's voyage as the calculation speed and calculates the time difference between the ship to be judged and the anchorage by combining the distance obtained in step 3 [formula (8)].
5) Calculate the estimated time for the ship to be judged to arrive at the anchorage (ATA), i.e., the estimated time for the ship to arrive at the anchorage is obtained by adding the update time of the current ship trajectory point to the time difference calculated in step 4 [formula (9)].

Under the principle that the inflexion point area should contain as many historical trajectory points as possible, the center of the circle of the inflexion point area selected in this article is the center of mass of clustering, and the radius is 1.61 km

TABLE II
CLUSTERING CENTER OF MASS AND RADIUS OF INFLEXION AREA

Anchorage Type	Center-of-mass coordinates	Area Radius
Cargo Ship Anchorage	(38.910,121.778)	1.61km
Tanker Anchorage	(38.912,121.768)	1.61km

$$s = \sin^2((\mathrm{lat}B - \mathrm{lat}A)/2)$$

$$c = \cos(\mathrm{lat}B)\cos(\mathrm{lat}A)\sin^2((\mathrm{lon}B - \mathrm{lon}A)/2)$$

$$d = 2r \cdot \arcsin(\sqrt{s + c}) \tag{7}$$

where r is the radius of the earth, taking the value of 6378 km, lat$A$, lat$B$ are the latitude of two trajectory points, lon$A$, lon$B$ are the longitude of two trajectory points

$$t_{\mathrm{anch}} = l_{\mathrm{anch}}/(0.5144 \cdot \overline{V}_{\mathrm{vessel}} \cdot 60) \tag{8}$$

where, t_{anch} indicates the time difference between the ship sailing from the inflexion point area to the anchorage in min, l_{anch} indicates the Hausdorff distance between the position coordinates of the ship arriving at the inflexion point area and the anchorage coordinates, and since the ship's sailing speed is basically unchanged or changes very little, the average sailing speed is used and indicated by $\overline{V}_{\mathrm{vessel}}$. The unit of the ship's speed in the AIS data is knot, so it needs to be converted to m/s, and its conversion. The relationship is 1 knot = 0.5144 m/s

$$\mathrm{ATA} = T_0 + t_{\mathrm{anch}} \tag{9}$$

where T_0 denotes the time when the ship reaches the inflexion point area.

Algorithm 1 ATA Prediction

```
Initialize: p ←[ ]; turn_p ←[ ]; anch_p ←[ ]
//Read in track points, inflexion point coordinates,
  anchorage coordinates, and threshold values
turn_p ←[(lon_1,lat_1),(lon_2,lat_2),...,(lon_n,lat_n)];
 anch_p ←[(lon_1,lat_1),...,(lon_n,lat_n)]; d←1.61
for i = 1,2,...,n do
   p_i ← (lon_i, lat_i)
end for
//Determining where the ship is at anchorage
for j = 0,1,2,...,n do
   If getdistance(p_j, turn_p) < d then
      l_anch ←getdistance(p_j, anch_p)
      V̄_vessel ← SOG̅
      t_anch ← l_anch/(0.5144 · V̄_vessel · 60)
      ATA ← T_0 + t_anch
   else
      the ship reached another anchorage
   end if
end for
```

IV. Simulation Experiments

A. Introduction of Experimental Conditions

In this article, we will apply the ship arrival time prediction model established above, conduct experiments on the AIS

TABLE III
SHIP ARRIVAL TIME PREDICTION RESULTS (PARTIAL)

TIME	MMSI	ATA	ATA_PRE	TIME_ERROR
2023-02-28 18:30:00	412208880	2023-02-28 19:30:00	2023-02-28 19:22:00	8min
2023-02-28 20:20:00	412764360	2023-02-28 21:30:00	2023-02-28 21:20:00	10min
2023-03-01 07:40:08	413213280	2023-03-01 08:30:43	2023-03-01 08:07:08	23min
2023-02-28 09:30:02	412473490	2023-02-28 10:30:26	2023-02-28 09:59:02	31min
2023-03-02 06:00:09	412355160	2023-03-02 07:00:33	2023-03-02 06:28:09	32min

TABLE IV
MODEL AND SHIP SCHEDULE SAILING TIME ERROR COMPARISON ANALYSIS TABLE

Model / Methodology	Maximum Error	Minimum Error	Average error
Model of this paper	**1h**	**8min**	**30min**
Vessel schedule	7days	1days	2.75days

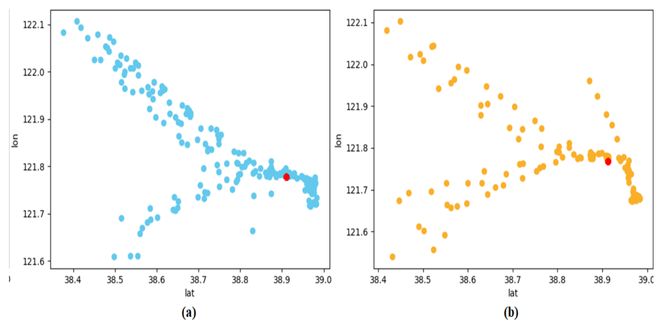


Fig. 7. Clustering results of historical ship track points at each anchorage.

dataset collected by the authors through the AIS big data platform, compare it with the ship schedule data provided by Dalian port, and then evaluate the practical application value of the model.

The experimental equipment conditions are as follows: the device CPU is R9 5950X 3.40 GHz, 16 G running memory, and the graphics card is NVIDIA RTX 3080.

B. Ship Arrival Time at Anchorage Prediction Experiment

1) Experiment on Finding Ship Trajectory Inflexion Points Based on Meanshift Clustering: The ship historical trajectories of cargo ship anchorage and tanker anchorage are clustered by the MeanShift clustering algorithm to obtain the dense area of trajectory points (inflexion points) through which the ship arrives at the anchorage and a circle of suitable radius (the circle contains as many trajectory points as possible) is drawn with the cluster center of mass in Table II as the center of the circle, as the inflexion point area leading to the corresponding anchorage, which is shown in Fig. 7. From the table, the center coordinates of the turning point area of the ship's historical trajectory toward the cargo ship anchorage are (38.910, 121.778), and the center coordinates of the turning point area of the ship's historical trajectory toward the tanker anchorage are (38.912, 121.768), and the radius of the turning point area is set to 1.61 km.

The clustering results of the historical trajectories of cargo ships heading to anchorages are shown in Fig. 7(a), and the clustering results of the historical trajectories of tankers heading to anchorages are shown in Fig. 7(b).

2) Experiment on Predicting the Arrival Time of Ships at Anchorage: In this article, after obtaining the specific inflexion point of the anchorage, the straight-line distance of the trajectory point of the ship to be predicted arriving at the inflexion point area from the destination of the anchorage is calculated, and the average value of the ship's SOG is selected as the speed for calculation. By using the straight-line distance and speed, the time difference of the ship arriving at the destination from the present position can be calculated, and thus, the predicted arrival time at anchorage is calculated by combining it with the current time. The authors used the difference between the predicted arrival time at anchorage and the actual arrival time of the ship as the model evaluation index to derive the prediction error and compared it with the schedule data provided by Dalian port, as shown in Tables III and IV.

In Table III, ATA_PRE represents the predicted time of arrival at the anchorage, and TIME_ERROR represents the error between the model-predicted time of arrival and the actual time of arrival at the anchorage.

Analysis of the two tables shows that the model proposed by the authors can predict the time of ship arrival at anchorage 1 h in advance, the minimum prediction error is about 8 min, the maximum error is about 1 h and the average prediction error is controlled at about 30 min; while the maximum error of the ship schedule is seven days, the minimum value is one day, and the average error is 2.75 days.

In summary, the vessel arrival time at anchorage obtained by this model is more accurate than the vessel arrival time at anchorage in the ship schedule, and this model is reasonable and effective, which has practical application significance for the improvement of the finger berthing efficiency and the development of the shore and bridge resource allocation scheme.

V. CONCLUSION AND OUTLOOK

At present, the low efficiency of designating berths and the unreasonable allocation of shore and bridge resources are important problems. The relevant scheduling work of the port is based on the shipping schedule sent by AIS equipment, which has a large error and is easy to cause a waste of resources; therefore, a more accurate time of ship arrival at anchorage can provide an auxiliary decision solution for the berthing system and shore and bridge resource allocation.

In this article, we propose to use the ship trajectory inflexion point to predict the ship's arrival time at the anchorage.

For the ship arrival time at anchorage prediction, first, MeanShift clustering is performed on the ship AIS data heading to each anchorage to find the dense trajectory area (i.e., the inflexion point) of the ship's route heading to each anchorage; second, the distance from the ship arrival position to the anchorage is calculated and the mean value of the ship-to-ground speed SOG is selected; finally, the time difference between the ship from the inflexion point to the anchorage is calculated and combined with the current ship trajectory The estimated time to anchorage (ATA) is calculated by combining the current ship's trajectory update time. The model proposed in this article can control the minimum prediction error at about 8 min and the average prediction error at about 30 min; the average error between the expected time to anchorage and the actual time to anchorage in the ship's schedule reaches 2.75 days, which shows that the model can effectively improve the berthing efficiency of the port and optimize the allocation of shore and bridge resources.

Acknowledgment

The author wishes to express his gratitude to his supervisor for his guidance and to the other authors for their assistance in writing this manuscript.

References

[1] J. Han and B. Y. Wang, "Application of MeanShift algorithm in image segmentation," *Mod. Comput.*, vol. 27, no. 33, pp. 71–76, Nov. 2021, doi: 10.3969/j.issn.1007-1423.2021.33.013.

[2] H. C. Li et al., "Study on price sensitivity based on electricity market reform," *J. North Univ. China, Natural Sci. Ed.*, vol. 41, no. 5, pp. 431–436, 2020, doi: 10.3969/j.issn.1673-3193.2020.05.008.

[3] H. Shen and S. Y. Li, "Underground positioning method based on WiFi and PDR fusion," *Coal Mine Machinery*, vol. 41, no. 10, pp. 202–204, Oct. 2020, doi: 10.13436/j.mkjx.202010064.

[4] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, Aug. 1995, doi: 10.1109/34.400568.

[5] C. Pani, T. Vanelslander, G. Fancello, and M. Cannas, "Prediction of late/early arrivals in container terminals—A qualitative approach," *Eur. J. Transp. Infrastruct. Res.*, vol. 15, no. 4, pp. 536–550, 2015.

[6] G. Fancello, C. Pani, M. Pisano, P. Serra, P. Zuddas, and P. Fadda, "Prediction of arrival times and human resources allocation for container terminal," *Maritime Econ. Logistics*, vol. 13, no. 2, pp. 142–173, Jun. 2011, doi: 10.1057/mel.2011.3.

[7] X. Xiao et al., "Ship trajectory clustering model based on AIS data and its application," *Navigat. China*, vol. 38, no. 2, pp. 82–86, Jun. 2015, doi: 10.3969/j.issn.1000-4653.2015.02.020.

[8] X. Xiao, "Study on ship's trajectory clustering model based on AIS data," M.S. thesis, Navigat. College, Jimei Univ., Fujian, China, 2015.

[9] G. L. Tang, "Arrival time forecasting of container vessels based on AIS data," *J. Chongqing Jiaotong Univ., Natural Sci.*, vol. 38, no. 2, pp. 1–5, May 2019.

[10] W. T. Dong, "Research on trajectory similarity analysis of fishing boats based on AIS," M.S. thesis, College Inf. Eng., Dalian Ocean Univ., Dalian, China, 2020.

[11] K. Park, S. Sim, and H. Bae, "Vessel estimated time of arrival prediction system based on a path-finding algorithm," *Maritime Transp. Res.*, vol. 2, Feb. 2021, Art. no. 100012, doi: 10.1016/j.martra.2021.100012.

[12] C. Zhang et al., "AIS data driven general vessel destination prediction: A random forest based approach," *Transp. Res. C, Emerg. Technol.*, vol. 118, Sep. 2020, Art. no. 102729, doi: 10.1016/j.trc.2020.102729.

[13] A. Dobrkovic et al., "Towards an approach for long term AIS-based prediction of vessel arrival times," in *Logistics and Supply Chain Innovation: Bridging the Gap Between Theory and Practice*. Switzerland: Springer, Jan. 2018, pp. 281–294, doi: 10.1007/978-3-319-22288-2_16.

[14] C. D. Zhao et al., "Ship trajectory prediction of RNN-Bi-LSTM based on characteristic attention mechanism," *J. Guangdong Ocean Univ.*, vol. 42, no. 5, pp. 102–109, Sep. 2022.

[15] H. Kwun and H. Bae, "Prediction of vessel arrival time using auto identification system data," *Int. J. Innov. Comput. Inf. Control*, vol. 17, no. 2, pp. 725–734, Apr. 2021, doi: 10.24507/ijicic.17.02.725.

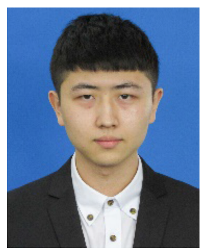
[16] Y. D. Zhao, "Forecasting model of ship arrival time considering AIS information," M.S. thesis, Dept. Harbor Ocean Eng., Dalian Univ. Technol., Dlian, China, 2018.

[17] M. Y. Chen, "Trajectory clustering algorithm based on improved Hausdorff distance," *Comput. Eng.*, vol. 38, no. 17, pp. 157–161, Sep. 2012, doi: 10.3969/j.issn.1000-3428.2012.17.044.

[18] O. Bodunov, F. Schmidt, A. Martin, A. Brito, and C. Fetzer, "Real-time destination and ETA prediction for maritime traffic," in *Proc. 12th ACM Int. Conf. Distrib. Event-based Syst.*, New York, NY, USA, Jun. 2018, pp. 198–201.

[19] D. Q. Xie et al., "The circular fitting algorithm based on MeanShift clustering and its application in welding spot detection," *Comput. Simul.*, vol. 36, no. 9, pp. 446–451, Sep. 2019, doi: 10.3969/j.issn.1006-9348.2019.09.093.

[20] M. Guan, Y. Cao, and X. Cheng, "Research on the recognition of multiple MMSI codes on a single vessel based on AIS datas," *IEEE Access*, vol. 11, pp. 106580–106586, 2023, doi: 10.1109/ACCESS.2023.3317706.



Meng Guan was born in Qingdao, Shandong, China, in 1998. He received the bachelor's degree in Internet of Things engineering from Qingdao University of Science and Technology, Qingdao, in 2021. He is currently pursuing the master's degree in computer technology with Liaoning Petrochemical University, Fushun, China.

During his graduate studies, his main research directions include big data analysis and the development and application of machine learning algorithms. He participated in and completed his advisor's project on maritime big data analysis, which was funded by the Liaoning Provincial Applied Basic Research Program. He has successfully applied for three software copyrights and has received school-level scholarships and the title of Outstanding Class Cadre multiple times.

Yu Cao received the Ph.D. degree from Northeastern University, Boston, MA, USA.

He is currently an Associate Professor with the College of Information and Control Engineering, Liaoning Petrochemical University. He has contributed to the publication of over 40 papers and has been granted four patents both domestically and internationally. Additionally, he has taken the lead in more than ten research projects. His research focus spans complex network dynamics, machine learning algorithms and applications, big data analysis, and cloud computing, as well as image processing.

Xu Cheng received the B.S. degree in automation and the M.S. and Ph.D. degrees in system engineering from Northeastern University, Shenyang, China, in 2007, 2009, and 2015, respectively.

Since 2017, she has been a Lecturer with Shenyang Agricultural University, Shenyang. Her research interests include integer programming and combinatorial optimization, intelligent optimization algorithms, and data analysis.