Vessel Motion Statistical Learning based on Stored AIS Data and Its Application to Trajectory Prediction

Lu Sun[1, 2, a], Wei Zhou[1, b]

[1] Research Institute of Information Fusion, Naval Aeronautical and Astronautical University, Yantai, 264001, China

[2] Institute of Electronics and Information Engineering, Naval Aeronautical and Astronautical University, Yantai, 264001, China

[a]email: sunlu825007@163.com, [b]email:yeaweam@163.com

Keywords: Maritime Surveillance; Vessel Motion; Statistical Learning; AIS; Trajectory Prediction

Abstract. A vessel motion statistical learning based on stored AIS data is proposed in this paper. This paper divide the region of interest into a uniformly sized grid, and analyze the stored AIS data messages according the vessel's position and index the motion information into the unique grid. The sailing state variation between messages are highlighted. Several predictors are designed to predict the vessel's position and the prediction error is get comparing the true position achieved from AIS messages. Experimental results show that the proposed model is credible and the prediction accuracy is higher.

I. Introduction

The effective management and control of vessels is one of the important functions of the maritime situation awareness systems. Predicting the position of vessels effectively can enhance the ability of situation awareness. However, limited surveillance resources constrain maritime domain awareness and compromise full security coverage at all times. This problem is exacerbated in areas covered by detecting systems with low SNR, low data rate, low resolution, and low information dimensions, especially in sea battlefield. This paper intend to explore some new approaches to locate the vessel's position and predict the future state accurately.

Automatic Identification System (AIS), provide a vast amount of near-real time information, calling for an ever increasingly degree of automation in transforming data to decision support elements. The AIS system was originally conceived for collision avoidance, allowing vessels to broadcast information on their location (position information). AIS also provides a wealth of valuable surveillance data and the motion messages which can be effectively archived in databases[1]. AIS contains static information such as vessel name and the unique number (maritime mobile service identity, MMSI), and dynamic information such as position, SOG (speed over ground), COG (course over ground) and true heading.

The database has stored a large amount of vessel sailing information in a long time, containing a lot of valuable "knowledge", which needs mining and analysis. The approach in this paper divide the region of interests into grids, and each AIS dynamic message will be indexed to a unique grid, accumulating the motion data, which will be the basis for trajectory prediction. The experiment results with a real data set show that the proposed vessel motion learning method is credible and the prediction accuracy is higher when applying to trajectory prediction.

The remainder of the paper is organized as follows. Section II reviews related work in the field of vessel motion analysis. Section III gives an overview of the proposed vessel motion learning process and the application to trajectory prediction is given in Section IV. Section V applies the methodology to a real data set and conclusions are reported in Section VI.

II. Related Work

Several methods have been proposed to derive motion patterns from a collection of trajectories as applied in video surveillance and image processing, where the traffic flows are constrained to stay in specific areas (see[2-3]). A trajectory prediction model based on Gaussian mixture models called GMTP for intelligent transportation systems is proposed in [4]. The GMTP algorithm is naturally a Gaussian nonlinear statistical probability model and the advantage of the proposed model is that the result is not only a predicted value, but also a whole distribution beyond the future trajectories, therefore making it possible to infer the location by using statistical probability distribution. The application of such techniques in maritime domain has gained a recent interest.

Some researchers focus on the anomaly detection from stored AIS data [5]-[7], the general similarity is that those methods get training result from historical trajectories to identify whether the target is abnormal.

In [9], interesting insights are obtained as well as an understanding of Dalian sea area such as high traffic density area and spatial pattern of traffic flow. In [10], the vessel trajectory is smoothed and predicted with least square method by using the recorded AIS vessel observation node data. In [11], the selection of grid size is discussed, identifying the optimal grid size for two different conditions – an open sea and a port area case. In series [12-14], an improved neurobiological inspired algorithm for situation awareness is proposed, producing prediction of future vessel location on the basis of current vessel behavior and obtaining better performance by using a multi-scale approach to representing spatial location that matches the spatial scale to the track behavior in a given region.

III. Vessel Motion Statistical Learning

A. Assigning Geographic Grids

In order to suppress the computational complexity, the surveillance area should be discretized into a uniformly sized grid to facilitate the integration of AIS data and spatial analysis. We use dimensions $0.001°\times 0.001°$ as the size of a grid to divide up the region of interest. A unique key will be given to each grid, which is a point structure consisting of longitude and latitude. For instance, a position point of (121.393735, 37.588135) will locate in a grid with key (121.393, 37.588).

With the geographic grids, each dynamic AIS message with MMSI, position, speed and heading, will be indexed to a unique grid by the position, and the motion structure in the grid will also accumulate sailing message from different vessels.

B. Motion Structure of Grid

Vessel's trajectory data from AIS is discrete, and the interval between two AIS messages from one vessel ranges from 2 seconds to 3 minutes, which is related to vessel state.

Although the message is discrete and the intervals are variable, each message contains spatial location, SOG, COG and other dynamic information, that is to say, each dynamic message can be indexed by a unique grid key according to the spatial location. In a motion structure of grid, SOG and COG in each message will be recorded, if next message of this vessel is recorded, the SOG variation rate and COG variation rate will be calculated, which are quite important, as the change of SOG (accelerate or decelerate) and COG (turn left or right) determine the next vessel state in a large part.

We proposed a motion structure named *GridMotion*, which is a set of dynamic messages receiving from different vessels. Each element in this set is consisting of MMSI, SOG, COG, SOG variation rate and COG variation rate, which are concerned with vessel motion. The spatial location in this message has worked on which grid it belongs to, which has been implied in it.

Assuming that the MMSI of target i is $mmsi_i$ and the state vector is $\vec{x}_k^{(i)} = \{lon_k^{(i)}, lat_k^{(i)}, sog_k^{(i)}, cog_k^{(i)}\}$ at time t_k, while the state vector at time t_{k+1} is $\vec{x}_{k+1}^{(i)} = \{lon_{k+1}^{(i)}, lat_{k+1}^{(i)}, sog_{k+1}^{(i)}, cog_{k+1}^{(i)}\}$, then SOG variation rate $\Delta sog_k^{(i)}$ and COG variation rate $\Delta cog_k^{(i)}$

can be calculated as follows: $\Delta sog_k^{(i)} = \dfrac{sog_{k+1}^{(i)} - sog_k^{(i)}}{t_{k+1} - t_k}$, $\Delta cog_k^{(i)} = \dfrac{cog_{k+1}^{(i)} - cog_k^{(i)}}{t_{k+1} - t_k}$.

C. Motion Learning Process

A dictionary structure is used during the motion learning process.

$GridMotions = \text{Dictionary} < GridKey, GridMotion >$

The key of the dictionary is the grid introduced in Section 3.1, and the value is the motion structure introduced in Section 3.2. The pseudo code of motion learning process is shown in Table. 1.

After the entire data learned complete, the motions will be frozen.

Table.1. The pseudo code of motion learning

Algorithm 1: Motion Learning

Input: all dynamic information for learning $\{\vec{x}\}$

foreach $\vec{x}_k^{(i)}$ in $\{\vec{x}\}$
    $GridKey = LocateInGrid(lon_k^{(i)}, lat_k^{(i)})$
    $\Delta sog_k^{(i)}, \ \Delta cog_k^{(i)} \leftarrow Delta(\vec{x}_k^{(i)}, \vec{x}_{k+1}^{(i)})$
    if $!GridMotions.Keys.Contains(GridKey)$
        $GridMotion =$ **new** $GridMotion(sog_k^{(i)}, cog_k^{(i)}, \Delta sog_k^{(i)}, \Delta cog_k^{(i)})$
        $GridMotions.Add(< GridKey, GridMotion >)$
    else
        $GridMotions[GridKey].Add(sog_k^{(i)}, cog_k^{(i)}, \Delta sog_k^{(i)}, \Delta cog_k^{(i)})$
    end if
end foreach

IV. Position Predictor

A. Prediction Directly

Prediction directly is reference to Kalman filter [10], which predict next state with current state.

Assuming the state vector of target i at time k is $\vec{x}_k^{(i)} = \{p_{x,k}^{(i)}, \dot{p}_{x,k}^{(i)}, p_{y,k}^{(i)}, \dot{p}_{y,k}^{(i)}\}$, the state vector at time $k+1$ is $\vec{x}_k^{(i)} = F_k \vec{x}_{k-1}^{(i)} + \omega_k$, where $F_k = \begin{pmatrix} I_2 & \Delta I_2 \\ 0 & I_2 \end{pmatrix}$ and process noise ω_k is Gaussian white noise with mean is zero and covariance matrix is $Q_k = \sigma_v^2 \begin{pmatrix} \frac{\Delta^4}{4} I_2 & \frac{\Delta^3}{2} I_2 \\ \frac{\Delta^3}{2} I_2 & \Delta^2 I_2 \end{pmatrix}$, where I_n, 0_n denote the $n \times n$ identity and zero matrices respectively, Δ is the sampling period (in this case, it is $t_{k+1} - t_k$), and σ_v is the standard deviation of the process noise.

This prediction method is the comparing method in the following experiments, noting *Predictor One*.

B. Position Predicting based on Learned Vessel Motion

As the vessel motions have been learned in Section 3, the position predictor can take full use of the learned grid motions. When the current vessel state $\vec{x}_k^{(i)} = \{lon_k^{(i)}, lat_k^{(i)}, sog_k^{(i)}, cog_k^{(i)}\}$ is get from AIS, the grid motion can be indexed in the learned grid motions by grid key easily, as the main change of SOG and COG has been analyzed, which will contribute to predicting the lately SOG and COG, making the predictor more truthful.

In order to make it simple, the predictor based on learned motions here only use mean SOG variation rate and mean COG variation rate to indicate the vessel change of SOG and COG, in condition of distinguishing the rough COG(there are two main headings near the lane area).

The pseudo code of this predictor is shown in Table. 2.

Table.2. The pseudo code of positing prediction

Algorithm 2: Position Predicting Algorithm
Input: current vessel state $\vec{x}_k^{(i)} = \{lon_k^{(i)}, lat_k^{(i)}, sog_k^{(i)}, cog_k^{(i)}\}$,
learned grid motions $GridMotions$
1: $GridKey = LocateInGrid(lon_k^{(i)}, lat_k^{(i)})$
2: **if** $!GridMotions.Keys.Contains(GridKey)$
3: $GridKey = FindNearestKey(GridMotions.Keys, lon_k^{(i)}, lat_k^{(i)})$
4: **end if**
5: $GridMotion = GridMotions[GridKey] / GridMotions[mmsi_i][GridKey]$
6: $sog = sog_k^{(i)} + GridMotion[GridKey].MeanDeltaSOG$
7: $cog = cog_k^{(i)} + GridMotion[GridKey].MeanDeltaCOG$
8: $PredicitonPosition = Predicitor(lon_k^{(i)}, lat_k^{(i)}, sog, cog)$
Output: $PredicitedPosition$

V. Experiment results

A. Data Set Description and Experimental Setup

Training data in experiment are collected from 2016-4-1 to 2016-6-4 by AIS base station on land which covers the *Yantai* harbor sea area. Then we extracted messages from the collected data in the area with bound of [121.39, 121.47]×[37.55, 37.60], which is noted as *dataset 1,* and extracted the messages from *dataset 1* sent from the vessels from *Bohai Lundu* Company, noted as *dataset 2*.

The total message number of *dataset 1* is 2984277, which is 95372 of *dataset 2*. Firstly, the *dataset 2* is imported, after the learning process introduced in Section 3, the dataset covers 654 grids, i.e. 654 *GridMotions* have been learned. Secondly, the *dataset 1* is imported, getting 2944 learned *GridMotions*.

The validation data are collected in 2016-6-5 from the same sea area and all the messages sent from the vessels of *Bohai Lundu* Company. Main information of the validation data is shown in Table. 3.

After motions learned, the predictors will work to trajectory predicting. In this paper, we uses four predictors, which are described in Table. 4.

When predictor works, the intervals between validation AIS data will be the prediction intervals, and regard the received AIS message as the true value of the vessel's position. The evaluation to the predictors is the Euclidean distance between predicted position and the true value, which contains the error mean and error standard deviation.

B. Prediction Result

The error mean and error standard deviation of prediction result of the four predictors applying to the validation data is calculated. The bar chart of all error means is shown in Fig. 1, and the bar chart of all error standard deviations is shown in Fig. 2.

The experiments show that the prediction error of *Predictor Two* and *Predictor Three* are smaller than others, i.e. the position predictor using motions learned from the vessel's own motion messages and the same kind of vessels can achieve higher accuracy. Comparing with *Predictor One,* when predicting positions, the variation of speed and course learned from the history motions contribute to the position predicting in *Predictor Two* and *Predictor Three*, which has been proved effective. In contrast, the prediction result of *Predictor Four* is not satisfactory, as the variation of speed and course used for predictor is learned from all the vessels in this area, containing different types of vessels with different sailing pattern and characteristics. It is not suitable for predicting position with such learned motions.

Two of the validation trajectories is shown in Fig. 3 and Fig. 4. In Fig. 3, the intervals between messages are small, and vessel turned its course often, resulting in poor performance of *Predictor One*. In Fig. 4, there are missing messages, so some individual intervals are larger. The performance of *Predictor Four* is poor because the motions of other vessel kinds influence the result a lot.

Table.3. Main information of the validation data

No.	MMSI	Data Count	Time duration	Initial position	End position
1	412330020	162	0:33:50--0:50:03	121.3923°E, 37.5890°N	121.4459°E, 37.5999°N
2	412328370	162	2:31:39--2:46:26	121.4463°E, 37.5999°N	121.3923°E, 37.5651°N
3	414095000	142	3:14:04--3:28:16	121.4472°E, 37.5998°N	121.3921°E, 37.5653°N
4	414210000	155	4:00:39--4:14:27	121.4497°E, 37.5996°N	121.3931°E, 37.5651°N
5	413408000	173	6:06:20--6:21:52	121.4480°E, 37.5997°N	121.3923°E, 37.5651°N
6	412328370	107	8:22:24--8:38:33	121.3924°E, 37.5647°N	121.4445°E, 37.5999°N
7	414095000	105	9:20:00--9:34:49	121.3931°E, 37.5657°N	121.4465°E, 37.5999°N
8	414210000	82	13:34:35--13:47:24	121.3957°E, 37.5669°N	121.4466°E, 37.5998°N
9	414096000	172	14:46:33--15:02:40	121.4491°E, 37.6000°N	121.3923°E, 37.5652°N
10	413408000	138	15:59:31--16:14:14	121.3921°E, 37.5651°N	121.4456°E, 37.5999°N
11	414211000	168	18:43:50--18:58:20	121.4490°E, 37.5997°N	121.3921°E, 37.5648°N
12	412330020	158	20:49:08--21:08:22	121.4493°E, 37.5998°N	121.3927°E, 37.5888°N
13	414096000	144	21:17:58--21:32:56	121.3925°E, 37.5652°N	121.4453°E, 37.6000°N
14	414211000	138	22:48:51--23:03:10	121.3921°E, 37.5652°N	121.4485°E, 37.5998°N

Table.4. The description of four predictions

Predictors	Description
Predictor One	Predict directly(introduced in part A of Section IV)
Predictor Two	the motions used for predictor learned from the vessel's own motion messages
Predictor Three	the motions used for predictor learned from the same kind of vessels, i.e. vessels from *Bohai Lundu*,
Predictor Four	the motions used for predictor learned from all vessels

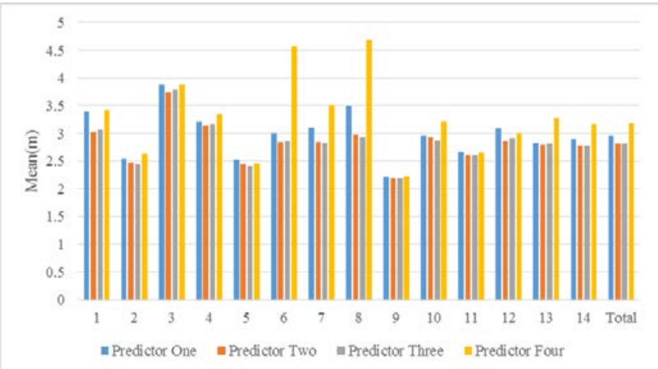

Fig.1. Bar chart of the error mean

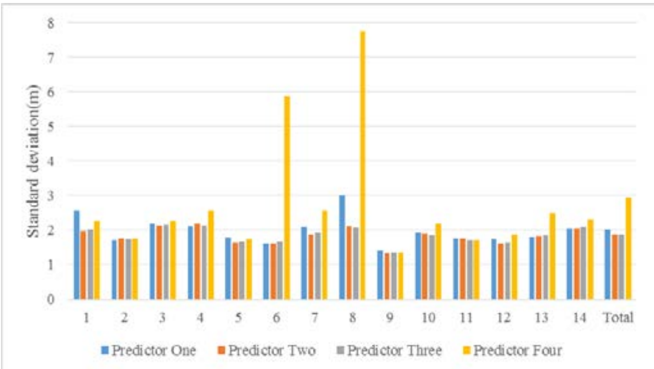

Fig.2. Bar chart of the error standard deviation

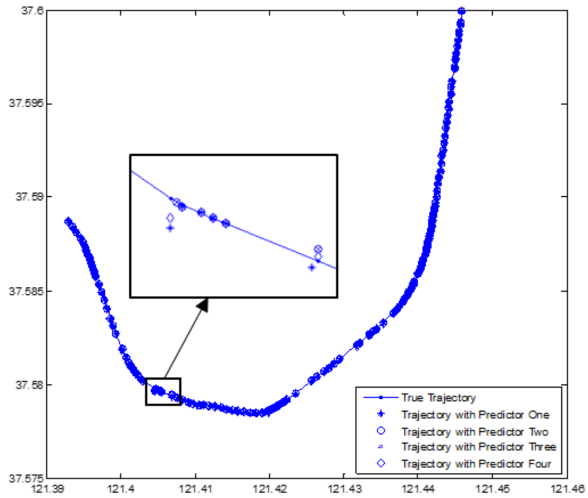

Fig.3. Prediction result of trajectory No.1

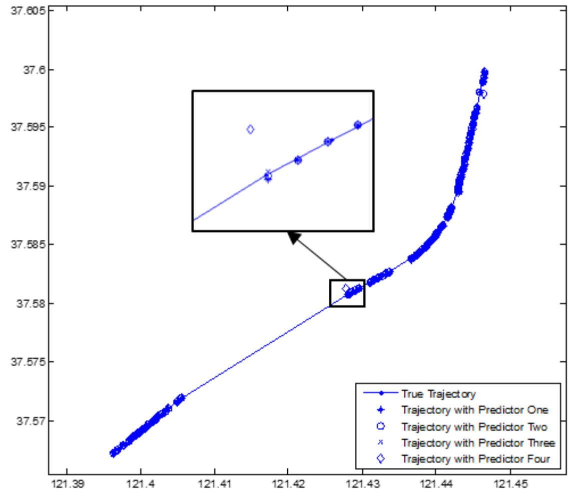

Fig.4. Prediction result of trajectory No.8

VI. Conclusions

In this paper, we proposed a vessel motion statistical learning model based on stored AIS data and applied it to vessel position predicting. The vessel motions are learned from stored AIS data in a unsupervised way, which is the basis of trajectory prediction. The experiment results show that the proposed vessel motion learning model is credible and the accuracy is higher when applying to trajectory prediction. The performance of predictor using motions learned from the vessel's own motion and the same type vessels is satisfactory. However, the vessel types in validation data (i.e. *Bohai Lundu*) are priori information, which can be hardly obtained from the raw AIS data automatically.

We will continue to look for new approaches to classify vessels by distinguishing the different motions and look for other mechanisms with which to further enhance the performance for position predicting with different time intervals.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China (Project No. 61531020) and by the Special Funds of Taishan Scholars Construction Engineering.

References

[1] G. Pallotta, M. Vespe, K. Bryan. Traffic knowledge discovery from AIS data[C], Information Fusion, 2013 16th International Conference on, pp. 1996-2003.

[2] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, S. Maybank. A system for learning statistical motion patterns[J]. IEEE Trans. Pattern Anal. Mach. Intell., 28, pp. 1450-1464, 2006.

[3] B. T. Morris, M.M. Trivedi. A Survey of Vision-Based Trajectory Learning and Analysis for Surveillance[J], IEEE Trans. Circuits Syst. Video Technol., 18, pp. 1114-1127, 2008.

[4] Qiao SJ, Jin K, Han N, Tang CJ, Gesangduoji, Gutierrez LA. Trajectory prediction algorithm based on Gaussian mixture model. Journal of Software, 2015, 26(5):1048-1063.

[5] M. Riveirom, G. Falkman. Detecting anomalous behavior in sea traffic: A study of analytical strategies and their implications for surveillance systems. International Journal of Information Technology and Decision Making, 2, pp. 317, 2014.

[6] Po-Ruey Lei. A framework for anomaly detection in maritime trajectory behavior. Knowledge and Information Systems, pp. 189-214, 2016.

[7] O. Ines, M. Mario, Krunoslav. Machine Learning Approaches to Maritime Anomaly Detection. Nase More, 5-6, pp. 96-101, 2014.

[8] D. Andrej, I. Maria-Eugenia, H. Jos Van. Using Machine Learning for Unsupervised Maritime Waypoint Discovery from Streaming AIS Data. 15th International Conference on Knowledge Technologies and Data-Driven Business, 2015.

[9] Zhu Feixiang. Mining ship spatial trajectory patterns from AIS database for maritime surveillance. IEEE International Conference on Emergency Management and Management Sciences, pp. 772-775, 2011.

[10] XU Tie, CAI Feng-jun, HU Qin-you, YANG Chun. Research on estimation of AIS vessel trajectory data based on Kalman filter algorithm. Modern Electronics Technique, 2014, 37(5): 97-100, 104.

[11] Ewa Osekowska, Henric Johnson, Bengt Carlsson. Grid Size Optimization for Potential Field based Maritime Anomaly Detection. Transportation Research Procedia, 2014, 720-729.

[12] N. A. Bomberger, B. J. Rhodes, M. Seibert, A. M. Waxman. Associative Learning of Vessel Motion Patterns for Maritime Situation Awareness. Information Fusion, 2006 9th International Conference on.

[13] B. J. Rhodes, N. A. Bomberger, M. Zandipour. Probabilistic associative learning of vessel motion patterns at multiple spatial scales for maritime situation awareness. Information Fusion, 2007 10th International Conference on.

[14] M. Zandipour, B. J. Rhodes, N. A. Bomberger. Probabilistic prediction of vessel motion at multiple spatial scales for maritime situation awareness. Information Fusion, 2008 11th International Conference on.