MODELING AND ANALYSIS OF MARINE TRAFFIC FOR ANOMALY DETECTION (A POSITION PAPER)

Wen-Jing Hsu^(a), Hui Fang^(b), and Shell-Ying Huang^(c)

^{(a),(b),(c)}School of Computer Engineering Nanyang Technological University Nanyang Avenue, Singapore 639798

^(a)hsu@ntu.edu.sg; ^(b)hfang@ntu.edu.sg; ^(c)assyhuang@ntu.edu.sg;

ABSTRACT

This paper presents a system for modeling and analyzing the patterns and anomalies in the vessel movement from collected sensor data. The major tasks involved are (a) Detection of macro-level traffic patterns, which includes identifying places of significance based on dwelling time or traffic density. It also includes construction of representative routes of real traffic as opposed to the nominal path on navigational maps. (b) Characterization of vessel movements: classification of journeys, transitional probability between significant places, ship move models, interaction model, etc. which describe the moves of the ships; (c) Tracking and prediction of ships, which enables prediction of the future physical position of a given ship, and the evolution of multiple interacting ships. (d) Anomaly detection: Vessels that behave anomalously in time, space, or mode are identified and monitored. We believe that the system developed will be a useful tool for marine safety and security.

Keywords: Anomaly detection, pattern analysis, machine learning, marine traffic

1. BACKGROUND AND MOTIVATIONS

Hundreds of thousands of vessels visit hub ports annually. At any point in time, there are more than 1,000 ships inside or around Singapore waters. Activities of these vessels are generally of high value to the nation's economy, and hence services must be provided to ensure their smooth operations; illicit or anomalous activities, on the other hand, must be prevented or minimized as much as possible.

To safeguard Singapore's interest and provide quality services, maritime and port authorities closely monitors the movements of these vessels. The authority typically deploys radar and vessel Automated Identification System (AIS) for this purpose and it collects large volume (multi-billion bytes annually) of data. Upon judicial analysis, the vessel data can reveal many kinds of useful information. For instance, Fig. 1 shows a time-lapsed view of the vessel data, which reveals the spatial distribution of vessels and usage of the navigational channels and anchorages.



Figure 1. Time lapsed vessel data showing the spatial distribution of vessels in and around Singapore waters

Fig. 2 shows circular movements of a particular vessel, which reveals the actions of the tide and currents on the vessel in the anchorage.



Figure 2. Vessel moving in circular fashion

It is also possible to trace particular vessels in rather fine-grained resolutions. Fig. 3 shows the trajectory of a particular vessel in a given visit, which enables closer examination, e.g. its nearest point to certain security-sensitive areas.



Figure 3. Trajectory of a particular vessel during a recent visit

Thus, it is possible to characterize the general traffic flows and movement patterns of various types of vessels, their temporal or spatial distributions, interactions and navigational behaviors over different segments of the navigational channels and their junctions. This information can be of use in the planning of navigational resources and marine operations. Given the normal activity patterns, it becomes possible to identify anomalous or suspicious behaviors which can be of use to marine security and safety.

2. RESEARCH OBJECTIVES

Presently the burden of analysis is placed mainly on specially-trained human operators and administrators. Machine-assisted data-analysis, however, are becoming inevitable because of the massive amount of data, short time to response, scarcity of qualified personnel, high cost of training. Machine tools are also valued for their consistence and dependability.

This project aims to develop computer-based tools to analyze the patterns and anomalies in the vessel movement from the archived sensor data. The following outlines the major tasks in this project.

- **Detection of macro-level traffic patterns**. This includes identifying places of significance (e.g. locations of long dwelling time or changes of speed or directions), reconstruction of actual trajectories as opposed to the designated path on navigational maps.
- Characterization of vessel movements: classification of journeys, transitional probability between SPs (Significant Places), ship move models, interaction model, etc. which describe the moves of the ships;
- **Tracking and prediction of ships**: In the simplest scenario, predict the future physical position of a

given ship, and in the more complicated scenario, the evolution of multiple interacting ships.

• Anomaly detection: Vessels that behave anomalously in time, space, or mode should be identified and monitored. For instance, vessels near Jurong Island should be monitored automatically.

3. APPROACH AND METHODOLOGY

To meet the challenges, we propose the following methodology from machine learning:

- Automated extraction of macro traffic patterns: The data log could be divided into training set and verification set; where the training set is used to extract patterns by using clustering and classification algorithms. The verification set is used to test and verify the rules extracted by the algorithms. The boundaries and the representative pathways of the traffic flows can be inferred automatically to generate the actual or de facto navigational network (as opposed to the stipulated, on-paper layout), and the distributions of the ships inside the actual routes can also be derived. General unsupervised clustering algorithms will be adapted to automatically identify the feature values that exhibit high concentrations. This will allow discovery of, e.g., places and time points of significance.
- Characterization of vessel journeys: Given the uncertain sensor data, paths linking the significant places can be constructed; the resulting paths are simplified representation that minimizes the expected errors. A journey may be decomposed into higher level "motifs" such as piecewise linear polygonal lines and semi-circular curves that link SPs. Statistics about vessel speed, directions, transitional probability between SPs (Significant Places) over the paths can be gathered for normal operations. Instantaneous traffic volume, vessel spatial distribution, flow rate can be inferred to find places of possible bottlenecks and congestion.
- Traction and Prediction of ship movements: For simple non-interacting vessel behaviors, ship moves can be modeled by using known algorithms such as Kalman Filter; For interacting behaviors, a probabilistic Bayesian model such as Particle Filters will be more adequate. Vessel interaction model will require identification of symptoms and signatures of various scenarios; With sufficiently detailed historical data, blow-by-blow evolution of multiple interacting ships can be modeled and analyzed.
- Anomaly detection: Once the patterns of normal ship journeys and movements are found, they are used to identify exceptional cases that deviate from

the norms. Vessels that behave anomalously in time, space, or mode can be identified automatically. Based on the high-level motif representation, journeys may be classified according to normal/suspicious categories for further analysis. Input from experienced ship captains and hydraulic experts will also be crucial.

Screening and cleansing of raw data: Because of interferences to radar signals or disturbances in electro-magneto fields, the sensor data may be erroneous and less than precise or reliable. For instance, the recorded data may have spurious instances of records which show vessels at an unlikely location (e.g. on land) or vessels traveling at exceedingly high speed. Certain data filed may also be missing. This pruning phase will require careful filtering through statistically justified method and the data should be discarded only after verification with domain experts. The sheer volume of the logged data and the complexities involved in the movement patterns dictate that only efficient algorithms that exhibit lower computational time and space complexity will be useful. It may eventually require a solution that runs on multi-core computers.

4. SYSTEM ARCHITECTURE

The architecture and the system components for vessel movement analysis are further described below:

(a) Vessel Journey Data Mining Engine: This is the software for finding "norms" from massive amount of data, and summarizing into higher-level abstraction (e.g. envelop or skeleton", list of significant places, paths and journeys); and automated construction of ship movement model.

(b). Vessel Tracking Engine: this software implements tracking of vessel movements by using Kalman filter or Particle filters. It handles inaccuracies in live sensor data and uncertainties in positions/locations by means of probabilistic distributions and recursive Bayesian inference methods.

(c). State Inference Engine: This software implements higher-level state inference schemes by using tools such as Partially Observable Markov Decision Process (POMDP).

(d). Anomaly Detection Engine: This software implements lower-level geometry matching (Digital Hough Transforms), sequence matching methods and recognition of activities by making reference to templates such as norms of same types of vessels or vessels with the same purpose of visit, or the vessel's own historical activities.

(e). Query Processing Engine and Console: This software handles the human computer interactions and operator control. It allows user to activate subsystem functions such as querying for vessel information, specifying regions, time, vessels, events, behaviors of interest.

(f). Visualization Engine: Software that shows output generated by the other subsystems, e.g.,

- Show vessel activities over a period of time in the past (which is an output from Data-Mining Engine or Query Processing Engine), present (from live sensors), and projection into future(from Vessel Tracking Engine);
- Show vessel activities within regions of interest from HCI-human computer interactions;
- Select vessels of interests either automatically from State Inference Engine/Anomaly Detection Engine or manually from HCI, to display/highlight according to types (e.g. VLCC, large passenger vessels), size (e.g. LOA> x m, GT > y tons), sensitive areas (e.g. inside junctions or near an crude oil refinery), projected situations (e.g. vessels whose domains are projected to overlap), vessel interactions: display vessels that come into vicinity; special purposes (tracking particular vessels) etc.

5. CONCLUSION

This paper presented a system design and a Bayesian model for simulating and analyzing the patterns and anomalies in the vessel movement from collected sensor data. The system is able to identify places of significance based on dwelling time or traffic density. It can also construct representative routes of real traffic as opposed to the nominal path on navigational maps. The system can classify journeys, characterize vessel movements between significant places, and derive transitional probabilities. Bayesian models are used to describe the moves of the ships; ship move models, interaction models. Upon completion, they will be able to track and predict ship moves and the evolution of multiple interacting ships and detect anomalies. We believe that the system will be a useful tool for marine safety and security.

AUTHORS BIOGRAPHY

Dr. Hsu Wen Jing is a faculty member of NTU. He is also an Affiliate Faculty Fellow of Singapore-MIT Alliance Program. He received research awards such as Ministry of Transportation Innovation Award (Merit), Best Paper Awards (PDCAT/IEEE Automation/PDCS), Gold Award and Grand Champion of Crayquest. He has about 200 publications in reviewed journals and conferences. He is presently Deputy Director of the Maritime Research Center of NTU.

Fang Hui is a Ph.D. candidate of NTU. His thesis developed a prediction system by using Bayesian networks. He is also a Research Associate working on a project related to simulation of marine traffic. His thesis research is on mining positional data for patterns.

Dr. Huang Shell Ying is an Associate Professor of NTU. Her research interests include intelligent agents, multi-agent systems, simulation based optimization and intelligent decision support systems. She has been the Principal Investigator or Co-PI for many projects funded by Maritime and Port Authority of Singapore, NSTB (now ASTAR), GINTIC (now SimTech), PSA Int'l, Jurong Port Pte Ltd, Surbana Consultants, APL and NTU. She has published many papers in international journals and international conferences.