Expert Systems with Applications 41 (2014) 2742–2753

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Knowledge discovery using genetic algorithm for maritime situational awareness

Chun-Hsien Chen [a,*], Li Pheng Khoo [a], Yih Tng Chong [b], Xiao Feng Yin [c]

[a] School of Mechanical and Aerospace Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore
[b] Department of Industrial and Systems Engineering, National University of Singapore, Faculty of Engineering, 1 Engineering Drive 2, Singapore 117576, Singapore
[c] Computing Science Department, Institute of High Performance Computing, 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632, Singapore

ARTICLE INFO

Keywords:
Genetic algorithm
Knowledge discovery
Machine learning
Defense
Maritime security
Decision support

ABSTRACT

Due to the large volume of data related to vessels, to manually pore through and to analyze the information in a bid to identify potential maritime threat is tedious, if at all possible. This study aims to enhance maritime situational awareness through the use of computational intelligence techniques in detecting anomalies. A knowledge discovery system based on genetic algorithm termed as GeMASS was proposed and investigated in this research. In the development of GeMASS, a machine learning approach was applied to discover knowledge that is applicable in characterizing maritime security threats. Such knowledge is often implicit in datasets and difficult to discover by human analysts. As the knowledge relevant to maritime security may vary from time to time, GeMASS was specified to learn from streaming data and to generate up-to-date knowledge in a dynamic fashion. Based on the knowledge discovered, the system functions to screen vessels for anomalies in real-time. Traditionally in maritime security studies, datasets that are applied as knowledge sources are related to vessels' geographical and movement information. This study investigated a novel leverage of multiple data sources, including Automatic Identification System, classification societies, and port management and security systems for the enhancement of maritime security. A prototype of GeMASS was developed and employed as a vehicle to study and demonstrate the functions of the proposed methodology.

1. Introduction

Maritime situational awareness, also known as maritime domain awareness, is an area of study that aims to use available data sources to create maximum awareness of activities in the maritime environment. There are currently various types of data that are relevant to maritime security, including Automatic Identification System (AIS) and data of classification societies.

The problem of identifying anomalous vessels involves analyzing multiple sets of data before a decision can be reached. In the case of having human analysts to study sets of information, qualitative reasoning techniques can be applied based on tacit knowledge within the analysts, usually derived from their experiences. In a decision-making process, such tacit knowledge serves to classify ships into different categories. As the information can be large in volume, disparate, and disorganized, it can be challenging for the analysts to pore through datasets, and to generate decisions. This problem is compounded if the time available for decision-making is limited.

Analogous to the tacit knowledge gained by domain experts through experiences, explicit knowledge can be computationally inducted from empirical data. In such approach, machine learning methodologies have been applied to knowledge acquisition. Machine learning techniques function to discover knowledge, for example, via inductive learning (Wong, Ziarko, & Li, 1986). Techniques such as decision tree learning (Quinlan, 1986), neural network learning (Fausett, 1994) and genetic algorithm-based learning (Goldberg, 1989) have also been developed by researchers to extract knowledge from historical datasets.

The general direction of this study is to enhance maritime security through a computational approach. It is known that intelligence that is applicable in detecting anomalous vessels potentially resides in data. Methodologies for extracting knowledge from multiple sets of maritime-related data are therefore studied. The next section briefly presents the state-of-the-art research in this domain.

2. A review of anomaly detection in maritime security

This section presents a review of decision support and anomaly detection systems in the field of maritime security. Fig. 1 shows an overview of the literature reviewed. As depicted in Fig. 1, there

* Corresponding author. Address: School of Mechanical & Aerospace Engineering, Nanyang Technological University, North Spine (N3), Level 2, 50 Nanyang Avenue, Singapore 639798, Singapore. Tel.: +65 6790 4888; fax: +65 6792 4062.
E-mail address: mchchen@ntu.edu.sg (C.-H. Chen).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.09.042

have been mainly two approaches of anomaly detection in the maritime security domain – signature-based and norm-based.

2.1. Signature-based anomaly detection (expert-knowledge driven)

Signature-based methodologies rely on experts' inputs for knowledge on anomalous behaviors and characteristics. Typically, workshops with domain experts are conducted for knowledge elicitation (Roy, 2008; van Laere & Nilsson, 2009). Knowledge elicited includes categories and specific manifestations of a vessel that would constitute an anomaly or a security threat. For instances, frequent change of flags, the allegiance between vessel owners with known terrorist/criminal organizations, loitering maneuvers, cargo types that do not match the port of call, and the shutting down of the AIS.

An anomaly detection system known as Collaborative Knowledge Exploitation Framework (CKEF) was developed to improve maritime domain awareness (Roy, 2008). The rule-based system uses experts' concepts of anomalies to derive signatures as knowledge for applications. More recently an overall framework known as Rule-Based Expert System (RBES) with rules defined by experts and encoded by knowledge engineers was proposed (Roy, 2010). In another example of a rule-based expert system, basic spatial and kinematical relations between objects for the deduction of different situations, e.g. smuggling, hijacking and piloting was developed (Edlund, Grönkvist, Lingvall, & Sviestins, 2006; Laxhammar, 2008). Separately, Fooladvandi, Brax, Gustavsson, and Fredin (2009) investigated whether or not the Bayesian networks acquired from expert knowledge has the ability to detect activities based on a signature-based detection approach.

2.2. Norm-based anomaly detection (data-driven)

It is common to use classification algorithms via a data-driven approach to perform anomaly detection. However, in the case of maritime security research, conventional classification algorithms cannot be directly applied due to the lack of adequate samples and known cases that should be classified as anomalous (Riveiro, Falkman, & Ziemke, 2008). A challenge faced in this research project was therefore that adequate instances of maritime security threats are not readily available in historical data for system training purpose.

An alternate data-driven approach is to build *normal models* via discovering knowledge from historical data (Brax & Niklasson, 2009). In an example, Ristic, Scala, Morelande, and Gordon (2008) presented a statistical analysis of vessel motion patterns (using adaptive Kernel Density Estimation) in ports and waterways based on AIS data. To build normal models, other methods employed by researchers include clustering (Laxhammar, 2008) and probability models. Johansson and Falkman (2007) used Bayesian network (based on probability theory) to build models of normal ship movements. Similarly, Jakob, Vaněk, Urban, Benda, and Pěchouček (2010) built probability models of vessels' locations and trajectories for detecting anomalous vessels. There were also reports of using conformal prediction method for detecting maritime anomalies, again based on normal training data (Laxhammar & Falkman, 2010). A Markov model was applied to define normal models based on historical AIS data, with an aim of detecting abnormal movements (Tun, Chambers, Tan, & Ly, 2007). In BAE Systems' study of learning normal vessel movements, AIS data was used in methodologies, including ARTMAP neural network for pattern identification (Rhodes, Bomberger, Seibert, & Waxman, 2005). BAE Systems also reported the development of a system that fuses radar, AIS, and video-based information for more accurate geographical data of vessels (Seibert et al., 2006).

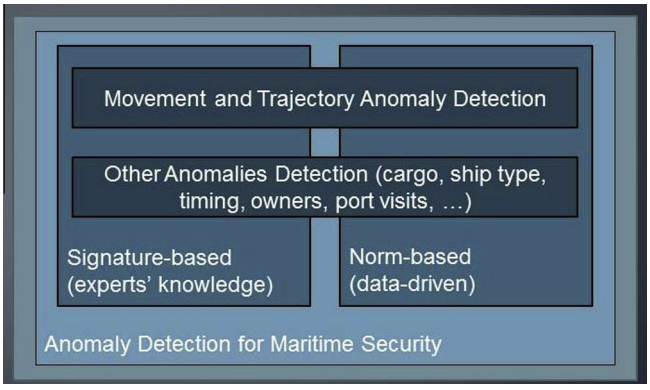

Fig. 1. Research of anomaly detection for maritime security.

3. Research problem and objective

There have been several key challenges for the signature-based anomaly detection approach. It is not straightforward for knowledge engineers to elicit, organize, and represent formal knowledge from experts. The process has to be on-going, as both real-world circumstances and experts' knowledge change with time. Signatures of illegal activities are open-ended, and may therefore be difficult for any system to obtain a robust coverage. Further, the boundaries between anomalous and normal behaviors are often difficult to be defined verbally during knowledge elicitation procedures. It is also a challenge to convert experts' verbatim into formal knowledge. As an alternative to the signature-based approach, there have been studies of building normal models, with the aim of detecting maritime anomalies.

Since anomaly can be defined as deviations from normality, normal models can be used to identify anomalies. This approach is akin to domain experts possessing knowledge of the norms derived from their experiences, and could therefore identify anomalies. As reflected in the reviewed literature, current studies of building normal models have been restricted to detecting anomalies in terms of vessels' physical movements, typically with AIS and/or radar-based datasets as inputs (Riveiro & Falkman, 2011). However, this study identified that information fields (e.g. cargo types, ship types, and port last visited) from AIS, port management systems, and classification societies' datasets potentially contain critical knowledge for identifying maritime anomalies. This research therefore investigates the extraction of knowledge from multiple datasets in supporting anomaly detection via a norm-based approach.

4. An approach of knowledge discovery using genetic algorithm

Knowledge discovery is a process of identifying patterns in a given training data set. These hidden patterns are useful as a basis for making decisions and predictions. Machine learning is an approach of knowledge discovery, whereby autonomous algorithms are prescribed for acquiring knowledge, and to improve the organization of the knowledge obtained (Tecuci & Kodratoff, 1995).

One important step in applying machine learning technique is to decide an effective representation scheme for both the training data and the knowledge extracted. Knowledge representation involves the modeling of knowledge in explicit schemes that facilitate the acquisition, learning, manipulation and application of knowledge. Schemes include mathematical expressions, predicate calculus, conceptual graphs, frames, scripts, objects, semantic networks and production rules (Chong, Chen, & Leong, 2009). In a