Expert Systems with Applications 40 (2013) 5719–5729

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Open data for anomaly detection in maritime surveillance

Samira Kazemi[a], Shahrooz Abghari[a], Niklas Lavesson[a,*], Henric Johnson[a], Peter Ryman[b]

[a] *School of Computing, Blekinge Institute of Technology, SE-371 79 Karlskrona, Sweden*
[b] *Swedish Coastguard, Sweden*

ARTICLE INFO

Keywords:
Open data
Anomaly detection
Maritime security
Maritime domain awareness

ABSTRACT

Maritime surveillance has received increased attention from a civilian perspective in recent years. Anomaly detection is one of many techniques available for improving the safety and security in this domain. Maritime authorities use confidential data sources for monitoring the maritime activities; however, a paradigm shift on the Internet has created new open sources of data. We investigate the potential of using open data as a complementary resource for anomaly detection in maritime surveillance. We present and evaluate a decision support system based on open data and expert rules for this purpose. We conduct a case study in which experts from the Swedish coastguard participate to conduct a real-world validation of the system. We conclude that the exploitation of open data as a complementary resource is feasible since our results indicate improvements in the efficiency and effectiveness of the existing surveillance systems by increasing the accuracy and covering unseen aspects of maritime activities.

1. Introduction

Maritime surveillance is the effective understanding of all maritime activities that could impact the security, safety, economy or environment.[1] Maritime transport handles over 80% of the volume of global trade.[2] Along with the development of the maritime transport system, the threats to maritime security such as illegal fishing and pollution, terrorism, smuggling activities and illegal immigration are increasing correspondingly. According to the Department of Homeland Security,[3] *anomaly detection* is one of several techniques available for improving the safety and security in the maritime domain. Furthermore, an efficient maritime surveillance system requires a complete *recognized maritime picture*, which can be defined as a composite picture of maritime activities over an area of interest (Lefebvre & Helleur, 2001). For national maritime sovereignty, this picture should include all activities within the 200 nautical miles wide *exclusive economic zone*. However, for some purposes such as the detection of illegal vessel transits, the recognized maritime picture could extend beyond this region (Ponsford, ĎSouza, & Kirubarajan, 2009). Using today's technology, continuous tracking of all maritime activities by a single sensor is insufficient since it cannot monitor everything that happens in the surveillance area. On the other hand, there are large amounts of data in the maritime domain that are gathered from a variety of sensors, databases and information systems. Therefore, by taking advantage of all the available data sources it would be possible to obtain a complete recognized maritime picture. The maritime surveillance systems generally use closed data sources that belong to the surveillance area of each country and are obtained from a variety of sensors and databases that are only accessible by the national authorities (see Section 2 for a definition of closed data). For detecting some of the anomalous activities such as smuggling, the maritime data beyond the surveillance area of each country are required. In order to assure security, maritime organizations in different countries need to exchange their privileged data and for this purpose they should deal with the diverse regulations of the data protection in each land. Exchanging data among countries is difficult, time-consuming and in some cases impossible because of the legislative issues. Moreover, there are activities that are neither reported to the maritime organizations, nor recorded in their data sources but these activities can be useful for surveillance purposes. The publicly accessible and reusable data that are free from the legislative issues are referred to as open data. Some of the open data sources may help in revealing previously unknown aspects of maritime activities. For example, there are different organizations such as ports that publish their vessel traffic data or their facility information online. In addition to the organizations, there are different online communities such as blogs, forums and social networks which provide the possibility of sharing information about maritime events. By exploiting the open data along with other confidential sources of data in the detection process, the anomaly detection can be done more wisely and the results can have more facts of interests for the maritime experts.

* Corresponding author. Tel.: +46 705383338.

[1] Integrating Maritime surveillance, common information sharing environment (cise), www.ec.europa.eu/maritimeaffairs/policy/integrated/maritime/surveillance/documents/integrating/maritime/surveillance/en.pdf.

[2] United Nations Conference on Trade and Development (UNCTAD), Review of Maritime Transport 2011, www.unctad.org/en/Docs/rmt2011/en.pdf.

[3] National plan to achieve maritime domain awareness for the national strategy for maritime security, www.dhs.gov/xlibrary/assets/HSPD_MDAPlan.pdf.

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.04.029

1.1. Contribution

This article contributes with a deeper understanding of open data as a complementary resource for establishing maritime surveillance operations. It provides a framework for anomaly detection based on the integration of open and closed data sources in the maritime surveillance domain. According to the framework, an anomaly detection system is developed which employs suitable algorithms to implement expert rules for detecting anomalies. Finally, this article contributes with a real-world validation of the developed anomaly detection system. The validation was performed by officers from the Swedish coastguard.

1.2. Outline

The remainder of this work is organized as follows: Section 2 reviews the background and related work regarding the open data and anomaly detection in the maritime surveillance domain. Sections 3 and 4 present the identified open data sources and describe the case study. The framework design and implementation described in Sections 5 and 6. Section 7 presents the system verification results and the validation results are shown in Section 8. Section 9 features a detailed discussion about the obtained results. Finally, Section 10 concludes the research with a discussion on the possible directions for future work.

2. Background

The idea behind open data has been established for a long time. Open data can be used in a variety of domains and can be obtained from any resource. The two major sources of open data are the open data in science and the open data in government. The long-standing concept of open data in science tries to overcome the difficulties in the current system of scientific publishing such as the inability to access data or usage limitation that is applied by the publishers or data providers (Molloy, 2011). Different groups, individuals and organizations are gathered to participate in a movement toward reforming the process of scientific publication (Molloy, 2011). One of the outcomes of the open data movement in science is the online availability of large number of scientific datasets for the public by different organizations. As well as the open data movement in science, governments for over a decade attempt to publish government data online and make them publicly accessible, readily available, understandable and usable (Alonso et al., 2009). The sharing of government data with the public can provide openness and transparency to citizens. It can also improve the degree of participation in the society activities and the efficiency and effectiveness of the government services and the operations within and between the governments (Dietrich et al., 2009).

According to one estimation (Dedijer & Jéquier, 1987), 90% of all information is open source, 9% is grey information (such as preprints of scientific articles, rumours in business circles, project proposals submitted to a research-funding agency, discussions with well-informed specialists, etc.), 0.9% is secret and 0.1% is non-existent information (i.e. the information you have, but you are not aware of it). Considering the large ratio of the open data sources, there should be a great value in using them in different domains. In the maritime surveillance systems, the majority of the exploited data are obtained from the confidential sources. However, in recent years the new concept of the Web, which takes the network as a platform for information sharing, interoperability and collaboration, has created new sources of data for maritime surveillance. There are organizations and communities that provide their maritime related data online and make them accessible for the public. Therefore, it would be beneficial for the maritime surveillance systems if they can take advantage of the open data to increase the safety and security in their surveillance area.

2.1. Terminology

Anomaly detection is widely used in the areas such as video surveillance, network security and military surveillance. Chandola, Banerjee, and Kumar (2009) define anomaly detection as: *The problem of finding patterns in data that do not conform to expected behavior.*

Depending on the domain of study, the non-conforming patterns are called by different names such as anomalies, outliers, exceptions, etc. In the maritime surveillance domain, these non-conforming patterns are referred as anomalies. Defense research and development Canada (Roy, 2008) provides the following definition for the term anomaly in the context of the maritime surveillance domain: *Something peculiar (odd, curious, weird, bizarre, atypical) because it is inconsistent with or deviating from what is usual, normal, or expected, or because it is not conforming to rules, laws or customs.*

The term *open data* refers to the idea of making data freely available to use, reuse or redistribute without any restriction. The open data movement follows the other open movements such as *open access* and *open source*. According to the Open Knowledge Foundation,[4] a community based organization that promotes open knowledge (whether it is content, data or information-based), an open work should be available as a whole, with a reasonable reproduction cost, preferably downloading via the Internet without charge and in a convenient and modifiable form. Furthermore, it should be possible to modify and distribute the work without any discrimination against persons, groups, fields or endeavour. In the scope of this study, the open data term refers to the publicly available data that may or may not require free registration.

2.2. Related work

In recent years, the number of studies that address the use of anomaly detection in the maritime surveillance domain is increasingly growing. Anomaly detection techniques are divided into two groups, namely data-driven and knowledge-driven approaches. There are a couple of works that proposed knowledge-based anomaly detection systems with different representation techniques and reasoning paradigms such as rule-based, description logic and case-based reasoning (Guyard & Roy, 2009; Nilsson, van Laere, Ziemke, & Edlund, 2008; Roy, 2010). A prototype for a rule-based expert system based on the maritime domain ontologies was developed by Edlund, Gronkvist, Lingvall, and Sviestins (2006) that could detect some of the anomalies regarding the spatial and kinematic relation between objects such as simple scenarios for hijacking, piloting and smuggling. Another rule-based prototype was developed by Defence R& D Canada (Roy, 2008, 2010). The aforementioned prototype employed various maritime situational facts about both the kinematic and static data in the domain to make a rule-based automated reasoning engine for finding anomalies. One of the popular data-driven anomaly detection approaches is the Bayesian network (Fooladvandi, Brax, Gustavsson, & Fredin, 2009; Johansson & Falkman, 2007; Lane, Nevell, Hayward, & Beaney, 2010). Johansson and Falkman (2007) used the kinematic data for creating the network; however, in the work that was done by Fooladvandi et al. (2009) expert's knowledge as well as the kinematic data was used in the detection process. Moreover Lane et al. (2010) presented the detection approaches for five unusual vessel behaviors and the estimation of the overall threat was

[4] Open definition, opendefinition.org/okd/.

performed by using a Bayesian network. Unsupervised learning techniques have been widely used for data-driven anomaly detection such as Trajectory Clustering (Dahlbom & Niklasson, 2007), self organizing map (Riveiro, Falkman, & Ziemke, 2008) and fuzzy ARTMAP neural network (Rhodes, Bomberger, Seibert, & Waxman, 2005). Some statistical approaches, such as Gaussian mixture model (Laxhammar, 2008), hidden Markov model (Andersson & Johansson, 2010), adaptive kernel density estimator (Ristic, La Scala, Morelande, & Gordon, 2008) and precise/imprecise state-based anomaly detection (Dahlbom & Niklasson, 2007) have been used in this context. The majority of the works that have been done in the context of anomaly detection only used transponder data from the Automatic Identification System (AIS).

There are a number of studies that employed data fusion techniques to fuse data from different sensors in anomaly detection systems (Carthel, Coraluppi, & Grignan, 2007; Guerriero, Willett, Coraluppi, & Carthel, 2008; Rhodes, Bomberger, Seibert, & Waxman, 2006; Vespe, Sciotti, Burro, Battistello, & Sorge, 2008). In these studies, the surveillance area was restricted to the coastal regions and the combination of data from AIS, synthetic aperture radar, infra-red sensors, video and other types of radar was used in the fusion process to obtain the vessel tracks. Furthermore, there are some other works that focused on the fusion of both sensor and non-sensor data (Andler et al., 2009; Ding, Kannappan, Benameur, Kirubarajan, & Farooq, 2003; Fooladvandi et al., 2009; Lefebvre & Helleur, 2004; Mano, 2010; Riveiro & Falkman, 2009). For example Lefebvre and Helleur (2004) and Riveiro and Falkman (2009) treated the expert's knowledge as the non-sensor data. Riveiro and Falkman (2009) introduced a normal model of vessel behavior based on AIS data by using self organizing map and a Gaussian mixture model. According to the model, the expert's knowledge about the common characteristic of the maritime traffic was captured as if – then rules and the anomaly detection procedure was supposed to find the deviation from the expected value in the data. Lefebvre and Helleur (2004) fused radar data with user's knowledge about the vessels of interests. The sensor data were modelled as track and the non-sensor data were modelled as templates. The track-template association was done by defining mathematical models for tracks and using fuzzy membership functions for association possibilities. Mano (2010) proposed a prototype for the maritime surveillance system that could collect data from different types of sensors and databases and regroup them for each vessel. Sensors like AIS, high frequency surface wave radar and classical radars and databases such as environmental database, Lloyd's Insurance and TF2000 Vessel database were included in this prototype. By using multi-agent technology an agent was assigned to each vessel and anomalies could be detected by employing a rule-based inference engine. When the combination of anomalies exceeded a threshold, vessel status was informed to the user as an anomaly. The work presented by Ding et al. (2003), proposed the architecture of a centralized integrated maritime surveillance system for the Canadian coasts. Sensors and databases included in this architecture were: high frequency surface wave radar, automatic dependant surveillance reports, visual reports, information sources, microwave radar, and radar sat. A common data structure was defined for storing data that were collected from different sensors. Andler et al. (2009), also described a conceptual maritime surveillance system that integrated all available information such as databases and sensor systems (AIS, long-range identification and tracking, intelligence reports, registers/databases of vessels, harbours, and crews) to help user to detect and visualize anomalies in the vessel traffic data in a worldwide scale. Furthermore, the authors suggested using open data in addition to other resources in the fusion process.

In conclusion, the main focus of the studies that have been done in the context of anomaly detection in the maritime surveillance domain was related to using sensors data and mainly the AIS data to find anomalies in the coastal regions. Detection of some suspicious activities such as smuggling requires vessel traffic data beyond the coastal region. Maritime authorities in each country have overall information of maritime activities in their surveillance area. But exchanging information among different countries is a complicated procedure because of the diverse regulation of data protection in each land. Therefore, using data sources that are free from legislative procedures can be a good solution for providing information that belongs to the regions outside the land territory. Furthermore, all the information about maritime activities is not recorded in the authorities' databases or reported to them. On the other hand, there are numerous open data sources consists of different websites, blogs and social networks that can be useful for observing the hidden aspects of maritime activities. Hence, this article will investigate the potential open data sources for maritime activities and exploit them to build an anomaly detection system. The aim of this system is to provide complementary decision support for coastguard operators when they analyze traditional closed data sources.

3. Open data in maritime surveillance

To obtain the applicable open data for anomaly detection, the first step is initiated by reviewing the information resources document[5] provided by the International Maritime Organization. This organization is the United Nations' specialized agency with responsibility for the safety and security of shipping and the prevention of marine pollution by vessels. The document introduces 29 governmental and intergovernmental organizations that work in different fields related to the maritime surveillance domain such as maritime safety, prevention of pollution from vessels, liability and insurance issues, shipping information, etc. All these 29 organizations' websites and the links provided by each of them are investigated and a list of online data sources is prepared. The obtained open data sources provide AIS data, information about vessel characteristics, ports, maritime companies, suppliers, weather, etc. Moreover, in the process of finding open data sources, an attempt is made to obtain sources of data that are related to the Baltic region and mostly Sweden by use of the previously observed data sources and also common search engines. This extensive collection of open data sources for the maritime surveillance domain is available for download.[6]

4. Case study

The case study is the employed research method in this article. The two important sources of information about maritime anomalies are reports of the workshops that were held in Canada (Roy, 2008) and Sweden (Andler et al., 2009; van Laere & Nilsson, 2009). In these two workshops attendees were experts in the maritime domain and a variety of maritime anomalies were identified.

According to the identified anomalies by the two workshops, a list of some potential maritime anomalies that can be detected by use of the available open data sources was prepared in the preliminary work of this study. Then, in a meeting with representatives of the Swedish coastguard the types of anomalies that are of high interest for the coastguard operators and the possibility of using open data for anomaly detection were discussed. During the

[5] Information resources on maritime security and ISPS code, www.imo.org/knowledgecentre/informationresourcesoncurrenttopics/maritimesecurityandispscode/documents/informationresourcesonmaritimesecurityandispscode.pdf.

[6] Open maritime-specific data collection, http://www.bth.se/com/nla.nsf/sidor/resources.

Table 1
Anomalies detectable through open data (confirmed by the coastguard).

No.	Expert rules	Anomaly
1	If a vessel destination does not exist in the port schedule then anomaly	VESSEL_NOT_INFORMED_PORT (A1)
2	If a vessel ETA does not match with the port ETA for the vessel then anomaly	ARRIVAL_TIME_MISMATCHED (A2)
3	If a vessel entered a port without informing the port then anomaly	VESSEL_ENTERED_PORT_ WITHOUT_NOTICE (A3)
4	If a vessel has requested a pilot but has not used the service then anomaly	VESSEL_NOT_USED_PILOT (A4)
5	If vessel A which normally travels between ports X and Y, suddenly goes to port Z then anomaly	UNUSUAL_TRIP_PATTERN (A5)
6	If a vessel has not left a port according to the port schedule then anomaly	VESSEL_NOT_LEFT_PORT (A6)
7	If a vessel exists in a port schedule but it has not entered the port then anomaly	VESSEL_NOT_ENTERED_PORT (A7)
8	If a vessel does not exist in the port schedule and the vessel has requested a pilot then anomaly	VESSEL_ORDERED_PILOT_ AND_NOT_INFORMED_PORT (A8)
9	If a vessel has moored in a port and has been observed somewhere else then anomaly	VESSEL_MOORED_IN_PORT (A9)
10	If vessel A has not entered a port according to the port schedule instead vessel B enters the port at the same time slot then anomaly	WRONG_VESSEL_ENTERED (A10)
11	If a vessel with the laid up status has been observed somewhere else then anomaly	VESSEL_LAID_UP (A11)

Note. ETA = estimated time of arrival.

meeting, the prepared list of anomalies was reviewed. Coastguard operators were asked about the possibility of the historical occurrence, and their degree of interest, for each anomaly. As an outcome of the meeting a number of scenarios were created and based on these scenarios, 11 expert rules were defined.

The first scenario refers to the anomalies related to the vessel static information such as name, owner, International Maritime Organization number, dimensions, type and the status (in service or laid up). For example, sailing a vessel with a draught of 22 meters over an area with a 9 meter depth or observing a vessel that should be laid up or changing the name or the owner of a vessel during its voyage indicate the existence of suspicious activities.

The second scenario is related to the prior arrival notification for vessels. Vessels should inform their arrival time to the ports at least 24 h in advance. Each port also provides an online timetable for the incoming vessels. Therefore, any mismatch between the reported AIS data regarding the destination or the arrival time of a vessel and the destination port timetable needs to be checked by the coastguards.

The third scenario is related to ordering pilots. Usually, large vessels because of their size and weight need to be guided by pilots through dangerous and congested waters. Therefore, vessels need to submit their request for a pilot and also inform the destination port. However, in some cases vessels order a pilot without informing the port. Such situations should be investigated.

The case study in this article comprises scenarios two and three, which are conducted in close collaboration with operators from the Swedish coastguard. The aim of the case study is to investigate the potential of open data about maritime activities (vessels, ports, and so on) as a complement to the closed data sources that are already used by the Swedish coastguard. They key questions posed in this study are:

- How can open data complement closed data for anomaly detection in the maritime surveillance domain?
- What is the positive and negative impacts of the open data sources? (that is, what is the increase in true negatives and positives in comparison to the increase in false negatives and positives?)

In the next meeting, the scenarios and the rules are presented to the representatives of the Swedish coastguard and they are asked to comment or suggest new scenarios or rules. By getting the final approval from the coastguard experts, one new rule (rule number 5) is added to the list. Table 1 shows the admitted rules by the experts. These identified maritime anomalies can be detected by use of AIS data, vessel traffic timetables in ports and pilots websites and the vessel characteristic data that are available in data sources such as Lloyd's. A capitalized name is provided to each anomaly that can be detected by the rules, and for the remainder of this article the anomalies will be referred to by these names.

5. Framework design

A new maritime surveillance framework and expert-based decision support system is presented in this article. The Open Data Anomaly Detection System (ODADS) is designed for traffic monitoring and detecting anomalies in the maritime domain by using open and closed data sources. Fig. 1 depicts the ODADS architecture. The framework is designed to be generalizable to similar applications in other domains; that is, for applications where the objective is to identify anomalous behavior through semi-automatic methods. The proposed framework is designed to provide decision-support based on knowledge-engineering-based or knowledge-discovery-based methods. It is focused on the extraction of information from open data sources. The setup of any implementation of the framework depends largely on the problem at hand. ODADS consists of three core modules:

1. Data Collector,
2. Anomaly Detector and,
3. Display Client.

The Data Collector module is responsible for collecting open data from the Internet, and for preprocessing and storing the data in the system database. The data can be related to vessel traffic (such as AIS reports, ports and pilots timetables), vessel characteristics, ports equipments and facilities, companies that are involved in maritime activities, news or reports about maritime events and activities available in different social media platforms (such as blogs and social networks), and so on. The Data Store comprises a set of databases that contain data belonging to different types of sensors, authorized databases and open data sources. The data in the Data Store can be fused or integrated before being used in the detection process. When the Data Collector completes its task, the Anomaly Detector becomes available. The Anomaly Detector module analyses the available open and closed data and detects possible anomalies by using both knowledge-driven and data-driven techniques. Different anomaly detection techniques are employed due to the distinct nature of anomalies and the complexity of the environment in the maritime surveillance domain. Previously known anomalies can be detected by knowledge-based techniques but in real-world situations it is desirable that an anomaly detection system can detect previously unseen anomalies as well. One of the potential benefits of using

Table 2
Confusion matrix for the nine classes of anomalies and the normal class.

		Predicted class										
		A1	A2	A3	A4	A5	A6	A7	A8	A9	Normal	Total
Actual class	A1	6	–	–	–	–	–	–	–	–	–	6
	A2	–	7	–	–	–	–	–	–	–	1	8
	A3	–	–	–	–	–	–	–	–	–	–	0
	A4	–	–	–	–	–	–	–	–	–	–	0
	A5	–	–	–	–	3	–	–	–	–	–	3
	A6	–	–	–	–	–	1	–	–	–	–	1
	A7	–	–	–	–	–	–	–	–	–	–	0
	A8	–	–	–	–	–	–	–	–	–	–	0
	A9	–	–	–	–	–	–	–	–	–	–	0
	Normal	–	–	–	–	–	–	–	–	–	192	192
	Total	6	7	0	0	3	1	0	0	0	193	17

Table 3
Validation results of the coastguard.

Anomaly	Alarms		Not checked
	True	False	
VESSEL_NOT_INFORMED_PORT	7	3	4
ARRIVAL_TIME_MISMATCHED	19	5	3
VESSEL_NOT_USED_PILOT	2	–	–
UNUSUAL_TRIP_PATTERN	7	6	–
VESSEL_NOT_LEFT_PORT	1	1	–
VESSEL_ORDERED_PILOT_AND_ NOT_INFORMED_PORT	2	–	–
VESSEL_MOORED_IN_PORT	–	3	–
UNDER_SURVEILLANCE_VESSEL	1	–	–
UNUSUAL_TRIP_PATTERN_AND_ VESSEL_NOT_ INFORMED_PORT	5	1	–
UNUSUAL_TRIP_PATTERN_AND_ VESSEL_ ARRIVAL_TIME_ MISMATCHED	4	–	–
UNUSUAL_TRIP_PATTERN_AND_ VESSEL_ORDERED_PILOT_ AND_NOT_INFORMED_PORT	–	1	–
VESSEL_ORDERED_PILOT_AND_ NOT_INFORMED_PORT_ AND_VESSEL_NOT_USED_PILOT	1	–	–
Total count	49	20	7
Total count (%)	64.47	26.32	9.21

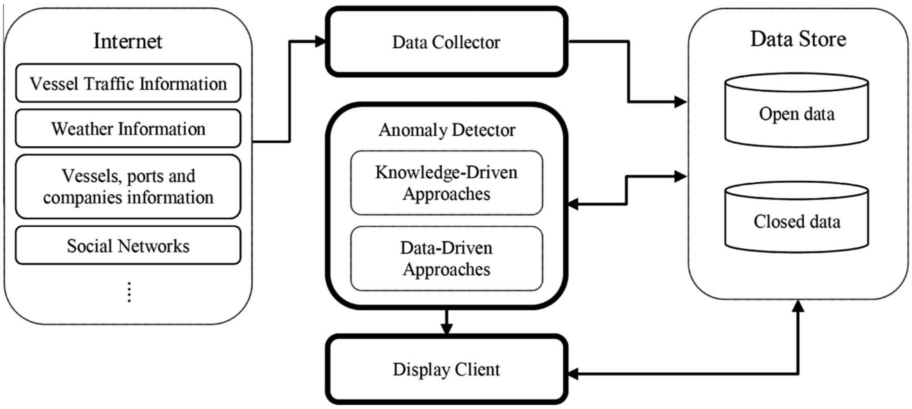


Fig. 1. The Open Data Anomaly Detection System (ODADS) architecture. The Data Collector module collects data from the Internet and stores them in the database. The Anomaly Detector module detects anomalies by taking advantage of different techniques. The Display Client module displays the detected anomalies to the user and enables system-user interaction.

data-driven methods such as machine learning algorithms is the possibility of detecting such unseen anomalies. However, it is difficult to evaluate how well today's data-driven systems manage to detect previously unseen cases. The proposed system in this article is deterministic and completely expert-based but our proposed framework general enough to allow data-driven detection methods in other applications. The Display Client module is the user interface of the system. This module represents the cognitive refinement level (Level 5) of the Joint Directors of Laboratories model. It is argued that the effectiveness of a system can be affected by the way that the system produced information is comprehended by the human user (Hall, Hall, & Tate, 2000). The cognitive refinement process involves traditional human computer interaction utilities such as geographical display or advanced methods that support functionalities such as cognitive aids, negative reasoning enhancement, focus/defocus of attention and

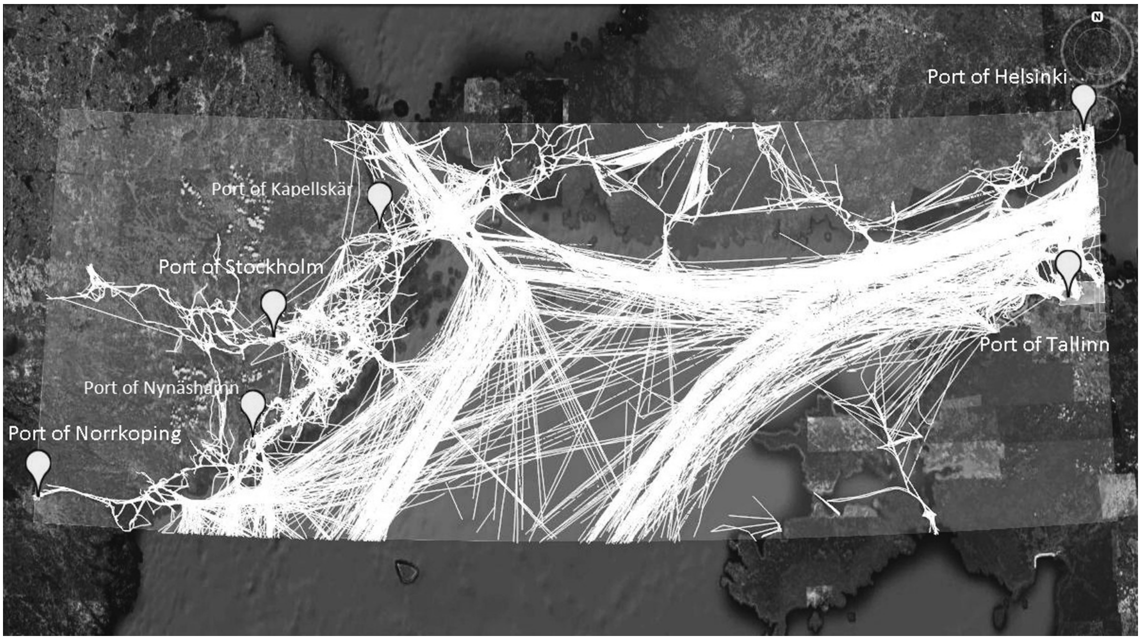


Fig. 2. The surveillance area and the vessels tracks extracted from AIS data for 6 days. The area is restricted to the north of the Baltic Sea and part of the Gulf of Finland. Ports from left to right are Norrköping, Nynäshamn, Stockholm, Kapellskär, Helsinki and Tallinn (the image is generated by Google Earth).

representing uncertainty. While designing the user interface, the six principles of user interface design that are based on the usage-centered design approach are considered. These six principles are: structure, simplicity, visibility, feedback, tolerance, and reuse (Constantine & Lockwood, 1999).

6. Implementation

ODADS is implemented by taking advantage of the identified maritime anomalies and the obtained open data sources. To limit the scope, four types of vessels (passenger, ferry, cargo, and tanker) are considered. Other types of vessels (such as fishing and sailing vessels) are omitted. Secondly, the rule related to the vessel static information is ignored. Furthermore, the WRONG_VESSEL_ENTERED anomaly is excluded due to its complexity. Moreover, in further collaboration with the coastguard representatives during the implementation phase, a new type of anomaly is proposed. This anomaly is called UNDER_SURVEILLANCE_VESSEL and occurs when a vessel of interest has any of the A1–A9 anomalies and the vessel exists in the vessels blacklist.

6.1. Data description

The required vessel traffic data can be obtained from AIS reports and ports and pilots timetables. The surveillance area is restricted to the north of the Baltic Sea and a part of the Gulf of Finland, the regional area between three European countries Sweden, Finland and Estonia. Fig. 2 shows the surveillance area where the geographic coordinates lie between latitudes 58.49–60.24° N and longitude 16.19–25.00° E. This region is one of the high-traffic regions in the Baltic Sea and is surrounded by the four highly used ports. The selected ports are: Stockholm group (Stockholm, Kapellskär and Nynäshamn) and Norrköping ports in Sweden, Helsinki port in Finland and Tallinn port in Estonia. Due to inaccessibility to the raw AIS data (a closed source) in the surveillance area, the AIS reports that are provided by the *MarineTraffic.com* website are exploited. These reports consist of both static and dynamic types of data for each vessel during its voyage such as name, type, year built, flag, call sign, maritime mobile service identity, International Maritime Organization identification number, origin, destination, Estimated Time of Arrival, speed (maximum and average), position (longitude and latitude), and heading. The pilot data belong to the Stockholm pilotage area in Sweden. Moreover, a common data representation format for ports and pilots data is defined that contains vessel name, vessel type, origin, destination, company name, vessel status and arrival/departure time. Table 4 in the appendix provides more details about these sources.

6.2. Detection methods

After investigating the nature of the anomalies and the potential techniques, it is determined that except for the UNUSUAL_TRIP_PATTERN anomaly, detection of other anomalies can be done by performing a search in the data for finding the desired match. If the match is not found then the vessel would be marked as an anomaly. Using exact string matching techniques for comparing vessel information from different data sources is inapplicable due to the potential errors that might occur because of different notations or human operator mistakes during data entry. Therefore, a metric should be used for measuring the degree of similarity between two vessels from different sources. After investigating the performance of different string matching techniques on the available data, the *JaroWinkler*[7] metric is chosen. For two strings if the JaroWinkler distance is less than or equal to a predefined threshold then the two strings are considered similar.

Detection of the UNUSUAL_TRIP_PATTERN anomaly requires data-driven approaches such as machine learning or statistical techniques. Unlike the other anomalies, detection of this anomaly requires a history of vessel traffic data for training the system. For this reason, data related to six months (September 15, 2011–March 15, 2012) of vessel traffic in the surveillance area are gathered. By monitoring the activities during this period, the system will be able to find the normal pattern of vessels trips in the area

[7] JaroWinkler (the variant of Jaro) measures the number and order of common characters in the two strings and also the number of transposition that needs to change one of the strings to the other.

Table 4
Data sources that are used in the implementation.

Website name	Data type
Marinetraffic.com	Real time information based on AIS systems[a]
Swedish Maritime Administration (Sjöfartsverket)	Stockholm pilotage area[b]
Ports of Stockholm, Kapellskär and Nynäshamn	Vessels in port and expected arrival[c]
Port of Norrköping	Vessels in port[d] Expected vessels arrival[e]
Port of Helsinki	Cargo vessels in port[f] Expected cargo vessels arrival[g] Expected passenger vessels departure[h] Expected passenger vessels arrival[i] Passenger vessels have visited the port before[j]
Port of Tallinn	Vessels in port[k] Expected cargo vessels arrival[l] Expected passenger vessels arrival[m] Expected passenger vessels departure[n]

[a] www.marinetraffic.com/ais/.
[b] www.sjofartsverket.se/sv/Infrastruktur-amp-Sjotrafik/Lotsning/Lotsinfo/.
[c] stockholmshamnar.se/en/Karta/Vessel-calls/.
[d] www.norrkoping-port.se/anlop.php?page=snabb_fih&link=110—111.
[e] www.norrkoping-port.se/anlop.php?page=snabb_fih_ank&link=110—111.
[f] www.portofhelsinki.fi/cargo_traffic/vessels_in_ports.
[g] www.portofhelsinki.fi/cargo_traffic/arrival_ships.
[h] www.portofhelsinki.fi/passengers/departure_times_and_terminals.
[i] www.portofhelsinki.fi/passengers/arrival_times_and_terminals.
[j] www.portofhelsinki.fi/passengers/cruise_ships_that_have_visited_the_port.
[k] www.ts.ee/?op=ships_in_port&lang=eng.
[l] www.ts.ee/?op=cargo_ships_arrivals&lang=eng.
[m] www.ts.ee/?op=passenger_ship_arrivals&lang=eng.
[n] www.ts.ee/?op=passenger_ship_departures&lang=eng.

of interest. For detecting this anomaly a simple statistical approach is used. A look up table is created and for each vessel the number of times that the vessel travels between two different ports is stored. For each vessel, if the frequency of travelling between its origin and destination is less than a predefined threshold (which is 2 in the current implementation) then the vessel will be reported as an anomaly.

There are situations that multiple anomalies can occur in the same time for a specific vessel. The combinations of anomalies that don't have any features in common are defined as new types of anomalies. For instance, a vessel has not informed its arrival to the destination port and its trip to the port is not common, in such situation a new type of anomaly UNUSUAL_TRIP_PATTERN_AND_VESSEL_NOT_INFORMED_PORT is defined.

7. System verification

To ensure that ODADS works properly, the system is tested manually with both real and manipulated data. The tests are performed during the implementation and also after completing the system. At first, a number of vessels with different types of anomalies are inserted to the real collected data to check whether all types of anomalies can be detected by ODADS. Then, the system is run for a period of time and the detected anomalies are checked manually against the available data to make sure about their correctness. A screenshot of the ODADS visualization of the maritime environment is shown in Fig. 3. During the test phase the Anomaly Detector module is updated and some of the detection conditions are narrowed down. The process is repeated until the system can detect all the anomalies correctly. Furthermore, before using ODADS in real-world situations, it is important to figure out to what extent the results of the system are accurate. For this reason, an experiment is conducted to measure the accuracy. Accuracy is the degree to which the estimates or measurements of a quantity correctly describe the exact value of that quantity. In other words, accuracy is the proportion of true results in the population. To evaluate the system accuracy, the number of True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) are needed. Accuracy is calculated by the following formula (Han, Kamber, & Pei, 2011):

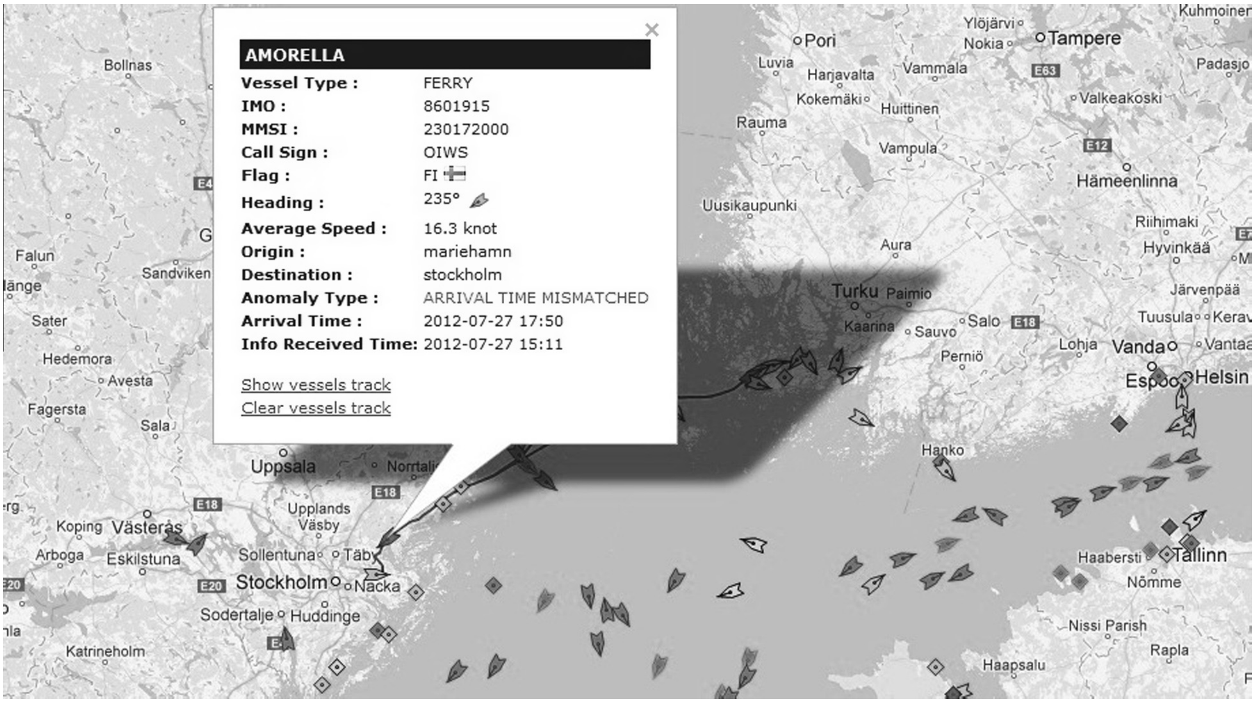
$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

The first step in designing the experiment is to identify the population. The population consists of the vessel traffic data in the surveillance area. Since the population is too large and it is impossible to look into all members manually to count the number of TP,FP,TN and FN, a sample should be taken from the population. The sampling frame is the vessel traffic data related to AIS, ports and pilots in the surveillance area, which are provided by ODADS. Due to the high volume of traffic through the surveillance area, it is expected that the majority of anomalies can be observed in one week execution of the system. Therefore, one week of vessel traffic data from April, 2012 is used as the sample frame (Tables 5 and 6 in the appendix present some information about the vessels traffic and detected anomalies during this week). A two stage random sampling is used in order to have unbiased and independent samples. In the first stage, a simple random sampling without replacement is done for selecting the time slots that ODADS attempts to collect and analyse the data. After selecting the time slots, the corresponding data for each time slot will be selected by a stratified sampling. Three strata are defined according to the type of vessels: ferry and passenger, cargo and tanker vessels. Selection of vessels is also limited to the vessels that are originated from or targeted to the four particular ports. The total number of time slots in the sample frame is 835. This means that on average, ODADS collects data 139 times a day. In the first stage of sampling a random timeslot is selected for each day, which results in 7 time slots for one week. Then, by considering the described limitation in the selection process, the average number of entire data in a selected time slot is about 100 records. Among these records, 30 records are selected by stratification. Almost 73% of the vessels in each time slot are moored. Since the majority of the anomalies are related to the vessels trips, a limitation on the number of moored vessels in the samples is defined. In this way, it can be possible to check more anomalies in the evaluation process. The second stage of sampling is repeated by taking into consideration that the number of moored vessel in the sample cannot exceed from the half of the sample size (in this case, 15).

After carrying out the sampling, all the samples are checked against the primary identified anomalies ($A_1 - A_9$). To compute the number of TP,FP,TN and FN, a confusion matrix is created based on the nine classes of anomalies and the normal class (Table 2). According to the matrix, the accuracy of the system is: 17 + 192/ 17 + 192 + 0 + 1 = 0.99. The existing FN for the ARRIVAL_TIME_MISMATCHED anomaly is due to the wrong provided AIS data by the vessel or possibly the MarineTraffic.com website and also the limitation of the system for considering all conditions. In this case, the vessel arrival time belongs to a couple of days before the current date and for this reason it is ignored by the system. However, it is quite possible to handle such situations if additional sources of AIS data are available.

8. System validity

The validation was made by an officer at the coastguard Headquarters office in Karlskrona, Sweden for four weeks (April 23,



AMORELLA

Vessel Type :	FERRY
IMO :	8601915
MMSI :	230172000
Call Sign :	OIWS
Flag :	FI
Heading :	235°
Average Speed :	16.3 knot
Origin :	mariehamn
Destination :	stockholm
Anomaly Type :	ARRIVAL TIME MISMATCHED
Arrival Time :	2012-07-27 17:50
Info Received Time:	2012-07-27 15:11

Fig. 3. A screenshot of the ODADS visualization of the maritime environment.

Table 5
Total number of vessels in the surveillance area during one week of the system execution.

	Count (Avg)[1]
Total number of vessels	673.29
Cargo, Tanker, Passenger and Ferry vessels	366.29
Cargo, Tanker, Passenger and Ferry vessels that are originated from or targeted to the specified ports	141.71

[1] The daily average count.

2012–May 18, 2012), at any time during working hours (08:00–17:00). The officers are supposed to evaluate the detected anomalies by checking them against the available data in the systems and data sources that are used during the normal operational activities at the coastguard. They are asked to provide weekly report about their evaluation results in order to decrease the possible malfunctioning of the system and the validation process. A detected anomaly for a vessel is true if it can be confirmed by the available data sources at the coastguard and consequently it is false if the authorized data sources provide any information that declines the detected anomaly. No further assessment is done regarding the classification of the detected anomalies to true and false alarms.

The sea monitoring system that is used by the coastguard officer is called SJöBASIS.[8] SJöBASIS aggregates the maritime data from different systems and agencies with the aim of improving the efficiency of maritime surveillance. In SJöBASIS, the required data that contain vessel position, speed, heading, arrival/departure time and trip, are obtained from the following sources SafeSeaNet,[9] SjöC, local AIS and HELCOM AIS.[10] The officer checks the validity of anomalies according to the priority that each anomaly has for him. During the four-week validation period, ODADS is used at the coastguard for 12 working days and in total 76 of the detected anomalies are evaluated. Table 3 presents the validation results. Among the evaluated anomalies, there are a number of anomalous vessels that remain unchecked due to a lack of corresponding data in the coastguard systems. A large number of detected anomalies are related to the ARRIVAL_TIME_MISMATCHED anomaly that in many cases can be due to the inconsistent time formats in different data sources and various settings in the AIS transmitters. In these cases, the detected anomalies by ODADS are correct detections based on the available data but the real-world situation that the data are used represent is considered normal. For example, a vessel that is going from Helsinki to Stockholm is reporting its arrival time according to the Finland local time instead of using the coordinated universal time format. Therefore, this artificial time difference results in that the arrival time reported by the vessel does not match the expected arrival time of the vessel at the destination port, which leads to set the ARRIVAL_TIME_MISMATCHED anomaly for that vessel. In addition to the comparative analysis in the validation process, modus operandi of using an Anomaly Detector that takes advantage of open data is investigated. The coastguard officer uses an analysis tool[11] to analyse the ODADS excel reports and draw conclusions regarding the modus operandi of the system in the emergency situations that can have impact on the maritime surveillance operations. One of the possible analyses of the system reports can be the investigation of vessels with multiple anomalies. Here are some examples of the vessels with multiple anomalies. A cargo vessel has recurring anomalies related to the arrival/departure time, trip and notification to the port. From the types of anomalies that are detected for this vessel, it can be concluded that the vessel behavior points to a higher threat concerning customs and border. For a passenger vessel the following anomalies are often detected: VESSEL_NOT_INFORMED_PORT, ARRIVAL_TIME_MISMATCHED, VESSEL_NOT_ENTERED_PORT and

[8] www.kustbevakningen.se/sv/granslos-samverkan/sjoovervakning-suppdraget/samverkan-sjoinformation/.

[9] www.emsa.europa.eu/operations/maritime-surveillance/safeseanet.html.

[10] www.helcom.fi/BSAP/ActionPlan/en_GB/SegmentSummary/.

[11] IBM i2 Analyst's Notebook, www.i2group.com/us/products/analysis-product-line/ibm-i2-analysts-notebook.

Table 6
The average number of detected anomalies during one week of execution.

Anomaly	Count (Avg.)[1]
ARRIVAL_TIME_MISMATCHED	23.86
VESSEL_NOT_INFORMED_PORT	14.71
VESSEL_NOT_LEFT_PORT	8.29
UNUSUAL_TRIP_PATTERN	3.71
VESSEL_NOT_USED_PILOT	2.00
UNUSUAL_TRIP_PATTERN_AND_VESSEL_ARRIVAL_TIME_MISMATCHED	1.71
VESSEL_MOORED_IN_PORT	1.29
UNUSUAL_TRIP_PATTERN_AND_VESSEL_NOT_INFORMED_ PORT	0.57
VESSEL_ENTERED_PORT_WITHOUT_NOTICE	0.29
VESSEL_NOT_ENTERED_PORT	0.29
UNDER_SURVEILLANCE_VESSEL	0.29
VESSEL_ARRIVAL_TIME_MISMATCHED_AND_VESSEL_NOT_USED_PILOT	0.29
VESSEL_ORDERED_PILOT_AND_NOT_INFORMED_PORT	0.14
UNUSUAL_TRIP_PATTERN_AND_VESSEL_NOT_USED_PILOT	0.14
VESSEL_ORDERED_PILOT_AND_NOT_INFORMED_PORT_AND_VESSEL_NOT_USED_PILOT	0.14
UNUSUAL_TRIP_PATTERN_AND_VESSEL_ORDERED_PILOT_AND_NOT_INFORMED_PORT	0.00
UNUSUAL_TRIP_PATTERN_AND_VESSEL_NOT_USED_PILOT_AND_VESSEL_ARRIVAL_TIME_MISMATCHED	0.00
VESSEL_ENTERED_PORT_WITHOUT_NOTICE_AND_NOT_LEFT_PORT_ON_TIME	0.00
VESSEL_NOT_ENTERED_PORT_AND_NOT_USED_PILOT	0.00
UNUSUAL_TRIP_PATTERN_AND_VESSEL_NOT_ENTERED_ PORT	0.00
UNUSUAL_TRIP_PATTERN_AND_VESSEL_NOT_ENTERED_PORT_AND_VESSEL_NOT_USED_PILOT	0.00
UNUSUAL_TRIP_PATTERN_AND_VESSEL_ORDERED_PILOT _AND_NOT_INFORMED_PORT_AND_VESSEL_NOT_USED_PILOT	0.00

[1] The daily average count.

VESSEL_NOT_LEFT_PORT. These anomalies may happen because of inaccurate or wrong provided data by the vessel. The conclusion and modus operandi for this vessel is that the duty officer will contact the vessel to highlight the importance of submitting accurate information. In real-world situations, the occurring anomalies such as VESSEL_NOT_ENTERED_PORT or VESSEL_NOT_LEFT_PORT for a passenger vessel can be related to serious issues such as an accident and it will result in increased scrutiny for that vessel. It is also possible to look into the relation between the types of vessels and the detected anomalies. Such assessment can be used for strategic and risk analysis. For tanker vessels the most common anomaly is VESSEL_NOT_INFORMED_PORT. This anomaly has a high priority for emergency preparedness for accidents involving tankers. Tankers with UNUSUAL_TRIP_PATTERN anomaly are a potential risk to other vessels and can cause accidents. From a risk assessment point of view the combination of this anomaly with the VESSEL_ NOT_ ENTERED_PORT or VESSEL_NOT_USED_PILOT anomalies can lead to high-risk situations. The most occurring anomaly for ferries and passenger vessels is ARRIVAL_TIME_MISMATCHED. In several cases this anomaly is detected incorrectly because of the wrong reported arrival time; however, this anomaly has great importance for the authorities to plan their operations regarding ferries and passenger vessels arrivals effectively. The most serious anomaly for ferries and passenger vessels is VESSEL_NOT_LEFT_PORT and the authorities should suspect that some form of accident or difficulty is arising regarding to the departure of the vessels. For cargo vessels, the most recurring anomalies are VESSEL_NOT_INFORMED_PORT and ARRIVAL_TIME_MISMATCHED. However, some of the ARRIVAL_TIME_MISMATCHED anomalies are false alarms because of the incorrect data. The VESSEL_NOT_INFORMED_PORT anomaly is important for ports security and safety. The prior notification to the ports is obligatory for vessels, but the fine for breaking this rule is negligible which lets the vessels that are involved in illegal activities such as smuggling disobey this rule. The most serious anomalies for the cargo vessels are the VESSEL_NOT_ENTERED_PORT and UNDER_SURVEILLANCE_VESSEL anomalies.

Furthermore, looking into the most frequent anomalies for different ports will assist the maritime authorities to make their decisions more efficiently. According to the report analysis, the most frequent anomaly in Stockholm and Nynäshamn ports is ARRIVAL_TIME_MISMATCHED, in port of Kapellskär is VESSEL_NOT_INFORMED_PORT and in Norrköping port is UNUSUAL_TRIP_PATTERN. One possible conclusion from the most popular anomalies for the ports is that the port authorities should be informed of the divergence in the traffic flow and the operational management functions in order to plan and allocate resources efficiently. On the other hand, in some cases the anomalies are too commonly occurring for a port because of the inaccurate provided data and they can be disregarded by the officer.

The received feedback from the coastguard representatives during the validation process indicates that ODADS complements the closed sources and assists the human operator in gaining a better understanding of the ongoing maritime activities. The representatives believe that the ODADS results are reliable and the quality of the open data that are used is good and can be used in real-world situations. The functionality, usability, and visualization tools in ODADS provide a simple and intuitive system for coastguard operators. In addition to illustrate the vessel traffic data in a simple, clear, and informative way, ODADS can provide statements about the anomalies and its statistical reports are beneficial when the authorities and freight companies conduct strategic analysis of maritime traffic and risk assessment. Finally, the capability of automatic detection of anomalies based on open data is considered a valuable asset to the coastguard.

9. Discussion

Taking advantage of anomaly detection systems will assist authorities to tighten security in the maritime surveillance domain. There are a number of studies which focused on developing anomaly detection systems by using knowledge-driven and data-driven approaches. For instance, Defence research and development Canada (Roy, 2008, 2010) developed a rule-based prototype for anomaly detection by exploiting maritime situational facts about both kinematic and static data of the domain. Edlund et al. (2006) developed another prototype for a rule-based expert system to detect the anomalies regarding spatial and kinematic relation between objects. Riveiro and Falkman (2009) proposed using a combination of data-driven and knowledge-driven approaches to detect anomalies by use of a normal model of vessel behavior based on AIS data and experts rules. In the majority of studies that addressed anomaly detection, the exploited data for the anomaly detection process were obtained from closed data sources and there is a lack of investigation on using open data sources for anomaly detection in the maritime surveillance domain. Therefore, in this article ODADS is implemented by employing expert rules to investigate the potential open data as a complement to the closed data for anomaly detection in the maritime surveillance domain.

The validity of the system is evaluated in real-world situations by the experts from the Swedish coastguard. Despite the inaccurate nature of open data and by considering the fact that only open data sources are used in the system, the high degree of true alarms (64.47%) in the validation process admits the validity of the system outcomes. Furthermore, there are no corresponding data in the authorized databases for 9.21% of the evaluated anomalies by the coastguard. This fact refers to a potential information gap in the closed data sources. However, the considerable number of false alarms (26.32%) for a surveillance system is still unsatisfactory. The number of false alarms indicates the difference between the accuracy of the system and the validity of the results. Even though the data that are used in ODADS are obtained from relatively trusted data sources such as ports, the false alarms occur mostly because of data inaccuracies. The open data that are exploited by ODADS suffer from these errors due to human operator mistakes, irregular data updates, data update latencies, and incompatible data formats. These are critical issues that unfortunately seem to concern many open data applications. In ODADS, there are situations where a detected anomaly disappeares in the next periods of system execution because of the arrival of revised and corrected data. Frequent occurrences of false alarms distract the operator's attention from real anomalies in the surveillance area.

To decrease the false alarms in ODADS, the main solution is to integrate open and closed data, which can cover the lack of information or inaccuracy in the open data. In addition, considering a probability for the detected anomalies can decrease the number of false alarms. This would be possible by analysing the history of vessels behavior as well as the current situation and defining a probability threshold to omit the anomalies that have a lower probability than the threshold. Furthermore, having extra information regarding vessels such as crew and cargo information, can affect the probability of being a real anomaly for a specific vessel. For example, if a vessel has the ARRIVAL_TIME_MISMATCHED anomaly and it has a crew member with a criminal record or a special cargo, then there is a possibility that the vessel is stopped somewhere to exchange something. Therefore, in such situation, the probability of being a true anomaly is high. According to the validation results, the UNUSUAL_TRIP_PATTERN anomaly creates the majority of false alarms. This is due in part to the statistical approach that is used for detecting this anomaly and also the wrong origin and destination information that the vessels provide. The lookup table that is created for storing the frequency of the trips between different places is not updated periodically. While populating the table, the ports timetables are used which can be incomplete. An alternative detection approach can be to use machine learning techniques, which attempt to detect anomalies according to the pattern of movements for individual vessels instead of the reported trip data by the vessels or ports.

The proposed framework is generalizable to similar applications in other domains due to its modularized and general design. This case study has investigated the potential of open data as a complement to closed data for anomaly detection in the maritime surveillance domain. The primary stakeholders in the case study are human operators from the Swedish coastguard. Since the main purpose of the case study was to focus analysis and investigation on two defined scenarios, it is not possible to draw conclusions about the generalizability of the results. Further investigations can shed light on this generalizability by conducting large-scale trials of the implemented expert-based system across additional scenarios and more maritime surveillance areas.

In local areas such as the surveillance area in this article, mainly because of large amount of quality assured data and the limited size of the surveillance area, it is easier for the maritime authorities to track and control the vessels activities. Therefore, the use of open AIS data in this region is not required and it should be prohibited to decrease the negative impacts of open data on the system results. On the other hand, when the vessel information beyond the exclusive economic zone is required, the value of open data becomes more obvious.

10. Conclusion and future work

This article investigated the potential open data as a complementary resource for anomaly detection in the maritime surveillance domain. A framework for anomaly detection was proposed based on the usage of open data sources along with other traditional sources of data. According to the proposed anomaly detection framework and the algorithms for implementing the expert rules, the Open Data Anomaly Detection System (ODADS) was developed. The validity of the results was investigated by the subject matter experts from the Swedish coastguard. The validation results showed that the majority of the ODADS evaluated anomalies were true alarms. Moreover, a potential information gap in the closed data sources was observed during the validation process. Despite the high number of true alarms, the number of false alarms was also considerable that was mainly because of the inaccurate open data. This article provided insights into the open data as a complement to the common data sources in the MS domain and is concluded that using open data will improve the efficiency of the surveillance systems by increasing the accuracy and covering unseen aspects of maritime activities.

In the future, it is important to investigate how the open data sources in the maritime domain can be used in a global perspective. In this article, the surveillance area was limited to a local area which is fully covered by the authorities' data sources. When the data beyond the exclusive economic zone are needed, it is more valuable to use open data sources. By taking advantage of the subject matter experts' knowledge about maritime surveillance, it would be possible to figure out how the global open data should be exploited for the surveillance purpose. Integration of the open data with maritime confidential data can improve the efficiency of maritime surveillance and should be considered as a further improvement of the system. Another improvement can be considering a probability for each detected anomaly according to the history of the vessels behavior and the current situation. Moreover, further investigation on the other sources of open data such as social data, which is created and shared through social media platforms, and online videos from the ports activities in the high risk regions, will be useful. The data that are used in ODADS are relatively trusted, but in case of using other open data sources in the maritime surveillance domain for anomaly detection, the quality assurance of the data should be investigated. As well as using knowledge-based systems, taking advantage of data-driven approaches such as machine learning techniques can increase the efficiency of the maritime surveillance systems. Finally, the next step for improving the maritime surveillance systems after being equipped with the anomaly detection functionality is to predict the future threats or incoming anomalies based on the analysis of the current situation.

References

Alonso, J. M., Ambur, O., Amutio, M. A., Azañón, O., Bennett, D., Flagg, R., McAllister, D., Novak, K., Rush, S., & Sheridan, J. (2009). Improving access to government through better use of the web. <www.w3.org/TR/egov-improving/>.

Andersson, M., & Johansson, R. (2010). Multiple sensor fusion for effective abnormal behavior detection in counter-piracy operations. In *Proceedings of international waterside security conference (WSS)*.

Andler, S. F., Fredin, M., Gustavsson, P. M., van Laere, J., Nilsson, M., & Svenson, P. (2009). SMARTracIn: a concept for spoof resistant tracking of vessels and detection of adverse intentions. In *Proceedings of international society for optical engineering*.

Carthel, C., Coraluppi, S., & Grignan, P. (2007). Multisensor tracking and fusion for maritime surveillance. In *Proceedings of 10th conference of the international society of information fusion*.

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys, 41*(3), 1–58.

Constantine, L. L., & Lockwood, L. A. D. (1999). *Software for use: A practical guide to the models and methods of usage-centred design*. Addison Wesley.

Dahlbom, A., & Niklasson, L. (2007). Trajectory clustering for coastal surveillance. In *10th conference of the international society of information fusion*.

Dedijer, S., & Jéquier, N. (1987). Intelligence for economic development: An inquiry into the role of the knowledge industry Berg, Oxford.

Dietrich, D., Gray, J., McNamara, T., Poikola, A., Pollock, P., Tait, J., & Zijlstra, T. (2009). Open data handbook open knowledge foundation logo. <www.opendatahandbook.org/en/>.

Ding, Z., Kannappan, G., Benameur, K., Kirubarajan, T., & Farooq, M. (2003). Wide area integrated maritime surveillance: An updated architecture with data fusion. In *Proceedings of sixth international conference on information fusion*.

Edlund, J., Gronkvist, M., Lingvall, A., & Sviestins, E. (2006). Rule-based situation assessment for sea surveillance. In *Proceedings of international society for optical engineering*.

Fooladvandi, F., Brax, C., Gustavsson, P., & Fredin, M. (2009). Signature-based activity detection based on Bayesian networks acquired from expert knowledge. In *Proceedings of 12th international conference on information fusion (FUSION)*.

Guerriero, M., Willett, P., Coraluppi, S., & Carthel, C. (2008). Radar/AIS data fusion and SAR tasking for maritime surveillance. In *Proceedings of 11th international conference on information fusion*.

Guyard, A. B., & Roy, J. (2009). Towards case-based reasoning for maritime anomaly detection: A positioning paper. In *Proceedings of 12th IASTED international conference on intelligent systems and control*.

Hall, M. J., Hall, S. A., & Tate, T. (2000). Removing the HCI bottleneck: How the human computer interface (HCI) affect the performance of data fusion systems. In *Proceedings of MSS national symposium on sensor and data fusion*.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques*. Elsevier.

Johansson, F., & Falkman, G. (2007). Detection of vessel anomalies – a Bayesian network approach. In *Third international conference on intelligent sensors, sensor networks and information*.

Lane, R. O., Nevell, D. A., Hayward, S. D., & Beaney, T. W. (2010). Maritime anomaly detection and threat assessment. In *Proceedings of 13th international conference on information fusion*.

Laxhammar, R. (2008). Anomaly detection for sea surveillance. In *11th international conference on information fusion*.

Lefebvre, E., & Helleur, C. (2001). Multisource information adaptive fuzzy logic correlator for recognized maritime picture. In *Proceedings of fourth international conference on information fusion*.

Lefebvre, E., & Helleur, C. (2004). Automated association of track information from sensor sources with non-sensor information in the context of maritime surveillance. In *Proceedings of international society of information fusion*.

Mano, J. P., Georgé, J. P., & Gleizes, M. P. (2010). Adaptive multi-agent system for multi-sensor maritime surveillance. Advances in Practical Applications of Agents and Multiagent Systems, 70, 285–290.

Molloy, J. C. (2011). The open knowledge foundation: Open data means better science. *Public Library of Science Biology, 9*(12), 1–4.

Nilsson, M., van Laere, J., Ziemke, T., & Edlund, J. (2008). Extracting rules from expert operators to support situation awareness in maritime surveillance. In *Proceedings of 11th international conference on information fusion*.

Ponsford, A. M., ĎSouza, I. A., & Kirubarajan, T. (2009). Surveillance of the 200 nautical mile EEZ using HFSWR in association with a spaced-based AIS interceptor. In *Proceedings of IEEE conference on technologies for homeland security*.

Rhodes, B. J., Bomberger, N. A., Seibert, M., & Waxman, A. M. (2005). Maritime situation monitoring and awareness using learning mechanisms. In *Proceedings of IEEE military communications conference*.

Rhodes, B. J., Bomberger, N. A., Seibert, M., & Waxman, A. M. (2006). SeeCoast: Automated port scene understanding facilitated by normalcy learning. In *Proceedings of IEEE military communications conference*.

Ristic, B., La Scala, B., Morelande, M., & Gordon, N. (2008). Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction. In *Proceedings of 11th international conference on information fusion*.

Riveiro, M., & Falkman, G. (2009). Interactive visualization of normal behavioral models and expert rules for maritime anomaly detection. In *Proceedings of sixth international conference on computer graphics, imaging and visualization*.

Riveiro, M., Falkman, G., & Ziemke, T. (2008). Improving maritime anomaly detection and situation awareness through interactive visualization. In *Proceedings of 11th international conference on information fusion*.

Roy, J. (2008). Anomaly detection in the maritime domain. In *Proceedings of international society for optical engineering*.

Roy, J. (2010). Rule-based expert system for maritime anomaly detection. In *Proceedings of international society for optical engineering*.

Roy, J., & Davenport, M. (2010). Exploitation of maritime domain ontologies for anomaly detection and threat analysis. In *Proceedings of international waterside security conference*.

van Laere, J., & Nilsson, M. (2009). Evaluation of a workshop to capture knowledge from subject matter experts in maritime surveillance. In *Proceedings of 12th international conference on information fusion*.

Vespe, M., Sciotti, M., Burro, F., Battistello, G., & Sorge, S. (2008). Maritime multi-sensor data association based on geographic and navigational knowledge. In *Proceedings of IEEE radar conference*.