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/.