Machine Learning Approaches to Maritime Anomaly Detection

Postupci strojnog učenja za otkrivanje pomorskih anomalija

Ines Obradović
Department of Electrical Engineering and Computing, University of Dubrovnik
e-mail: ines.obradovic@unidu.hr

Mario Miličević
Department of Electrical Engineering and Computing, University of Dubrovnik
e-mail: mario.milicevic@unidu.hr

Krunoslav Žubrinić
Department of Electrical Engineering and Computing, University of Dubrovnik
e-mail: krunoslav.zubrinic@unidu.hr

UDK 656.61 : 004.942
Review / *Pregledni članak*
Paper accepted / *Rukopis primljen:* 7. 10. 2014.

Summary

Topics related to safety in maritime transport have become very important over the past decades due to numerous maritime problems putting both human lives and the environment in danger. Recent advances in surveillance technology and the need for better sea traffic protection led to development of automated solutions for detecting anomalies. These solutions are based on generating normality models from data gathered on vessel movement, mostly from AIS. This paper provides a presentation of various machine learning approaches for anomaly detection in the maritime domain. It also addresses potential problems and challenges that could get in the way of successful automation of such systems.

KEY WORDS
maritime traffic
anomaly detection
situational awareness
machine learning
AIS

Sažetak

Teme vezane uz sigurnost u pomorskom prometu dobile su na važnosti tijekom proteklih desetljeća uslijed brojnih pomorskih problema kojima se ugrožavaju ljudi i okoliš. Nedavna dostignuća u tehnologiji nadzora i potreba za boljom zaštitom pomorskog prometa dovele su do razvoja automatiziranih rješenja za otkrivanje anomalija. Ta rješenja temelje se na generiranju modela normaliteta iz prikupljenih podataka o kretanju plovila, uglavnom iz AIS-a. Ovaj rad daje pregled različitih pristupa strojnog učenja sa ciljem otkrivanja anomalija u pomorstvu. Također se bavi potencijalnim problemima i izazovima koji bi mogli biti prepreka uspješnoj automatizaciji takvih sustava.

KLJUČNE RIJEČI
pomorski promet
otkrivanje anomalije
situacijska svijest
strojno učenje
AIS

INTRODUCTION / *Uvod*

Maritime transport is the backbone of international trade, as more than three quarters of the global trade volume is carried by sea [6]. The complexity and increasing volume of sea traffic make maritime a sensitive and a high-risk transport sector in terms of security. Topics related to safety in maritime transport have become very important over the past decades. This is mostly due to numerous maritime problems, which put both human lives and the environment in danger, such as collision, grounding, illegal fishing, smuggling, pollution, and piracy.

In the past, surveillance of maritime traffic was difficult due to a lack of data. However, since the electronic tracking systems emerged, the amount of available data has dramatically grown, transforming the problem into one of overabundance, leading to a need for automated analysis.

A lot of current research is aimed at determining the best way to exploit this wealth of data, in order to improve situational awareness in the maritime domain. Maritime situational awareness (MSA) is an area of study that aims to use available data sources to create maximum awareness of activities in the maritime environment.

Moreover, in order to enhance MSA, the European Commission has developed the common information sharing environment for the European Union maritime domain (Maritime CISE). It is a voluntary collaborative process in the European Union with the objective of ensuring that maritime surveillance information collected by one maritime authority and considered necessary for the operational activities of others can be shared and be subject to multiuse, rather than collected and kept for a single purpose [5].

The main goal of the efforts to improve MSA is to be able to detect anomalies. The idea behind this is to use all of the collected data on vessel movements to distinguish certain patterns. By definition, a pattern is composed of recurring events that repeat in a predictable manner. Here, what is predictable is considered normal and the rest will be considered as anomalies. To generate normality models from vessel movement data,

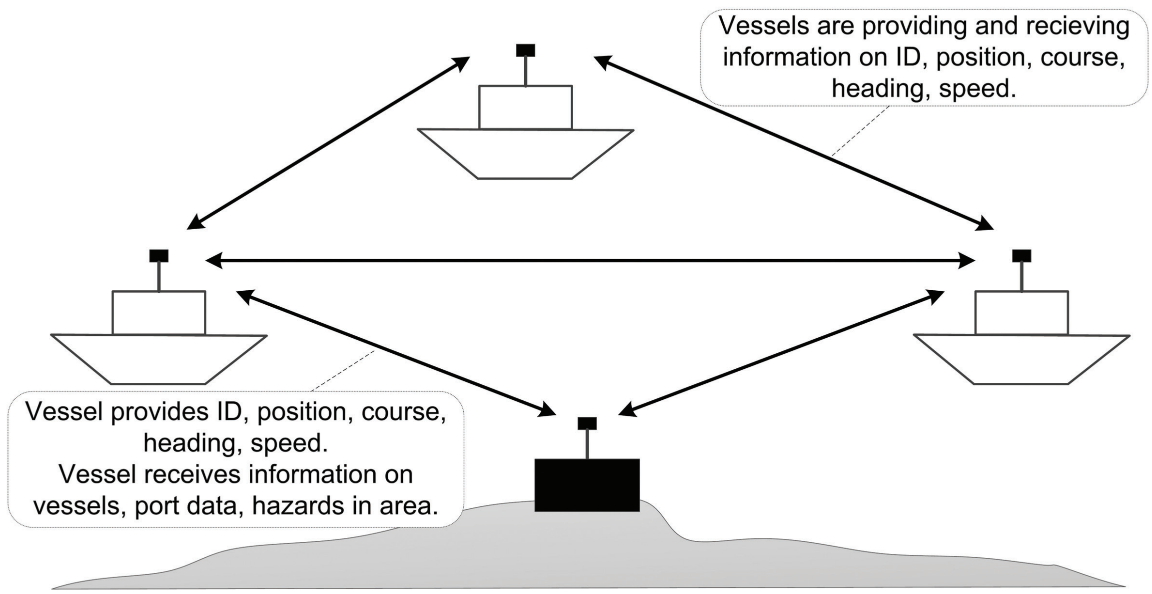

Figure 1 AIS system overview [21]
Slika 1. Pregled AIS sustava [21]

researchers use many different machine learning techniques.

In the following sections, we present how safety at sea can benefit from using various machine learning techniques for maritime anomaly detection.

AIS AS THE DATA SOURCE / *AIS kao izvor podataka*

Data are the raw material of anomaly detection, so it is reasonable to put the data acquisition at the beginning of the process. Anomalies are derived from data, so the success of the analyses depends largely on data collection. A variety of sensors has been used over the years to provide information on vessel movements. The most important self-reporting maritime system is the Automatic Identification System (AIS). The AIS has been made compulsory by the International Convention for Safety of Life at Sea (SOLAS) for vessel of above 300 gross tons and most commercial vessels such as cargo, passenger vessel, tankers, etc. [17,20]. In addition, AIS is required for fishing vessels with a length greater than 15 m and sailing in water under the jurisdiction of Member States of the European Union [4]. The obligation of AIS usage is causing the drastic increase in the amount of information available to analysts.

The AIS transponder sends dynamic messages every two to thirty seconds, depending on the vessel speed, and every three minutes when the vessel is at the anchor. The messages include the vessel unique identifier (Maritime Mobile Service Identity - MMSI), location, course, speed, destination, navigational status, and other details (Figure 1).

The AIS is intended to enhance safety of life at sea, the safety and efficiency of navigation, and the protection of the marine environment. Although it is used primarily for collision avoidance, it can also be utilized as a source of data for maritime surveillance and anomaly detection. In order to provide a better perception and situational awareness, AIS data can be combined with data from other sources, such as data from coastal radars, video and infrared surveillance systems and synthetic-aperture radar (SAR) systems.

APPLYING MACHINE LEARNING TECHNIQUES / *Primjena tehnika strojnog učenja*

Anomaly detection is an important problem that has been researched within diverse research areas and application domains. It refers to the problem of finding patterns in data that do not conform to expected behavior. Its importance is due to the fact that anomalies in data translate to significant, and often critical, actionable information in various application domains. Anomaly detection finds extensive use in a variety of applications such as fraud detection for credit cards, insurance, or health care, intrusion detection for cyber-security, fault detection in safety critical systems, and military surveillance for enemy activities [3]. This section describes how several machine learning techniques are used to detect anomalies in the maritime domain.

SUPPORT VECTOR MACHINES / *Potporni vektorski strojevi*

It is described in [7] how the Support Vector Machines (SVMs) can be utilized for analyzing the AIS raw data to detect vessel anomaly behavior. The SVMs are a set of supervised methods that need some prior knowledge before classification. The SVMs method is implemented as a pattern classification technique that measures the similarity between input tracking data and the tracking data stored in the data base. The anomaly detection system consists of two phases: the training phase, where the models of AIS raw data are developed; and testing phase, which is a system performance evaluation process. The visual analysis method is used for detection of vessel anomaly behavior.

Figure 2 shows the scenario of vessel anomaly behavior in the case of a U-turn route. The normal route is shown with blue stripes, whereas the anomaly route is shown with a black circle.

NEURAL NETWORKS / *Neuronske mreže*

In [1] and [16] it is shown how neural networks methods can be used for maritime situational awareness at a variety of conceptual, spatial, and temporal levels. The papers report

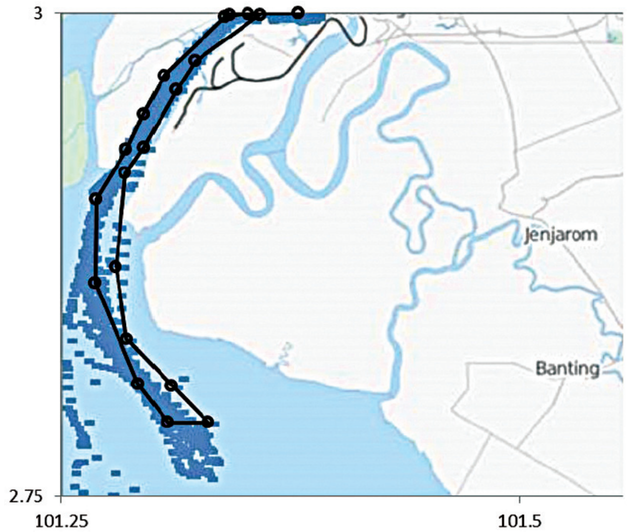

Figure 2 Vessel anomaly behavior with the U-turn route [7]
Slika 2. Anomalija u ponašanju plovila na polukružnoj ruti [7]

successful learning for detection of anomalous vessel event behavior and to predict the future vessel location, both based on AIS data. At the core of the system lies a significantly modified version of the fuzzy ARTMAP neural network classifier.

For event-level anomaly detection, the learning system must initially be presented with a series of observations that are known to reflect routine activity. As normalcy is learned, new observations can be judged for normalcy. Events considered unusual can then be flagged as alerts to cue human operator attention. Operators may guide learning by confirming or rejecting alerts raised by the system. As activities or contexts change, learning proceeds in a semi-supervised fashion, benefiting from operator experience.

The system enhanced to produce predictions of future vessel location on the basis of current vessel position (latitude and longitude) and velocity (course and speed) is reported in [1]. The system associates different geographical grid locations through Hebbian learning corresponding to the position of a vessel at constant time intervals. The implementation places a uniform square grid over the area of interest to discretize the vessel location. The system is able to place the vessel in a grid location and give it a velocity state, for each report. Weights are attributed to pairs of grid locations/speed and change via Hebbian learning. To detect an anomalous vessel position, the system uses the previous grid position, and if the weight is not strong enough, an alert is raised. Figure 3 illustrates the prediction of the future position of the vessel (ID 31988500) in Miami Harbor.

BAYESIAN NETWORKS / *Bayesove mreže*

A lot of work has been done in researching applications of Bayesian Networks (BNs) as a tool for detecting anomalies in vessel tracks based on AIS data. BNs potentially have two substantial advantages in this domain over other types of models: 1) possibility to easily include expert knowledge into the model, and 2) possibility for non-specialists to understand and interpret the learned model [9].

Helldin and Riveiro [8] examine how the reasoning capabilities of a BN can assist surveillance system operators. They focus specifically on finding and building suitable explanations from BNs outcomes, when BNs are used for detecting anomalous behavior in maritime traffic. Explanations are provided in the form of Explanation and Causal Explanation Trees (ET and CET).

Lane et al. [12] identify five anomalous ship behaviors that can be monitored using AIS transmissions: deviation from standard routes, unexpected AIS activity, unexpected port arrival, close approach, and zone entry. For each behavior, a

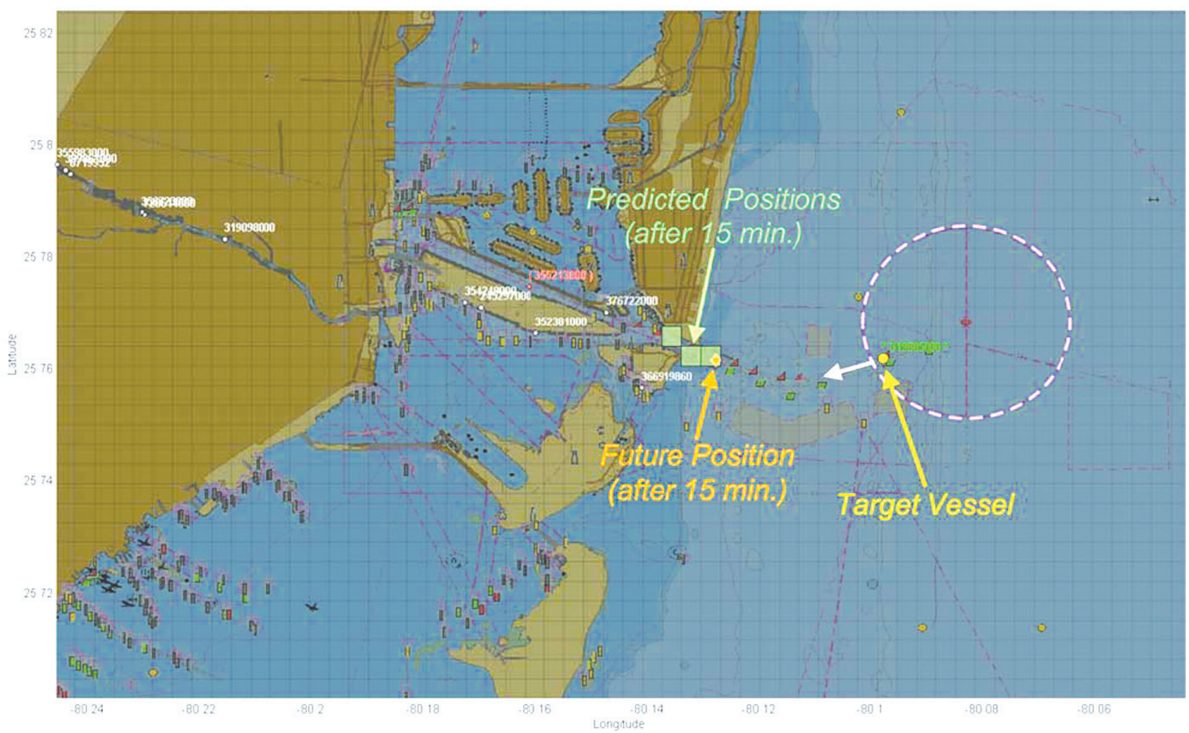

Figure 3 Predicting future vessel position [1]
Slika 3. Predviđanje budućeg položaja plovila [1]

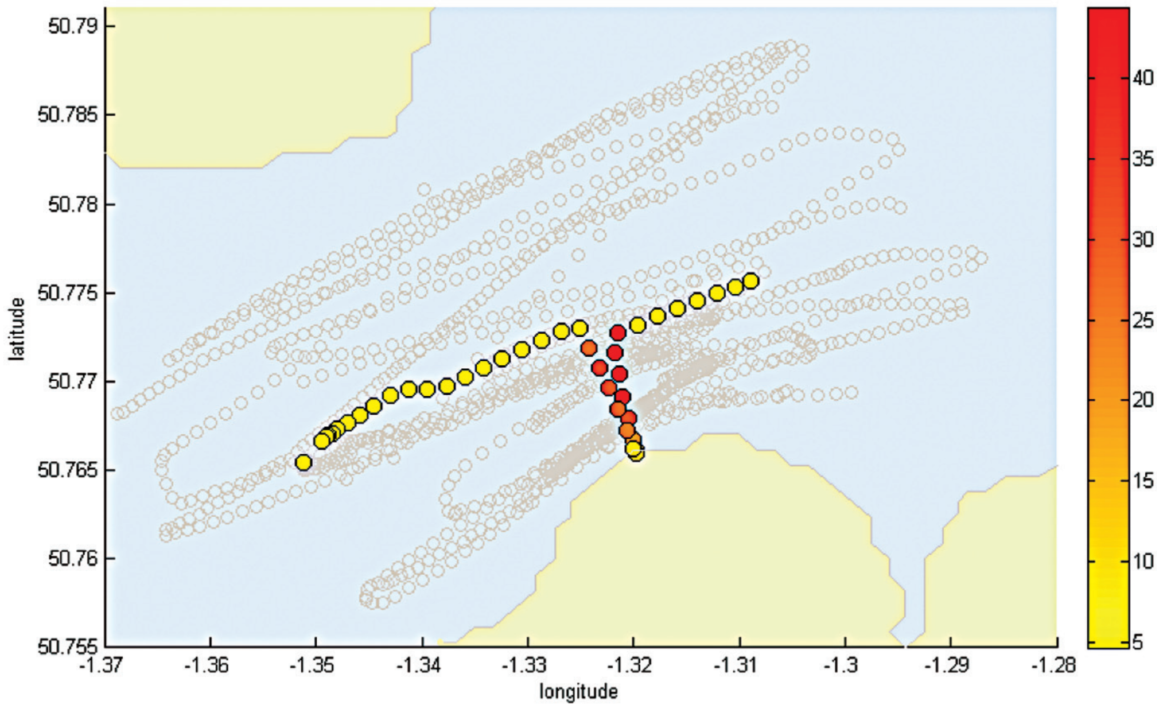

Figure 4 People smuggling scenario [11]
Slika 4. Scenarij krijumčarenja ljudi [11]

process is described for determining the probability that it is anomalous. Individual probabilities are combined using a BN to calculate the overall probability that a specific threat is present.

Mascaro et al. in [14] and [15] produced networks at two different time scales, in the form of the time series and track summary models. They did so by using the machine learner CaMML on AIS data combined with additional real world data such as weather and time, as well as vessel interactions. The models demonstrated distinct and complementary strengths in identifying anomalies, paving the way to an improvement in the field of anomaly detection by combining their assessments.

GAUSSIAN PROCESSES / *Gaussijanski procesi*

In [11] a model of normality is created from historical AIS data using Gaussian Processes (GPs), thus codified expert knowledge is not required. An advantage in GPs is that the model is non-parametric so it is not necessary to build in features of anomalous behavior. The model uses an Active Learning paradigm that allows selection of an optimal training sample from AIS data, which accurately represents the entire set. The resultant model allows a measure of normality to be calculated for each newly-observed transmission according to its velocity given its current latitude and longitude. Using this measure of normality, ships can be identified as potentially anomalous and prioritized for further investigation.

Figure 4 illustrates a people smuggling scenario. A vessel that, during a NE/SW Channel run, breaks off to head to shore, before returning to its original path. A vessel follows a normal track, then speeds to the shore to pick up some people and quickly returns to its original path. The color of the test points represents the level of abnormality detected using the model.

GAUSSIAN MIXTURE MODEL / *Gaussijanski model mješavine*

Gaussian Mixture Model (GMM) is a common model for approximating continuous multi-modal distributions when knowledge regarding the structure is limited. It has been used in many anomaly detection applications [3]. Laxhammar proposed and implemented unsupervised clustering of normal vessel traffic patterns using multivariate GMM as cluster model in [13]. The patterns are represented as the momentary location, speed and course of tracked vessels. The learnt cluster models are used for anomaly detection in sea traffic. The implemented models are trained and evaluated using recorded sea traffic. A qualitative analysis reveals that the most distinguishing anomalies found in the traffic are vessels crossing sea lanes and vessels travelling close to and in the opposite direction of sea lanes. The generality of the proposed model is stressed, as it is potentially applicable to other domains involving surveillance of moving objects.

THE ROLE OF VISUALIZATION / *Uloga vizualizacije*

Riveiro et al. in [18] and [19] argue that visualization and interaction play the key role in improving anomaly detection performance in general, and in particular, in supporting human involvement. Visualization and interaction are key to perform an adequate data analysis, construct understandable normal models, update and validate such models and create useful and comprehensible output. This cannot only generate suitable response from operators, but also improve the whole anomaly detection process.

Figure 5 presents a visualization of normal vessel behavioral models built from real AIS data. The model is built using a statistical method that combines Self Organizing Maps (SOMs) and GMM. Eight-dimensional space is used to represent such probability density functions (position, speed, course over ground, heading, length, width and draught).The probability function is projected over a 2 dimensional map. High values of probability are represented in red, while blue represents lower probability values [18].

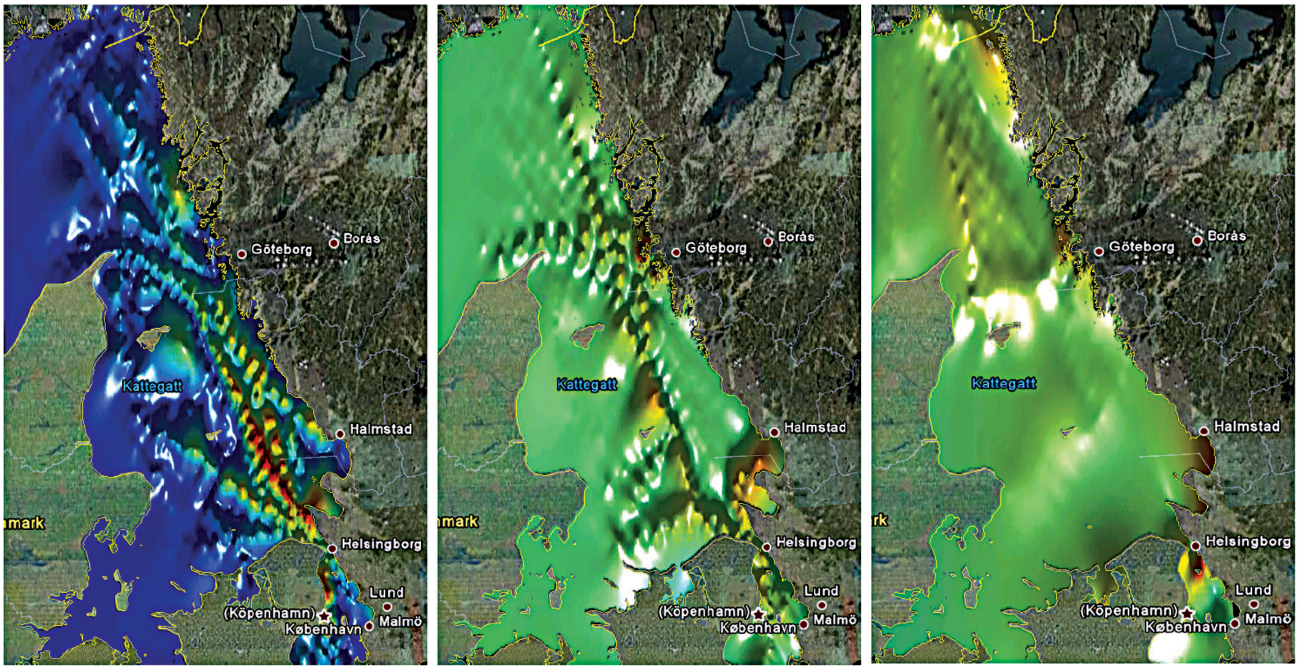

Figure 5 Visualizations of normal behavioral models for cargo (left), tanker (middle) and passenger (right) vessels [18]
Slika 5. Vizualizacija uobičajenih modela ponašanja za teretne (lijevo), tankere (sredina) i putničke (desno) brodove [18]

POTENTIAL PROBLEMS AND CHALLENGES / *Potencijalni problemi i izazovi*

FALSIFIED AIS DATA / *Falsificirani ais podaci*

Given that the AIS is a self-reporting system, the trustworthiness of positional information depends on data being reported by the vessel, rather than measured by a sensor. Any self-reporting system is prone to "spoofing" or the intentional reporting on incorrect information. There are also several other problems pertaining to the AIS data transmission, reception and exploitation. Rather than spoofing, vessels could simply turn off their AIS transmitters, possibly periodically, in order to hinder the surveillance systems and their operators from detecting illicit activities. Paper [10] deals with the problem of determining whether a vessel is transmitting falsified AIS data.

INSUFFICIENT TRAINING DATA / *Nedostatak podataka za učenje*

Although the AIS produces vast amounts of data reporting on the information about movements of the vessels, there are no known anomalous tracks, nor are there any standardized or publicly available vessel track data sets containing anomalies. This obstacle makes it impossible for machine learning algorithms to train on anomalous data. Still, there are many ways to create anomalous data, e.g. anomalous data can be generated by modifying selected attributes to random values within their ranges [2], or by using anomalous models [11].

FALSE ALARMS / *Lažne uzbune*

The anomaly threshold is a central parameter in all anomaly detection systems, since it regulates the sensitivity to true anomalies and the rate of false alarms. This is an issue which arises also due to substantial input. Since the data amounts to millions of data objects, if the threshold is badly calibrated, a large percent of false alarms can make the analysis overwhelming for the operator. On the other hand, if the threshold is unbalanced in the opposite direction, the risk of missing true anomalies increases.

LACK OF EXPLANATION / *Nedostatak objašnjenja*

Many maritime anomaly detection systems only inform their users about detected anomalies but do not provide further explanations. The ability of explaining the reasoning behind an alarm is important for an operator to understand the advice the system gives. Deficient explanations can have a negative effect on users' confidence in the detection system. Paper [8] gives the review of different explanation methods for BNs and empirical tests conducted with ET and CET methods in maritime scenario.

CONCLUSION / *Zaključak*

Anomaly detection is a problem of finding patterns in data that do not conform to expected behavior and it finds extensive use in a wide variety of applications. Here, we give an overview of several machine learning techniques that can be used to detect anomalies in the maritime domain.

AIS data, especially when combined with data from other sources, provides a high level of perception and situational awareness. Selected methods, namely, support vector machines, neural networks, Bayesian networks, Gaussian processes and Gaussian mixture model are presented. It is also explained how they can be used on such data to identify anomalous behavior, such as deviation from standard routes, unexpected AIS activity, unexpected port arrival, close approach, and zone entry, to predict future vessel position etc.

Nevertheless, the process of anomaly detection cannot be fully automated as it requires the attention of the human operator. Human involvement can be supported through the incorporation of visualization and interaction in the anomaly detection system.

Application of machine learning techniques for maritime anomaly detection is a promising field that could enhance security and safety at sea, however, there are still some problems that need to be solved.

REFERENCES / *Literatura*

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

[2] Cansado, A., Soto, A., 2008, "Unsupervised anomaly detection in large databases using Bayesian networks", Applied Artificial Intelligence 22 (4), pp. 309-330.

[3] Chandola, V., Banerjee, A., Kumar, V., 2009, "Anomaly detection: A survey", ACM Comput. Surv., 41(3), pp.1–58.

[4] Commission of the European Communities, DocumentCOM (2008) 310 final - 2005/0239 COD, Brussels 2008.

[5] European Commission, "Better situational awareness by enhanced cooperation across maritime surveillance authorities: next steps within the Common Information Sharing Environment for the EU maritime domain", Communication from the Commission to the European Parliament and the Council, 2014.

[6] Grech, M., Horberry, T., Koester, T., 2008, "Human factors in the maritime domain", CRC Press, Boca Raton

[7] Handayani, D.O.D., Sediono, W., Shah, A., 2013 . "Anomaly Detection in Vessel Tracking Using Support Vector Machines (SVMs)", Advanced Computer Science Applications and Technologies (ACSAT), pp. 213 – 217.

[8] Helldin, T., Riveiro, M., 2009, "Explanation Methods for Bayesian Networks: review and application to a maritime scenario", Proceedings of the 3rd Annual SkövdeWorkshop on Information Fusion Topics, pp. 11-16.

[9] Johansson, F., Falkman, G., 2007, "Detection of vessel anomalies - a Bayesian network approach", Int. Conf. on Intelligent Sensors, Sensor Networks and Information Processing, pp. 395-400.

[10] Katsilieris, F., Braca, P., Coraluppi, S., 2013, "Detection of malicious AIS position spoofing by exploiting radar information", Proceedings of the 16th International Conference on Information Fusion.

[11] Kowalska, K., Peel, L., 2012, "Maritime anomaly detection using Gaussian process active learning", 15th Conference on Information Fusion, Singapore, pp. 1164–1171.

[12] Lane, R. O., Nevell, D. A., Hayward, S. D., Beaney, T. W., 2010, "Maritime Anomaly Detection and Threat Assessment", 13th Intl. Conf. on Information Fusion.

[13] Laxhammar, R., 2008 , "Anomaly detection for sea surveillance", Proceedings of the 11th International Conference on Information Fusion, Germany.

[14] Mascaro, S., Korb K. B., Nicholson A. E., 2010, "Learning Abnormal Vessel Behaviour from AIS Data with Bayesian Networks at Two Time Scales", Technical ReportTR 2010/, Bayesian Intelligence.

[15] Mascaro, S., Nicholson, A. E., Korb, K. B, 2014, "Anomaly Detection in Vessel Tracks using Bayesian networks", International Journal of Approximate Reasoning (55).

[16] Rhodes, B. J., Bomberger, N. A., Seibert, M. C., Waxman, A. M., 2005, "Maritime situation monitoring and awareness using learning mechanisms". In Proceedings of IEEE MILCOM 2005 Military Communications Conference, Atlantic City, USA.

[17] Ristic, B., La Scala, B., 2008, "Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction," 11th International Conference on Information Fusion.

[18] Riveiro, M., Falkman, G., 2011, "The role of visualization and interaction in maritime anomaly detection", Proc. of SPIE 7868, Visualization and Data Analysis.

[19] Riveiro, M., Falkman, G., Ziemke, T., 2008, "Improving Maritime Anomaly Detection and Situation Awareness Through Interactive Visualization", Proceedings of the 11th International Conference on Information Fusion.

[20] Safety of Life at Sea (SOLAS) convention Chapter V. Regulation 19.

[11] SOLAS, Guidance on Chapter V – Safety of Navigation (Revision), 2007. Available: https://mcanet.mcga.gov.uk/public/c4/solasv/index.html (21.9.2014.)