VILNIUS UNIVERSITY

Andrius
DARANDA

Machine learning-based prediction of the behavior of marine traffic participants and discovering non-standard marine traffic situations

SUMMARY OF DOCTORAL DISSERTATION

Technological Sciences,
Informatics Engineering (T 007)

VILNIUS 2021

This dissertation was written between 2015 and 2020 in the Vilnius University.

Academic supervisor:
Prof. Habil. Dr. Gintautas Dzemyda (Vilnius University, Technological
Sciences, Informatics Engineering – T 007).
Academic consultant:
Prof. Dr. Arūnas Andziulis (Klaipėda University, Technological Sciences, Informatics Engineering T 007).

This doctoral dissertation will be defended in a public meeting of the Dissertation Defence Panel:

Chairman – **Prof. Dr. Julius Žilinskas** (Vilnius University, Technological
Sciences, Informatics Engineering – T 007).

Members:
Prof. Habil. Dr. Rimantas Barauskas (Kaunas University of Technology, Tecnological Sciences, Informatics Engineering – T 007),
Assoc. Prof. Dr. Nikolaj Goranin (Vilnius Gediminas Technical University, Tecnological Sciences, Informatics Engineering – T 007),
Prof. Dr. Audris Mockus (University of Tennessee, USA, Tecnological Sciences, Informatics Engineering – T 007),
Prof. Dr. Aistis Raudys (Vilnius University, Technological Sciences, Informatics Engineering – T 007).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 12:00 p. m. on 25th of June, 2021 in Room 203 of the Institute of Data Science and Digital Technologies of Vilnius University.
Address: Akademijos str. 4, LT-04812, Vilnius, Lithuania

The summary of the doctoral dissertation was distributed on the 24th of May 2021.
The text of this dissertation can be accessed at the library of Vilnius University, as well as on the website of Vilnius University:
www.vu.lt/lt/naujienos/ivykiu-kalendorius

VILNIAUS UNIVERSITETAS

Andrius
DARANDA

Mašininiu mokymusi grindžiamas laivybos eismo dalyvių elgsenos prognozavimas bei nestandartinių laivybos srauto situacijų atradimas

DAKTARO DISERTACIJOS SANTRAUKA

Technologijos mokslai,
Informatikos inžinerija (T 007)

VILNIUS 2021

Disertacija rengta 2015– 2020 metais Vilniaus universitete.

Mokslinis vadovas:
prof. habil. dr. Gintautas Dzemyda (Vilniaus universitetas, technologijos mokslai, informatikos inžinerija – T 007).
Mokslinis konsultantas:
prof. dr. Arūnas Andziulis (Klaipėdos universitetas, technologijos mokslai, informatikos inžinerija – T 007).

Gynimo taryba:
Pirmininkas – **prof. dr. Julius Žilinskas** (Vilniaus universitetas, technologijos mokslai, informatikos inžinerija – T 007).
Nariai:
prof. habil. dr. Rimantas Barauskas (Kauno technologijos universitetas, technologijos mokslai, informatikos inžinerija – T 007),
doc. dr. Nikolaj Goranin (Vilniaus Gedimino technikos universitetas, technologijos mokslai, informatikos inžinerija – T 007),
prof. dr. Audris Mockus (Tenesio universitetas, JAV, technologijos mokslai, informatikos inžinerija – T 007),
prof. dr. Aistis Raudys (Vilniaus universitetas, technologijos mokslai, informatikos inžinerija – T 007).

Disertacija ginama viešame Gynimo tarybos posėdyje 2021 m. birželio mėn. 25 d. 12 val. Vilniaus universiteto Duomenų mokslo ir skaitmeninių technologijų instituto 203 auditorijoje.
Adresas: Akademijos g. 4, LT-04812 Vilnius, Lietuva.
Disertacijos santrauka išsiuntinėta 2021 m. gegužės 24 d.
Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje ir VU interneto svetainėje adresu: *https://www.vu.lt/naujienos/ivykiu-kalendorius*

1. INTRODUCTION

1.1. Research area

In recent decades, marine traffic has dramatically increased. Since ancient times until the present, vessel safety has depended on the watch officer's (navigator) experience and decisions in any maneuvering situation. Even in modern times, marine navigation technologies could not ensure analysis and control of marine traffic. The watch officer must constantly supervise and analyze the marine traffic for dangerous situations or any unusual behavior to ensure safe navigation. It is essential to anticipate and predict the unusual behavior of another monitored vessel. It allows the navigator to make the right decision in time and plan actions to ensure safe marine navigation.

In this thesis, we investigate the possibility of applying machine learning methods for maneuver modeling and threat assessment of maneuvering situations to ensure safe navigation. This work aims to create machine learning methods for predicting and threat assessment of maneuvering situations to ensure safe navigation.

This thesis is based on methods of machine learning. The aim is to apply machine learning to modern marine solutions. A historical marine traffic data clustering method was proposed and investigated. The result of the clustering was applied to further experiments on machine learning algorithm. Two different methods were proposed and examined to predict the next turning point and route of the vessel. Such prediction allows us to plan and evaluate actions to ensure safe navigation. The method to detect anomalies in traffic flow was proposed and examined.

In this thesis, a novel method based on contextual knowledge was proposed. This method makes evaluation and threat assessment of the predictions of the turning point. The purpose of this prediction is to evaluate an unfamiliar maneuvering situation. The proposed method is designed to process real-time marine traffic data.

The analysis of the proposed algorithms was performed by solving the clustering tasks and classification of the historical marine navigation data.

1.2. Relevance of the problem

In today's world, cargo shipping by vessels is an essential mode of transport. Most of the cargo in the world is transported by vessels. Delivery of goods by vessel is the most economical of the available modes of transportation and is also safer and much more environmentally friendly. Even though vessels are safer, accidents or incidents involving vessels are pretty common. The consequences of these disasters are much more painful and tragic compared to accidents in other areas of transport. The reasons for it is:

1. Incomparably more people can die at sea. This is especially true of passenger or cruise vessels;
2. Loss of cargo transported in large quantities by vessels results in significantly higher losses.
3. An ecological catastrophe can occur.

Over the last 50 years, navigation in the marine traffic has been regulated by various legal acts, compliance with which must ensure safety in the marine traffic: International Regulations for Preventing Collisions at Sea 1972) [1], Safety of Life at Sea [23], Standards of Training, Certification, and Watchkeeping for Seafarers (STCW) etc. These conventions aim to ensure the safety of human life at sea by establishing uniform principles, standards, and rules for work on the vessels.

Significant progress has also been made in the application of technologies to vessels, such as the Automatic Identification System (AIS), the Electronic Chart Display and Information System (ECDIS), the Global Maritime Distress and Safety System (GMDSS) etc. These technologies improve safety in marine traffic but can not replace the person – the navigator who makes decisions and is responsible for all possible consequences. This decision is based on all available navigation information about the current situation, analyzing and modeling possible solutions. It allows taking a proper decision on how to proceed in a particular maneuvering situation. Due to the large dimensions of vessels and their restricted maneuverability, the navigator must anticipate and evaluate the maneuvers of other vessels. By assessing the risk and making a decision, the navigator assumes full responsibility for the possible consequences. Therefore, navigators are needed for artificial intelligence decision support methods based on machine learning [27]. Such methods must be able to learn and improve the accuracy of prediction without additional human intervention. It is essential to maneuver situations to recommend possible solutions.

1.3. Aim and tasks of the research

This research aims to develop a machine learning approach to predict and evaluate the maneuvering of monitored vessels to ensure safe navigation in marine traffic.

The object of this research is machine learning algorithms for modeling vessel maneuvers and assessing the maneuvering situation.

The main tasks have established to achieve the aim of the research:

- To analyze the specialized methods and algorithms used in vessels to ensure safe navigation in marine traffic; to experimentally assess application limitations.
- To investigate the possibilities of applying machine learning methods to predict and evaluate the movement and possible maneuver of the monitored vessels, predict the future turning points of the monitored vessel, and detect deviations from the expected route (anomalies).
- To develop a complex method intelligence based on machine learning for the analysis of maneuvers of passing vessels, the maneuver of the monitored vessel could be predicted according to the circumstances, but also that maneuver is

assessed based on contextual knowledge obtained from the historical marine traffic data.

- To substantiate the effectiveness of the proposed methods and solutions to perform experimental studies with large-scale marine traffic historical data.

1.4. Scientific novelty

1. Proposed and experimentally investigated algorithms for historical navigation data clustering.
2. The proposed and experimentally investigated methods are intended to predict the future nearest turning points of the monitored vessel. It could be used to predict the future maneuver of the monitored vessel.
3. A method has been proposed and experimentally investigated to detect a deviation from the planned or predicted route to a particular port.
4. The dissertation presents a new, complex method of machine learning based on contextual information-based prediction and assessment of the future turning point. This prediction is evaluated for warning an unusual situation in the absence of any indication of it. This is particularly important to ensure safe marine navigation. The method is designed to work with real-time navigational navigation data.

1.5. Statements to be defended

- Machine learning algorithms are practicable for:
    - To predict the future nearest turning point of the monitored vessel;
    - To model the entire route of the vessel to the port of destination.
    - To assess the presence of abnormally behaving vessels in the overall marine traffic flow to a given port of destination and identify them.

- Contextual information and a combination of machine learning-based methods can comprehensively predict the future nearest turning point of a monitored vessel. The predicted turning point is based on contextual data. This predicted maneuver could be assessed for different marine traffic actors.

1.6. Approbation of the research

The research results have been published in three peer-reviewed journals, in three peer-reviewed conference proceedings, and were presented and discussed at three national and international conferences. Intermediary results and discussions were presented at one national workshop.

1.7. Outline of the dissertation

The dissertation consists of five chapters and a list of references. The dissertation chapters are as follows: Introduction; Methods ensuring safe navigation in marine traffic; Vessel route prediction model; Prediction model based on the contextual knowledge; Conclusions. This work contains 116 pages that include 53 figures and one table; the list of references consists of 138 sources.

2. VESSEL ROUTE PREDICTION AND ANOMALY DETECTION

2.1 The problem of navigation in the marine traffic

Navigation data sent by the Automatic Identification System (AIS) significantly improves safety in marine traffic [1], [2]. However, an analysis of these navigation data is always necessary to assess the current maneuvering situation. The analysis is always carried out with the means available to the navigator by depicting the surrounding vessels on paper or electronic charts. During the examination, the overrun with other vessels or obstacles is calculated. Also, the vessel traffic situation is assessed. Another essential aspect of evaluating the situation is to predict the maneuver of the observed vessel or the route along which the monitored vessel could continue its voyage. Based on the analysis performed, the passage of the own vessel is adjusted by taking into account other influencing factors such as surrounding vessels, obstacles, weather conditions, etc. However, in the absence of analysis, incomplete analysis, or lack of attention to the current maneuvering situation, disasters often occur. Usually, such disasters are caused by the human factor. [3], [4], [5].

Detecting unusual maneuvers or anomalies of marine traffic participants is important not only for the navigator but also for vessel control services. The services have an obligation to detect and prevent various illegal activities, such as illegal fishing, oil spill detection, criminal offenses, etc. [6], [7], [8]. Anomaly detection helps discover fake counterfeit navigation data transmitted through AIS [9], [10].

Detecting anomalies is a complex task. It is difficult to determine which vessel is behaving abnormally because:

1. All marine traffic participants behave "correctly" until the accident occurs and is followed by a detailed analysis of the incident;

2. The assessment of other marine traffic participants is a subjective assessment of the navigator;
3. It is difficult to assess the maneuver itself if there is no information why a particular maneuver was performed;

Various methods for detecting anomalies have been developed [11], [12], [13]. However, at present, the most practical way for the vessel navigator to find out about the maneuvering actions of the vessel being monitored is to communicate directly with other marine traffic participants via a special watch channel on Very High-Frequency radio. Another more practical way is to assess the route it is taking and compare it with the destination port.

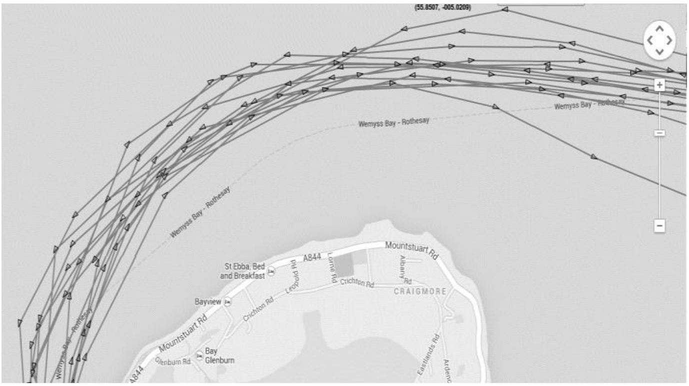

Fig. 1 Example of different vessel movements to the same port of destination

The planned route from port A to port B can always vary slightly, as shown in **Fig. 1**. The reasons for it could be very different:

1. How the watchman on duty follows the planned route. Weather conditions. The weather is especially noticeable when strong winds blow, and the when there are strong headwinds.
2. Navigators have a different point of view of the current situation. Significant deviations from the planned route are possible when vessels pass each other.

3. Condition and operation of mechanical components (for example, steering feather)

2.2 Clustering of navigation data

Navigation data are received through AIS every 2–6 seconds. It depends on the speed of the vessel and the type of navigation data. The navigation data obtained at this frequency are combined to form the vessel's route. However, working with such data is not easy because:

1. The vessel is updating the navigation data every 2 to 6 seconds. If vessels are moving at, for example, 25 km/h, it yields a point about 14 to 42 meters apart. At the same time, the length of the vessel could reach 100 meters and more.
2. It is challenging to distinguish turning points. A vessel of a length of 100 meters can have a turning radius (vessel circulation) of up to 400 meters (4–7 vessel lengths). It depends on the change in course.
3. The planned route is a straight line between the turning points. The distance between the turning points could reach several or several hundred kilometers.
4. The position of the turning point of the vessel is variable with the planned route. It depends on the weather conditions and the physical characteristics of the vessel,
5. The resulting navigational data forms an extensive array of data. It can be an issue because of computational resource constraints.

The DBSCAN (Density-based spatial clustering of applications with noise) clustering algorithm was used [14] to solve the listed problems. This algorithm is designed to work with large amounts of [15] geospatial data [16], [17], [18]. This algorithm is based on data density clustering, grouping data according to the density of spaced geographic points, not just distance.

2.3 Model of monitored vessel route prediction

In the general case, the vessels' route consists of many turning points. The positions (coordinates) of these turning points depend on how the route was planned. The route is planned individually by each navigator, but general trends remain when traveling to a particular port.

The turning points of the route are described by coordinates – latitude and longitude. However, the course and speed at which the turning point is reached become an essential factor, especially for big vessels. The reason for this is vessel inertia. It is necessary to consider the vessels' maneuverability characteristics because the vessel must follow the planned route.

The monitored vessel route prediction model consists of two main parts:

1. Preparation of the navigational data. The navigational data preparation is clustering turning points based on the DBSCAN algorithm. The next step is to calculate the coordinates (longitude and latitude) of the centers of these clusters, as shown in **Fig 2**.
2. Training of the Multi-Layer Neural Network (MLNN) model according to the prepared (clustered) data

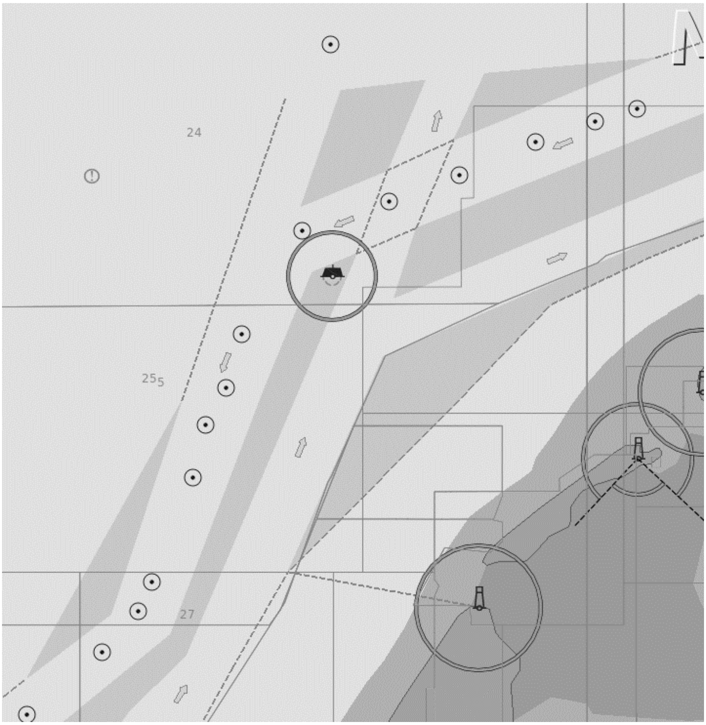

Fig. 2 Positions of the clusters in marine traffic

The second component of the proposed monitored vessel route prediction model is trained MLNN. The MLNN training is based on prepared (clustered) historical navigational data. As mentioned above, the main parameters that must be taken into account are the following: the speed of the vessel, the direction of movement, and the currently available coordinates (longitude and latitude). These parameters determine the coordinates (latitude and longitude) of the next turning point. Based on these prepared and navigational data, multilayer MLNN was trained as shown in **Fig. 3**

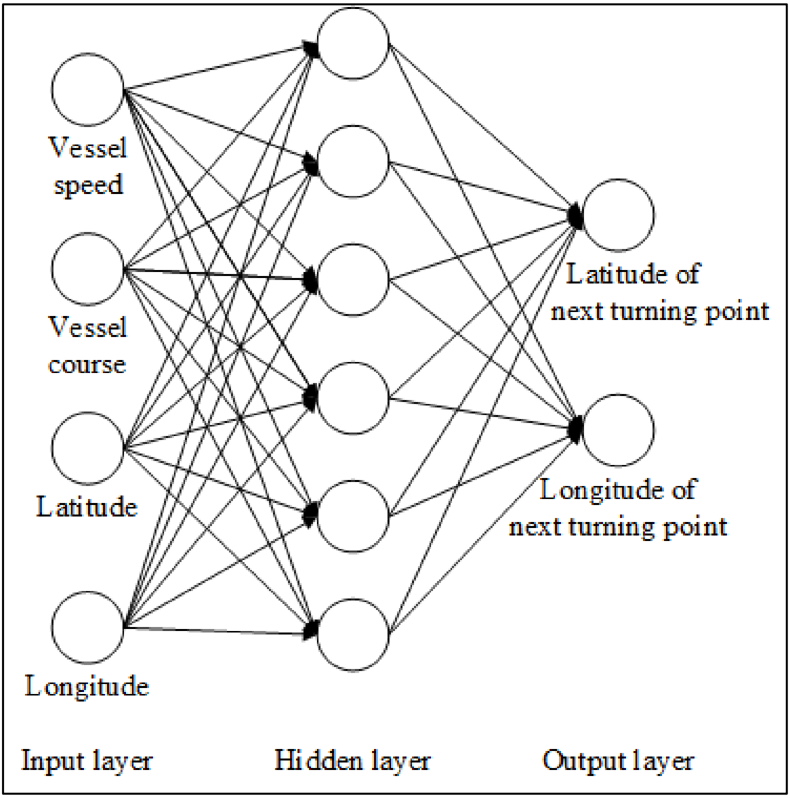

Fig. 3 Applied Multi-Layer Neural Network

The MLNN input layer enters the current position (latitude and longitude) of the vessel, the speed of the vessel, and the direction of movement. The coordinates of the predicted cluster center (turning point) are predicted in the output layer. The data entered in the experiments were normalized in the range [min, max]. The prepared data was divided by the following proportions – 80% of the data is submitted for learning, the rest for testing. The ReLU activation function was used. As many individual MLNN as there were clusters calculated with DBSCAN were trained. It means each route turning point (cluster) has a separate, trained MLNN. The result of the training progress of a specific turning point is shown in **Fig. 4.:** 36360 turning points were provided for this MLNN training, an accuracy of 97.53 % was achieved.

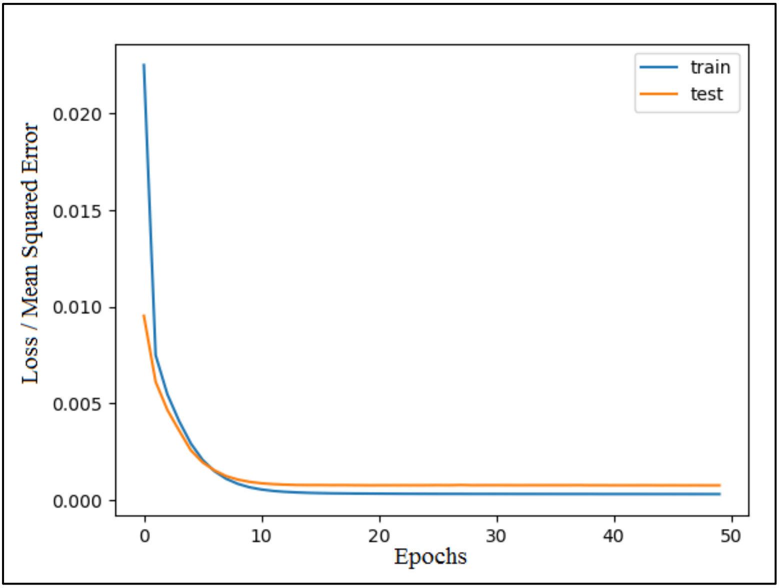

Fig. 4 MLNN learning progress

The created MLNN cannot predict the route directly. MLNN, as mentioned above, predicts the future coordinates of the monitored vessel's turning point. It can be done based on the current position, course, and speed of the vessel being monitored. The most significant advantage of such prediction is that MLNN recalculates the forecast every time based on updated navigational data. In this way, a hypothetical route of the monitored vessel can be concluded by consistent prediction of the expected turning points as shown in **Fig. 5**.

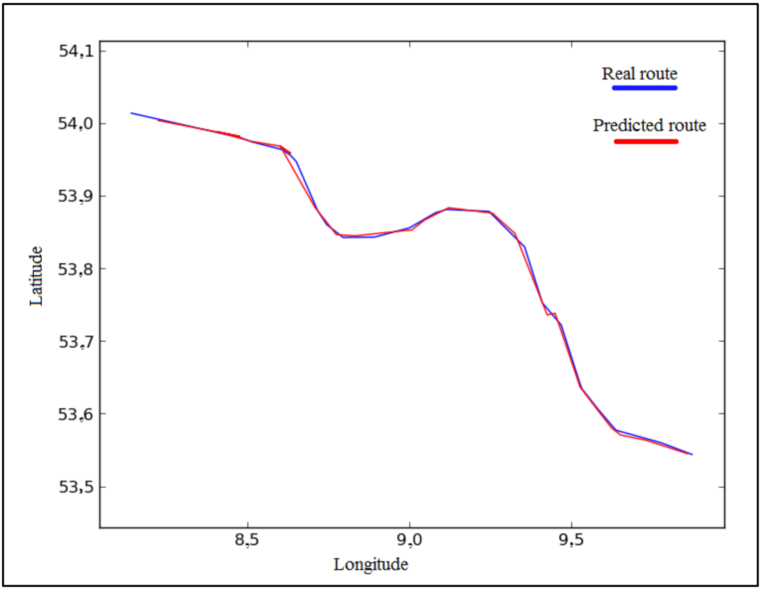

Fig. 5 Comparison of the predicted and real routes

However, it is recommended to use MLNN prediction for the next few steps and to recalculate the prediction when receiving updated navigation data over AIS.

2.3 Detection of anomalies in the marine traffic

The outliers are discarded in preparation navigational data with the DBSCAN algorithm. Such outliers are not used for further MLNN training. However, these outliers refer to particular participants in marine traffic who proceed to the same port of destination. Such outliers are illogical or even contrary to the overall flow of maritime traffic. It is especially apparent for a particular port of destination. Since navigation data is filtered by the destination port, the reasons for anomalies can be very different: the unchanged destination port in AIS transmitter settings; the vessel is lost due to technical problems; illegal activities hidden behind the incorrectly identified port of destination (e.g., illegal fishing etc.). The outlier points (marked with red rectangles) visually stand out from the typical marine traffic flow as shown in **Fig. 6**.

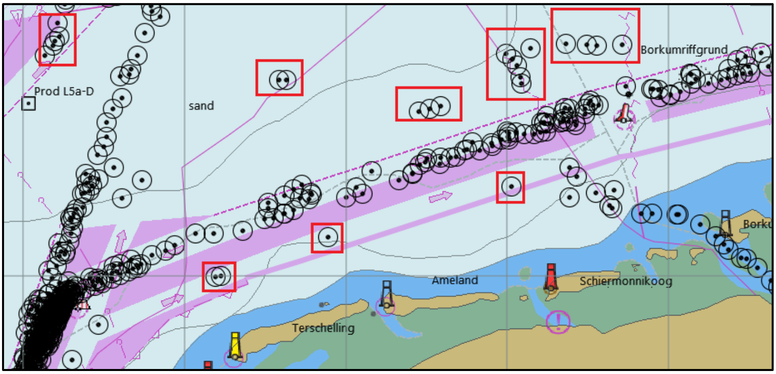

Fig. 6 Examples of anomalies in marine traffic

Outliers points do not affect the prediction of the observed vessel route by the method proposed in Section 2.2. However, assessing them is very important to ensure safety in marine traffic. Such anomalous movements of vessels endanger safe navigation because:

1. Vessels that are outside marine traffic and intend to reach the same port of destination will sooner or later have to join the overall marine traffic. However, this may not happen in the most safe way;
2. It is difficult to predict the actions of such a vessel because the purpose of such maneuvers is not clear;
3. There is a need for constant monitoring of changes in the situation. It is mainly for actors who are not in the overall marine traffic flow;
4. The radar or electronic charts may be misleading due to the unpredictable actions of monitored vessels.

The proposed route prediction model of the monitored vessel is designed to predict the movement of the vessel. This vessel should be in the expected vessel traffic and moving to a particular port. For the vessel which is outside such marine traffic flow, the future turning point can also be predicted, but with a more significant error. The prediction is possible to apply if the movement of the monitored vessel is directed to the typical marine traffic flow. This flow coincides with the turning point according to which the MLNN is trained. In the event of an error, it is difficult to determine whether that particular vessel moves towards merging into the overall marine traffic flow or whether the destination of such a vessel is still another port. It is crucial to determine such an anomaly. The method is proposed below based on an estimate of the distance between the nearest neighboring points. An example of such a flow is given in **Fig. 7**. The centers of the turning point clusters are marked with points. Random points in the marine traffic flow are selected. These points are numbered and marked with the × symbol.

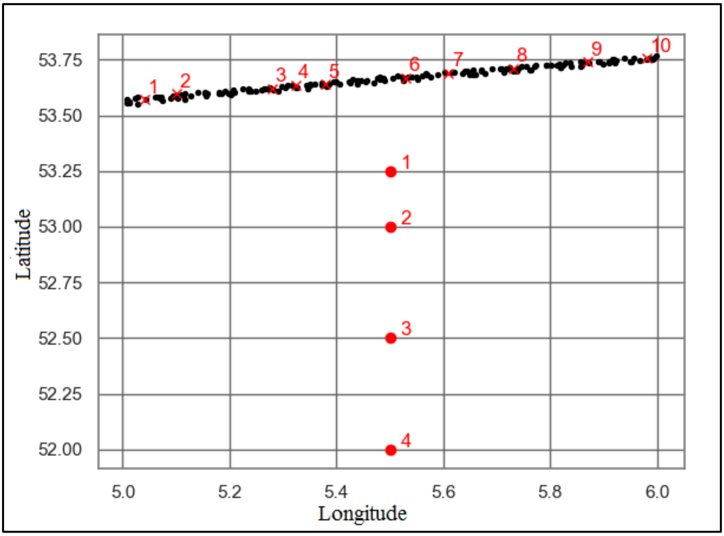

Fig. 7 Anomalous vessel movement toward expected marine traffic

Let us analyze the data in **Fig. 7** in more detail using the *k*-nearest neighbors' comparison. For each point, we find the *k*-nearest points and compute the average distance between this point and the remaining *k* points. Afterward, the results are averaged for all points of the traffic flow. As a result, we get one average value for a fixed *k*.

In **Fig. 7**, we present an average distance among points for different k = 1, …, 8. The average distance increases depending on the number of neighbors. Some regression may approximate this dependence. Linear, second-order polynomial, and logarithmic functions were examined. We see the linear dependence with an R-squared statistical measure near 1.

The k-nearest neighbors from k = 1 to 8 for each of 14 marked points from **Fig. 7**, were calculated and dependencies on k of the average distance were determined. We noticed differences among the dependencies on the average distance for the particular point from the normal traffic flow and the anomaly cases. The examples of dependencies of the average distance among the neighboring points are presented in **Fig. 8** and **Fig. 9**.

Table 1 summarizes the results: variances of estimates, when k runs from 1 to 8, i.e., each variance was evaluated using eight numbers. In the anomaly cases, the variances are much smaller than in the normal traffic flow. Moreover, Table 1 leads to the criteria to detect the vessels outside the typical marine traffic pattern:

1) the distance between the abnormally acting vessel and the nearest neighboring point from the normal traffic flow obtained by DBSCAN is much larger than the average distance among the points of the expected traffic flow;
2) the variance of the average distance between neighboring points obtained for k = 1, …, 8 is much higher for the vessels from the normal traffic flow as compared with an abnormally acting vessel;
3) the variance decreases with the growing distance between the vessel and the normal traffic flow.

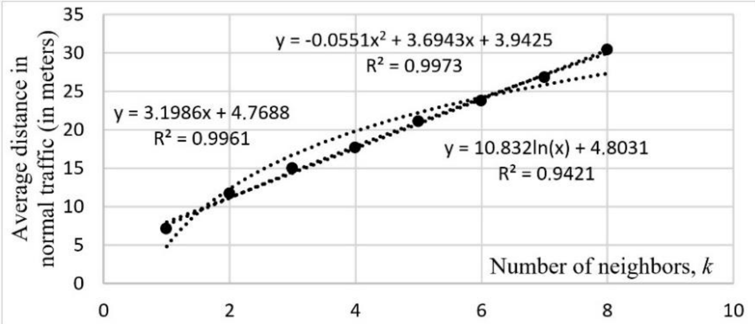

Fig. 8 Average distance among the neighboring points in normal traffic is dependent on *k*

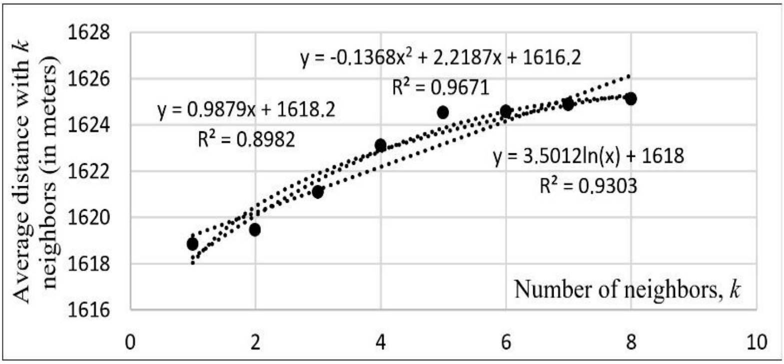

Fig. 9 Average distance among the anomaly case No. 4 and the neighboring points in expected traffic in dependence on *k*

Table 1. Variances of estimates

Points		Variance in meters
Points in traffic	1	72.71
	2	52.83
	3	25.49
	4	52.51

	5	38.72
	6	28.09
	7	139.36
	8	22.72
	9	70.82
	10	84.57
Points out of traffic	1	15.65
	2	10.16
	3	8.00
	4	5.70
Average for normal traffic		53.92

Anomaly detection is crucial to predicting dangerous situations. It is essential to make the right decision for safe marine navigation. An analysis of the *k*-nearest neighbors is applied for the marine traffic anomaly detection.

This method could be beneficial for modeling marine traffic flow. Besides, this method could detect marine traffic anomalies on many vessels at once and in real-time. The presented approach provides a solution for marine surveillance and marine traffic situational awareness.

3. NOVEL MACHINE LEARNING APPROACH FOR SELF-AWARE PREDICTION BASED ON CONTEXTUAL KNOWLEDGE

3.1 Decision tree algorithm for turning point prediction

The decision tree (DT) algorithm is a method for non-parametric supervised machine learning. This method could be used for pattern recognition and classification problems [19] [20] [21]. DT processes the data by looping partitions to achieve the uniform classification of the target value. At each split, DT reduces the entropy of the target value. This is achieved by adjusting the optimal split of independent variables. DT has unique abilities to create a sequence of decisions. This feature makes DT high-speed problem solvers. Moreover, DT is easy to understand and interpret. However, DT is hard to optimize.

In our case, DT is used to learn the routes by data aggregated using Ordering points to identify the clustering structure algorithm (OPTICS). DT primary purpose is to predict the future turning point (point C in **Fig. 10**). The main problem is that DT has a wide choice of possible turn points, but it should recommend the most suitable one (C). The data item for DT training consists of six features: (x_1, x_2) the current vessel position (Latitude & Longitude) describing the point A, (x_3, x_4) the kinematic data (SOG – speed over ground, COG – course over ground), (x_5, x_6) the following position (Latitude & Longitude) describing the point B, and the target data (Latitude & Longitude) describing the point C. See **Fig. 10** for an example of training data.

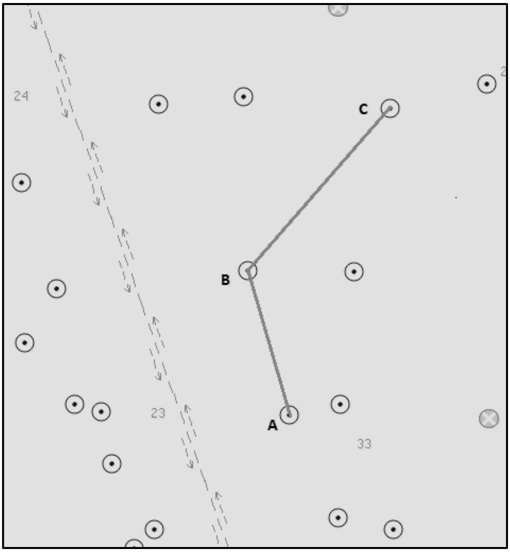

Fig. 10 Example of visual presentation of training data

DT tries to learn the context of marine traffic. The trained model is devoted to a particular geographical location because of the data used for learning.

The DT has a *maxdepth* parameter. Small *maxdepth* size may decrease the number of features used for decision. DT is realized in the integrated *scikit-learn* tool [24]. The DT is learning by splitting nodes with samples. It starts from the Root Node (with 393124 samples in our case). The mean squared error (*mse*) is used to split the node into sub-nodes. DT calculates *mse* for each divided subset and chooses the result with the smallest *mse* value. Each sub-node consists of a different number of samples.

3.2 Novelty detection based on SVM

Several different methods that involve the navigation experience to predict the next waypoint have been developed. These methods are based on historical marine traffic data. One of the most complicated problems is to filter anomalies, other unusual events, or dangerous situations from the entire historic marine traffic data. The simplest way to filter such data is to search and analyze every accident or dangerous maneuvering. It is challenging for humans because they must find or have prior knowledge about such cases or filter and extract the respective data by hand.

Despite this, the Novelty detection principle was used to tackle this task. Novelty detection is the ability of an intelligent organism to identify an utterly unknown pattern. The pattern should be sufficiently salient or associated with high positive or negative usefulness. This principle came from neurophysiology proposed by Yevgeny Sokolov [22] in the 1950s. This method is widely used to detect anomalies or outliers. The principles of Novelty detection are given in [23] and [24].

The SVM One-Class algorithm may serve Novelty detection, i.e., to filter the dangerous or unusual maneuvering situations from navigation data. In general, SVM arranges the dataset on a hyperplane or a set of hyper-planes in a high or infinite-dimensional space. This space is used for classification, regression, or other tasks. The best results could be achieved by the hyperplane, which has the most considerable distance to any class nearest data points (operating margin). The SVM One-Class algorithm learns to classify new data as similar or different from the training set. [25] has suggested a method for solving the One-Class classification problem using SVM.

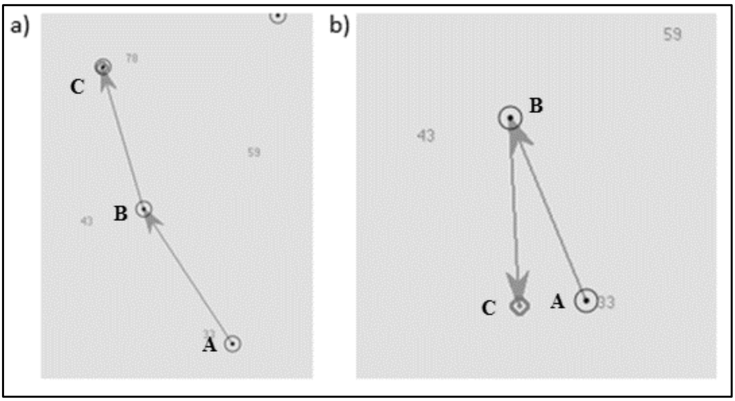

Fig. 11 Comparison of normal (a) and unusual (b) vessel track

The typical vessel path should be low-pitched with the lowest possible angles, as shown in **Fig. 11**. There are many practical reasons for it – fuel economy, vessel size, inertia etc. Nevertheless, there could be many reasons for an unusual path, as shown in **Fig. 11b**. For example, it could be the sailing vessel going by changing downwind, the fishing vessel with nets, law enforcement vessel on duty etc.

As mentioned before, DT is trained, and prediction is based on the historical vessel routes. The One-Class SVM algorithm was trained for a completely different task. SVM should evaluate the DTs prediction, i.e., the future turning point C predicted by DT should be assessed by SVM.

The data item for SVM training consists of (1) the distance $dist_1$ between points A and B, (2) the distance $dist_2$ between points B and C, (3) the angle ∠ABC. See **Fig. 12** for an example of training data.

$dist_1$	$dist_2$	$\angle ABC$
6174.52	1287.87	151.20
671.35	121.82	129.92
1526.95	1590.68	57.44
1024.27	24.30	82.41
6029.97	44602.39	163.37
...	...	...
2500.94	310.07	100.69
2762.53	1864.93	144.60
8158.88	24575.09	172.39
4293.26	20709.93	172.25
5300.63	4825.68	160.69

Fig. 12 Example of One-Class SVM classification

One-Class SVM [25] is realized in the integrated tool *scikit-learn* [26].

3.3 Training and results of the proposed approach

The proposed approach in Section 4 consists of the integral use of two different machine learning methods. We need to train the corresponding classifiers. As it is shown in **Fig. 13**, firstly we need is to get historical marine traffic data. The next is the aggregation of a dataset for better training results. The historical marine data are aggregated by OPTICS. The result is a dataset of the centers of clusters of turning points is generated. After this, the training process is divided into two main parts:

1. DT is trained to predict the future point.
2. One-Class SVM classifier is trained to detect unusual patterns. This step needs additional clustered data processing:
    a. Calculate distances between pairs of turning points (A, B) and (B, C);
    b. Calculate the angle ∠ABC.
3. The trained models are saved for future usage.

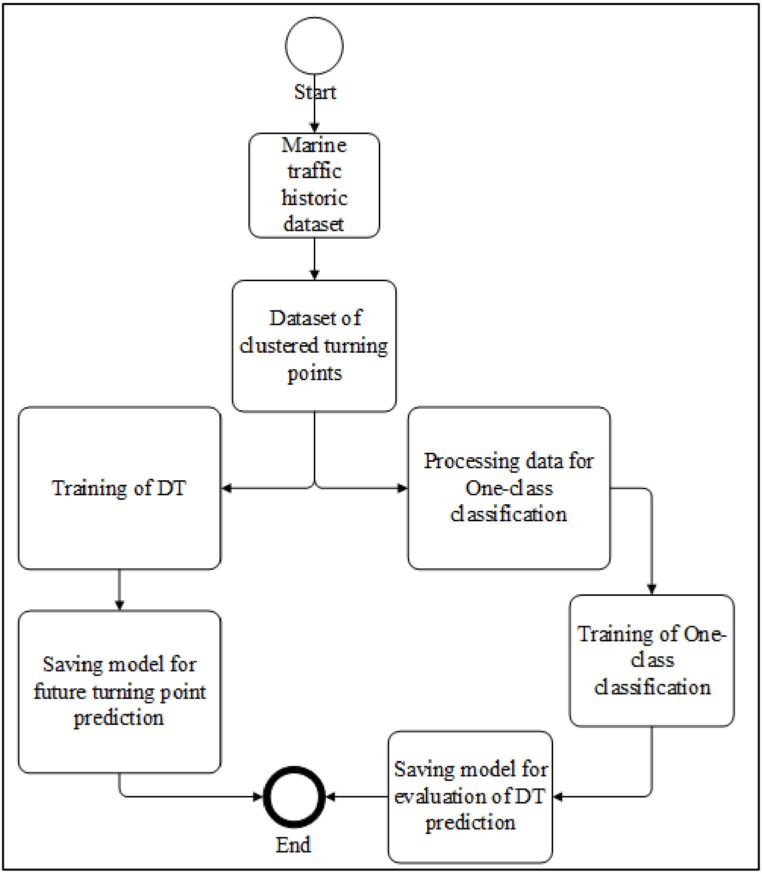

Fig. 13 Training process

The main aim of training DT is to learn a particular traffic pattern in a particular geographical area. It is quite a complex problem because the DT method should predict the future turn point (C) by the current position (A) of the vessel, its course, and speed. The experiments were carried out to evaluate the impact of parameter *maxdepth* on the prediction accuracy. *maxdepth* values were varied from 3 to 37. The results are given in **Fig. 14**. The DT with a *maxdepth* equal to 26 reaches more than 90% accuracy. The maximum accuracy is 94% as *maxdepth* equal to 37.

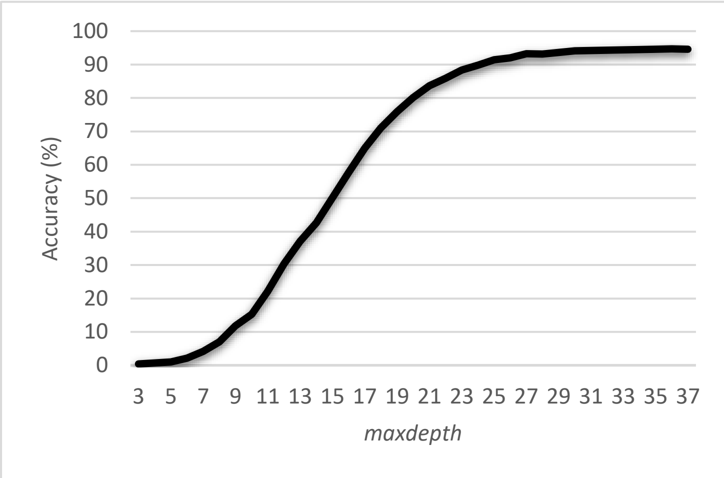

Fig. 14 Precision of the DT model with different *maxdepth*

An example of an anomaly detected by the One-Class SVM classifier is given in **Fig. 15**. In general, the One-Class SVM classifier assigned 5.86% of DT predictions as unusual – 27,125 cases from 462,499.

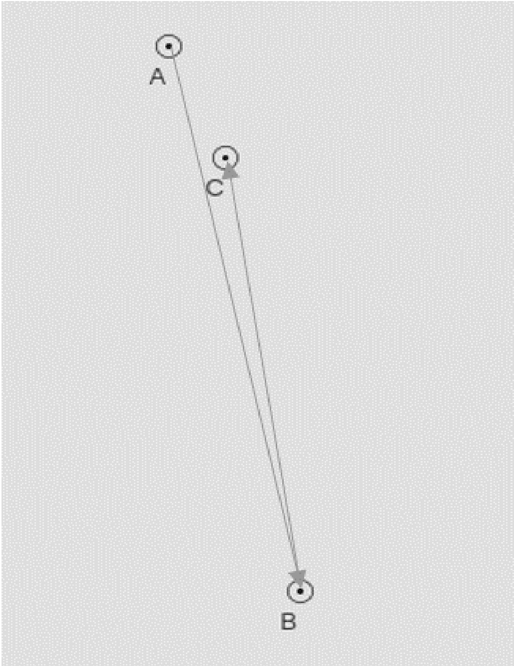

Fig. 15 Example of an anomaly detected by One-Class SVM classifier as a result of the analysis

4. SUMMARY AND CONCLUSIONS

The different methods were examined and proposed in the dissertation. These methods are designed to solve the tasks of ensuring safe navigation. The dissertation offers various methods to predict the future turning point or route of the monitored vessel and evaluate the vessels' maneuvers with other vessels. These methods are based on machine learning.

Main results:

- An analytical review was made of vessel navigation technologies, methods, and route planning algorithms which are designed to ensure safe navigation in the marine traffic and route planning. At present, marine traffic safety is ensured by a small number of dedicated technologies for marine situation surveillance. In marine navigation, intelligent technologies are not applied to decision making due to the complex operating environment. The focus concentrates on informing the navigator, who has to make decisions. However, there is a strong need for intelligent technologies in ensuring safe navigation, and this need will only increase soon. To ensure safe navigation, the marine traffic environment is too complex for the application of route planning or obstacle avoidance designed. These algorithms do not plan the best route but calculate the shortest one. However, such a solution is not good because the algorithm plane does not include other factors that determine safe navigation or maneuvering.
- Predicting a vessel turning point or route is an important task to ensure safe navigation. Knowing the future turning point or route of the vessel being monitored makes it possible to predict a possible future dangerous situation. It creates an opportunity to plan the maneuver and decide how to avoid a collision or such a dangerous situation safely. Each navigator must evaluate and make decisions independently. The

methods have been developed to predict the future nearest turning points of the monitored vessel to facilitate decision-making. A method has also been developed to detect an anomaly in the overall marine traffic flow. The application of this method helps assess the current situation in real-time and predict the actions of other participants. Such assessment is an essential condition for ensuring safe vessel navigation.

- It is not enough to predict the future maneuvers of the monitored vessel to ensure safe vessel navigation in marine traffic. These future maneuvers need to be assessed. In the general case, it is difficult to evaluate the maneuver performed because it is complicated to identify and link the cause-and-effect relationships of the particular maneuver. Such connections would explain why a maneuver is being performed and what subsequent maneuvers in that situation would be. The method proposed in the dissertation is intended for estimating the future maneuver of a vessel before it occurs. It is based on the application of contextual knowledge to assess a specific, future maneuvering situation. This machine learning method combines two different methods of machine learning. The uniqueness of the proposed method is that it is designed to assess the future maneuver based on contextual knowledge before it occurs. Methods similar to the proposed contextual knowledge-based method have not been identified in the scientific literature, making comparisons difficult. Nevertheless, the experiments performed proved the applicability of the proposed method.

After the experimental research, the following conclusions were formulated:

- Experiments were performed to test the effectiveness of route planning, obstacle avoidance, and deep reinforced learning algorithms. The tests have shown that those algorithms do not provide safe route planning as they cannot consider many factors affecting navigation. The navigators must critically evaluate the results of those algorithms. The suggested decision can be an obviously illogical route for the destination.

- The performing experiments with a deep reinforced learning algorithm, the agent achieves a good enough 87.7% route planning accuracy. After complicating the task for this algorithm, with the addition of (a) obstacles and (b) the changing positions of the destination point at each stage, only got: (a) accuracy of 48.1% and (b) accuracy of 50.77%, respectively.
- The prediction accuracy of 97.53 % is achieved by predicting the vessel route turning points and routes of monitored vessels using MLNN. Therefore the proposed method is suitable for practical use. With this method, the turning points of the monitored vessel were predicted. It is possible to create preliminary routes of the monitored vessels based on the prediction. The purpose of such predictions is not to make the possible route as accurate as possible but to predict the forthcoming maneuvers of the observed vessel.
- A relationship was found between the average distance between *k* neighboring vessels and whether the vessels are in the expected traffic or are outside the expected marine traffic. In the case of navigational data, the dependence of the average distance between neighboring vessels *k* on *k* is linear to non-linear (logarithmic) for vessels that are outside the total marine traffic flow. This dependency can be applied to real-time marine traffic monitoring.
- In the case of automatic monitoring of marine traffic flows, it is crucial to detect abnormally moving vessels algorithmically. These features can detect a vessel beyond the overall marine traffic flow:
    - The average distance between a vessel not in the marine traffic flow and several nearest neighboring

vessels in the marine traffic flow is significantly greater than the average distance calculated for vessels in the entire marine traffic flow;
    - The average variance of the distance between the selected vessels and several nearest neighboring vessels is significantly higher for vessels selected from the marine traffic flow than vessels outside the total marine traffic flow. This variance decreases with increasing distance between the vessel and the typical marine traffic flow.
- By applying data filtering and clustering with the OPTICS algorithm, the sample of data for further experiments was significantly reduced. The initial data sample was 19 GB, and 237 MB of data sample was obtained after clustering. The performed experiments showed that the OPTICS algorithm is more suitable for working with geographic marine traffic data than DBSCAN. Also, to obtain more accurate predictions, it is expedient to use the most significant possible sample of historical navigational data, filtering them according to the need, for example, type of vessels, route, seasons etc. Such filtering improves clustering and prediction.
- Experiments have shown that the decision tree algorithm is suitable for predicting the future turning point of the monitored vessel. Experiments were performed to evaluate the influence of the decision tree parameter *maxdepth* on the prediction accuracy. By changing the estimate of the *maxdepth* parameter from 3 to 37, the maximum prediction accuracy of 94 % was achieved.
- By applying the One-class SVM algorithm to detecting unusual situations among 462,499 situations, as many as 27,125 unusual situations were detected, which accounted for 5.86 % of all investigated situations.

REFERENCES

[1] Tsou M. C.: Online analysis process on Automatic Identification System data warehouse for application in vessel traffic service. *Proceedings of the Institution of Mechanical Engineers Part M: Journal of Engineering for the Maritime Environment* 230, 1, 199–215, 2016. Doi: 10.1177/1475090214541426

[2] Eriksen T., Høye G., Narheim B., Meland B. J.: Maritime traffic monitoring using a space-based AIS receiver. *International Astronautical Federation - 55th International Astronautical Congress 2004*, vol. 8, 5276–89, 2004.

[3] Fujii Y., Shiobara R.: The analysis of traffic accidents. *Journal of Navigation* 24, 4, 534–43, 1971. Doi: 10.1017/S0373463300022372

[4] Celik M., Cebi S.: Analytical HFACS for investigating human errors in shipping accidents. *Accident Analysis and Prevention* 41, 1, 66–75, 2009. Doi: 10.1016/j.aap.2008.09.004

[5] Mazaheri A., Montewka J., Kujala P.: Correlation between the ship grounding accident and the ship traffic – A case study based on the statistics of the Gulf of Finland. *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation* 7, 2, 119–24, 2013. Doi: 10.12716/1001.07.01.16

[6] Eide M. S., Endresen Ø., Brett P. O., Ervik J. L., Røang K.: Intelligent ship traffic monitoring for oil spill prevention: Risk based decision support building on AIS. *Marine Pollution Bulletin* 54, 2, 145–8, 2007. Doi: 10.1016/j.marpolbul.2006.11.004

[7] Jin M., Shi W., Lin K., Li K. X.: Marine piracy prediction and prevention: Policy implications. *Marine Policy* 108, 2019. Doi: 10.1016/j.marpol.2019.103528

[8] Prabowo A. R., Bae D. M.: Environmental risk of maritime territory subjected to accidental phenomena: Correlation of oil spill and ship grounding in the Exxon Valdez's case. *Results in Engineering* 4, 100035, 2019. Doi: 10.1016/j.rineng.2019.100035

[9] Jakovlev S., Daranda A., Voznak M., Lektauers A., Eglynas T., Jusis M.: Analysis of the Possibility to Detect Fake Vessels in the Automatic Identification System. *2020 61st International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS)*, IEEE, 1–5, 2020.

[10] Longépé N., et al.: Completing fishing monitoring with spaceborne Vessel Detection System (VDS) and Automatic Identification System (AIS) to assess illegal fishing in Indonesia. *Marine Pollution Bulletin* 131, 33–9, 2018. Doi: 10.1016/j.marpolbul.2017.10.016

[11] Hodge V. J., Austin J.: A survey of outlier detection methodologies. *Artificial Intelligence Review* 22, 2, 85–126, 2004. Doi: 10.1007/s10462-004-4304-y

[12] Venskus J., Treigys P., Bernatavičienė J., Tamulevičius G., Medvedev V.: Real-time maritime traffic anomaly detection based on sensors and history data embedding. *Sensors* 19, 17, 3782, 2019. Doi: 10.3390/s19173782

[13] Wang Y., Han L., Liu W., Yang S., Gao Y.: Study on wavelet neural network based anomaly detection in ocean observing data series. *Ocean Engineering* 186, 2019. Doi: 10.1016/j.oceaneng.2019.106129

[14] Ester M., Kriegel H., Sander J., Xu X.: A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, 226–31, 1996.

[15] Heidari S., et al: Big data clustering with varied density based on MapReduce. *Journal of Big Data* 6, 1, 77, 2019. Doi: 10.1186/s40537-019-0236-x

[16] Tang J., Liu F., Wang Y., Wang H.: Uncovering urban human mobility from large scale taxi GPS data. *Physica A: Statistical Mechanics and Its Applications* 438, 140–53, 2015. Doi: 10.1016/j.physa.2015.06.032

[17] Zhou A.: Approaches for scaling DBSCAN algorithm to large spatial databases. *Journal of Computer Science and Technology*

15, 6, 509–26, 2000. Doi: 10.1007/BF02948834

[18] Daszykowski M., Walczak B., Massart D.: Looking for natural patterns in data. Part 1. Density-based approach. *Chemometrics and Intelligent Laboratory Systems* 56, 2, 83–92, 2001. Doi: 10.1016/S0169-7439(01)00111-3

[19] Quinlan J.: Induction of decision trees. *Machine Learning* 1, 1, 81–106, 1986. Doi: 10.1007/bf00116251

[20] Wang X., Chen B., Qian G., Ye F.: On the optimization of fuzzy decision trees. *Fuzzy Sets and Systems* 112, 1, 117–25, 2000. Doi: 10.1016/S0165-0114(97)00386-2

[21] Murthy S., Kasif S., Salzberg S.: A System for Induction of Oblique Decision Trees. *Journal of Artificial Intelligence Research* 2, 1–32, 1994. Doi: 10.1613/jair.63

[22] Sokolov E.: Neuronal models and the orienting reflex. In: Mary A.B. Brazier, editor. *The Central Nervous System and Behavior*, New York: Macey, 187–276, 1960.

[23] Markou M., Singh S.: Novelty detection: A review - Part 1: Statistical approaches. *Signal Processing* 83, 12, 2481–97, 2003. Doi: 10.1016/j.sigpro.2003.07.018

[24] Markou M., Singh S.: Novelty detection: A review - Part 2:: Neural network based approaches. *Signal Processing* 83, 12, 2499–521, 2003. Doi: 10.1016/j.sigpro.2003.07.019

[25] Schölkopf B., Platt J. C, Shawe-Taylor J., Smola A. J., Williamson R. C.: Estimating the support of a high-dimensional distribution. *Neural Computation* 13, 7, 1443–71, 2001. Doi: 10.1162/089976601750264965

[26] Pedregosa F., et al: Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 2825–30, 2011.

[27] Machine Learning. Available at: https://www.ibm.com/cloud/learn/machine-learning/ [Date accessed: February 7, 2021]

LIST OF PUBLICATIONS ON THE TOPIC OF THE DISSERTATION

The main results of the research have been presented at international and republican conferences.

Research results have been presented at these international conferences:

1. Databases and Information Systems: 12th International Baltic Conference, DB&IS 2016, Riga, Latvia, July 4-6, 2016.
2. The 61TH international scientific conference of Riga technical university: IEEE Section of Information Technology and Management Science, Riga, Latvia, October 15-16, 2016.
3. WorldCist'21 - 9th World Conference on Information Systems and Technologies, in Terceira Island, Azores, Portugal, 30-31 March to 1-2 April 2021.

Articles printed in peer-reviewed scientific journals, peer-reviewed Web of Science Clarivate Analytics, with citation rate:

1. Jakovlev, S., Andziulis, A., Daranda, A., Voznak, M., & Eglynas, T.: Research on ship autonomous steering control for short-sea shipping problems. *Transport*, 32, 2, 198–208, 2017. https://doi.org/10.3846/16484142.2017.1286521
2. Daranda A., Dzemyda G.: Navigation decision support: discover of vessel traffic anomaly according to the historic marine data. *International Journal of Computers Communications & Control*, 15, 3, 2020. ISSN 1841-9844. https://doi.org/10.15837/ijccc.2020.3.3864.
3. Daranda A., Dzemyda G.: Novel machine learning approach for self-aware prediction based on the contextual knowledge. *International Journal of Computers Communications & Control*. 2021 (Accepted).

Articles printed in peer-reviewed scientific journals referenced in Web of Science Clarivate Analytics:

4. Daranda A.: Neural network approach to predict marine traffic. Baltic J. Modern Computing, 4, 483–495, 2016.

Articles in peer-reviewed international conference proceedings:

5. Daranda A., Dzemyda G: Artificial intelligence based strategy for vessel decision support system. In: *Proceedings of the WorldCist'21 - 9th World Conference on Information Systems and Technologies*, Terceira, pp. 49–58, 2021.
6. Jakovlev S., Daranda A., Voznak M., Lektauers A., Eglynas T., Jusis M.: Analysis of the possibility to detect fake vessels in the automatic identification system. In: *2020 61st International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS)*, IEEE, pp. 1–5, 2020.

Abstracts in conference proceedings:

7. Daranda, Andrius; Dzemyda, Gintautas: A Marine Traffic Prediction using Recurrent Neural Networks, *9th International workshop on Data Analysis Methods for Software Systems (DAMSS)*, Druskininkai, Lithuania, November 30 - December 2, 2017. Vilnius: Vilnius University, 2017. ISBN 9789986680642. p. 14-15.

Dirbtinio intelekto metodų panaudojimas saugiai laivybai užtikrinti

Santrauka

Šiuolaikiniame pasaulyje laivyba yra ypač svarbi transporto rūšis. Didžioji pasaulyje transportuojamų krovinių dalis gabenama jūriniu transportu. Krovinių pristatymas laivais yra ekonomiškiausias iš esamų transportavimo būdų, taip pat yra saugesnis ir ekologiškesnis. Nepaisant to, kad krovinių gabenimas laivais yra saugesnis, gana dažnai pasitaiko su laivais susijusių nelaimingų atsitikimų ir avarijų. Šių nelaimių pasekmės būna žymiai skaudesnės ir tragiškesnės lyginant su kitų transporto priemonių avarijomis, nes:

1. Gali žūti nepalyginamai daugiau žmonių. Tai yra ypač aktualu keleiviniams / kruiziniams laivams.
2. Prarandamas krovinys, kuris laivais gabenamas dideliais kiekiais, todėl nuostoliai būna žymiai didesni.
3. Gali kilti ekologinė katastrofa.

Per pastaruosius 50 metų laivyba buvo reglamentuota įvairiais teisiniais aktais, kurių laikymasis turi užtikrinti saugią laivybą: konvencija dėl tarptautinių taisyklių, padedančių išvengti laivų susidūrimų jūroje (angl. *International Regulations for Preventing Collisions at Sea 1972*), tarptautinė konvencija dėl žmogaus gyvybės apsaugos jūroje (angl. *Safety of Life at Sea*), konvencija dėl jūrininkų rengimo, atestavimo ir budėjimo normatyvų (angl. *Standards of Training, Certification and Watchkeeping for Seafarers*) (STCW) ir t. t. Šiomis konvencijomis siekiama užtikrinti žmonių gyvybių saugumą jūroje, nustatant tam taikomus vienodus darbo principus, standartus bei taisykles laivuose.

Taip pat didelė pažanga įgyvendinta pritaikant technologijas laivams – Automatinę identifikacinę sistemą (AIS), elektroninių jūrlapių vaizdavimo ir informacijos sistemą (ECDIS), globalinę jūrų avarinio ryšio ir saugos sistemą (GMDSS) ir t. t. Šios technologijos

gerina laivybos saugumą, tačiau niekaip negali pakeisti laivavedžio, kuris priima sprendimus ir yra atsakingas už galimas pasekmes. Laivavedys remiasi visa prieinama informaciją apie esamą situaciją, analizuoja ir modeliuoja galimus sprendimus bei priima sprendimą, kaip elgtis konkrečioje situacijoje. Dėl didelių laivo gabaritų bei jo manevravimo savybių laivavedys turi iš anksto numatyti ir įvertinti kitų laivų manevrus. Laivavedys, įvertinęs riziką ir priimdamas sprendimą, prisiima visišką atsakomybę už galimas pasekmes. Todėl yra reikalingi dirbtinio intelekto metodai, pagrįsti mašininiu mokymusi ir skirti laivavedžio sprendimų paramai. Tokie metodai turi gebėti mokytis ir gerinti prognozavimo tikslumą be papildomo žmogaus įsikišimo. Tai ypač svarbu modeliuojant manevravimo situacijas ir rekomenduojant galimus sprendimus, siekiant išvengti pavojingų situacijų.

3.1. Tyrimų objektas

Disertacijos tyrimo objektas – mašininio mokymosi algoritmų taikymas laivų manevrams modeliuoti ir situacijai vertinti.

3.2. Darbo tikslas ir uždaviniai

Tikslas:

- Sukurti mašininiais mokymosi algoritmais grindžiamą metodą, kuris leistų prognozuoti ir įvertinti stebimų laivų manevravimą, siekiant užtikrinti saugią laivybą.

Uždaviniai:

- Analitiškai apžvelgti laivyboje naudojamus metodus ir algoritmus, skirtus saugiai laivybai užtikrinti; eksperimentiškai įvertinti taikymo ribotumus.
- Ištirti galimybes pritaikyti mašininio mokymosi algoritmus stebimų laivų judėjimui bei galimam manevrui prognozuoti ir įvertinti: prognozuoti būsimus stebimo laivo posūkio taškus bei aptikti nukrypimus nuo bendro maršruto (anomalijas).

- Sukurti mašininiu mokymusi grįstą kompleksinį metodą, skirtą prasilenkiančių laivų manevrų analizei, kai stebimo laivo manevras ne tik prognozuojamas pagal susidariusias aplinkybes, bet ir vertinamas remiantis kontekstinėmis žiniomis, gautomis iš istorinių laivybos duomenų.
- Siūlomiems metodams ir sprendimų efektyvumui pagrįsti, atlikti eksperimentinius tyrimus su didelės apimties laivybos istoriniais navigacijos duomenimis.

3.3. Tyrimų metodika

Šios disertacijos tyrimas pagrįstas šiais metodais:

1. Darbe naudojami analizės, lyginamosios analizės ir apibendrinimo metodai.
2. Darbo tikslui pasiekti ir uždaviniams spręsti analizuojami ir vertinami šiuo metu laivyboje taikomi metodai ir algoritmai, skirti saugiai laivybai užtikrinti, taip pat algoritmai, skirti planuoti maršrutą bei išvengti esamų kliūčių.
3. Mašininis mokymasis iš didelių istorinių laivybos navigacijos duomenų.

3.4. Mokslinis darbo naujumas

Disertacija yra moksliškai reikšminga dėl šių priežasčių:

5. Pasiūlyti ir eksperimentiškai ištirti algoritmai, skirti istoriniams navigacijos duomenims klasterizuoti.
6. Pasiūlyti ir eksperimentiškai ištirti metodai, skirti prognozuoti stebimo laivo būsimus artimiausius posūkio taškus, kuriais remiantis galima numatyti būsimą laivo manevrą.
7. Pasiūlyti ir eksperimentiškai ištirti metodai, skirti aptikti nukrypimą (anomaliją) nuo planuoto arba prognozuojamo maršruto į tam tikrą uostą.
8. Disertacijoje pateikiamas naujas kompleksinis atraminiais vektoriais ir sprendimo medžiais grindžiamas mašininio mokymosi metodas, paremtas kontekstine informacija

grindžiamu būsimo posūkio taško prognozavimu ir vertinimu. Ši prognozė vertinama siekiant įspėti apie neįprastą situaciją dar nesant jos požymiams. Tai ypač svarbu norint užtikrinti saugią laivybą. Metodas skirtas darbui su realiojo laiko laivybos navigacijos duomenimis.

3.5. Praktinė darbo reikšmė

Šiuo metu informaciją, kuri reikalinga užtikrinti saugiai laivybai, turi apdoroti ir įvertinti už laivo saugumą atsakingas asmuo. Esamos kompiuterinės sistemos naudojamos tik palengvinti navigacinės informacijos keitimuisi tarp laivybos dalyvių ir (arba) užtikrinti navigacinės informacijos vizualizavimui. Esant tiesioginio susidūrimo pavojui, tokios kompiuterinės sistemos perspėja laivavedį, tačiau šis perspėjimas paremtas laivų judėjimo krypties vektorių apskaičiavimu. Gilesnė analizė apie situaciją nėra atliekama kompiuterinėmis sistemomis. Todėl saugiai laivybai užtikrinti sprendimus visada priima atsakingas asmuo. Vis dėlto žmogiškasis faktorius yra viena iš pagrindinių nelaimingų atsitikimų jūroje priežasčių. Esant šiuolaikiniam laivybos intensyvumui, nebepakanka vien esamų technologinių sprendimų, skirtų tik laivybos navigacijos informacijai vaizduoti. Pasiūlyti metodai kompleksiškai sprendžia saugios laivybos užtikrinimo problemą. Naudojant pasiūlytus metodus galima išspręsti navigacijos duomenų analizės ir sprendimo priėmimo paramos uždavinius. Šie uždaviniai dėl autonominių laivų kūrimo poreikio tampa ypač aktualūs.

3.6. Ginamieji teiginiai

Disertacijos ginamieji teiginiai:

- Pasitelkus mašininio mokymosi algoritmais grįstus metodus, galima:
    - prognozuoti stebimo laivo būsimus artimiausius posūkio taškus;
    - modeliuoti visą laivo maršrutą į paskirties uostą;

- įvertinti, ar yra neįprastai judančių laivų bendrame laivybos sraute į tam tikrą paskirties uostą bei juos identifikuoti.

- Pasitelkus kontekstinę informaciją ir sujungus mašininiu mokymusi grįstus metodus galima ne tik kompleksiškai prognozuoti stebimo laivo būsimą artimiausią posūkio tašką, bet ir įvertinti to laivo būsimą manevrą atsižvelgiant į kontekstinius duomenis, bei jį įvertinti įvairių laivybos dalyvių atžvilgiu.

3.7. Darbo rezultatų aprobavimas

Pagrindiniai tyrimo rezultatai pristatyti tarptautinėse bei respublikinėse konferencijose.

Pranešimai skaityti šiose tarptautinėse konferencijose:

4. Databases and Information Systems: 12th International Baltic Conference, DB&IS 2016, Riga, Latvia, July 4-6, 2016.
5. The 61TH international scientific conference of Riga technical university: IEEE Section of Information Technology and Management Science, Riga, Latvia, October 15-16, 2020.
6. WorldCist'21 - 9th World Conference on Information Systems and Technologies, in Terceira Island, Azores, Portugal, 30-31 March to 1-2 April 2021.

Straipsniai recenzuojamuose mokslo žurnaluose, referuojamuose „Web of Science Clarivate Analytics“ ir turinčiuose citavimo rodiklį:

7. Jakovlev, S., Andziulis, A., Daranda, A., Voznak, M., & Eglynas, T.: Research on ship autonomous steering control for short-sea shipping problems. *Transport*, 32, 2, 198–208, 2017. https://doi.org/10.3846/16484142.2017.1286521
8. Daranda A., Dzemyda G.: Navigation decision support: discover of vessel traffic anomaly according to the historic

marine data. *International Journal of Computers Communications & Control*, 15, 3, 2020. ISSN 1841-9844. https://doi.org/10.15837/ijccc.2020.3.3864.

9. Daranda A., Dzemyda G.: Novel machine learning approach for self-aware prediction based on the contextual knowledge. *International Journal of Computers Communications & Control*. 2021 (Accepted).

Straipsniai recenzuojamuose mokslo žurnaluose, referuojamuose „Web of Science Clarivate Analytics“:

10. Daranda A.: Neural network approach to predict marine traffic. Baltic J. Modern Computing, 4, 483–495, 2016.

Straipsniai recenzuojamuose tarptautinių konferencijų mokslo darbuose :

11. Daranda A., Dzemyda G: Artificial intelligence based strategy for vessel decision support system. In: *Proceedings of the WorldCist'21 - 9th World Conference on Information Systems and Technologies*, Terceira, pp. 49–58, 2021.
12. Jakovlev S., Daranda A., Voznak M., Lektauers A., Eglynas T., Jusis M.: Analysis of the possibility to detect fake vessels in the automatic identification system. In: *2020 61st International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS)*, IEEE, pp. 1–5, 2020.

Pranešimų santraukos konferencijų leidiniuose:

13. Daranda, Andrius; Dzemyda, Gintautas: A Marine Traffic Prediction using Recurrent Neural Networks, *9th International workshop on Data Analysis Methods for Software Systems (DAMSS)*, Druskininkai, Lithuania,

November 30 - December 2, 2017. Vilnius: Vilniaus universitetas, 2017. ISBN 9789986680642. p. 14-15.

3.8. Disertacijos apimtis ir struktūra

Darbą sudaro įvadas, trys skyriai, išvados, literatūros sąrašas, autoriaus publikacijų disertacijos tema sąrašas.

Įvade pateikiami tyrimų sritis, objektas, darbo tikslas ir uždaviniai, tyrimų metodai, mokslinis darbo naujumas, praktinė darbo reikšmė, ginamieji teiginiai.

Pirmame skyriuje aptariamas pasirinktos temos aktualumas ir bendra problematika. Pateikiami laivyboje naudojamų metodų aprašymai ir algoritmai, skirti saugiai laivybai užtikrinti. Eksperimentiškai įvertinami tų algoritmų taikymo ribotumai.

Antrame skyriuje pateikiamas stebimo laivo būsimo artimiausio posūkio taško prognozavimo metodas, paremtas dirbtiniu neuroniniu tinklu, ir tiriamas jo taikymas modeliuojant laivo judėjimą. Taip pat pateikiamas metodas įvertinti, ar yra neįprastai judančių laivų bendrame laivybos sraute į tam tikrą paskirties uostą ir juos identifikuoti.

Trečiame skyriuje aprašomas mašininiu mokymusi grindžiamas metodas, gebantis prognozuoti stebimo laivo būsimą artimiausią posūkio tašką ir įvertinti to laivo būsimą manevrą atsižvelgiant į kontekstinius duomenis.

Disertacijos apimtis: 116 puslapių, 1 lentelė, 53 iliustracija ir 1 priedas. Disertacijoje remtasi 138 literatūros šaltiniais.

APIBENDRINIMAS IR BENDROSIOS IŠVADOS

Disertacijoje išnagrinėti ir pasiūlyti metodai, skirti spręsti saugios laivybos užtikrinimo uždavinius. Šiems uždaviniams spręsti disertacijoje pasiūlyti įvairūs metodai, leidžiantys prognozuoti būsimą stebimo laivo posūkio tašką arba maršrutą ir įvertinti laivo manevrus kitų laivų atžvilgiu. Šie metodai paremti mašininio mokymosi ir dirbtinio intelekto metodais.

Pagrindiniai rezultatai:

- Analitiškai apžvelgtos laivyboje naudojamos technologijos, metodai ir maršrutų planavimo algoritmai, skirti užtikrinti saugią laivybą bei maršrutų planavimą. Šiuo metu laivybos saugumas užtikrinamas nedideliu skaičiumi technologijų, skirtų laivybai stebėti. Laivyboje išmaniosios technologijos sprendimams priimti netaikomos dėl sudėtingos veikimo aplinkos. Pasirenkamas paprasčiausias būdas, t. y. susitelkiama į laivavedžio, kuris turi priimti sprendimus, informavimą. Tačiau pastebimas didelis poreikis taikyti išmaniąsias technologijas laivyboje, o ateityje šis poreikis tik didės. Saugiai laivybai užtikrinti maršruto planavimo / kliūčių išvengimo algoritmų taikymas yra sudėtingas, nes algoritmai neplanuoja geriausio maršruto, o apskaičiuoja trumpiausią. Tačiau tai nėra gerai, nes toks planavimas neįtraukia kitų faktorių, lemiančių saugią laivybą.
- Laivo posūkio taško ar maršruto prognozavimas yra svarbus uždavinys, siekiant užtikrinti saugią laivybą. Žinant būsimą stebimo laivo posūkio tašką arba maršrutą, įmanoma numatyti būsimą pavojingą situaciją. Atsiranda galimybė iš anksto planuoti manevrą ir priimti sprendimą, kaip saugiai išvengti susidūrimo ar pavojingos situacijos susidarymo. Kiekvienas laivavedys privalo vertinti ir priimti sprendimus savarankiškai. Sprendimo priėmimui palengvinti sukurti metodai, skirti prognozuoti būsimus artimiausius stebimo

laivo posūkio taškus. Taip pat sukurtas metodas, leidžiantis aptikti anomaliją bendrame laivybos sraute. Šio metodo taikymas padeda įvertinti esamą situaciją realiu laiku bei prognozuoti kitų dalyvių veiksmus. O tai yra svarbiausia sąlyga, siekiant užtikrinti saugią laivybą.

- Saugiai laivybai užtikrinti nepakanka vien tik numatyti būsimus stebimo laivo manevrus. Šiuos būsimus manevrus būtina įvertinti. Įprastai įvertinti atliekamą manevrą yra sudėtinga, nes nelengva nustatyti ir susieti priežastinius ir pasekminius atliekamo manevro ryšius. Tokie ryšiai paaiškintų, kodėl yra atliekamas manevras ir kokie galimi kiti manevrai toje situacijoje. Disertacijoje pasiūlytas metodas skirtas būsimam laivo manevrui įvertinti prieš jam įvykstant. Jis paremtas kontekstinių žinių taikymu, kuriuo siekiama įvertinti konkrečią būsimą manevravimo situaciją. Dirbtinio intelekto metodas sujungia du skirtingus mašininio mokymosi metodus. Pasiūlyto metodo išskirtinumas yra tas, kad jis yra skirtas būsimam laivo manevrui įvertinti prieš jam įvykstant, remiantis kontekstinėmis žiniomis. Panašių metodų į pasiūlytą kontekstinėmis žiniomis paremtą metodą mokslinėje literatūroje identifikuoti nepavyko, todėl sudėtinga atlikti palyginimą. Nežiūrint to, atlikti eksperimentai įrodė pasiūlyto metodo taikymo galimybes.

Atlikus eksperimentinius tyrimus, suformuluotos šios išvados:

- Atlikti eksperimentai, skirti patikrinti maršruto planavimo, kliūčių išvengimo arba gilaus sustiprinto mokymosi algoritmų efektyvumą, parodė, kad tie algoritmai neužtikrina saugaus maršruto planavimo, kadangi negali atsižvelgti į daugelį laivybai įtaką darančių faktorių. Tų algoritmų rezultatus laivavedžiui reikia vertinti kritiškai, nes gali būti pasiūlytas net ir nelogiškas maršrutas kelionės tikslo atžvilgiu.
- Atliekant eksperimentus su giliuoju sustiprintu mokymosi algoritmu, agentas pasiekia pakankamai gerą 87,7 % maršruto

planavimo tikslumą. Apsunkinus šiam algoritmui uždavinį, t. y. pridėjus: a) papildomų kliūčių ir b) pradėjus keisti gauto taško padėtį kiekviename etape, atitinkamai pasiektas tik: a) 48,1 % ir b) 50,77 % tikslumas. Šie eksperimentai parodė, kad esant sudėtingesnėms sąlygoms, giliojo sustiprinto mokymosi algoritmas netinka saugiam manevravimui planuoti ir atlikti.

- Prognozuojant stebimų laivų posūkio taškus ir maršrutus naudojantis DNT, pasiekiamas 97,53 % prognozavimo tikslumas, todėl pasiūlytas metodas yra tinkamas praktiniam panaudojimui. Šiuo metodu prognozuoti stebimo laivo posūkio taškai, o remiantis prognozėmis, galima net sudaryti preliminarius stebimų laivų maršrutus. Tokio prognozavimo tikslas yra ne kuo tiksliau sudaryti galimą maršrutą, o numatyti, kokie bus artimiausi stebimo laivo manevrai.
- Atrasta priklausomybė tarp vidutinio *k* kaimyninių laivų atstumo ir to, ar laivai yra tik bendrame sraute, ar jų yra ir už bendro laivybos srauto. Navigacijos duomenyse esant laivų, kurie yra už bendro laivybos srauto, kinta vidutinio atstumo tarp *k* kaimyninių laivų priklausomybė nuo *k* iš tiesinės į netiesinę (logaritminę). Šią priklausomybę galima pritaikyti vykdant laivybos stebėjimą realiuoju laiku.
- Vykdant automatinį laivybos srautų stebėjimą, svarbu algoritmiškai užtikrinti neįprastai judančių laivų aptikimą. Laivą, esantį už bendro laivybos srauto, galima aptikti pagal šiuos požymius:
  - Vidutinis atstumas tarp laivybos sraute nesančio laivo ir kelių artimiausių laivybos sraute esančių kaimyninių laivų yra daug didesnis už tokį vidutinį atstumą, apskaičiuotą laivams, esantiems bendrame laivybos sraute;
  - Vidutinė atstumo tarp pasirinkto laivo ir kelių artimiausių kaimyninių laivų dispersija yra daug didesnė iš laivybos srauto pasirinktiems laivams nei laivams, kurie yra už bendro laivybos srauto. Ši

dispersija mažėja didėjant atstumui tarp laivo ir bendro laivybos srauto.

- Pritaikius duomenų filtravimą ir klasterizavimą OPTICS algoritmu, sumažinta duomenų imtis tolesniems eksperimentams. Pradinė duomenų imtis buvo 19 GB, po klasterizavimo gauta 237 MB duomenų imtis. Atlikti eksperimentai parodė, kad OPTICS algoritmas yra tinkamesnis darbui su geografiniais laivybos duomenimis nei DBSCAN. Taip pat siekiant tikslesnių prognozių, tikslinga naudoti kuo didesnes navigacijos duomenų imtis, jas filtruojant pagal poreikį, pavyzdžiui, laivų tipą, maršrutą, metų laikus. Filtravimas pagerina klasterizavimą ir prognozavimą.
- Atlikus eksperimentus nustatyta, kad sprendimų medžio algoritmas yra tinkamas prognozuoti stebimo laivo būsimą posūkio tašką. Atlikti eksperimentai, siekiant įvertinti sprendimo medžio parametro *maxdepth* įtaką prognozavimo tikslumui. Keičiant parametro *maxdepth* įvertį nuo 3 iki 37 pasiektas maksimalus 94 % prognozavimo tikslumas.
- Pritaikius *One-Class SVM* algoritmą neįprastoms situacijoms aptikti tarp 462499 situacijų, aptiktos net 27125 neįprastos situacijos, kurios sudarė 5,86 % visų tirtų situacijų.

Andrius Daranda

MACHINE LEARNING-BASED PREDICTION OF THE BEHAVIOR OF MARINE TRAFFIC PARTICIPANTS AND DISCOVERING NON-STANDARD MARINE TRAFFIC SITUATIONS

Summary of Doctoral Dissertation
Technological Sciences
Informatics Engineering (T 007)
Editor Zuzana Šiušaitė

Andrius Daranda

MAŠININIU MOKYMUSI GRINDŽIAMAS LAIVYBOS EISMO DALYVIŲ ELGSENOS PROGNOZAVIMAS BEI NESTANDARTINIŲ LAIVYBOS SRAUTO SITUACIJŲ ATRADIMAS

Daktaro disertacijos santrauka
Technologijos mokslai
Informatikos inžinerija (T 007)
Redaktorė Jorūnė Rimeisytė-Nekrašienė

Vilniaus universiteto leidykla
Saulėtekio al. 9, LT-10222 Vilnius
El. p. info@leidykla.vu.lt,
www.leidykla.vu.lt
Tiražas 45 egz.