Baltic J. Modern Computing, Vol. 4 (2016), No. 3, 483-495

Neural Network Approach to Predict Marine Traffic

Andrius DARANDA

Vilnius University, Institute of Mathematics and Informatics
Akademijos St. 4, LT 08663, Vilnius, Lithuania

andrius.daranda@gmail.com

Abstract. The marine traffic has been significantly rising fast during the last period of time. One of the most important problems for prediction of the marine traffic is to find certain patterns movements of particular vessels. Even the modern navigation devices could not ensure full maritime picture in narrow waterways because of geographical structures and various disturbances. Moreover, all vessels have could not use autopilot in straits for these reasons; they have to operated manually according to rules. In this paper, investigate the problem of maritime traffic prediction by artificial neural network.

Keywords: Maritime traffic, AIS, Artificial neural network, DBSCAN algorithm, prediction.

1. Introduction

The increased demand for marine traffic safety and security issues is extremely important to predict the maritime traffic. The intelligent monitoring and predicting systems for marine traffic become necessary due to vessels congested zones. The main issue of such intelligent maritime traffic monitoring and predicting system is to know vessel route for the entire journey. For such intelligent maritime system, the information from vessel Automatic Identification System (AIS) could be used. AIS sends periodically messages about vessel port of destination. However, AIS has no information about vessel's turning points or planned route. So, we need to predict vessel turning regions in the planned route. It could decrease dramatically the risks of accident at marine traffic. Consequently, despite the considerable amount of intelligent features developed for air (Prandini, et al., 2011) and land transportation (Franke & Heinrich, 2002) systems to improve the navigation safety and security, those features are still under developing for marine vessel navigational systems, because the maritime surveillance has its specific (Hall & Linas, 1997).

In this paper, the model of intelligent maritime traffic predicting system is proposed. The aim of the system is to track and predict the vessel's turning points and the route. Clustering the turning regions is performed by DBSCAN (density-based spatial clustering of applications with noise) algorithm (Ester, et al., 1996). The route is computed by the artificial neural network using data from DBSCAN. The proposed model is tested by a huge amount of real marine traffic data.

2. Maritime traffic

2.1. Marine traffic monitoring challenges

The Vessel Traffic Services (VTS) is a marine traffic controlled service. The main purpose of VTS is to ensure safe navigation, safety of life at sea and the protection of the marine environment. The VTS are monitoring the vessels which are passing traffic separation schemes and warning them against any possible collision and potential accident. The VTS are always seeking new ways to attain the maximum benefit from information technology (Yip, 2008). The VTS are equipped with modern marine surveillance mechanisms for the marine traffic monitoring purpose: radars, automatic plotting aid (ARPA), AIS and long-range identification and tracking (LRIT) systems. These marine traffic surveillance systems are incorporated into Electronic Chart Display & Information System (ECDIS), marine vessels satellite tracking, network for maritime data exchange and other different navigational systems. Definitely, there are many challenges for such maritime surveillance systems (Lokukaluge, et al., 2012):

- Huge amount of surveillance territory;
- Large number of targets;
- Multi target observations;
- Noisy signal and geographical structures;
- Propagation of radio waves.

The marine traffic congestion zones are biggest challenge for maritime surveillance systems. Such zones are usually in harbors, straits and channels with overloaded marine traffics. Therefore, the marine traffic safety depends on basic AIS operation principles.

2.2. Background of AIS

The AIS is automated vessels tracking and identifying system. This system is an electronically based system that exchanges the navigational data with nearby vessels, AIS base stations, and satellites. The exchange of navigational data is especially important for traffic safety and navigation efficiency. The AIS periodically broadcasts vessels navigational data by the AIS messages over Very high frequency (VHF) radio electromagnetic waves. The AIS are integrated with vessel's positioning systems (GPS), gyrocompass, rate of turn indicator and with others marine navigational sensors. The AIS messages consist of the vessel International Maritime Organization number (MMSI number), current vessel position (latitude and longitude), course, speed and etc. These messages are transmitted directly to AIS receivers which are in vessels or VTS Base stations. The received navigational data could be displayed on ECDIS and on the Radar systems. Consequently, the vessels could be precisely tracked along the coast line by VTS. Typically, the AIS signal range is about 20-30 nautical miles and it is depending on VHF electromagnetic radio characteristics. Otherwise, the vessels could be tracked worldwide by the special satellites (Lokukaluge, et al., 2012). The AIS requirements and implementation plan are outlined in Subparagraph 2.4 of Regulation 19 of Chapter V of the International Convention for the Safety of Life at Sea (SOLAS) (Anon., 1974/1980). The system has been mandatory to all new ships in international traffic since 1 July

2002, and ended on 2004 include all passenger ships, tankers and other ships of 300 tons engaged in international voyages.

3. DBSCAN algorithm

3.1. DBSCAN algorithm

As a kind of density-based clustering, the DBSCAN algorithm was first introduced by (Ester, et al., 1996). The key idea in DBSCAN is that for each data object of a cluster, the neighbor-hood of a given radius (ε) has to contain at least a minimum number (MinPts) of objects (Zhou, et al., 2000). The clustering is unsupervised classification of patterns (observations, data items, or feature vectors) into groups (Güngör & Ünler, 2007). This classification process does not need prior any knowledge about the database data. The clustering process a set of data objects into clusters. The objects are more similar to each other in the cluster than the objects in the different clusters according to some predefined criteria (Güngör & Ünler, 2008). Also, such data objects are called data points or points, and the database is usually referred as a data set. The algorithm grows regions with sufficiently high density into clusters and discovers clusters of arbitrary shape in spatial databases with noise. Density-based clustering defines cluster as region, the objects of the region are dense. The clusters are separated from one another by low-density regions (Han & Kamber, 2006). The reason we choose density-based clustering is that it has significant advantages over partitioned and hierarchical clustering algorithms. It can discover clusters of arbitrary shapes. The computational complexity can be reduced by building some special data structures. In addition it is able to effectively identify noise points (Cao, et al., 2006) (Fig. 1).

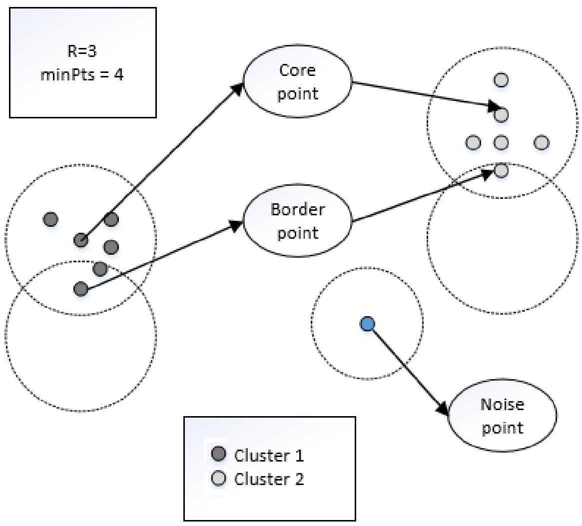

Fig. 1. DBSCAN clustering algorithm

Otherwise, the density-based clustering algorithms easily lead to memory problem when process the large databases.

4. Artificial Neural Network

4.1. Artificial Neural Network development

The Artificial Neural Networks (ANN) simulates the processes of the human brain. The ANN makes associations between varieties of the information. The ANN realizes as intuitive reasoning rather than the logical reasoning normally executed by machine. So, the ANN has the capability to learn with any complex structure of data. The main advantage of ANN is to incorporate uncertainty as well as data which is dynamic character has led to a number of studies to establish its applicability to predict marine traffic.

One of the ANN scheme is the back-propagation (BP) network (Haykin, 1999). The back-propagation neural network architecture is designed by fully interconnected layers or rows of processing units. The neural network learns the relationship between data using an algorithm known as "training" to modify weights according to the error between the predicted and the actual values until it learns the best relationship between inputs and outputs. The errors are calculated during this process. These errors are used to back-propagated from the output neurons to all the hidden neurons. All weights are adjusted by the errors. This learning process continues till the error is minimized to specified minimum value. After this process the weights are saved as ANN knowledge. The weights are used to perform information processing operation by back-propagation algorithm. The ANN may have as large numbers of neurons as it needed but it depends on system calculation capabilities. Likewise, the number of layer is exchangeable. The ANN behavior depends on the simple activation function which could be a linear or a non-linear. The ANN knowledge is training by pre-mined data. The ANN is fully connected by every neuron to every layer of neuron (Mehdi & Mehdi, 2010).

The output of each neuron was calculated by an activation function. All the inputs were summed by neurons with threshold. To develop the ANN, the pre-mined data was divided into two parts: 90% of turning points were used to train the ANN and the rest of observations were used for validation. All data was normalized to interval of [-1, 1].

The ANN is capable to implement a learning algorithm and made decision support ability of marine traffic prediction. A three-layer neural network model (Fig**.** 2) was developed to predict the turning regions of vessels in marine traffic. The ANN inputs were:

- Latitude of turning regions;
- Longitude of turning regions;
- Speed;
- Course;
- Vessel MMSI number;
- Vessel Dimension (length & width);
- Type of vessel.

The ANN outputs were used to predict the next turning regions. The ANN outputs:

- Latitude of next turning region;
- Longitude of next turning region;

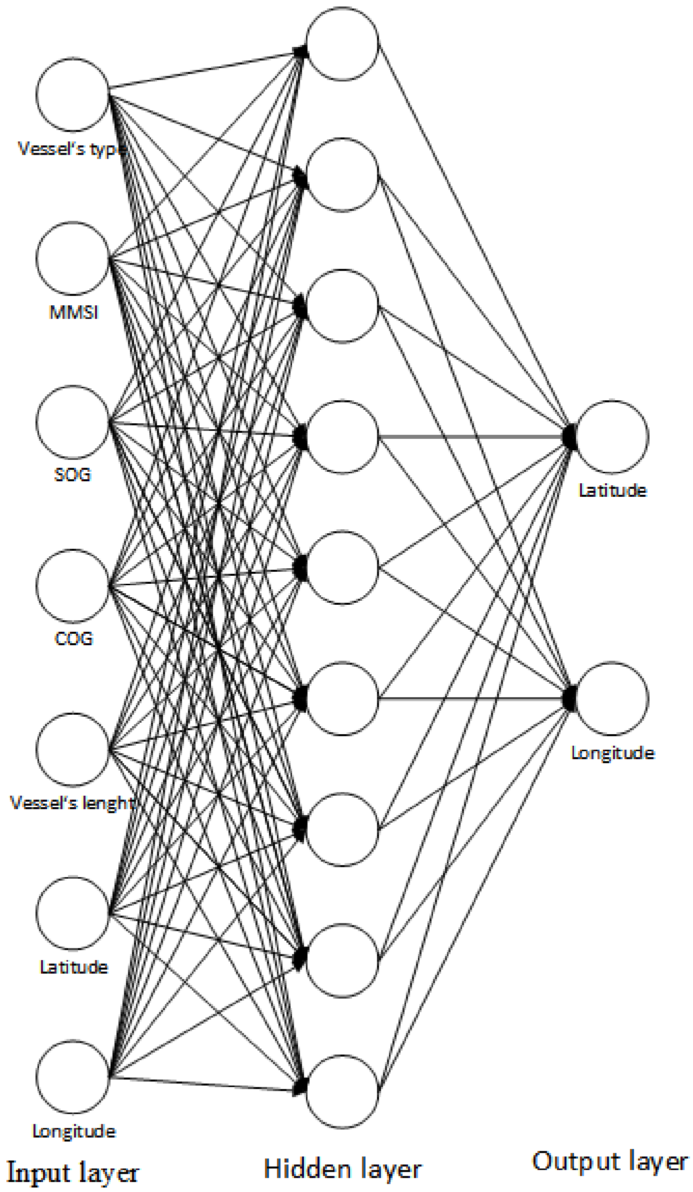

Fig. 2. Model of Artificial Neural network for prediction of next turning region coordinates

The algorithm was solved by python 2.7 (used the implementation of (Jones, et al., 2001)) with neuron package (Wojciechowski, 2010).

Every neuron in the hidden or output layer will firstly act as a summing junction which combines and modifies the inputs to each neuron from the previous layer, using Eq. (1):

$$S_i = \sum_{j=1}^{m} x_i w_{ij} + b_j \quad (1)$$

where S_i is the net input to node (neuron) j in hidden or output layer, x_i is the inputs to node j (or the outputs of the previous layer), w_{ij} is the weights representing the strength of the connection between the i-th node and j-th node, i is the number of nodes and b_j is the bias associated with node j. Then, the sum of bias (activation thresholds) and previously weighted inputs is passing through a suitable transfer (activation) function to determine its output (M.A. Razavi, 2003). There are various kinds of transfer function for artificial neural networks such as linear (purelin), hyperbolic tangent sigmoid (tansig), logarithmic sigmoid (logsig) and radial basis (radbas) transfer functions. The most popular transfer function for a non-linear relationship is the *logsig* function, that produces output values in the range of [−1, 1], respectively. The general forms of this function are indicated below:

Logarithmic sigmoid (logsig): $f\left(S_j\right) = \frac{1}{1+ e^{-Sj}}$ (2)

Where $f\left(S_j\right)$ is the output of node j. j is also an element of the inputs to the nodes in the next layer. A very common algorithm for training the network is the feed forward back-propagation algorithm. Fitting procedure from which the weights are determined is performed using a least-squared minimization routine in which the sum of the square of the errors between the calculated and the actual (target) data must be minimized.

In the neural network architecture, one hidden layer with sigmoid activation function has been used. The numbers of hidden neurons and hidden layers have been found by trial and error using collected data to obtain the best performance. Training and predicting processes are realized at 6 steps as explained below:

1. Collecting the whole data in one place.
2. Determining the train and test sets.
3. Converting the data into the ANN inputs.
4. Determining, training and testing the network topology.
5. Repeating the 1st, 2nd, 3rd and the 4th steps as long as it is required to determine the optimal model.
6. Application of the optimal ANN model.

In the proposed ANN model, MMSI number, vessel speed (SOG), vessel course (COG), length and vessel position (latitude & longitude) were chosen as input variables, while predicted vessel turning position (latitude & longitude) was selected as output variable. Since, other parameters (for example weather conditions, passing of other vessels and other obstacles at sea) were constant under all studied conditions. So the ANN was trained with the six input parameters to explore the best relationship between these parameters and vessel turning position (latitude & longitude) (as output) by the ANN.

4.2. ANN training with collected data

The data was collected from marinetraffic.com website and was stored into MySQL database. The stored data was used for training of the proposed model of the ANN. The data covered whole Baltic Sea. The mean count of vessels was 3000 on a time step of 180 seconds. The data was filtered by change of course to calculate vessel turning regions and vessel's speed. If vessel's course was changed more than 4° it is turning point of vessel's route. Likewise, vessel's which have less than 4 knots of speed weren't monitored. The monitoring lasted 10 days. More than 300,000 turning points, 30,000 vessels and 6,000 destinations were stored during this period. So, turning points data like latitude and longitude, vessel's type, course, speed, size and MMSI was stored in database. This data was used to train the ANN. The trained ANN has been used to predict next position of turning point (latitude & longitude). Moreover, the trained ANN has been used to predict whole route of vessel to port of destination. All stored data considered to decide model structure and used for training of ANN. The sampling stored period of 180 seconds was chosen because of maneuver performance of vessels. In this respect, three-minute-ahead positions are predicted by utilizing past data of vessel. The following ANN model has been trained and analyzed for the best performance.

5. Proposed model development

The proposed model has two main parts:

Clustering turning regions by DBSCAN algorithm;

The ANN learning clustered turning points;

The clustering data was implemented by Scikit-learn's (Pedregosa, 2011) python library. This library requires the input of a distance matrix showing pairwise distances between all turning regions in the dataset. So, the DBSCAN algorithm requires two parameters: ε (epsilon) and the minimum number of points required to form a cluster (minPts). Parameter ε depends on radius of cluster. It could not be equal for whole route because of narrow channel, harbors and etc. (Fig. 3).

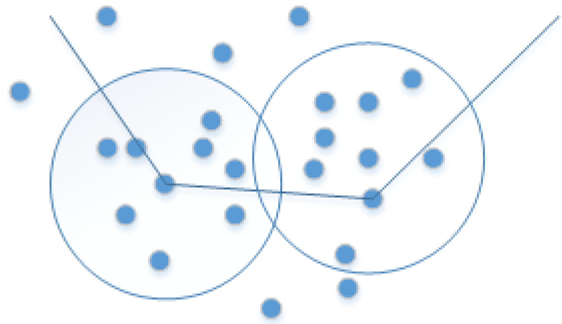

Fig. 3. Clustering turning regions with the some ε parameter

Therefore, the various parameter ε was chosen for open seas, harbors, straits and channels. In order to get optimal clusters, the proposed model calculates various parameters ε for four times. It calculates ε from biggest possible value and reduced on every cycle in order to minimize possible error.

As a second part of propose model is the ANN. The model was created by the three-layer neural network structure with a sufficient number of neurons in the hid-den layer and enough learning time can theoretically model any nonlinear functions at any accuracy. However, the questions in practical implementation are the selection of training algorithm, number of hidden neurons, learning rate, and momentum, to guarantee that the error converges to a desired value. The learning rate was chosen by trial and error method and it equal to 0.0009. If the learning rate was too large, oscillation or divergence occurred, whereas a small learning rate resulted in a slow training process without improvement in the accuracy of the proposed model. The training process was terminated when the half-sum of the squared error reached a threshold value of 23. This threshold corresponded to prediction accuracies above 99% for the training data-set.

For the proposed three-layer neural network model, the number of input neurons was determined by six particular vessel parameters (MMSI number, position (latitude and longitude), speed, course and port of destination) that affect turning points. The numbers of neurons in the hidden layer can be selected with flexibility but the number determines how well a data-set is trained. Too many hidden neurons lead to a data-set-specific relationship between inputs and outputs, or an “over-fitting” problem. With smaller numbers of hidden neurons, the network may not have adequate power to learn the patterns with satisfactory accuracy, resulting in an “under fitting” problem.

5.1. Prediction marine traffic with proposed model

The ANN was completed after training step. The input-output vectors have been modified as explained above for prediction. The proposed ANN model is using calculated weights for prediction of turning points. The next step is to calculate the nearest turning point (coordinates). It could be done by calculating distance (3, 4, 5, 6 and 10) to nearest cluster on the map. This uses the “haversine” formula to calculate the great-circle distance between two points – that is, the shortest distance over the earth’s surface – giving an ‘as-the-crow-flies’ distance between the points.

$$a = sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos(\varphi_1) * \cos(\varphi_2) * \ sin^2\left(\frac{\Delta\lambda}{2}\right) \tag{3}$$

$$c = 2 * atan2(\sqrt{a}, \sqrt{(1-a)}); \tag{4}$$

$$d = R * c \tag{5}$$

$$\theta = atan2(\sin(\Delta\lambda) * \cos(\varphi_2), \cos(\varphi_1) * \sin(\varphi_2) - \sin(\varphi_1) * \cos(\varphi_2) * \cos(\Delta\lambda)); \tag{6}$$

Midpoints of cluster are used to find nearest cluster and current vessel position. Midpoint is calculated by:

$$B_x = \cos(\varphi_2) * \cos(\Delta\lambda); \tag{7}$$

$$B_y = \cos(\varphi_2) * \sin(\Delta\lambda); \tag{8}$$

$$\varphi_m = atan2(\sin(\varphi_1) + \sin(\varphi_2), \sqrt{((\cos(\varphi_1) + B_x)^2 + B_y^2)}; \quad (9)$$

$$\lambda_m = \lambda_1 + atan2(B_y, \cos(\varphi_1) + B_x); \quad (10)$$

Where φ is latitude, λ is longitude, R is earth's radius (mean radius = 6,371 km);

Moreover, course must coincide with bearing (8, 9 and 10) to calculate nearest cluster. Determinate cluster is used to load the ANN for calculation. Therefore, every cluster has own ANN with calculated weights. The ANN loaded with vessel parameters (current position, course, speed, MMSI number, vessel type & vessel particulars) predict next turning point (latitude & longitude).

Usually, the vessel route consists of many turning points. The method has particularly been formed for prediction of vessel turning points. Moreover, the same way could be used for predicting all vessel routes. As shown in Fig. 4, the ANN could be used as many times as it need.

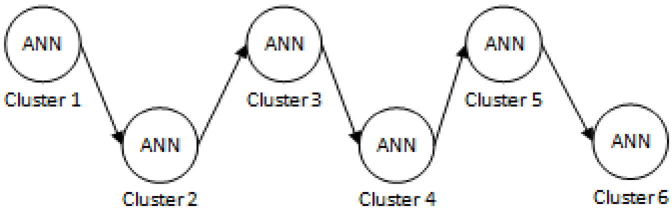

Fig. 4. Vessel's route prediction by clusters sequence

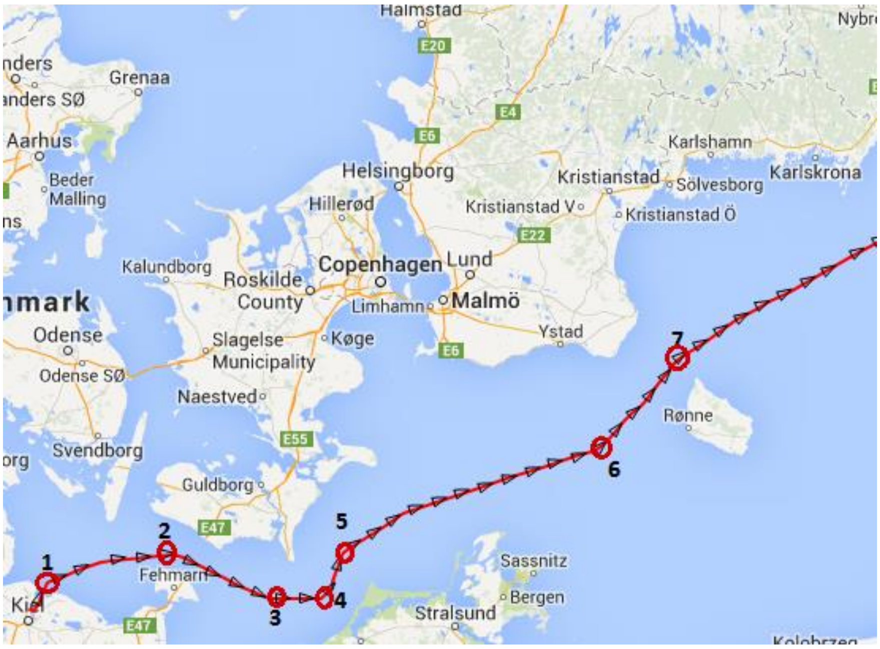

Fig. 5. Route prediction by proposed model

So, the proposed model could be used to predict turning point for predict next turning point (Fig. 5) and for whole vessel route.

As shown in Fig. 5 the route could be divided by turning regions. This region is different for vessels but it could be united by clusters.

6. Results

6.1. Clustering results

The large amount marine traffic data was used. This real live data collected from the vessel AIS transmitters. The collected data was recorded and accumulated in the database in order to test the proposed model. However, the data was filtered by turning points (4° change of vessel course). The data preprocessed to sort by port of destination. The set of data contains turning point by port of destination as shown in Fig. 6.

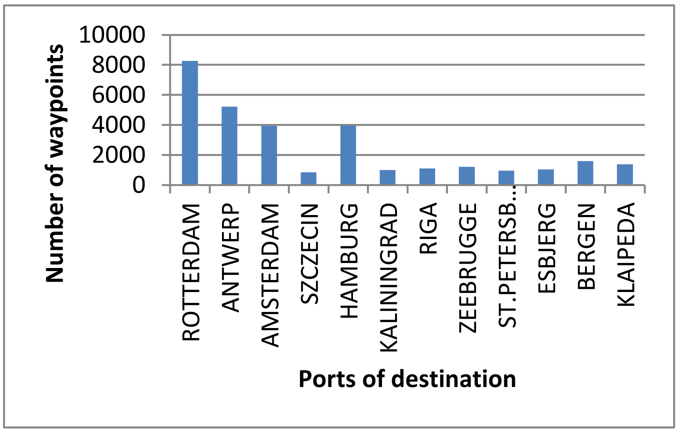

Fig. 6. The number of turning regions by port of destination

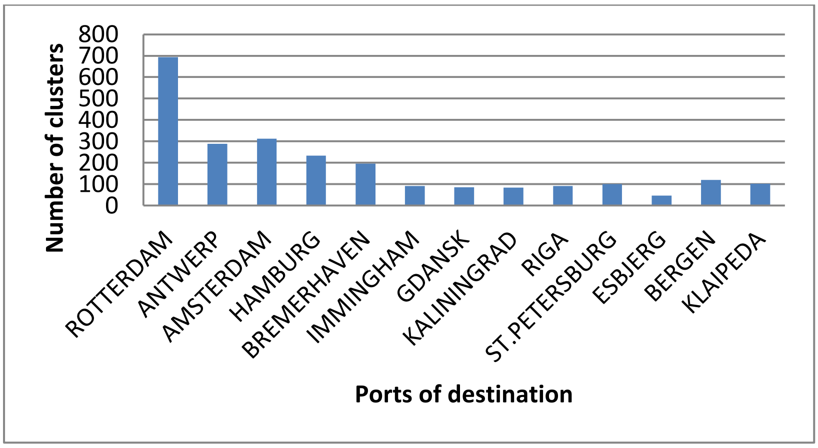

Fig. 7. The number of clusters by port of destination

This data set was clustered by DBSCAN algorithm. As a result we got clusters sorted by port of destination as shown in Fig. 7. The Rotterdam port have largest amount of clusters.

The number of turning points depends on DBSCAN clustering parameter ε. The cluster size depends on turning points count.

6.2. Artificial Neural Network results

The neural network model correctly predicted vessel route. The precision of ANN prediction depends on cluster size. Also it depends on DBSCAN algorithm parameter ε. The distance between real turning point and predicted turning point do not exceed cluster diameter (parameter ε) (Fig. 8).

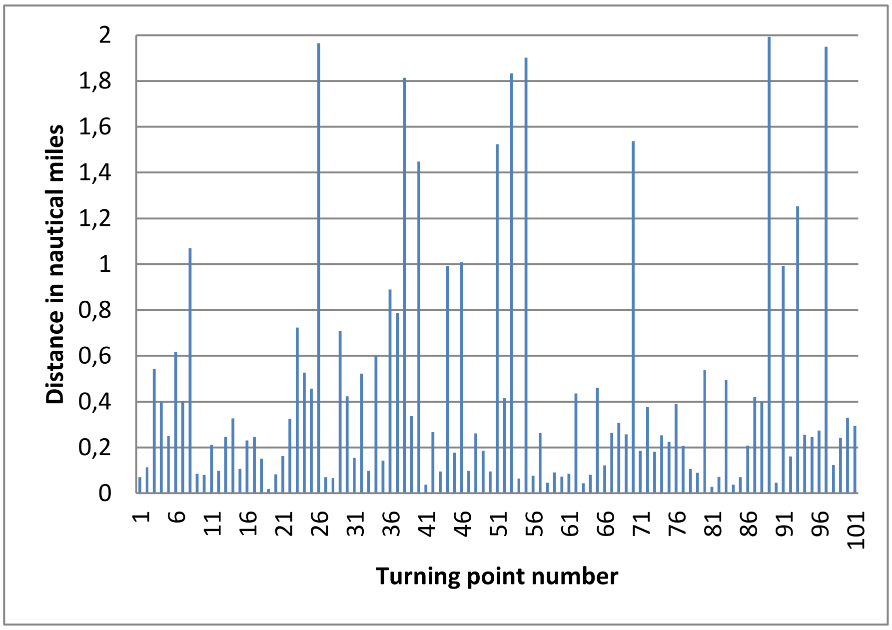

Fig. 8. The precision of predicted turning points

Therefore, the ANN could predict turning point with different precision in narrow channel or open seas. The precision could be increase by adjusting parameter ε.

7. Conclusion and future development

As mentioned before, the turning points are most important data for prediction and experiments. The proposed model satisfied the goal to predict vessel route.

The proposed prediction model is used DBSCAN algorithm and Artificial Neural Network. It proves the effectiveness to predict marine traffic. This model could be used for early warning and guidance systems before any dangerous situations occur. The proposed model could be used for any specific vessel to predict the route. Route prediction is shown in Fig. 9.

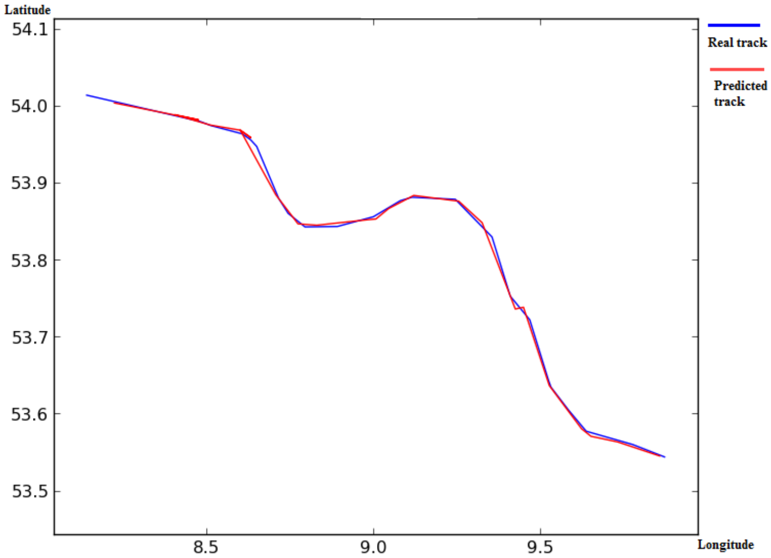

Fig. 9. The real and predicted route for 24 hours

The proposed model was developed and demonstrated to show its prosperous aspects. Conclusions can be drawn as follows.

The marine traffic prediction model is proposed and developed to demonstrate its effectiveness.

Marine traffic can be statistically analyzed and used for prediction.

The proposed model could be used for marine traffic flow estimation.

The proposed model can easily to predict future marine traffic situation. The system can be utilized to plan VTS and on-land AIS stations and to assess the influence of various waterway-area development projects.

The sampling should be stored at 60 seconds period to increase turning point precision.

References

Anon., 1974/1980. International convention for the Safety of Life at Sea (SOLAS), Chapter V Safety of Navigation, Regulation 19. *IMO.*

Cao, F., Ester, M., Qian, W., Zhou, A., 2006. *Desity-based clustering over an evolving data stream with noise.* Bethesda, s.n., p. 328–339.

Comaniciu, D., Meer, P., 1999. Distribution free decomposition of multivariate data. *Pattern Analysis, 2,* pp. 22-30.

Ester, M., Kriegel, H. P., Sander, J., Xu, X., 1996. *A density-based algorithm for dis-covering clusters in large spatial databases with noise.* Portland, s.n.

Franke, U., Heinrich, S., 2002. Fast obstacle detection for urban traffic situations. *IEEE Trans. Intell. Transp. Syst., vol. 3, no. 3,* p. 186–197.

Franke, U., Heinrich, S., 2002. Fast obstacle detection for urban traffic situations. *IEEE Trans. I ntell. Transp. Syst., vol. 3,* pp. 186-197.

Güngör, Z., Ünler, A., 2007. K-harmonic means data clustering with simulated annealing heuristic. *Applied Mathmatics and Computation, 184,* p. 199–209.

Güngör, Z., Ünler, A., 2008. K-harmonic means data clustering with tabu-search method. *Applied Mathematical Modelling,* pp. 1115-1125.

Haykin, S., 1999. *Neural Networks: A Comprehensive Foundation, second ed.*. Englewood Cliffs, NJ: Prentice-Hall.

Hall, D. L., Linas, J., 1997. An introduction to multisensor data fusion. *Proc. IEEE , vol. 85, no. 1,* p. 6–23.

Han, J., Kamber, M., 2006. *Data Mining: Concepts and Techniques.* s.l.:Morgan Kaufmann.

Hoye, G., Narheim, B., Eriksen, T. & Meland, B., 2004. Feasibility Study On Space Based AIS For Large-Area Surveillance Of Norwegian Waters. *FFI/RAPPORT-2004/01190.*

Yip, T. L., 2008. Transportation Research E44. *Port traffic risks – A study of accidents in Hong Kong waters,* pp. 921-931.

Jones, E., Oliphant, T., Peterson, P., 2001. *SciPy: Open source scientific tools for Python.* [Online] Available at: `http://www.scipy.org/`

Lokukaluge, P. P., Oliveira, P., Soares, C. G., 2012. Maritime Traffic Monitoring Based on Vessel Detection, Tracking, State Estimation, and Trajectory Prediction. *EEE Transactions on Intelligent Transportation Systems, vol. 13 , no. 3 ,* pp. 1188 -1200.

M.A. Razavi, A. M. M. M., 2003. Dynamic modeling of milk ultrafiltration by artificial neural network. *J. Membr. Sci.,* p. 47–58.

Mehdi, K., Mehdi, B., 2010. An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with Applications, Vol. 37,* p. 479–489.

Pedregosa, 2011. Scikit-learn: Machine Learning in Python. *JMLR 12,* pp. 2825-2830.

Prandini, M., Piroddi, L., Puechmorel, S., Brázdilová, S. L., 2011. Toward air traffic com-plexity assessment in new generation air traffic management systems. *IEEE Trans. Intell. Transp. Syst.,* p. 809–818.

Tran, T. N., Wehrens, R., Buydens, L., 2006. Buydens KNN-Kernel density-based clustering for high-dimensional multivariate data. *Computational Statistics and Data Analysis,* pp. 513-525.

Wojciechowski, M., 2010. *Feed-forward neural network for python, Technical University of Lodz (Poland), Department of Civil Engineering.* [Online] Available at: `http://ffnet.sourceforge.net/` [Accessed 2014].

Xu, J., Wunshc, D., 2005. Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks,* p. 645–678.

Zhou, A. et al., 2000. Approaches for Scaling DBSCAN Algorithm to Large Spatial Databases. *Journal of Computer Science and Technology, 15(6),* pp. 509-526.

Received May 26, 2016, accepted May 26, 2016