METEOROLOGICAL DATA INFLUENCE ON MISSING VESSEL TYPE DETECTION USING DEEP MULTI-STACKED LSTM NEURAL NETWORK

J. Venskus, P. Treigys
Vilnius University
Vilnius, LITHUANIA
e-mail: julius.venskus@mii.vu.lt

Abstract

Highly-loaded seaports have extremely complex and intensive marine vessel traffic, which generates large volumes of traffic data. Meteorological conditions and maritime vessel type influence maritime traffic and they must also be taken into account in order to train the model capable of recognizing the abnormal movement of the sea transport. Real data often misses some data values, such as type of vessel or its status. This paper reviews method of obtaining vessel traffic and meteorological data and filling missing vessel type data in Rotterdam port region. A deep multi-stacked LSTM neural network model is trained to fill the missing vessel type data. This model is trained with available vessel type data and used to predict missing values. This paper describes creation and evaluation of this model. Results of the experiment show it is expedient to use traffic data of a vessel in conjunction with meteorological data.
Keywords: data science, LSTM neural network, Vessel type detection, meteorology

1 Introduction

Maritime transport is one of the most important and intense sectors of human activity, accounting for about 90% of total trade. The high volume of vessel traffic generates large amounts of data, which overload various information systems and sensors [3]. Assistive systems are developed to facilitate the task, which extract the necessary information from the big data. One of the systems is an unusual traffic detection system, which requires full data for accurate detection. Unfortunately, the data that comes from different systems such as AIS, radars or satellite systems, is not full at all times [1]. The lack of such data prevents the creation of a sufficiently accurate model for detection of unusual vessel traffic. It is therefore necessary to develop smart systems for filling in the missing data, especially with the increased development of new methods for the detection of unusual traffic, which is essential for safety at sea [5]. This article offers a way to fill in the missing data for missing vessel types, which would allow for improved prediction of abnormal maritime traffic. The first part of the article introduces the developed method used to fill in the missing vessel type information in the data, and the second part describes the experiments with this method using vessel traffic data in the Rotterdam harbour. This research is continuous work in field of abnormal maritime traffic detection [4].

2 Proposed Method

Deep neural neural network: The main purpose of the model being developed is to determine the type of vessel by the available or received sets of vessel positions so that the missing information fields can be filled. The model input consists of a sequence of vessel position vectors, and the model predicts the type of vessels sailing under these sets. The model for vessel type prediction uses historical sets of vessel position vectors sorted by time, which can be represented as follows: $X_T = [X_{T-(n-1)}, X_{T-(n-2)}, \ldots, X_{T-1}, X_T]$, where X_T is the set of the vessel's positions, T is the sequence number of a vessel set, which was received at a certain time, n is the predefined length of the set. The input vector X consists of the positioning elements of the vessel, such as the latitude, longitude, heading, speed, time, state reported by the vessel, weather conditions in the geographical location. We can describe the full input vector as a matrix:

$$X_T^p = \begin{bmatrix} X^1 \\ X^2 \\ \vdots \\ X^p \end{bmatrix} = \begin{bmatrix} X^1_{T-(n-1)} & X^1_{T-(n-2)} & \cdots & X^1_{T-1} & X^1_T \\ X^2_{T-(n-1)} & X^2_{T-(n-2)} & & X^2_{T-1} & X^2_T \\ & \vdots & \ddots & \vdots & \\ X^p_{T-(n-1)} & X^P_{T-(n-2)} & \cdots & X^p_{T-1} & X^p_T \end{bmatrix},$$

Where p is the number of elements in the vessel's position vector. The output vector consists of the predicted distribution of probability classes of vessel types calculated by Softmax function. LSTM Deep Neural Network [2] with fully connected multilayer percepron is used to train the model at work. The simplified network architecture is shown in Figure 1. The deep network architecture diagram shows a network structure consisting of 6 constituent layers. The first layer is input layer In with a number of inputs that equals to the length of the vessel's position sequence n. The input layer is connected in series to the first n cells from A1 to an in LSTM (A) layer. The LSTM layer may have more than n cells. The total number of cells is expressed in k when k = n. LSTM (A) layer is connected in series to the LSTM (B) layer. Each output of layer A is connected to Layer B inputs. The total number of cells in LTSM (B) is expressed in k. Both LSTM layers use ReLu activation function. The last cell in B is connected to the multilayer fully connected layer of perceptron. The layer of perceptron consists of two hidden layers of neurons and one output layer. The hiddens layers use ReLu activation function. A number j of outputs constitutes an output layer where each output describes the probability of a particular class classification, which is calculated by Softmax function. Adam's stochastic optimizer with a training factor $\alpha = 0.001$ and a decay factor $\delta = 10^{-6}$ are used for network training. The termination of epoch training cycles is set in accordance with the validation set. The training uses the Sparse Categorical Cross Entropy [2] for loss function. **Data preparation:** Duplicate position entries are cleared out and then data is parsed based on desired data types. The same actions are performed to meteorological data. Technical data fields of a vessel are assigned to each position vector of a vessel based on vessel MMSI identificator. Meteorological data is assigned to a position data vector by using the method of the closest neighbor, depending on the closest time and geolocation of the forecast. Model train-

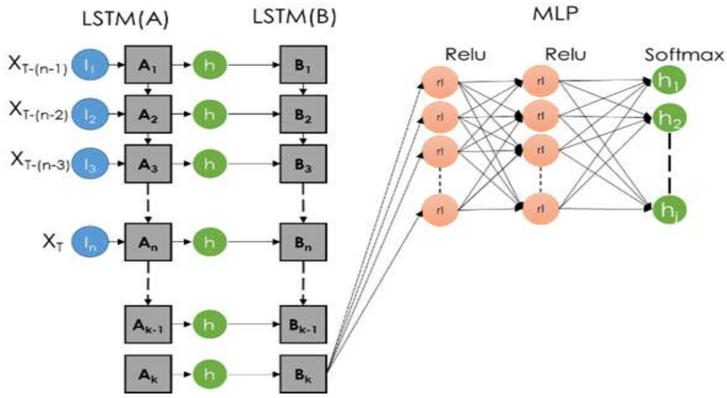

Figure 1: Multi-Stacked LSTM deep neural network architecture

ing data is formed by joining all the data to one vector. **Creation of Vector Sets:** After the preparation of data the vector sets that will be provided to train, validate and test the model, are further formed. The data sets must correspond to the model input matrix, which is described by formula above. To do this, the available vessel data is grouped by their MMSI identifier. All consecutive position of a vessel is cut in sequences of 12 positions by step of 3 positions. All these formed sequences are used to construct training matrix described in this article above.

3 Experiment and Results

To test proposed method Rotterdam harbour area was chosen. The data for model training are collected from several sources such as AIS vessel traffic monitoring system, vessel parameter information system, meteorological observation system, and geographic information system. This information comes from several data sources. The marine traffic data was colected from shipfinder.co: Colected data is: geolocation, speed, direction and type of vessel, length, width, draft, etc. The Meteorological data colected from worldweatheronline.com, provides meteorological data in given geolocations: wind direction and speed, wave, swell and other data of a particular location in 3h intervals. Two separate set was formed to test influence of meteorogical data. One set with meteo data, another without. A total of $2.90 * 10^7$ vessel traffic vectors were collected in one set from the Rotterdam harbour from November 1, 2018 till November 30, 2018, of which $2.78 * 10^7$ do not have information about vessel type. This represents 95.88% of all available data. A set with vessel type information consists of $1.195 * 10^6$ vessel traffic vectors from Rotterdam harbour. These vectors were collected and created using the methods mentioned above, and they constitute 4.12% of all data. The data are randomly divided into three sets: 50% of the data are used for training, 30% for validation, and 20% for testing. Training data set is used to train models. Validation Set - is designed to select the number of LSTM layers in the model and LSTM cells in the layer. The test set is used to evaluate accuracy of the final model. In this

Table 1: Trend of classification accuracy for different network settings

	Meteorological data excluded		Meteorological data included	
Layers	Cells	Accuracy	Cells	Accuracy
2	245	0.78	290	0.78
3	215	0.79	265	0.81
4	195	0.77	250	0.93

article precision, recall, and accuracy are calculated using a test set in order to evaluate the accuracy of the classifier for different numbers of deep multi-stacked LSTM neural network layers and cells. Table 1 first part shows the results of the experiment for different values of the model parameters without meteorological data. We see that the best result was achieved using 3 LSTM layers made of 215 cells. Table 1 second part shows the results of the experiment with meteorological data. Best result was achieved using 4 LSTM layers with 250 cells.

4 Conclusion

According to the results of the experiment, the proposed method of combining vessel traffic data with meteorological data leads to an improved classification. Based on the results the best model configuration is chosen, then checked using continued data with classification accuracy 0.93, recall 0.92, and precision 0.93.

References

[1] Alvarez A.A. et al. (2016). Mining Vessel Tracking Data for Maritime Domain Applications. *Data Mining Workshops (ICDMW), 2016 IEEE 16th International Conference.* pp. 361-367.

[2] Ergen T., Kozat S.S. (2018). Online Training of LSTM Networks in Distributed Systems for Variable Length Data Sequences. *IEEE Transactions on Neural Networks and Learning Systems.* Vol. **29**. pp. 5159-5165.

[3] Sun F., Deng Y., and Deng F. (2015). Unsupervised maritime traffic pattern extraction from spatio-temporal data. 11th International Conference on Natural Computation (ICNC). pp. 1218-1223.

[4] Venskus J. at al. (2017) Integration of a self-organizing map and a virtual pheromone for real-time abnormal movement detection in marine traffic. *Informatica. Vilniaus Universitetas.* Vol. **28**. pp. 359-374.

[5] Zhen R., et al. (2017). Maritime Anomaly Detection within Coastal Waters Based on Vessel Trajectory Clustering and Naïve Bayes Classifier. *The Journal of Navigation.*, pp. 1-23.