

Vessel behaviour classification from AIS without geographical biases

Raphael Sturgis, Valentin Emiya, Basile Couëtoux, Pierre Garreau

▶ **To cite this version:**

Raphael Sturgis, Valentin Emiya, Basile Couëtoux, Pierre Garreau. Vessel behaviour classification from AIS without geographical biases. IEEE International Conference on Intelligent Transportation Systems, Oct 2022, Macau, China. hal-03775942

HAL Id: hal-03775942
https://amu.hal.science/hal-03775942

Submitted on 13 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vessel behaviour classification from AIS without geographical biases

Raphael Sturgis[1] [2], Valentin Emiya[1] and Basile Couetoux[1] and Pierre Garreau[2]

***Abstract*— The automatic detection of vessel behaviours from Automatic Identification System (AIS) data is a challenging aspect of designing intelligent systems and aiding maritime situational awareness. The development of such systems remains limited to some activities like fishing, and by geographical biases that prevent systems to generalise to other areas than that used for training.**

To contribute to these questions, we investigate how to treat raw data or engineered features so that they do not convey such biases at training time and we propose methods for point-wise behaviour detection in the context of container vessels with four target behaviours. Several systems are studied, with raw data or engineered features as inputs, followed by shallow or deep learning classifiers. While good performances are obtained by several of them, we observe that a decision tree classifier with engineered features outperforms an LSTM in areas where no labelled data is available for training.

I. INTRODUCTION

The Automatic Identification System (AIS) is a real-time emission system giving position and other static and dynamic information for every commercial vessel over 500 gross tonnage or transporting passenger. Initially introduced for collision avoidance, AIS data has been increasingly used to aid maritime situational awareness in tasks such as trajectory prediction or autonomous navigation. To perform these tasks efficiently, the ability to differentiate between different behaviours is necessary.

Behaviour classification from AIS data has motivated a number of recent works. Among all possible behaviours of interest, it has been primarily studied in the context of fishing activity detection [1], [2], [3], [4], [5], [6]. Behaviours may be general —e.g., anchor, at port, underway, fishing— or specific —e.g., trawling, longlining [7]. Behaviour classification in contexts other than fishing is also challenging. For instance, behaviour detection for container vessels and commercial transport vessels in general paves the way for improved capabilities of autonomous transport, CO_2 calculations, or logistics with routing capabilities. Therefore some works have been proposed, including discrimination between static, manoeuvring and cruise to help autonomous vessels avoid collision and predicting trajectories of other vessels [8], [9], and anomaly detection for automatic route detection [10]. Existing methods have employed various machine learning approaches to classify behaviours: hidden Markov model and biology inspired models [1], conditional random fields [2], partition-wise gated recurrent units [3], convolutional neural network (CNN) applied to images generated using AIS position data and speed [8], [9], [4], [7], 1D-CNN [5], k-nearest neighbours, dynamic time warping and multilayer perceptron [6].

A challenge observed in the literature is that AIS data is heterogeneous [3], making it difficult for machine learning approaches. Indeed, AIS contains angular data such as heading and course over ground (COG) and non-angular data such as speed over ground (SOG). To get around this issue, multiple methods apply machine vision algorithms to images produced by plotting AIS positions [8], [9], [4], [7]. Removing angular data is another option proposed in [10] where a simple threshold on the speed is used.

Another challenge is the ability of the system to generalise well, and especially to avoid *geographical biases*. We introduce this term to refer to situations where a system performs well only on one geographical area where it has been trained. In such situations, the system may have learned geographical elements specific to this area, like the absolute position or orientation of docks. Many of the aforementioned works are actually dedicated to one geographical region [8], [5], [11] or consider multiple locations but train and test on the same regions without evaluating inter-region performance [4]. Limitations due to geographical biases include not only the inability to transfer prediction capacities to other areas than those used for training, but also the lack of adaptability to changes of geographical elements over time. As a consequence, expensive and time-consuming manual annotation work, as well as repeated learning and validation tasks are required.

Our objective is to propose efficient systems for point-wise trajectory classification without geographical biases. Our systems take a vessel trajectory obtained from AIS data and produce a label for each individual point on this trajectory. These labels describe the behaviour of the vessel at that position, among four options: underway, docked, moored and drifting. To avoid geographical biases, the systems are designed to learn the intrinsic factors that characterise a behaviour rather than geographical elements. This can be achieved by engineering features such that any sensitive attribute that could convey geographical biases is removed. The use of raw AIS data is compared to custom engineered features. Furthermore, decision tree (DT) classifiers and long short-term memory (LSTM [12]) networks are compared when using these possible input settings. The former classifiers are simple, shallow models while the latter are state-of–the-art methods for point-wise sequence classification, e.g., in natural language processing [13], [14], [15], as well as in tasks using AIS data such as trajectory prediction [16] or Vessel classification and anomaly detection [17].

The research on this article was supported by Region SUD

[1]Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France `firstname.lastname@lis-lab.fr`

[2]Searoutes SAS, Marseille, France, `contact@searoutes.com`

From our experiments we show that even if an LSTM outperforms a DT in the same area in which it is trained, a DT trainable in a few seconds has similar performance to an LSTM model in some geographical areas, or even outperforms LSTMs in regions with different data distribution.

This paper is organised as follows. In section II, our method is introduced. Experiments to validate our approach are presented in III. Lastly, conclusions on the performances of our system and further discussions are proposed in section IV.

II. Designing systems for ship behaviour detection

The proposed systems result from the application of machine learning methods to raw or engineered features. A large effort is dedicated to study engineered features, starting from raw AIS contents (Section II-A) and using elementary operations (Section II-B) in order to design appropriate features (Section II-C). The goal is two-fold: (i) contributing to the classical question about the performance when using engineered features vs. learned representations Y from raw data; and (ii) proposing features with invariant properties with respect to translations and rotations of the trajectories, in order to avoid geographical biases and to obtain systems that can generalise to unseen areas.

A. Raw AIS feature description

Automatic identification system (AIS) is a system that transmits vessel information at regular time intervals. This information is shared over high frequency radio waves with all other vessels and ground operations. The primary usage of this system is to avoid collisions between vessels. AIS contains both static and dynamic information about the vessel. For our purposes we are only focused on dynamic data, described below in Table I (see [18] for further details).

Field	unit
Timestamp	UTC
MMSI	ID
Position	degrees for latitude and longitude
Speed over ground (SOG)	knots
Course over ground (COG)	degrees from north
Heading	degrees from north
Rate of turn (ROT)	degrees per second

TABLE I: AIS data content

Hence a typical AIS trajectory is a time series composed with multivariate heterogeneous data. It includes angular data, which should be handled carefully when used as inputs for machine learning models.

B. Elementary operations on features

We introduce elementary operations that will be subsequently applied to design engineered features from raw AIS data. The elementary operations are the rate of change and the first two statistical moments defined locally over a time window.

We consider each feature as a univariate time series with sampling period τ.

For a regular real-valued time series $f(t)$ such as the SOG, the rate of change between $f(t)$ and $f(t+\tau)$ is simply defined as:

$$\Delta_f(t) = \frac{f(t) - f(t-\tau)}{\tau} \tag{1}$$

In order to track local statistical moments in time series, we use a sliding window with radius r centred in $f(t)$. The mean $\bar{f}(t)$ and variance $\sigma_f^2(t)$ are defined as:

$$\bar{f}(t) = \frac{1}{2r+1} \sum_{n=-r}^{r} f(t+n\tau)$$

$$\sigma_f^2(t) = \frac{1}{2r+1} \sum_{n=-r}^{r} \left(f(t+n\tau) - \bar{f}(t)\right)^2$$

Angular features are more delicate to handle: the rate of change and statistical moments introduced previously cannot be computed straightforwardly from angular data like COG, heading, ROT, latitude and longitude. We adopt variants of these operations that are suitable for angular data [19]. In order to avoid discontinuity issues around 2π, the rate of change for angular data is computed by applying eq. (1) to the unwrapped version of any angular feature $\theta(t)$.

For statistical moments, unwrapping is not appropriate. For any angular feature $\theta(t)$, we adapt the definition of the mean direction and circular variance of angular data [19] to the case of a sliding window with radius r. Based on the distribution of $\theta(t)$ on the unit circle $z_\theta(t) = e^{i\theta(t)}$, one can compute the empirical mean over a sliding window $\bar{z}_\theta(t) = \frac{1}{2r+1} \sum_{n=-r}^{r} z_\theta(t+n\tau)$. The mean direction and the circular variance are defined as

$$\bar{\theta}(t) = \text{angle}\left(\bar{z}_\theta(t)\right) \tag{2}$$

$$\sigma_\theta^2(t) = 1 - \left|\bar{z}_\theta(t)\right| \tag{3}$$

where $\text{angle}(z)$ denotes the argument of a complex number $z \in \mathbb{C}$. Note that while the samples $z_\theta(t)$ lie on the unit circle, their mean $\bar{z}_\theta(t)$ is located inside the unit circle: its modulus is close to 1 (resp. 0) when the samples are concentrated around a given direction (resp. spread over the circle).

Last, in order to characterise the spread of a ship trajectory in space, we also define the position variance over a sliding window as

$$\sigma_{pos}^2(t) = \frac{1}{2r+1} \sum_{n=-r}^{r} h^2\left(p(t+n\tau), \bar{p}(t)\right)$$

where the Haversine function h computes the great circle distance between two positions, $p(t)$ is the position at a time t given as a pair (latitude, longitude) and $\bar{p}(t)$ is the mean direction for the latitude and the longitude at time t computed using eq. (2).

C. Proposed engineered features

Using the elementary operations defined in Section II-B, we design features that may facilitate the discrimination between ship behaviours. We also pay attention to produce

features that are invariant to translations and rotations of the trajectories in order to avoid geographical biases. This is achieved by first removing from the feature selection any sensitive attribute than may convey geographical biases, namely position, COG and heading, which directly encode the absolute positions or orientations of specific instances of geographical elements like docks; we then derive engineered features, possibly from the previous sensitive features, by applying the above operations, ensuring that the obtained quantities are not prone to inducing geographical biases.

The retained features are:

- sog is an obvious choice for behaviour detection as it correlates heavily with some behaviours; for instance, docked ships must have speed of 0 and vessels underway cannot.
- $drift$ is defined as the absolute difference $|heading\,(t) - COG\,(t)\,|$ between the heading and COG [20] and is illustrated in Figure 1. It may give information on the type of manoeuvre a ship is performing.
- Δ_{sog}, as a change in speed, reports acceleration of a vessel.
- Δ_{cog} and $\Delta_{heading}$ give some insight into the rate of turn of a vessel for a manoeuvre.
- σ_{cog} and $\sigma_{heading}$ may indicate whether, e.g., a ship is moving erratically or not.
- $\sigma_{\Delta_{cog}}$ and $\sigma_{\Delta_{heading}}$ might give insights into behaviours where the rate of turn of the vessel changes.
- σ_{pos}, as the spread of the vessel position, may contribute to characterise some behaviours.
- σ_{sog} indicates if the vessel speed is changing a lot.

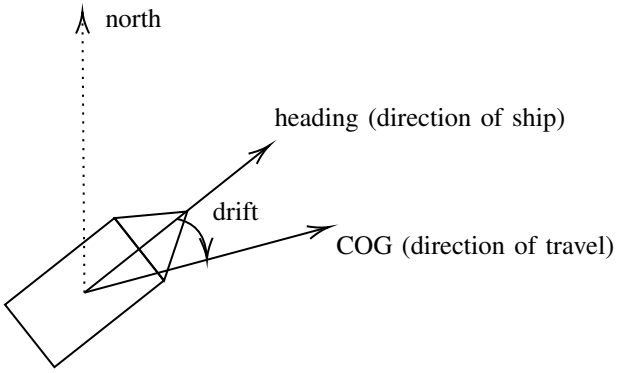


Fig. 1: Graphical representation of $drift$

III. Experiments

A. Data

dataset	underway	docked	moored	drifting
Western Mediterranean	51.6%	36.0%	7.1%	5.3%
Baltic	52.3%	42.4%	5.0%	0.4%
Suez	77.9%	15.8%	6.2%	0%

TABLE II: Label proportions

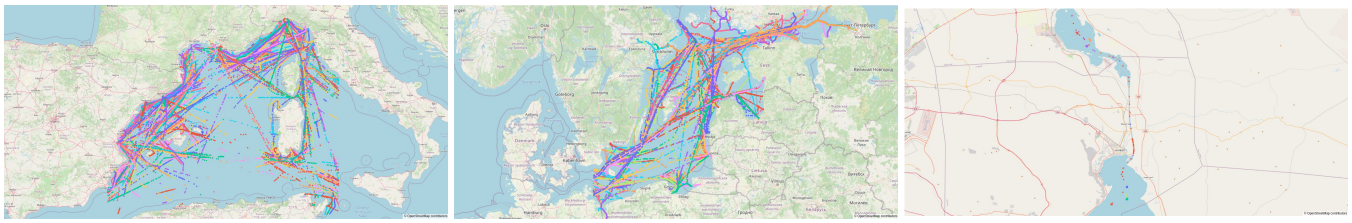
Three datasets have been developed for this study, including one for training, validation and testing purposes, and two other ones for testing only. All three datasets use AIS data manually labelled with four target behaviours: underway, docked, moored and drifting. These behaviours can be determined visually by a human: an example of the behaviours docked, moored and drifting can be seen in Figure 3. The western Mediterranean dataset, used for training, validation and testing, is built using AIS data from 334 individual container ships in the western Mediterranean. The data is shown visually in Figure 2a. This dataset contains $194,443$ AIS records from 1 March 2020 to 6 March 2020. The trajectories from the dataset are split into 3 subsets. Firstly a training dataset containing 60% of trajectories. A validation dataset we used for evaluating different configurations during a parameter selection phase containing 20% of trajectories. The last subset is a test set, containing the remaining 20% of trajectories, used exclusively for getting final performance metrics. The second dataset is data from 100 container ships in the Baltic Sea as seen in Figure 2b. This data contains records from 10 October 2021 to 31 October 2021 and has $172,431$ AIS records. The last dataset is trajectories from 12 container ships near the Suez Canal from 23 March 2021 to 30 March 2021 during the 2021 Suez Canal obstruction, as seen in Figure 2c. It contains 2178 records. This last dataset has much fewer data points than the other ones because during this crisis, ships would turn off AIS. The proportions of the labels are presented inTable II. We can see slight differences in proportion of behaviours between the Western Mediterranean and Baltic datasets, but starker differences with the Suez dataset.

All the datasets are composed of raw AIS data points that need some prepossessing to be useful. The first treatment applied to the data is to remove points with missing features and aberrant speeds. Then, trajectories are built from the data by isolating each ship by its identifier (vessel MMSI) and separating this trace into multiple trajectories. One vessel trace might not be usable as one trajectory as it could be split and have holes of multiple hours or day with no data. Traces are cut into multiple trajectories if the time between 2 position reports is less than a threshold time γ, set to $\gamma = 1h$.

Once trajectories have been calculated, we have multiple trajectories per vessel but with non-uniform sampling, which is not appropriate for further learning on time series. Trajectories are thus interpolated in order to have a uniform sampling period equal to $\tau = 5min$. We use cubic interpolation on all raw features. After interpolation of the raw data, the selected features are computed. The characteristics of the three datasets after the prepossessing steps are summarized inTable III.

Dataset	Number of points	Number of trajectories
Western Mediterranean	193712	2017
Baltic	171982	1582
Suez	2173	248

TABLE III: Label proportions



(a) Western Mediterranean dataset (b) Baltic dataset (c) Suez dataset

Fig. 2: Illustration of the three datasets



(a) Docked (green) (b) Moored (magenta) (c) Drifting (yellow)

Fig. 3: Example of behaviours patterns from the western Mediterranean dataset (red dots are underway)

B. System description

1) Decision Tree classifiers: DTs are simple models that work by classifying individual points without considering the time series aspect of the data. Nonetheless they are much faster to learn than more complex methods such as LSTMs. DTs trees need to be optimised for the specific data we use. To do so a grid search optimisation was performed on 3 criteria: the splitting criterion (Gini or entropy), splitter (best or random) and the max depth of the decision tree (1, 2, 5, 10). These parameters have been tuned by finding the model with the best performance on the validation set.

2) LSTM networks: LSTMs are more complex than DTs, but also more expressive. They also natively exploit the sequences aspect of the data.

a) Normalisation: Empirical experiments not reported here have shown that normalisation of the data is necessary for our LSTM model, providing from 6% to 40% of accuracy improvements. Nonetheless normalisation is not straightforward since the data is heterogeneous and thus depends on the feature. Regular scalar values Δ_{sog}, σ_{cog}, $\sigma_{heading}$, $\sigma_{\Delta_{cog}}$, $\sigma_{\Delta_{heading}}$, σ_{pos}, σ_{sog} are normalised by the standardisation method. SOG values are scaled by the inverse of their third quartile since they are non-negative. Angles are normalized using their maximum possible value: $Drift$, Δ_{cog}, $\Delta_{heading}$ and $longitude$ have been divided by $180\,\text{deg}$ and $latitude$ by $90\,\text{deg}$.

b) Training process: The LSTM networks proposed here are composed of n bi-LSTM layers, with m units for each layer. In order to find the best LSTM architecture for our problem a grid search was performed for values of n between 1 and 6 and for $m \in \{5, 10, 20, 50, 100\}$. For each combination 5 runs were performed. The average f1 score on the 5 runs for each architecture is calculated and the architecture with the best f1 score on the validation set is selected. LSTMs were trained on a maximum of 100 epochs, the validation set was used for early stopping with a patience of 10 epochs. The batch size was set to 1. The Adam optimizer was chosen and the categorical cross entropy loss is used.

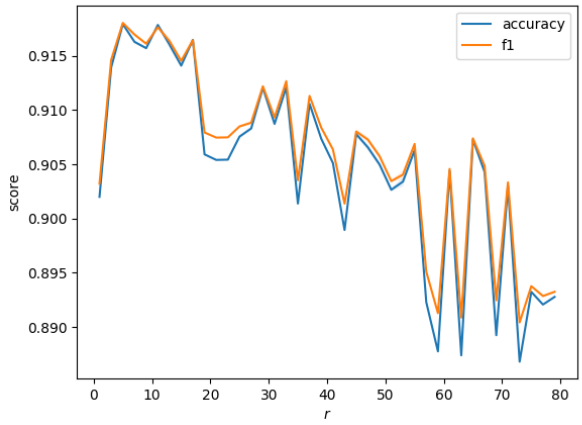
3) Time window radius selection: Some operations presented in II-C have a time window radius parameter r. The wider the time window the wider the time period our models have access in order to make a decision, the smoother the changes will be for our statistical features, but we might lose in specificity.

To study the impact of the selection of r and find the best parameters, a DT classifier with r between 2 and 80 and an LSTM for values between 1 and 15 were trained, for both using the proposed engineered features as inputs. In order to conserve as much exploitable data as possible, the time window will ignore the out of bound values and just compute means and standard deviations on smaller time windows. In extreme cases, the time window will completely consume the whole trajectory, resulting in a single value for the trajectories.

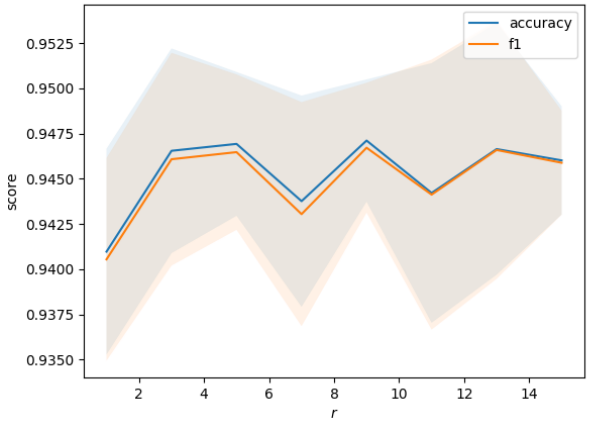
For each time window 5 LSTMs are trained for each architecture. For all these LSTMs the accuracy on the validation set is recorded. The architecture with the best average accuracy is then selected and the 5 models with this architecture are tested on the test set, the accuracy and f1 score where recorded and showed on Figure 4b.

Since the training process for DT is deterministic we only train one model with each combination of parameters. The model with the best performance on validation set is evaluated on the test set. Both f1 score and accuracy are presented in Figure 4a. In the case of the DT, performances on both metrics increase for small sizes, but then decrease for values greater than 11. This shows that the size of the window has an impact on performance, and also that the optimal value seems to be $r = 11$ for DT classifiers. On the other hand, for the LSTM network, the maximum is observed at a value of $r = 9$. Nonetheless no trend can be seen from

the data and all variations are within the standard deviation.



(a) DT



(b) LSTM (shaded area: standard deviation)

Fig. 4: Accuracy and f1 score on the test set as a function of the radius of the time window used in engineered features.

C. System performances

Our final experiment's objective is to compare the classification performance of the proposed DT and LSTM systems on the three datasets. Two main questions underlie this experiment: (i) which benefits can we expect from a DT on engineered features vs. an LSTM on raw data? And (ii) can we eventually achieve learning with no geographical biases so that classification can be successful in other areas?

To achieve these objective we consider 4 different configurations for LSTM networks. The first input configuration, we consider only the *sog*, this configuration as no geographical biases but is limited in information. In the second configuration, we considered the *sog* and the raw position of a vessel, which can induce geographical biases. For the third configuration, we consider the same features, but we apply random translations of longitude and latitude values. This treatment is applied to trajectories in the train set only, the model is used to predict behaviour of trajectories that haven't been treated. Translation is performed in order to limit learning of geographical biases. For the last configuration, we consider the 11 engineered features proposed in section II-C. Each of these configurations used normalized data.

The best model architectures for each configurations are summarized in Table IV.

configuration	n	m	r
sog	3	100	n/a
sog&position	4	100	n/a
sog&shifted position	4	100	n/a
engineered	2	10	9

TABLE IV: Best LSTM architecture per configuration

The best model is selected and then tested on the test set and on the other two datasets. We test on the Baltic and Suez dataset in order to measure the performances in areas not seen during training. We applied the same methodology for DTs, but did not consider the third configuration. Indeed, since the DT model only classifies single vectors, it cannot exploit the shifted trajectory; it would be completely equivalent to only using sog, which has been verified experimentally. The configuration with a DT trained only with SOG will be used a baseline.

The results are summarized in Table V.

model, configuration	Mediterranean	Baltic	Suez
DT sog	0.849	0.936	0.379
DT sog&pos	0.888	0.440	0.644
DT engineered	0.913	0.966	**0.942**
LSTM sog	0.923	0.960	0.570
LSTM sog&pos	0.934	0.959	0.664
LSTM sog&shifted pos	0.921	0.954	0.507
LSTM engineered	**0.947**	**0.970**	0.789

TABLE V: F1-score on the three test sets for all the models and input configurations (bold shape denotes best performance for each dataset).

Results confirm that a shallow classifier like the DT model performs best with engineered features, while a deep classifier like the LSTM networks can reach high performance on raw or engineered data. This also validates the relevance of the proposed engineered features. Indeed, LSTMs or DT trained with these features shows better performances on all datasets than models trained with other features.

Performances on the Mediterranean test set is the better with LSTMs than DTs. Nonetheless, on the Baltic dataset, DTs with engineered features perform on part with the LSTM models with a good capacity to generalize to areas with similar behaviours distributions. On the Suez dataset we can see that the DT with engineered features outperforms all other models. This suggests that this model is best for generalizing to any area.

Finally, we detail the performance of the two models using engineered features by reporting the accuracy per class (one versus all) in Table VI. Both models perform similarly to each other for vessels underway and docked, nonetheless, on the other 2 behaviours the DT is better by more than 10%.

For areas with potentially more of these two behaviors, such as congested ports, DTs seem to be a better choice.

Model, configuration	Underway	Docked	Moored	Drifting
DT engineered	0.955	0.968	0.931	0.760
LSTM engineered	0.980	0.977	0.882	0.653

TABLE VI: Performances per label on Baltic dataset

IV. CONCLUSIONS

In this paper, several systems for behaviour classification from AIS data have been designed and studied, resulting in good classification performance for container vessels on a four-behaviour detection task. Experiments suggest that a deep LSTM network applied to raw AIS data generally perform on part with other proposed solutions when testing on data from the same area. Nonetheless, this model may learn some geographical biases and would require data augmentation at training time in a more diverse way than with the proposed random position shifts. When an area has no available labels for training, the best solution appears to be a decision trained on the custom engineered features. This model performs similarly to LSTM in areas similar to the training set and significantly better in areas with different data distributions than the training set. This is promising to improve maritime situational awareness on a global scale using limited training data.

The engineered features presented here improve performance in the case of decision tree for multiple reasons. Firstly, they remove biases from the local geography of the training set. They also give access to a higher level of abstraction than the raw features. Also, the use of a time window of appropriate length allows for a single data point to have access to some characteristics from the trajectory. This last point also gives us clues as to why the LSTMs do not perform well with these features, as LSTMs are already able to perform high level abstraction from the data and also consider natively the sequence aspect of the data.

We have shown that our LSTM do not perform well on moored behaviours. This can have many causes, but one certainly is the relative rarity of this behaviour in our training dataset. The loss function we chose for this could be changed for a function that takes into account better the relative distribution of labels. Other classification methods could be explored, in the same vein as LSTMs, transformer networks could be explored, potentially with other loss functions.

REFERENCES

[1] E. N. de Souza, K. Boerder, S. Matwin, and B. Worm, "Improving fishing pattern detection from satellite AIS using data mining and machine learning," *PLOS ONE*, vol. 11, p. e0158248, jul 2016.

[2] B. Hu, X. Jiang, E. Souza, R. Pelot, and S. Matwin, "Identifying fishing activities from AIS data with conditional random fields," in *Annals of Computer Science and Information Systems*, IEEE, oct 2016.

[3] X. Jiang, X. Liu, E. N. de Souza, B. Hu, D. L. Silver, and S. Matwin, "Improving point-based AIS trajectory classification with partition-wise gated recurrent units," in *2017 International Joint Conference on Neural Networks (IJCNN)*, IEEE, may 2017.

[4] I. Kontopoulos, K. Chatzikokolakis, K. Tserpes, and D. Zissis, "Classification of vessel activity in streaming data," in *Proceedings of the 14th ACM International Conference on Distributed and Event-based Systems*, ACM, jul 2020.

[5] K. il Kim and K. M. Lee, "Convolutional neural network-based gear type identification from automatic identification system trajectory data," *Applied Sciences*, vol. 10, no. 11, 2020.

[6] B. Chuaysi and S. Kiattisin, "Fishing vessels behavior identification for combating IUU fishing: Enable traceability at sea," *Wireless Personal Communications*, vol. 115, pp. 2971–2993, feb 2020.

[7] I. Kontopoulos, A. Makris, and K. Tserpes, "A deep learning streaming methodology for trajectory classification," *ISPRS International Journal of Geo-Information*, vol. 10, p. 250, apr 2021.

[8] X. Chen, A. Kamalasudhan, and X. Zhang, "An application of convolutional neural network to derive vessel movement patterns," in *2019 5th International Conference on Transportation Information and Safety (ICTIS)*, pp. 939–944, 2019.

[9] X. Chen, Y. Liu, K. Achuthan, and X. Zhang, "A ship movement classification based on automatic identification system (ais) data using convolutional neural network," *Ocean Engineering*, vol. 218, p. 108182, 2020.

[10] G. Pallotta, M. Vespe, and K. Bryan, "Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction," *Entropy*, vol. 15, pp. 2218–2245, jun 2013.

[11] F. Mazzarella, M. Vespe, D. Damalas, and G. Osio, "Discovering vessel activities at sea using ais data: Mapping of fishing footprints," in *17th International conference on information fusion (FUSION)*, pp. 1–7, IEEE, 2014.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735–1780, nov 1997.

[13] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks*, vol. 18, pp. 602–610, jul 2005.

[14] Z. Huang, W. Xu, and K. Yu, "Bidirectional lstm-crf models for sequence tagging," *arXiv preprint arXiv:1508.01991*, Aug. 2015.

[15] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based lstm for aspect-level sentiment classification," in *Proceedings of the 2016 conference on empirical methods in natural language processing*, pp. 606–615, 2016.

[16] M. Gao, G. Shi, and S. Li, "Online prediction of ship behavior with automatic identification system sensor data using bidirectional long short-term memory recurrent neural network," *Sensors*, vol. 18, no. 12, 2018.

[17] D. Nguyen, R. Vadaine, G. Hajduch, R. Garello, and R. Fablet, "A multi-task deep learning architecture for maritime surveillance using AIS data streams," in *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, IEEE, oct 2018.

[18] M. Series, "Technical characteristics for an automatic identification system using time-division multiple access in the vhf maritime mobile band," *Recommendation ITU: Geneva, Switzerland*, pp. 1371–1375, 2014.

[19] A. D. Rao and S. Girija, *Angular statistics*. Place of publication not identified: CRC Press, 2019.

[20] M. Gao and G.-Y. Shi, "Ship-handling behavior pattern recognition using AIS sub-trajectory clustering analysis based on the T-SNE and spectral clustering algorithms," *Ocean Engineering*, vol. 205, p. 106919, jun 2020.

Code is available on our gitlab