Ship length estimation using common radar field entries

Eftychios Protopapadakis*[a], Matthaios Bimpas[a], Dimitrios Dres[a]
[a] *Telesto Technologies, Cholargos 15561, Greece*

ABSTRACT

In this paper, we focus on identifying the length of a ship given the track history. The problem entails to a traditional classification problem, on which multiple Bayesian optimized classifiers are tested. The input values, i.e. speed, course, and position, are provided by a data fusion algorithm from multiple sources, including over-the-horizon radar and AIS. Results indicate that using track history, spanning a five-minute period or less, can provide adequate information to estimate correctly the length of a ship.

Keywords: Classification, ship length, hyperparameter optimization

1. INTRODUCTION

Maritime Situational Awareness (MSA) is a challenging research field. Marine routes represent a huge portion of commercial and human trades; therefore, surveillance, security and environmental protection themes are gaining increasing importance. There are various criteria that can provide significant information regarding the behavior of a ship: these criteria can be related to position, speed, course, existence of AIS or other type of attributes.

Existing literature employs various computer vision algorithms for the ship detection. These approaches are RGB based [1], multispectral based [2], or SAR based [3]. Many of these approaches utilize more than one sources of data as inputs. There are, also, radar based approaches: Over-the-horizon high frequency doppler radars [4], compact high frequency radar [5] or other types, e.g. MIMO radars [6]. The former case, i.e. vison based, allows for more details regarding the ship's characteristics. The latter case, i.e. radar based, provides less detailed information on a larger scale.

This case study investigates the refinement possibilities of common radar field entries, for the extraction of specific ship's traits. Thus, in this paper we emphasize on ship's length estimation. Such information can be beneficial in many ways, and can be easily incorporated to advanced systems for ship behavior analysis [7].

2. PROPOSED METHODOLOGY

In this paper, we present a simple and efficient method to estimate the category of a ship's size, given the track history and common fields provided by any type of radar. The proposed methodology exploits fussed track data; i.e. data entries originating from multiple sensors [8]. These data entries are stored locally. Once a specific number of track entries is gathered, a low-level feature extraction occurs and then the feature values are fed to a classifier.

Let as denote as $\boldsymbol{x}_i^{(t)} = \{ln_i\, , lt_i, c_i, s_i\}^{(t)}$, the common filed values, i.e. longitude, latitude, course and speed, for ship i at a time t. Then, the input to the classifiers has the form of $\boldsymbol{x}_i = \left[\boldsymbol{x}_i^{(t)},\ \boldsymbol{x}_i^{(t-1)}, \dots, \boldsymbol{x}_i^{(t-l)}\right]$, $l = 1, \dots, \ell$, where ℓ denotes the number of past moments to be considered. All values are normalized to $[0,1]$. Let us also denote as $\boldsymbol{y}_i \in \mathbb{Z}^{k\times 1}$, a vector indicating the corresponding class $\boldsymbol{C_k}$, i.e. the range category for the length. We would like to learn a function $f: \boldsymbol{\mathcal{X}} \rightarrow \boldsymbol{\mathcal{Y}}$ so that f is expected to be a good predictor on future data.

2.1 Employed classifiers

We have scrutinized the effectiveness of a series of well-known classifiers in ship's length recognition, using track history data entries. The techniques considered for constructing linear and non-linear models were:

Naïve Bayes classifier, a family of simple probabilistic classifiers based on applying Bayes' theorem with strong independence assumptions between the features [9]. Abstractly, naive Bayes is a conditional probability model: given a problem instance to be classified, represented i.e. vector $\boldsymbol{x}$ representing some $\boldsymbol{\ell} \times 4$ features (independent variables), it assigns to this instance probabilities $\boldsymbol{p}(\boldsymbol{C_k}|\boldsymbol{x_1}, \dots, \boldsymbol{x_n})$ for each of $\boldsymbol{K}$ possible outcomes or classes $\boldsymbol{C_k}$.

Linear discriminant analysis, a statistical analysis method useful in determining whether a set of variables is effective in predicting category membership. Discriminant analysis (Discr) classifiers assume that different classes generate data based on different Gaussian distributions so that $p(\boldsymbol{x}|y = \boldsymbol{C}_k) \sim N(\boldsymbol{\mu}_k, \boldsymbol{S}), k = 1, \dots, \boldsymbol{K}$. In order to train such a classifier, we need to estimate the parameters of a Gaussian distribution for each class. Then, to predict the classes of new data, the trained classifier finds the class with the smallest misclassification cost: $\hat{y} = arg \max_k \left\{ \left(\boldsymbol{x} - \frac{\mu_k}{2} \right)^T \beta_k + \log \pi_k \right\}$, where $\beta_k = \boldsymbol{S}^{-1} \boldsymbol{\mu}_k$.

k-nearest neighbors, a non-parametric method used for classification [10]. A majority vote of its neighbors classifies an object, with the object being assigned to the class most common among its k nearest neighbors; it is, therefore, a type of instance-based learning, where the function is only approximated locally, and all computation is deferred until classification.

Classification trees use a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value. In classification tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels [11]. Each internal (non-leaf) node is labeled with an input feature. The arcs coming from a node labeled with a feature are labeled with each of the possible values of the feature. Each leaf of the tree is labeled with a class or a probability distribution over the classes.

Support vector machines, a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap (margin) that is as wide as possible [12]. New examples are then mapped into that same space and predicted to belong to a category based on which side of the margin they fall on. The mappings used by SVM schemes are defined through a kernel function $K(\boldsymbol{x}, \boldsymbol{y})$ selected to suit the problem.

The above methods all lead to standalone models. In addition, ensembles are also used, which are based on combinations of multiple models resulting from a single learning algorithm. For the purposes of this analysis, we consider three ensemble algorithms, namely boosting and random forests using classification trees and discriminant analysis as the base learning algorithms.

2.2 Parameterization of learning algorithms

Typically, machine learning models have a set of parameters that need to be defined before the learning process (i.e. training using available data) begins. All these parameters are known as hyperparameters. Some indicative examples are the reqularization parameter c in SVMs, the number of k nearest neighbors and the number of children (max depth) in decision trees. In this work, the hyperparameter optimization process, for each of the classifiers, was based on grid-search optimization algorithm, which was limited to three hours [13]. Grid search approach is an exhaustive search through a specified subset of hyperparameters. Given the relatively small amount of training data (see sec. 3.1) and the complexity of the models, grid search was a viable solution for parameters fine tuning.

3. EXPERIMENTAL EVALUATION

The length of a ship is estimated using five categories: i) <10 meters, ii) 10-15 meters, iii) 15-20 meters, iv) 20-25 meters and v) >25 meters. The number of feature values, used as inputs to the classifiers, varied, depending on how many past track records we use. The following scenarios applied in our case:

1) 2 past moments: 1.2 minutes duration, 8 values per ship
2) 4 past moments: 2.6 minutes duration, 16 values per ship
3) 6 past moments: 4 minutes duration, 24 values per ship
4) 8 past moments: 5.3 minutes duration, 32 values per ship

The data transmission times had an approximately 40 seconds interval rate.

3.1 Dataset description

The utilized dataset pertains to approximately 7 hours of data captured from the Mediterranean coast of southern France. Only ships that have AIS were considered, since we need the actual length to use it as ground truth for performance evaluation. Despite the large number of ships in the vicinity, only a small subset provided the actual length values. Also, the number of ships with length less that 20 meters was limited, resulting in unbalanced dataset. The final training sets (after the data balancing approach) are shown in Table 1.

Table 1. Train set length distribution

Class / Past Moments	2	4	6	8
<10 meters	14	12	12	7
10-15 meters	42	25	19	14
15-20 meters	43	40	33	23
20-25 meters	49	41	35	22
>25	45	40	40	25

3.2 Performance scores

The results are analyzed through standard measures of predictive performance for binary classification tasks. First, the confusion matrices (i.e. 5 × 5 matrices) are formed for each of the suggested combinations of the classifier algorithm with the past moments to use. If we apply the one against all paradigm, we can divide the original 5 × 5 to five 2 × 2 matrices. The elements of the confusion (2 × 2) matrices in that case involve: (1) the number of ships, with length different that the investigated one, correctly classified as ships with different length (true negatives, TN), (2) the number of ships in the investigated length category misclassified as different length category (false negatives, FN), (3) the number of ships outside investigated length ranges, misclassified as within investigated range (false positives, FP), and (4) the number of ships within the investigated range correctly classified (true positives, TP). Using these elements, four well-known classification performance indices are calculated:

1. Accuracy (ACC) represents the percentage of correct classification for all classes. It is defined as:

$$ACC = (TP + TN) / (P + N)$$

2. Recall (Re) indicates the fraction of the specific length cases identified by a model. It is defined as:

$$Re = TP / P$$

3. Precision (Pr) indicates the correct ship's length predictions. It is defined as:

$$Pr = TP / (TP + FP),.$$

4. F1-score (F1) is the harmonic mean of precision and recall. It is defined as:

$$F_1 = 2 \cdot \frac{\mathrm{Pr} \cdot Re}{\mathrm{Pr} + Re}$$

Figure 1 illustrates the confusion matrices generated by different type of classifiers. An ideal scenario would be a diagonal matrix. In that case, all ships would be classified correctly. Depicted cases achieve overall accuracy score of, approximately, 90%. Yet, a closer look indicates that the capability of identifying small length ships is less than 50%.

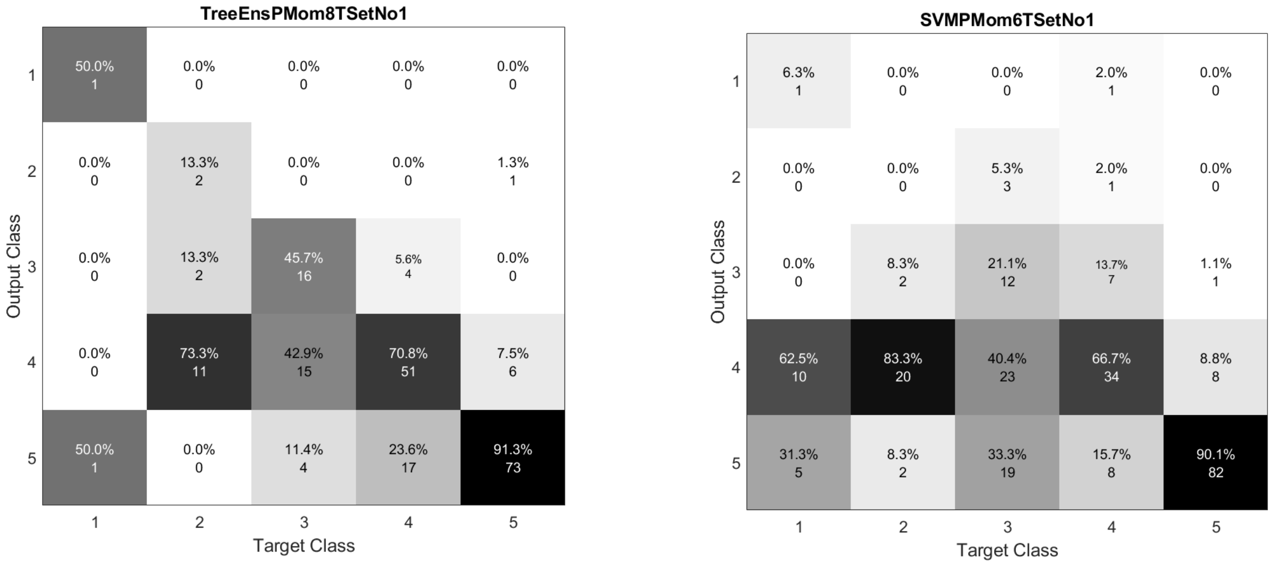

Figure 1: (Left) Tree ensemble classifier confusion matrix, over test set number 2, using as inputs 8 past moments. (Right) SVM classifier confusion matrix, over test set number 1, using as inputs 6 past moments.

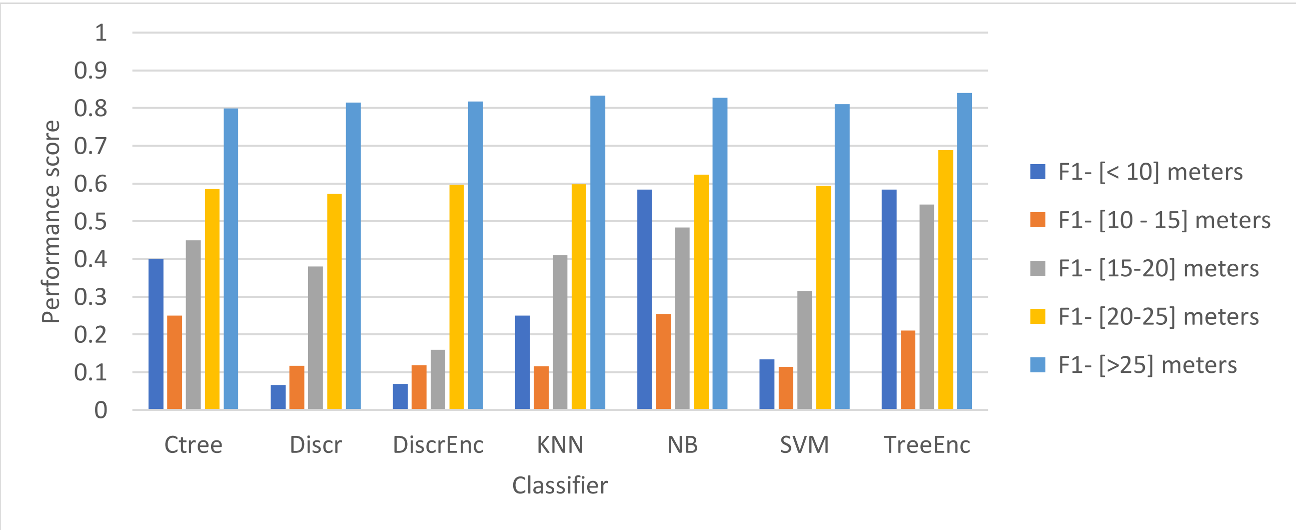

Figure 2. Average F1 scores, for each of the investigated categories.

Figure 2 provides details on the classifier’s suitability for the length estimation. All classifiers achieve a good performance for the over 25 meters class. As the length of the ship declines, so does the performance score. Less than 10-meter ships are a peculiar case, for which the ensemble of classification trees appear to be the best solution.

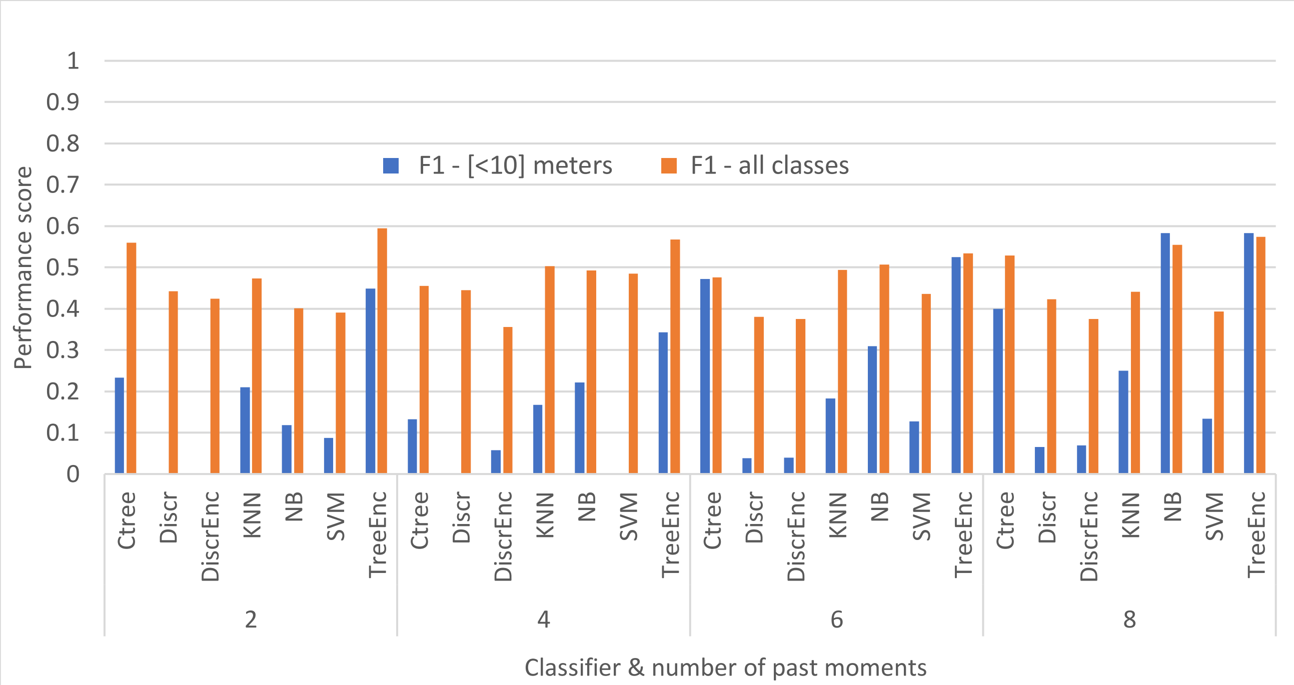

Figure 3. Analytical assessment for the detection capabilities when we have less than 10 meters size of ships.

Figure 3 provides a further insight on the less than 10 meters case. There are two points worth mentioning. At first, increasing the number of past moments (i.e. more input values) can lead to an increase in detection performance. Secondly, there is a trade-off between the detection in less than 10 meter class and the rest. An overall suggestion would be the utilization of a tree ensemble classifier for the length estimation problem.

4. CONCLUSIONS

The applicability of traditional classification schemes for the identification of a ship's length has been investigated. Results indicate that the track history of a vehicle, in terms of location, speed and course can provide adequate information for the estimation of its length. Despite the relatively small length sizes, i.e. four out of five categories were below 25 meters, an ensemble of tree classifiers using five-minute track history, provide an average of 60% in F1 score. The impact of additional values and how the time intervals among radar transmission affects the performance, will be further investigated.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Commission's H2020 research and innovation programme, under grant agreement no 700478 (RANGER project). The authors would like to thank all project partners for their collaboration

REFERENCES

[1] K. Makantasis, E. Protopapadakis, A. Doulamis, and N. Matsatsinis, "Semi-supervised vision-based maritime surveillance system using fused visual attention maps," *Multimed. Tools Appl.*, vol. 75, no. 22, pp. 15051–15078, 2016.

[2] M. Gianinetto *et al.*, "OBIA ship detection with multispectral and SAR images: A simulation for Copernicus security applications," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, pp. 1229–1232.

[3] F. Mazzarella, M. Vespe, and C. Santamaria, “SAR Ship Detection and Self-Reporting Data Fusion Based on Traffic Knowledge,” *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 8, pp. 1685–1689, Aug. 2015.
[4] F. R. G. Cruz *et al.*, “An Interactive and Plotted Data Visualization of High Frequency Doppler Radar Ship Detection and Tracking Software for Maritime Surveillance,” in *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM)*, 2018, pp. 1–6.
[5] S. Park, C. J. Cho, B. Ku, S. Lee, and H. Ko, “Compact HF Surface Wave Radar Data Generating Simulator for Ship Detection and Tracking,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 6, pp. 969–973, Jun. 2017.
[6] J. N. Bathurst, M. Hefnawi, J. R. Bray, and Y. M. M. Antar, “Design considerations for a shipboard MIMO radar for surface target detection,” in *2018 IEEE Radar Conference (RadarConf18)*, 2018, pp. 0316–0321.
[7] E. Protopapadakis, A. Voulodimos, A. Doulamis, N. Doulamis, D. Dres, and M. Bimpas, “Stacked autoencoders for outlier detection in over-the-horizon radar signals,” *Comput. Intell. Neurosci.*, vol. 2017, 2017.
[8] G. Soldi *et al.*, “Heterogeneous Information Fusion for Multitarget Tracking Using the Sum-product Algorithm,” in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 5471–5475.
[9] A. C. Antonakis and M. E. Sfakianakis, “Assessing naïve Bayes as a method for screening credit applicants,” *J. Appl. Stat.*, vol. 36, no. 5, pp. 537–545, May 2009.
[10] N. Bhatia and Vandana, “Survey of Nearest Neighbor Techniques,” *ArXiv10070085 Cs*, Jul. 2010.
[11] W. Stuetzle, “Estimating the Cluster Tree of a Density by Analyzing the Minimal Spanning Tree of a Sample,” *J. Classif.*, vol. 20, no. 1, pp. 025–047, May 2003.
[12] S. Abe, *Support Vector Machines for Pattern Classification*. Springer, 2010.
[13] E. Protopapadakis, D. Niklis, M. Doumpos, A. Doulamis, and C. Zopounidis, “Sample selection algorithms for credit risk modelling through data mining techniques,” *Int J Data Min. Model. Manag.*, vol. 11, no. 2, pp. 103–128, 2019.