Behavioral Learning of Vessel Types with Fuzzy-Rough Decision Trees

Rafael Falcon and Rami Abielmona
Research & Engineering, Larus Technologies Corporation
58 Antares Drive Unit 1, Ottawa ON, Canada K2E 7W6
{rafael.falcon, rami.abielmona}@larus.com

Erik Blasch
Air Force Research Lab
Rome, NY, USA, 13441
erik.blasch.1@us.af.mil

Abstract—A reliable and efficient characterization of vessel activities along coastal regions is of crucial importance for maritime domain awareness. With increased navigational flows across all waterways and the worldwide dissemination of active and passive vessel tracking modalities, learning a vessel’s behavior is becoming a strategic priority for maritime operators and decision makers. In this paper, we propose an interpretable computational model based on fuzzy-rough decision trees (FRDTs) to predict the vessel type given a summary vector in the form of descriptive track features that include kinematic, static and environmental information. The track summaries are generated from the fusion of Automatic Identification System (AIS), Synthetic Aperture Radar (SAR) and Canada weather reports. Our methodology uses fuzzy rough sets to discard irrelevant features on the basis of their dependency of the vessel type, prior to the iterative construction of the FRDT. Empirical results with a real-world data set in the east coast of North America confirm that the proposed approach is able to accurately assign the correct label (i.e., type) to previously unseen vessels in over 80% of the cases.

Index Terms—maritime domain awareness; behavioral learning; information fusion; track summaries; machine learning; automatic identification system; fuzzy decision trees; fuzzy rough sets

I. INTRODUCTION

In a world where more than 40% of the population lives within 100 kilometers of a coast [1] and where traditional and asymmetric threats to physical and cyber infrastructures continue to rise each year, countries are becoming increasingly aware of the importance in their ability to achieve persistent surveillance and continuous awareness of their maritime domains. In a vast and mostly uninhabited country such as Canada, which borders the Atlantic, Pacific and Arctic oceans, Maritime Domain Awareness (MDA) provides knowledge of potential threats from maritime activities. MDA is defined as the situational understanding of activities that impact maritime security, safety, economy and/or the environment [2]. MDA involves a system of people, processes and technological tools that discover, sense, analyse and react to events and perform physical and virtual defence of the country’s borders. It includes the capture and storage of domain knowledge obtained along with the actions, effects and outcomes for use in planning future surveillance operations. The outcome expected from MDA is the effective tasking of joint and interagency forces to respond to offensive/illegal activities, disasters and rescue scenarios in the maritime domain.

In order to accurately and effectively monitor a maritime area, the vast depth and breadth of incoming data must be interpreted and managed. Often referred to as the Big Data Problem, this challenge is best handled through the creation and maintenance of a real-time representative model of the world. Early solutions tried to resolve this challenge through Low-Level Information Fusion (LLIF) modules that used complex mathematical formulations or brute force number crunching; however, these solutions were inadequate because of the complexity created by the 4-dimensional vector subject to variety, volume, velocity and veracity, which quickly increased to the point where LLIF modules were overwhelmed. LLIF was only capable of performing fusion when the data itself was limited in volume, involved few types (low variety), did not frequently change in mission-critical applications (low speed), and was fairly trustworthy (high veracity).

To address the challenges of Big Data, High-Level Information Fusion (HLIF), which in the Joint Director of Laboratories (JDL) / Data Fusion Information Group (DFIG) models [3] [4] is defined as Level 2 Fusion and above, has become the focus of recent research and development efforts. HLIF capabilities continue to evolve so as to alleviate the challenges presented by Big Data including (i) anomaly detection, a process by which patterns are detected in a given dataset that do not conform to a pre-defined typical behavior (e.g., outliers), (ii) trajectory prediction, a process by which future positions (i.e., states) and motions (i.e., trajectories) of an object are estimated, (iii) intent assessment, a process by which object behaviors are characterized based on their purpose of action, and (iv) threat assessment, a process by which object behaviors are inferred based on their capability, opportunity and intent.

Behavioral learning provides a complex, yet innovative and adaptive, framework that involves learning by analyzing the changes in the system that occur as a result of stimulus-response experiences, as opposed to the knowledge-processing and memory-intensive cognitive learning paradigm. Behavioral learning has been used in intelligent systems and allows for an adaptive emergence of HLIF capabilities for MDA.

Inferring the type of a vessel given the observable evidence about its track has several important applications. First, it allows us to characterize the navigational patterns of all vessels
in a region of interest, which could be used to elucidate causality models (e.g., what are the likely reasons for tugboat tracks in this region to exhibit these features?). Second, once a vessel type has been learned from its track behavior, maritime operators can ascertain what its capabilities are, which in turn leads to a more precise intent and threat assessment. Third, if the inference process assigns confidence degrees for all vessel types, then borderline patterns can be easily identified and further analyzed (e.g., why is this cargo vessel behaving more like a pleasure craft?). Finally, if there is a mismatch between the reported and the predicted vessel types, this might be an indication of an incorrect transmission, a malicious intent of advertising the wrong vessel type, or another anomalous event.

In this paper, we propose an interpretable computational model based on fuzzy-rough decision trees (FRDTs) to learn the vessel type given a summary vector in the form of descriptive track features that include kinematic, static and environmental information. The track summaries are generated from the fusion of Automatic Identification System (AIS), Synthetic Aperture Radar (SAR) and Environment Canada reports. Our methodology uses fuzzy rough sets [5] to discard irrelevant features on the basis of their dependency of the vessel type, prior to the iterative construction of the FRDT. Empirical results with a real-world data set in the east coast of North America confirm that the proposed approach is able to accurately assign the correct type to previously unseen vessels.

The rest of the paper is structured as follows. Section II briefly reviews relevant works. Section III unveils the proposed Machine Learning approach while Section IV sheds light on the empirical evaluation. Finally, some conclusions and pointers to future work are provided in Section V.

II. RELATED WORK

This Section briefly reviews several relevant studies in maritime behavioral learning, HLIF of AIS/SAR for MDA as well as fuzzy rough sets and FRDTs.

A. Maritime Behavioral Learning

Maritime behavioral learning includes event-level learning from motion analysis to determine normalcy of situational behaviors. In 2007, Rhodes et al. [6] predicted behaviors based on the learning of the target motions within a scene. They used the AIS reports from an MDA system. The analysis of the track and classification (threats as identifications) results support inter-event activity analysis. In 2008, they followed up with a multiscale analysis for event-level detection [7]. The AIS ship data was put into categories and normalcy activities were determined with fine and coarse model learning. In 2009, they developed an adaptive neural network (ANN) for behavior monitoring with a recursive version of the expectation maximization (EM) algorithms to minimize the Kullback-Leibler (KL) information metric [8]. The goal was to shift the focus of the user from monitoring to activity assessment for timely response. The adaptive mixture-based neural network classifier algorithm is composed of five main stages, namely, class (category) choice, class (category) match, recursive learning, information based pruning and anomaly detection. They used the Bayesian Information Criteria (BIC) to simultaneously choose the number of categories (clusters) and estimate the model parameters with respect to model accuracy, BIC utility, and number of categories created. Another example uses motion pattern estimation from the AIS data using a kernel density estimation technique [9]. Normalcy models are based on the contextual operating conditions of targets applied to many domains [10].

Mascaro et al. [11] [12] developed two anomaly detection schemes that are trained on vessel tracks (coming from AIS and weather data) using Bayesian networks. The time series approach works at the contact level while the track summary approach encompasses all contacts in the track. Both schemes output anomaly scores that are reflective of the oddity of a certain contact/track w.r.t. a learned normalcy model and offer complementary views of a vessel’s behavior. Despite the encouraging results reported in terms of accuracy and interpretability, the learning process is quite complicated and slow (as evidenced by the entangled topology of the Bayesian networks obtained) while the level of involvement of a domain expert in the definition of the multi-tier causal structures that prevail among the network nodes is not trivial.

Our track-level methodology, though not geared towards anomaly detection but rather vessel type prediction, improves upon all these aspects while retaining crucial properties like model interpretability and simplicity as well as the generation of confidence degrees for the classifications made. Additionally, the track summaries in our work take into account SAR reports which are associated with the AIS tracks and the environmental information.

B. AIS/SAR HLIF for Maritime Domain Awareness

Recent efforts in HLIF techniques for MDA have utilized clustering techniques for SAR and AIS data. Clustering is a widely popular data mining technique that helps reveal meaningful knowledge structures (clusters) amid the seemingly disordered distribution of data points in the feature space. In 2007, Laxhammar [13] applied two unsupervised learning methods (clustering algorithm and self-organizing neural network) to the anomaly detection problem of vessels in the southern Swedish coastline using only AIS reports. For training data, only AIS reports pertaining to normal vessel activity were considered. Anomalous behavior was used for the testing data. The chief idea was to model normal activity and then flag any significant deviation from what is normal as anomalous. The surveillance region is discretized as a grid and anomalous activity recognition is performed in each cell. The author proposes two feature spaces for each vessel: \( F_1 = (\text{Vel}_x, \text{Vel}_y) \) and \( F_2 = (\text{Vel}_x, \text{Vel}_y, \text{Lat}, \text{Lon}) \).

The first clustering technique is the Mixture of Gaussians (MoG) model. It tries to identify a number \( c \) of Gaussian functions \( N(\mu, \Sigma) \) that could have generated the observed data distribution. The estimation of the Gaussian centers \( \mu \) and the covariance matrices \( \Sigma \) is iteratively done via the Expectation-Maximization (EM) algorithm. A greedy version of EM is used
to construct the clusters incrementally instead of having the value of $c$ set beforehand. Once the MoG has been achieved, any future observation (data point) that significantly deviates from the modeled pattern is flagged as anomalous.

The second clustering technique put forth in [13] is a neural network that is able to dynamically learn the number of clusters in the system. The Fuzzy Adaptive Resonance Theory (FuzzyART) network is trained with normal data patterns. If a pattern is deemed to resemble fairly well an existing category (output layer unit, cluster), the category itself is updated; otherwise, if the pattern does not represent any existing category, then a new category is created on the fly. At this point, the data pattern is regarded as abnormal. Notice, however, that the order in which the data patterns are presented to the network will greatly influence the categorization process.

The anomaly detection frameworks (MoG, FuzzyART) put forward in [13] can only detect very simply types of anomalies (e.g., vessel crossing sea lane or traveling in opposite direction of the sea lane) because of the limited information drawn from AIS. The author observed that the inclusion of vessel latitude and longitude in the second feature space did not appear to enhance the detection capabilities of such anomalies.

Like many pattern learning frameworks, [13] is general enough for multiple domains involving generic motion in the 2-D plane, requiring minimal adaptation and no need for specific domain knowledge given its unsupervised nature. MoG is more suitable when training data contains noise or anomalies. FuzzyART implies fast learning yet it is more sensitive to noise.

Recently, Shao et al. [14] employed computational intelligence methods such as Fuzzy K-Nearest Neighbors to correlate SAR/AIS/Ground-Moving Target Indicator (GMTI) vessel reports and subsequently associate them via Fuzzy C-Means. They also augmented the traditional Kalman Filter for vessel track prediction with an Echo State Network to better model nonlinearity.

C. Fuzzy Rough Sets and Fuzzy-Rough Decision Trees

Decision trees (DT) [15] are among the most popular and successful data mining techniques. These computational models are able to extract meaningful patterns from large bodies of data through an inductive process that gradually connects the observed evidence with some conclusion (e.g., a class label) drawn about it. Although DT are better known in their role as classifiers, they also have prominent applications in regression, clustering and feature selection. One of DT’s main appeals is their interpretability, i.e., the decision rules that are extracted from the tree topology are usually easy to understand by the end user.

Fuzzy logic has come to boost the explanatory capabilities of standard decision trees by modeling the system attributes as linguistic variables and injecting linguistic terms (such as tall, high, poor or excellent) into the inductive generation of the tree structure. Fuzzy decision trees (FDTs) [16] are better equipped to deal with the uncertainty, vagueness and noise that permeates real life through the fuzzy representation of the variables at play. In [17], the authors dissect several metrics that are typically used for ranking attributes during the recursive construction of an FDT, such as fuzzy information gain, fuzzy Gini index, or fuzzy ambiguity.

More recently, fuzzy rough sets [5] have become an attractive option for the development of FDTs. This is because, as in rough set theory [18], the imprecision of a concept can be captured through its lower and upper approximations, with the added advantage that both approximations are modeled after fuzzy sets, hence also capturing concept vagueness and uncertainty, as shown in Fig. 1.

\[ U = \text{Set of Objects} \]

Fig. 1. A fuzzy rough set representing a concept $X$ through its fuzzy lower approximation $B_L X$ and fuzzy upper approximation $B_U X$ induced by a subset of attributes $B \subseteq A$.

The term “fuzzy-rough decision tree” (FRDT) often refers to an FDT enhanced with properties borrowed from rough set theory. Jensen and Shen [19] employ the fuzzy-rough dependency as the attribute selection criterion within Fuzzy ID3. They contrast this metric with other well-known criteria and realize it is indeed competitive and advantageous. The authors in [20] and [21] follow the same idea, except that the former scheme uses the fuzzy lower approximation to derive the optimal branching point along the domain of a numerical attribute and the latter relies on the fuzzy-rough dependency to define the termination criteria during the FDT construction.

Our proposed approach is neither a generalization of Fuzzy-ID3 as in [19] nor does it only render a binary FRDT like that in [20]. It is closer to the approach in [21] and yet different in the sense that the fuzzy-rough uncertainty is used as the discriminative factor whenever the fuzzy-rough dependency is unable to discern among two or more candidate attributes for expansion. We also perform feature selection based on fuzzy-rough sets prior to the top-down recursive FRDT generation in order to remove irrelevant attributes and hence refine the data-mined predictive model.

III. PROPOSED METHODOLOGY

This Section sheds light on the data and the algorithmic framework used in this study to predict a vessel’s type from its track summary for MDA.

A. Experimental Data

Our experimental dataset consists of (1) a 5-day ExactEarth\(^1\) feed of 227,299 AIS contacts along the eastern coast of Canada

\(^1\)http://www.exactearth.com
and northeast USA; (2) a SAR feed of 188 contacts from Canada’s RADARSAT-2 satellite for the same area and period of interest; and (3) Environment Canada’s publicly available repository³ of hourly weather reports from 30 stations in October 2012, totalling 162,389 records.

After discarding contacts with missing or invalid values, the remaining AIS records were correlated by their Maritime Mobile Service Identity (MMSI) field (unique vessel identifier). Multiple tracks for the same vessel were created if it was either stopped or did not transmit for at least 6 hours. Fuzzy K-Nearest Neighbors [14] was used to associate SAR contacts to AIS tracks, as displayed in Fig. 2. All SAR contacts were correctly associated with their respective AIS tracks, i.e. no dark targets were spotted. These associated contacts were augmented with weather reports if their issue date was within 1 hour of the contact’s time stamp and the corresponding weather station was at most 50 km away from the contact’s received position. Once singleton-contact tracks were removed, 8,526 tracks representing 173,457 contacts were available.

Fig. 2. An example of SAR contacts (white dots) being associated with AIS tracks (in red and orange) via Fuzzy KNN.

The training data for our supervised learning scheme comes in the form of a decision system $DS = (U, A \cup \{d\}, V, f)$ where $U$ is the set of all examples (track summaries generated from the above AIS/SAR/weather tracks) which are described by the conditional attributes in $A$ and whose vessel type is the decision attribute $d$. The domains $V_i$ of all attributes are contained in $V$ and $f : U \times A \rightarrow V$ is the information function that assigns a particular value to a training example in a certain attribute. Following the idea in [11], we created track summaries out of the 19 numerical attributes in Table I which include static, positional, dynamic, and environmental attributes. The domain $V_d$ of the decision attribute $d$ is the set of 17 decision classes (vessel types) depicted in Table II.

### B. Vessel Type Prediction

Our goal is to learn a model from the available data that can predict the type of vessel that most likely resembles a given track summary. It is important that the inference process leads to confidence degrees (regarding how likely it is that a track summary belongs to a particular vessel type, say cargo or tug) and that it provides insights as to how that conclusion was reached (interpretabiliy). For the reasons mentioned in Section II-C, we picked a fuzzy-rough decision tree as our classifier of choice. Algorithm 1 outlines the “Behavioral Learning using Fuzzy-Rough Decision Tree” (BLuFuRoDT) method.

1) **Attribute Selection:** Our first step is the removal of irrelevant or low-quality attributes before we start inducing the FRDT in a top-down manner. This will reduce the complexity of the subsequent data-mined model and will boost its interpretability without sacrificing its predictive power. Three filter approaches will be used to weigh the conditional attributes based on their ability to infer the value of the decision attribute. Section IV-A elaborates on this crucial preprocessing step.

2) **Fuzzy Information System (FIS) Generation:** Before we start constructing the FRDT from the available data, our approach requires an a priori fuzzification of all conditional attributes $a \in A$ and of the decision attribute $d$ in order to reduce the granularity of the attribute domains and identify the representative values, which shall then be modeled as linguistic
Algorithm 1 BluFuRoDT( )

Input: A filtered fuzzy information system $FIS_α = (U, A \cup \{d\}, V, f)$, minimum significance level $α$, truth level $β$, maximum depth $ρ$

Output: A FRDT and its ensuing fuzzy rule base

1: create a new FDT with a single root node;
2: if stop condition met then
3:   label FDT either as null node or as a leaf node with the decision class $d^*$; return FDT;
4: end if
5: $a^*$ ← the attribute to expand in the FDT;
6: for each linguistic term $l \in L(a^*)$ do
7:   $FDT_{child}$ ← generate branch and delete it if empty;
8:   add $FDT_{child}$ as child of $FDT$ with edge labeled $l$;
9: end for
10: return FDT;

The depth of the alpha cut is indicated by the decision attribute $d^*$.

\[ \mu^{(j)}_{iv} = \begin{cases} \mu_{iv}^{(j)} & \text{if } \mu_{iv}^{(j)} \geq \alpha \\ 0 & \text{otherwise} \end{cases} \]  
\[ \forall v \in L(a_i), 1 \leq i \leq |A \cup \{d\}|, 1 \leq j \leq N \]

4) Iterative Attribute Expansion: In our top-down FRDT induction scheme, the attribute that possesses the highest discriminative power w.r.t the decision $d$ among all the attributes in the recursive subsample of the original training set is selected for expansion. It is well-documented [15] that the criterion used to judge among attributes will have a pivotal influence on the efficacy and efficiency of the resultant tree. We use the fuzzy-rough dependency measure $τ_a(d)$ as opposed to the more traditional fuzzy information gain or ambiguity in light of the encouraging results reported in the literature [20][19][21]. This $τ_a(d)$ metric, defined by Equation (2), is interpreted as the ratio of the cardinality of the fuzzy positive region $μ_{POS_{a}}(d)$ induced by fuzzy attribute $a$ on the fuzzy decision attribute $d$ to the number of examples in the data set.

\[ τ_a(d) = \frac{\sum_{x \in U} μ_{POS_{a}}(d)(x)}{|U|} \]

The fuzzy positive region of the example $x$ induced by fuzzy attribute $a$ on $d$ is calculated as follows:

\[ μ_{POS_{a}}(d)(x) = \sup_{c \in L(d)} μ_{gc}(x) \]

where $c$ is a fuzzy decision class belonging to the fuzzy decision attribute $d$ and $μ_{gc}(x)$ is the fuzzy lower approximation of concept $c$ according to example $x$ induced by the fuzzy attribute $a$:

\[ μ_{gc}(x) = \sup_{l \in L(a)} \min \left( μ_l(x), \inf_{y \in U} I(μ_l(y), μ_{gc}(y)) \right) \]

and $I(x,y) = \min(1-x+y,1)$ is Łukasiewicz’s fuzzy implicator.

At every iteration, let $a^*$ denote the attribute $a \in A$ with the highest value of $τ_a(d)$. If two or more attributes possess the highest fuzzy-rough dependency degree $τ_a^*(d)$, we need further information to make the decision regarding which of them should be expanded. In this paper, we use the fuzzy-rough uncertainty measure $δ_a(d)$ as the discriminating factor. We compute the fuzzy-rough uncertainty for each of those attributes as defined in (5) and select the one with the smallest value, where ties are broken arbitrarily.

\[ δ_a(d) = \frac{\sum_{x \in U} μ_{BND_{a}}(d)(x)}{|U|} \]

where $μ_{BND_{a}}(d)(x)$ is the fuzzy boundary region induced by fuzzy attribute $a$ on the fuzzy decision attribute $d$ according to example $x$ and can be computed as in expression (6).

\[ μ_{BND_{a}}(d)(x) = \sup_{c \in L(d)} μ_{pc}(x) - μ_{gc}(x) \]

4For the sake of brevity, we omit the details here and refer the interested reader to [22].
with \( \mu_{ac}(x) \) as defined in (4) and \( \mu_{d}(x) \) being the fuzzy upper approximation induced by fuzzy attribute \( a \) on the fuzzy decision attribute \( d \) according to example \( x \) and whose expression is given below:

\[
\mu_{d}(x) = \sup_{y \in U} \min \left( \mu_{a}(x), \sup_{y \in U} \min ( \mu_{a}(y), \mu_{c}(y) ) \right) \tag{7}
\]

The attribute \( a^* \) that is finally selected for expansion will generate \( |L(a^*)| \) branches, i.e. one per linguistic term associated with the fuzzy variable \( a^* \). Each branch will be labeled after the linguistic term \( l \in L(a^*) \) from which it sprung. If none of the examples in a branch reaches the minimum significance level \( \alpha \) in its membership grade to the underlying fuzzy concept, the branch is deleted.

5) Termination Criteria: The inductive FRDT generation will stop if any of the following criteria is met:

- no more conditional attributes can be expanded, i.e. \( FIS_{\alpha} = (U, \{ d \}, V, f) \);
- a user-specified maximum tree depth \( \rho \) is reached.
- the classification truth level \( \sigma_l(c) \) of the linguistic term \( l \in L(a^*) \) that originated this branch node with respect to any fuzzy decision class \( c \in L(d) \) exceeds a user-specified threshold \( \beta \); or
- no further expansion is possible as the remaining attributes only generate empty branches.

The classification truth level is defined in (8). The numerator quantifies the fuzzy dependency of the decision attribute \( d \) on the linguistic term \( l \in L(a) \) and the denominator is the fuzzy cardinality (or sigma-count) of \( l \).

\[
\sigma_l(c) = \frac{\sum_{x \in U \in L(d)} \sup_{y \in U} \min \left( \mu_l(x), \inf_{y \in U} I \left( \mu_l(y), \mu_c(y) \right) \right)}{\sum_{x \in U} \mu_l(x)} \tag{8}
\]

If the first stop criterion fires, the branch is terminated as a null node. In the rest of the cases, the branch is terminated as a leaf node labeled with the fuzzy decision class \( c^* \) having the highest truth level, i.e. \( c^* = \arg \max_{c \in L(d)} \sigma_l(c) \). Ties are broken in favor of the class with the highest sigma-count.

6) Classifying New Examples: Unlike a classical decision tree (in which the label value for a test example is obtained after following a path from the root to one leaf), in an FDT each path may yield a different truth level for a certain label value and they all have to be considered. Hence, the following steps have to be taken [15] to classify an unlabeled example:

1) Fuzzify the test example according to the scheme in Section III-B2.
2) For each path from the root to a leaf, calculate the joint membership grade of the example through all its intermediate nodes (rule antecedents) by applying a t-norm operator, e.g., the minimum or product operators. This membership grade will be associated with the label of the corresponding leaf node (rule consequent).
3) For each label, apply a t-conorm (e.g., maximum) to aggregate the membership grades in all its leaves.

4) Assign the label with the highest aggregated membership grade. Alternatively, the example may be assigned different labels with the membership grades calculated in the previous step.

IV. EXPERIMENTAL EVALUATION

In this Section, we empirically validate the proposed Blu-FuRoDT vessel type prediction scheme described in Section III-B using the data set outlined in Section III-A. All simulations were conducted in RapidMiner\(^5\) 5.3 with a Core i7-3840QM @ 2.80 GHz and 20 GB RAM under Windows 7.

A. Attribute Selection

Three filter approaches, namely correlation-based weak association (CBWA) [26], maximum relevance (MR) [26] and fuzzy-rough dependency degree (FRDD) [19] are employed to weight each of the 19 conditional attributes according to the strength of their predictive relationship with the decision attribute (vessel type). A stability analysis was carried out to determine how resilient each method is when faced with data sets of different cardinalities. Fig. 3 displays the robustness (calculated after the Jaccard index) of the feature subsets returned by each filter method over 10 independent runs with bootstrapped example sets of varying size.

![Fig. 3. Robustness of each univariate attribute weighting scheme as the examples in the original data set are bootstrapped with stratified sampling.](http://www.rapidminer.com)

One may notice that FRDD is always superior to CBWA and preferable to MR when the number of vessel track summaries for the training set is scarce. As more track summaries are available, both FRDD and MR are more robust since they return feature subsets with 80-85% elements in common over multiple runs with variable numbers of training examples. On average, FRDD exhibits 74% robustness index, MR 69.9% and CBWA 66.8%. Therefore, the fuzzy-rough-set-based attribute selector becomes our choice to reduce the dimensionality of the vessel track summary feature space.

Table III reports the attribute weight vector elicited by FRDD on the original data set. We retained 10 out of the 19 conditional attributes after the preprocessing phase. Ship length stands as the most discriminative attribute, followed

\(^5\)http://www.rapidminer.com
TABLE III
FRDD: Normalized attribute weights

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Weight</th>
<th>Relevant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>ship_length</td>
<td>1.0</td>
<td>Yes</td>
</tr>
<tr>
<td>avg.speed</td>
<td>0.183</td>
<td>Yes</td>
</tr>
<tr>
<td>max.speed</td>
<td>0.183</td>
<td>Yes</td>
</tr>
<tr>
<td>speed_st_dev</td>
<td>0.183</td>
<td>Yes</td>
</tr>
<tr>
<td>course_st_dev</td>
<td>1.00</td>
<td>Yes</td>
</tr>
<tr>
<td>heading_st_dev</td>
<td>0.097</td>
<td>Yes</td>
</tr>
<tr>
<td>duration</td>
<td>0.082</td>
<td>Yes</td>
</tr>
<tr>
<td>end_point_lat</td>
<td>0.055</td>
<td>Yes</td>
</tr>
<tr>
<td>start_point_lat</td>
<td>0.052</td>
<td>Yes</td>
</tr>
<tr>
<td>max_lat</td>
<td>0.051</td>
<td>Yes</td>
</tr>
<tr>
<td>start_point_lon</td>
<td>0.050</td>
<td>No</td>
</tr>
<tr>
<td>end_point_lon</td>
<td>0.049</td>
<td>No</td>
</tr>
<tr>
<td>max_lon</td>
<td>0.047</td>
<td>No</td>
</tr>
<tr>
<td>avg_min_temp</td>
<td>0.037</td>
<td>No</td>
</tr>
<tr>
<td>avg_max_temp</td>
<td>0.037</td>
<td>No</td>
</tr>
<tr>
<td>avg_dewpoint_temp</td>
<td>0.027</td>
<td>No</td>
</tr>
<tr>
<td>avg_rel_humidity</td>
<td>0.014</td>
<td>No</td>
</tr>
<tr>
<td>avg_pressure</td>
<td>0.007</td>
<td>No</td>
</tr>
<tr>
<td>avg_wind_speed</td>
<td>0.0</td>
<td>No</td>
</tr>
</tbody>
</table>

by the speed, course, heading and duration indicators. The geographical attributes (latitude, longitude) do not seem to arbitrate very well among the different vessel types. On the other hand, the six environmental attributes have little to say in terms of mapping a vessel type to its track summary, which is caused by relatively uniform weather conditions reported in the short (5-day) period of interest that affect all vessel trajectories. We should expect an increased influence of these weather-related attributes as their variability spikes under a prolonged time period (e.g., summer vs. winter).

B. Predicting Vessel Types

We compare our BluFuRoDT method with other traditional classifiers such as the C4.5 decision tree, Naïve Bayes, K-Nearest Neighbor (KNN), Multi-Layer Perceptron (MLP), Support Vector Machines (SVM) and a Random Forest made up of 10 decision trees. Table IV reports the average accuracy over 10 folds of a stratified cross validation of each classifier with the track summary data set after attribute selection.

TABLE IV
Vessel type prediction accuracy (in %)

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-NN</td>
<td>62.89 ± 6.07</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>50.38 ± 4.76</td>
</tr>
<tr>
<td>C4.5</td>
<td>54.93 ± 4.10</td>
</tr>
<tr>
<td>Random Forest</td>
<td>51.17 ± 2.13</td>
</tr>
<tr>
<td>MLP</td>
<td>81.50 ± 5.71</td>
</tr>
<tr>
<td>SVM</td>
<td>62.75 ± 5.06</td>
</tr>
<tr>
<td>BluFuRoDT</td>
<td>80.74 ± 7.73</td>
</tr>
</tbody>
</table>

Despite the removal of the attributes with little informative bearing on the vessel type, several class distributions still exhibit significant overlap, e.g., tanker/cargo or fishing/pleasure craft, which gives rise to complicated decision boundaries that, in some cases, the classifiers are not able to well delineate in an attempt to avoid overfitting. BluFuRoDT and MLP are the only two classifiers rising above 80% accuracy threshold in their predictions. Yet MLP is unable to provide a clear interpretation of the inference process leading to the recommended vessel class, unlike BluFuRoDT which can easily induce a fuzzy rule base to link its conclusions to the initial evidence (track summary indicators).

We could ascribe BluFuRoDT's superiority over five well-known classification models (including Random Forest, a meta-classifier) to the combination of (a) the removal of non-influential fuzzy sets associated to the linguistic terms of the pruned attributes, which reduces the sensitivity of the fuzzy-rough dependency measure and (b) the selection of the conditional attribute with the smallest fuzzy-rough uncertainty w.r.t. the decision attribute among all those having the highest dependency degrees during the iterative attribute expansion. This latter step helps curb the detrimental effect caused by random selection of one of those attributes for split. Instead of blind selection, BluFuRoDT makes an informed choice to reduce the gap between the fuzzy upper and lower approximations of a decision class.

C. Parametric Configuration

To discover a near-optimal parameter configuration for our FRDT-based prediction model, we leaned upon an evolutionary strategy (ES) algorithm to navigate across the parameter space \((k, \alpha, \beta)\) defined by the following bounds: \(k \in \{2, 10\}\), \(\alpha \in [0, 0.5]\) and \(\beta \in [0.5, 1]\). The ES runs for 50 generations with 8 population members and a binary tournament selection scheme. Each candidate parameter vector endures crossover and a Gaussian mutation with 90% and 10% probabilities, respectively. The best parametric configuration found by the ES method is as follows: \(k = 4, \alpha = 0.185\) and \(\beta = 0.875\).

Fig. 4 portrays a subtree of the resultant model induced by BluFuRoDT. The ship length attribute is acting as the root given its predominant fuzzy-rough dependency degree with respect to the vessel type (decision) attribute.

V. Conclusions

In this paper, we have developed a behavioral learning model rooted on FRDTs to ascertain the vessel type given a set of descriptive track features that include kinematic, static and environmental information. The track summaries are generated from the fusion of AIS, SAR and Canada weather reports. Our methodology uses fuzzy rough sets to discard irrelevant features on the basis of their dependency of the vessel type, prior to the iterative construction of the FRDT. The attribute to be expanded at each iteration is the one with the smallest fuzzy-rough uncertainty among those showing the highest fuzzy-rough dependency of the vessel type. Empirical results with a real-world data set in the east coast of North America confirm that the proposed approach is able to accurately assign the correct label (i.e., type) to previously unseen vessels in about 80% of the cases while providing an interpretable model with confidence levels per class.

For future work, we will investigate the influence of different fuzzification methods on the ensuing FRDT and will test the approach with longer and more involved data sets. We are
Fig. 4. A subtree of the FRDT induced by BluFuRoDT. Each numerical attribute has been fuzzified into four linguistic terms, namely, low, medium, moderate and high. The leaves are labeled after the vessel type with the highest classification truth level (in brackets).

also presently working on online vessel type prediction, i.e. as new contacts arrive and the track is being formed, an online track summary is elicited. Finally, using BluFuRoDT for vessel track anomaly detection is another promising endeavour.

REFERENCES


