Turing Detection in Sandbar Sharks through Accelerometer Data

A thesis submitted in partial fulfillment of the requirement for the degree of Bachelor of Science in Computer Science Department from The College of William and Mary

by

Benjamin Powell

Accepted for Honors

Gang Zhou
Type in the name, Director

Denys Poshvyvanyk
Type in the name

Timothy Davis
Type in the name

Kevin Weng
Type in the name

Laurie Sanderson
Type in the name

Williamsburg, VA
April 30, 2018
Abstract

This paper introduces a novel method for classifying turning in sandbar sharks from accelerometer data. Because marine organisms are difficult to observe visually, attachable tags are often used instead, typically including only an accelerometer for power reasons; measuring complex motions, such as turning, is difficult without a gyroscope. Six features, including a novel metric designed to detect turning direction, are used. To deal with sample imbalance, a mixture of SMOTE and modified bootstrap sampling is employed. The classifier has 79% accuracy and can highlight periods of heightened turning activity as well as determine the direction of swimming.
List of Figures

1  Still from footage of the tank. The top is covered in plastic, and the toroidal shape constricts the movement of the shark further. All work with sharks was conducted in accordance with an approved animal care protocol, IACUC-2017-05-26-12133-kcweng. Photo credit: Dan Crear, VIMS. .......................................................... 8

2  Frequency of each behavior in the labeled data. .......................................................... 9

3  Amount of data labeled for each shark. .......................................................... 9

4  Diagram showing the location of the tag on the shark as well as its coordinate system. For the sake of clarity, the X, Y, and Z axes are referred to as “forward”, “vertical”, and “lateral” acceleration in this paper. .......................................................... 10

5  Diagram showing the centrifugal force exerted on the tag (red arrow) as the shark turns (yellow arrow). .......................................................... 10

6  Histograms comparing the occurrences of lateral acceleration and modified lateral acceleration $z'$. .......................................................... 11

7  Comparison of Power Spectral Distributions (PSD) of straight swimming and turning. The straight swimming spectrum is more complex because it was made from a longer sample. .......................................................... 12

8  The orientation of the shark over the course of a single right turn. The time scale of this figure differs from figure 9. .......................................................... 13

9  The orientation of the shark as it completes one counterclockwise revolution around the tank. .......................................................... 13

10 Plot comparing $\gamma$ and $C$ values for each bootstrap sample after optimization. Three outliers were discarded with $C$ values of 51.899, 947.15, and 0.002, respectively. .......................................................... 18

11 GUI mockup of application. Histograms show at what times right turns are occurring. .......................................................... 20

B.1 Pitch and roll of the shark over one left turn. The first two and a half seconds cover the turn itself, while the rest have the shark return to straight swimming. .......................................................... 24

B.2 Acceleration in each axis as the shark completes one counterclockwise revolution around the tank. Gravity is found in the vertical axis, but there is a significant offset to the lateral acceleration as well, suggesting centrifugal force or a slight roll. .......................................................... 24

B.3 Acceleration in each axis as the shark completes two turns in quick succession. Label times are approximate, but it is difficult to see turning behaviors in the accelerometer without further analysis. .......................................................... 25

B.4 Spectrogram of lateral acceleration for over an hour of mostly straight swimming. The fundamental frequency of the tailbeat frequency is visible as a strong band around 1Hz. Some lower-frequency signal might also exist, but it is difficult to discern due to the existence of a DC offset as seen in Figure A.2. .......................................................... 25
List of Tables

1. Mean and standard deviation of ODBA for each class. ........................................ 13
2. Modified cost matrix. Correct classifications have a cost of 0; incorrect classifications are penalized differently depending on their importance in successfully detecting turning. ................................................................. 16
3. Comparison of performance over different sampling techniques. AUROC is macro-average discrimination – the average discrimination for each class. ............................. 17
4. Confusion matrix for final optimized classifier. Rows correspond with actual class, and columns predicted class. ................................................................. 20
Acknowledgements

The author thanks the advisor for this project, Dr. Gang Zhou, as well as his grad students for their boundless patience and invaluable assistance.

The Virginia Institute of Marine Science (VIMS) and the VIMS Eastern Shore Lab facilities were integral to this work, and the author thanks Dan Crear, Rich Brill, Peter Bushnell, Gail Schwieterman, Sean Fate, PG Ross, Justin Paul, Read Boniwell, Stephanie Boniwell, Richard Snyder and Jim Brister for assistance with experiments.

The author would also like to thank Wouter Deconinck, Kevin Weng, and the members of the Sharkduino project for their help.

The author would like to thank the members of his honors committee: Gang Zhou, Kevin Weng, S. Laurie Sanderson, Timothy Davis, and Denys Poshlyvanyk.

Finally, the author thanks his parents, whom he loves, as well as the four sharks who swam for hours on end for the benefit of science.

This work was supported by NSF CNS1253506 (CAREER), the William and Mary Makerspace, and VIMS.
# Contents

1 Introduction ........................................... 7

2 Data Sources and Collection .......................... 8

3 Sensors and Feature Selection ......................... 9
   3.1 Windowing ........................................ 9
   3.2 Conditional Segmented-Mean Lateral Acceleration 10
   3.3 Tailbeat Features .................................. 11
   3.4 Orientation ....................................... 12
   3.5 ODBA .............................................. 13
   3.6 Generic Features .................................. 14

4 Classification .......................................... 14
   4.1 Standardization And Split ......................... 14
   4.2 Automatic Feature Selection ...................... 15
   4.3 Resampling ........................................ 15
      4.3.1 Cost Matrix .................................. 15
      4.3.2 Bootstrapping ................................ 16
      4.3.3 SMOTE ........................................ 16
      4.3.4 Our Approach ................................ 17
      4.3.5 Performance Comparison ...................... 17
   4.4 Classifier Design .................................. 17

5 System Implementation .................................. 18

6 Performance ........................................... 19

7 Discussion and Future Work .......................... 19

8 Related Works ......................................... 21

9 Conclusion ............................................ 22

A Data and Code ........................................ 23

B Additional Figures .................................... 24
1 Introduction

Marine wildlife is difficult to track or observe through cameras, so attachable tags, typically incorporating accelerometers or gyroscopes, are often used instead. Thus, the problem of classifying behaviors by looking at acceleration and/or angular velocity is well-studied. Gyroscopes are ideal for measuring complex movements, and can be combined with an accelerometer to very accurately estimate orientation, but they can draw 1,000 times as much current at similar sample rates. Tags must be small and light compared to the animal, limiting battery size and thus deployment time.

Measuring turning through an accelerometer in particular is a difficult proposition. The pitch and roll of the animal can be estimated through the gravity vector (static acceleration), and pitch has been used to estimate the orientation of many animals [1] [2], but the yaw – the critical component of turning – can only be estimated from a gyroscope or magnetometer. However, the use of secondary behaviors associated with turning, such as inward rolling and centrifugal force, provide a way to determine the direction and strength of the turn.

In this paper, we introduce a novel method for classifying the turning of sandbar sharks in a toroidal tank through supervised machine learning on triaxial accelerometer data, distinguishing both between the presence of the turn and the direction of swimming. The accelerometer data is broken into regular windows, and features are derived for each window. The data is then resampled through a combination of SMOTE and bootstrap sampling to deal with sample imbalance, and is fed into a Support Vector Machine with an Error-Correcting Output Codes model to solve the multiclass problem. The classifier is 94% accurate and, while not precise enough to reliably identify individual turns, can detect the direction of swimming and highlight areas of frequent turning.

The use of Support Vector Machines to classify the behavior of animals is not new [3], but classifiers are typically not tuned to accurately detect short “transitional” behaviors; furthermore, no papers have attempted to classify turning, and only one other paper [4] has classified the behavior of sandbar sharks. The degree of imbalance between turning and straight-swimming samples is also very high, necessitating the use of extreme resampling techniques.

The remainder of this paper is structured as follows. In section 2, we cover the context under which the data were collected and our labeling process. We then discuss our windowing process, the features extracted from each window, and our rationale for each in section 3. The automatic feature selection, resampling, and pre-processing techniques are explained in section 4, as well as the classifier and optimizer used. We briefly demonstrate in section 5 the application created to let marine biologists apply this system to their own data. We discuss the performance of the classifier, the implications of that performance, and possible future work in turning estimation in sections 6 and 7. Finally, we briefly cover related work in the field in section 8 and give our conclusions in section 9.
2 Data Sources and Collection

Our data comes from hypoxia trials conducted on three juvenile sandbar sharks (*Carcharhinus plumbeus*) from June to August 2015 by the Gloucester Point Virginia Institute of Marine Science [4]. The trials measured the maximum and routine metabolic rates (energy expended per unit time) of the sharks. Each shark was caught and placed in a toroidal tank and was tested over two days. On the first day, the sharks were periodically prodded (an action known as the “chase protocol”) to find the maximum metabolic rate of the shark. Over the night, they were allowed to reach their routine metabolic rate, the lowest metabolic rate they can attain while swimming. On the second day, the shark was starved of oxygen until it could no longer maintain its metabolic rate [4].

Each shark was about 80cm in length, and was tested individually. The tank, shown in figure 1 was small enough to constrict the movement of the shark, forcing it to either swim around the tank or make tight 180-degree turns. Unlike some species of shark, sandbar sharks do not rest on the bottom, but instead swim continuously. Ultimately, we decided to detect four behaviors: clockwise swimming, counterclockwise swimming, tight left turns, and tight right turns. Categorizing the behavior by both turning and direction has two purposes. The direction of the swimming or turning will result in different signs for certain features, such as lateral acceleration. It is also more detailed and thus may be of greater use to researchers.

As shown in figure 2, sharks spent only a small amount of time (2.722%) turning. They were more likely to swim clockwise (56.95%) than counterclockwise (40.32%) in the labeled data, although the labeled data made up a small portion of the total collected data, increasing the likelihood of sampling error. The imbalance in the data set presented serious problems in successful classification; solutions explored to mitigate this imbalance are presented in section 4.2.
<table>
<thead>
<tr>
<th>Behavior</th>
<th>Frequency</th>
<th>Shark Name</th>
<th>Data Labeled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clockwise</td>
<td>56.95%</td>
<td>SB28</td>
<td>02:58:14</td>
</tr>
<tr>
<td>Counterclockwise</td>
<td>40.33%</td>
<td>SB30</td>
<td>05:43:03</td>
</tr>
<tr>
<td>Right Turn</td>
<td>2.059%</td>
<td>SB34</td>
<td>03:01:05</td>
</tr>
<tr>
<td>Left Turn</td>
<td>0.662%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Frequency of each behavior in the labeled data.

Figure 3: Amount of data labeled for each shark.

An overhead camera recorded the sharks during the day, giving us a ground truth by which we could label our data. Ideally, we would like to use data from the night, when the metabolic rate of the shark was typical. Unfortunately, the camera could not see the shark at night time, making labeling impossible. Our training data instead comes from the first day during periods when the shark was not being prodded. This was done in the hopes that the increased metabolic rate would lead to more instances of tight turning.

We ultimately labeled 7 hours and 35 minutes of data between all 3 sharks, marking the time at which a shark ended one behavior and began a new one. The times for each shark are seen in figure 3. A shark was said to start turning as soon as its body started to bend in the direction of a turn, and was said to start swimming straight as soon as its body fully straightened out. These were chosen only for the sake of consistent labeling, and did not follow any standard from the field of marine biology. The mean turn duration was 1.81 seconds with a standard deviation of 3.9 seconds. Labels were accurate to the second because of the windowing system discussed in section 3.1.

3 Sensors and Feature Selection

A Gulf Coast Data Concepts, LLC X16-mini tri-axial accelerometer was attached to each shark on the right side of the dorsal fin, as shown in figure 4. Wiggling in the dorsal fin can be problematic, particularly when determining the orientation (pitch and roll) of the shark from static acceleration as documented by Brownscombe et al. [2]. All vertebrate animal research was conducted in accordance with an approved animal care protocol at the College of William & Mary, IACUC-2017-05-26-12133-kcweng. Each accelerometer collected data at 25 samples per second and had a maximum range of 16g.

3.1 Windowing

Classifying shark behavior at a given moment requires knowledge of the data surrounding that moment. Because of this requirement, the data were grouped into 60 sample (2.4 second) windows, 30 of which were shared with the windows around it (15 for each side). Each window was given
the label that occupied most of the window. For example, if a window consisted of 20 samples of right turning followed by 40 samples of clockwise swimming, then the window would be classified as clockwise.

The size and overlap were selected by comparing the precision and recall of several options in an ad-hoc fashion, though the choices were motivated by Fida et al [5], specifically the degree of overlap.

### 3.2 Conditional Segmented-Mean Lateral Acceleration

It was predicted that the turning of the shark would produce a small but measurable lateral acceleration. In the reference frame of the tag, this would appear as a centrifugal (outward) acceleration along the lateral axis. The sign of the acceleration, in particular, was important for determining the direction of swimming: it would be negative for left turns (counterclockwise swimming) and positive for right turns (clockwise swimming).

To exaggerate the significance of the sign, a modified version of mean lateral acceleration was added as a feature $\bar{Z}'$. First, a threshold $t$ was calculated as the mean lateral acceleration of the entire data set. Then, for each window $Z_w$, the positive and negative samples ($Z_+$ and $Z_-$) were isolated. The final value for window $w$ was

$$\bar{Z}'_w = \begin{cases} 
Z_+ & \text{if } |Z_+| \geq |Z_-| \\
Z_- & \text{otherwise}
\end{cases}$$

As the plots in figure 6 show, this resulted in fairly distinct separation between counterclockwise and clockwise swimming when compared to the unweighted mean. Similar results were also obtained for distinguishing between left and right turns.
### 3.3 Tailbeat Features

Straight swimming (defined as swimming clockwise or counterclockwise around the tank without turning) is characterized in the frequency domain by a strong peak around 1Hz, as seen in figure 7, due to the tailbeat of the shark. Turning behaviors typically lack this peak and are more constant, so the amplitude of the peak during straight swimming should be high compared to the mean amplitude of the entire window.

To find the tailbeat frequency, the window is first converted into the frequency domain via a Fast Fourier Transform. Frequencies outside of the 0.8-1.6Hz range are discarded, and the peak is chosen by the highest point in the remaining frequencies.

To determine if the tailbeat is significantly larger than the surrounding environment, an additional tailbeat distinctiveness feature is derived for each window $W$:

$$\text{distinctiveness} = \frac{\max W}{\frac{1}{|W|} \sum W}$$

This is the ratio of the amplitude of the tailbeat frequency to the mean amplitude of the window. If the tailbeat is not very distinct, it is likely non-existent, and thus would point to a turning behavior.

In practice, as shown in figure 7, the spectra of turning and straight swimming is not too different, but the 1Hz spike in the straight swimming is absent in the turning.
Figure 7: Comparison of Power Spectral Distributions (PSD) of straight swimming and turning. The straight swimming spectrum is more complex because it was made from a longer sample.

3.4 Orientation

While imperfect, the pitch and roll of the shark can be estimated with an accelerometer through static acceleration. Static acceleration, associated with the gravity vector, is separated from dynamic acceleration by a 3-second central moving average [6]. As in Brownscombe et al [2], instantaneous pitch $P$ is derived from static forward acceleration $y_s$ and mean forward acceleration $\mu$:

$$P = \sin^{-1}(y_s - \mu)$$

Oddly, the roll $R$ is not a common feature in many papers, but can be derived from the lateral static acceleration $z_s$ and vertical static acceleration $y_s$:

$$R = \tan^{-1}(-z_s/y_s)$$

The patterns in the orientation of the shark change for straight swimming and turning. As shown in figure 8, turns have larger extremes, with pitches and rolls in excess of 20 degrees. Figure 9 shows that straight swimming has smaller fluctuations but also different means.

Since one of the sharks tended to roll inward during turns, the change in orientation over time $\Delta R$ and $\Delta P$ were also considered. Delta-pitch was already used in Brownscombe to detect bonefish behaviors [2]. Since $R$ and $P$ are discrete, for a sample $i$:

$$\Delta P_i = P_i - P_{i-1}$$
Figure 8: The orientation of the shark over the course of a single right turn. The time scale of this figure differs from figure 9.

Figure 9: The orientation of the shark as it completes one counterclockwise revolution around the tank.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Turn</td>
<td>1.23</td>
<td>0.10</td>
</tr>
<tr>
<td>Right Turn</td>
<td>1.22</td>
<td>0.11</td>
</tr>
<tr>
<td>Counterclockwise</td>
<td>1.21</td>
<td>0.10</td>
</tr>
<tr>
<td>Clockwise</td>
<td>1.20</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Table 1: Mean and standard deviation of ODBA for each class.

\[
\Delta R_i = R_i - R_{i-1}
\]

\(P\) and \(R\) were converted into features for each window by taking their mean, standard deviation, and maximum and minimum values, as well as their “peak-to-peak” \((P_{\text{max}} - P_{\text{min}})\) values. \(\Delta P\) and \(\Delta R\) were converted into features by taking their mean, minimum, maximum, and skewness.

### 3.5 ODBA

Overall Dynamic Body Acceleration (ODBA) is a simple but widely-used metric in marine biology that is widely correlated with energy expenditure [7]. It is the orthogonal sum of the dynamic acceleration of the animal on all three axes:

\[
\text{ODBA} = |x_d| + |y_d| + |z_d|
\]

It was believed that turning behaviors may have slightly lower ODBA than straight swimming behaviors. As shown in figure 1, the ODBA of each class were within one standard deviation of each other, limiting the usefulness of the feature somewhat. Regardless, the mean ODBA of each window was included as a feature.
3.6 Generic Features

Along with features designed for this dataset, “generic” features used in many behavior classification papers (like Martiskainen et al. [3]) were also included. The following features were added for every axis of the acceleration of each window:

- Mean $\mu$
- Standard deviation $\sigma$
- Skewness $s = \frac{E(x-\mu)^3}{\sigma^3}$
- Kurtosis $k = \frac{E(x-\mu)^4}{\sigma^4}$
- Maximum
- Minimum
- Position of the maximum in the window
- Position of the minimum in the window
- Mean of squared acceleration $S = \frac{1}{w} \sum_{i=1}^{n} |A_i|^2$

where $E(t)$ is the expected value of $t$.

Additionally, the Pearson correlation coefficient between the $xz$, $xy$, and $yz$ axes were calculated, as seen in Martiskainen et al [3]. The coefficient $r$ between two samples $x$ and $y$ is

$$r = \frac{\sum_{i=1}^{n} (x_i - \hat{x})(y_i - \hat{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \hat{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \hat{y})^2}}$$

where $\hat{x}$ and $\hat{y}$ are the means of each sample.

These feature choices were arbitrary, but as discussed in Section 4.2, feature selection is employed to automatically vet each feature.

4 Classification

4.1 Standardization And Split

Before feature selection or classification, each feature $F$ was standardized by subtracting the mean $\mu$ and dividing by the standard deviation $\sigma$, giving each feature equal weight:

$$F' = \frac{F - \mu}{\sigma}$$

The data is then shuffled and split into training and test sets. 80% of the data is placed in the training set and 20% is placed in the test set.
4.2 Automatic Feature Selection

In total, 58 features were derived, but not all of them were likely to be helpful to the classifier; in fact, the presence of “noise” features could increase variance and lead to overfitting. Sequential forward selection [8] was used for automatic feature selection. In sequential forward selection, a criterion (in this case, error percentage) is obtained by training and testing a model from a subset of all features, starting with 1 feature and adding additional features to the set if they decrease the criterion. Five-fold cross validation is used to split training samples into training and test samples. Six features were ultimately selected:

- Conditionally-segmented forward acceleration
- Mean forward acceleration
- Mean vertical acceleration
- Maximum lateral acceleration
- Vertical mean of squares
- Lateral mean of squares

Feature selection increased F measure from 66% to 75% and improved the classification of right turns, anticlockwise swimming, and clockwise swimming. In particular, the classifier was extremely likely to classify clockwise samples as anticlockwise before feature selection.

4.3 Resampling

Our two “minority classes” – left turns and right turns – make up only 2.7% of our dataset, so the first instinct of the classifier is to choose one of the “majority classes” – clockwise or counterclockwise swimming – in every case. Imbalanced datasets are a common problem in machine learning, with a variety of proposed solutions. We tested the effects of SMOTE [9], bootstrap sampling [10], and cost matrix modification [10] on the performance of the classifier, eventually deciding to use a modified approach mixing both bootstrap sampling and SMOTE.

4.3.1 Cost Matrix

The cost matrix provides a simple way to affect the cost function of a supervised classifier by describing the cost of predicting a sample of class $i$ to be of class $j$. We tested the effects of modifying the cost matrix to punish the misclassification of left and right turns as clockwise and counterclockwise swimming; at the same time, we made the classifier less concerned with distinguishing
### Table 2: Modified cost matrix.

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
<th>L-Turn</th>
<th>R-Turn</th>
<th>CC-Wise</th>
<th>C-Wise</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.2</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Correct classifications have a cost of 0; incorrect classifications are penalized differently depending on their importance in successfully detecting turning.

between left and right swimming and between clockwise and counterclockwise swimming, as shown in figure 2.

Several similar cost matrices were also evaluated with different degrees of misclassification cost for left and right turns (5, 20, and 100). Ultimately, the unweighted classification matrix achieved the best performance; see section 4.3.5.

#### 4.3.2 Bootstrapping

Wallace et al [10] found that undersampling the majority class tends to improve classifier discrimination for many cost-based classifiers, including SVMs. However, the loss of data led to greater variance in each classifier.

To combat this, Wallace et al created multiple undersampled “bootstrap” samples by drawing with replacement from the training data, training a classifier for each bootstrap, and using a majority vote to make predictions about new samples. We use a modified version of bootstrap sampling, described in section 4.3.4, with 11 bootstraps.

#### 4.3.3 SMOTE

Chawla et al [9] solved the data imbalance problem by creating synthetic minority samples from existing samples. Their algorithm was fairly simple:

1. For every minority sample, pick $k$ of its nearest neighbors.
2. Randomly choose $m$ neighbors from $k$.
3. For each chosen neighbor, create a new sample $N$ by linearly interpolating between the original sample $A$ and its neighbor $B$ in feature space by a random factor $0 \leq c \leq 1$ such that

   $N = A + (N - A) \times c$

We arbitrarily chose $k = 6$ and $m = 3$. 

16
### 4.3.4 Our Approach

Our resampling algorithm is a combination of modified bootstrap sampling and SMOTE. For each bootstrap, we first take $n$ minority samples without replacement; traditional bootstrapping samples with replacement, but due to the low number of samples it is believed that sampling without replacement will reduce the number of duplicates. We then generate $3n$ new minority samples with SMOTE using $k = 6$. Finally, we take $4n$ majority samples with replacement (since the number of majority samples is far greater than $4n$).

This approach leads to an imbalance between the number of left and right turn samples as well as an imbalance between clockwise and counterclockwise swimming samples. However, undersampling each class until they had the same number of samples decreased f measure by 5% and did not help in the detection of left turns.

### 4.3.5 Performance Comparison

Figure 3 shows the general performance of each sampling technique over the test set. Unmodified classification has a higher overall F measure, which would seem to indicate that it is the superior approach. However, the raw method failed to classify any left or right turn samples correctly, so it is not useful for our purposes. The same is also true for the modified cost matrix. Bootstrapping correctly classified some right turns, but the number of false positives for turning samples was high. Modified bootstrapping reduced the number of false positives without significantly increasing the number of false negatives.

### 4.4 Classifier Design

Behavior classification papers [11] [3] employ a variety of algorithms, but most commonly use decision trees, SVMs, or an unsupervised method such as k-means clustering. Each of the seven bootstrap samples were classified with a Support Vector Machine (SVM) with a Gaussian kernel, using an Error-Correcting Codes model for the purpose of multiclassing. Each bootstrap underwent hyperparameter optimization separately from the others via a 30-evaluation Bayesian optimization; three parameters were optimized:

- Box constraint $C$ (effectively the maximum cost of misclassification; higher box constraints

<table>
<thead>
<tr>
<th></th>
<th>Raw</th>
<th>Cost</th>
<th>B-Strap</th>
<th>Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>F Measure</td>
<td>0.85</td>
<td>0.85</td>
<td>0.60</td>
<td>0.75</td>
</tr>
<tr>
<td>AUROC</td>
<td>0.79</td>
<td>0.78</td>
<td>0.82</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Table 3: Comparison of performance over different sampling techniques. AUROC is macro-average discrimination – the average discrimination for each class.
Figure 10: Plot comparing $\gamma$ and $C$ values for each bootstrap sample after optimization. Three outliers were discarded with $C$ values of 51.899, 947.15, and 0.002, respectively.

- Kernel scale $\gamma$ (related to the construction of the separating hyperplane, tends to be proportional to the average distance between samples)
- Coding method (method of multi-classification, either one-versus-one or one-versus-all)

Perhaps due to the small number of samples, variance in choices of $C$ between bootstraps were high. Even after discarding outliers, $C$ values ranged from 0.602 to 20.597, with the majority of bootstraps either near 1 or near 20. Kernel scale was more consistent, with a mean $\gamma$ of 1.006 and a standard deviation of 0.3914. A scatter plot of these parameters can be found in figure 10. The classifiers were evenly divided between coding method, with 6 favoring one-versus-one and 5 favoring one-versus-all.

Enabling optimization resulted in significantly longer training times, but did not significantly improve accuracy; f measure and AUROC are similar, and turning samples were not much better classified. The final results are given in section 6.

5 System Implementation

To provide non-technical users with the ability to classify their own shark data using our model, a standalone desktop app was written in Haskell. Our data is loaded and classifiers are trained externally in MATLAB, producing 11 bootstrap classifiers which are then embedded into the app.
Within the app, users upload raw sample data through CSV files, supplying the accelerometer data and times at each sample. The application is specifically designed for Gulf Coast Data Concepts tags, but will take any CSV as long as the data has a 25Hz sample rate and the start time is provided. Other sample rates should not be used, as the classifiers are trained specifically on 25Hz data.

The app imports the data, windows it, runs the classifiers, and gives both a prediction for each window and a prior probability score for each class. The prediction is given by the majority vote of each bootstrap. The probability score for a given class is the mean of the score of each bootstrap for that class.

Since windows overlap, many time frames have two predictions and sets of probabilities. All windows in the data frame (save the last one) are thus broken into two sub-windows: the window proper, and the overlapping area immediately following it. The overlapping windows are treated as having 21 bootstraps – two sets of 11 bootstraps, one from each surrounding window, with one bootstrap discarded at random to eliminate the possibility of a tie.

The predictions and probabilities can then be exported as a CSV or viewed through one of several displays:

- A summary set of statistics, including the total durations and frequencies of each class of behavior
- Histograms detailing when different behaviors occur as a function of time of day or another feature (for example, mean ODBA)
- A timeline showing individual turning events, as well as probability scores of each window

A mockup of the first two views can be seen in figure 11.

6 Performance

The confusion matrix, shown in figure 4, reveals that the classifier had a high number of false positives for both left and right turns; despite this, the actual recall for both is rather low, as previously mentioned. Counter-clockwise and clockwise swimming have better performance, particularly in distinguishing between one another.

The F measure of the data is 79.3%.

7 Discussion and Future Work

The classifier was good at distinguishing between clockwise and counterclockwise swimming and fairly good at distinguishing between right and left turns. Of the samples correctly classified as
turning samples, 86% were correctly predicted to be of a certain direction. The directions of 93% of the samples classified as non-turning were also correctly predicted.

The classifier was not, however, able to consistently detect individual left and right turns. The confusion matrix in figure 4 shows that a prediction of a turning behavior (either right or left turn) has only a 9% likelihood of actually being a turning behavior; similarly, 68% of turning samples were classified as non-turning samples.

Ultimately, about 320 turning samples (and 15,000 non-turning samples) were fed into the classifiers, and while additional samples were generated through SMOTE, Brodley et al [10] noted that the new samples are still within the boundaries of the existing samples, limiting its usefulness. More turning samples (or possibly better SMOTE techniques) may have made it easier for the classifier to find meaningful patterns in the data.
A higher ratio of majority (straight swimming) samples to minority (turning) samples in each bootstrap may have improved performance, since there was a heavy tendency to misclassify clockwise and counterclockwise samples as right turns. Using a different ratio for each bootstrap, or a different degree of SMOTE, may yield better results – it would combine balanced and imbalanced approaches to deal with different situations.

One of the major issues with detecting turning behaviors may come from our windowing system. Turning behaviors may not be wholly encapsulated in a single window even with overlap, and if the window is too large, these kinds of transitional behaviors are often lost when features are derived. Our current windowing system is commonly used in real-time systems, but our offline system could benefit from the combined use of unsupervised clustering or novelty detection. Scholkopf et al describes an algorithm for novelty detection [12] which could be used to break data into “usual” and “unusual” sets, each having its own class weights or resampling degrees.

If it is possible to classify turning behaviors with an accelerometer, is is possible that a regression-based technique could be used to identify the degree of turning. A more power-intensive but proven turning detection method, such as a gyroscope or magnetometer, could be used as a ground truth.

The constricting nature of the tank likely assisted greatly in classifying behaviors. Classifying the behavior of fish in the wild is much more difficult, but it has been done before for non-turning behaviors [13]. Our results speak to the possibility of detecting the heading of fish by observing secondary “tells” and some degree of statistical guesswork.

8 Related Works

Brownscombe et al [2] used accelerometers and decision trees to study and classify a set of bonefish activities, including resting, swimming, and coasting. Their data was not as imbalanced, and they did not detect turning behaviors, but they went beyond our work in using a regression model to predict how often a fish spent doing a particular activity based on factors like time of day and water temperature.

Martiskainen et al [3] used nine features to classify the behaviors of cows via accelerometers using Support Vector Machines, using a one-against-all method to solve the multiclass problem. They used 10-second windows, which led to good accuracy in detecting constant behaviors. Our paper studies sharks and is more interested in shorter “transitional” behaviors, namely turning.

Whitney et al [14] were the first researchers to study shark behavior, specifically mating behavior, with accelerometers, but didn’t use supervised machine learning or classification, and didn’t study turning. However, they did use k-means clustering to recognize interesting parts of the data.

Noda et al [13] studied Japanese amberjacks with a 9DOF IMU containing a gyroscope, accelerometer, and magnetometer, allowing them to carefully describe the behavior of the fish in
various scenarios. They described the data rather than classifying it.

McClune [11] studied the Eurasian badger with an accelerometer. Two-second windows were enough to capture the “repetitive cycles of movement” of the badger. K-nearest neighbors classification was used to describe each behavior statistically, and a decision tree was constructed based on those statistics. Our paper studies sharks, but shares the goal of automatically interpreting accelerometer data.

Tsuda et al [15] monitored the spawning behavior of chum salmon with a two-axis accelerometer in their natural environments. Instead of using machine learning to classify behavior, characteristics of each behavior were found by cross-referencing accelerometer and video data.

9 Conclusion

In this paper, a novel method for classifying shark turning behaviors was proposed. Data from a tri-axial accelerometer was broken into overlapping windows and used to train multiple bootstrapped Support Vector Machines, which collectively made predictions about new samples. The classifiers were able to distinguish the direction of swimming. While they were unable to reliably detect individual turns, they could give useful information about the general frequency and causes of the erratic or unusual behavior indicated by turns.

The sampling method, a novel combination of SMOTE and modified bootstrap sampling, makes the most of rare turning information. It leads to superior recall and comparable precision and accuracy when compared to bootstrap sampling and training on unbalanced data.

No general solution exists to detect turning behaviors about the vertical axis (yaw) with one accelerometer, even turning in place, because the gravity vector measured by static acceleration does not change with yaw. Detecting turning through a gyroscope or magnetometer is simple and fairly accurate, but the high current draw of gyroscopes and magnetometers limits the lifetime of tags using them significantly. The ability to detect turning with only an accelerometer could be used to save power by only activating a gyroscope when frequent turning is present – or, if improved, could replace a gyroscope entirely in limited cases.
A Data and Code

The raw data for this project belongs to Dan Crear and Kevin Weng at VIMS. The author thanks them for allowing the data to be used for this project.

The MATLAB code used for this project can be found at https://github.com/TheBen27/Honors-Project. The raw accelerometer data needs to be downloaded separately.

The app covered in section 5 is available at https://github.com/TheBen27/Shark-Turning-App. It requires the Haskell Tool Stack, which is available at https://docs.haskellstack.org/en/stable/README/.
Figure B.1: Pitch and roll of the shark over one left turn. The first two and a half seconds cover the turn itself, while the rest have the shark return to straight swimming.

Figure B.2: Acceleration in each axis as the shark completes one counterclockwise revolution around the tank. Gravity is found in the vertical axis, but there is a significant offset to the lateral acceleration as well, suggesting centrifugal force or a slight roll.
Figure B.3: Acceleration in each axis as the shark completes two turns in quick succession. Label times are approximate, but it is difficult to see turning behaviors in the accelerometer without further analysis.

Figure B.4: Spectrogram of lateral acceleration for over an hour of mostly straight swimming. The fundamental frequency of the tailbeat frequency is visible as a strong band around 1Hz. Some lower-frequency signal might also exist, but it is difficult to discern due to the existence of a DC offset as seen in Figure A.2.
References


