Modeling Ostracods Motility Through Movement Using Machine Learning Based Probability Density

Mushfika Rahman

12/18/2024

1 Introduction

Predicting the motility of the underwater spices has been a matter of interest among the researchers. The benefit of predicting the movement of small underwater species is improved ecosystem monitoring, which aids in conservation efforts, sustainable species management, and understanding the impact of environmental changes on biodiversity. Our goal is to use machine learning based model to predict the path of the ostracods which will help us distinguish between alive & dead samples. The hypothesis that alive samples will exhibit active locomotion, enabling them to navigate against or across water currents, whereas deceased samples will move in accordance with the water flow. Thus the dead samples will likely to have a distribution similar to waterflow's distribution in their motility. The probability density function is essential to quantify uncertainty associated with random movement based on observations. Given N number of observations for one object $O_1, O_2, O_3, ..., O_N$ where O_i stands for $[x_i, y_i, t_i]$; $[x_i, y_i]$ is their position and t_i is the frame where it is located. Hence the path for object will be

$$\begin{aligned}
P(path) &= P(O_1, O_2, O_3, \ldots, O_N) && (1)\\
&= P(O_N|O_1, \ldots, O_{N-1})P(O_{N-1}|O_1, \ldots, O_{N-2}) && (2)\\
&= \prod_{i=2}^{N} P(O_i|O_1, \ldots, O_{i-1}) && (3)\\
&\approx \prod_{i=2}^{N} P(O_i|O_{i-1}) \ where \ P(O_i|O_{i-1}) && (4)\\
&\sim \mathcal{N}(\mu, \Sigma) && (5)
\end{aligned}$$

Furthermore, μ, Σ are the parameters using to describe the distribution of the displacement D vector which contains displacements $d_1, d_2, d_3, ..., d_N$ where $d_i = O_{i+1} - O_i$.Additionally, we will take advantage of $n \times n$ grid search of the movement across the video sequence. To learn our parameters we will use the coordinates of the tracked objects in the video sequence. Here, μ, Σ are defined as below:

$$\begin{aligned}
\mu &= \mathbb{E}[d_i] && (6)\\
&= \mathbb{E}[O_{i+1} - O_i] && (7)
\end{aligned}$$

$$\begin{aligned}\Sigma &= \mathbb{E}[(d_i-\mu)(d_i-\mu)^T] \qquad (8)\\ &= \mathbb{E}[O_{i+1}-O_i] \qquad (9)\end{aligned}$$

Our goal is to model an anamoly detector where the dead sample will be modeled with multivariate normal distribution and the alive ones will treated as an anomaly. The grid based approach will give us insight to object's entry exit points, direction of movement thus indicating presence and absence of objects in each grid. Thus the leanred parameters μ, Σ from observations across different grid cells will indicate the average location of object aiding to locate central tendency. Furthermore, the Σ will provide information about the spread and orientations. Thus, by analyzing the object's movement across grid cells will help us estimate the future position and common movement patterns.

2 Method

Our data is collected through imaging technique where microscopic imaging technique was utilized. The video sequence is broken down with multiple frames where each object is detected and tracked. From this tracking we get the sequence of (x,y) coordinates of the moving objects, we will calculate displacement across 2 dimensions for each objects, the displacement d_i is calculated for objects

$$(dx_i, dy_i) = [(x_{i+1} - xi), (y_{i+1} - yi)] \tag{10}$$

In some instances, the $i+1^{\text{th}}$ observation and i^{th} observation doesn't belong to consequtive frames, then 10 will be

$$(dx_i, dy_i) = [(x_{i+1} - xi/(t_{i+1} - t_i)), (y_{i+1} - yi/(t_{i+1} - t_i))] \tag{11}$$

Since our approach take advantage of a grid based approach we additionally calculate the grid cell where the displacement is located essentially essentially i^{th} observation lies.For instance, consider the following table for 3×3 grid search.

$\mu_{0,0}$ $\Sigma_{0,0}$	$\mu_{0,1}$ $\Sigma_{0,1}$	$\mu_{0,2}$ $\Sigma_{0,2}$
$\mu_{1,0}$ $\Sigma_{1,0}$	$\mu_{1,1}$ $\Sigma_{1,1}$	$\mu_{1,2}$ $\Sigma_{1,2}$
$\mu_{2,0}$ $\Sigma_{2,0}$	$\mu_{2,1}$ $\Sigma_{2,1}$	$\mu_{2,2}$ $\Sigma_{2,2}$

Thus, our training parameters μ, Σ will take account into the displacements in particular location (p, q) cell. The μ, Σ for (p, q) grid cell having displacement vector d_i vector containing $[dx_i, dy_i]$ displacements for m observations:

$$\mu_{(p,q)} = [(\frac{1}{m-1}\sum_{k=0}^{m} d_i)] \tag{12}$$

$$\Sigma_{p,q} = \begin{pmatrix} \text{Var}(dx_i) & \text{Cov}(dx_i, dy_i) \\ \text{Cov}(dy_i, dx_i) & \text{Var}(dy_i) \end{pmatrix} \tag{13}$$

Thus, for R objects, where one object $1 \leq r \leq R$ has n_r displacements and d_i $1 \leq i \leq n_r$. Therefore, for (p, q) grid cell where d_i lies we can find the number of displacements A with the help of indicator function I in particular cell has in the following way:

$$A = \sum_{i=1}^{R} \sum_{j=1}^{n_r} I[d_{ri} \in (p, q)] \tag{14}$$

Additionally, Our mu using the equation of 12 becomes:

$$\mu_{(p,q)} = [(\frac{1}{A} \sum_{r=1}^{R} \sum_{i=1}^{n_r} I[d_{ri} \in (p, q) \cdot d_{ri})] \tag{15}$$

Furthermore,an object having displacements $\langle d_1, d_2, \dots, d_n \rangle$ which has k dimensions, we calculate the probability density $\langle P(d_1), P(d_2), \dots, P(d_n) \rangle$ function across the grids. The probability density for $\langle d_j, \rangle$ located in p, q cell is calculated :

$$f(\mathbf{d_j}) = \frac{1}{(2\pi)^{k/2} |\Sigma_{p,q}|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{d_j} - \boldsymbol{\mu_{p,q}})^T \Sigma_{p,q}^{-1} (\mathbf{d_j} - \boldsymbol{\mu_{p,q}}) \right) \tag{16}$$

The equation 16 is used as our training and testing evaluation. Our model learns the parameters using the dead species examples and testing it with alive example likely to give us probability density value different than dead examples. Additionally, with the help of discrepant probability density values of alive and dead samples, we can set a limit to detect anomalous movement among the species. Furthermore, it helps us to predict the motility of the samples.

2.1 Predictive Task

Our predictive task is to distinguish between alive & dead samples based on their movement. Upon computing the probability density value of the displacements with the dead sample's statistics difference in their movement pattern is observed. Firstly, the assumption is that the dead sample follows a distribution differing from the alive sample since the alive samples have added velocity to flow against the current. Thus, by imposing a certain threshold, alive samples could be detected as outliers. The threshold range is between the minimum & maximum of the dead & alive sample's probability density values.

$$\text{threshold\_range} = [\min(f(d_{ja}), f(d_{jd})), \max(f(d_{ja}), f(d_{jd}))] \tag{17}$$

By exploring the threshold_range, whichever produces corrected classified points for all the object's select that for the desired threshold. For example, an object X_i having probability density $\langle P(d_1), P(d_2), \dots, P(d_n) \rangle$. Let w denote the size of the sliding window applied to the probability density sequence. Define W_k as the sliding window k of size w applied to the probability density sequence of X_i:

$$W_k = \langle P(d_k), P(d_{k+1}), \dots, P(d_{k+w-1}) \rangle, \quad \text{where } k \in [1, n - w + 1]$$

Define cls_{X_i}, the classification label for object X_i, as:

$$\mathrm{cls}_{X_i} = \begin{cases} \text{"a"} & \text{if } \forall W_k, (W_k) \leq \text{threshold} \\ \text{"d"} & \text{otherwise} \end{cases}$$

Furthermore, the model predicts motility alive vs dead using the Bayesian theorem. For example, an object X_i we calculate,

$$P(\mathbf{X_i} \mid C) = \log P(\mathbf{X_i} \mid C) = \sum_{j=1}^{n} \log P(d_j)$$

For the alive class, we compute the $P(\mathbf{X_i} \mid A)$ using the alive displacement statistics.

$$P(A \mid \mathbf{X_i}) = \frac{P(\mathbf{X_i} \mid A) \cdot P(A)}{P(\mathbf{X_i})}$$

For the dead class, we compute the $P(\mathbf{X_i} \mid D)$ using the dead displacement statistics.

$$P(D \mid \mathbf{X_i}) = \frac{P(\mathbf{X_i} \mid D) \cdot P(D)}{P(\mathbf{X_i})}$$

$$\mathrm{Class}(\mathbf{X_i}) = \begin{cases} \text{"a"} & \text{if } P(A \mid \mathbf{X_i}) > P(D \mid \mathbf{X_i}) \\ \text{"d"} & \text{if } P(A \mid \mathbf{X_i}) \leq P(D \mid \mathbf{X_i}) \end{cases}$$

3 Dataset Visualization

Our data set contains numerous living and dead samples and their path coordinates. Each alive or dead examples have variable size of observations which is across different frames how it moves in different position. In 17 & 18 we can see different movement patterns of the alive & dead samples

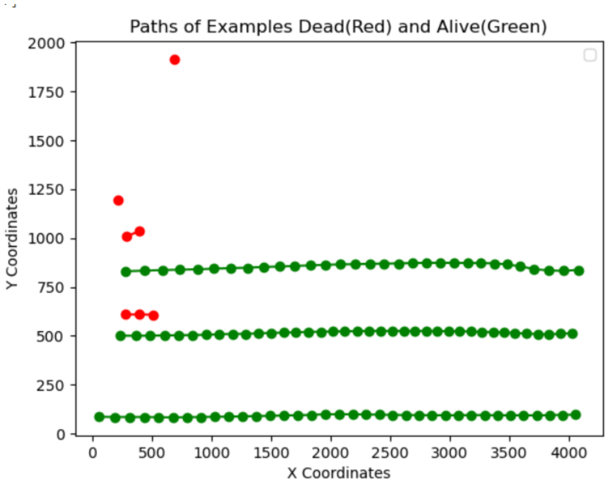

Figure 1: movement of dead & alive objects

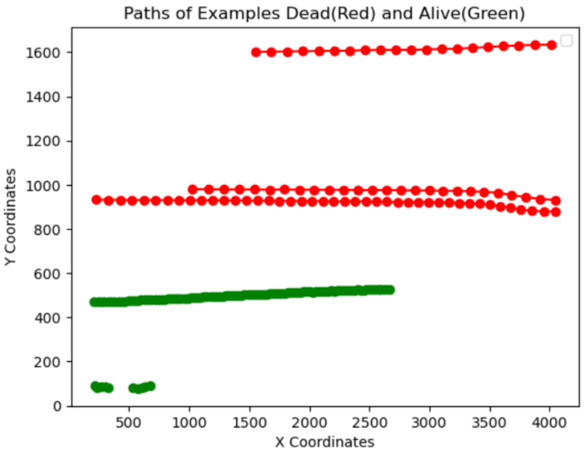

Figure 2: movement of dead & alive objects

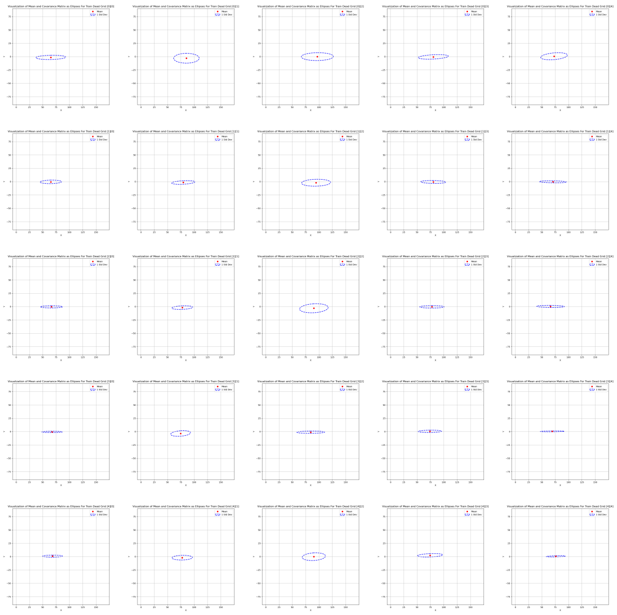

Figure 3: mu and sigma of trainning set's dead examples grid by grid displacementes

4 Result Analysis

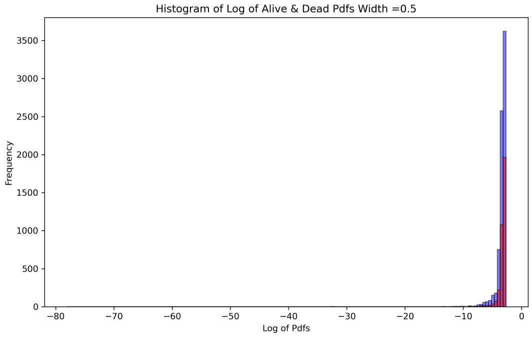

Figure 7: histogram of dead alive object's probability density values modeled with dead examples

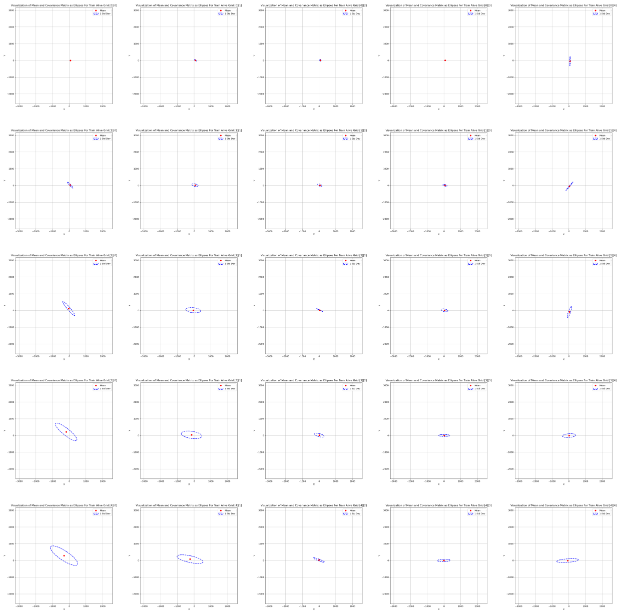

Figure 4: mu and sigma of trainning set's alive examples grid by grid displacementes

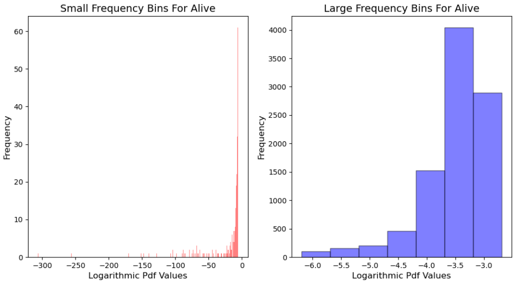

Figure 8: largest & smallest values of alive's probability density values modeled with dead

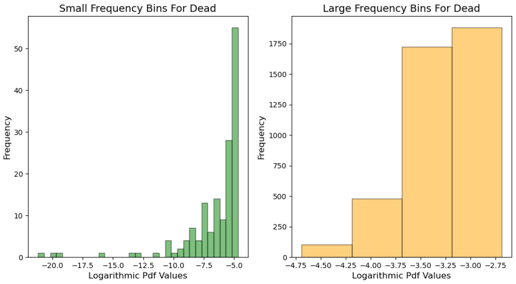

Figure 9: largest & smallest values of dead's probability density values modeled with dead

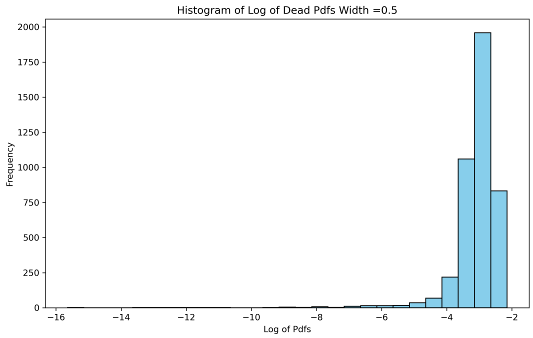

Figure 5: histogram of dead object's probability density log values modeled with dead examples

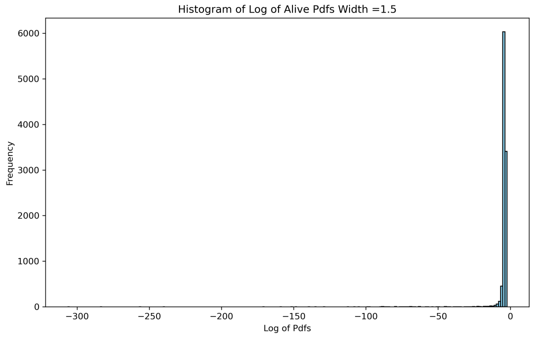

Figure 6: histogram of alive object's probability density log values modeled with dead examples

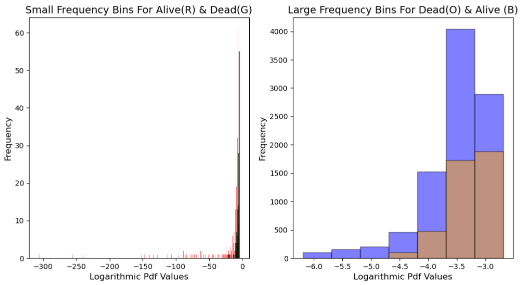

Figure 10: dead & alive's largest & smallest values of probability density values modeled with dead

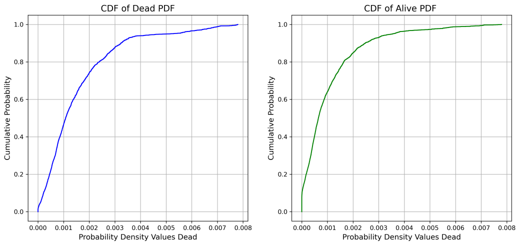

Figure 11: cdf of alive & dead with probability density values

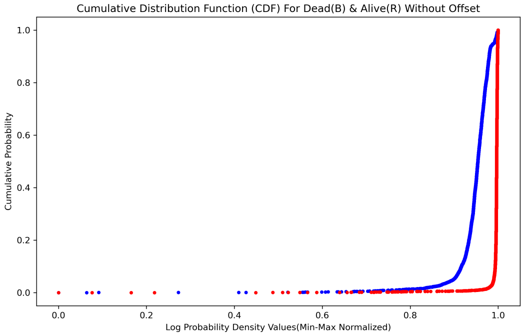

Figure 12: cdf of alive & dead with probability density values

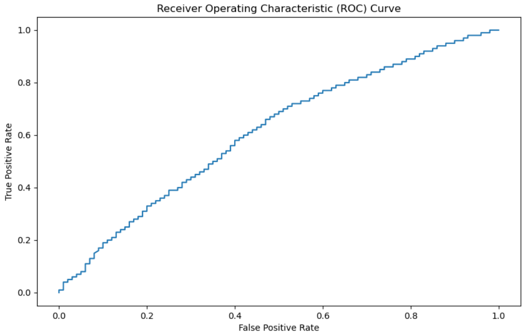

Figure 13: ROC curve of the thresholds

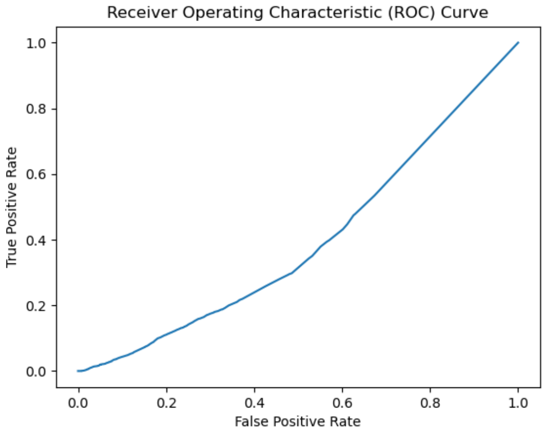

Figure 14: ROC curve by taking the minimum of object's probability density values

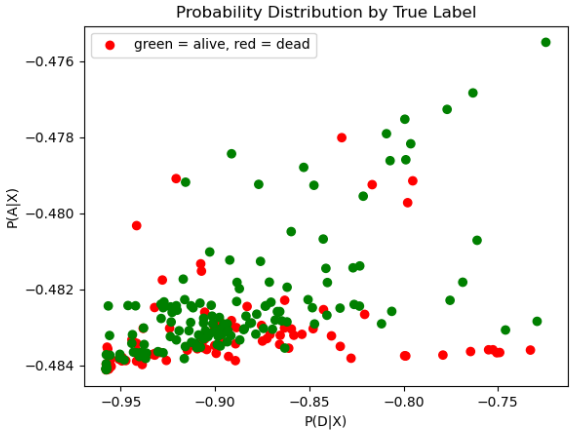

Figure 15: Points plotted with their probability distribution values

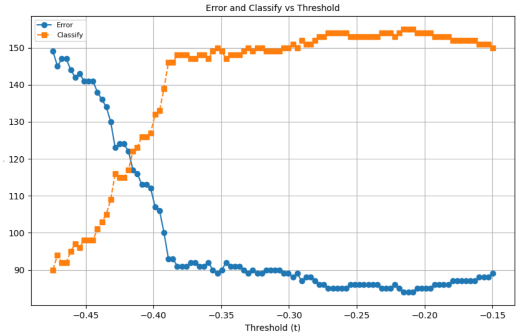

Figure 16: classification & errors for thresholds using bayesian

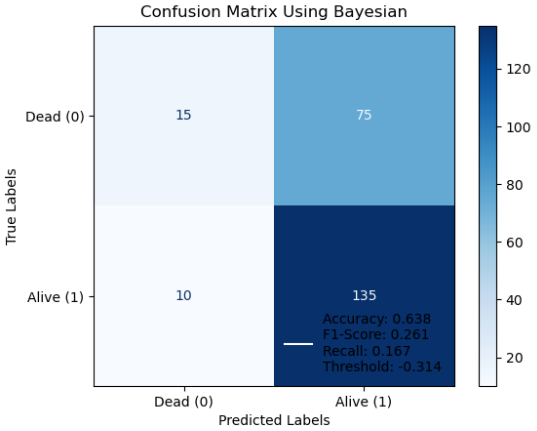

Figure 17: Confusion Matrix Training Set Using Bayesian Classifier

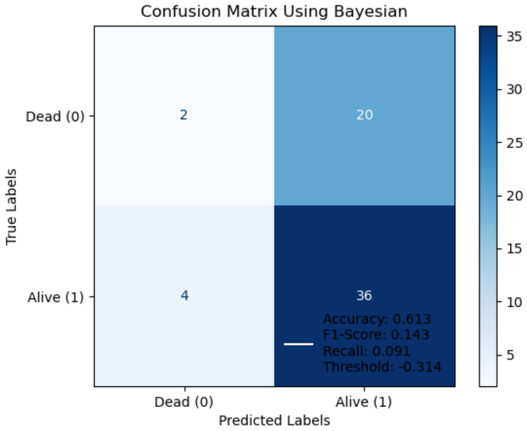

Figure 18: Confusion Matrix Testing Set Using Bayesian Classifier

References