On the determination of threshold of POM

Zijie Zhao

2018/10/11

On the determination of threshold of POM

This is a brief introduction about how to apply a simple model to daily POM - induced mortality data and hourly sea surface temperature to determine the threshold of a POM break.

The key functon used in this example is *oyster_accuracy(oyster_file,temp_file,limit,step,show_plot)*. This function is coded in R.

Some details of this function have been provided here. The running of this function requires five variables, which are: 1). *oyster_file* should be a string, indicating the name of csv file containing the POM - induced mortality. In this case, the csv file has been assumed to contain at least two columns: *Measurement Time*, indicating the sampling time for each observation, and *Value*, which is a binary variable where 0 indicates there is no significant POM - induced mortality and 2 otherwise.

2). *temp_file* should be a string, indicating the name of csv file containing the water temperature data. In this case, the csv file has been assumed to contain at least two columns: *Measurement Time*, indicating the measuring time for each observation, and *Value*, which should be a numerical vector containing water temperature in each time step.

3). *limit* should be a single scalar, indicating the maximum test threshold.

4). *step* should be a single scalar, indicating the step length of changing threshold. The length of test thresholds is therefore (limit-0+1)/step

5). *show_plot* should be a logical variable (T or F) to indicate if this function would return a sample figure to show the change of accuracy along test thresholds.

The output of this variable would be a list. This list includes:

1). *test_threshold* is a numerical vector containing all test thresholds determined by *limit* and *step* in inputs.

2). *best_threshold* is the threshold corresponding to the highest accuracy.

3). *accuracy* is a numerical vector containing accuracy along test_threshold

4). *plot* is a simple ggplot illustrating the whole process.

Let's do some exampples to see how it works. Here we used all float water temperature and wind inflection POM data.

Running the function

```
source('Oyster_accuracy.R')
library(ggplot2)
h<-oyster_accuracy(temp_file='All_Float.csv',oyster_file = 'Comb_POM.csv',limit=25,step=0.5,show_plot =
```

Showing the accuaracy,

```
h$accuracy
```

```
##  [1] 0.3952569 0.3952569 0.3952569 0.3952569 0.3952569 0.3952569 0.3952569
##  [8] 0.3952569 0.3952569 0.3952569 0.3952569 0.3952569 0.3952569 0.3952569
## [15] 0.3952569 0.3952569 0.3952569 0.3952569 0.3952569 0.3952569 0.3952569
## [22] 0.3952569 0.3952569 0.3952569 0.3952569 0.3952569 0.3952569 0.3952569
```

```
## [29] 0.3992095 0.4031621 0.4090909 0.4051383 0.4209486 0.4486166 0.4960474
## [36] 0.5375494 0.5770751 0.5968379 0.6245059 0.6363636 0.6561265 0.6660079
## [43] 0.6699605 0.6719368 0.6561265 0.6422925 0.6284585 0.6185771 0.6106719
## [50] 0.6086957 0.6086957
```

test thresholds,

```
h$test_threshold
```

```
##  [1]  0.0  0.5  1.0  1.5  2.0  2.5  3.0  3.5  4.0  4.5  5.0  5.5  6.0  6.5
## [15]  7.0  7.5  8.0  8.5  9.0  9.5 10.0 10.5 11.0 11.5 12.0 12.5 13.0 13.5
## [29] 14.0 14.5 15.0 15.5 16.0 16.5 17.0 17.5 18.0 18.5 19.0 19.5 20.0 20.5
## [43] 21.0 21.5 22.0 22.5 23.0 23.5 24.0 24.5 25.0
```

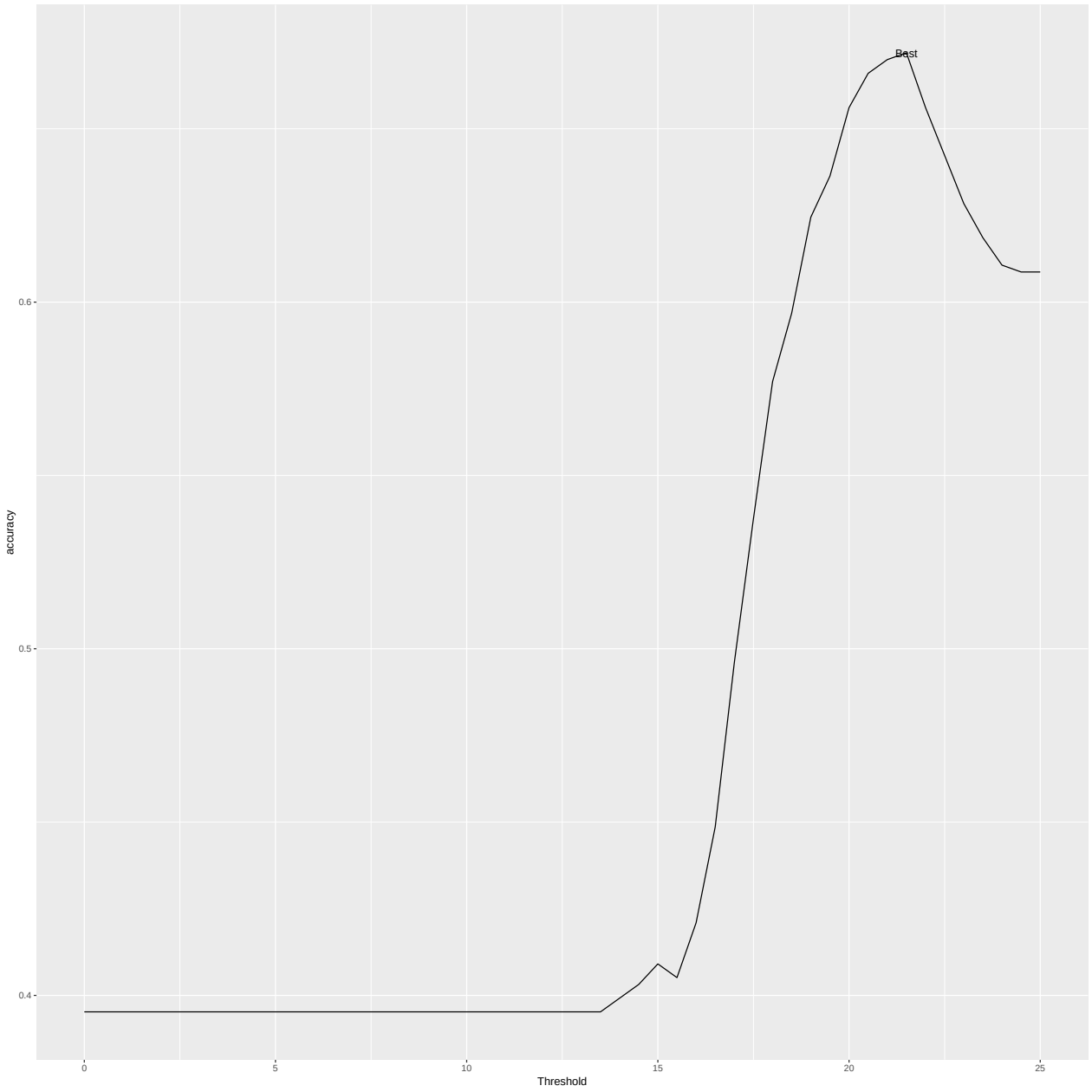
best threshold,

```
h$best_threshold
```

```
## [1] 21.5
```

plot,

```
h$plot
```


Best
0.6
0.5
0.4
accuracy
0
5
10
15
20
25
Threshold