chapter 2 - RF model

Dajun Wang

7/20/2020

Introduction

Material and methods

We used wildlife tracking collars (e-obs gmbh) equipped with tri-axial accelerometers sampling continuously at 10Hz to monitor behaviour in domestic and free-roaming dogs. The wildlife tracking collar (hence tri-axial accelerometer) was mounted such that the x-, y-, and z- axes were parallel to the median (surge), the dorsal (sway), and the dorsal (heave) planes of the animal, respectively. Eleven adult dogs (of German Shepard and Golden Retriever breeds) were selected and collared to obtain the training dataset required for the construction of the random forest model. All eleven dogs were well-trained individuals that could perform the selected repertoire of movement behaviours (see Table 1 for more information) through the verbal instructions of their trainer while being off-leash. Each dog was tasked to perform the entire repertoire of behaviours at least twice (with consent from their trainer), and approximately 300 minutes of video-recorded behavioural observations were obtained.

The acceleration measurements collected from the collared dogs were binned into two-second windows (or 'bursts') and were labelled accordingly with the associated dog behaviour. Labelling was done by matching with the video footage time-stamps of the performed dog behaviour to the time-stamps of the collected acceleration measurements. All performed (and observed) dog behaviours were limited into the seven categories as described in Table 1, and behavioural transitions were not catalogued since they occurred very quickly (i.e., less than 1s). Labelled acceleration measurements were binned into two-second bursts as this duration was found to accomodate at least two full strides for motion-based dog behaviours (i.e., walking, foraging) without much influence from un-intentional behavioural transitions.

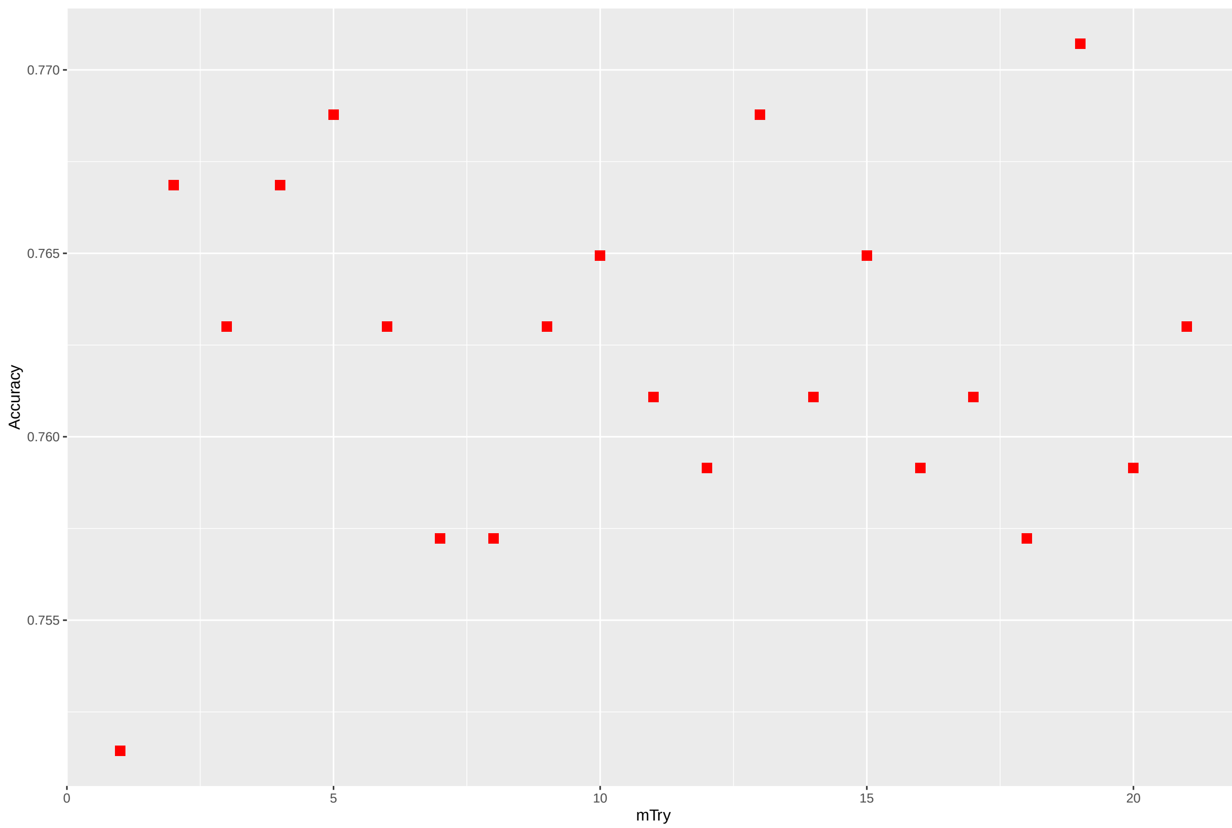
Random Forests (RF) [@Lush2016b;@Graf2015d] was selected as the modeling tool for predicting unobserved behaviours in wild animals based on measurements of observed behaviours in captive animals. Random Forests is a relatively novel and powerful machine learning algorithm that has been reported to work well with complex ecological data that are not easily fitted by traditional methods such as generalized linear models [@Cutler2007]. Random Forests are capable of making accurate predictions from datasets with highly correlated predictor variables and identify the importance measures of each conditional variable. This capability identifies the extent of which a specific predictor variable influences the algorithm's classification accuracy [@Cutler2007]; a higher measure of importance in a predictor variable demonstrates the greater influence it has in comparison to the other predictor variables used in the model [@Cutler2007].

To construct the RF model to predict the seven classes of dog behaviour (Table 2), a series of summary statistics (mean, min, max, kurtosis, skewness, range, standard deviation) for all three axes were applied onto each labelled burst of accelerometer measurements to describe the predictor variables of the random forest model. Subsequently, a range of mtry (i.e., number of predictor variables available for splitting at each tree node; different mTry parameter values can affect model sensitivity and stability) and ntree (i.e., number of decision 'trees'; larger number of trees produce more stable models) parameters were examined to identify the values that have the strongest influence on the predictive capacity of the model. Predictions made by all trees for each observation were then tallied and prediction accuracies were calculated by comparing the predicted and actual classifications. The 'randomForest' package [@Liaw2002] in the R (version 3.0.1, R Core Team 2013) statistical program was used to construct and fit all models, and to derive the variable importance estimates of each used predictor variable.

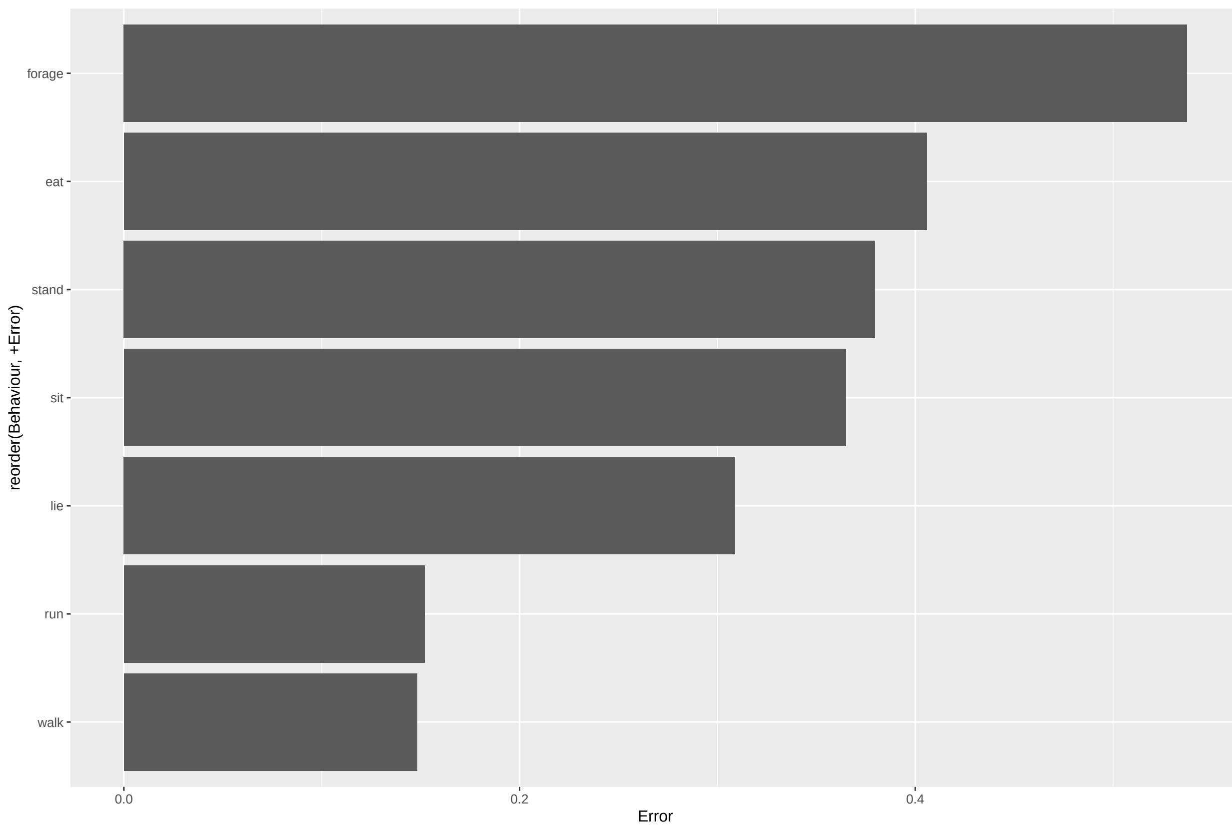
Statistics Label	Predictor Variables	Descriptions
Mean	mean.x mean.y mean.z	The calculated mean of the acceleration measurements within each burst for axes x, y and z.
Min	min.x min.y min.z	The smallest acceleration measurement within each burst for axes x, y and z.
Max	max.x max.y max.z	The largest acceleration measurement within each burst for axes x, y and z.
Kurtosis	kurt.x kurt.y kurt.z	The relative flatness of the acceleration measurements within each burst for axes x, y and z.
Skew	skew.x skew.y skew.z	The relative skewness of the acceleration measurements within each burst for axes x, y and z.
Range	range.x range.y range.z	The calculated difference between the largest and smallest acceleration measurements of axes x, y, and z.
Standard Deviation	sd.x sd.y sd.z	The calculated standard deviations of the acceleration measurements of axes x, y, and z.

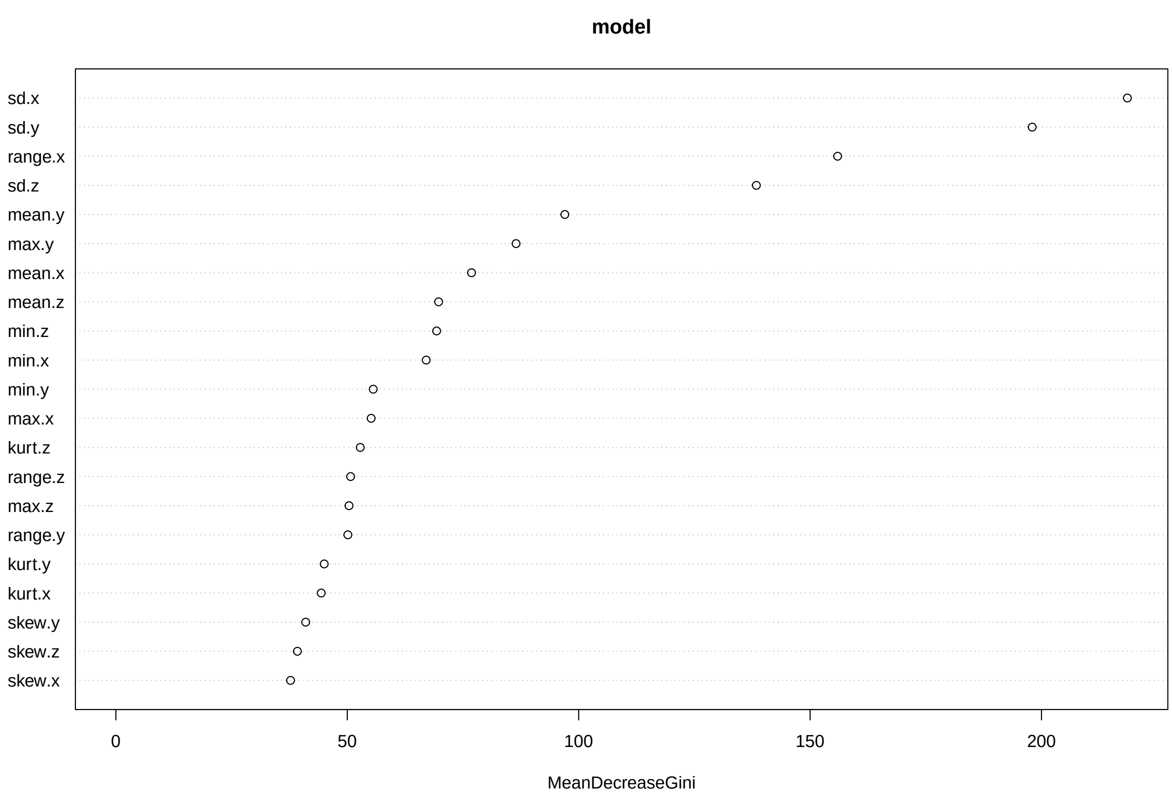
i used rf to find the variance of identified movement behaviours (which becomes a table) and picked an arbitrary value from an axis that best represents motion in a dog. After which, I ground-truth the acc-thr value by incorporating the desired value into a tracking collar and brought four different dogs for a walk. All dogs were walked continuously for approximately ten minutes, and with the studied dog resting (with the collar on) only in the beginning and end of each walk. The dogs were rested in a non-motion movement behaviour (e.g., lying, sitting or standing) to simulate the resting behaviour of free-roaming dogs in the wild. The GPS and ACC dataset were then retrieved from the collar, and the occurence of short- and long-burst gps relocations were plotted sequentially against the variance of the accelerometer measurements collected from all three axes.

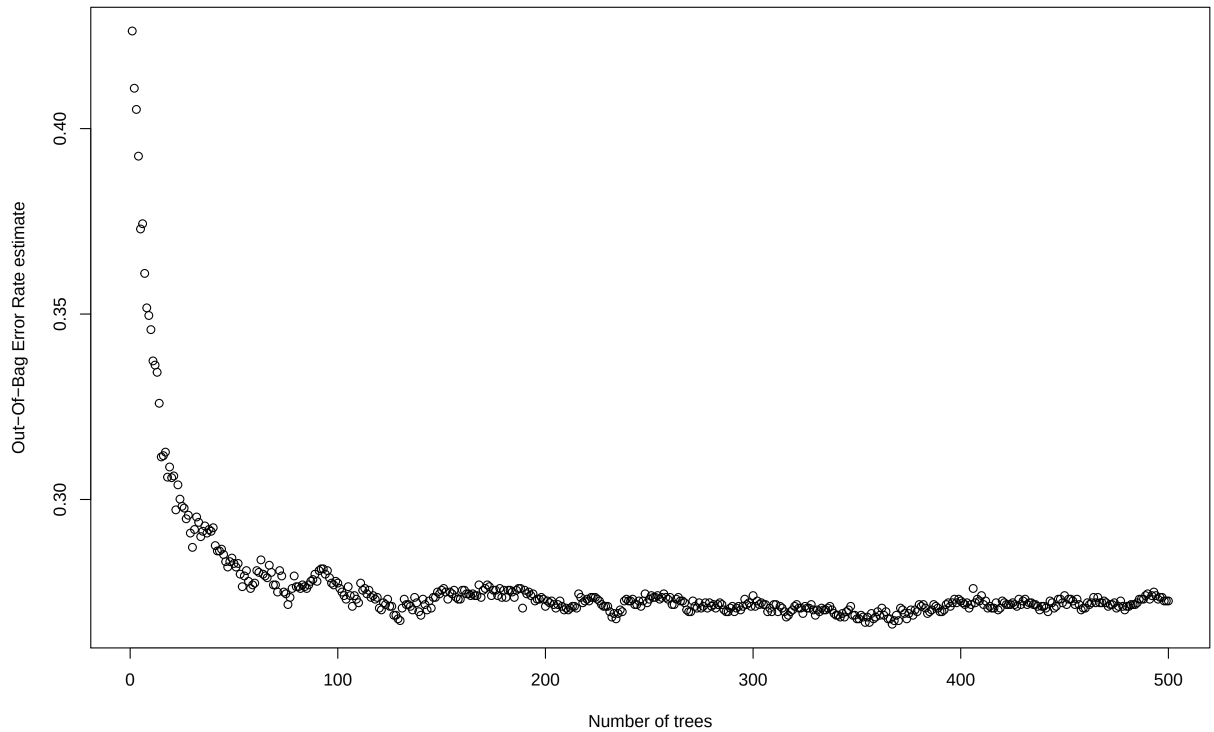
test test



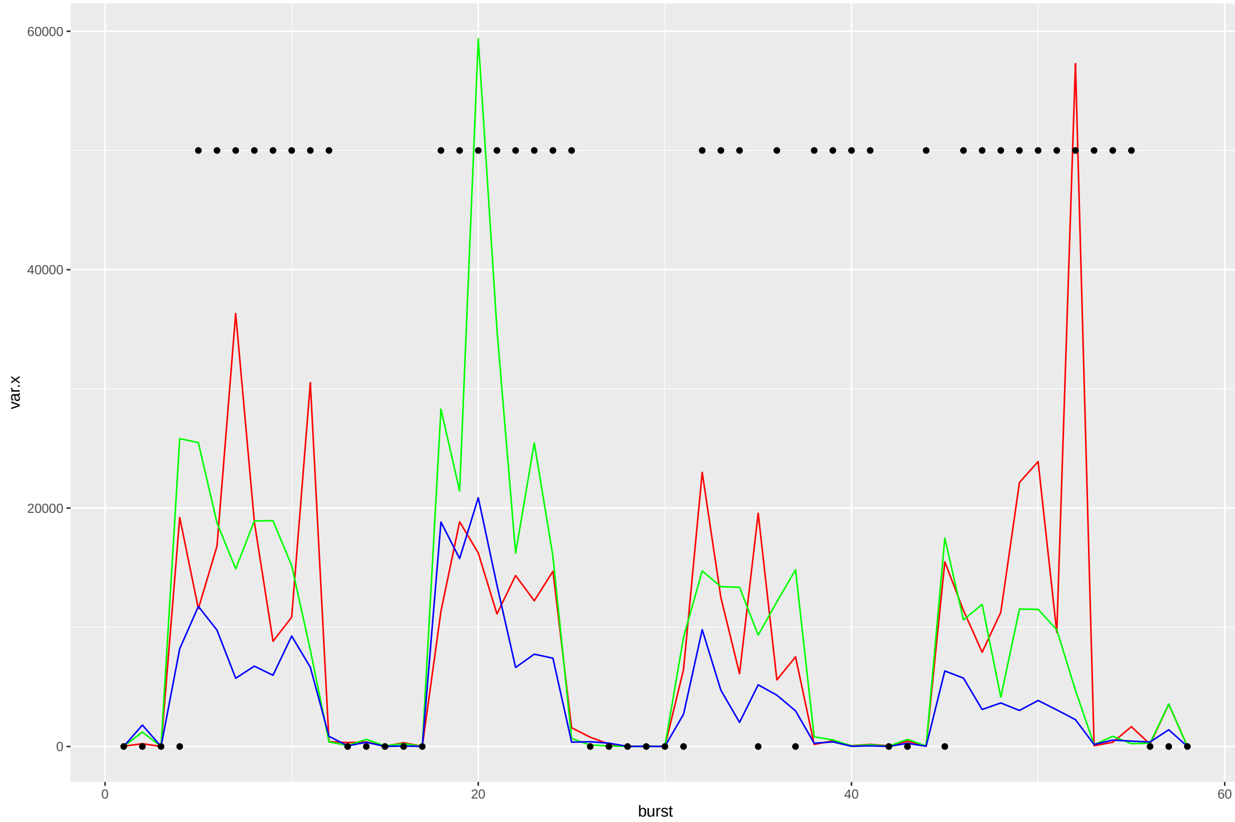

```
##
## ================================================
##         eat forage lie run sit stand walk accuracy
## ------------------------------------------------
## eat     41     7     12   0   3     2    4    59.40%
## forage   1    62     10   5   1     2   53    46.30%
## lie      7     7    293   1  63    26   27    69.10%
## run      0     1      2 346   0     0   59    84.80%
## sit      2     7     85   0 221    17   16    63.50%
## stand    1     6     28   2  14   116   20    62.00%
## walk     0    11      3  49   7     5  431    85.20%
## ------------------------------------------------
```


forage
eat
stand
sit
lie
run
walk
reorder(Behaviour, +Error)
0.0
0.2
0.4
Error


model
sd.x
sd.y
range.x
sd.z
mean.y
max.y
mean.x
mean.z
min.z
min.x
min.y
max.x
kurt.z
range.z
max.z
range.y
kurt.y
kurt.x
skew.y
skew.z
skew.x
0
50
100
150
200
MeanDecreaseGini


Out−Of−Bag Error Rate estimate
0.30
0.35
0.40
0
100
200
300
400
500
Number of trees

```
## [1] 76.30%
```



Results

Discussion

Conclusion

References