Boruta

Kaitlyn Hohmeier

2022-07-11

R Markdown

```
#install.packages("Boruta")
library(Boruta)
```

```
## Warning: package 'Boruta' was built under R version 4.0.5
```

```
library(RColorBrewer)
```

```
## Warning: package 'RColorBrewer' was built under R version 4.0.3
```

```
# 2021 data
data_sampling = read.csv("SRER21_dataset_v1.csv", header = T)
data_sampling = na.omit(data_sampling)

data_sampling$Veg_class = as.factor(data_sampling$Veg_class)

str(data_sampling)
```

```
## 'data.frame':    6805 obs. of  19 variables:
##  $ ï..OID_     : int  2 3 4 5 6 7 8 9 10 11 ...
##  $ Id          : int  2 3 4 5 6 7 8 9 10 11 ...
##  $ gridcode    : int  2 3 4 5 6 7 8 9 10 11 ...
##  $ Shape_Length: num  20.4 8.2 55.8 15.2 27.2 ...
##  $ Shape_Area  : num  2.38 1.32 24.43 1.88 9.29 ...
##  $ CH_mean     : num  1 1 0.92 1 0.143 ...
##  $ ARVI_mean   : num  0.2739 0.5302 0.478 0.3429 0.0444 ...
##  $ ARVI_med    : num  0.2739 0.5302 0.501 0.3429 -0.0089 ...
##  $ ARVI_max    : num  0.274 0.53 0.617 0.343 0.266 ...
##  $ EVI_mean    : num  0.323 0.461 0.381 0.347 0.182 ...
##  $ EVI_med     : num  0.323 0.461 0.385 0.347 0.147 ...
##  $ EVI_max     : num  0.323 0.461 0.447 0.347 0.316 ...
##  $ NDVI_mean   : num  0.428 0.612 0.576 0.469 0.224 ...
##  $ NDVI_med    : num  0.428 0.612 0.597 0.469 0.179 ...
##  $ NDVI_max    : num  0.428 0.612 0.674 0.469 0.412 ...
##  $ SAVI_mean   : num  0.312 0.416 0.36 0.324 0.178 ...
##  $ SAVI_med    : num  0.312 0.416 0.364 0.324 0.147 ...
##  $ SAVI_max    : num  0.312 0.416 0.414 0.324 0.299 ...
```

```
##  $ Veg_class   : Factor w/ 2 levels "non-woody","woody": 2 2 2 2 1 2 2 1 2 2 ...
##  - attr(*, "na.action")= 'omit' Named int [1:962] 21 22 37 55 66 71 78 79 90 91 ...
##   ..- attr(*, "names")= chr [1:962] "21" "22" "37" "55" ...
```

```
table(data_sampling$Veg_class)
```

```
##
## non-woody     woody
##      2230      4575
```

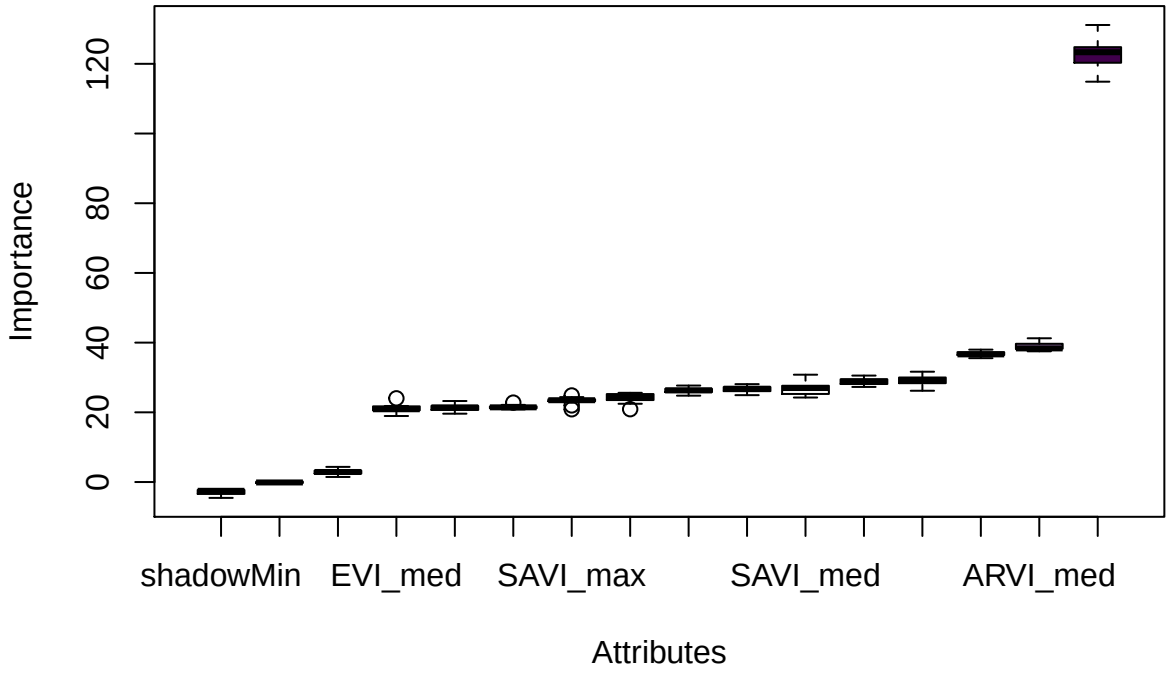
```
# set seed and run the Boruta algorithm
set.seed(200)
traindata = subset(data_sampling, select = -c(ï..OID_,Id,gridcode,Shape_Length,Shape_Area))
boruta.test <- Boruta(Veg_class~., data = traindata, doTrace = 2)
```

```
##  1. run of importance source...

##  2. run of importance source...

##  3. run of importance source...

##  4. run of importance source...

##  5. run of importance source...

##  6. run of importance source...

##  7. run of importance source...

##  8. run of importance source...

##  9. run of importance source...

##  10. run of importance source...

##  11. run of importance source...

## After 11 iterations, +29 secs:

##  confirmed 13 attributes: ARVI_max, ARVI_mean, ARVI_med, CH_mean, EVI_max and 8 more;

##  no more attributes left.
```

```
print(boruta.test)  # print final output of Boruta
```

```
## Boruta performed 11 iterations in 29.38833 secs.
##  13 attributes confirmed important: ARVI_max, ARVI_mean, ARVI_med,
## CH_mean, EVI_max and 8 more;
##  No attributes deemed unimportant.
```

```
plot(boruta.test, col = brewer.pal(11, "PRGn"))  # box plot
```




```
attStats(boruta.test)  # table
```

```
##             meanImp medianImp    minImp    maxImp normHits  decision
## CH_mean   122.96926 123.32399 114.86973 131.13801        1 Confirmed
## ARVI_mean  36.80034  36.62767  35.53964  37.99233        1 Confirmed
## ARVI_med   38.91326  38.34335  37.50898  41.24633        1 Confirmed
## ARVI_max   26.59549  26.78405  24.92877  28.09224        1 Confirmed
## EVI_mean   24.14288  24.43542  20.90736  25.60534        1 Confirmed
## EVI_med    21.12333  20.88274  18.94872  24.02037        1 Confirmed
## EVI_max    21.35716  21.41091  19.61141  23.24885        1 Confirmed
## NDVI_mean  26.26490  26.31163  24.78469  27.68538        1 Confirmed
## NDVI_med   28.84034  28.86728  27.28709  30.54341        1 Confirmed
## NDVI_max   21.52566  21.43104  20.83191  22.78843        1 Confirmed
## SAVI_mean  29.14574  29.26053  26.19459  31.65369        1 Confirmed
## SAVI_med   26.84346  26.97480  24.24477  30.80097        1 Confirmed
## SAVI_max   23.29228  23.47327  20.88388  24.83504        1 Confirmed
```

```
getConfirmedFormula(boruta.test)  # model
```

```
## Veg_class ~ CH_mean + ARVI_mean + ARVI_med + ARVI_max + EVI_mean +
##     EVI_med + EVI_max + NDVI_mean + NDVI_med + NDVI_max + SAVI_mean +
##     SAVI_med + SAVI_max
## <environment: 0x0000000020523268>
```

```r
# 2017 data
data_sampling17 = read.csv("SRER_2017_training_bi.csv", header = T)
data_sampling17 = na.omit(data_sampling17)


data_sampling17$Veg_class = as.factor(data_sampling17$Veg_class)

str(data_sampling17)
```

```
## 'data.frame':    4339 obs. of  19 variables:
##  $ OID_        : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Id          : int  2 3 5 6 7 8 9 10 11 12 ...
##  $ gridcode    : int  2 3 5 6 7 8 9 10 11 12 ...
##  $ Shape_Length: num  41.6 31.4 33.4 32.2 28.8 ...
##  $ Shape_Area  : num  17.41 5.05 4.84 14.2 10.73 ...
##  $ CH_mean     : num  0.1907 0.045 0.01 0.0117 0.0138 ...
##  $ ARVI_max    : num  0.6091 0.3878 0.0388 0.1898 0.3227 ...
##  $ ARVI_mean   : num  0.34418 0.2453 0.00404 0.03342 0.08032 ...
##  $ ARVI_med    : num  0.3794 0.2749 0.0268 0.019 0.0615 ...
##  $ EVI_max     : num  0.443 0.332 0.151 0.199 0.248 ...
##  $ EVI_mean    : num  0.292 0.253 0.132 0.142 0.159 ...
##  $ EVI_med     : num  0.291 0.244 0.135 0.138 0.142 ...
##  $ NDVI_max    : num  0.666 0.526 0.208 0.341 0.457 ...
##  $ NDVI_mean   : num  0.477 0.407 0.182 0.207 0.247 ...
##  $ NDVI_med    : num  0.508 0.434 0.205 0.197 0.228 ...
##  $ SAVI_max    : num  0.407 0.329 0.15 0.197 0.249 ...
##  $ SAVI_mean   : num  0.286 0.255 0.132 0.142 0.159 ...
##  $ SAVI_med    : num  0.288 0.248 0.135 0.139 0.143 ...
##  $ Veg_class   : Factor w/ 2 levels "non-woody","woody": 2 1 1 1 1 2 1 1 2 2 ...
```

```r
table(data_sampling17$Veg_class)
```

```
##
## non-woody     woody
##      2088      2251
```

```r
# set seed and run the Boruta algorithm
set.seed(200)
traindata17 = subset(data_sampling17, select = -c(OID_,Id,gridcode,Shape_Length,Shape_Area))
boruta.test17 <- Boruta(Veg_class~., data = traindata17, doTrace = 2)
```

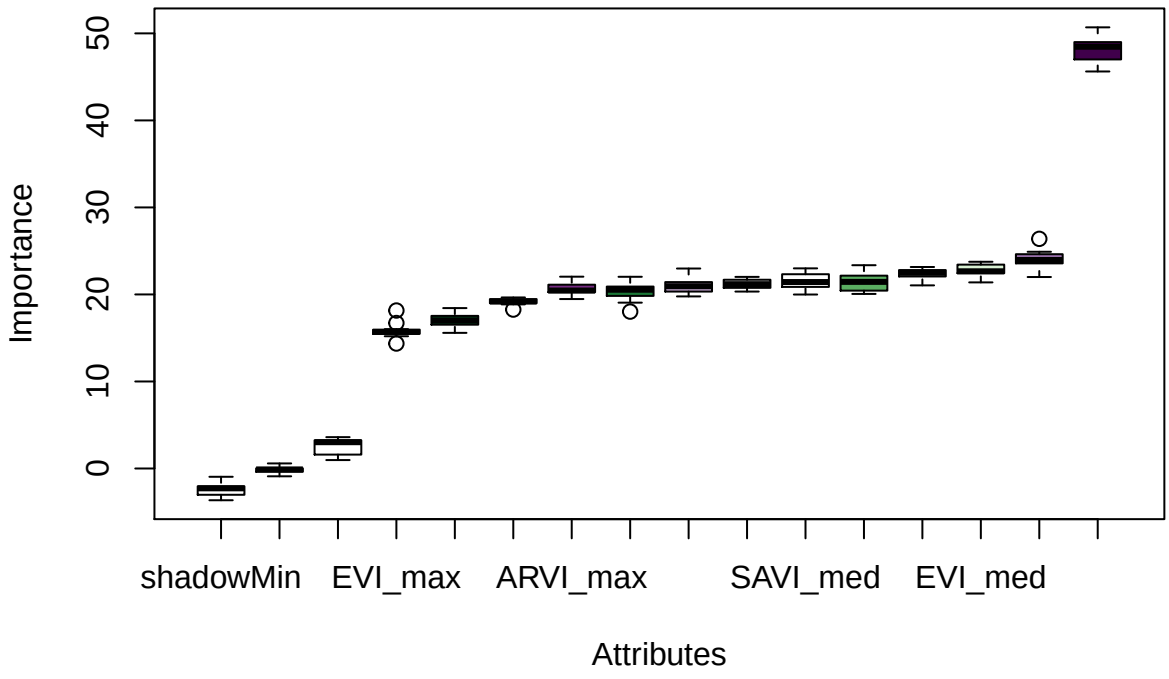
```
##  1. run of importance source...

##  2. run of importance source...

##  3. run of importance source...

##  4. run of importance source...

##  5. run of importance source...
```

```
##  6. run of importance source...

##  7. run of importance source...

##  8. run of importance source...

##  9. run of importance source...

##  10. run of importance source...

##  11. run of importance source...

## After 11 iterations, +24 secs:

##  confirmed 13 attributes: ARVI_max, ARVI_mean, ARVI_med, CH_mean, EVI_max and 8 more;

##  no more attributes left.
```

```
print(boruta.test17)  # print final output of Boruta
```

```
## Boruta performed 11 iterations in 23.67566 secs.
##  13 attributes confirmed important: ARVI_max, ARVI_mean, ARVI_med,
## CH_mean, EVI_max and 8 more;
##  No attributes deemed unimportant.
```

```
plot(boruta.test17, col = brewer.pal(11, "PRGn"))  # box plot
```



```
attStats(boruta.test17)  # table
```

```
##            meanImp medianImp   minImp   maxImp normHits  decision
## CH_mean   48.17478  48.46436 45.62041 50.69171        1 Confirmed
## ARVI_max  20.61580  20.44995 19.47104 22.04327        1 Confirmed
## ARVI_mean 24.09615  23.96154 22.00115 26.39015        1 Confirmed
## ARVI_med  21.00214  20.94443 19.77373 22.97900        1 Confirmed
## EVI_max   15.84628  15.68850 14.35863 18.15406        1 Confirmed
## EVI_mean  21.21644  21.18202 20.33218 22.01612        1 Confirmed
## EVI_med   22.73651  22.65984 21.38691 23.75615        1 Confirmed
## NDVI_max  19.16496  19.18637 18.26038 19.65658        1 Confirmed
## NDVI_mean 21.43923  21.44484 20.05836 23.36056        1 Confirmed
## NDVI_med  20.28855  20.54305 18.02811 22.02879        1 Confirmed
## SAVI_max  17.06834  17.00413 15.58676 18.43359        1 Confirmed
## SAVI_mean 22.35667  22.48347 21.04472 23.15362        1 Confirmed
## SAVI_med  21.50312  21.42756 19.99077 22.99684        1 Confirmed
```

```
getConfirmedFormula(boruta.test17)  # model
```

```
## Veg_class ~ CH_mean + ARVI_max + ARVI_mean + ARVI_med + EVI_max +
##     EVI_mean + EVI_med + NDVI_max + NDVI_mean + NDVI_med + SAVI_max +
##     SAVI_mean + SAVI_med
## <environment: 0x00000000201332c8>
```