Package 'steprf'

October 14, 2022

Title Stepwise Predictive Variable Selection for Random Forest

Version 1.0.2

Date 2022-6-28

Description An introduction to several novel predictive variable selection methods for random forest. They are based on various variable importance methods (i.e., averaged variable importance (AVI), and knowledge informed AVI (i.e., KIAVI, and KIAVI2)) and predictive accuracy in stepwise algorithms. For details of the variable selection methods, please see: Li, J., Siwabessy, J., Huang, Z. and Nichol, S. (2019) <doi:10.3390/geosciences9040180>. Li, J., Alvarez, B., Siwabessy, J., Tran, M., Huang, Z., Przeslawski, R., Radke, L., Howard, F., Nichol, S. (2017). <DOI:10.13140/RG.2.2.27686.22085>.

Depends R (>= 4.0)

Imports spm, randomForest, spm2, psy

License GPL (>= 2)

RoxygenNote 7.1.1

Encoding UTF-8

Suggests knitr, rmarkdown, lattice, reshape2

NeedsCompilation no

Author Jin Li [aut, cre]

Maintainer Jin Li <jinli68@gmail.com>

Repository CRAN

Date/Publication 2022-06-29 11:20:02 UTC

R topics documented:

cran-comments *Note on notes*

Description

This is an updated and extended version of ‘spm‘ package. The change in package name from ‘spm‘ to ‘spm2‘ is due to the change in Author’s support from Geoscience Australia to Data2Action Australia.

R CMD check results 0 errors | 0 warnings | 0 notes

Author(s)

Jin Li

RFcv2 *Cross validation, n-fold for random forest (RF)*

Description

This function is a cross validation function for random forest. It is for functions ’steprf’, ’steprfAVI’, ect.

Usage

```
RFcv2(
  trainx,
  trainy,
  cv.fold = 10,
 mtry = if (!is.null(trainy) && !is.factor(trainy)) max(floor(ncol(trainx)/3), 1) else
    floor(sqrt(ncol(trainx))),
  ntree = 500,
  predacc = "VEcv",
  ...
)
```

Arguments

trainx	a dataframe or matrix contains columns of predictor variables.
trainy	a vector of response, must have length equal to the number of rows in trainx.
cv.fold	integer; number of folds in the cross-validation. if > 1, then apply n-fold cross validation; the default is 10, i.e., 10-fold cross validation that is recommended.
mtry	a function of number of remaining predictor variables to use as the mtry parameter in the randomForest call.

ntree	number of trees to grow. This should not be set to too small a number, to ensure that every input row gets predicted at least a few times. By default, 500 is used.
predacc	"VEcv" for vecv for numerical data, or "ccr" (i.e., correct classification rate) or "kappa" for categorical data.
...	other arguments passed on to randomForest.

Value

A list with the following component: vecv for numerical data: ; or ccr (correct classification rate) for categorical data: .

Note

This function is largely based on rf.cv (see Li et al. 2013) and rfcv in randomForest.

Author(s)

Jin Li

References

Li, J., J. Siwabessy, M. Tran, Z. Huang, and A. Heap. 2013. Predicting Seabed Hardness Using Random Forest in R. Pages 299-329 in Y. Zhao and Y. Cen, editors. Data Mining Applications with R. Elsevier.

Li, J. 2013. Predicting the spatial distribution of seabed gravel content using random forest, spatial interpolation methods and their hybrid methods. Pages 394-400 The International Congress on Modelling and Simulation (MODSIM) 2013, Adelaide.

Liaw, A. and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18-22.

Examples

```
library(spm)
data(hard)
data(petrel)

rfcv1 <- RFcv2(petrel[, c(1,2, 6:9)], petrel[, 5], predacc = "VEcv")
rfcv1

rfcv2 <- RFcv2(hard[, -c(1, 17)], hard[, 17], predacc = "ccr")
rfcv2

rfcv3 <- RFcv2(hard[, -c(1, 17)], hard[, 17], predacc = "kappa")
rfcv3

n <- 10 # number of iterations, 60 to 100 is recommended.
VEcv <- NULL
for (i in 1:n) {
rfcv1 <- RFcv2(petrel[, c(1,2,6:9)], petrel[, 5], predacc = "VEcv")
```

```
VEcv [i] <- rfcv1
}
plot(VEcv ~ c(1:n), xlab = "Iteration for RF", ylab = "VEcv (%)")
points(cumsum(VEcv) / c(1:n) ~ c(1:n), col = 2)
abline(h = mean(VEcv), col = 'blue', lwd = 2)

n <- 10 # number of iterations, 60 to 100 is recommended.
measures <- NULL
for (i in 1:n) {
rfcv1 <- RFcv2(hard[, c(4:6)], hard[, 17], predacc = "ccr")
measures <- rbind(measures, rfcv1)
}
plot(measures ~ c(1:n), xlab = "Iteration for RF", ylab = "Correct
classification Rate  (%)")
points(cumsum(measures) / c(1:n) ~ c(1:n), col = 2)
abline(h = mean(measures), col = 'blue', lwd = 2)
```

`steprf` *Select predictive variables for random forest by various variable importance methods and predictive accuracy in a stepwise algorithm*

Description

This function is to select predictive variables for random forest by various variable importance methods (i.e., AVI, Knowledge informed AVI (KIAVI), KIAVI2) and predictive accuracy. It is implemented via the functions 'steprfAVI' and 'steprfAVIPredictors'.

Usage

```
steprf(
  trainx,
  trainy,
  method = "KIAVI",
  cv.fold = 10,
  ntree = 500,
  rpt = 20,
  predacc = "VEcv",
  importance = TRUE,
  maxk = c(4),
  nsim = 100,
  delta.predacc = 0.001,
  min.n.var = 2,
  corr.threshold = 0.5,
  ...
)
```

Arguments

trainx	a dataframe or matrix contains columns of predictor variables.
trainy	a vector of response, must have length equal to the number of rows in trainx.
method	a variable selection method for 'RF'; can be: "AVI", "KIAVI" and "KIAVI2". If "AVI" is used, it would produce tha same results as 'steprfAVI'. By default, "KIAVI" is used.
cv.fold	integer; number of folds in the cross-validation. if > 1, then apply n-fold cross validation; the default is 10, i.e., 10-fold cross validation that is recommended.
ntree	number of trees to grow. This should not be set to too small a number, to ensure that every input row gets predicted at least a few times. By default, 500 is used.
rpt	iteration of cross validation.
predacc	"VEcv" for vecv for numerical data, or "ccr" (i.e., correct classification rate) or "kappa" for categorical data.
importance	imprtance of predictive variables.
maxk	maxk split value. By default, 4 is used.
nsim	iteration number. By default, 100 is used.
delta.predacc	minimum changes between the accuracies of two consecutive predictive models.
min.n.var	minimum number of predictive variables remained in the final predictive model the default is 2. If 1 is used, then warnings: 'invalid mtry: reset to within valid range' will be issued, which should be ignored.
corr.threshold	correlation threshold and the defaults value is 0.5.
...	other arguments passed on to randomForest.

Value

A list with the following components: 1) steprfPredictorsFinal: the variables selected for the last RF model, whether it is of the highest predictive accuracy need to be confirmed using 'max.predictive.accuracy' that is listed next; 2) max.predictive.accuracy: the predictive accuracy of the most accurate RF model for each run of 'steprfAVI', which can be used to confirm the model with the highest accuracy, 3) numberruns: number of runs of 'steprfAVI'; 4) laststepAVI: the outpouts of last run of 'steprfAVI'; 5) steprfAVIOutputsAll: the outpouts of all 'steprfAVI' produced during the variable selection process; 6) steprfPredictorsAll: the outpouts of 'steprfAVIPredictors' for all 'steprfAVI' produced during the variable selection process; 7) KIAVIPredictorsAll: predictors used for all 'steprfAVI' produced during the variable selection process; for a method "AVI", if the variables are different from those in the traning dataset, it suggests that these variables should be tested if the predictive accuracy can be further improved.

Note

In 'steprf', 'steprfAVI' is used instead of 'steprfAVI1' and 'steprfAVI2'. This is because: 1) 'avi' is expected to change with the removal of each predictor, but in 'steprfAVI1' the averaged variable importance is calculated only once and is from the full model only, so its use is expected to produce a less optimal model, hence not used; and 2) the 'steprf' would lead to the same set of predictors as that for 'steprfAVI2' if 'steprfAVI2' is used, so it is not used either.

Author(s)

Jin Li

References

Li, J. (2022). Spatial Predictive Modeling with R. Boca Raton, Chapman and Hall/CRC.

Li, J. (2019). "A critical review of spatial predictive modeling process in environmental sciences with reproducible examples in R." Applied Sciences 9: 2048.

Li, J. 2013. Predicting the spatial distribution of seabed gravel content using random forest, spatial interpolation methods and their hybrid methods. Pages 394-400 The International Congress on Modelling and Simulation (MODSIM) 2013, Adelaide.

Li, J., Alvarez, B., Siwabessy, J., Tran, M., Huang, Z., Przeslawski, R., Radke, L., Howard, F. and Nichol, S. (2017). "Application of random forest, generalised linear model and their hybrid methods with geostatistical techniques to count data: Predicting sponge species richness." Environmental Modelling & Software 97: 112-129.

Li, J., Siwabessy, J., Huang, Z., and Nichol, S. (2019). "Developing an optimal spatial predictive model for seabed sand content using machine learning, geostatistics and their hybrid methods." Geosciences 9 (4):180.

Li, J., Siwabessy, J., Tran, M., Huang, Z. and Heap, A., 2014. Predicting Seabed Hardness Using Random Forest in R. Data Mining Applications with R. Y. Zhao and Y. Cen. Amsterdam, Elsevier: 299-329.

Li, J., Tran, M. and Siwabessy, J., 2016. Selecting optimal random forest predictive models: a case study on predicting the spatial distribution of seabed hardness. PLOS ONE 11(2): e0149089.

Liaw, A. and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18-22.

Smith, S.J., Ellis, N., Pitcher, C.R., 2011. Conditional variable importance in R package extendedForest. R vignette [http://gradientforest.r-forge.r-project.org/Conditional-importance.pdf].

Examples

```
library(spm)
data(petrel)
set.seed(1234)
steprf1 <- steprf(trainx = petrel[, c(1,2, 6:9)], trainy =
petrel[, 5], method = "KIAVI", rpt = 2, predacc = "VEcv", importance = TRUE,
 nsim = 3, delta.predacc = 0.01)
names(steprf1)
steprf1$steprfPredictorsFinal$variables.most.accurate
steprf1$max.predictive.accuracy
```

`steprfAVI` *Select predictive variables for random forest by AVI and accuracy in a stepwise algorithm*

Description

This function is to select predictive variables for random forest by their averaged variable importance (AVI) that is calculated for each model after excluding the least important variable, and returns the corresponding predictive accuracy. It is also developed for 'steprf' function.

Usage

```
steprfAVI(
  trainx,
  trainy,
  cv.fold = 10,
  ntree = 500,
  rpt = 20,
  predacc = "VEcv",
  importance = TRUE,
  maxk = c(4),
  nsim = 100,
  min.n.var = 2,
  corr.threshold = 0.5,
  ...
)
```

Arguments

`trainx`	a dataframe or matrix contains columns of predictor variables.
`trainy`	a vector of response, must have length equal to the number of rows in trainx.
`cv.fold`	integer; number of folds in the cross-validation. if > 1, then apply n-fold cross validation; the default is 10, i.e., 10-fold cross validation that is recommended.
`ntree`	number of trees to grow. This should not be set to too small a number, to ensure that every input row gets predicted at least a few times. By default, 500 is used.
`rpt`	iteration number of cross validation.
`predacc`	"VEcv" for vecv for numerical data, or "ccr" (i.e., correct classification rate) or "kappa" for categorical data.
`importance`	importance of predictive variables.
`maxk`	maxk split value. By default, 4 is used.
`nsim`	iteration number for 'avi'. By default, 100 is used.
`min.n.var`	minimum number of predictive variables remained in the final predictive model the default is 2. If 1 is used, then warnings: 'invalid mtry: reset to within valid range' will be issued, which should be ignored.
`corr.threshold`	correlation threshold and the defaults value is 0.5.
`...`	other arguments passed on to randomForest.

Value

A list with the following components: 1) variable.removed: variable removed based on AVI, 2) predictive.accuracy: averaged predictive accuracy of the model after excluding the variable.removed, 3) delta.accuracy: contribution to accuracy by each variable.removed, and 4) predictive.accuracy2: predictive accuracy matrix of the model after excluding the variable.removed for each iteration.

Author(s)

Jin Li

References

Li, J. (2022). Spatial Predictive Modeling with R. Boca Raton, Chapman and Hall/CRC.

Li, J. 2013. Predicting the spatial distribution of seabed gravel content using random forest, spatial interpolation methods and their hybrid methods. Pages 394-400 The International Congress on Modelling and Simulation (MODSIM) 2013, Adelaide.

Li, J., Alvarez, B., Siwabessy, J., Tran, M., Huang, Z., Przeslawski, R., Radke, L., Howard, F. and Nichol, S. (2017). "Application of random forest, generalised linear model and their hybrid methods with geostatistical techniques to count data: Predicting sponge species richness." Environmental Modelling & Software 97: 112-129.

Li, J., Siwabessy, J., Huang, Z., Nichol, S. (2019). "Developing an optimal spatial predictive model for seabed sand content using machine learning, geostatistics and their hybrid methods." Geosciences 9 (4):180.

Li, J., Siwabessy, J., Tran, M., Huang, Z. and Heap, A., 2014. Predicting Seabed Hardness Using Random Forest in R. Data Mining Applications with R. Y. Zhao and Y. Cen. Amsterdam, Elsevier: 299-329.

Liaw, A. and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18-22.

Smith, S.J., Ellis, N., Pitcher, C.R., 2011. Conditional variable importance in R package extendedForest. R vignette [http://gradientforest.r-forge.r-project.org/Conditional-importance.pdf].

Examples

```
library(spm)

data(petrel)
set.seed(1234)
steprf1 <- steprfAVI(trainx = petrel[, c(1,2, 6:9)], trainy = petrel[, 5],
 rpt = 2, predacc = "VEcv", importance = TRUE, nsim = 3, min.n.var = 2)
steprf1

steprf2 <- steprfAVI(trainx = hard[, -c(1, 17)], trainy = hard[, 17],
 rpt = 2, predacc = "ccr", importance = TRUE, nsim = 3, min.n.var = 2)
steprf2

#plot steprf1 results
library(reshape2)
```

```
pa1 <- as.data.frame(steprf1$predictive.accuracy2)
names(pa1) <- steprf1$variable.removed
pa2 <- melt(pa1, id = NULL)
names(pa2) <- c("Variable","VEcv")
library(lattice)
par (font.axis=2, font.lab=2)
with(pa2, boxplot(VEcv~Variable, ylab="VEcv (%)", xlab="Predictive variable removed"))

barplot(steprf1$delta.accuracy, col = (1:length(steprf1$variable.removed)),
names.arg = steprf1$variable.removed, main = "Predictive accuracy vs variable removed",
font.main = 4, cex.names=1, font=2, ylab="Increase in VEcv (%)")
```

`steprfAVI1` *Select predictive variables for random forest by avi and accuracy in a stepwise algorithm*

Description

This function is to select predictive variables for random forest by their averaged variable importance which is derived from the full model and returns the corresponding predictive accuracy. That is, in comparison with 'steprfAVI', the averaged variable importance is calculated only once and is from the full model only.

Usage

```
steprfAVI1(
  trainx,
  trainy,
  cv.fold = 10,
 mtry = if (!is.null(trainy) && !is.factor(trainy)) max(floor(ncol(trainx)/3), 1) else
    floor(sqrt(ncol(trainx))),
  ntree = 500,
  rpt = 2,
  predacc = "VEcv",
  importance = TRUE,
  maxk = c(4),
  nsim = 100,
  min.n.var = 2,
  corr.threshold = 0.5,
  ...
)
```

Arguments

`trainx`	a dataframe or matrix contains columns of predictor variables.
`trainy`	a vector of response, must have length equal to the number of rows in trainx.

cv.fold	integer; number of folds in the cross-validation. if > 1, then apply n-fold cross validation; the default is 10, i.e., 10-fold cross validation that is recommended.
mtry	a function of number of remaining predictor variables to use as the mtry parameter in the randomForest call.
ntree	number of trees to grow. This should not be set to too small a number, to ensure that every input row gets predicted at least a few times. By default, 500 is used.
rpt	iteration of cross validation.
predacc	"VEcv" for vecv for numerical data, or "ccr" (i.e., correct classification rate) or "kappa" for categorical data.
importance	imprtance of predictive variables.
maxk	maxk split value. By default, 4 is used.
nsim	iteration number. By default, 100 is used.
min.n.var	minimum number of predictive variables remained in the final predictive model the default is 1.
corr.threshold	correlation threshold and the defaults value is 0.5.
...	other arguments passed on to randomForest.

Value

A list with the following components: variable removed based on avi (variable.removed), averaged predictive accuracy of the model after excluding variable.removed (predictive.accuracy), contribution to accuracy by each variable.removed (delta.accuracy), and predictive accuracy matrix of the model after excluding variable.removed for each iteration (predictive.accuracy2)

Author(s)

Jin Li

References

Li, J. (2022). Spatial Predictive Modeling with R. Boca Raton, Chapman and Hall/CRC.

Li, J. 2013. Predicting the spatial distribution of seabed gravel content using random forest, spatial interpolation methods and their hybrid methods. Pages 394-400 The International Congress on Modelling and Simulation (MODSIM) 2013, Adelaide.

Li, J. (2019). "A critical review of spatial predictive modeling process in environmental sciences with reproducible examples in R." Applied Sciences 9: 2048.

Li, J., Siwabessy, J., Huang, Z., Nichol, S. (2019). "Developing an optimal spatial predictive model for seabed sand content using machine learning, geostatistics and their hybrid methods." Geosciences 9 (4):180.

Li, J., Siwabessy, J., Tran, M., Huang, Z. and Heap, A., 2014. Predicting Seabed Hardness Using Random Forest in R. Data Mining Applications with R. Y. Zhao and Y. Cen. Amsterdam, Elsevier: 299-329.

Li, J., Tran, M. and Siwabessy, J., 2016. Selecting optimal random forest predictive models: a case study on predicting the spatial distribution of seabed hardness. PLOS ONE 11(2): e0149089.

Liaw, A. and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18-22.

Smith, S.J., Ellis, N., Pitcher, C.R., 2011. Conditional variable importance in R package extendedForest. R vignette [http://gradientforest.r-forge.r-project.org/Conditional-importance.pdf].

Examples

```
library(spm)
data(petrel)
set.seed(1234)
steprfAVI1.1 <- steprfAVI1(trainx = petrel[, c(1,2, 6:9)], trainy =
petrel[, 5], predacc = "VEcv", nsim = 10)
steprfAVI1.1

#plot steprf1 results
library(reshape2)
pa1 <- as.data.frame(steprfAVI1.1$predictive.accuracy2)
names(pa1) <- steprfAVI1.1$variable.removed
pa2 <- melt(pa1, id = NULL)
names(pa2) <- c("Variable","VEcv")
library(lattice)
par (font.axis=2, font.lab=2)
with(pa2, boxplot(VEcv~Variable, ylab="VEcv (%)", xlab="Predictive variable removed"))

barplot(steprfAVI1.1$delta.accuracy, col = (1:length(steprfAVI1.1$variable.removed)),
names.arg = steprfAVI1.1$variable.removed, main = "Predictive accuracy vs variable removed",
font.main = 4, cex.names=1, font=2, ylab="Increase rate in VEcv (%)")

barplot(steprfAVI1.1$delta.accuracy, col = (1:length(steprfAVI1.1$variable.removed)),
names.arg = steprfAVI1.1$variable.removed,  main = "Predictive accuracy vs variable removed",
font.main = 4, cex.names=1, font=2, ylab="Increase in VEcv (%)")
```

`steprfAVI2` *Select predictive variables for random forest by AVI and accuracy in a stepwise algorithm*

Description

This function is similar to 'steprfAVI'; the only difference is that 'set.seed()' is added before each code line that involves randomness and such additions alter the results considerably.

Usage

```
steprfAVI2(
  trainx,
  trainy,
```

```
  cv.fold = 10,
  ntree = 500,
  rpt = 20,
  predacc = "VEcv",
  importance = TRUE,
  maxk = c(4),
  nsim = 100,
  min.n.var = 2,
  corr.threshold = 0.5,
  rseed = 1234,
  ...
)
```

Arguments

trainx	a dataframe or matrix contains columns of predictor variables.
trainy	a vector of response, must have length equal to the number of rows in trainx.
cv.fold	integer; number of folds in the cross-validation. if > 1, then apply n-fold cross validation; the default is 10, i.e., 10-fold cross validation that is recommended.
ntree	number of trees to grow. This should not be set to too small a number, to ensure that every input row gets predicted at least a few times. By default, 500 is used.
rpt	iteration of cross validation.
predacc	"VEcv" for vecv for numerical data, or "ccr" (i.e., correct classification rate) or "kappa" for categorical data.
importance	imprtance of predictive variables.
maxk	maxk split value. By default, 4 is used.
nsim	iteration number. By default, 100 is used.
min.n.var	minimum number of predictive variables remained in the final predictive model the default is 2. If 1 is used, then warnings: 'invalid mtry: reset to within valid range' will be issued, which should be ignored.
corr.threshold	correlation threshold and the defaults value is 0.5.
rseed	random seed. By default, 1234 is used.
...	other arguments passed on to randomForest.

Value

A list with the following components: 1) variable.removed: variable removed based on AVI, 2) predictive.accuracy: averaged predictive accuracy of the model after excluding the variable.removed, 3) delta.accuracy: contribution to accuracy by each variable.removed, and 4) predictive.accuracy2: predictive accuracy matrix of the model after excluding the variable.removed for each iteration.

Author(s)

Jin Li

References

Li, J. (2022). Spatial Predictive Modeling with R. Boca Raton, Chapman and Hall/CRC.

Li, J. 2013. Predicting the spatial distribution of seabed gravel content using random forest, spatial interpolation methods and their hybrid methods. Pages 394-400 The International Congress on Modelling and Simulation (MODSIM) 2013, Adelaide.

Li, J., Alvarez, B., Siwabessy, J., Tran, M., Huang, Z., Przeslawski, R., Radke, L., Howard, F. and Nichol, S. (2017). "Application of random forest, generalised linear model and their hybrid methods with geostatistical techniques to count data: Predicting sponge species richness." Environmental Modelling & Software 97: 112-129.

Li, J., Siwabessy, J., Huang, Z., Nichol, S. (2019). "Developing an optimal spatial predictive model for seabed sand content using machine learning, geostatistics and their hybrid methods." Geosciences 9 (4):180.

Li, J., Siwabessy, J., Tran, M., Huang, Z. and Heap, A., 2014. Predicting Seabed Hardness Using Random Forest in R. Data Mining Applications with R. Y. Zhao and Y. Cen. Amsterdam, Elsevier: 299-329.

Liaw, A. and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18-22.

Smith, S.J., Ellis, N., Pitcher, C.R., 2011. Conditional variable importance in R package extendedForest. R vignette [http://gradientforest.r-forge.r-project.org/Conditional-importance.pdf].

Chang, W. 2021. Cookbook for R. http://www.cookbook-r.com/.

Examples

```
library(spm)
data(petrel)
steprf1 <- steprfAVI2(trainx = petrel[, c(1,2, 6:9)], trainy = petrel[, 5],
 rpt = 2, predacc = "VEcv", importance = TRUE, nsim = 3, min.n.var = 2)
steprf1

#plot steprf1 results
library(reshape2)
pa1 <- as.data.frame(steprf1$predictive.accuracy2)
names(pa1) <- steprf1$variable.removed
pa2 <- melt(pa1, id = NULL)
names(pa2) <- c("Variable","VEcv")
library(lattice)
par (font.axis=2, font.lab=2)
with(pa2, boxplot(VEcv~Variable, ylab="VEcv (%)", xlab="Predictive variable removed"))

barplot(steprf1$delta.accuracy, col = (1:length(steprf1$variable.removed)),
names.arg = steprf1$variable.removed, main = "Predictive accuracy vs variable removed",
font.main = 4, cex.names=1, font=2, ylab="Increase in VEcv (%)")
```

`steprfAVIPredictors` *Extract names of the selected predictive variables by steprf*

Description

This function is to extract names of the selected predictive variables by steprfAVI.

Usage

```
steprfAVIPredictors(steprf1, trainx)
```

Arguments

`steprf1`	a list of output of 'steprf' function.
`trainx`	a dataframe or matrix contains columns of predictor variables.

Value

A list with the following components: 1) variables.most.accurate: a list of predictive variables contained in the most accurate RF model, 2) PABV: a list of predictive variables with positive contributions to the predictive accuracy of RF models, that is, predictive accuracy boosting variable (PABV), 3) PARV: a list of predictive variables with negative contributions to the predictive accuracy of RF models, that is, predictive accuracy reducing variable, and 4) max.predictive.accuracy: the predictive accuracy of the most accurate RF model.

Author(s)

Jin Li

References

Li, J. (2022). Spatial Predictive Modeling with R. Boca Raton, Chapman and Hall/CRC.

Li, J. (2019). "A critical review of spatial predictive modeling process in environmental sciences with reproducible examples in R." Applied Sciences 9: 2048.

Li, J., Siwabessy, J., Huang, Z., and Nichol, S. (2019). "Developing an optimal spatial predictive model for seabed sand content using machine learning, geostatistics and their hybrid methods." Geosciences 9 (4):180.

Examples

```
library(spm)
data(petrel)
set.seed(1234)
steprf1 <- steprfAVI(trainx = petrel[, c(1,2, 6:9)], trainy = petrel[, 5],
 nsim = 10, min.n.var = 2)
```

```
steprfAVIPredictors(steprf1, trainx = petrel[, c(1,2, 6:9)])
```

Index