

Package ‘FrechForest’

December 4, 2019

Title Frechet Random Forests

Version 0.8.1

Description Random forests are a statistical learning method widely used in many areas of scientific research essentially for its ability to learn complex relationships between input and output variables and also its capacity to handle high-dimensional data. However, current random forests approaches are not flexible enough to handle longitudinal heterogeneous data. In this package, we introduce Fréchet trees and Fréchet random forests, which allow to manage data for which input and output variables are curves. To this end, a new way of splitting the nodes of trees is introduced and the prediction procedures of trees and forests are generalized.

Depends R (>= 3.6.0)

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

Imports stats, kmlShape, foreach, doRNG, doParallel, graphics

Suggests testthat

NeedsCompilation no

Author Louis Capitaine [aut, cre]

Maintainer Louis Capitaine <louis.capitaine@u-bordeaux.fr>

Repository CRAN

Date/Publication 2019-12-04 14:20:02 UTC

R topics documented:

DataGenCurves	2
Frechforest	2
FrechTree	4
predict.Frechforest	5
predict.FrechTree	5
Index	7

DataGenCurves *Curves data generation function*

Description

This function allows to ...

Usage

```
DataGenCurves(n)
```

Arguments

n [numeric]: Number of trajectories

Value

A list with the followig elements:

- X:
- Y:
- id:
- time:

Examples

```
data <- DataGenCurves(50)
```

Frechforest *Frechet random forest*

Description

This function builds Frechet random Forest introduced by Capitaine et.al, this includes the OOB predictions, OOB errors and variable importance computations.

Usage

```
Frechforest(X, Y, id, time, mtry = ceiling(ncol(X)/3), ntree = 100,  
ncores = 1, ERT = FALSE, timeScale = 0.1, imp = TRUE, ...)
```

Arguments

X	[matrix]: a data frame or a matrix of trajectories predictors. Each column codes for a trajectory predictor.
Y	[vector]: a vector containing the output trajectories (same length as nrow(X)).
id	[vector (factor)]: identifier, one for each trajectory to attribute each measurement of X and Y to one of the trajectories (same length as Y).
time	[vector]: time measurements for the observations of both X and Y (same length as Y).
mtry	[numeric]: number of variables randomly chosen at each split. Default is ncol(X)/3
ntree	[numeric]: number of randomized Frechet trees composing the Frechet random forest. Default is 100.
ncores	[numeric]: number of cores used in parallel. Default is 1.
ERT	[logical]: if TRUE uses Extremely randomized Frechet trees to build the Frechet forest.
timeScale	[numeric]: allow to modify the time scale, increasing or decreasing the cost of the horizontal shift. If timeScale is very big, then the Frechet mean tends to the Euclidean distance. If timeScale is very small, then it tends to the Dynamic Time Warping.
imp	[logical]: TRUE to compute the variables importance FALSE otherwise (default imp=TRUE)
...	: optional parameters to be passed to the low level function.

Value

a Frechet random forest which is a list of the following elements:

- rf: a list of the ntree randomized Frechet trees that compose the forest.
- mse : a vector containing the OOB prediction error of each randomized Frechet tree composing the forest.
- OOB.err: a vector containing the OOB prediction error of each trajectory in the learning sample.
- OOB.pred: a list of the OOB prediction of each trajectory in the learning set.
- Importance: A vector containing the p variables importance.
- rsq: “pseudo R-squared”: $1 - \text{mse}/\text{Var}(y)$.

Examples

```
## Not run:
set.seed(10)
data <- DataGenCurves(50)
FRF <- Frechforest(data$X,data$Y, data$id,data$time, ntree=40,ncores=2, toPlot="none")

## End(Not run)
```

 FrechTree

FrechTree

Description

This function runs Frechet tree for longitudinal data using some shape respecting distance and mean.

Usage

```
FrechTree(X, Y, id, time, select = "CV", ...)
```

Arguments

X	[matrix]: a data frame or a matrix of trajectories predictors. Each column codes for a trajectory predictor.
Y	[vector]: a vector containing the output trajectories (same length as nrow(X)).
id	[vector (factor)]: identifier, one for each trajectory to attribute each measurement of X and Y to one of the trajectories (same length as Y).
time	[vector]: time measurements for the observations of both X and Y (same length as Y).
select	[character]: a character string indicating which criteria to select the optimal pruned sub-tree, either "CV" for cross validation on the prediction error or "Hubert" for the Hubert's statistics. Default is "CV".
...	: optional parameters to be passed to the low level function

Value

a Frechet tree which is a list of the following elements :

- `feuilles`: a vector indicating in which leaf is each measurement.
- `V_split`: a matrix of two columns that describes the structure of the tree. Each row codes for a split, the first column indicates the node and the second column gives the associated splitting variable.
- `Y_curves`: a list of the predicted trajectories (Frechet mean) for each leaf of the optimal Frechet tree.
- `hist_nodes`: a list of the representative trajectories for each node according to the associated splitting variable.

Examples

```
## Not run:
set.seed(10)
data <- DataGenCurves(40)
Ft <- FrechTree(data$X,data$Y, data$id,data$time,select = "Hubert", toPlot="none")

## End(Not run)
```

predict.Frechforest *Frechet random forests prediction function*

Description

Given a Frechet tree and new input trajectories predictors X, this function allows to predict output trajectories Y associated with the new X given a Frechet random forest

Usage

```
## S3 method for class 'Frechforest'
predict(object, X, time, id, timeScale = 0.1, ...)
```

Arguments

object	[list]: a list of the randomized Frechet tree composing the Frechet random forest.
X	[matrix]: a data frame or a matrix of trajectories predictors.
time	[vector]: time measurements of the new trajectories to predict.
id	[vector]: identifier, one for each trajectory to attribute each measurement of X to one of the trajectories.
timeScale	[numeric]: allow to modify the time scale, increasing or decreasing the cost of the horizontal shift. If timeScale is very big, then the Frechet mean tends to the Euclidean distance. If timeScale is very small, then it tends to the Dynamic Time Warping.
...	: optional parameters to be passed to the low level function.

Examples

```
## Not run:
set.seed(10)
data <- DataGenCurves(30)
FRF <- Frechforest(data$X,data$Y, data$id,data$time, ntree=40,ncores=2, toPlot="none")
pred <- predict(FRF, data$X, data$time, data$id)

## End(Not run)
```

predict.FrechTree *Frechet Tree prediction*

Description

Given a Frechet tree and new input trajectories predictors X, this function returns the identifier of the leaf in which each observation falls.

Usage

```
## S3 method for class 'FrechTree'  
predict(object, X, time, id, timeScale = 0.1, ...)
```

Arguments

<code>object</code>	: Frechet tree obtained with the function <code>FrechTree</code> .
<code>X</code>	[matrix]: a data frame or a matrix of trajectories predictors.
<code>time</code>	[vector]: time measurements of the new trajectories to predict.
<code>id</code>	[vector]: identifier, one for each trajectory to attribute each measurement of <code>X</code> to one of the trajectories.
<code>timeScale</code>	[numeric]: allow to modify the time scale, increasing or decreasing the cost of the horizontal shift. If <code>timeScale</code> is very big, then the Frechet mean tends to the Euclidean distance. If <code>timeScale</code> is very small, then it tends to the Dynamic Time Warping.
<code>...</code>	: optional parameters to be passed to the low level function.

Value

a matrix of the identifier of the leaf (second column) in which each individual (first column) falls.

Examples

```
## Not run:  
set.seed(10)  
data <- DataGenCurves(40)  
Ft <- FrechTree(data$X, data$Y, data$id, data$time, select = "Hubert", toPlot="none")  
  
## End(Not run)
```

Index

DataGenCurves, [2](#)

Frechforest, [2](#)

FrechTree, [4](#)

predict.Frechforest, [5](#)

predict.FrechTree, [5](#)