

# Package ‘erhcv’

October 7, 2018

**Type** Package

**Title** Equi-Rank Hierarchical Clustering Validation

**Version** 0.1.3

**License** GPL (>= 2)

**Author** Simon-Pierre Gadoury <spgadou@me.com>

**Maintainer** Simon-Pierre Gadoury <spgadou@me.com>

**Description** Assesses the statistical significance of clusters for a given dataset through bootstrapping and hypothesis testing of a given matrix of empirical Spearman's rho, based on the technique of S. Gaiser et al. (2010) <doi:10.1016/j.jmva.2010.07.008>.

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 6.0.1

**Depends** R (>= 3.5.0)

**Suggests** HAC, knitr, rmarkdown

**Imports** igraph, stringr, stringi, utils, Rdpack

**RdMacros** Rdpack

**Collate** 'VerifyTree.R' 'ClusterNodeSelection.R' 'EliminateCluster.R'  
'GetLeaves.R' 'GetPairs.R' 'hclust2tree.R' 'tree2plot.R'

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2018-10-07 14:50:03 UTC

## R topics documented:

ClusterNodeSelection . . . . .	2
EliminateCluster . . . . .	3
GetLeaves . . . . .	3
GetPairs . . . . .	4
hclust2tree . . . . .	4
tree2plot . . . . .	5
VerifyTree . . . . .	5

---

ClusterNodeSelection *Subcluster significance test*

---

### Description

Test the significance of a subcluster directly under a given cluster. If it is not significant, the subcluster is eliminated and its element are merged in the main cluster. Bootstrap samples of the underlying data needs to be provided, as well as the structure under consideration.

### Usage

```
ClusterNodeSelection(cluster, testPos, alpha, data, BootData)
```

### Arguments

cluster	the main cluster (of the form provided by the function hclust2tree)
testPos	the position of the subcluster to test, directly under the main cluster
alpha	the confidence level for the tests
data	the underlying data
BootData	the dataframe of bootstrap samples of Spearman rho, with columns named "(i,j)", where "i" and "j" are different leaves

### Details

The hypothesis testing is made with the matrix of Spearman's rho for a given dataset, see (Gaisser and Schmid 2010).

### Value

The main cluster, with or without the node under test, whether the hypothesis can be rejected or not.

### Author(s)

Simon-Pierre Gadoury

### References

Gaisser S, Schmid F (2010). "On testing equality of pairwise rank correlations in a multivariate random vector." *Journal of Multivariate Analysis*, **101**(10), 2598–2615.

---

EliminateCluster	<i>Subcluster elimination</i>
------------------	-------------------------------

---

**Description**

Eliminate a subcluster and merge its elements to the main cluster.

**Usage**

```
EliminateCluster(cluster, pos)
```

**Arguments**

cluster	the cluster under consideration (of the form provided by the function <code>hclust2tree</code> )
pos	the position of the subcluster, directly under the main cluster, to eliminate

**Value**

The main cluster, without the subcluster that was eliminated.

**Author(s)**

Simon-Pierre Gadoury

---

GetLeaves	<i>Leaves extractions</i>
-----------	---------------------------

---

**Description**

Extract the leaves (nodes in the bottom of the hierarchy) under (not necessarily directly under) a cluster.

**Usage**

```
GetLeaves(cluster)
```

**Arguments**

cluster	a cluster (of the form provided by the function <code>hclust2tree</code> )
---------	--

**Value**

Vector of leaves.

**Author(s)**

Simon-Pierre Gadoury

---

GetPairs	<i>Obtain pairs of leaves</i>
----------	-------------------------------

---

**Description**

Find all the pairs of leaves between a specific subcluster and all the other subclusters under the main cluster. Pairs between the subclusters of the specific subcluster are also given.

**Usage**

```
GetPairs(cluster, pos)
```

**Arguments**

cluster	main cluster (of the form provided by the function hclust2tree)
pos	position of the subcluster directly under the main cluster. This subcluster is the one every other subcluster will be paired with.

**Value**

Matrix of pairs of leaves.

**Author(s)**

Simon-Pierre Gadoury

---

hclust2tree	<i>hclust object transformation</i>
-------------	-------------------------------------

---

**Description**

Transforms a hclust object into a tree (cluster) used in other functions of this package.

**Usage**

```
hclust2tree(clustering)
```

**Arguments**

clustering	hclust object
------------	---------------

**Author(s)**

Simon-Pierre Gadoury

---

tree2plot	<i>Plot of a tree structure</i>
-----------	---------------------------------

---

**Description**

Provide a plot or a graph.data.frame object for a given tree structure.

**Usage**

```
tree2plot(tree, plot = TRUE, ...)
```

**Arguments**

tree	the tree under consideration of the form provided by the function hclust2tree)
plot	show a basic plot the tree ?
...	extra parameters passed to igraph.plot

**Value**

a plot or graph.data.frame object

**Author(s)**

Simon-Pierre Gadoury

---

VerifyTree	<i>Verify tree structure</i>
------------	------------------------------

---

**Description**

Given a matrix of data, where the rows are observations and the columns are variables, it verifies the statistical significance of hierarchical nodes provided by hclust, through the use of the empirical matrix of Spearman's rho.

**Usage**

```
VerifyTree(data, alpha = 0.95, nboot = 500, distance.method = "maximum",
           hclust.method = "complete")
```

**Arguments**

data	data used for the clustering
alpha	the confidence level for the tests
nboot	the number of bootstrap samples to use
distance.method	method for the distance matrix
hclust.method	method for the clustering

**Details**

The hypothesis testing, as well as the clustering, is made with the matrix of Spearman's rho for a given dataset, see (Gaisser and Schmid 2010).

**Value**

A list, containing the bootstrap samples and the initial tree structure, modified, according to the results of the tests

**Author(s)**

Simon-Pierre Gadoury

**References**

Gaisser S, Schmid F (2010). "On testing equality of pairwise rank correlations in a multivariate random vector." *Journal of Multivariate Analysis*, **101**(10), 2598–2615.

**Examples**

```
require(HAC)
str <- hac(type = 1, tree = list(list(list("X4", "X5", 6),
                                       "X6", 3), "X1", list("X2", "X3", 10), 1))

set.seed(2018)
U.. <- rHAC(1000, str)
U.. <- U..[,c(4, 5, 6, 1, 2, 3)]

## Tree via hclust
spear <- cor(U.., method = "sp")
clust <- hclust(dist(spear, method = "maximum"),
               method = "complete")
tree1 <- hclust2tree(clust)

## Tree after verification
tree2 <- VerifyTree(U.., alpha = 0.95,
                    distance.method = "maximum",
                    hclust.method = "complete")$Tree

## Comparison
par(mfrow = c(1, 3))
tree2plot(tree1)
tree2plot(tree2)
plot(str)
par(mfrow = c(1, 3))
```

# Index

ClusterNodeSelection, [2](#)

EliminateCluster, [3](#)

GetLeaves, [3](#)

GetPairs, [4](#)

hclust2tree, [4](#)

tree2plot, [5](#)

VerifyTree, [5](#)