Package 'CrossClustering'

February 3, 2015

Type Package

Title CrossClustering: a partial clustering algorithm with automatic estimation of the number of clusters and identification of outliers.

Version 1.0

Date 2014-09-04

Author@R c(person(``Paola", ``Tellaroli", email = ``tellaroli@stat.unipd.it"), person(``Marco", ``Bazzi", email = ``bazzi@stat.unipd.it"), person(``Michele", ``Donato", email = ``michele.donato@wayne.edu"))

Author Paola Tellaroli, Marco Bazzi, Michele Donato

Maintainer Paola Tellaroli <tellaroli@stat.unipd.it>

Description CrossClustering is a partial clustering algorithm that combines the Ward's minimum variance and Complete Linkage algorithms, providing automatic estimation of a suitable number of clusters and identification of outlier elements.

License GPL (>=3)

URL http://CRAN.R-project.org/package=CrossClustering

Depends cluster

R topics documented:

`CrossClustering` *CrossClustering: a partial clustering algorithm with automatic estimation of the number of clusters and identification of outliers*

Description

This function performs the CrossClustering algorithm. This method combines the Ward's minimum variance and Complete Linkage algorithms, providing automatic estimation of a suitable number of clusters and identification of outlier elements.

Usage

```
CrossClustering(d, k.w.min = 2, k.w.max, k.c.max, out = TRUE)
```

Arguments

d	a dissimilarity structure as produced by the function dist
k.w.min	minimum number of clusters for the Ward's minimum variance method. By default is set equal 2
k.w.max	maximum number of clusters for the Ward's minimum variance method (see details)
k.c.max	maximum number of clusters for the Complete-linkage method. It can not be equal or greater than the number of elements to cluster (see details)
out	logical. If TRUE (default) outliers must be searched (see details)

Details

See cited document for more details.

Value

A list of objects describing characteristics of the partitioning as follows:

Optimal.cluster	number of clusters
Cluster.list	a list of clusters; each element of this lists contains the indices of the elemenents belonging to the cluster
Silhouette	the average silhouette witdh over all the clusters
n.total	total number of input elements
n.clustered	number of input elements that have actually been clustered

Author(s)

Paola Tellaroli, <tellaroli@stat.unipd.it>; Marco Bazzi, <bazzi@stat.unipd.it>; Michele Donato, <michele.donato@wayne.edu>

References

Tellaroli, P., Bazzi, M., Brazzale, A. R., Donato, M., Draghici, S. Cross Clustering: a partial clustering algorithm with automatic estimation of the number of clusters (manuscript in preparation)

Examples

```
### Generate simulated data
toy <- matrix(NA, nrow=10, ncol=7)
colnames(toy) <- paste("Sample", 1:ncol(toy), sep="")
rownames(toy) <- paste("Gene", 1:nrow(toy), sep="")
set.seed(123)
toy[,1:2] <- rnorm(n=nrow(toy)*2, mean=10, sd=0.1)
toy[,3:4] <- rnorm(n=nrow(toy)*2, mean=20, sd=0.1)
toy[,5:6] <- rnorm(n=nrow(toy)*2, mean=5, sd=0.1)
toy[,7] <- runif(n=nrow(toy), min=0, max=1)
```

```
### toy is transposed as we want to cluster samples (columns of the original matrix)
d <- dist(t(toy), method="euclidean")

### Run CrossClustering
toyres <- CrossClustering(d, k.w.min=2, k.w.max=5, k.c.max=6, out=TRUE)
```

Index