Package 'ClusBoot'

December 13, 2019

Type Package

Title Bootstrap Clustering

Version 1.0

Date 2019-11-11

Author Sugnet Lubbe

Maintainer Sugnet Lubbe <slubbe@sun.ac.za>

Description Clustering algorithms are designed to cluster objects into a number of clusters. Any clustering algorithm provides the 'best'
grouping of objects according to some criterion.
This does not guarantee a 'good' clustering solution in the sense that some allocations were not simply the result of chance.
This package allows the user to apply any clustering algorithm to a data set. The cluster allocations are subjected to a
bootstrap analysis
to determine the extent to which the clustering structure is stable and fundamental to the data set. For more information
see <https://slubbe.wixsite.com/academic-cv/conference-presentations>.

License AGPL-3

NeedsCompilation no

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2019-12-13 15:20:02 UTC

R topics documented:

ClusBoot-package   *Performs bootstrap on a cluster analysis output*

Description

Any clustering output is subjected to a bootstrap procedure to determine the stability of the clustering solution. The results are displayed in the form of a Multi-dimensional scaling plot and a silhouette plot.

Details

The DESCRIPTION file:

Package:	ClusBoot
Type:	Package
Title:	Bootstrap Clustering
Version:	1.0
Date:	2019-11-11
Author:	Sugnet Lubbe
Maintainer:	Sugnet Lubbe <slubbe@sun.ac.za>
Description:	Clustering algorithms are designed to cluster objects into a number of clusters. Any clustering algorithm provid
License:	AGPL-3

Index of help topics:

```
ClusBoot-package        Performs bootstrap on a cluster analysis output
boot.silhouette         Construct silhouette plot from bootstrap
                        replicates
case_study_psychiatrist
                        Patient by psychiatric symptom data
clusboot                Performs bootstrap on a cluster analysis output
complete.linkage        Wrapper function to return only the clustering
                        allocation.
plot.clusboot           Constructs a Multi-dimensional scaling plot to
                        represent the bootstrap replications.
```

A package to perform bootstrap on any cluster analysis output and visualise the results. For more infomation see https://slubbe.wixsite.com/academic-cv/conference-presentations

Author(s)

Sugnet Lubbe slubbe@sun.ac.za

Maintainer: Sugnet Lubbe slubbe@sun.ac.za

References

Lubbe, S. Visualisations associated with bootstrapping cluster analysis. Data Science, Statistics and Visualisation conference, July 2017, Lisbon, Portugal.

---

boot.silhouette	*Construct silhouette plot from bootstrap replicates*

Description

A silhouette plot is constructed, indicating the proportion of times the cluster members cluster together in the same cluster.

Usage

```
boot.silhouette(clusboot.out, ...)
```

Arguments

clusboot.out	an object of class clusboot, usually, a result of a call to clusboot
...	more plotting parameters, e.g. col

Details

The clustering tightness is computed for each cluster, compared to the 'nearest' alternative cluster. The cluster tightness is computed as the mean of the proportion of times each pair of objects are clustered together in the same cluster. The tightness to other clusters is computed as the mean of the proportion an item of this cluster and one from the other cluster as clustered together. The 'nearest' alternative cluster is defined as the alternative cluster with the numerically largest tightness to the current cluster.

Value

a vector with number of components equal to the number of classes, providing the silhouette value for each class.

Author(s)

Sugnet Lubbe slubbe@sun.ac.za

References

Lubbe, S. Visualisations associated with bootstrapping cluster analysis. Data Science, Statistics and Visualisation conference, July 2017, Lisbon, Portugal.

See Also

clusboot for performing bootstrap on a cluster analysis output and plot.clusboot to visually represent the bootstrap replications

Examples

```
data(case_study_psychiatrist)
boot.out <- clusboot (scale(case_study_psychiatrist), B=100, k=6, clustering.func=complete.linkage)
plot(boot.out)
boot.silhouette (boot.out)
```

---

case_study_psychiatrist

Patient by psychiatric symptom data

Description

Abstract: Presence/absence ratings of 24 psychiatric symptoms in 30 psychiatric inpatients made by an individual psychiatrist. Subject matter background: The data have been collected in a case study of an individual psychiatrist to identify his implicit taxonomy. Data structure: object x variables data matrix

Usage

```
data(case_study_psychiatrist)
```

Format

A data frame with 30 observations on the following 28 variables.

V1 inappropriate affect, appearance or behavior; binary vector

V2 interview belligerence - negativism; binary vector

V3 agitation - excitement; binary vector

V4 retardation; binary vector

V5 lack of emotions; binary vector

V6 speech disorganization; binary vector

V7 grandiosity; binary vector

V8 suspicion - ideas of persecution; binary vector

V9 hallucinations - delusions; binary vector

V10 overt anger; binary vector

V11 depression; binary vector

V12 anxiety; binary vector

V13 obsession - compulsion; binary vector

V14 suicide; binary vector

V15 self injury; binary vector

V16 somatic concerns; binary vector

V17 social isolation; binary vector

V18 daily routine impairment; binary vector

V19 leisure time impairment; binary vector

V20 antisocial impulses or acts; binary vector

V21 alcohol abuse; binary vector

V22 drug abuse; binary vector

V23 disorientation; binary vector

V24 memory impairment; binary vector

V25 rating on Global Assessment Scale, a 101-point scale for overall severity of psychiatric disturbance; a numeric vector

V26 Affective (Affective Disorder or Anxiety Disorder); binary vector

V27 Psychotic (Schizophrenic Disorder or Paranoid Disorder); binary vector

V28 Substance abuse (Substance Use Disorder or Substance-Induced Disorder); binary vector

Details

The data set forms part of the International Federation of Classification Societies Cluster Benchmark Data Repository

Source

Van Mechelen, I., & De Boeck, P. (1989). Implicit taxonomy in psychiatric diagnosis: A case study. Journal of Social and Clinical Psychology, 8, 276-287.

References

http://ifcs.boku.ac.at/repository/data/case_study_psychiatrist/index.html

clusboot	*Performs bootstrap on a cluster analysis output*

Description

B bootstrap samples are drawn with replacement from the data and cluster analysis is performed on the bootstrap samples.

Usage

```
clusboot(datmat, B = 1000, clustering.func = complete.linkage, ...)
```

Arguments

datmat	a samples by variables data matrix or a distance object, whatever input is required for clustering.func
B	number of bootstrap replicates
clustering.func	
	any function that performs cluster analysis which returns a single vector of cluster allocations as output
...	any other arguments to be sent to clustering.func

Value

proportions	an nxn matrix with (i,j)-th element the proportion of times objects i and j clustered in the same cluster)
clustering	vector of cluster allocations of the input data

Author(s)

Sugnet Lubbe slubbe@sun.ac.za

References

Lubbe, S. Visualisations associated with bootstrapping cluster analysis. Data Science, Statistics and Visualisation conference, July 2017, Lisbon, Portugal.

See Also

[plot.clusboot](#) to visually represent the bootstrap replications and [boot.silhouette](#) for a silhouette summary of the bootstrap replicates as well as [kmeans](#)

Examples

```
data(case_study_psychiatrist)
clusboot (scale(case_study_psychiatrist), B=100, k=6, clustering.func=complete.linkage)
```

---

complete.linkage	*Wrapper function to return only the clustering allocation.*

Description

Wrapper function to return only the clustering allocation from the hclust function with default method = "complete".

Usage

```
complete.linkage(X, k)
```

Arguments

X	numeric matrix of data, or an object that can be coerced to such a matrix (such as a numeric vector or a data frame with all numeric columns).
k	either the number of clusters, say k, or a set of initial (distinct) cluster centres. If a number, a random set of (distinct) rows in x is chosen as the initial centres.

Value

a single vector of cluster allocations, the output of cutree(hclust(dist(X)), k)

Author(s)

Sugnet Lubbe slubbe@sun.ac.za

References

hclust()

Examples

```
data(case_study_psychiatrist)
complete.linkage(scale(case_study_psychiatrist), k=6)
```

---

plot.clusboot	*Constructs a Multi-dimensional scaling plot to represent the bootstrap replications.*

Description

Constructs a Multi-dimensional scaling plot to represent the bootstrap replications.

Usage

```
## S3 method for class 'clusboot'
plot(x, col = NULL, ...)
```

Arguments

x	an object of class clusboot, usually, a result of a call to clusboot
col	an optional vector equal in length to the number of clusters with colours to represent different clusters
...	more plotting parameters

Details

Constructs a Multi-dimensional scaling plot to represent the bootstrap replications.

Value

a two-column matrix with rows containing the MDS coordinates of the samples

Author(s)

Sugnet Lubbe slubbe@sun.ac.za

References

Lubbe, S. Visualisations associated with bootstrapping cluster analysis. Data Science, Statistics and Visualisation conference, July 2017, Lisbon, Portugal.

See Also

clusboot for performing bootstrap on a cluster analysis output and plot.clusboot to visually represent the bootstrap replications

Examples

```
data(case_study_psychiatrist)
boot.out <- clusboot (scale(case_study_psychiatrist), B=100, k=6, clustering.func=complete.linkage)
plot(boot.out)
```

Index