Package 'clusteringPkg'

May 15, 2015

Type Package

Title Clusterng and reporting cluster memberships

Version 1.0

Date 2015-05-14

Author Georgia Tsiliki

Maintainer Georgia Tsiliki <gtsiliki@central.ntua.gr>

Depends R (>= 2.10), RCurl, pmml, jsonlite, blockcluster, Matrix,
vegan, jpeg

Description This package employs hierarchical clustering and bi-clustering

License GPL-2

R topics documented:

---

clusteringPkg-package *Generate cluster memberships*

Description

Estimate the clustering memberships for colmns and rows of a given matrix respectively. Data are clustered based on hierarchical clustering algorithm from the vegan R package or bi-clustering from the blockcluster R package.

Details

Package:	clusteringPkg
Type:	Package
Version:	1.0
Date:	2015-05-14
License:	GPL-2

Estimate cluster memberships for a given data matrix. Important functions are generate.biclust.model and generate.hierar.model.

Author(s)

Georgia Tsiliki

Maintainer: Georgia Tsiliki <gtsiliki@central.ntua.gr>

References

Help files of blockcluster and vegan packages

Examples

```
data("dat11")

predF<- list()

required.param<- list(xID="proteins", yID=list(), distMethod="manhattan",
hclustMethod="ward.D2",nORh="mean")

hierar1<- generate.hierar.model(dat11,predF,required.param)
```

---

dat1 *A sample data object*

Description

The dataset for this test is a data frame

Usage

```
data("dat1")
```

Format

A list of two objects

datasetURI a character vector- ambit data set uri

dataEntry a data frame containing two columns: compound and values. Compound is a character vector with all compound anbit uris, and calues is a data frame with all numberic values of the data set (compounds by features)

Details

There are no more details

Source

The source of this function is in the

References

There are no references

Examples

```
data(dat1)
## maybe str(dat1) ; plot(dat1) ...
```

---

dat11   *A sample data object*

Description

The dataset for this test is a data frame

Usage

```
data("dat11")
```

Format

A list of two objects

datasetURI a character vector- ambit data set uri

dataEntry a data frame containing two columns: compound and values. Compound is a character
vector with all compound anbit uris, and calues is a data frame with all numberic values of the
data set (compounds by features)

Details

There are no more details

Source

The source of this function is in the

References

There are no references

Examples

```
data(dat11)
## maybe str(dat11) ; plot(dat11) ...
```

---

dat11i *Information for hierarchical clustering*

Description

A list with information for hierarchical given by vegan package

Usage

```
data("dat11i")
```

Format

A list with 3 objects:

Summary Summary of hierarchical clustering model

modelImages A serialized jpeg file of a dendrogram

predictedFeatures The character string 'clusterMemberships' to indicate the new vector created
including all predicted clustering memberships

Details

Example dataset to predict with dat1p, dat1m

Source

The source of this function is in the

References

There are no references

Examples

```
data(dat11i)
## maybe str(dat11i) ; plot(dat11i) ...
```

---

dat11m *Serialized hierarchical clustering model file*

Description

A character string for a serialized hierarchical clustering model, i.e. a vector including cluster memberships

Usage

```
data("dat11m")
```

Format

A character string

Details

Example hierarchical clustering model based on dat11

Source

The source of this function is in the

References

There are no references

Examples

```
data(dat11m)
## maybe str(dat11m) ; plot(dat11m) ...
```

---

dat11p *A sample data object*

Description

The dataset for this test is a data frame

Usage

```
data("dat11p")
```

Format

A list of two objects

datasetURI a character vector- ambit data set uri

dataEntry a data frame containing two columns: compound and values. Compound is a character vector with all compound anbit uris, and calues is a data frame with all numberic values of the data set (compounds by features)

Details

Data set for prediction with dat11m

Source

The source of this function is in the

References

There are no references

Examples

```
data(dat11p)
## maybe str(dat11p) ; plot(dat11p) ...
```

---

dat1i   *Information for bi-clustering*

Description

A list with information for bi-clustering given by blockcluser package

Usage

```
data("dat1i")
```

Format

A list with 4 objects:

ICL Integrated complete likelihood value as returned by blockcluster package

Summary Summary of bi-clust model

modelImages A serialized jpeg file of a heatmap produced by blockcluster

predictedFeatures The character string 'clusterMemberships' to indicate the new vector created including all predicted clustering memberships

Details

Example dataset to predict with dat1p, dat1m

Source

The source of this function is in the

References

There are no references

Examples

```
data(dat1i)
## maybe str(dat1i) ; plot(dat1i) ...
```

---

dat1m   *Serialized bi-clust model file*

Description

A character string for a serialized bi-clust model, i.e. a list including two vectors cluster membership for columns and rows respectively

Usage

```
data("dat1m")
```

Format

A character string

Details

Example bi-clust model based on dat1

Source

The source of this function is in the

References

There are no references

Examples

```
data(dat1m)
## maybe str(dat1m) ; plot(dat1m) ...
```

dat1p	*A sample data object*

Description

The dataset for this test is a data frame

Usage

```
data("dat1p")
```

Format

A list of two objects

datasetURI a character vector- ambit data set uri

dataEntry a data frame containing two columns: compound and values. Compound is a character vector with all compound anbit uris, and calues is a data frame with all numberic values of the data set (compounds by features)

Details

Data set for prediction with dat1m

Source

The source of this function is in the

References

There are no references

Examples

```
data(dat1p)
## maybe str(dat1p) ; plot(dat1p) ...
```

---

generate.biclust.model

Generate bi-clustering model based on blockcluster package.

Description

Estimate the clustering memberships for colmns an drows of a given matrix respectively. Data are clustered based on bi-clustering algorithm from the blockcluster R package using default values. The user needs to specify the number of clusters for each axes.

Usage

```
generate.biclust.model(dataset, predictionFeature, parameters)
```

Arguments

dataset   list of 2 objects, datasetURI:= character sring, code name of dataset, dataEntry:=
                    data frame with 2 columns

predictionFeature
                    character string specifying which is the prediction feature in dataEntry

parameters   list with parameter values for biclustering. 4 objects should be included, i.e.
                    'xID' the ID of values in x axis, 'yID' the ID of values in y axis, 'datatype'
                    the data type of the dataset (one of "binary" , "contingency", "continuous" or
                    "categorical".), 'nclust' a numeric vector indicating the number of clusters for x
                    axis and y axis (e.g. c(5,4)).

Details

No details required

Value

A List

rawModel   A serialized GO descriptors object (class raw) giving the cluster memberships
                    of proteins/genes in the data.

pmmlModel   A pmml object - now empty

predictedFeatures
                    A character vector with names for the new descriptors

independentFeatures
                    A list with Ambit names for all genes/ proteins features included in the model

additionalInfo   A list including the following: ICL (Integrated complete likelihood value as re-
                    turned by blockcluster package), Summary (Summary of bi-clust model), mod-
                    elImages (A serialized jpeg file of a heatmap produced by blockcluster), pre-
                    dictedFeatures (The character string 'clusterMemberships' to indicate the new
                    vector created including all predicted clustering memberships).

Note

No notes for this function

Author(s)

Georgia Tsiliki

References

The help file of blockcluster package

See Also

generate.hierar.model

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

data("dat1")

predF<- list()

required.param<- list(xID='proteins',yID='NPs',datatype='categorical',nclust=c(1,2))

biclust1<- generate.biclust.model(dat1,predF,required.param)
```

---

generate.hierar.model   *Generate bi-clustering model based on vegan package.*

Description

Estimate the clustering memberships for colmns and rows of a given matrix respectively. Data are clustered based on hierarchical clustering algorithm from the vegan R package using default values. The user needs to specify the number of clusters or the height of the dendrogram.

Usage

```
generate.hierar.model(dataset, predictionFeature, parameters)
```

Arguments

dataset	list of 2 objects, datasetURI:= character sring, code name of dataset, dataEntry:= data frame with 2 columns
predictionFeature	
	character string specifying which is the prediction feature in dataEntry
parameters	list with parameter values for hierarchical clustering. 5 objects should be included, i.e. 'xID' the ID of values in x axis, 'yID' the ID of values in y axis, distance method (could be one of those provided via vegan R package), 'hclust-Method' (could be one of those provided via vegan R package), 'nORh' either a numeric value or character giving number of clusters or a function to define height respectively.

Details

No details required

Value

A List

rawModel	A serialized GO descriptors object (class raw) giving the cluster memberships of proteins/genes in the data.
pmmlModel	A pmml object - now empty
independentFeatures	
	A list with Ambit names for all genes/ proteins features included in the model
predictedFeatures	
	A character vector with names for the new descriptors
additionalInfo	A list including the following: ICL (Integrated complete likelihood value as returned by blockcluster package), Summary (Summary of hierarchical clustering model), modelImages (A serialized jpeg file of a dedrogram), predictedFeatures (The character string 'clusterMemberships' to indicate the new vector created including all predicted clustering memberships).

Note

No notes for this function

Author(s)

Georgia Tsiliki

References

The help file of skeleton

See Also

generate.biclust.model

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

data("dat11")

predF<- list()

required.param<- list(xID="proteins", yID=list(), distMethod="manhattan",
hclustMethod="ward.D2",nORh="mean")

hierar1<- generate.hierar.model(dat11,predF,required.param)
```

---

pred.clusters   *Returns clustermemberships*

Description

Returns cluster memberships as estimeated by generate.hierar.model or generate.biclust.model functions.

Usage

```
pred.clusters(dataset, rawModel, additionalInfo)
```

Arguments

dataset	Data for prediction. A list of two objects: datasetURI (a character string), dataEntry (a data frame).
rawModel	R model serialized (cluster memberships for one or two dimensions of the data matrix supplied).
additionalInfo	Any additional information needed for rawModel. Here the list generated by generate.hierar.model or generate.biclust.model functions. The list should contain a field named 'predictedFeatures' which should be exactly the same as that returned by generate.hierar.model or generate.biclust.model functions.

Details

No further details required

Value

A list of one objected called 'predictions' which is also a list of one cell data-frames each containing the cluster membership of the particular row or column clustered. Columns or rows maintain their original order. In case of biclustering column memberships appera first and then row memberships.

Note

No notes for this function

Author(s)

Georgia Tsiliki

References

No references required

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

data("dat1p")
data("dat1m")
data("dat1i")

pred.res<- pred.clusters(dat1p, dat1m, dat1i)
```

Index