Package 'superbiclust'

October 14, 2022

Type Package

Title Generating Robust Biclusters from a Bicluster Set (Ensemble Biclustering)

Version 1.2

Date 2021-05-26

Author Tatsiana Khamiakova

Maintainer Tatsiana Khamiakova <t.tavita@gmail.com>

Depends biclust, fabia

Imports methods, Matrix, graphics, stats

Description Biclusters are submatrices in the data matrix which satisfy certain conditions of homogeneity. Package contains functions for generating robust biclusters with respect to the initialization parameters for a given bicluster solution contained in a bicluster set in data, the procedure is also known as ensemble biclustering. The set of biclusters is evaluated based on the similarity of its elements (the overlap), and afterwards the hierarchical tree is constructed to obtain cut-off points for the classes of robust biclusters. The result is a number of robust (or super) biclusters with none or low overlap.

License GPL (>= 2)

LazyLoad yes

NeedsCompilation no

Repository CRAN

Date/Publication 2021-05-28 06:00:02 UTC

Repository/R-Forge/Project superbiclust

Repository/R-Forge/Revision 14

Repository/R-Forge/DateTimeStamp 2014-11-27 15:39:29

R **topics documented:**

---

superbiclust-package	*generating robust biclusters form the set of biclusters*

Description

The package contains a number of functions for computing similarity matrix of the biclusters obtained by a variety of methods, initialization seeds or various parameter settings. It uses biclustering output as generated by biclust or fabia. isa2 package can be used to generate the biclusters as well, however, a prior conversion is needed to a biclust object by using isa2.biclust() function. The matrix is used for the construction of hierarchical tree based on overall similarity, row similarity or column similarity to obtain cut-off points for the similarity metric of choice. Various statistics are output per bicluster set: a number of a given gene(compound) or gene (compound) set has been present in any bicluster of output or per run. After the tree is cut, the robiust or super biclusters are obtained in a form of biclust object, which can further be used for plotting of biclusters. Biclusters are submatrices in the data which satisfy certain conditions of homogeneity. For more details on biclusters and biclustering see Madeira and Oliveira (2004).

Details

Package:	superbiclust
Type:	Package
Version:	0.99
Date:	2012-08-23
License:	GPL
LazyLoad:	yes

Author(s)

Tatsiana Khamiakova <tatsiana.khamiakova@uhasselt.be>

References

Madeira and Oliveira (2004) Biclustering algorithms for biological data analysis: a survey. IEEE/ACM Trans Comput Biol Bioinform. 2004 Jan-Mar;1(1):24-45. Shi et al. (2010) A bi-ordering approach to linking gene expression with clinical annotations in gastric cancer. BMC Bioinformatics. 11. pages 477.

BiclustSet-class	*Class* BiclustSet

Description

BiclustSet Class contains the biclustering result in a form: bicluster rows and bicluster columns

Objects from the Class

Objects can be created by calls of the form new("BiclustSet", ...). The variety of inputs variety of inputs (isa2, fabia, biclust,...) can be used.

Slots

GenesMembership: logical, object of class "matrix", with row membership within a bicluster

ColumnMembership: logical, object of class "matrix", with column membership within a bicluster

Number: code"numeric", number of biclusters in the set

Author(s)

Tatsiana Khamiakova

BiclustSet-methods   *Constructor of BiclustSet object*

Description

The method extract relevant information from a variety of biclustering input and constructs a BiclustSet object

Methods

signature(x = "ANY")

signature(x = "Biclust") Converts Biclust objects into BiclustSet object

signature(x = "Factorization") Converts FABIA Factorization object into BiClustSet

signature(x = "list") Converts a list with biclustering results into BiClustSet

See Also

[BiclustSet](#)

Examples

```
 test <- matrix(rnorm(5000), 100, 50)
 test[11:20,11:20] <- rnorm(100, 3, 0.1)
 test[17:26,21:30] <- rnorm(100, 3, 0.1)
#Run FABIA
 set.seed(1)
 FabiaRes1 <- fabia(test)
#construct BiclustSet object from FABIA output
 FabiabiclustSet <- BiclustSet(FabiaRes1)
 FabiabiclustSet
```

---

combine   *Combine two Biclust objects into one*

Description

Combine two Biclust objects into one

Usage

```
combine(x,y)
```

Arguments

x   1st Biclust object containing bicluster results

y   2nd Biclust object containing bicluster results

Details

If a biclust function returns empty set, joined result contains only results of non-empty object. Info and Parameters slots of a "Biclust" object contain information about both biclustering runs.

Value

object of a class Biclust

Author(s)

Tatsiana Khamiakova

See Also

[BiclustSet](#)

Examples

```
#combine output of two biclust objects
 test <- matrix(rnorm(5000), 100, 50)
 test[11:20,11:20] <- rnorm(100, 3, 0.1)
 test[17:26,21:30] <- rnorm(100, 3, 0.1)
 set.seed(1)
 PlaidRes1 <- biclust(x=test, method=BCPlaid())
 set.seed(2)
 PlaidRes2 <- biclust(x=test, method=BCPlaid())
 combinedRes <- combine(PlaidRes1,PlaidRes2)
 summary(combinedRes)
```

| getStats | *Get frequency statistic for the columns and rows membership* |

Description

For a given Bicluster set, for each row and column in data, compute frequency of apperance within a bicluster

Usage

```
getStats(x)
```

Arguments

x   Biclust object containing bicluster results

Value

a list of column and row frequencies

Author(s)

Tatsiana Khamiakova

HCLtree   *Hierarchical structure of bicluster output*

Description

Constructs and plots hierarchical tree of biclusters output based on the similarity matrix

Usage

```
HCLtree(x)
```

Arguments

x   Similarity object containing pairwise similarity indices for all biclusters in the output

Details

This function operates on a similarity matrix, which is converted to the distance between biclusters according to $dist(a, b) = 1 - sim(a, b)$, where the smaller the distance, the higher is overlap in terms of rows and columns. The tree is constructed using complete method and plotted. Further, the structure must be explored and robust or super-biclusters obtained after cutting the tree. identify function can be applied to the hierarchical tree to see the partition and get the plots of biclusters.

Value

tree

Author(s)

Tatsiana Khamiakova

See Also

similarity, plotSuper

Examples

```
#compute sensitivity for BiMAX biclusters
 test <- matrix(rnorm(5000), 100, 50)
 test[11:20,11:20] <- rnorm(100, 3, 0.1)
 test[17:26,21:30] <- rnorm(100, 3, 0.1)
 testBin <- binarize(test,2)
 res <- biclust(x=testBin, method=BCBimax(), minr=4, minc=4, number=10)
 BiMaxBiclustSet <-  BiclustSet(res)
```

```
 SensitivityMatr<- similarity(BiMaxBiclustSet,index="sensitivity")
#construct hierarchical clustering based on the sensitivities
HCLsensitivity <- HCLtree(SensitivityMatr)
plot(HCLsensitivity, main="structure of bicluster solution")
```

---

jaccardMat   *Jaccard similarity Matrix for bicluster output*

Description

computes jaccard similarity matrix for biclusters in two bicluster sets

Usage

```
jaccardMat(x, y, type=c("rows", "cols", "both"))
```

Arguments

x	BiclustSet object containing row and column indicators of bicluster membership, number of biclusters
y	BiclustSet object containing row and column indicators of bicluster membership, number of biclusters
type	whether to compute Jaccard index in two dimensions, row dimension or column dimension

Details

This function operates on BiclustSet objects and computes pairwise similarity based on the common elements between biclusters. The Jaccard similarity score ja for two biclusters A and B is computed as

$$ja = \frac{|A \cap B|}{|A \cup B|}$$

Value

matrix of pairwise Jaccard indices

Author(s)

Tatsiana Khamiakova

See Also

[similarity](#),kulczynskiMat, [ochiaiMat](#), [sensitivityMat](#), [specificityMat](#),sorensenMat

---

| kulczynskiMat | *Kulczynski similarity Matrix for bicluster output* |

Description

computes Kulczynski similarity matrix for biclusters in two bicluster sets

Usage

```
kulczynskiMat(x, y, type=c("rows", "cols", "both"))
```

Arguments

x	BiclustSet object containing row and column indicators of bicluster membership, number of biclusters
y	BiclustSet object containing row and column indicators of bicluster membership, number of biclusters
type	whether to compute Kulczynski index in two dimensions, row dimension or column dimension

Details

This function operates on BiclustSet objects and computes pairwise similarity based on the common elements between biclusters. Kulczynski similarity score ku for two biclusters A and B is computed as

$$ku = 2|A \cap B| \left(\frac{1}{|A|} + \frac{1}{|B|} \right)$$

Value

matrix of pairwise Kulczynski indices

Author(s)

Tatsiana Khamiakova

See Also

[similarity](), [jaccardMat](), [ochiaiMat](), [sensitivityMat](), [specificityMat](), sorensenMat

---

ochiaiMat *Ochiai similarity Matrix for bicluster output*

Description

Computes Ochiai similarity matrix for biclusters in two bicluster sets

Usage

```
ochiaiMat(x, y, type=c("rows", "cols", "both"))
```

Arguments

x	BiclustSet object containing row and column indicators of bicluster membership, number of biclusters
y	BiclustSet object containing row and column indicators of bicluster membership, number of biclusters
type	whether to compute Ochiai index in two dimensions, row dimension or column dimension

Details

This function operates on BiclustSet objects and computes pairwise similarity based on the common elements between biclusters. The Ochiai similarity score ja for two biclusters A and B is computed as

$$oc = \frac{|A \cap B|}{\sqrt{|A||B|}}$$

Value

matrix of pairwise Ochiai indices

Author(s)

Tatsiana Khamiakova

See Also

similarity, jaccardMat, ochiaiMat, sensitivityMat, specificityMat, sorensenMat

plotProfilesAcrossAllSamples

Plot Gene Expression Profiles Across All Samples of the Original Data

Description

Plot Gene Expression Profiles Across All Samples of the Original Data

Usage

```
plotProfilesAcrossAllSamples(x, coreBiclusterGenes, coreBiclusterSamples)
```

Arguments

x   data

coreBiclusterGenes

vector of genes belonging to bicluster

coreBiclusterSamples

vector of samples belonging to bicluster

Details

The plot re-sorts the samples by bicluster membership and highlights them in red. Only the genes of a bicluster are plotted.

Value

no return value; a plot is drawn to the current device

Author(s)

Tatsiana Khamiakova

See Also

BiclustSet ,plotSuperAll

---

plotProfilesWithinBicluster

Plot Gene Expression Profiles within a (Core) Bicluster

Description

Plot Gene Expression Profiles within a (Core) Bicluster

Usage

```
plotProfilesWithinBicluster(x, main = "", sampleNames, geneNames = NULL)
```

Arguments

x	expression matrix (of class 'matrix') for the subset of genes and samples corresponding to the bicluster under study
main	main title for the graph
sampleNames	names of the samples to be used for annotating the x axis (character vector of length equal to the number of columns of the expression matrix 'x' (representing the bicluster)
geneNames	names of the genes to be plotted in a legend (character vector of length equal to the number of rows of the expression matrix 'x'); only suitable for biclusters containing a small number of genes

Value

no return value; a plot is drawn to the current device

Author(s)

Tatsiana Khamiakova

See Also

plotSuper

plotSuper *Plot gene profiles within biclusters*

Description

Function for plotting gene profiles for compounds within constructed super-bicluster

Usage

```
plotSuper(x, data, BiclustSet)
```

Arguments

x	a vector, containing indices of biclusters, to be joined for obtaining the robust bicluster
data	matrix, dataset, from which the bicluster results are obtained
BiclustSet	a BiclustSet object containing bicluster output

Details

This function constructs a robust bicluster from a set of biclusters identified in a hierarchical tree and plots gene profiles for columns in a robust bicluster. Each line represents a gene from a bicluster. The bicluster is saved as Biclust object which can be further plotetd by available functions from biclust package. The information about the number of biclusters used to generate the resulting robust bicluster is saved in Call slot of the object. This information is important to see how often the bicluster has been discovered under different parameter settings (e.g. initialization seeds) Indices used as an input can be obtained by identify function or by cutting the tree.

Value

biclust object containing bicluster and the information about bicluster subset used to generate it

Author(s)

Tatsiana Khamiakova

See Also

[HCLtree](), [plotSuperAll](), [plotProfilesWithinBicluster]()

---

plotSuperAll	*Plot gene profiles for all samples in the data*

Description

Function for plotting bicluster gene profiles for all samples in the data

Usage

```
plotSuperAll(x, data, BiclustSet)
```

Arguments

x	a vector, containing indices of biclusters, to be joined for obtaining the robust bicluster
data	matrix, dataset, from which the bicluster results are obtained
BiclustSet	a BiclustSet object containing bicluster output

Details

This function constructs a robust bicluster from a subset of biclusters specified in x argument and plots the expression profiles

Value

biclust object

Author(s)

Tatsiana Khamiakova

See Also

[HCLtree](#), [plotProfilesAcrossAllSamples](#)

---

sensitivityMat	*Sensitivity Matrix for bicluster output*

Description

Computes sensitivity matrix for biclusters in two bicluster sets

Usage

```
sensitivityMat(x, y, type=c("rows", "cols", "both"))
```

Arguments

x	BiclustSet object containing row and column indicators of bicluster membership, number of biclusters
y	BiclustSet object containing row and column indicators of bicluster membership, number of biclusters
type	whether to compute sensitivity in two dimensions, row dimension or column dimension

Details

This function operates on BiclustSet objects and computes pairwise similarity based on the common elements between biclusters. Sensitivity inclusion score sen of biclusters A and B is computed as

$$sen = \frac{|A \cap B|}{|A|}$$

Value

matrix of pairwise sensitivities

Author(s)

Tatsiana Khamiakova

See Also

[similarity](#),[jaccardMat](#), [ochiaiMat](#), [kulczynskiMat](#), [specificityMat](#),[sorensenMat](#)

---

similarity	*Similarity Matrix for bicluster output*

Description

computes similarity matrix for the biclustering output based on one of the pairwise similarity indices of biclusters in a given bicluster set

Usage

```
similarity(x, index = "jaccard", type="rows")
```

Arguments

x	BiclustSet object containing row and column indicators of bicluster membership, number of biclusters
index	similarity index for the biclusters in output
type	whether to perform similarity in two dimensions, "both" (recommended for biclustering), row dimension, "rows" (default, requires less computations) or column dimension "cols"

Details

This function operates on BiclustSet object and computes pairwise similarity based on the common elements between biclusters. type variable controls whether similarity index is constructed for all elements, or in one dimension (rows or columns) only. In general, similarity indices for one dimension (row or column) are higher than for two-dimensions. Several options for similarity indices are offered: jaccard (default), kulczynski, sensitivity, specificity, sorensen and ochiai indices.

Value

a "similarity" object containing similarity matrix

Author(s)

Tatsiana Khamiakova

See Also

[HCLtree](), [plotSuper]() , [jaccardMat](),[kulczynskiMat](), [ochiaiMat](), [sensitivityMat](), [specificityMat](),[sorensenMat]()

Examples

```
#compute sensitivity for BiMAX biclusters
 test <- matrix(rnorm(5000), 100, 50)
 test[11:20,11:20] <- rnorm(100, 3, 0.1)
 test[17:26,21:30] <- rnorm(100, 3, 0.1)
 testBin <- binarize(test,2)
 res <- biclust(x=testBin, method=BCBimax(), minr=4, minc=4, number=10)
 BiMaxBiclustSet <-  BiclustSet(res)
 SensitivityMatr<- similarity(BiMaxBiclustSet,index="sensitivity", type="rows")
 SensitivityMatr
```

---

sorensenMat	*Sorensen similarity Matrix for bicluster output*

Description

Computes Sorensen similarity matrix for biclusters in two bicluster sets

Usage

```
sorensenMat(x, y, type=c("rows", "cols", "both"))
```

Arguments

x	BiclustSet object containing row and column indicators of bicluster membership, number of biclusters
y	BiclustSet object containing row and column indicators of bicluster membership, number of biclusters
type	whether to compute Sorensen index in two dimensions, row dimension or column dimension

Details

This function operates on BiclustSet objects and computes pairwise similarity based on the common elements between biclusters. Sorensen similarity score *so* for two biclusters A and B is computed as

$$so = \frac{2|A \cap B|}{|A| + |B|}$$

Value

matrix of pairwise Sorensen indices

Author(s)

Tatsiana Khamiakova

See Also

similarity,jaccardMat, ochiaiMat, sensitivityMat, specificityMat,kulczynskiMat

---

specificityMat   *Specificity Matrix for bicluster output*

Description

Computes specificity matrix for biclusters in two bicluster sets

Usage

```
specificityMat(x, y, type=c("rows", "cols", "both"))
```

Arguments

x	BiclustSet object containing row and column indicators of bicluster membership, number of biclusters
y	BiclustSet object containing row and column indicators of bicluster membership, number of biclusters
type	whether to compute specificity in two dimensions, row dimension or column dimension

Details

This function operates on BiclustSet objects and computes pairwise similarity based on the common elements between biclusters. Sensitivity inclusion score *spe* of biclusters A and B is computed as

$$spe = \frac{|A \cap B|}{|B|}$$

Value

matrix of pairwise specificities

Author(s)

Tatsiana Khamiakova

See Also

[similarity](#),[jaccardMat](#), [ochiaiMat](#), [kulczynskiMat](#), [sensitivityMat](#),[sorensenMat](#)

Index