Package ‘s4vd’

November 19, 2015

Version 1.1-1
Title Biclustering via Sparse Singular Value Decomposition
   Incorporating Stability Selection
Date 2015-11-11
Author Martin Sill, Sebastian Kaiser
Maintainer Martin Sill <m.sill@dkfz.de>
Depends biclust,methods,irlba,foreach
Description The main function s4vd() performs a biclustering via sparse
   singular value decomposition with a nested stability selection.
   The results is an biclust object and thus all methods of the
   biclust package can be applied.
License GPL-2
LazyLoad yes
Repository CRAN
Repository/R-Forge/Project s4vd
Repository/R-Forge/Revision 109
Repository/R-Forge/DateTimeStamp 2015-11-18 10:48:14
Date/Publication 2015-11-19 17:15:53
NeedsCompilation no

R topics documented:

BCheatmap .......................................................... 2
BCs4vd ........................................................... 3
BCssvd ........................................................... 5
jaccardmat ......................................................... 7
lung200 ........................................................... 8
stabpath ........................................................... 9

Index 11

1
BHeatmap

Overlap Heatmap for the visualization of overlapping biclusters

Description

Heatmap function to plot biclustering results. Overlapping biclusters are indicated by colored rectangles.

Usage

BHeatmap(X, res, cexR = 1.5, cexC = 1.25, axisR = FALSE, axisC = TRUE, heatcols = maPalette(low="blue", mid="white", high="red", k=50), clustercols = c(1:5), allrows = FALSE, allcolumns = TRUE)

Arguments

X the data matrix
res the biclustering result
cexR relativ font size of the row labels
cexC relativ font size of the column labels
axisR if TRUE the row labels will be plotted
axisC if TRUE the column labels will be plotted
heatcols a character vector specifying the heatcolors
clustercols a character vector specifying the colors of the rectangles that indicate the rows and columns that belong to a bicluster
allrows if FALSE only the rows assigned to any bicluster will be plotted
allcolumns if FALSE only the columns assigned to any bicluster will be plotted

Author(s)

Martin Sill <m.sill@dkfz.de>

Examples

#lung cancer data set Bhattacharjee et al. 2001
data(lung200)
set.seed(12)
res1 <- biclust(lung200, method=BCs4vd(), pcer=.5, pceru=0.01, ss.thr=c(0.6,0.65), start.iter=3, size=0.632, cols.nc=TRUE, steps=100, pointwise=TRUE, merr=0.0001, iter=100, nbiclust=10, col.overlap=FALSE)
BHeatmap(lung200, res1)
**Description**

The function performs biclustering of the data matrix by sparse singular value decomposition with nested stability selection.

**Usage**

```r
## S4 method for signature 'matrix,BCs4vd'
biclust(x, method=BCs4vd(),
  steps = 100,
  pcerv = 0.05,
  pceru = 0.05,
  ss.thr = c(0.6,0.65),
  size = 0.632,
  gamm = 0,
  iter = 100,
  nbiclust = 10,
  merr = 10^-4),
  cols.nc=FALSE,
  rows.nc=TRUE,
  row.overlap=TRUE,
  col.overlap=TRUE,
  row.min=4,
  col.min=4,
  pointwise=TRUE,
  start.iter=0,
  savepath=FALSE)
```

**Arguments**

- **x** The matrix to be clustered.
- **method** calls the BCs4vd() method
- **steps** Number of subsamples used to perform the stability selection.
- **pcerv** Per comparison wise error rate to control the number of falsely selected right singular vector coefficients (columns/samples).
- **pceru** Per comparison wise error rate to control the number of falsely selected left singular vector coefficients (rows/genes).
- **ss.thr** Range of the cutoff threshold (relative selection frequency) for the stability selection.
- **size** Size of the subsamples used to perform the stability selection.
gamm  Weight parameter for the adaptive LASSO, nonnegative constant (default = 0, LASSO).
iter  Maximal number of iterations to fit a single bicluster.
nbiclust  Maximal number of biclusters.
nerr  Threshold to decide convergence.
cols.nc  Allow for negative correlation of columns (samples) over rows (genes).
rows.nc  Allow for negative correlation of rows (genes) over columns (samples).
row.overlap  Allow rows to overlap between biclusters.
col.overlap  Allow columns to overlap between biclusters.
row.min  Minimal number of rows.
col.min  Minimal number of columns.
pointwise  If TRUE performs a fast pointwise stability selection instead of calculating the complete stability path.
start.iter  Number of starting iterations in which the algorithm is not allowed to converge.
savepath  Saves the stability path in order plot the path with the stabpathplot function. Note that pointwise needs to be TRUE to save the path. For extreme high dimensional data sets (e.g. the lung cancer example) the resulting biclust object may exceed the available memory.

Value

Returns an object of class Biclust.

Author(s)

Martin Sill \<m.sill@dkfz.de>

References

Martin Sill, Sebastian Kaiser, Axel Benner and Annette Kopp-Schneider "Robust biclustering by sparse singular value decomposition incorporating stability selection", Bioinformatics, 2011

See Also

biclust, Biclust

Examples

# example data set according to the simulation study in Lee et al. 2010
# generate artificial data set and a correspondig biclust object
u <- c(10,9,8,7,6,5,4,3,rep(2,17),rep(0,75))
v <- c(10,-10,8,-8,5,-5,rep(3,5),rep(-3,5),rep(0,34))
u <- u/sqrt(sum(u^2))
v <- v/sqrt(sum(v^2))
d <- 50
set.seed(1)
BCssvd

Biclustering via sparse singular value decomposition

Description

The function performs a biclustering of the data matrix by sparse singular value decomposition.

Usage

```r
## S4 method for signature 'matrix,BCssvd'
biclust(x, method=BCssvd(), K=10, threu = 1, threv = 1, gamu = 0, gamv = 0, u0 = svd(X)$u[,1], v0 = svd(X)$v[,1], merr = 10^(-4), niter = 100)
```

Arguments

- `x`: the matrix to be clustered
- `method`: calls the BCssvd() method
- `K`: number of SSVD-layers

Example:

```r
X <- (d*u%*%t(v)) + matrix(rnorm(100*50),100,50)
params <- info <- list()
RowxNumber <- matrix(rep(FALSE,100),ncol=1)
NumberxCol <- matrix(rep(FALSE,50),nrow=1)
RowxNumber[u!=$('#0,1] <- TRUE
NumberxCol[1,v!=$('#0] <- TRUE
Number <- 1
ressim <- BiclustResult(params,RowxNumber,NumberxCol,Number,info)

#perform s4vd biclustering
ress4vd <- biclust(X,method=BCs4vd,p cerv=0.5,p ceru=0.5,pointwise=FALSE,nbiclust=1, savepath=TRUE)
#perform s4vd biclustering with fast pointwise stability selection
ress4dpw <- biclust(X,method=BCs4vd,p cerv=0.5,p ceru=0.5,pointwise=TRUE,nbiclust=1)
#perform ssvd biclustering
resssvd <- biclust(X,BCssvd,K=1)
#agreement of the results with the simulated bicluster
jaccardind(ressim,ress4vd)
jaccardind(ressim,ress4dpw)
jaccardind(ressim,resssvd)
```
type of penalty (thresholding rule) for the left singular vector, 1 = (Adaptive) 
LASSO (default) 2 = hard thresholding

type of penalty (thresholding rule) for the right singular vector, 1 = (Adaptive) 
LASSO (default) 2 = hard thresholding

weight parameter in Adaptive LASSO for the left singular vector, nonnegative 
constant (default = 0, LASSO)

weight parameter in Adaptive LASSO for the right singular vector, nonnegative 
constant (default = 0, LASSO)

initial left singular vector

initial right singular vector

threshold to decide convergence

maximum number of iterations

Returns an Biclust object.

Adaptation of original code from Mihee Lee by Martin Sill <m.sill@dkfz.de>

Mihee Lee, Haipeng Shen, Jianhua Z. Huang and J. S. Marron "Biclustering via Sparse Singular 
Value Decomposition", Biometrics, 2010

biclust, Biclust

# example data set according to the simulation study in Lee et al. 2010
# generate artificial data set and a correspondig biclust object
u <- c(10,9,8,7,6,5,4,3,rep(2,17),rep(0,75))
v <- c(10,-10,8,-8,5,-5,rep(3,5),rep(-3,5),rep(0,34))
u <- u/sqrt(sum(u^2))
v <- v/sqrt(sum(v^2))
d <- 50
set.seed(1)
X <- (d*u%*%t(v)) + matrix(rnorm(100*50),100,50)
params <- info <- list()
RowxNumber <- matrix(rep(FALSE,100),ncol=1)
NumberxCol <- matrix(rep(FALSE,50),nrow=1)
RowxNumber[u!=0,1] <- TRUE
NumberxCol[1,v!=0] <- TRUE
Number <- 1
ressim <- BiclustResult(params,RowxNumber,NumberxCol,Number,info)
# perform s4vd biclustering
ress4vd <- biclust(X, method=BCs4vd, pceru=0.5, pceru=0.5, pointwise=FALSE, nbiclust=1, savepath=TRUE)
# perform s4vd biclustering with fast pointwise stability selection
ress4vdpw <- biclust(X, method=BCs4vd, pceru=0.5, pceru=0.5, pointwise=TRUE, nbiclust=1)
# perform ssvd biclustering
resssvd <- biclust(X, BCssvd, K=1)
# agreement of the results with the simulated bicluster
jaccardind(ressim, ress4vd)
jaccardind(ressim, ress4dpw)
jaccardind(ressim, resssvd)

---

**jaccardmat**  
**jaccard matrix**

### Description
The function calculates the pairwise jaccard coefficients between the biclusters of two biclustering results.

### Usage
```
jaccardmat(res1, res2)
```

### Arguments
- `res1`: A biclustering result as an object of class Biclust
- `res2`: A biclustering result as an object of class Biclust

### Details
The result is matrix of pairwise jaccard coefficients between the biclusters of `res1` and `res2`.

### Author(s)
Martin Sill \<m.sill@dkfz.de>\

### See Also
- `jaccardind`
Examples

```r
#lung cancer data set  Bhattacharjee et al. 2001
data(lung200)
set.seed(12)
res1 <- biclust(lung200, method=BCSvd(), pcer=.5, pceru=.01, ss.thr=c(0.6, 0.65),
                start.iter=3, size=0.632, cols.nc=TRUE, steps=100, pointwise=TRUE,
                merr=0.0001, iter=100, nbicluster=10, col.overlap=FALSE)
res2 <- biclust(lung200, method=BCPlaid())
jaccardmat(res1, res2)
```

Description

Lung cancer gene expression data set

Usage

data(lung200)

Format

This data set contain 56 samples and gene expression values of a subset of 200 genes showing the highest variance of the 12,625 genes measured using the Affymetrix 95av2 GeneChip. The samples comprise 20 pulmonary carcinoid samples (Carcinoid), 13 colon cancer metastasis samples (Colon), 17 normal lung samples (Normal) and 6 small cell lung carcinoma samples (SmallCell). The rownames are affymetrix gene ids.

Source

http://www.pnas.org/content/98/24/13790/suppl/DC1

References

**stabpath**

---

**Stability paths plot**

**Description**

The function plots the stability path of a S4VD result.

**Usage**

```r
stabpath(res, number)
```

**Arguments**

- `res`: the S4VD result
- `number`: the bicluster for which the stability path shall be plotted

**Details**

Plots the stability path for the rows and the columns regarding the last iteration of the s4vd algorithm. Note that if the pointwise error control was used or if savepath=FALSE the final selection probabilities for the rows and the columns will be plotted.

**Author(s)**

Martin Sill <m.sill@dkfz.de>

**Examples**

```r
# example data set according to the simulation study in Lee et al. 2010
# generate artificial data set and a correspondig biclust object
u <- c(10, 9, 8, 7, 6, 5, 4, 3, rep(2, 17), rep(0, 75))
v <- c(10, -10, 8, -8, 5, -5, rep(3, 5), rep(-3, 5), rep(0, 34))
u <- u/sqrt(sum(u^2))
v <- v/sqrt(sum(v^2))
d <- 50
set.seed(1)
X <- (d*u**%*%(v)) + matrix(rnorm(100*50), 100, 50)
params <- info <- list()
RowxNumber <- matrix(rep(FALSE, 100), ncol=1)
NumberxCol <- matrix(rep(FALSE, 50), nrow=1)
RowxNumber[u!=0,1] <- TRUE
NumberxCol[1,v!=0] <- TRUE
Number <- 1
ressim <- BiclustResult(params, RowxNumber, NumberxCol, Number, info)
#perform s4vd biclustering
ress4vd <- biclust(X, method=BCs4vd, pcer=0.5,
pceru=0.5, ss.thr=c(0.6, 0.65), steps=500,
pointwise=FALSE, nbiclust=1, savepath=TRUE)
```
#perform s4vd biclustering with fast pointwise stability selection

```
ress4vd<-.biclust(X,method=BCs4vd,pcer=0.5,
pceru=0.5,ss.thr=c(0.6,0.65),steps=500,
pointwise=TRUE,nbiclust=1)
```

#stability paths

```
stabpath(ress4vd,1)
```

#selection probabilities for the pointwise stability selection

```
stabpath(ress4vd<.,1)
```
Index

*Topic **biclustering**
  BCs4vd, 3
*Topic **classif**
  BCssvd, 5
*Topic **cluster**
  BCssvd, 5
*Topic **datasets**
  lung200, 8
*Topic **stability selection**
  BCs4vd, 3
*Topic **svd**
  BCs4vd, 3
  BCssvd, 5

BCheatmap, 2
BCs4vd, 3
BCs4vd-class (BCs4vd), 3
BCssvd, 5
BCssvd-class (BCssvd), 5
Biclust, 4, 6
biclust, 4, 6

jaccardmat, 7

lung cancer microarray data set200
  (lung200), 8
lung200, 8

OverlapHeat (BCheatmap), 2

s4vd (BCs4vd), 3
ssvd (BCssvd), 5
stabilitypath (stabpath), 9
stabpath, 9