

Package ‘flowPeaks’

April 24, 2012

Version 0.8

Date 1-12-2012

Title An R package for flow data clustering

Author Yongchao Ge<yongchao.ge@gmail.com>

Maintainer Yongchao Ge<yongchao.ge@gmail.com>

Depends R (>= 2.12.0)

Description A fast and automatic clustering to classify the cells into subpopulations based on finding the peaks from the overall density function generated by K-means.

bioViews Flow cytometry, Clustering, Gating, Bioinformatics

License Artistic-1.0

R topics documented:

adjust.flowPeaks	2
assign.flowPeaks	2
barcode	3
concave	4
evalCluster	4
flowPeaks	5
plot.flowPeaks	7
print.flowPeaks	8
summary.flowPeaks	8
Index	10

adjust.flowPeaks	<i>Adjusting the smoothing and mering behavior of the flowPeaks results</i>
------------------	---

Description

Adjusting the smoothing and merging behavior of the flowPeaks results by changing the multipliers of the covariance matrix and the tolerance level for joining two peaks

Usage

```
adjust.flowPeaks(object,tol,h0,h,...)
```

Arguments

object	The output from the function flowPeaks
tol	See flowPeaks
h0	See flowPeaks
h	See flowPeaks
...	Optional additional arguments. At present no additional arguments are used.

Value

It returns an updated object of class flowPeaks, the details of which can be seen in [flowPeaks](#).

Author(s)

Yongchao Ge <yongchao.ge@gmail.com>

See Also

[flowPeaks](#)

assign.flowPeaks	<i>Obtain the flowPeaks cluster id with the option of identify outliers and to a new data set</i>
------------------	---

Description

The function takes a flowPeaks output and a new data set (or could be the same dataset that generated the flowPeaks), and compute the cluster label assignment

Usage

```
assign.flowPeaks(fp,A,tol=0.01,fc=0.8)
```

Arguments

fp	the output from the function flowPeaks
A	A data matrix with the same number of columns as the data that generated fp
tol	All points that have density is less than tol (default is 1%) of the peak density of that cluster are labeled as outliers. If tol is set 0, no outliers according to this rule. The details can be seen in the first equation of the section 2.5 in the flowPeaks manuscript (Ge et al 2012)
fc	All points that have less than fc (default is 80%) of overall density that is contributed from the classified cluster are labeled as outliers. If fc is set to 100%, no outliers can be found according to this rule. The details can be seen in the second equation of the section 2.5 in the flowPeaks manuscript (Ge et al 2012)

Value

It returns the class label assignment of each data point, where -1 indicates outliers. If tol and fc are both set to 1. The returned labels are the same as `fp$peaks.cluster`.

Author(s)

Yongchao Ge <yongchao.ge@gmail.com>

References

Ge Y. et al, flowPeaks: a fast unsupervised clustering for flow cytometry data via K-means and density peak finding, Manuscript, 2012.

See Also

[flowPeaks](#)

barcode

The barcode dataset

Description

A flowcytometry data that is used to barcode to measure many samples simultaneously

Usage

```
data(barcode)
```

Format

An object (barcode) of data frame with 180912 rows and 3 columns and a vector (barcode.cid) for the cluster membership with the manual gating.

Source

The data is a random subset of the full data set for Figure 3A of the paper, "Misty Mountain clustering: application to fast unsupervised flow cytometry gating, Sugar I. P. and Sealfon S. C., BMC Bioinformatics, 2010, 11:502". This subset was used to do all comparisons in the paper with other clustering algorithms.

References

Sugar I. P. and Sealfon S. C., Misty Mountain clustering: application to fast unsupervised flow cytometry gating, *BMC Bioinformatics*, 2010, 11:502.
 Ge Y. et al, flowPeaks: a fast unsupervised clustering for the flow cytometry data via K-means and density peak finding, Manuscript, 2012.

concave

The concave dataset

Description

A simulated flowcytometry data with two concave shapes

Usage

```
data(concave)
```

Format

An object (concave) of data frame with rows and 3 columns and a vector (concave.cid) for the true membership

References

Ge Y. et al, flowPeaks: a fast unsupervised clustering for the flow cytometry data via K-means and density peak finding, Manuscript, 2012.

evalCluster

evaluate the clustering results to be compared with the gold standard

Description

This function takes the cluster lables of the two clustering, one is based on the gold standard, the other is a candidate clusterign, and compute one of the three metrics to assess the candidate clustering performance.

Usage

```
evalCluster(gs,cand,method=c("Rand.index","Fmeasure","Vmeasure"),
            rm.gs.outliers=TRUE)
```

Arguments

gs	A integer-value vector of length n for the cluster label of the gold standard clustering, where negative numbers such as -1 is for the outliers
cand	A integer-value vector of length n for the cluster label of a candidate clustering, where -1 is for the outliers
rm.gs.outliers	Determining whether the outliers of the gold standard clustering should be used in the comparison
method	A single character to indicate which one of three metrics should be used to evaluate the clustering. The details are described in Ge (2012) and references mentioned in that paper

Rand.index The adjusted Rand.index

Fmeasure F-measure

Vmeasure V-measure

Author(s)

Yongchao Ge <yongchao.ge@gmail.com>

References

Ge Y. et al, flowPeaks: a fast unsupervised clustering for flow cytometry data via K-means and density peak finding, Manuscript, 2012.

See Also

[flowPeaks](#)

flowPeaks

Doing the flowPeaks analysis

Description

This is the core function in the flowPeaks package. It generates the output of the cluster and information associated with each cluster, which can be used by the function plot for visualization

Usage

```
flowPeaks(x,tol=0.1,h0=1,h=1.5)
```

Arguments

<code>x</code>	a data matrix for the flow cytometry data, it needs to have at least two rows, and the names for each column should be unique
<code>tol</code>	The tolerance (between 0 and 1) when neighboring clusters should be considered to be merged
<code>h0</code>	The multiplier of the vaiarance matrix S_0
<code>h</code>	The multiplier of the variance matrix S

Value

It returns an object of class `flowPeaks`, which is a list of the following variables:

<code>peaks.cluster</code>	An integer shows the cluster id (between 1 and K for K clusters) for each cell. The clustering is based on the <code>flowPeaks</code> algorithm
<code>peaks</code>	A summary of the cluster information. It is also a list with the following three variables: <ul style="list-style-type: none"> • <code>cid</code>: cluster id, should always be 1:K; • <code>w</code>: the weights of the cluster 1 to cluster K; • <code>mu</code>: The mean of all cells in the K clusters; • <code>S</code>: The variance matrix of the K clusters. Note each variance matrix for each cluster has been stacked as a column vector
<code>kmeans.cluster</code>	An integer shows the cluster id for the initial kmeans clustering
<code>kmeans</code>	A summary of the initial kmeans clustering. The meaning of the variables can be seen in the description of <code>peaks</code> above
<code>info</code>	The information that can be used for plot, and how the initial kmeans clustering and the final <code>flowPeaks</code> clustering are connected
<code>x</code>	The input data <code>x</code>

Author(s)

Yongchao Ge <yongchao.ge@gmail.com>

References

Ge Y. et al, `flowPeaks`: a fast unsupervised clustering for flow cytometry data via K-means and density peak finding, Manuscript, 2012.

See Also

[plot.flowPeaks](#)

Examples

```
data(barcode)
fp<-flowPeaks(barcode[,c(1,3)])
plot(fp)
##to compare it with the gold standard
evalCluster(barcode.cid,fp$peaks.cluster,method="Vmeasure")
#to remove the outliers
fpc<-assign.flowPeaks(fp,fp$x)
plot(fp,classlab=fpc,drawboundary=FALSE,
     drawvor=FALSE,drawkmeans=FALSE,drawlab=TRUE)
```

plot.flowPeaks

Plot the results generated by flowPeaks

Description

This function takes the results generated from flowPeaks as an input, and plot the data in 2D. These plots display the clustering structure

Usage

```
## S3 method for class 'flowPeaks'
plot(x,idx=c(1,2),drawlab=FALSE,
     cols=c("red","green3","blue","cyan","magenta","yellow","gray"),drawvor=TRUE,
          drawlocalpeaks=FALSE,drawkmeans=TRUE,drawboundary=TRUE,
          classlab, negcol, negpch,...)
```

Arguments

x	The output from the function flowPeaks
idx	The index of the columns will be used to plot the clustering. idx needs to be at least length 2, and have no duplicate elements, and the values can only take from 1 to d, where d is the number of columns for the input matrix x that is used as an input of the function flowPeaks
drawlab	The option to decide whether we should draw the cluster labels
cols	The color specification for plotting the points in each cluster. Please note, "white" and "black" are not allowed, which are reserved for other purpose
drawvor	Deciding whether the voronoi diagram should be drawn, only good for 2D data
drawlocalpeaks	Deciding whether the local peaks with a triangle symbol should be drawn
drawkmeans	Deciding whether the kmeans center with a filled circle should be drawn
drawboundary	Deciding whether the boundary between clusters should be drawn, only good for 2D data
classlab	Use this to replace the default class labels from x\$peak.cluster, for example, the classlab may come from assignflowPeaks
negcol	Deciding the color of the negative, which are outliers
negpch	Deciding the symbols for the outliers
...	Optional additional arguments. At present no additional arguments are used.

Author(s)

Yongchao Ge <yongchao.ge@gmail.com>

See Also

[flowPeaks](#)

print.flowPeaks	<i>The display of the flowPeaks results</i>
-----------------	---

Description

The display of the flowPeaks results

Usage

```
## S3 method for class 'flowPeaks'  
print(x,...)
```

Arguments

x	The output from the function flowPeaks
...	Optional additional arguments. At present no additional arguments are used.

Author(s)

Yongchao Ge <yongchao.ge@gmail.com>

See Also

[flowPeaks](#)

summary.flowPeaks	<i>The summary of the flowPeaks results</i>
-------------------	---

Description

The summary of the flowPeaks results

Usage

```
## S3 method for class 'flowPeaks'  
summary(object,...)
```


Arguments

object	The output from the function flowPeaks
...	Optional additional arguments. At present no additional arguments are used.

Author(s)

Yongchao Ge <yongchao.ge@gmail.com>

See Also

[flowPeaks](#)

Index

- *Topic **adjust**
 - adjust.flowPeaks, [2](#)
- *Topic **cluster,multivariate,smooth**
 - flowPeaks, [5](#)
- *Topic **cluster,multivarite**
 - assign.flowPeaks, [2](#)
- *Topic **cluster,utiliteis**
 - evalCluster, [4](#)
- *Topic **datasets**
 - barcode, [3](#)
 - concave, [4](#)
- *Topic **plot,multivariate,smooth**
 - plot.flowPeaks, [7](#)
- *Topic **print**
 - print.flowPeaks, [8](#)
- *Topic **summary**
 - summary.flowPeaks, [8](#)

adjust.flowPeaks, [2](#)
assign.flowPeaks, [2](#)

barcode, [3](#)

concave, [4](#)

evalCluster, [4](#)

flowPeaks, [2](#), [3](#), [5](#), [5](#), [7–9](#)

plot.flowPeaks, [6](#), [7](#)
print.flowPeaks, [8](#)

summary.flowPeaks, [8](#)