

Package ‘flowPeaks’

May 17, 2012

Version 0.99.0

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)

Enhances flowCore

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.

biocViews 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	9
Index	10

adjust.flowPeaks	<i>Adjusting the smoothing and merging 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 detail definition 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 labels with the option of identifying outliers and applying 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	an object of class flowPeaks, the output from the function flowPeaks or adjust.flowPeaks
A	A data matrix with the same number of columns as the data that generated fp
tol	All points where the probability 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 Section 2.5 in the flowPeaks manuscript (Ge et al 2012)
fc	All points where the classified cluster contributes less than fc (default is 80%) of overall density are labeled as outliers. If fc is set to 0%, no outliers can be found according to this rule. The details can be seen in the second equation of 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. When A is the same data that generated fp, If tol is 1 and fc is 0, 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, 2012, Bioinformatics, in press.

See Also

[flowPeaks](#)

barcode

The barcode dataset

Description

A flow cytometry data that is used 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 labels according to the manual gating.

Source

The data is a random subset of the full data set for Figure 3A of the paper (Sugar et al 2010), This subset was used to do all comparisons in the paper (Ge et al 2012) 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 flow cytometry data via K-means and density peak finding, 2012, Bioinformatics, in press

concave	<i>The concave dataset</i>
---------	----------------------------

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 cluster labels.

References

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

evalCluster	<i>evaluate the result of a clustering algorithm by comparing it with the gold standard</i>
-------------	---

Description

This function takes the cluster labels of the two clusterings, 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

<code>gs</code>	A integer-valued vector of length <code>n</code> for the cluster labels of the gold standard clustering, where negative numbers such as -1 is for the outliers
<code>cand</code>	A integer-valued vector of length <code>n</code> for the cluster label of a candidate clustering, where -1 is for the outliers
<code>rm.gs.outliers</code>	Determining whether the outliers of the gold standard clustering should be removed in the comparison
<code>method</code>	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, 2012, Bioinformatics, in press.

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

x	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. For a flowFrame data, its expression matrix slot should be used as x, where only channels of interest are selected (see the example below).
tol	The tolerance (between 0 and 1) when neighboring clusters should be considered to be merged
h0	The multiplier of the variance matrix S0
h	The multiplier of the variance matrix S

Value

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

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

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, 2012, Bioinformatics, in press.

See Also

[plot.flowPeaks](#)

Examples

```
##demonstrate how to use a flowFrame
## Not run:
require(flowCore)
samp <- read.FCS(system.file("extdata","0877408774.B08",
package="flowCore"))
##do the clustering based on the asinh transforamtion of
##the first two FL channels
fp<-flowPeaks(asinh(samp@exprs[,3:4]))
plot(fp)

## End(Not run)

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)

#to adjust the cluster by increasing the tol,h0, h, which results
#in a smaller number of clusters
fp2<-adjust.flowPeaks(fp,tol=0.5,h0=2,h=2)
summary(fp2)
print(fp) #an alternative of using summary(fp)
```

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	An object of class flowPeaks, e.g., the output from the functions flowPeaks or adjust.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 assign.flowPeaks
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, [9](#)

adjust.flowPeaks, [2](#), [3](#), [8](#)
assign.flowPeaks, [2](#), [8](#)

barcode, [3](#)

concave, [4](#)

evalCluster, [4](#)

flowPeaks, [2](#), [3](#), [5](#), [5](#), [8](#), [9](#)

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

summary.flowPeaks, [9](#)