Package 'flowMap'

April 9, 2015

Type Package

Title Mapping cell populations in flow cytometry data for cross-sample comparisons using the Friedman-Rafsky Test

Version 1.4.0

Author Chiaowen Joyce Hsiao, Yu Qian, and Richard H. Scheuermann

Maintainer Chiaowen Joyce Hsiao <joyce.hsiao1@gmail.com>

Description flowMap quantifies the similarity of cell populations across multiple flow cytometry samples using a nonparametric multivariate statistical test. The method is able to map cell populations of different size, shape, and proportion across multiple flow cytometry samples. The algorithm can be incorporate in any flow cytometry work flow that requires accurat quantification of similarity between cell populations.

License GPL (>=2)

VignetteBuilder knitr

Depends R (>= 3.0.1), ade4(>= 1.5-2), doParallel(>= 1.0.3), abind(>= 1.4.0), reshape2(>= 1.2.2), ggplot2(>= 0.9.3.1), igraph (>= 0.7.1), scales(>= 0.2.3), Matrix(>= 1.1-4), gplots(>= 2.14.1), methods (>= 2.14)

Suggests BiocStyle, knitr

LazyData true

biocViews MultipleComparison, FlowCytometry

R topics documented:

`flowMap-package` *Flow cytometry data cross-sample comparison*

Description

This package implements a method for matching cell populations across multiple flow cytometry samples.

Author(s)

Chiaowen Hsiao <joyce.hsiao1@gmail.com>

`FRstats-class` *FR statistics and p-values generated from one single draw of the population pair comparison*

Description

This class stores the statistics required to compute median FR statistics across random draws.

FIXME

usage ## Accessors getFRstats(object) getPnorm(object)

Author(s)

Chiaowen Joyce Hsiao <joyce.hsiao1@gmail.com>

Examples

```
## see vignettes
```

getFR *FR test for a single cell population pair comparison*

Description

Compute FR statistic for any two cell populations in flow cytometry data. Runtime of the FR test is quadratic in the number of events (nodes) totaling a single population pair comprison.

Usage

```
getFR(xx1, xx2)
```

Arguments

xx1	events of a single cell population, organized in a matrix or a data.frame of events (rows) by features (columns).
xx2	events of a single cell population, organized in a matrix or a data.frame of events (rows) by features (columns).

Value

ww FR statistic.

runs number of within-group subtrees (large number of runs indicate high degree of dissimilarity between the two cell populations being compared.

mu expected number of runs (when the two cell populations are similarly distributed).

sigma2 variance of runs.

pNorm p-values of the FR statitic assuming large sample asymptotic normal assumption.

Author(s)

Chiaowen Joyce Hsiao <joyce.hsiao1@gmail.com>

Examples

```
## see vignettes
```

getFRest *FR tests to compare two flow cytometry samples*

Description

Estimate FR statistics comparing cell populations across two flow cytometry samples. The estimates are generated from getFRmat and are dependent on the number of random draws as well as the size of each random sample. For every cell population pair comparison, we take the median of the FR statistics across all random samples as the estimated similarity. Parallel computing is used to minimize the runtime of the algorithm (**doParallel**). Users can specify the number of cores to be used depending on the available computing power. Default uses all available processing cores in the system.

Visualize minimum spanning tree (MST) of the pooled data combined from the two cell populations. MST is the basis of the FR statistics. Two populations are similar if their respective events congregate with events of the same population membership. Runs is calculated as the number of edges connecting nodes of different cell population membership plus 1 ((or equivalently, the number of subtrees of homogeneous cell population membership).

Usage

```
getFRest(XX1, XX2, sampleMethod = "proportional", sampleSize = 200,
  estStat = "median", ndraws = 200, ncores = NULL)

makeFRMST(mat, node.colors = c("blue", "red"))
```

Arguments

XX1	a flow cytometry sample of cell populations, organized in a matrix or a data.frame of events (rows) by features (columns) where cell populaiton memberships are indexed in the last column by a variable named id.
XX2	a flow cytometry sample of cell populations, organized in a matrix or a data.frame of events (rows) by features (columns) where cell populaiton memberships are indexed in the last column by a variable named id.
ndraws	number of random samples. Runtime is linear in the number of random samples. (default: 200)
sampleMethod	downsampling method, options include *equalSize* or *proportional*. Both methods sample events without replacement from the combined events in a single cell population pair comparison. Using *equalSize*, each sample includes an equal number of events from the two cell populations being compared. Using *proportional*, the ratio of the event membership is same as the ratio of event membershiop prior to sampling. (default: *proportional*)
sampleSize	specifies *S*,the number of events to be included in each sample. For *equalSize* sampling , S/2 is sampled from each population. For *proportional* sampling, the ratio of event membership is the same as the ratio of event membership prior to sampling. (default: 200)

estStat	statistic that used to estimate FR statistic of each population pair comparison across random samples. (default: median)
mat	a matrix or data.frame of the pooled data of a cell population pair. The rows contain events in the pooled data. The columns are the expression markers. The population membership is indexed in the last column of the matrix or data.frame. The membership IDs are for plotting purposes, and hence need to uniquely identify the two cell populations in the pooled data.
node.colors	colors to label cell population membership of the events. (default: c("blue","red"))

Value

wmat a matrix of estimated FR statistics for each XX1 by XX2 population comparisons.

runsmat a matrix of estimated runs for each XX1 by XX2 population comparisons.

mumat a matrix of estimated expected number of runs for each XX1 by XX2 population comparisons.

sigma2mat a matrix of estimated variance of runs for each XX1 by XX2 population comparisons.

pNormat a matrix of one-sided p-values of the estimated FR statistic for each XX1 by XX2 population comparisons under the asymptotic normality assumption of the FR statistic.

g an *igraph* object that contains graph adjacency matrix of the minimum spanning tree calculated from euclidean distance between nodes in the pooled data combined from the two cell populations.

gall an *igraph* object that contains graph adjacency matrix containing euclidean distance between nodes in the pooled data from the two cell population. Complete graph is assumed here.

ww FR statistics for the cell population comparison (formula: (runs-mu)/sqrt(sigma2))

runs observed number of runs for the cell population comparison

mu expected number of runs for the cell population comparison

sigma2 variance of runs for the cell population comparison

pNorm p-value of the FR statistic for the cell population comparison

distmat a matrix defining Euclidean distance between events (nodes) across the two cell populations

mstree minimum spannin tree defined in an adjacency matrix. 1 indicates an edge between nodes and 0 indicates non-existent edge.

C number of edge pairs (pairs of edges that share a common node) in the minumum spanning tree

m number of events in the first cell population

n number of events in the second cell population

Author(s)

Chiaowen Joyce Hsiao <joyce.hsiao1@gmail.com>

Chiaowen Joyce Hsiao <joyce.hsiao1@gmail.com>

Examples

```
## see vignettes
## see vignettes
```

getFRmat *FR tests to compare two flow cytometry samples*

Description

Perform FR tests to compare cell populations across two flow cytometry samples. The FR statistics are estimated based on a downsampling scheme designed to optimize runtime as well as precision and accuracy of the statistics. The downsampling scheme samples from the pooled data of events across a single cell population comparison. The events in the sample maintain the same cell population membership ratio as the events in the pooled data (*proportional*) or include an equal number of events from the two cell populations in the comparison.

Usage

```
getFRmat(XX1, XX2, sampleMethod, sampleSize, i = NULL)
```

Arguments

XX1	a flow cytometry sample of cell populations, organized in a matrix or a data.frame of events (rows) by features (columns) where cell populaiton memberships are indexed in the last column by a variable named id.
XX2	a flow cytometry sample of cell populations, organized in a matrix or a data.frame of events (rows) by features (columns) where cell populaiton memberships are indexed in the last column by a variable named id.
sampleMethod	downsampling method, options include *equalSize* or *proportional*. Both methods sample events without replacement from the combined events in a single cell population pair comparison. Using *equalSize*, each sample includes an equal number of events from the two cell populations being compared. Using *proportional*, the ratio of the event membership is same as the ratio of event membershiop prior to sampling. (default: *proportional*)
sampleSize	specifies *S*,the number of events to be included in each sample. For *equalSize* sampling , S/2 is sampled from each population. For *proportional* sampling, the ratio of event membership is the same as the ratio of event membership prior to sampling.
i	dummy variable to initialize parallel computing (**doParallel**).

Value

wmat a list of matrices containing sample FR statistics for each XX1 by XX2 population comparisons.

runsmat a list of matrices containing sample runs for each XX1 by XX2 population comparisons.

mumat a list of matrices containing sample expected number of runs for each XX1 by XX2 population comparisons.

sigma2mat a list of matrices containing sample estimated variance of runs for each XX1 by XX2 population comparisons.

pNormat a list of matrices containing sample one-sided p-value associated with the FR statistics of XX1 by XX2 population comparisons under the asymptotic normality assumption of the FR statistic.

Author(s)

Chiaowen Joyce Hsiao <joyce.hsiao1@gmail.com>

`makeDistmat` *Generate a similarity matrix quantifying similarity between cell populations across flow cytmetry samples*

Description

Perform `getFRest` for multiple sample comparisons. For a two-sample comparison of n1 and n2 cell populations, we generate a similarity matrix of dimension (n1+n2)-by-(n1+n2).

Usage

```
makeDistmat(samples, sampleMethod = "proportional", sampleSize = 200,
  ndraws = 200)
```

Arguments

`samples`	a list of data.frames or matrices of flow cytometry samples. Each data.frame or matrix consists of columns of features (flow cytometry channels), followed by a column of population membership (*id*).
`sampleMethod`	downsampling method, options include *equalSize* or *proportional*. Both methods sample events without replacement from the combined events in a single cell population pair comparison. Using *equalSize*, each sample includes an equal number of events from the two cell populations being compared. Using *proportional*, the ratio of the event membership is same as the ratio of event membershiop prior to sampling. (default: *proportional*)
`sampleSize`	specifies *S*,the number of events to be included in each sample. For *equalSize* sampling , S/2 is sampled from each population. For *proportional* sampling, the ratio of event membership is the same as the ratio of event membership prior to sampling. (default: 200)
`ndraws`	number of random samples. Runtime is linear in the number of random samples. (default: 200)

Value

distmat a matrix of estimated FR statistics across multiple samples. Rows and columns are indexed in the order in which each sample appears in the input list. For example, a two sample comparison entails rows and columns of *distmat* indexed as 1.x follwed by 2.y where x is the Sample 1 population IDs, and 2.y is the Sample 2 population IDs.

Author(s)

Chiaowen Joyce Hsiao <joyce.hsiao1@gmail.com>

`statCrossLists` *summary statistic across elements of lists*

Description

Compute an estimate of F-R statistics across random draws of cell population comparison.

Usage

```
statCrossLists(obj, STAT)
```

Arguments

`obj`	an object of one or more lists. List elements must be matrices of sample dimensions
`STAT`	the statistic to be computed for each entry in the matrix across lists. Default value: median.

Value

est a matrix of the estimated statistic across random draws.

Index