Package: rOCEAN (via r-universe)

February 11, 2025

Type Package

Title Two-Way Feature Set Testing for Multi-Omics

Version 1.0

Maintainer Mitra Ebrahimpoor <mitra.ebrahimpoor@gmail.com>

Description For any two way feature-set from a pair of pre-processed omics data, pairwise-TDP, Column-TDP and row-TDP are calculated. Due to embeded closed testing procedure, the choice of feature-sets can be changed infinite times and even after seeing the data withuot any change in type I error rate. For more details refer to the refrence article.

License GPL (>= 2)

Date 2024-10-29

Encoding UTF-8

RoxygenNote 7.3.2

Imports ff

Repository https://mitra-ep.r-universe.dev

RemoteUrl https://github.com/mitra-ep/rocean

RemoteRef HEAD

RemoteSha 4c44201690d1773e0a3f10a738fca4ba4241e933

Contents

corPs *Calculate pairwise p-value*

Description

Calculates pairwise matrix of p-values based on Pearson's correlation test for two matrices. To gain speed and manage RAM usage, the matrices are split into several smaller chunks.

Usage

```
corPs(pm1, pm2, type = c("Mat", "Vec"), pthresh = 0.05)
```

Arguments

pm1, pm2	Subsets of two omics data sets where rows are the features and columns are samples. The rows of the two matrices would define the two-way feature set of interest.
type	Two options are available. Mat: Calculate the correlation of subsets and return a matrix; Vec: calculate the correlation matrix, subset by the given threshold and return a vector of p-values.
pthresh	Only relevant for type="Vec". The threshold by which the p-values are filtered (p>pthresh is removed). Default value is 0.05.

Value

Either a matrix or vector of pairwise p-values, as indicated by type parameter.

getCat *Calculate p-categories*

Description

Calculates cumulative p-categories for a given matrix of p-values.

Usage

```
getCat(mps, gCT, scale = c("col", "row"))
```

Arguments

mps	Matrix of p-values, representing pairwise associations between two feature sets.
gCT	Parameters of the global closed testing, which is the output of simesCT function.
scale	Scale of the quantification, a character string. Possible choices are "col" and "row".

Value

Matrix of p-categories.

See Also

simesCT

`ocean` *OCEAN algorithm*

Description

Calculates heuristic and lower bound for the true discovery proportion (TDP) in 3 scales for a specified two-way feature set (Algorithm 1 in the reference). The input is either two omics data sub-matrices or the pre-calculated matrix of p-values for pairwise associations. In case the result is not exact, the function adopts branch and bound (Algorithm 2 in the reference), if `nMax` allows.

Usage

```
ocean(
  pm1,
  pm2,
  gCT,
  scale = c("pair", "row", "col"),
  mps,
  nMax = 100,
  verbose = TRUE
)
```

Arguments

`pm1`, `pm2`	Matrix; Subsets of two omics data sets where rows are the features and columns are samples. The rows of the two matrices would define the two-way feature set of interest.
`gCT`	Vector; Parameters of the global closed testing, output of simesCT function.
`scale`	Optional character vector; It specifies the scale of TDP quantification. Possible choices are "pair" (pair-TDP), "col" (col-TDP) and "row" (for row-TDP'). If not specified, all three scales are returned.
`mps`	Optional matrix of p-values; A sub-matrix of pairwise associations, representing the two-way feature set of interest. If provided, `pm1` and `pm2` are not required. If not provided, matrix of pairwise associations will be derived from `pm1` and `pm2` based on Pearson's correlation.
`nMax`	Maximum number of steps for branch and bound algorithm, if set to 1 branch and bound is skipped even if the result is not exact. The default value is a 100. The algorithm may stop before the `nMax` is reached if it converges sooner.
`verbose`	Logical; if `TRUE`, progress messages will be displayed during the function's execution. Default is `TRUE`.

Value

TDP is returned for the specified scales, along with number of steps taken and convergence status for branch and bound algorithm.

See Also

simesCT pairTDP runbab

Examples

```
#number of feature per omic data set
n_cols<-1000
n_rows<-1200

#random matrix of p-values
set.seed(1258)
pvalmat<-matrix(runif(n_rows*n_cols, min=0, max=1)^3, nrow=n_rows, ncol=n_cols)

#calculate CT parameters
gCT<-simesCT(mps=pvalmat, m=nrow(pvalmat)*ncol(pvalmat))

#calculate TDPs for a random feature set
subpmat<-pvalmat[1:400,100:750]
#Note: it can take loner to run this script if nMax is large
out<-ocean(mps=subpmat, gCT=gCT, nMax=2)
out
```

---

pairTDP *pairwise true discoveries proportion*

Description

Calculates the TDP over pairs; based on SEA algorithm

Usage

```
pairTDP(mps, n, gCT)
```

Arguments

mps	Matrix or vector of pairwise associations.
n	Number of pairs; may not be the size of p if a threshold is used to remove large p-values.
gCT	Parameters of the global closed testing, output of simesCT function.

Value

Proportion of true discoveries out of n pairs of features.

See Also

SEA, simesCT

`runbab` *Branch and bound algorithm implementation*

Description

Performs B&B when the bound are not exact

Usage

```
runbab(sCat, ssh, ssb, nMax = 100)
```

Arguments

`sCat`	Category matrix, output of getCat function
`ssh`	current Heuristic as provided by SingleStep function
`ssb`	current Bound as provided by SingleStep function
`nMax`	Maximum number of steps for the algorithm, the algorithm may stop sooner if it converges.

Value

A list, including the heuristic and the bound for the number of true discoveries, along with number of steps taken and convergence status.

See Also

getCat singleStep

`simesCT` *Closed testing with Simes*

Description

Calculates five parameters from closed testing with Simes local tests based on raw data. These parameter are unique per data/alpha-level combination and do not depend on feature sets. Calculation may be somewhat long depending on the size of data sets and PC configurations.

Usage

```
simesCT(om1, om2, mps, m, alpha = 0.05)
```

Arguments

om1, om2	Two omics data sets where rows are features and columns are samples.
mps, m	Optional, pre-calculated matrix/vector of pairwise associations and the size. To save time in calculation of parameters, a threshold such as the type I error may be applies to remove larger p-values. If a threshold is used, size of matrix and m will not match. m should always be the size of the matrix of associations (number of features in om1 X number of features in om2).
alpha	type I error rate, default value is 0.05.

Value

Vector of integers: grand H value, concentration p-value, size of concentration set z, size of the original pair-wise associations matrix and the type I error level used in calculations.

References

See more details in "Hommel's procedure in linear time" doi: 10.1002/bimj.201700316.

---

singleStep *Single step algorithm*

Description

Calculates heuristic and upper-bound for the number of true discoveries based on the Algorithm 1 introduced in paper.

Usage

```
singleStep(sCat, B)
```

Arguments

sCat	p-categories matrix, output of getCat function.
B	Optional, to identify rows to be fixed (1) or removed (0) while splitting the search space.

Value

A list of two objects, the heuristic and the lower bound for true number of discoveries

See Also

getCat

Index