Package 'ccml'

December 14, 2022

Type Package

Title Consensus Clustering for Different Sample Coverage Data

Version 1.2.0

Description
Consensus clustering, also called meta-clustering or cluster ensembles, has been increasingly used in clinical data. Current consensus clustering methods tend to ensemble a number of different clusters from mathematical replicates with similar sample coverage. As the fact of common variety of sample coverage in the real-world data, a new consensus clustering strategy dealing with such biological replicates is required. This is a two-step consensus clustering package, which is used to input multiple predictive labels with different sample coverage (missing labels).

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.2.1

Depends R (>= 3.5.0)

Imports ggplot2, diceR, parallel, tidyr, SNFtool, plyr, ConsensusClusterPlus (>= 1.56.0)

Suggests spelling, testthat (>= 3.0.0)

Language en-US

Config/testthat/edition 3

NeedsCompilation no

Author Chuanxing Li [aut, cre], Meng Zhou [aut]

Maintainer Chuanxing Li <chuan-xing.li@ki.se>

Repository CRAN

Date/Publication 2022-12-14 07:30:02 UTC

R topics documented:

`callNCW` *Calculate normalized consensus weight(NCW) matrix based on permutation.*

Description

Calculate normalized consensus weight(NCW) matrix based on permutation.

Usage

```
callNCW(
  title = "",
  label,
  nperm = 10,
  ncore = 1,
  seedn = 100,
  stability = TRUE,
  plot = NULL
)
```

Arguments

`title`	A character value for output directory. Directory is created only if not existed. This title can be an abosulte or relative path.
`label`	A matrix or data frame of input labels, columns=different clustering results and rows are samples.
`nperm`	A integer value of the permutation numbers, or nperm=10(default), which means `nperm`*1000 times of permutation.
`ncore`	A integer value of cores to use, or ncore=1 (default). It's the input core numbers for the parallel computation in this package `parallel`.
`seedn`	A numerical value to set the start random seed for reproducible results, or seedn=100 (default). For every 1000 iteration, the seed will +1 to get repeat results.
`stability`	A logical value. Should estimate the stability of normalized consensus weight based on permutation numbers (default stability=TRUE), or not?
`plot`	character value. NULL(default) - print to screen, 'pdf', 'png', 'pngBMP' for bitmap png, helpful for large datasets, or 'pdf'. Input for `randConsensusMatrix`.

Value

A matrix of normalized consensus weights.

Examples

```
# load data
data(example_data)
label=example_data

# if plot is not NULL, results will be saved in "result_output" directory
title="result_output"


# run ncw
ncw<-callNCW(title=title,label=label,stability=TRUE,nperm=4,ncore=1)
```

---

ccml *A two-step consensus clustering inputing multiple predictive labels with different sample coverages (missing labels)*

Description

A two-step consensus clustering inputing multiple predictive labels with different sample coverages (missing labels)

Usage

```
ccml(
  title,
  label,
  output = "rdata",
  nperm = 10,
  ncore = 1,
  seedn = 100,
  stability = TRUE,
  maxK = 15,
  reps = 1000,
  pItem = 0.9,
  plot = NULL,
  clusterAlg = "spectralClusteringAffinity",
  innerLinkage = "complete",
  ...
)
```

Arguments

title	A character value for output directory. Directory is created only if not existed. This title can be an abosulte or relative path. Input for callNCW, plotCompareCW, ConsensusClusterPlus::ConsensusClusterPlus, ConsensusClusterPlus::calcICL
label	A matrix or data frame of input labels or a character value of input file name, columns=different clustering results and rows are samples. label could be import as '.rdata', '.rda', or '.csv'. Input for callNCW, plotCompareCW
output	A character value for output format, or "rdata"(default) as save to .rdata when both output and plot are not NULL, others will return to workspace.
nperm	A integer value of the permutation numbers, or nperm=10(default), which means nperm*1000 times of permutation. Input for callNCW
ncore	A integer value of cores to use, or ncore=1 (default). It's the input core numbers for the parallel computation in this package parallel. Input for callNCW
seedn	A numerical value to set the start random seed for reproducible results, or seedn=100 (default). For every 1000 iteration, the seed will +1 to get repeat results. Input for callNCW, ConsensusClusterPlus::ConsensusClusterPlus
stability	A logical value. Should estimate the stability of normalized consensus weight based on permutation numbers (default stability=TRUE), or not? Input for callNCW
maxK	integer value. maximum cluster number to evaluate. Input for ConsensusClusterPlus::ConsensusClus for the consensus clustering based on normalized consensus weights.
reps	integer value. number of subsamples. Input for ConsensusClusterPlus::ConsensusClusterPlus
pItem	numerical value. proportion of items to sample. Input for ConsensusClusterPlus::ConsensusClusterP
plot	character value. NULL(default) - print to screen, 'pdf', 'png', 'pngBMP' for bitmap png, helpful for large datasets. Input for ConsensusClusterPlus::ConsensusClusterPlus, ConsensusClusterPlus::calcICL,callNCW,plotCompareCW
clusterAlg	character value. cluster algorithm. 'spectralClusteringAffinity' for spectral clustering of similarity/affinity matrix(default), other methods for clustering of distance matrix, 'hc' heirarchical (hclust), 'pam' for paritioning around medoids, 'km' for k-means upon data matrix, 'kmdist' for k-means upon distance matrices (former km option), or a function that returns a clustering. Input for ConsensusClusterPlus::ConsensusClusterPlus.
innerLinkage	heirarchical linkage method for subsampling, or "complete"(default). Input for ConsensusClusterPlus::ConsensusClusterPlus
...	Other input arguments for ConsensusClusterPlus::ConsensusClusterPlus

Value

A list of three items

- ncw - A matrix of normalized consensus weights. Output from callNCW.
- fcluster - A list of length maxK. Each element is a list containing consensusMatrix (numerical matrix), consensusTree (hclust), consensusClass (consensus class asssignments). ConsensusClusterPlus also produces images. Output from ConsensusClusterPlus::ConsensusClusterPlus
- icl a list of two elements clusterConsensus and itemConsensus corresponding to cluster-consensus and item-consensus. Output from ConsensusClusterPlus::ConsensusClusterPlus

Examples

```
# load data
data(example_data)
label=example_data

# if plot is not NULL, results will be saved in "result_output" directory
title="result_output"


# not estimate stability of permutation numbers.
res_1=ccml(title=title,label=label,nperm = 3,ncore=1,stability=FALSE,maxK=5,pItem=0.8)

# other methods for clustering of distance matrix
res_2<-ccml(title=title,label=label,nperm = 10,ncore=1,stability=TRUE,maxK=3,
            pItem=0.9,clusterAlg = "hc")

# set the start random seed
res_3<-ccml(title=title,label=label,output=FALSE,nperm = 5,ncore=1,seedn=150,stability=TRUE,maxK=3,
           pItem=0.9)
```

`example_data` *The input data for example*

Description

In this matrix, columns represent the results of different clustering results and rows are samples.

Usage

```
example_data
```

Format

A matrix with 10 rows and 5 columns.

`plotCompareCW` *Plot of original consensus weights vs. normalized consensus weights grouping by the number of co-appeared percent of clustering(non-missing).*

Description

Plot of original consensus weights vs. normalized consensus weights grouping by the number of co-appeared percent of clustering(non-missing).

Usage

```
plotCompareCW(title, label, ncw, plot = NULL)
```

Arguments

title	A character value for output directory.
label	A matrix or data frame of input labels, columns=different clustering results and rows are samples.
ncw	A matrix of normalized consensus weights with sample-by-sample as the order of sample(rows) in label.
plot	character value. NULL(default) - print to screen, 'pdf', 'png', 'pngBMP' for bitmap png, helpful for large datasets, or 'pdf'.

Value

A ggplot point in PDF format with x-axis: original consensus weights; y-axis: normalized consensus weights; color: percent of co-appeared of clustering; size: number of duplicates sample .

Examples

```
# load data
data(example_data)
label=example_data

# if plot is not NULL, results will be saved in "result_output" directory
title="result_output"


ncw<-callNCW(title=title,label=label,stability=TRUE)
plotCompareCW(title=title,label=label,ncw=ncw)
```

---

randConsensusMatrix *Calculate consensus weight matrix based on the permuted input label matrix. Internal function used by* callNCW

Description

Calculate consensus weight matrix based on the permuted input label matrix. Internal function used by callNCW

Usage

```
randConsensusMatrix(
  l.seed,
  l.label = label,
  l.ns = ns,
  l.nc = nc,
  l.nv = nv,
  l.index = index,
  l.pair.ind = pair.ind,
  l.ppath = ppath,
  l.plot = plot
)
```

Arguments

`l.seed`	A numerical value to set the random seed for reproducible results, 1000 random label matrix will be generated based on this seed number.
`l.label`	A matrix or data frame of input labels, columns=different clustering results and rows are samples.
`l.ns`	A integer value of number of samples, =`nrow(l.label)`
`l.nc`	A integer value of number of samples, =`ncol(l.label)`
`l.nv`	A integer vector of the number of non missing values for each column in `l.label`
`l.index`	A list of index with length of `l.nc` of non missing values for each column in `l.label`
`l.pair.ind`	A n-by-2 index matrix of array indices of upper triangular of `l.label` with non missing values
`l.ppath`	A character value for output directory.
`l.plot`	character value. NULL(default) - print to screen, 'pdf', 'png', 'pngBMP' for bitmap png, helpful for large datasets, or 'pdf'.

Value

A character of finished seed.

Write a binary file of 1000 random consensus weight matrix(as a vector n-by-1, n= nrow(`l.pair.ind`)) with the seed `l.seed`, output file name: paste0("s",`l.seed`,"rcw").

`spectralClusteringAffinity` *Perform spectral clustering algorithms for an affinity matrix, using SNFtool::spectralClustering.*

Description

Perform spectral clustering algorithms for an affinity matrix, using SNFtool::spectralClustering.

Usage

```
spectralClusteringAffinity(affi_matrix, k, type = 3)
```

Arguments

affi_matrix	A numerical similarity or affinity matrix.
k	A number value of clusters
type	The variants of spectral clustering to use. See SNFtool::spectralClustering

Value

A vector consisting of cluster labels of each sample.

Index