Package ‘tightClust’

June 12, 2018

Type Package
Title Tight Clustering
Version 1.1
Date 2018-06-12
Author George C. Tseng <ctseng@pitt.edu>, Wing H. Wong <whwong@stanford.edu>
Maintainer Chi Song <song.1188@osu.edu>
Depends R (>= 2.10.1), base, utils, stats
Description The functions needed to perform tight clustering Algorithm.
License GPL (>= 2)
NeedsCompilation no
Repository CRAN
Date/Publication 2018-06-12 21:09:18 UTC

R topics documented:

  tightClust-package .................................................. 1
  plot.tight.clust ..................................................... 2
  tclust.test.data ................................................... 3
  tight.clust ......................................................... 3
  write.tight.clust ................................................... 5

Index 7

tightClust-package  Tight Clustering Package

Description

  This package could perform tight clustering algorithm proposed by George C. Tseng and Wing H. Wong.

Details

1
plot.tight.clust

Package: tightClust
Type: Package
Version: 1.0
Date: 2012-08-28
License: GPL (>=2)

Author(s)
George C. Tseng <ctseng@pitt.edu>, Wing H. Wong <whwong@stanford.edu>
Maintainer: Chi Song <chs108@pitt.edu>

References

plot.tight.clust  Plot tight cluster result

Description
A function to plot the heatmap of the tight cluster result.

Usage
## S3 method for class 'tight.clust'
plot(x, standardize.gene = TRUE, order.sample = FALSE, plot.noise=TRUE, ...)

Arguments
x Return value of the tight.clust function
standardize.gene If standardize each gene vector to mean 0 and sd 1.
order.sample It specifies whether to order samples (features) using the hierachical clustering method.
plot.noise It specifies whether to plot the remaining noise genes (objects).
... Arguments to image.

Author(s)
Chi Song <chs108@pitt.edu>
References


---

tclust.test.data  

test data for tight clustering package

Description

Sample microarray data

Usage

data(tclust.test.data)

Format

The data is a list of 3 items:

GeneID  ID of each gene  
Annotation  Annotation information of each gene  
Data  Data matrix of gene expression: each row represent one gene; each column represent one sample

---

tight.clust  
Tight Clustering

Description

This function could perform the tight clustering algorithm.

Usage

tight.clust(x, target, k.min, alpha = 0.1, beta = 0.6,  
top.can = 7, seq.num = 2, resamp.num = 10,  
samp.p = 0.7, nstart = 1, remain.p = 0.1,  
k.stop = 5, standardize.gene=TRUE, random.seed=NULL)
Arguments

- **x**: Input data, should be a matrix. Each row represents a gene (object) to be clustered. Gene (object) names are usually represented in the row names and sample (feature) names are in the column names of the matrix.

- **target**: The total number of clusters that the user aims to find.

- **k.min**: The starting point of k0. See 'Details' for more information.

- **alpha**: The threshold of comembership index. Default value is suggested to be used.

- **beta**: The threshold of clusters stably found in consecutive k0. Default value is suggested to be used.

- **top.can**: The number of top (size) candidate clusters for a specific k0. Default value is suggested to be used.

- **seq.num**: The number of subsequent k0 that finds the tight cluster. Default value is suggested to be used.

- **resamp.num**: Total number of resampling to obtain comembership matrix. Default value is suggested to be used.

- **samp.p**: Percentage of subsamples selected. Default value is suggested to be used.

- **nstart**: Number of different random initial for K-means. Default value is suggested to be used.

- **remain.p**: Stop searching when the percentage of remaining points <= remain.p. Default value is suggested to be used.

- **k.stop**: Stop decreasing k0 when k0<=k.stop. Default value is suggested to be used.

- **standardize.gene**: It specifies whether to standardize each gene vector to mean 0 and sd 1. Default value is suggested to be used.

- **random.seed**: If random.seed is NULL no random seed will be set. If random.seed is a number, it will be used as the random seed. This parameters should be used to get the same result for different runs.

Details

Tight clustering method is a resampling-evaluated clustering method that aims to directly identify tight clusters in a high-dimensional complex data set and allow a set of scattered objects without being clustered. The method was originally developed for gene cluster analysis in microarray data but can be applied in any complex data. The most important parameter is k.min. A large k.min results in smaller and tighter clusters. Normally k.min>=target+5 is suggested. All other parameters do not affect the quality of final clustering results too much and are suggested to remain unchanged.

Value

Returned value is a "tight.clust" object (list). The first element is the original data matrix. The second element is a vector representing the cluster identity (-1: scattered gene set; 1: the first cluster; 2: the second cluster; ...). The third element is a vector of the size of each tight cluster.

Author(s)

Chi Song <chs108@pitt.edu>
References

Examples

```r
## load the test dataset
data(tclust.test.data)
## find 10 tight clusters
ptm<-proc.time()
## k.min=25, tighter clusters will be found
## target=1 is used to save time, target=10 is recommended
tclust1<-tight.clust(tclust.test.data$Data, target=1, k.min=25, random.seed=12345)
proc.time()-ptm
## plot the heat map of cluster result
plot(tclust1)
## write the cluster result
write.tight.clust(tclust1)
ptm<-proc.time()
## k.min=10, looser clusters will be found
## target=1 is used to save time, target=5 is recommended
tclust2<-tight.clust(tclust.test.data$Data, target=1, k.min=10, random.seed=12345)
proc.time()-ptm
## plot the heat map of cluster result
plot(tclust2)
## write the cluster result
write.tight.clust(tclust2)
```

Description
A function to print the tight cluster result to a file or connection.

Usage

```r
write.tight.clust(x, ...)
```

Arguments

- `x` : Return value of the `tight.clust` function
- `...` : Arguments to `write.table`.

Author(s)
Chi Song <chs108@pitt.edu>
References

Index

*Topic cluster
  plot.tight.clust, 2
  tight.clust, 3
  write.tight.clust, 5
*Topic datasets
  tclust.test.data, 3
*Topic package
  tightClust-package, 1

plot.tight.clust, 2

tclust.test.data, 3
tight.clust, 3
tightClust (tightClust-package), 1
tightClust-package, 1

write.tight.clust, 5