

Package ‘boclust’

December 4, 2017

Title A Clustering Method Based on Boosting on Single Attributes

Version 0.1.2

Description An overlap clustering algorithm for categorical ultra-dimension data.

Depends R (>= 3.3)

License GPL-2

LazyData true

RoxygenNote 6.0.1

biocViews genefilter

Imports reshape2 (>= 1.4.2), psych (>= 1.7.3.21), Rtsne (>= 0.13),
d3heatmap (>= 0.6.1.1), venn (>= 1.2), DT (>= 0.2), shiny (>=
1.0.5), ggplot2 (>= 2.2.1), genefilter

NeedsCompilation no

Author Peng Zhang [aut],
Kaixiu Jin [aut, cre]

Maintainer Kaixiu Jin <jinkaixiu@zju.edu.cn>

Repository CRAN

Date/Publication 2017-12-04 05:12:29 UTC

R topics documented:

bo.simu.data	2
BossaClust	2
BossaSimi	3
bossa_interactive	4
OverlapClust	5
Index	6

bo.simu.data	<i>A data of 300 cells and 200 genes.</i>
--------------	---

Description

A simulation data used to show the cluster result of sparse and high-dimension data. There are 6 true clusters.

Format

A dataframe with 300 rows and 200 variables:

a_1_1 one of the fictitious marker genes ...

BossaClust	<i>Bossa Clustering</i>
------------	-------------------------

Description

With the previous calculated similarity matrix or the original categorical dataframe, the results of both overlap clustering and hierarchical clustering are obtained with several recommended cluster numbers(k) after processing the merge cluster step.

Usage

```
BossaClust(data, data.pre = NULL, alpha = 1, p = c(0.9, 0.75, 0.5),
  lin = 0.25, is.pca = TRUE, pca.sum.prop = 0.95, n.comp = 50,
  fix.pca.comp = FALSE, cri = 1, lintype = "ward.D2", perplexity = 30)
```

Arguments

<code>data</code>	an original categorical data with n observations and p variables.
<code>data.pre</code>	an list obtained by BossaSimi including original categorical data, similarity matrix, dissimilarity matrix and transformed data, Bossa scores. It is recommended to calculate the <code>data.pre</code> first and then do BossaClust in order to save time when trying to change parameters of this function.
<code>alpha</code>	A power scaling for Bossa scores, representing the weight of variable sigma value.
<code>p</code>	A set of quantiles(90 similarity matrix to form clusters at different levels of within-cluster similarity.
<code>lin</code>	A tuning parameter to control the size of each overlap cluster before merging, smaller <code>lin</code> leads to larger cluster size.
<code>is.pca</code>	A logical variable indicating if the Bossa scores should transformed to principle components and then calculate the similarity matrix. It is recommended when processing the ultra-dimension data.

<code>pca.sum.prop</code>	A numeric indicating how many components should be reserved in order to make this proportion of variance. The default is <code>pca.sum.prop = 0.95</code> .
<code>n.comp</code>	The number of components of PCA. The default is <code>n.comp = 50</code> .
<code>fix.pca.comp</code>	A numeric variable indicating whether choosing the fixed number of components or the fixed proportion of variance and the default is to choose fixed proportion.
<code>cri</code>	A tuning parameter, if p value smaller than <code>cri</code> , then reject the NULL hypothesis and merge overlap sub-clusters. And <code>cri</code> can be any numeric less than 1, if <code>cri = 1</code> then the criteria will be reset to $0.05/N$ (N is the number of all overlap sub-clusters), and if <code>cri = 2</code> then the criteria $0.05/N(N-1)$.
<code>lintype</code>	The agglomeration method to be used in <code>hclust</code> . This should be (an unambiguous abbreviation of) one of "ward.D", "ward.D2", "single", "complete", "average" and so on. The default is "ward.D2".
<code>perplexity</code>	A parameter of <code>tsne</code>

Value

An object including overlap clusters after merging and non-overlap clusters, which can be showed by function `bossa_interactive`

Examples

```
{
data(bo.simu.data)
object <- BossaClust(bo.simu.data)
}
```

Bossasimi

*Bossa Similarity***Description**

Calculate the similarity matrix of Bossa scores which are obtained by boosting on single attribute.

Usage

```
BossaSimi(data, is.pca = TRUE, pca.sum.prop = 0.95, fix.pca.comp = FALSE,
n.comp = 50, alpha = 1, pro.show = FALSE)
```

Arguments

<code>data</code>	A <code>data.frame</code> or <code>matrix(n*p)</code> of original categorical data.
<code>is.pca</code>	A logical variable indicating if the Bossa scores should transformed to principle components and then calculate the similarity matrix. It is recommended when processing the ultra-dimension data.

pca.sum.prop	A numeric indicating how many components should be reserved in order to make this proportion of variance. The default is <code>pca.sum.prop = 0.95</code> .
fix.pca.comp	A numeric variable indicating whether choosing the fixed number of components or the fixed proportion of variance and the default is to choose fixed proportion.
n.comp	The number of components of PCA. The default is <code>n.comp = 50</code> .
alpha	A power scaling for Bossa scores, representing the weight of variable sigma value.
pro.show	A logical indicator whether show the details of the process.

Value

An object including Bossa scores, Bossa dissimilarity and Bossa similarity(for [OverlapClust](#)).

Examples

```
{
## generate sparse data from the toy model of CIDR
sparse.data <- data.frame(g.1 = c(0, 5, 0, 6, 8, 6, 7, 7), g.2 = c(5, 0, 0, 0, 5, 7, 5, 7))

## with low-dimensional data, pca is unnecessary
bossa.change <- BossaSimi(sparse.data, is.pca = FALSE)

## data after normalization
data.after <- bossa.change$U.score.non.pca

## similarity matrix of normalized data
data.simi <- bossa.change$bossa.simi
}
```

`bossa_interactive` *Opens BOSSA results in an interactive session in a web browser.*

Description

Runs interactive shiny session of BOSSA based on pre-computed clusterings.

Usage

```
bossa_interactive(object)
```

Arguments

`object` a output of `BossaClust` function

Value

Opens a browser window with an interactive shiny app and visualize all pre-computed clusterings.

Examples

```
{
  data(bo.simu.data)
  object <- BossaClust(bo.simu.data)
  if(interactive()){bossa_interactive(object)}
}
```

OverlapClust

Overlap Clustering

Description

Do overlap clustering with Bossa similarity in different levels of p.

Usage

```
OverlapClust(data.simi, p = c(0.9, 0.75, 0.5), lin = 0.25,
  pro.show = FALSE)
```

Arguments

data.simi	The similarity matrix of Bossa scores obtained by BossaSimi
p	A set of quantiles(90 similarity matrix to form clusters at different levels of within-cluster similarity.
lin	A tuning parameter to control the size of each overlap cluster before merging, smaller lin leads to larger cluster size.
pro.show	A logical indicator whether show the details of the process.

Value

A list including two data.frame: overlap sub-clusters and cluster center for each.

Examples

```
{
  data(bo.simu.data)

  ## calculate the similarity matrix
  bossa.simi <- BossaSimi(bo.simu.data)$bossa.simi

  overlap.clust <- OverlapClust(bossa.simi)
}
```

Index

*Topic **datasets**

bo.simu.data, [2](#)

bo.simu.data, [2](#)

bossa_interactive, [3](#), [4](#)

BossaClust, [2](#), [2](#)

BossaSimi, [2](#), [3](#), [5](#)

hclust, [3](#)

OverlapClust, [4](#), [5](#)