Package ‘dynamicTreeCut’

March 13, 2008

Version 1.11-3

Date 2007-03-12

Title Dynamic Tree Cut Methods for detection of clusters in hierarchical clustering dendrograms.

Author Peter Langfelder <Peter.Langfelder@gmail.com> and Bin Zhang <binzhang.ucla@gmail.com>, with contributions from Steve Horvath <SHorvath@mednet.ucla.edu>

Maintainer Peter Langfelder <Peter.Langfelder@gmail.com>

Depends R, stats

ZipData no

License GPL version 2 or newer

Description Contains methods for detection of clusters in hierarchical clustering dendrograms.

Title Dynamic Tree Cut Methods for detection of clusters in hierarchical clustering dendrograms.

URL http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/BranchCutting/

R topics documented:

cutreeDynamic ............................................................ 2
cutreeDynamicTree .................................................... 4
cutreeHybrid ............................................................. 5
dynamicTreeCut-package ............................................. 7
indentSpaces .............................................................. 7
printFlush ................................................................. 8

Index 10
**cutreeDynamic**  
Adaptive branch pruning of hierarchical clustering dendrograms.

**Description**

This wrapper provides a common access point for two methods of adaptive branch pruning of hierarchical clustering dendrograms.

**Usage**

```r
cutreeDynamic(dendro, cutHeight = NULL, minClusterSize = 20, method = "hybrid", distM = NULL,
deepSplit =@ifelse(method=="hybrid", 1, FALSE),
maxCoreScatter = NULL, minGap = NULL,
maxAbsCoreScatter = NULL, minAbsGap = NULL, clusterTrim = 0,
labelUnlabeled = TRUE,
useMedoids = FALSE, maxDistToLabel = cutHeight,
respectSmallClusters = TRUE,
verbose = 2, indent = 0)
```

**Arguments**

dendo  
A hierarchical clustering dendrogram such as one returned by `hclust`.

cutHeight  
Maximum joining heights that will be considered. For `method="tree"` it defaults to 0.99. For `method="hybrid"` it defaults to 99 percentile and the maximum of the joining heights on the dendrogram.

minClusterSize  
Minimum cluster size.

method  
Chooses the method to use. Recognized values are "hybrid" and "tree".

distM  
Only used for method "hybrid". The distance matrix used as input to `hclust`. If not given and method == "hybrid", the function will issue a warning and default to method = "tree".

deepSplit  
For method "hybrid", can be either logical or integer in the range 0 to 3. For method "tree", must be logical. In both cases, provides a rough control over sensitivity to cluster splitting. The higher the value (or if TRUE), the more and smaller clusters will be produced. For the "hybrid" method, a finer control can be achieved via `maxCoreScatter` and `minGap` below.

maxCoreScatter  
Only used for method "hybrid". Maximum scatter of the core for a branch to be a cluster, given as the fraction of `cutHeight` relative to the 5th percentile of joining heights. See Details.

minGap  
Only used for method "hybrid". Minimum cluster gap given as the fraction of the difference between `cutHeight` and the 5th percentile of joining heights.

maxAbsCoreScatter  
Only used for method "hybrid". Maximum scatter of the core for a branch to be a cluster given as absolute heights. If given, overrides `maxCoreScatter`.

minAbsGap  
Only used for method "hybrid". Minimum cluster gap given as absolute height difference. If given, overrides `minGap`. 

clustTrim Only used for method "hybrid". Specifies the fraction of the cluster height that will be trimmed before stage 2; all objects joining above the $1 - \text{ClusterTrim}$ fraction of the cluster height will be trimmed. If set to 0 or less, nothing will be trimmed.

labelUnlabeled Only used for method "hybrid". if TRUE, the second (PAM-like) stage will be performed.

useMedoids Only used for method "hybrid" and only if labelUnlabeled==TRUE. If TRUE, the second stage will be used object to medoid distance; if FALSE, it will use average object to cluster distance. The default (FALSE) is recommended.

maxDistToLabel Only used for method "hybrid" and only if labelUnlabeled==TRUE. Maximum object distance to closest cluster that will result in the object assigned to that cluster.

respectSmallClusters Only used for method "hybrid" and only if labelUnlabeled==TRUE. If TRUE, branches that failed to be clusters in stage 1 only because of insufficient size will be assigned together in stage 2. If FALSE, all objects will be assigned individually.

verbose Controls the verbosity of the output. 0 will make the function completely quiet, values up to 4 gradually increase verbosity.

indent Controls indentation of printed messages (see verbose above). Each unit adds two spaces before printed messages; useful when several functions’ output is to be nested.

Details

This is a wrapper for two related but different methods for cluster detection in hierarchical clustering dendrograms.

In order to make the shape parameters maxCoreScatter and minGap more universal, their values are interpreted relative to cutHeight and the 5th percentile of the merging heights (we arbitrarily chose the 5th percentile rather than the minimum for reasons of stability). Thus, the absolute maximum allowable core scatter is calculated as $\text{maxCoreScatter} \times (\text{cutHeight} - \text{refHeight}) + \text{refHeight}$ and the absolute minimum allowable gap as $\text{minGap} \times (\text{cutHeight} - \text{refHeight})$, where refHeight is the 5th percentile of the merging heights.

Value

A vector of numerical labels giving assignment of objects to modules. Unassigned objects are labeled 0, the largest module has label 1, next largest 2 etc.

Author(s)

Peter Langfelder, \{Peter.Langfelder@gmail.com\}

References


See Also

\url{hclust, cutreeHybrid, cutreeDynamicTree}.
Description

Detect clusters in a hierarchical dendrogram using a variable cut height approach. Uses only the information in the dendrogram itself is used (which may give incorrect assignment for outlying objects).

Usage

cutreeDynamicTree(dendro, maxTreeHeight = 1, deepSplit = TRUE, minModuleSize = 50)

Arguments

dendro  
Hierarchical clustering dendrogram such produced by `hclust`.

maxTreeHeight  
Maximum joining height of objects to be considered part of clusters.

deepSplit  
If TRUE, method will favor sensitivity and produce more smaller clusters. When FALSE, there will be fewer bigger clusters.

minModuleSize  
Minimum module size. Branches containing fewer than minModuleSize objects will be left unlabeled.

Details

A variable height branch pruning technique for dendrograms produced by hierarchical clustering. Initially, branches are cut off at the height maxTreeHeight; the resulting clusters are then examined for substructure and if subclusters are detected, they are assigned separate labels. Subclusters are detected by structure and are required to have a minimum of minModuleSize objects on them to be assigned a separate label. A rough degree of control over what it means to be a subcluster is implemented by the parameter `deepSplit`.

Value

A vector of numerical labels giving assignment of objects to modules. Unassigned objects are labeled 0, the largest module has label 1, next largest 2 etc.

Author(s)

Bin Zhang, ⟨binzhang.ucla@gmail.com⟩, with contributions by Peter Langfelder, ⟨Peter.Langfelder@gmail.com⟩.

References

http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/BranchCutting

See Also

cutreeHybrid
cutreeHybrid  
Hybrid adaptive tree cut for hierarchical clustering dendrograms.

Description

Detect clusters in a dendrogram produced by the function hclust.

Usage

cutreeHybrid(dendro, distM, cutHeight = NULL, minClusterSize = 20,  
depthSplit = 1,  
maxCoreScatter = NULL, minGap = NULL,  
maxAbsCoreScatter = NULL, minAbsGap = NULL, clusterTrim = 0,  
labelUnlabeled = TRUE,  
useMedoids = FALSE, maxDistToLabel = cutHeight,  
respectSmallClusters = TRUE, verbose = 2, indent = 0)

Arguments

dendro  
a hierarchical clustering dendrogram such as one returned by hclust.
distM  
Distance matrix that was used as input to hclust.
cutHeight  
Maximum joining heights that will be considered. It defaults to 99 of the range between the 5th percentile and the maximum of the joining heights on the dendrogram.
minClusterSize  
Minimum cluster size.
depthSplit  
Either logical or integer in the range 0 to 3. Provides a rough control over sen-
sitivity to cluster splitting. The higher the value, the more and smaller clusters will be produced. A finer control can be achieved via maxCoreScatter and minGap below.
maxCoreScatter  
Maximum scatter of the core for a branch to be a cluster, given as the fraction of cutHeight relative to the 5th percentile of joining heights. See Details.
minGap  
Minimum cluster gap given as the fraction of the difference between cutHeight and the 5th percentile of joining heights.
maxAbsCoreScatter  
Maximum scatter of the core for a branch to be a cluster given as absolute heights. If given, overrides maxCoreScatter.
minAbsGap  
Minimum cluster gap given as absolute height difference. If given, overrides minGap.
clusterTrim  
Specifies the fraction of the cluster height that will be trimmed before stage 2; all objects joining above the 1–ClusterTrim fraction of the cluster height will be trimmed. If set to 0 or less, nothing will be trimmed.
labelUnlabeled  
If TRUE, the second (PAM-like) stage will be performed.
useMedoids  
if TRUE, the second stage will be use object to medoid distance; if FALSE, it will use average object to cluster distance. The default (FALSE) is recommended.
maxDistToLabel
Maximum object distance to closest cluster that will result in the object assigned to that cluster.

respectSmallClusters
If TRUE, branches that failed to be clusters in stage 1 only because of insufficient size will be assigned together in stage 2. If FALSE, all objects will be assigned individually.

verbose
Controls the verbosity of the output. 0 will make the function completely quiet, values up to 4 gradually increase verbosity.

indent
Controls indentation of printed messages (see verbose above). Each unit adds two spaces before printed messages; useful when several functions’ output is to be nested.

Details
The function detects clusters in a hierarchical dendrogram based on the shape of branches on the dendrogram. For details on the method, see http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/BranchCutting.

In order to make the shape parameters maxCoreScatter and minGap more universal, their values are interpreted relative to cutHeight and the 5th percentile of the merging heights (we arbitrarily chose the 5th percentile rather than the minimum for reasons of stability). Thus, the absolute maximum allowable core scatter is calculated as maxCoreScatter * (cutHeight - refHeight) + refHeight and the absolute minimum allowable gap as minGap * (cutHeight - refHeight), where refHeight is the 5th percentile of the merging heights.

Value
A list containing the following elements:

labels
Numerical labels of clusters, with 0 meaning unassigned, label 1 labeling the largest cluster etc.

cores
Numerical labels indicating cores of found clusters.

smallLabels
Numerical labels for branches that failed to be recognized clusters only because of insufficient number of objects.

trimmed
Numerical labels indicating objects that have been trimmed from clusters.

branches
A list detailing the detected branch structure.

Author(s)
Peter Langfelder, (Peter.Langfelder@gmail.com)

References

See Also
hclust, as.dist
### dynamicTreeCut-package

*Methods for detection of clusters in hierarchical clustering dendrograms.*

### Description

Contains methods for detection of clusters in hierarchical clustering dendrograms.

### Details

<table>
<thead>
<tr>
<th>Package</th>
<th>dynamicTreeCut</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td>1.11-3</td>
</tr>
<tr>
<td>Date</td>
<td>2007-03-12</td>
</tr>
<tr>
<td>Depends</td>
<td>R, stats</td>
</tr>
<tr>
<td>ZipData</td>
<td>no</td>
</tr>
<tr>
<td>License</td>
<td>GPL version 2 or newer</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/BranchCutting/">http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/BranchCutting/</a></td>
</tr>
<tr>
<td>Packaged</td>
<td>Wed Mar 12 12:32:38 2008; plangfelder</td>
</tr>
<tr>
<td>Built</td>
<td>R 2.4.1; ; 2008-03-12 12:48:47; unix</td>
</tr>
</tbody>
</table>

### Index:

- cutreeDynamic: Adaptive branch pruning of hierarchical clustering dendrograms.
- cutreeDynamicTree: Dynamic dendrogram pruning based on dendrogram only.
- cutreeHybrid: Hybrid adaptive tree cut for hierarchical clustering dendrograms.
- indentSpaces: Spaces for indented output.
- printFlush: Print arguments and flush the console.
- treecut-package: Methods for detection of clusters in hierarchical clustering dendrograms.

### Author(s)

Peter Langfelder <Peter.Langfelder@gmail.com> and Bin Zhang <binzhang.ucla@gmail.com>, with contributions from Steve Horvath <SHorvath@mednet.ucla.edu>

Maintainer: Peter Langfelder <Peter.Langfelder@gmail.com>

### Description

Returns a character string containing two times `indent` spaces.
Usage

indentSpaces(indent = 0)

Arguments

indent Desired level of indentation. The number of returned spaces will be twice this argument.

Value

A character string containing spaces, of length twice indent.

Author(s)

Peter Langfelder, ⟨Peter.Langfelder@gmail.com⟩

Examples

spaces = indentSpaces(0);
print(paste(spaces, "This output is not indented..."));
spaces = indentSpaces(1);
print(paste(spaces, "...while this one is.")

printFlush  

Print arguments and flush the console.

Description

Passes all its arguments unchanged to the standard print function; after the execution of print it flushes the console, if possible.

Usage

printFlush(...) 

Arguments

... Arguments to be passed to the standard print function.

Details

Passes all its arguments unchanged to the standard print function; after the execution of print it flushes the console, if possible.

Value

Returns the value of the print function.

Author(s)

Peter Langfelder, ⟨Peter.Langfelder@gmail.com⟩
printFlush

See Also

print
Index

*Topic **cluster**
  *cutreeDynamic, 1*
  *cutreeHybrid, 5*
*Topic **misc**
  *cutreeDynamicTree, 4*
*Topic **package**
  *dynamicTreeCut-package, 7*
*Topic **print**
  *indentSpaces, 7*
  *printFlush, 8*

**as.dist, 6**

cutreeDynamic, 1
cutreeDynamicTree, 3, 4
cutreeHybrid, 3, 4, 5

dynamicTreeCut
  *(dynamicTreeCut-package, 7)*
dynamicTreeCut-package, 7

**hclust, 3, 4, 6**

**indentSpaces, 7**

**print, 8, 9**
**printFlush, 8**