Package 'dendsort'

Title Modular Leaf Ordering Methods for Dendrogram Nodes

Version 0.3.4

Description An implementation of functions to optimize ordering of nodes in a dendrogram, without affecting the meaning of the dendrogram. A dendrogram can be sorted based on the average distance of subtrees, or based on the smallest distance value. These sorting methods improve readability and interpretability of tree structure, especially for tasks such as comparison of different distance measures or linkage types and identification of tight clusters and outliers. As a result, it also introduces more meaningful reordering for a coupled heatmap visualization. This method is described in ``dendsort: modular leaf ordering methods for dendrogram representations in R'', F1000Research 2014, 3: 177 <doi:10.12688/f1000research.4784.1>.

License GPL-2 | GPL-3

Encoding UTF-8

Suggests knitr, RColorBrewer, gplots, seriation, gapmap, rmarkdown

VignetteBuilder knitr

URL https://github.com/evanbiederstedt/dendsort

BugReports https://github.com/evanbiederstedt/dendsort/issues

RoxygenNote 7.1.1

NeedsCompilation no

Maintainer Evan Biederstedt <evan.biederstedt@gmail.com>

Author Ryo Sakai [aut],
Evan Biederstedt [cre, aut]

Repository CRAN

Date/Publication 2021-04-20 11:40:02 UTC

R topics documented:

---

dendsort-package   *Modular Leaf Ordering Methods for Dendrogram Nodes*

Description

Modular Leaf Ordering Methods for Dendrogram Nodes

Details

This package includes functions to optimize ordering of nodes in a dendrogram, without affecting the meaning of the dendrogram. A dendrogram can be sorted based on the average distance of subtrees, or based on the smallest distance value. These sorting methods improve readability and interpretability of tree structure, especially for tasks such as comparison of different distance measures or linkage types and identification of tight clusters and outliers. As a result, it also introduces more meaningful reordering for a coupled heatmap visualization.

Author(s)

Ryo Sakai <ryo.sakai@esat.kuleuven.be>

---

dendsort   *Sorting and reordering dendrogram nodes*

Description

dendsort sorts a dendrogram object which is typically a result of hierarchical clustering (hclust). The subtrees in the resulting dendrogram are sorted based on the average distance of subtrees at every merging point. The tighter cluster, in other words the cluster with smaller average distance, is placed on the left side of branch. When a leaf merge with a cluster, the leaf is placed on the right side.

Usage

```
dendsort(d, isReverse = FALSE, type = "min")
```

Arguments

d	a dendrogram or hclust object.d
isReverse	logical indicating if the order should be reversed.Defaults to FALSEisReverse
type	character indicating the type of sorting. Default to "min" type

Value

output A sorted dendrogram or hclust.

Examples

```
#generate sample data
set.seed(1234); par(mar=c(0,0,0,0))
x <- rnorm(10, mean=rep(1:5, each=2), sd=0.4)
y <- rnorm(10, mean=rep(c(1,2), each=5), sd=0.4)
dataFrame <- data.frame(x=x, y=y, row.names=c(1:10))
#calculate Euclidian distance
distxy <- dist(dataFrame)
#hierachical clustering "complete" linkage by default
hc <- hclust(distxy)

#sort dendrogram
dd <- dendsort(as.dendrogram(hc))
hc_sorted  <- as.hclust(dd)

#sort in reverse, you can also pass hclust object
plot(dendsort(hc, isReverse=TRUE))

#sort by average distance
plot(dendsort(hc, type="average"))

#plot the result
par(mfrow = c(1, 3), mai=c(0.8,0.8,2,0.8))
plot(x, y, col="gray", pch=19, cex=2)
text(x, y, labels=as.character(1:10), cex=0.9)
plot(hc,main="before sorting", xlab="", sub="")
plot(hc_sorted, main="after sorting", xlab="", sub="")
```

---

sample_tcga	*Sample data matrix from the integrated pathway analysis of gastric cancer from the Cancer Genome Atlas (TCGA) study*

Description

a multivariate table obtained from the integrated pathway analysis of gastric cancer from the Cancer Genome Atlas (TCGA) study. In this data set, each column represents a pathway consisting of a set of genes and each row represents a cohort of samples based on specific clinical or genetic features. For each pair of a pathway and a feature, a continuous value of between 1 and -1 is assigned to score positive or negative association, respectively.

Usage

```
data(sample_tcga)
```

Format

A data frame with 215 rows and 117 variables

Details

We would like to thank Sheila Reynolds and Vesteinn Thorsson from the Institute for Systems Biology for sharing this sample data set.

Index