Package ‘dendsort’

December 14, 2015

Type Package
Title Modular Leaf Ordering Methods for Dendrogram Nodes
Version 0.3.3
Date 2015-12-14
Author Ryo Sakai
Maintainer Ryo Sakai <ryo@vda-lab.be>
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.
License GPL-2 | GPL-3
URL https://bitbucket.org/vda-lab/dendsort/,
    http://f1000research.com/articles/3-177/v1
Suggests knitr, RColorBrewer, gplots, heatmap.plus, seriation, gapmap
VignetteBuilder knitr
RoxygenNote 5.0.1
NeedsCompilation no
Repository CRAN
Date/Publication 2015-12-14 12:39:17

R topics documented:

dendsort-package . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2
dendsort . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2
sample_tcg . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

Index 5
Modular Leaf Ordering Methods for Dendrogram Nodes

Description

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.
isReverse logical indicating if the order should be reversed. Defaults to FALSE.
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(0.8,0.8,0.8,0.8))
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 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

*Topic datasets
  sample_tcga, 3

*Topic dendrogram
  dendsort, 2

*Topic package
  dendsort-package, 2

dendsort, 2
dendsort-package, 2

sample_tcga, 3