

Package ‘MDSMClust’

June 15, 2021

Title Multidimensional Scaling for Microbiome Clustering

Version 0.1.0

Date 2021-03

Description Tidying up the code.

License CC0

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.1

VignetteBuilder knitr

Depends R (>= 2.10)

Imports SpiecEasi (>= 1.1.0),
BiocManager (>= 1.30.10),
philr (>= 1.14.0),
fpc (>= 2.2.8),
MiSPU (>= 1.0),
GUniFrac (>= 1.1),
miProfile (== 1.0),
dirmult (>= 0.1.3.4),
cluster (>= 2.1.0),
clues (>= 0.6.2.2),
DirichletMultinomial (>= 1.30.0),
MASS (>= 7.3.53),
vegan (>= 2.5.6),
reshape2 (>= 1.4.4)

Suggests knitr, rmarkdown

R topics documented:

DMM	2
DM_throat_info	3
LN_throat_info	3
Martinez	4
MDS	4
Smits	5
Index	7

DMM

Fit Dirichlet Multinomial Models to count Data and Evaluate the Result with Adjusted ARI

Description

This function fits Dirichlet Multinomial Models (DMM) to a count matrix and then calculate agreement indices between any two partitions for the data set.

Usage

```
DMM(otu.count, k, group, randMethod = "MA",
    seed = 2021, trim = FALSE, threshold)
```

Arguments

otu.count	OTUs with count numbers or a data matrix. The samples are the rows.
k	an integer or a numeric vector. Specify the numbers of Dirichlet components to fit.
group	a numeric vector of class labels. The length of group should be the same as the number of samples.
randMethod	specifies the preferred external index to measure the agreement between the clustering result and the given group. Available indices are: "Rand", "HA" (Hubert and Arabie's adjusted Rand index), "MA" (Morey and Agresti's adjusted Rand index), "FM" (Fowlkes and Mallows's index), "Jaccard" (Jaccard index). By default, "MA" will be output.
seed	set.seed with the given number. By default, set.seed(2021).
trim	logical. By default, it's FALSE. If TRUE, please indicate the threshold.
threshold	trim the matrix with <code>otu.count[, which(colSums(otu.count) > threshold)]</code> .

Value

A list of 3 containing k (the numbers of Dirichlet components), ARI (the corresponding adjusted Rand Indices) and cluster (the clustering results).

Examples

```
data(Martinez)
tree = Martinez$tree
otu.table = Martinez$otutable
otu.count = t(otu.table)
DMM(otu.count, group = as.numeric(as.factor(Martinez$sampleinfo)),
    k = c(2,3,5), trim = T, threshold = 5)
```

DM_throat_info*The sample template data for simulation - Dirichlet Multinomial Model*

Description

The data before it's simulated with DM model in three scenarios: common lineage, rare lineage, random OTUs.

Usage

```
data(DM_throat_info)
```

Format

List of 4.

Examples

```
data(DM_throat_info)
```

LN_throat_info*The sample template data for simulation - Logit-normal Multinomial*

Description

The data before it's simulated with LN model in three scenarios: common lineage, rare lineage, random OTUs.

Usage

```
data(LN_throat_info)
```

Format

List of 4.

Examples

```
data(LN_throat_info)
```

Martinez

*Martinez Data***Description**

Martinez data is a data set from the R package "MicrobiomeCluster" including OTU table, tree and sample geographic information.

Usage

```
data(Martinez)
```

Format

A large list with 3 components which are tree, sampleinfo and otutable.

References

Martinez, I., Stegen, J. C., Maldonado-Gomez, M. X., Eren, A. M., Siba, P. M., Green-hill, A. R., and Walter, J. (2015). The gut microbiota of rural Papua New Guineans: composition, diversity patterns, and ecological processes. *Cell reports*, 11(4):527-538.

Examples

```
data(Martinez)
table(Martinez$sampleinfo)
```

MDS

*Cluster OTU Data and Evaluate the Result with Adjusted ARI***Description**

This function can conduct classical multidimensional scaling (MDS) with seven different distances and then calculate agreement indices between any two partitions for a data set.

Usage

```
MDS(otu.count, distance, tree, mdsrank, k, est, threshold = 0.9,
    K.max, group, randMethod = "MA", seed = 2021)
```

Arguments

otu.count	OTUs with count numbers or a data matrix. The samples are the rows.
distance	character string specifying the distance to be used for calculating dissimilarities between observations. The currently available options are "bray", "jaccard", "unweighted UniFrac", "weighted UniFrac", "generalized UniFrac", "PhILR" and "clr" which are corresponding to Bray-Curtis distance, Jaccard Index, unweighted, weighted, generalized UniFrac, Euclidean after Phylogenetic ILR (PhILR) transformation and center logratio (CLR) transformation.

tree	specifies the corresponding tree if choose "unweighted UniFrac", "weighted UniFrac", "generalized UniFrac" or "PhILR" as the distance.
mdsrnk	the maximum dimension of the space which the data are to be represented in when conduct the MDS; must be less than or equal to n-1.
k	true number of the clusters. If not specified, please choose one of the three methods using parameter "est" to estimate the number.
est	specifies the preferred methods to estimate the number of clusters. Available choices are "gap" (Gap statistic), "ps" (Prediction Strength), "si" (Silhouette Index).
threshold	the threshold for prediction strength method mentioned in "est" taking value from 0 to 1. By default, it's 0.9. In practice, 0.8 or a higher one is preferred.
K.max	integer. Specify the maximum number of clusters if use some methods to estimate k.
group	a numeric vector of class labels. The length of group should be the same as the number of samples.
randMethod	specifies the preferred external index to measure the agreement between the clustering result and the given group. Available indices are: "Rand", "HA" (Hubert and Arabie's adjusted Rand index), "MA" (Morey and Agresti's adjusted Rand index), "FM" (Fowlkes and Mallows's index), "Jaccard" (Jaccard index). By default, "MA" will be output.
seed	set.seed with the given number. By default, set.seed(2021).

Value

A list of 3 containing bestk (the best number of clusters), ARI (the corresponding adjusted Rand Index) and cluster (the clustering results).

Examples

```
data(Martinez)
tree = Martinez$tree
otu.table = Martinez$otutable
otu.count = t(otu.table)
MDS(otu.count, distance = "bray", mdsrnk = 5, est = "gap", K.max = 16,
group = as.numeric(as.factor(Martinez$sampleinfo)), randMethod = "MA")
```

Smits

Smits Data

Description

Smits data is a data set from the R package "MicrobiomeCluster" including OTU table, tree and sample seasonal information. The samples are from the early wet season and the late dry season.

Usage

```
data(Smits)
```

Format

A large list with 3 components which are tree, sampleinfo and otutable.

References

Smits, S. A., Leach, J., Sonnenburg, E. D., Gonzalez, C. G., Lichtman, J. S., Reid, G., Knight, R., Manjurano, A., Chagalucha, J., Elias, J. E., Dominguez-Bello, M. G., and Sonnenburg, J. L. (2017). Seasonal cycling in the gut microbiome of the hadza hunter- gatherers of Tanzania. *Science*, 357(6353):802-806.

Examples

```
data(Smits)
table(Smits$sampleinfo)
```

Index

- * **ARI**

- DMM, [2](#)

- MDS, [4](#)

- * **DMM,**

- DMM, [2](#)

- * **MDS,**

- MDS, [4](#)

- * **datasets**

- DM_throat_info, [3](#)

- LN_throat_info, [3](#)

- Martinez, [4](#)

- Smits, [5](#)

DM_throat_info, [3](#)

DMM, [2](#)

LN_throat_info, [3](#)

Martinez, [4](#)

MDS, [4](#)

Smits, [5](#)