Package 'PROPS'

October 18, 2022

Type Package

Title PRObabilistic Pathway Score (PROPS)

Version 1.18.0

Date 2017-10-23

Author Lichy Han

Maintainer Lichy Han <lhan2@stanford.edu>

Description This package calculates probabilistic pathway scores using gene expression data. Gene expression values are aggregated into pathway-based scores using Bayesian network representations of biological pathways.

License GPL-2

NeedsCompilation no

Imports bnlearn, reshape2, sva, stats, utils, Biobase

Suggests knitr, rmarkdown

VignetteBuilder knitr

biocViews Classification, Bayesian, GeneExpression

git_url https://git.bioconductor.org/packages/PROPS

git_branch RELEASE_3_15

git_last_commit 0829b35

git_last_commit_date 2022-04-26

Date/Publication 2022-10-18

R topics documented:

PROPS-package *PRObabilistic Pathway Score (PROPS)*

Description

This package calculates probabilistic pathway scores using gene expression data. Gene expression values are aggregated into pathway-based scores using Bayesian network representations of biological pathways.

Details

The DESCRIPTION file: This package was not yet installed at build time.

Index: This package was not yet installed at build time.
Calculates PRObabilistic Pathway Scores (PROPS), which are pathway-based features, from gene-based data.

Author(s)

Lichy Han

Maintainer: Lichy Han <lhan2@stanford.edu>

References

Lichy Han, Mateusz Maciejewski, Christoph Brockel, William Gordon, Scott B. Snapper, Joshua R. Korzenik, Lovisa Afzelius, Russ B. Altman. A PRObabilistic Pathway Score (PROPS) for Classification with Applications to Inflammatory Bowel Disease.

Examples

```
#Load in randomly generated example data
#Each row is a sample
#Each column is a gene, named with Entrez Gene ID
data(example_healthy)
data(example_data)

#Run PROPS with default KEGG pathway edges
props_features <- props(example_healthy, example_data)

#Run PROPS with user input edges
data(example_edges)
props_features2 <- props(example_healthy, example_data, example_edges)
```

example_data *Example data, 50 samples, 22600 genes.*

Description

Example data, consisting of 50 samples, 22600 genes. Example data were randomly generated. Disease data such as this example data should be formatted either as a data frame or ExpressionSet. Data frames should have rows corresponding to patients and columns as genes with Entrez ID column names. ExpressionSet probes should be mapped to Entrez ID first before proceeding.

Usage

```
data("example_data")
```

example_edges *Example pathway edges. Contains 3 randomly generated pathways.*

Description

Example pathway edges. Contains 3 randomly generated pathways. User input edges should be a data frame with 3 columns, where columns 1 and 2 are the source (from) and sink (to) of the edge, and column 3 is the pathway ID or pathway name the edge belongs to. Gene IDs should be Entrez ID.

Usage

```
data("example_edges")
```

Format

A data frame with 300 observations on the following 3 variables.

`from` a character vector

`to` a character vector

`pathway_ID` a character vector

example_healthy *Example healthy data, 100 samples, 22600 genes.*

Description

Example healthy data, 100 samples, 22600 genes. Example healthy data were randomly generated. Healthy data such as this example healthy data should be formatted either as a data frame or ExpressionSet. Data frames should have rows corresponding to patients and columns as genes with Entrez ID column names. ExpressionSet probes should be mapped to Entrez ID first before proceeding.

Usage

```
data("example_healthy")
```

kegg_pathway_edges *KEGG pathway edges*

Description

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway edges, obtained using the KEGGgraph package and parsing the xml files available from KEGG. Pathway edge data are formatted as a data frame with 3 columns, where columns 1 and 2 are the source (from) and sink (to) of the edge, and column 3 is the pathway ID or pathway name the edge belongs to. KEGG is the default pathway database to be used for calculating probabilistic pathway scores. Users may instead choose to provide their own pathway edges if desired.

Usage

```
data("kegg_pathway_edges")
```

Format

A data frame with 80642 observations on the following 3 variables.

`props` *PRObabilistic Pathway Scores (PROPS)*

Description

Calculates PRObabilistic Pathway Scores (PROPS), which are pathway-based features, from gene-based data.

Usage

```
props(healthy_dat, dat, pathway_edges = NULL, batch_correct = FALSE, healthy_batches = NULL, dat_batche
```

Arguments

`healthy_dat`	Data frame or ExpressionSet of healthy data used to parameterize the pathways. Healthy data frame should be formatted as one sample per row, and columns should correspond to genes and be named by the gene Entrez ID. If using ExpressionSet, map row.names probes to corresponding Entrez IDs before proceeding.
`dat`	Data frame or ExpressionSet of patient data for which to calculate pathway-based scores. Data frame should be formatted as one sample per row, and columns should correspond to genes and be named by the gene Entrez ID. If using ExpressionSet, map row.names probes to corresponding Entrez IDs before proceeding.
`pathway_edges`	Optional user specified pathway edges. If no pathway edges are provided, KEGG pathways are used by default. User input edges should be a data frame with 3 columns, where columns 1 and 2 are the source (from) and sink (to) of the edge, and column 3 is the pathway ID or name the edge belongs to. Genes should be named by Entrez ID.
`batch_correct`	Optional flag to do batch correction using ComBat. If TRUE, then batch numbers must be provided.
`healthy_batches`	Batch covariate numbers, as a numeric vector, corresponding to the healthy data. For example, given 5 healthy patients from two batches, an example input would be (1, 2, 2, 1, 1).
`dat_batches`	Batch covariate numbers, as a numeric vector, corresponding to the patient data. For example, given 10 patients from three batches, an example input would be (1, 3, 2, 1, 3, 3, 2, 1, 2, 1). Batch "1" is the same for healthy data and patient data, indicating in this example there are 3 healthy samples and 4 patient samples from the same batch.

Value

Returns a data frame of pathway-based log-likelihood values, where each row corresponds to a pathway. The first two columns are the KEGG pathway ID and name, and the remaining columns correspond to each sample's pathway features.

Author(s)

Lichy Han

References

Lichy Han, Mateusz Maciejewski, Christoph Brockel, William Gordon, Scott B. Snapper, Joshua R. Korzenik, Lovisa Afzelius, Russ B. Altman. A PRObabilistic Pathway Score (PROPS) for Classification with Applications to Inflammatory Bowel Disease.

Examples

```
#Load in randomly generated example data
#Each row is a sample
#Each column is a gene, named with Entrez Gene ID
data(example_healthy)
data(example_data)

#Run PROPS with default KEGG pathway edges
props_features <- props(example_healthy, example_data)

#Run PROPS with user input edges
data(example_edges)
props_features2 <- props(example_healthy, example_data, example_edges)
```

Index